

## Simulation of elliptic and hypo-elliptic conditional diffusions

Bierkens, Joris; van der Meulen, Frank; Schauer, Moritz

**DOI**

[10.1017/apr.2019.54](https://doi.org/10.1017/apr.2019.54)

**Publication date**

2020

**Document Version**

Accepted author manuscript

**Published in**

Advances in Applied Probability

**Citation (APA)**

Bierkens, J., van der Meulen, F., & Schauer, M. (2020). Simulation of elliptic and hypo-elliptic conditional diffusions. *Advances in Applied Probability*, 52(1), 173-212. <https://doi.org/10.1017/apr.2019.54>

**Important note**

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

**Copyright**

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

**Takedown policy**

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

# SIMULATION OF ELLIPTIC AND HYPO-ELLIPTIC CONDITIONAL DIFFUSIONS

JORIS BIERKENS, FRANK VAN DER MEULEN, AND MORITZ SCHAUER

ABSTRACT. Suppose  $X$  is a multidimensional diffusion process. Assume that at time zero the state of  $X$  is fully observed, but at time  $T > 0$  only linear combinations of its components are observed. That is, one only observes the vector  $LX_T$  for a given matrix  $L$ . In this paper we show how samples from the conditioned process can be generated. The main contribution of this paper is to prove that guided proposals, introduced in (Schauer et al. (2017)), can be used in a unified way for both uniformly and hypo-elliptic diffusions, also when  $L$  is not the identity matrix. This is illustrated by excellent performance in two challenging cases: a partially observed twice integrated diffusion with multiple wells and the partially observed FitzHugh-Nagumo model.

## 1. INTRODUCTION

Let  $X = (X_t, t \in [0, T])$  be a  $d$ -dimensional diffusion process satisfying the stochastic differential equation (SDE)

$$dX_t = b(t, X_t) dt + \sigma(t, X_t) dW_t, \quad X_0 = x_0, \quad t \in [0, T]. \quad (1)$$

Here  $b: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ ,  $\sigma: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d'}$  and  $W$  is a  $d'$ -dimensional Wiener process with all components independent. Stochastic differential equations are widely used for modelling in engineering, finance and biology, to name a few fields of applications. In this paper we will not only consider *uniformly elliptic* models, where it is assumed that there exists an  $\varepsilon > 0$  such that for all  $(t, x) \in [0, T] \times \mathbb{R}^d$  and  $y \in \mathbb{R}^d$  we have  $\|\sigma(t, x)'y\|^2 \geq \varepsilon\|y\|^2$ , but also hypo-elliptic models. These are models where the randomness spreads through all components - ensuring the existence of smooth transition densities of the diffusion, even though the diffusion is possibly not uniformly elliptic (for example because the Wiener noise only affects certain components.) Such models appear frequently in application areas; many examples are given in the introductory section of Clairon and Samson (2017). A rich subclass of nonlinear hypo-elliptic diffusions that is included in our setup is specified by a drift of the form

$$b(t, x) = Bx + \beta(t, x), \quad (2)$$

where

$$B = \begin{bmatrix} 0_{k \times k'} & I_{k \times k} \\ 0_{k' \times k'} & 0_{k' \times k} \end{bmatrix} \quad \beta(t, x) = \begin{bmatrix} 0_{k \times 1} \\ \underline{\beta}(t, x) \end{bmatrix} \quad \sigma = \begin{bmatrix} 0_{k \times d'} \\ \underline{\sigma}(t) \end{bmatrix} \quad (3)$$

---

*Date:* September 24, 2019.

*2000 Mathematics Subject Classification.* Primary: 60J60, Secondary: 65C30, 65C05.

*Key words and phrases.* Diffusion bridge, FitzHugh-Nagumo model, guided proposal, Langevin sampler, partially observed diffusion, twice-integrated diffusion.

and  $\underline{\sigma}: [0, \infty) \rightarrow \mathbb{R}^{k' \times d'}$ ,  $\underline{\beta}: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^{k'}$  and  $k + k' = d$ . This includes several forms of integrated diffusions.

