

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

Frank van der Meulen and Moritz Schauer,

*Delft Institute of Applied Mathematics (DIAM)*

*Delft University of Technology*

*Mekelweg 4*

*2628 CD Delft*

*The Netherlands*

*e-mail: [m.r.schauer@tudelft.nl](mailto:m.r.schauer@tudelft.nl)*

*e-mail: [f.h.vandermeulen@tudelft.nl](mailto:f.h.vandermeulen@tudelft.nl)*

**Abstract.** Bayesian 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. Data-augmentation has been proposed for obtaining draws from the posterior distribution of the parameters. Within this approach, the discrete time observations are augmented with diffusion bridges connecting these observations. This poses two challenges: (i) efficiently generating diffusion bridges; (ii) if unknown parameters appear in the diffusion coefficient, then direct implementation of data-augmentation results in an induced Markov chain which is reducible. In this paper we show how both challenges can be addressed in continuous time (before discretisation) by using *guided proposals*. These are Markov processes with dynamics described by the stochastic differential equation of the diffusion process with an additional term added to the drift coefficient to guide the process to hit the right end point of the bridge. The form of these proposals naturally provides a mapping that decouples the dependence between the diffusion coefficient and diffusion bridge using the driving Brownian motion of the proposals. As the guiding term has a singularity at the right end point, care is needed when discretisation is applied for implementation purposes. We show that this problem can be dealt with by appropriately time changing and scaling of the guided proposal process. In two examples we illustrate the performance of the algorithms we propose. The second of these concerns a diffusion approximation of a chemical reaction network with a four-dimensional diffusion driven by an eight-dimensional Brownian motion.

*Keywords:* Discretely observed diffusion process, multidimensional diffusion bridge; data augmentation; linear processes; innovation process; Chemical Langevin Equation; reparametrisation.

Primary 62M05, 60J60; secondary 62F15, 65C05.

## 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 et al. \(2001\)](#) and [Roberts and Stramer \(2001\)](#). This algorithm is a special form of the Metropolis-Hastings (MH) algorithm 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  $\theta$ . Since in both steps it is usually not possible to draw directly from the distribution of interest, each step can be replaced by generating a proposal which is then accepted with probability defined by the MH-algorithm. There are two challenges with this approach:

*Challenge 1: generating “good” proposals for the missing segments.* The problem of simulating diffusion bridges has received a lot of attention in the past. 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 et al. \(2001\)](#)), (iii) forward simulated processes derived from representations of the Brownian bridge in discrete time ([Durham and Gallant \(2002\)](#)), (iv) coupling arguments ([Bladt and Sørensen \(2012\)](#)), (v) a constrained sequential Monte Carlo algorithm with a resampling scheme guided by backward pilots [Lin et al. \(2010\)](#), and (vi) exact simulation ([Beskos et al. \(2006\)](#)).

[Delyon and Hu \(2006\)](#) extended the work of Durham and Gallant 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*. The proposals of this type have the drawback of a possible mismatch between drift and guiding term which prevents their use for routine practice. Instead a variant of the proposal without the target drift, also due to [Delyon and Hu \(2006\)](#), is used. When discretised, the latter is closely related to the Modified Brownian Bridge proposed in [Durham and Gallant \(2002\)](#). In [Schauer et al. \(2013\)](#) a general class of proposal processes for simulating diffusion bridges was introduced. The proposals in [Schauer et al. \(2013\)](#) 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 substantially reduce the mismatch of drift and guiding, 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 et al. \(2013\)](#). Further, this method only requires forward simulation and applies to diffusion processes which cannot be transformed to have unit diffusion coefficient as well.

This comes with another challenge, as in any implementation, ultimately all proposals have to be evaluated on a finite number of grid points. Now the pulling term added to the drift for guided proposals has a singularity near the endpoint. This renders direct Euler discretisation inaccurate. Furthermore, integrals that appear in the acceptance probability of bridges potentially suffer from this problem as well.

*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. From a second perspective, diffusion bridges corresponding to different  $\theta$  have mutually singular measures. For illuminating figures that illustrate this problem, we refer to [Roberts and Stramer \(2001\)](#). This difficulty can be overcome, when 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 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  $X^* = g(\theta, Z^*)$ . The process  $Z^*$  will be called the “innovation process”, as was done in related approaches (Cf. [Chib et al. \(2004\)](#) and [Golightly and Wilkinson \(2010\)](#)). Defining  $g$  and  $Z^*$  appropriately is rather subtle and we postpone a detailed discussion to Sections 2 and 4. To be successful, the tight dependence between  $\theta$  and  $X^*$  should be decoupled. In a more general set-up, decouplings of similar forms are discussed under the keyword *non-centered parameterisation* ([Papaspiliopoulos et al. \(2003\)](#)). It turns out that likelihood ratios between diffusion bridges with different parameters driven by the same innovations  $Z^*$  can be defined and used to update the diffusion parameter. In an elementary form, where  $\sigma_\theta(t, x) = \theta$  this technique was used by [Roberts and Stramer \(2001\)](#). [Chib et al. \(2004\)](#) and [Golightly and Wilkinson \(2010\)](#) extended this technique to general diffusions in a discretised setting. A construction for time homogeneous diffusions based on [Delyon and Hu \(2006\)](#) was explored by [Fuchs \(2013\)](#) (in particular section 7.4).

Papaspiliopoulos et al. (2013) have an approach where the missing data is initially considered in continuous time using Delyon and Hu (2006) bridge proposals, but the degeneracy problem is tackled after discretisation.

### 1.1. Contribution

In this article we show how the guided proposals introduced in Schauer et al. (2013) address the outlined challenges and can be used to define an efficient MCMC procedure for generating draws from the posterior distribution of  $\theta$  given discrete time data. The procedure can be seen as extension and unification of previous approaches within a continuous time framework. Specific features of our approach include:

- We use in each data augmentation step “adapted” bridge proposals which take 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. This example concerns a diffusion approximation of a chemical reaction network and it turns out that our methods are often well suited for estimation of diffusions obtained in this way.
- 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 innovation process is defined using the proposal process. As a result, in our algorithm (Cf. see algorithm 1), somewhat surprisingly, 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.
- 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 space-scaling of the proposal process that allows for numerically accurate discretisation and evaluation of the likelihood.

### 1.2. Outline

In Section 2 we clarify the aforementioned difficulties in a toy example. Here, we set some general notation, introduce some key ideas used throughout and further motivate the advantage of using guided proposals. In Section 3 we recap some key results from Schauer et al. (2013) on guided proposals. The core of this paper is Section 4 where we precisely state our algorithm. In Section 5 we relate our approach to earlier work by Golightly and Wilkinson (2010), Fuchs (2013) and Papaspiliopoulos et al. (2013).

In Section 6 we introduce a time-change and space-scaling of the proposal process to deal with the aforementioned numerical discretisation issues. Next, in Section 7 we discuss two examples. In the second example we estimate the parameters in a prokaryotic auto-regulation network example of Golightly and Wilkinson (2010). The appendices contain a few postponed proofs and some implementation details.

## 2. A toy problem

In this section we consider a toy example used 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 the main idea 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  $\pi(\theta \mid X_T)$ . 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

$$X_t^\circ = u(1 - t/T) + vt/T + \theta Z_t^\circ, \quad (2.1)$$

where  $Z^\circ$  is a standard Brownian bridge on  $[0, T]$  (i.e.  $Z^\circ(0) = 0$  and  $Z^\circ(T) = 0$ ). This is the bridge process corresponding to the unconditioned diffusion process  $d\tilde{X}_t = \theta dW_t$ ,  $X_0 = u$ .

Denoting the laws of  $X^*$  and  $X^\circ$  by  $\mathbb{P}_\theta^*$  and  $\mathbb{P}_\theta^\circ$  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,  $p$  and  $\tilde{p}$  denote the transition densities of the processes  $X$  and  $\tilde{X}$  respectively. 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

$$X_t^\circ = g(\theta, Z_t^\circ) \quad \text{for} \quad g(\theta, x) = u(1 - t/T) + vt/T + \theta x.$$

Define the process  $Z^*$  by the relation

$$X_t^* = g(\theta, Z_t^*). \quad (2.3)$$

This construction can be viewed by first choosing a process  $Z^\circ$  whose law does not depend on  $\theta$  (this is crucial, as we will shortly explain). Here, we have chosen  $Z^\circ$  to be a standard Brownian bridge on  $(0, T)$ . Second, setting  $X^\circ = g(\theta, Z^\circ)$  defines the mapping  $g$ . Third, equation (2.3) defined the process  $Z^*$ .

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  $Z^\circ$  by  $\mathbb{Q}_\theta^*$  and  $\mathbb{Q}^\circ$  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:

$$\begin{array}{c|cc|cc} \text{Process} & Z^* & \xrightarrow{g(\theta, \cdot)} & X^* & Z^\circ & \xrightarrow{g(\theta, \cdot)} & X^\circ \\ \hline \text{Measure} & \mathbb{Q}_\theta^* & & \mathbb{P}_\theta^* & \mathbb{Q}^\circ & & \mathbb{P}_\theta^\circ \end{array} \quad (2.4)$$

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

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

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

$$\begin{aligned} \frac{d\mathbb{Q}_{\theta^\circ}^*(Z^*)}{d\mathbb{Q}_\theta^*(Z^*)} &= \frac{d\mathbb{Q}_{\theta^\circ}^*(Z^*)}{d\mathbb{Q}^\circ(Z^*)} \bigg/ \frac{d\mathbb{Q}_\theta^*(Z^*)}{d\mathbb{Q}^\circ(Z^*)} \\ &= \frac{d\mathbb{P}_{\theta^\circ}^*(g(\theta^\circ, Z^*))}{d\mathbb{P}_{\theta^\circ}^\circ(g(\theta^\circ, Z^*))} \bigg/ \frac{d\mathbb{P}_\theta^*(g(\theta, Z^*))}{d\mathbb{P}_\theta^\circ(g(\theta, Z^*))} \end{aligned}$$

and thus results from absolute continuity of  $\mathbb{P}_\theta^*$  and  $\mathbb{P}_\theta^\circ$ . Being able to relate the likelihood ratio of  $\mathbb{Q}_{\theta^\circ}^*$  and  $\mathbb{Q}_\theta^*$  in this way is possible as  $Q^\circ$  does not depend on  $\theta$ . By equation (2.2), we now get

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

Substituting this expression into equation (2.5) yields

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

and all terms containing the unknown transition density cancel.

The 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 4.3 we give guidelines on appropriate choice of  $q$  if the drift is of a specific structure with respect to  $\theta$ . In the toy-example,  $X^\circ$  is defined in equation (2.1). This is a convenient choice, as it easily yields absolute continuity of  $\mathbb{P}_\theta^*$  and  $\mathbb{P}_\theta^\circ$ . Moreover, generating an exact realisation of  $X^\circ$  at any set of points in  $[0, T]$  is straightforward. Nevertheless, there are at least two disadvantages to this approach:

1. The proposal does not take the drift  $b$  into account. Depending on the precise form of  $b$ , this means that proposal bridges can look completely different from the bridge we wish to draw. Figures that illustrate this phenomenon in case  $dX_t = \sin(X_t - \theta) dt + \theta dW_t$  can be found in Lin et al. (2010) (figure 4). This results in low MH-acceptance probabilities.
2. If the dispersion coefficient takes the more general form  $\sigma_\theta(t, X_t)$ , then  $\mathbb{P}_\theta^*$  and  $\mathbb{P}_\theta^\circ$  are singular. Taking  $X^\circ$  to be the process defined by  $d\tilde{X}_t = \sigma_\theta(t, X_t^\circ) dW_t$ , conditioned on starting in  $u$  and ending in  $v$  at time  $T$ , may seem to solve this problem. Unfortunately this introduces another problem as for almost all choices of  $\sigma_\theta$  it will be impossible to generate realisations from  $X^\circ$  conditioned on hitting the endpoint  $v$  at time  $T$ . Moreover, the transition densities  $\tilde{p}$  will be completely intractable in this case.

### 3. Guided proposals

A flexible class of proposal processes was developed and studied in Schauer et al. (2013). We will use this framework throughout and provide a recap of the relevant results in this section. For precise statements of these results we refer the reader to Schauer et al. (2013).

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

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

where

$$b^*(t, x) = b(t, x) + a(t, x) \nabla_x \log p(t, x; T, v) \quad (**)$$

and  $a(t, x) = \sigma(t, x) \sigma'(t, x)$ . As the true transition density  $p(t, x; T, v)$  (this is the density of the process starting in  $x$  at time  $s$ , ending in  $v$  at time  $T$ ) of the diffusion process is generally intractable, we propose to replace it by the transition density of a diffusion process  $\tilde{X}$  for which it is known in closed form. Denote the transition density of  $\tilde{X}$  by  $\tilde{p}(s, x; T, v)$ . Define the process  $X^\circ$  as the solution of the SDE

$$dX_t^\circ = b^\circ(t, X_t^\circ) dt + \sigma(t, X_t^\circ) dW_t, \quad X_0^\circ = u, \quad (\circ)$$

where

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

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  $T$ ).

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

$$\tilde{R}(s, x) = \log \tilde{p}(s, x), \quad \tilde{r}(s, x) = \nabla \tilde{R}(s, x), \quad \tilde{H}(s, x) = -\Delta \tilde{R}(s, x), \quad (3.1)$$

where  $\nabla$  and  $\Delta$  denote gradient and Laplacian with respect to  $x$ . In [Schauer et al. \(2013\)](#) sufficient conditions for absolute continuity of  $\mu_{X^\circ}$  and  $\mu_{X^*}$  are established together with a closed form expression for the Radon-Nikodym derivative. It turns out that

$$\frac{d\mathbb{P}^*}{d\mathbb{P}^\circ}(X^\circ) = \frac{\tilde{p}(0, u; T, v)}{p(0, u; T, v)} \exp \left( \int_0^T D(s, X_s^\circ) ds \right), \quad (3.2)$$

where  $D$  is given by

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

and hence does not depend on any intractable objects. The class of linear processes,

$$d\tilde{X}_t = \tilde{B}(t)\tilde{X}_t dt + \tilde{\beta}(t) dt + \tilde{\sigma}(t) dW_t, \quad (3.3)$$

is a flexible class with known transition densities and its induced guided proposals satisfy the conditions for absolute continuity derived in [Schauer et al. \(2013\)](#) under weak conditions on  $\tilde{B}$ ,  $\tilde{\beta}$  and  $\tilde{\sigma}$ . Proposal processes  $X^\circ$  derived by choosing a linear process as in (3.3) 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  $\tilde{a}(T) = (\tilde{\sigma}\tilde{\sigma}^T)(T) = a(T, v)$ . Another requirement is existence of an  $\varepsilon > 0$  such that for all  $s \in [0, T]$   $x \in \mathbb{R}^d$  and  $y \in \mathbb{R}^d$  we have  $\min(y'a(s, x)y, y'\tilde{a}(s)y) \geq \varepsilon\|y\|^2$  (uniform ellipticity). Therefore, a simple type of guiding proposals is obtained upon choosing  $d\tilde{X}_t = \tilde{\beta} dt + \tilde{\sigma}(T, v) dW_t$ . For this particular choice

$$dX_t^\circ = \left( b(t, X_t^\circ) + a(t, X_t^\circ)a(T, v)^{-1} \left[ \frac{v - X_t^\circ}{T - t} - \tilde{\beta} \right] \right) dt + \sigma(t, X_t^\circ) dW_t, \quad X_0^\circ = u.$$

Depending on the precise form of  $b$  and  $\sigma$  it can nevertheless be advantageous to use guided proposals induced for nonzero  $\tilde{B}$ .

For later reference, we write

$$\Psi(X^\circ) = \exp \left( \int_0^T D(s, X_s^\circ) ds \right) \quad (3.4)$$

and  $\Psi_\theta(X^\circ)$  in the case where  $b$  or  $\sigma$  depend on a parameter  $\theta$ .



## 4. Proposed MCMC algorithm

### 4.1. Innovation process

Using the notation introduced in the previous section, the dynamics of the proposal process can be described by the stochastic differential equation

$$dX_t^\circ = (b_\theta(t, X_t^\circ) + a_\theta(t, X_t^\circ)\tilde{r}_\theta(t, X_t^\circ)) dt + \sigma_\theta(t, X_t^\circ) dW_t, \quad X_0^\circ = u. \quad (4.1)$$

Here  $\tilde{r}$  is determined by  $\tilde{\beta}$ ,  $\tilde{B}$  and  $\tilde{\sigma}$  appearing in equation (3.3). Although we conceptually follow the same approach as in Section 2, we now take  $Z^\circ$  to be the driving Brownian motion instead of a Brownian bridge and set  $Z^\circ = W$ . Define the mapping  $g$  by

$$X^\circ = g(\theta, Z^\circ). \quad (4.2)$$

Such a mapping exists if  $X^\circ$  is a strong solution to equation (4.1) (Cf. chapter V-10, p. 127 of [Rogers and Williams \(2000\)](#)). Sufficient conditions for existence of a strong solution are detailed in [Schauer et al. \(2013\)](#). Define the process  $Z^*$  by

$$Z_t^* = W_t - \int_0^t \sigma'_\theta(s, X_s^*) (r_\theta(t, X_t^*) - \tilde{r}_\theta(s, X_s^*)) ds \quad (4.3)$$

Now it's easily verified that

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

We refer to the process  $Z^*$  as the *innovation process* corresponding to  $X^\circ$  (by analogy of the terminology of [Golightly and Wilkinson \(2010\)](#) and [Chib et al. \(2004\)](#)). From equations (4.2) and (4.3) it follows that  $X^*$  is related to  $Z^*$  just like  $X^\circ$  is related to  $Z^\circ$ . Note however that while  $Z^*$  does not depend on  $\theta$ ,  $Z^\circ$  does. Denote the laws of  $X^\circ$ ,  $X^*$ ,  $Z^\circ$  and  $Z^*$  by  $\mathbb{P}_\theta^\circ$ ,  $\mathbb{P}_\theta^*$ ,  $\mathbb{Q}^\circ$ ,  $\mathbb{Q}_\theta^*$  respectively and recall the diagram in (2.4). We have

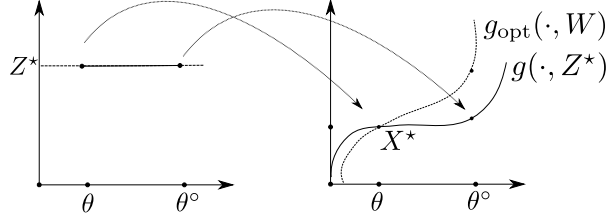
$$\frac{d\mathbb{Q}_\theta^*}{d\mathbb{Q}^\circ}(Z^*) = \frac{d\mathbb{P}_\theta^*}{d\mathbb{P}_\theta^\circ}(g(\theta, Z^*)) = \frac{\tilde{p}_\theta(0, u; T, v)}{p_\theta(0, u; T, v)} \Psi_\theta(g(\theta, Z^*)). \quad (4.4)$$

Here, the final equality follows from equation (3.2), which takes over the role of (2.2) in the toy-example.