Suppose  $L$  is a  $m \times d$  matrix with  $m \leq d$ . We aim to simulate the process  $X$ , conditioned on the random variable

$$V = LX_T.$$

The conditional process is termed a diffusion bridge, albeit its paths do not necessarily end at a fixed point but in the set  $\{x: V = Lx\}$ . Besides being an interesting mathematical problem on its own, simulation of such diffusion bridges is key to parameter estimation of diffusions from discrete observations. If the process is observed at discrete times directly or through an observation operator  $L$ , data-augmentation is routinely used for performing Bayesian inference (see for instance [Golightly and Wilkinson \(2006\)](#), [Papaspiliopoulos et al. \(2013\)](#) and [van der Meulen and Schauer \(2017a\)](#)). Here, a key step consists of the simulation of the “missing” data, which amounts to simulation of diffusion bridges.

Another application is nonlinear filtering, where at time  $t$  the state  $X_t$  was observed and at time  $t + T$  a new observation  $LX_{t+T}$  comes in. Interest then lies in sampling from the distribution on  $X_{t+T}$ , conditional on  $(X_t, LX_{t+T})$ . The simulation method developed in this paper can then be used for constructing efficient particle filters. We leave the application of our methods to estimation and filtering to future research, although it is clear that our results can be used directly within the algorithms given in [van der Meulen and Schauer \(2017a\)](#). Finally, rare-event simulation is a third application area for which our results are useful.

We aim for a unified approach, by which we mean that the bridge simulation method applies simultaneously to uniformly elliptic and hypo-elliptic models. This is important, as in the aforementioned estimation problems either one of the two types of ellipticity may apply to the data. While the sample paths of uniformly- and hypo-elliptic diffusions are very different, the corresponding distributions of the observations can be very similar if the diffusion coefficients are close. Algorithms which are invalid for hypo-elliptic diffusions will therefore be numerically unstable if the model is close to being hypo-elliptic, and it may be a priori unknown if this is the case.

**1.1. Literature review.** In case the diffusion is uniformly elliptic and the endpoint is fully observed, i.e.  $L = I$ , the problem has been studied extensively. Cf. [Clark \(1990\)](#), [Durham and Gallant \(2002\)](#), [Beskos et al. \(2006\)](#), [Delyon and Hu \(2006\)](#), [Beskos et al. \(2008\)](#), [Hairer et al. \(2009\)](#), [Lin et al. \(2010\)](#), [Lindström \(2012\)](#), [Bayer and Schoenmakers \(2013\)](#), [Bladt et al. \(2016\)](#), [Schauer et al. \(2017\)](#) and [Whitaker et al. \(2017a\)](#).

Much less is known when either  $L \neq I$  or when the diffusion is not assumed to be uniformly elliptic. In [Beskos et al. \(2008\)](#) and [Hairer et al. \(2009\)](#) a Langevin MCMC sampler is constructed for sampling diffusion bridges when the drift is of the form  $b(x) = Bx + \sigma\sigma'\nabla V(x)$  and  $\sigma$  is constant, assuming uniform ellipticity. Subsequently, this approach was extended to hypo-elliptic diffusions of the form

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} Y_t \\ f(X_t) - Y_t \end{bmatrix} dt + \begin{bmatrix} 0 \\ 1 \end{bmatrix} dW_t$$

in Hairer et al. (2011). However, no simulation results were included to the paper as “these simulations proved prohibitively slow and the resulting method does not seem like a useful approach to sampling” (Hairer et al. (2011), page 671).

We will shortly review in more detail the works Delyon and Hu (2006), Marchand (2012) and van der Meulen and Schauer (2018), as the present work builds upon these. The first of these papers includes some forms of hypo-elliptic diffusions, whereas the latter two papers consider uniformly elliptic diffusions with  $L \neq I$ .