If we had taken as different proposal process from (4.1) then the mapping  $g$  would change according to (4.2). At this point we want to stress that while different choices for the map  $g$  formally yield valid algorithms, updating  $\theta$  to  $\theta^\circ$  given  $Z^*$  is difficult if  $g(\theta^\circ, Z^*)$  does not resemble a draw from  $X^*$  conditional on  $\theta^\circ$ . Ideally, one would take  $g = g_{\text{opt}}$ , where  $g_{\text{opt}}$  is defined by the relation  $X^* = g_{\text{opt}}(\theta, W)$ . As this is not feasible, using  $g$  as derived from  $(\circ)$  instead of  $(\star)$  is a very reasonable choice, see figure 1.

### 4.2. 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


 FIGURE 1. Cartoon:  $g(\theta^\circ, Z^\star)$  should resemble a draw from  $X^\star$  conditional on  $\theta^\circ$ .

on  $\theta$  by  $\pi_0$ . The idea is to define a Metropolis–Hastings sampler on  $(\theta, Z^\star)$  instead of  $(\theta, X^\star)$  where  $Z^\star$  is the innovation process from the previous section. More precisely, we construct a Markov chain for  $(\theta, (Z_i^\star)_{1 \leq i \leq n})$ , where each  $Z_i^\star$  is an innovation process corresponding to the bridge  $X_i^\star$  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  $Z_i^\star = W_i$ .
2. **Update**  $Z^\star \mid (\theta, \mathcal{D})$ . 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

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

Set

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

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

$$A_2 = \prod_{i=1}^n \frac{\tilde{p}_{\theta^\circ}(t_{i-1}, x_{i-1}; t_i, x_i)}{\tilde{p}_\theta(t_{i-1}, x_{i-1}; t_i, x_i)} \frac{\Psi_{\theta^\circ}(g(\theta^\circ, Z_i^\star))}{\Psi_\theta(g(\theta, Z_i^\star))} \frac{q(\theta \mid \theta^\circ)}{q(\theta^\circ \mid \theta)} \frac{\pi_0(\theta^\circ)}{\pi_0(\theta)}$$

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  $Z^\star$  from  $X^\star$ . This is a consequence of adapting the definition of the innovations to the bridge proposals being used. It is essential, as  $Z^\star$  depends on  $r$  which is intractable (see equation (4.3)).

Step 2 constitutes a step of a MH-sampler with independent proposals. The expression for  $A_1$  follows directly from equation (3.2). 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.)

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

In this subsection we study the specific case for which

$$b_{\vartheta}(x) = \sum_{i=1}^N \vartheta_i \varphi_i(x) \quad (4.5)$$

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 Van der Meulen et al. (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 4.2.

**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

$$A_3 = \prod_{i=1}^n \frac{\tilde{p}_{(\gamma^{\circ}, \vartheta)}(t_{i-1}, x_{i-1}; t_i, x_i)}{\tilde{p}_{(\gamma, \vartheta)}(t_{i-1}, x_{i-1}; t_i, x_i)} \frac{\Psi_{(\gamma^{\circ}, \vartheta)}(g((\gamma^{\circ}, \vartheta), Z_i^*))}{\Psi_{(\gamma, \vartheta)}(g((\gamma, \vartheta), Z_i^*))} \frac{q(\gamma \mid \gamma^{\circ})}{q(\gamma^{\circ} \mid \gamma)} \frac{\pi_0(\gamma^{\circ})}{\pi_0(\gamma)}$$

Set

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

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

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

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

*Proof.* Suppose  $(\vartheta, \gamma, Z^\star) \sim \pi$ , where  $\pi$  denotes the posterior distribution. Consider the map  $f: (\vartheta, \gamma, Z^\star) \mapsto (\vartheta, \gamma, X^\star)$ , where  $X^\star = g((\vartheta, \gamma), Z^\star)$ . Here  $X^\star$  solves  $(\star)$  by (4.3). We show that step 3.2 preserves  $\pi$ . The distribution of  $(\vartheta, \gamma, X^\star)$  is the image measure of the posterior distribution  $\pi$  of the tuple  $(\vartheta, \gamma, Z^\star)$  under  $f$  and coincides with the posterior distribution of  $(\vartheta, \gamma, X^\star)$ . 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  $(\gamma, X^\star)$ . The latter preserves  $\pi \circ f^{-1}$ . Hence  $(\vartheta^\circ, \gamma, X^\star) \sim \pi \circ f^{-1}$ . In step 3.2(c) we compute  $(\vartheta^\circ, \gamma, Z^\circ)$  as pre-image of  $(\vartheta^\circ, \gamma, X^\star)$  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 (4.5) 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  $W_\vartheta^{-1}$ . 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**  $\vartheta \mid (\gamma, Z^\star, \mathcal{D})$ .

- (a) Set  $X^\star = g(\vartheta, Z^\star)$ .
- (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

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

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 average acceptance probability equal to 0.234.

## 5. Discussion on related work

In this section we point out similarities and differences of this paper with respect to related approaches in [Golightly and Wilkinson \(2006\)](#), [Golightly and Wilkinson \(2010\)](#), [Delyon and Hu \(2006\)](#), [Fuchs \(2013\)](#) and [Papaspiliopoulos et al. \(2013\)](#). [Delyon and Hu \(2006\)](#) introduced proposals with dynamics governed by the SDE

$$dX_t^\circ = \left( \lambda b(t, X_t^\circ) + \frac{v - X_t^\circ}{T - t} \right) dt + \sigma(t, X_t^\circ) dW_t, \quad X_0^\circ = u, \quad (\circ')$$

where  $\lambda \in \{0, 1\}$ . Sufficient conditions for absolute continuity for both cases are given in [Delyon and Hu \(2006\)](#). The pulling term  $(v - X_t^\circ)/(T - t)$  in the SDE forces  $X^\circ$  to hit  $v$  at time  $T$  and the case  $\lambda = 1$  takes the drift of the process into account when proposing bridges. However, for both  $\lambda = 0$  and  $\lambda = 1$  the resulting bridges may not resemble true bridges.

Without loss of generality we may assume we have one observation  $X_T = v$ . If  $\lambda = 0$  in  $(\circ')$ , a slight adjustment of the Euler discretisation of the bridge (introduced by [Durham and Gallant \(2002\)](#)) sets

$$X_{t_i}^\circ - X_{t_{i-1}}^\circ \approx \frac{v - X_{t_{i-1}}^\circ}{T - t_{i-1}} \delta_i + \frac{T - t_i}{T - t_{i-1}} \sigma(t_{i-1}, X_{t_{i-1}}^\circ) (W_{t_i} - W_{t_{i-1}}),$$

where  $0 = t_0 < t_1 < \dots < t_N = T$  and  $\delta_i = t_i - t_{i-1}$  (the adjustment being the addition of the term  $(T - t_i)/(T - t_{i-1})$  in front of  $\sigma$ ). If  $\delta := \max_{1 \leq i \leq N} \delta_i$  is small, the right hand side defines a discrete time Markov chain approximation  $\underline{X}_i^\circ$  to  $X_{t_i}^\circ$  termed the Modified Brownian Bridge. We can define a mapping  $g$  by the relation

$$(\Delta \underline{X}_1^\circ, \dots, \Delta \underline{X}_N^\circ) = g(\theta, \Delta W_1, \dots, \Delta W_N),$$

where  $\Delta X_i^\circ = X_i^\circ - X_{i-1}^\circ$  and  $\Delta W_i = W_{t_i} - W_{t_{i-1}}$ . This implies that parameters in the diffusion coefficient can be updated conditional on  $(\Delta W_1, \dots, \Delta W_N)$ , instead of  $(\Delta X_1^\circ, \dots, \Delta X_N^\circ)$ ; thereby preventing degeneracy of the data-augmentation algorithm. This is the approach of [Golightly and Wilkinson \(2010\)](#). [Chib et al. \(2004\)](#) use the same algorithm using Euler discretisation of  $X^\circ$  in case  $\lambda = 0$  (so no correction in the diffusion coefficient appears in the discretisation). Note that there is no formal proof that these methods work as  $\delta \downarrow 0$ .

[Fuchs \(2013\)](#) explores how to use the driving Brownian motion in  $(\circ')$  as innovation process within a continuous time setup. MH-acceptance probabilities are obtained using results in [Delyon and Hu \(2006\)](#) who provide expressions for the likelihood ratio of the laws of  $X^\star$  and  $X^\circ$ . However, the absolute continuity results of [Delyon and Hu \(2006\)](#) do not provide the proportionality constants in the derived likelihood ratio. For generating diffusion bridges using a MH-sampler these constants are irrelevant. However, as these constants do depend on  $\theta$  they cannot be neglected.

We shortly review the derivation of this constant in the one-dimensional case if  $\lambda = 0$  (the argument for  $\lambda = 1$  is similar). Let  $t < T$  and denote the restrictions of the measures  $\mathbb{P}^\star$  and  $\mathbb{P}^\circ$  by  $\mathbb{P}_t^\star$  and  $\mathbb{P}_t^\circ$  respectively. By equation (4.1) in [Schauer et al. \(2013\)](#) and Girsanov's theorem,

$$\frac{dP_t^\star}{dP_t^\circ}(X^\circ) = \frac{p(t, X_t^\circ; T, v)}{p(0, u; T, v)} \times G_t(X^\circ) \times \exp \left( - \int_0^t \kappa(s, X_s^\circ) a^{-1}(s, X_s^\circ) dX_s^\circ + \frac{1}{2} \int_0^t \kappa(s, X_s^\circ)^2 a^{-1}(s, X_s^\circ) ds \right),$$

where  $\kappa(s, x) = (v - x)/(T - s)$  and

$$G_t(X^\circ) = \exp \left( \int_0^t b(s, X_s^\circ) a^{-1}(s, X_s^\circ) dX_s^\circ - \frac{1}{2} \int_0^t b(s, X_s^\circ)^2 a^{-1}(s, X_s^\circ) ds \right).$$

A careful argument by [Delyon and Hu \(2006\)](#) (also outlined in [Papaspiliopoulos et al. \(2013\)](#)) shows that this likelihood ratio can be rewritten as

$$\begin{aligned} G_t(X^\circ) \times \exp \left( - \frac{1}{2} \int_0^t (T - s) \kappa(s, X_s^\circ)^2 \diamond da^{-1}(s, X_s^\circ) \right) \\ \times \frac{p(t, X_t^\circ; T, v)}{\varphi(v; X_t^\circ, (T - t)a(t, X_t^\circ))} \frac{\alpha}{\sqrt{a(t, X_t^\circ)}}. \end{aligned} \quad (5.1)$$

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

$$\alpha = \frac{1}{p(0, u; T, v)} \frac{1}{\sqrt{2\pi T}} \exp \left( - \frac{1}{2T} a(0, u)^{-1} (v - u)^2 \right).$$

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. [Delyon and Hu \(2006\)](#) provide sufficient conditions such that for solutions to  $(\circ')$  there exists an almost surely finite random variable  $C$  such that

$$|v - X_t^\circ|^2 \leq C(T - t) \log \log(1/(T - t) + e). \quad (5.2)$$

This ensures that the second term in equation (5.1) is well defined in the limit  $t \uparrow T$ . Intuitively, when  $t$  is close to  $T$  and  $X_t^\circ$  is close to  $v$ ,  $p(t, X_t^\circ; T, v)$  and  $\varphi(v; X_t^\circ, (T - t)a(t, X_t^\circ))$  cancel each other. This is true under the expectation, for completeness we give the result in A.2, as to the best of our knowledge a rigorous derivation of the constant appearing in the likelihood ratio for the Delyon and Hu (2006) type proposals is missing in the literature. In this derivation, we naturally assume the conditions stated in section 4 of Delyon and Hu (2006).

Papaspiliopoulos et al. (2013) extend the reparametrisation of the diffusion bridges in Roberts and Stramer (2001) (cf. section 2) to reducible diffusions and otherwise decouple parameter and bridge by first discretising and then using the transformation formula in Euclidean coordinates for the discretised bridges. The use of the driving Brownian motion of the continuous time bridge proposal ( $\phi'$ ) as innovation process was proposed by Fuchs (2013).

## 6. Time changing and scaling of linear guided proposals

Simulation of  $X^\circ$  and numerical evaluation of  $\Psi$  as defined in (3.4) is numerically cumbersome since the drift of  $X^\circ$  and the integrand  $D$  explode for  $s$  near the endpoint  $T$ . As an example, suppose  $\sigma$  is constant and we take  $\tilde{X} = \sigma dW_t$ . Then we have  $\tilde{r}(s, x) = \tilde{a}^{-1}(v - x)/(T - s)$  and

$$\log \Psi(X^\circ) = \int_0^T D(s, X_s^\circ) ds = \int_0^T b(s, X_s^\circ)' \tilde{r}(s, X_s^\circ) ds.$$

In this section we explain how these numerical problems can be dealt with using a time and scaling of the proposal process.

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 as a 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 intuition about our choice of the time change is as follows: as shown in Schauer et al. (2013) up to a logarithmic term  $\|v - X_s^\circ\| \sim \sqrt{T - s}$  for  $s$  close to  $T$ . Therefore,  $\|\tilde{r}(s, X_s^\circ)\| \sim \|\frac{v - X_s^\circ}{T - s}\| \sim 1/\sqrt{T - s}$ . Now if we wish to remove the singularity in

the integral  $\int_0^T (T-s)^{-1/2} ds$ , we can substitute  $s = u(2 - u/T)$  to get  $\int_0^T ((T-s)^{-1/2} ds = 2T^{-1/2} \int_0^T 1 du = 2\sqrt{T}$ .

To make this intuitive idea more precise, we need a few more results from [Schauer et al. \(2013\)](#). If  $\tilde{X}$  is a linear process (satisfying equation (3.3)), then

$$\tilde{r}(s, x) = \tilde{H}(s)(v(s) - x), \quad (6.1)$$

where

$$v(s) = \Phi(s, T)v - \int_s^T \Phi(s, \tau)\tilde{\beta}(\tau) d\tau \quad (6.2)$$

( $\tilde{r}$  and  $\tilde{H}$  were defined in equation (3.1)). Here  $\Phi(t, s) = \Psi(t)\Psi(s)^{-1}$  with  $\Psi$  the fundamental  $d \times d$  matrix that satisfies

$$\Psi(t) = I + \int_0^t \tilde{B}(\tau)\Psi(\tau) d\tau.$$

Define the time change  $\tau: [0, T] \rightarrow [0, T]$  by

$$\tau(s) = s(2 - s/T)$$

and define the process  $U$  by

$$U_s := \frac{1}{2} \frac{v(\tau(s)) - X_{\tau(s)}^\circ}{T - \tau(s)} \tau'(s) = \frac{v(\tau(s)) - X_{\tau(s)}^\circ}{T - s}. \quad (6.3)$$

This implies

$$X_{\tau(s)}^\circ = v(\tau(s)) - (T - s)U_s =: G(s, U_s). \quad (6.4)$$

In the following, we denote the derivatives of  $v$  and  $\tau$  by  $\dot{v}$  and  $\dot{\tau}$  respectively.

**Proposition 6.1.** The time changed process  $U = (U_s, s \in [0, T])$  satisfies the stochastic differential equation

$$\begin{aligned} dU_s &= \frac{2}{T} \dot{v}(\tau(s)) ds - \frac{2}{T} b(\tau(s), G(s, U_s)) ds \\ &\quad + \frac{1}{T-s} \left( I - 2a(\tau(s), G(s, U_s))J(s) \right) U_s ds \\ &\quad - \sqrt{\frac{2}{T}} \frac{1}{\sqrt{T-s}} \sigma(\tau(s), G(s, U_s)) dW_s, \quad U_0 = \frac{v-u}{T} \end{aligned} \quad (6.5)$$

where

$$J(s) = \tilde{H}(\tau(s))(T-s)^2/T. \quad (6.6)$$

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

If we simulate  $U$  on an equidistant grid we can recover  $X^\circ$  on a nonequidistant grid from equation (6.4). 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  $U_s = g(\theta, Z_s^\circ)$ , where  $Z^\circ$  is the driving Brownian Motion for  $U$ .



At first, it may seem that the SDE given by equation (6.5) has not resolved the singularity problem at time  $T$ . However, note that if  $b$  and  $\sigma$  are smooth, then for  $s \approx T$ ,

$$dU_s \approx -\frac{2}{T}b(T, v) ds - \frac{1}{T-s}U_s ds - \sqrt{\frac{2}{T}} \frac{1}{\sqrt{T-s}}\sigma(T, v) dW_s. \quad (6.7)$$

In the following subsection we study discretisation of this SDE and show that no discretisation error is made at the final discretisation step near  $T$ .