Stramer and Roberts (2007) consider Bayesian estimation of nonlinear continuous-time autoregressive (NLCAR) processes using a data-augmentation scheme. This is a specific class of hypo-elliptic models included by the specification (2)–(3). The method of imputation is however different from what is proposed in this paper.

*Estimation* of discretely observed hypo-elliptic diffusions has been an active field over the past 10 years. As we stated before, within the Bayesian approach a data-augmentation strategy where diffusion bridges are imputed is natural. However, this is by no means the only way for doing estimation. Frequentist approaches to the estimation problem include Sørensen (2012), Ditlevsen and Samson (2017), Lu et al. (2016), Comte et al. (2017), Samson and Thieullen (2012), Pokern et al. (2009), Clairon and Samson (2017) and Melnykova (2018).

**1.2. Review of Delyon and Hu (2006) and Schauer et al. (2017).** To motivate and explain our approach, it is useful to review shortly the methods developed in Delyon and Hu (2006) and Schauer et al. (2017). The method that we propose in this article builds up on these papers. Both of these are restricted to the setting  $L = I$  (full observation of the diffusion at time  $T$ ) and uniform ellipticity. Their common starting point is that under mild conditions the diffusion bridge, obtained by conditioning on  $LX_t = v$ , is a diffusion process itself, governed by the SDE

$$dX_t^* = (b(t, X_t^*) + a(t, X_t^*)r(t, X_t^*)) dt + \sigma(t, X_t^*) dW_t, \quad X_0^* = x_0. \quad (4)$$

Here  $a = \sigma\sigma'$  and  $r(t, x) = \nabla_x \log p(t, x; T, v)$ . We implicitly have assumed the existence of transition densities  $p$  such that  $\mathbb{P}^{(t,x)}(X_T \in A) = \int_A p(t, x; T, \xi) d\xi$  and that  $r(t, x)$  is well defined. The SDE for  $X^*$  can be derived from either Doob’s  $h$ -transform or the theory of initial enlargement of the filtration. Unfortunately, the “guiding” term  $a(t, X_t^*)r(t, X_t^*)$  appearing in the drift of  $X^*$  is intractable, as the transition densities  $p$  are not available in closed form. Henceforth, as direct simulation of  $X^*$  is infeasible, a common feature of both Delyon and Hu (2006) and Schauer et al. (2017) is to simulate a tractable process  $X^\circ$  instead of  $X^*$ , that resembles  $X^*$ . Next, the mismatch can be corrected for by a Metropolis-Hastings step or weighting. The *proposal*  $X^\circ$  (the terminology is inherited from  $X^\circ$  being a proposal for a Metropolis-Hastings step) is assumed to solve the SDE

$$dX_t^\circ = b^\circ(t, X_t^\circ) dt + \sigma(t, X_t^\circ) dW_t, \quad X_0^\circ = x_0, \quad t \in [0, T], \quad (5)$$

where the drift  $b^\circ$  is chosen such that the process  $X_t^\circ$  hits the correct endpoint (say  $v$ ) at the final time  $T$ . Delyon and Hu (2006) proposed to take

$$b^\circ(t, x) = \lambda b(t, x) + (v - x)/(T - t), \quad (6)$$

where either  $\lambda = 0$  or  $\lambda = 1$ , the choice  $\lambda = 1$  requiring the drift  $b$  to be bounded. If  $\lambda = 0$ , a popular discretisation of this SDE is the Modified Diffusion Bridge introduced by [Durham and Gallant \(2002\)](#). A drawback of this method is that the drift is not taken into account. In [Schauer et al. \(2017\)](#) it was proposed to take

$$b^\circ(t, x) = b(t, x) + a(t, x)\tilde{r}(t, x) \quad (7)$$

Here  $\tilde{r}(t, x) = \nabla_x \log \tilde{p}(t, x; T, v)$ , where  $\tilde{p}(t, x)$  is the transition density of an auxiliary diffusion process  $\tilde{X}$  that has tractable transition densities. In this paper, we always assume  $\tilde{X}$  to be a linear process, i.e. a diffusion satisfying the SDE

$$d\tilde{X}_t = \tilde{b}(t, \tilde{X}_t) dt + \tilde{\sigma}(t) dW_t, \quad \text{where } \tilde{b}(t, x) = \tilde{B}(t)x + \tilde{\beta}(t). \quad (8)$$

The process  $X^\circ$  obtained in this way will be referred to as a *guided proposal*.