Finally we give the time changed version of formula (3.2), which follows from the same elementary, if tedious, calculations as in the proof of Proposition 6.1:

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

### 6.1. Discretisation error

In this section we show in a simple setting the benefits of discretising  $U$  instead of  $X^\circ$ . To this end, we consider the diffusion  $dX_s = \sigma dW_s$ . In this case it is natural to take  $d\tilde{X} = \sigma dW_s$  which yields proposals satisfying the stochastic differential equation (SDE)

$$dX_s^\circ = \frac{v - X_s^\circ}{T-s} ds + \sigma dW_s, \quad X_0^\circ = u$$

(note the resemblance with equation (6.7)). Of course, for this particular example it is trivial to obtain an exact realisation at a finite set of points in  $[0, T]$ , but suppose for a moment we're not aware of this fact. Now there are at least two approaches for obtaining an approximate discrete skeleton of the process:

1. Apply the Euler discretisation directly to the SDE for  $X^\circ$ .
2. First transform  $X^\circ$  to the process  $U$  as defined in (6.3), then apply Euler discretisation on that process and subsequently transform this discretisation back to  $X^\circ$ .

In this section we will show that the second approach yields improved accuracy. We assume in both cases we wish to obtain a skeleton at  $m-1$  (with  $m \geq 2$ ) time points in  $(0, T)$ . Let  $h = T/m$ .

*Approach 1.* Direct application of Euler discretisation to  $X^\circ$  gives that the law of  $X_{s+h}^\circ$  conditional on  $X_s^\circ = x$  is approximately normal with mean  $x + \frac{v-x}{T-s}h$  and covariance  $C_{\text{Euler}} = ha$ . While the mean is exact, the true covariance equals  $C_{\text{true}} = h \frac{T-s-h}{T-s}a$ . Therefore, for  $s \in [0, T-h]$

$$d(s) := \|C_{\text{true}} - C_{\text{Euler}}\| = \frac{h^2}{T-s} \|a\| \leq h \|a\|.$$

Note that the largest difference is attained for  $s = T - h$ .

*Approach 2.* Straightforward calculus gives that  $U$  satisfies the SDE

$$dU_s = -\frac{1}{T-s}U_s ds + \sqrt{\frac{2}{T}} \frac{1}{\sqrt{T-s}} \sigma dW_s, \quad U_0 = \frac{v-u}{T}.$$

If we denote the Euler discretisation applied to this equation by  $\underline{U}$ , then

$$\underline{U}_{s+h} \mid \underline{U}_s = c \sim N \left( \frac{T-s-h}{T-s}c, \frac{2}{T} \frac{h}{T-s}a \right).$$

As  $X_{\tau(s)}^\circ = v - (T-s)U_s$ , we define  $\underline{X}_{\tau(s)}^\circ = v - (T-s)\underline{U}_s$ . Using the result of the previous display, some tedious calculations reveal that

$$\underline{X}_{\tau(s+h)}^\circ \mid \underline{X}_{\tau(s)}^\circ = x \sim N \left( x + 2h \frac{v-x}{T-s} - h^2 \frac{v-x}{(T-s)^2}, C'_{\text{Euler}} \right),$$

where

$$C'_{\text{Euler}} = \frac{2h}{T} \frac{(T-s-h)^2}{T-s} a.$$

The true law of  $X_{\tau(s+h)}^\circ$ , conditional on  $X_{\tau(s)}^\circ = x$  has exactly the same mean, while the true covariance equals

$$C'_{\text{true}} = \frac{2h}{T} \frac{(T-s-h)^2}{T-s} a - \frac{h^2}{T} \frac{(T-s-h)^2}{(T-s)^2} a.$$

Therefore,

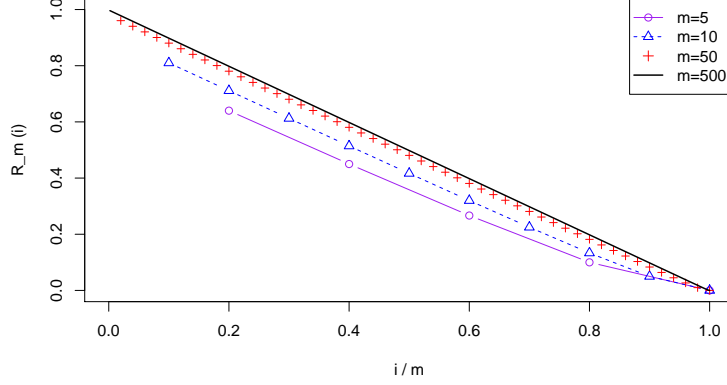
$$\begin{aligned} d'(s) &:= \|C'_{\text{true}} - C'_{\text{Euler}}\| = \frac{h^2}{T} \left( 1 - \frac{h}{T-s} \right)^2 \|a\| \\ &\leq \frac{h^2}{T} \left( \frac{T-h}{T} \right)^2 \|a\| \leq \frac{h^2}{T} \|a\| = \frac{h}{N} \|a\|. \end{aligned}$$

Note that this difference is zero if  $s = T - h$ ; while its maximal value is attained for  $s = 0$ .

Comparing both approaches we see that

$$R_m(i) := \frac{d'((i-1)h)}{d((i-1)h)} = \frac{(m-i)^2}{m(m-i+1)}, \quad i = 1, \dots, m.$$

This ratio is always smaller than 1 and  $i \mapsto R(i)$  decreases roughly linearly to zero. This implies that it is always advantageous in this example to use the second approach, the closer we come to the endpoint  $T$ , the more we gain. As we apply the discretisation iteratively over all points it is important that the ratio is uniformly smaller than 1. Further, from figure 2 it is apparent that relative gain of approach 2 is higher for small values of  $m$ . Simulation results indicate that this advantage also holds more generally, though a proof for this is not available yet.

FIGURE 2. Behaviour of  $R_m(i)$  with respect to  $m$  and  $i$ .

## 7. Examples

### 7.1. Example for one-dimensional diffusion

In this section we consider an example. 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 time change and scaling proposed in the previous section reduces discretisation error.

We consider the diffusion process driven by the SDE

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

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. See figure 3.

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 random walk proposals for  $q(\sigma^\circ | \sigma)$  of the form

$$\log \sigma^\circ := \log \sigma + u, \quad u \sim \mathcal{U}(-0.1, 0.1).$$

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  $1000$ .

Figures 4 and 5 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

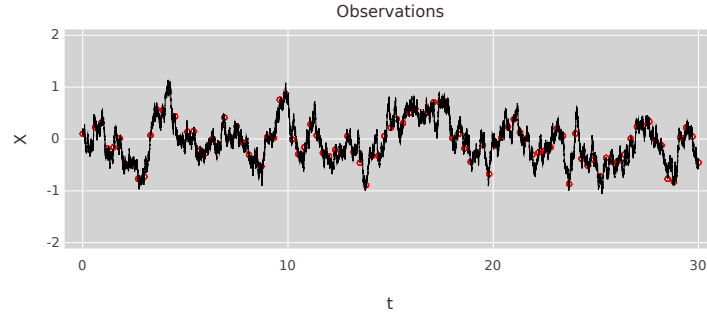


FIGURE 3. Diffusion path and discrete observations (small circles).

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, as we anticipated in section 6.1.)

## 7.2. Prokaryotic auto-regulation

### 7.2.1. Model and data

In this section we apply our methods to an example presented in section 4 of [Golightly and Wilkinson \(2010\)](#) on prokaryotic auto-regulation. We briefly introduce the example and focus on estimation; the interested reader can consult [Golightly and Wilkinson \(2010\)](#) for more details and background on this example. A gene is expressed into a protein  $P$  in several steps including transcription of the DNA encoding the protein into mRNA by RNA-polymerase and synthesis of the protein in the ribosome as specified by the mRNA. In the autoregulation model, pairs of the protein may reversibly form dimers  $P_2$  which bind to certain regions of the DNA resulting in bound DNA  $\cdot P_2$  which is not transcribed into RNA for  $P$ , thus providing an auto-regulation mechanism for the synthesis of this particular protein  $P$ . The total amount of bound and free DNA is constant  $\text{DNA} + \text{DNA} \cdot P_2 = K$  so the quantities of RNA,  $P$ ,  $P_2$  and DNA describe the state of the system. These reactions can be modelled as Markov jump process. While it is straightforward to exactly simulate sample paths of the jump process with values  $(\text{RNA}, P, P_2, \text{DNA}) \in \mathbb{N}^4$ , estimation can be carried out by exploiting a diffusion approximation to this process. This approximating diffusion process with values in  $\mathbb{R}^4$  solves the *Chemical Langevin Equation*

$$dX_t = Sh_\theta(X_t) dt + S \text{diag}(\sqrt{h_\theta(X_t)}) dW_t \quad (7.1)$$

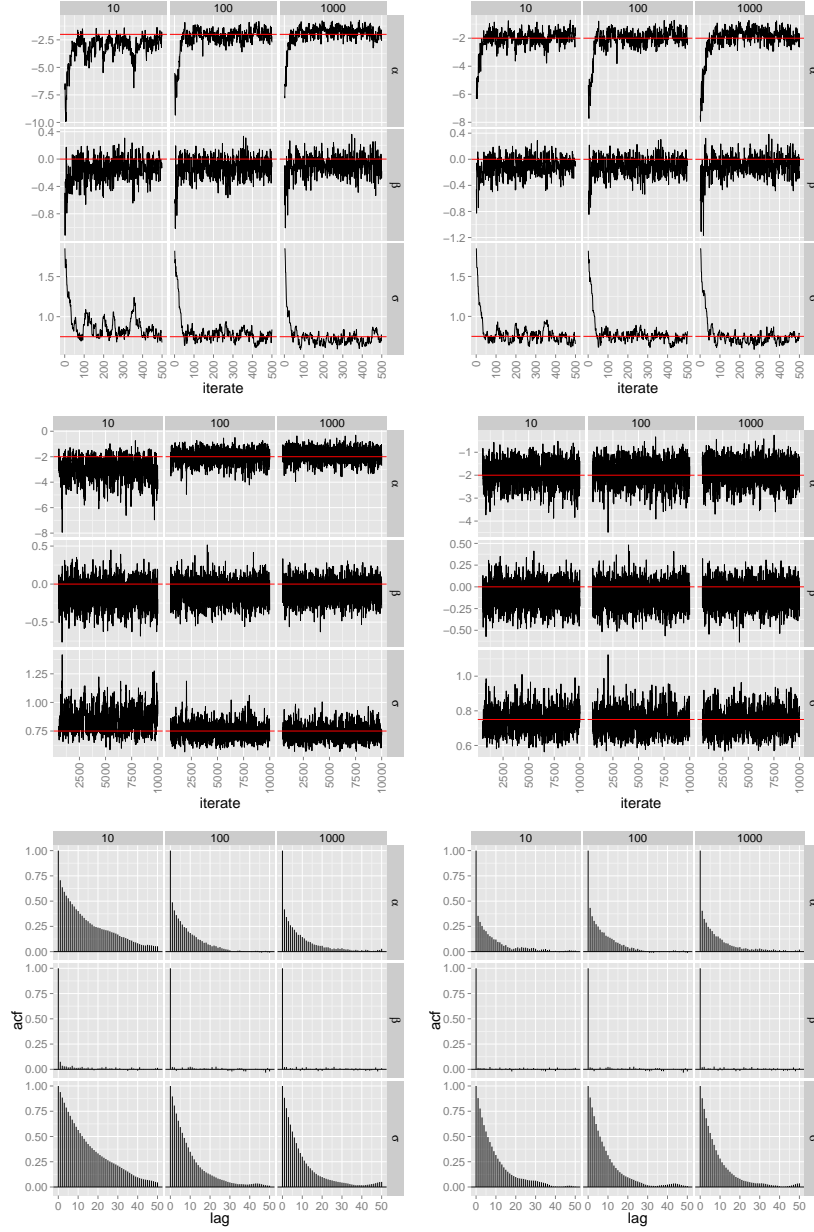


FIGURE 4. Panels comparing different numbers of imputed points ( $m = 10, 100, 1000$ ). Left: without time change. Right: with time change. Top: first 500 iterates. Middle: iterates 501-10.000. Bottom: ACF-plots based on iterates 501-10.000.

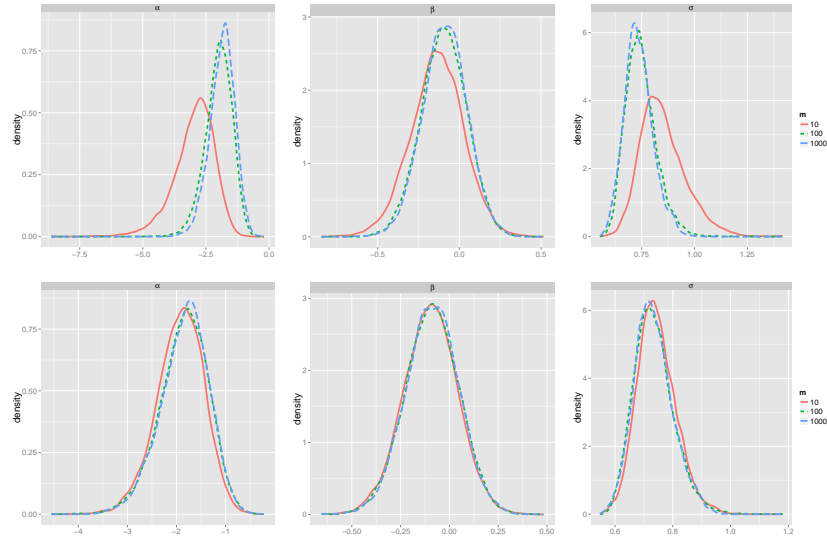


FIGURE 5. Kernel density estimates of the parameters based on iterates 501-10,000. Top: non time-changed. Bottom: time-changed. The bias from using only a small number of imputed points ( $m = 10$ , red curve) is clearly smaller for the time changed process.

driven by a  $\mathbb{R}^8$ -valued Brownian motion, where

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

is the stoichiometry matrix of the system describing the chemical reactions and

$$h_\theta(x) = \theta \circ h(x)$$

is a function describing the hazard for a particular reaction to happen. Here  $\circ$  denotes the Hadamard (or entrywise) product of two vectors and

$$\theta = [\theta_1, \theta_2, \dots, \theta_8]'$$

is the vector of the unknown rates of each of the 8 distinct reactions which are known to happen proportionally to the amount of

$$h(x) = [x_3x_4, K - x_4, x_4, x_1, \frac{1}{2}x_2(x_2 - 1), x_3, x_1, x_2]'$$

(respectively) in the system. Note that this is a natural situation in which both drift and diffusion coefficient depend on the same parameters. Golightly and Wilkinson (2010) proceed from (7.1) to a SDE driven by a 4-dimensional Brownian motion with  $\sigma(x) = (Sh_\theta(x)S')^{\frac{1}{2}}$ . In our case, except for algorithm 2 which does not apply here, this is not necessary.

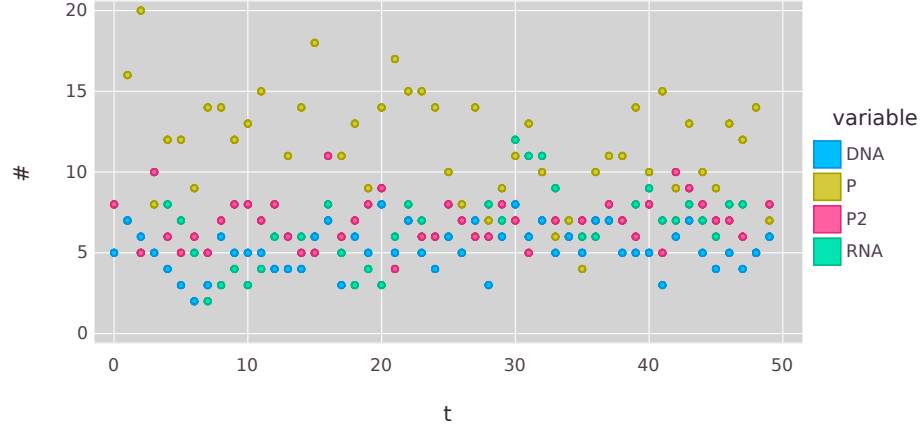


FIGURE 6. *Prokaryotic auto-regulation: Quantities of (RNA, P, P<sub>2</sub>, DNA) at integer times (simulated from discrete model).*

Golightly and Wilkinson (2010) simulated the continuous time Markov jump process with parameter values  $\theta = (0.1, 0.7, 0.35, 0.2, 0.1, 0.9, 0.3, 0.1)$  and  $K = 10$  and recorded the quantities of each of (RNA, P, P<sub>2</sub>, DNA) at integer times  $0, 1, \dots, 49$ , see figure 6.

For inference, Golightly and Wilkinson (2010) use independent  $\mathcal{U}[-5, 5]$  priors on  $\log \theta_i$  and a Gaussian random walk proposal updating the components of  $\theta$  jointly. They find that the few observations taken allow for remarkably good recovery of the true parameters.

### 7.2.2. Choice of guided proposals

We perform an analysis of the same data, this time with the guided innovation scheme using the time change from section 6 using (7.1) as model. 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  $b_\theta$ . 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

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

Except for its first and fifth entry,  $x \mapsto h(x)$  is linear. Therefore, for  $i \in \{2, 3, 4, 6, 7, 8\}$  we set  $\tilde{h}_i = h_i$ . For  $i = 1$ , we take  $\tilde{h}_1 = c_1 + u_{1,3}x_3 + u_{1,4}x_4$ . Values for  $c_1, u_{1,3}$  and  $u_{1,4}$  are obtained from a weighted linear regression of  $x_3x_4$  on  $x_3$  and  $x_4$ , with weights proportional to  $x_3x_4$ . Similarly, for  $i = 5$ , we take  $\tilde{h}_5 = c_5 + u_{5,2}x_2$ . Values for  $c_5$  and  $u_{5,2}$  are obtained from a weighted linear regression of  $\frac{1}{2}x_2(x_2 - 1)$  on  $x_2$  and  $x_4$ , with weights proportional to  $x_2(x_2 - 1)$ .

We take a *weighted* regression in this way because for a good proposal the error matters more if the corresponding dispersion component is small. This automatic procedure gives  $u_{5,2} = 18$ ,  $c_5 = -82$ ,  $u_{1,3} = 5$ ,  $u_{1,4} = 6$ , and  $c_1 = -32$  for the data obtained.

### 7.2.3. Rejection of nonpositive bridges

Contrary to the underlying jump process, the chemical Langevin approximation may have positive probability of leaving the positive cone, even before Euler discretisation [Szpruch and Higham \(2009/10\)](#). This is also the case for this example, to intuitively see this, note for example that  $a_{11}(x) = \theta_3^2 x_4 + \theta_7^2 x_1$ . So even if the first coordinate of  $X$  approaches 0 at time  $t$ , the diffusivity of  $X$  in that direction does not vanish if  $X_t^{(4)} > 0$ , so  $X_{t+h}^{(1)} < 0$  with positive probability.

As we are naturally not interested in values of  $\theta$  which explain the data via bridges  $X^*$  violating the condition  $X^* \geq 0$ , we resample proposals with  $X_s^* < 0$  for some  $s$ . In this way we effectively draw from the conditional distribution of  $\theta$ , conditional on the discrete time data and  $X_s^* \geq 0$  for all values of  $s$  that correspond to imputed points.

### 7.2.4. Results from applying algorithm 1

As priors for  $\log \theta_i$  we use independent  $\mathcal{U}[-7, 7]$  distributions. We used algorithm 1 with step 2(a) replaced with

- Sample a Wiener process  $Z_i^\circ$ , conditional on  $g(\theta, Z_i^\circ) \geq 0$  (using rejection sampling).

and step 3(a) replaced with

- Sample  $\theta^\circ \sim q(\cdot \mid \theta)$ , conditional on  $g(\theta^\circ, Z_i^\circ) \geq 0$  for all  $1 \leq i \leq n$  (using rejection sampling).

We took random-walk type proposals for  $\theta$  on the log scale, where conditional of the current value  $\theta$  a new value  $\theta^\circ$  is sampled from

$$\log \theta^\circ \sim N(\log \theta, (0.12)^2).$$

We choose  $\theta = (0.05, \dots, 0.05)'$  as first iterate. A run of 100,000 iterations using  $m = 20, 50$  imputed points (including endpoints) for each bridge yields the estimates for  $\theta_1, \dots, \theta_8$  given in the upper panel of table 1.

In case  $m = 50$ , on average 76% of the sampled bridges for a segment where accepted in step 2(b) and the average acceptance probability for  $\theta^\circ$  in step 3 was 24%.

### 7.2.5. Results from applying algorithm 3

In this example one can clearly see from figure 8 that both  $\theta_1$  and  $\theta_2$  as  $\theta_4$  and  $\theta_5$  are highly correlated. This worsens mixing of the Markov chain for these parameters.



Alg. 1	Mean	Standard dev.	Autocorrelation time
$\theta_1$	0.0957	0.0277	180.05
$\theta_2$	0.7523	0.2187	176.88
$\theta_3$	0.3777	0.1120	171.18
$\theta_4$	0.2342	0.0516	75.18
$\theta_5$	0.0726	0.0206	187.59
$\theta_6$	0.7016	0.1958	169.24
$\theta_7$	0.3095	0.0901	169.36
$\theta_8$	0.1381	0.0297	87.62

Alg. 3	Mean	Standard dev.	Autocorrelation time
$\theta_1$	0.0976	0.0289	97.23 (54%)
$\theta_2$	0.7664	0.2268	100.29 (57%)
$\theta_3$	0.3811	0.1039	20.44 (12%)
$\theta_4$	0.2343	0.0523	15.47 (21%)
$\theta_5$	0.0731	0.0215	127.90 (68%)
$\theta_6$	0.7048	0.2032	121.05 (72%)
$\theta_7$	0.3130	0.0845	21.30 (13%)
$\theta_8$	0.1376	0.0295	15.74 (18%)

TABLE 1

*Prokaryotic auto-regulation: Estimated parameters of the posterior distribution using algorithms 1 and 3. The estimated autocorrelation time was computed using R package coda.*

Since

$$b_\theta(x) = S\theta \circ h(x) = \sum_j \theta_j \varphi_j(x) \quad \text{with} \quad \varphi_j(x) = (s_{ij} h_j(x))_i \quad (7.2)$$

we see that  $b$  is of the form (4.5). Therefore, we can apply algorithm 3. We make a slight adjustment, in which step 3.2'(c) is replaced with

- Sample  $\theta^\circ \sim \mathcal{N}(\theta, \alpha^2 W_\theta^{-1})$  under the condition  $g(\theta^\circ, Z_i^*) \geq 0$  for all  $1 \leq i \leq n$  (using rejection sampling).

A run of 100,000 iterations using  $m = 20, 50$  imputed points (including endpoints) for each bridge yields the estimates for  $\theta_1, \dots, \theta_8$  given in the lower panel of table 1. We took the same priors as in subsection 7.2.4. The scaling parameter of the random walk proposal was chosen according to the remark after algorithm 3. The corresponding acceptance probability for the proposals for  $\theta$  was 26% (for  $m = 50$ ). This compares well with the anticipated value of 0.234.

From table 1 we see that indeed application of algorithm 3 greatly reduces autocorrelation times over algorithm 1.

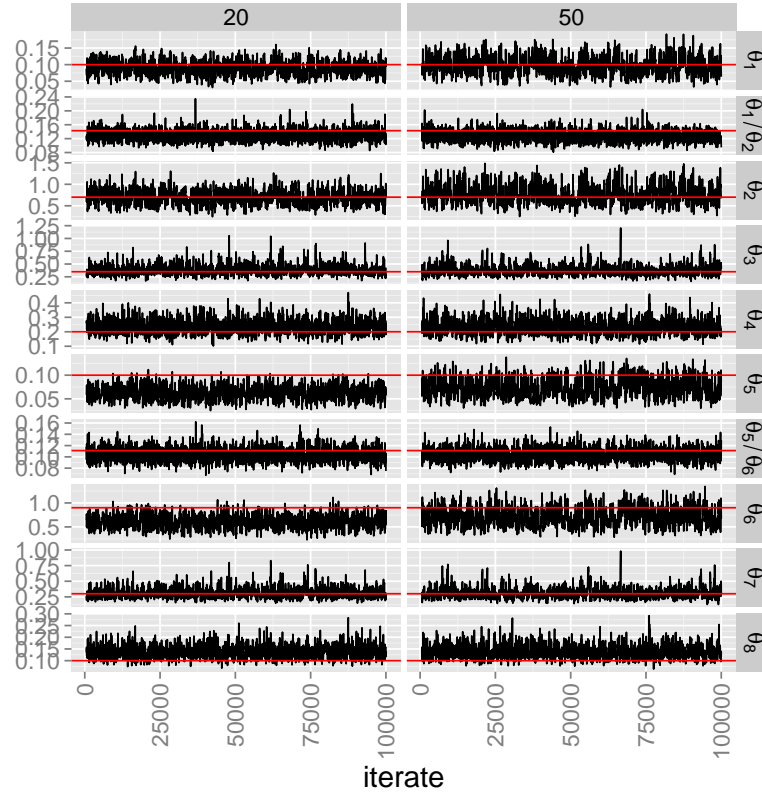


FIGURE 7. Prokaryotic auto-regulation: Iterates of the MCMC chain for  $m = 20, 50$  using 3.a. Note the shift in the distribution of  $\theta_5$  and  $\theta_6$ .

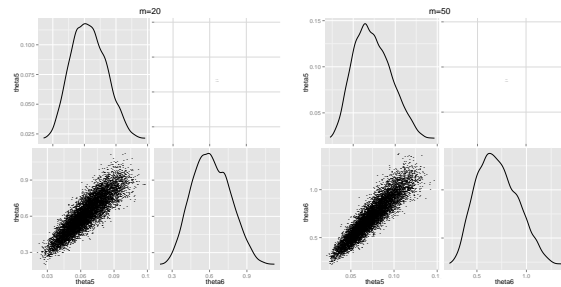


FIGURE 8. Scatterplots of  $\theta_5$ ,  $\theta_6$  and marginal densities for  $m = 20, m = 50$  (including burn in).

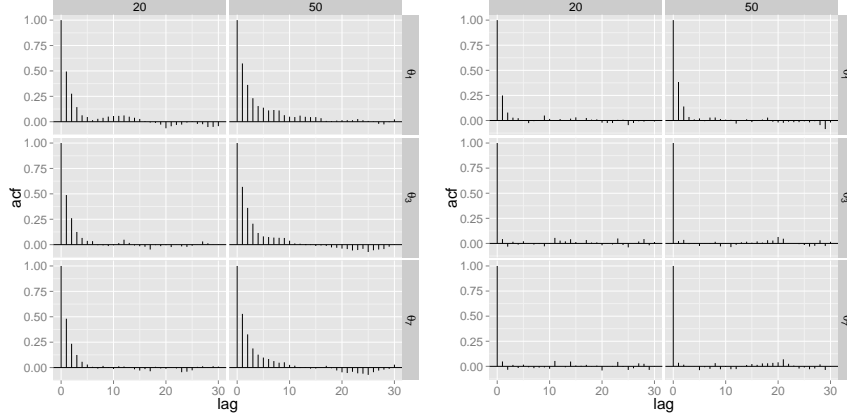


FIGURE 9. ACF plots of the thinned samples of  $\theta_1$ ,  $\theta_3$ ,  $\theta_7$  (taking every 50th iterate after burn in). Left: algorithm 1, right: algorithm 3.

## Appendix A: Proofs

### A.1. Proof of Proposition 6.1

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

$$dX_s = b(s, X_s) ds + \sigma(s, X_s) dW_s$$

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

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

Applying this to  $X^\circ$  (defined in equations (o) and (oo)) 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

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

By equation (6.1),

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

The result now follows from substituting this expression and using the relation  $X_\tau^\circ = G(U_s)$ . The final statement follows from  $\lim_{t \rightarrow T} \tilde{H}(t)(T - t) = \tilde{a}^{-1}$ , see (Schauer et al., 2013, Lemma 8).

### A.2. Proportionality constant for the Delyon-Hu type proposal

Denote the product in equation (5.1) by  $\xi_t(X^\circ)$ . Delyon and Hu (2006) provide a rigorous proof that  $X^\circ$  and  $X^\star$  are absolutely continuous on  $[0, T]$  with density  $c \xi_T(X^\circ)$  for some (not specified) constant  $c > 0$ . Hence  $\mathbb{E} \xi_T(X^\circ) = 1/c$ .

For  $t < T$  we have

$$\mathbb{E} \left[ \frac{\varphi(v; X_t^\star, (T-t)a(t, X_t^\star))}{p(t, X_t^\star; T, v)} \sqrt{a(t, X_t^\star)} \right] = \alpha \mathbb{E} [\xi_t(X^\circ)].$$

The right-hand-side of this expression converges to  $\alpha \mathbb{E} [\xi_T(X^\circ)] = \alpha/c$ . The left-hand-side equals

$$\int \frac{1}{\sqrt{2\pi(T-t)}} \exp \left( -\frac{1}{2} \frac{(v-x)^2}{2a(t, x)(T-t)} \right) \frac{p(0, u; t, x)}{p(0, u; T, v)} dx.$$

From here, we can use the substitution  $y = (x - v)/\sqrt{T - t}$  in the integral and use the same arguments as in the proof of Lemma 7 in Delyon and Hu (2006) to establish that this converges to  $\sqrt{a(T, v)}$  as  $t \uparrow T$ . We conclude that  $c = \alpha/\sqrt{a(T, v)}$  is the proportionality constant.

## Appendix B: Implementation details

The source code of the examples will be available online<sup>1</sup> It is written in the programming language Julia (Bezanson et al. (2012)). In this section we provide some computational details in case  $\tilde{X}$  is a time homogeneous process with linear drift  $\tilde{b}(x) = \tilde{B}x + \tilde{\beta}$  and constant diffusion coefficient  $\tilde{a}$ . Simulation of  $U$  as defined in equation (6.5) requires evaluation of both  $\dot{v}$  and  $J$ , where  $v$  and  $J$  are defined in equations (6.2) and (6.6) respectively. If  $\tilde{B} \equiv 0$ , then this is straightforward. Else, if we let  $\lambda$  be the solution to continuous time Lyapunov (matrix) equation

$$\tilde{B}\lambda + \lambda\tilde{B}' + \tilde{a} = 0,$$

then

$$J(s) = \frac{(T-s)^2}{T} \left( e^{-\tilde{B}(T(1-s/T)^2)} \lambda e^{-\tilde{B}'(T(1-s/T)^2)} - \lambda \right)^{-1}$$

$$\dot{v}(\tau(s)) = e^{-\tilde{B}(T(1-s/T)^2)} (\tilde{B}v + \tilde{\beta}).$$

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

Note that both  $J$  and  $\dot{v}$  do not depend on  $x$ . This implies that these functions only need to be computed once on the grid of a single bridge in each iteration. In case  $\bar{B}$  does not depend on  $\theta$  then these functions can be precomputed on a grid in advance to the MCMC-algorithm. As computing matrix exponentials is expensive, this reduces computing time substantially.

## Acknowledgement

We thank Andrew Golightly for sharing the simulated data used in the prokaryotic auto-regulation example with us.

## 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**(3), 333–382. With discussions and a reply by the authors.
- Bezanson, J., Karpinski, S., Shah, V. B. and Edelman, A. (2012). Julia: A fast dynamic language for technical computing. *CoRR* **abs/1209.5145**.
- Bladt, M. and Sørensen, M. (2012). Simple simulation of diffusion bridges with application to likelihood inference for diffusions. Preprint.
- Chib, S., Pitt, M. K. and Shephard, N. (2004). Likelihood based inference for diffusion driven models. Economics Papers 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*, pp. 1418–1420. IEEE.
- Delyon, B. and Hu, Y. (2006). Simulation of conditioned diffusion and application to parameter estimation. *Stochastic Processes and their Applications* **116**(11), 1660 – 1675.
- Durham, G. B. and Gallant, A. R. (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *J. Bus. Econom. Statist.* **20**(3), 297–338. With comments and a reply by the authors.
- Elerian, O., Chib, S. and Shephard, N. (2001). Likelihood inference for discretely observed nonlinear diffusions. *Econometrica* **69**(4), 959–993.
- Eraker, B. (2001). MCMC analysis of diffusion models with application to finance. *J. Bus. Econom. Statist.* **19**(2), 177–191.
- Fuchs, C. (2013). *Inference for diffusion processes*. Springer, Heidelberg. With applications in life sciences, With a foreword by Ludwig Fahrmeir.
- Golightly, A. and Wilkinson, D. J. (2006). Bayesian sequential inference for nonlinear multivariate diffusions. *Stat. Comput.* **16**(4), 323–338.
- Golightly, A. and Wilkinson, D. J. (2010). *Learning and Inference in Computational Systems Biology*, chapter Markov chain Monte Carlo algorithms for SDE parameter estimation, pp. 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**(2), 609–635.

- Lin, M., Chen, R. and Mykland, P. (2010). On generating Monte Carlo samples of continuous diffusion bridges. *J. Amer. Statist. Assoc.* **105**(490), 820–838.
- Papaspiliopoulos, O., Roberts, G. O. and Sköld, M. (2003). Non-centered parameterizations for hierarchical models and data augmentation. In *Bayesian statistics, 7 (Tenerife, 2002)*, pp. 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.
- Papaspiliopoulos, O., Roberts, G. O. and Stramer, O. (2013). Data Augmentation for Diffusions. *J. Comput. Graph. Statist.* **22**(3), 665–688.
- Pedersen, A. R. (1995). Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. *Bernoulli* **1**(3), 257–279.
- Roberts, G. O. and Stramer, O. (2001). On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. *Biometrika* **88**(3), 603–621.
- 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.
- Rosenthal, J. S. (2011). *Handbook of Markov Chain Monte Carlo (Chapman & Hall/CRC Handbooks of Modern Statistical Methods)*, chapter Optimal proposal distributions and adaptive MCMC. Chapman and Hall/CRC, 1 edition.
- Schauer, M. R., Van der Meulen, F. H. and Van Zanten, J. H. (2013). Guided proposals for simulating multi-dimensional diffusion bridges. *ArXiv e-prints*.
- Sørensen, H. (2004). Parametric Inference for Diffusion Processes Observed at Discrete Points in Time: a Survey. *Internat. Statist. Rev.* **72**, 337–354.
- Szpruch, L. and Higham, D. J. (2009/10). Comparing hitting time behavior of Markov jump processes and their diffusion approximations. *Multiscale Model. Simul.* **8**(2), 605–621.
- Van der Meulen, F. H., Schauer, M. R. and Van Zanten, J. H. (2014). Reversible jump MCMC for nonparametric drift estimation for diffusion processes. *Comput. Statist. Data Anal.* **71**, 615–632.
- Van der Meulen, F. H. and Van Zanten, J. H. (2013). Consistent nonparametric Bayesian inference for discretely observed scalar diffusions. *Bernoulli* **19**(1), 44–63.
- Van Zanten, J. H. (2013). Nonparametric bayesian methods for one-dimensional diffusion models. *Mathematical biosciences* **243**(2), 215–222.