We denote the laws of  $X$ ,  $X^*$  and  $X^\circ$  viewed as measures on the space  $C([0, t], \mathbb{R}^d)$  of continuous functions from  $[0, t]$  to  $\mathbb{R}^d$  equipped with its Borel- $\sigma$ -algebra by  $\mathbb{P}_t$ ,  $\mathbb{P}_t^*$  and  $\mathbb{P}_t^\circ$  respectively. [Delyon and Hu \(2006\)](#) provided sufficient conditions such that  $\mathbb{P}_T^*$  is absolutely continuous with respect to  $\mathbb{P}_T^\circ$  for the proposals derived from (6). Moreover, closed form expressions for the Radon-Nikodym derivative were derived. For the proposals derived from (7), it was proved in [Schauer et al. \(2017\)](#) that the condition  $\tilde{\sigma}(T)' \tilde{\sigma}(T) = a(T, v)$  is necessary for absolute continuity of  $\mathbb{P}_T^*$  with respect to  $\mathbb{P}_T^\circ$ . We refer to this condition as the *matching condition*, as the diffusivity of  $X$  and  $\tilde{X}$  need to match at the conditioning point. Under that condition (and some additional technical conditions), it was derived that

$$\frac{d\mathbb{P}_T^*(X^\circ)}{d\mathbb{P}_T^\circ(X^\circ)} = \frac{\tilde{p}(0, x_0; T, v)}{p(0, x_0, T, v)} \exp\left(\int_0^T \mathcal{G}(s, X_s^\circ) ds\right),$$

where  $\mathcal{G}(s, x)$  is tractable. A great deal of work in the proof is concerned with proving that  $\|X_t^\circ - v\| \rightarrow 0$  at the ‘‘correct’’ rate.

**1.3. Approach.** We aim to extend the results in [Delyon and Hu \(2006\)](#) and [Schauer et al. \(2017\)](#) by lifting the restrictions of

- (1) uniform ellipticity;
- (2)  $L$  being the identity matrix.

1.3.1. *Extending [Delyon and Hu \(2006\)](#).* We first explain the difficulty in extending this approach beyond uniform ellipticity. To see the problem, we fix  $t < T$ . Absolute continuity of  $\mathbb{P}_t^*$  with respect to  $\mathbb{P}^\circ$  requires the existence of a mapping  $\eta(s, x)$  such that

$$\sigma(s, x)\eta(s, x) = b^*(s, x) - b^\circ(s, x), \quad s \in [0, t], \quad x \in \mathbb{R}^d, \quad (9)$$

which follows from Girsanov’s theorem ([Liptser and Shiryaev \(2001\)](#), Section 7.6.4). However, for the choice of [Delyon and Hu \(2006\)](#) (as given in equation (6)) this mapping  $\eta$  need not exist both in case  $\lambda = 0$  and  $\lambda = 1$ . If  $\lambda = 1$  then we have

$$b^*(s, x) - b^\circ(s, x) = \frac{v - x}{T - s},$$

and therefore  $\eta(s, x)$  only exists if  $v - x$  is in the column space of  $\sigma(s, x)$ . A similar argument applies to the case  $\lambda = 0$ . From these considerations, it is not surprising that

Delyon and Hu (2006) need additional assumptions on the form of the drift to deal with the hypo-elliptic case. More specifically, they consider

$$dX_t = (\sigma(t)h(t, X_t) + B(t)x + \beta(t)) dt + \sigma(t) dW_t, \quad (10)$$

with  $\sigma(t)$  admitting a left-inverse. Then they show that bridges can be obtained by simulating bridges corresponding to this SDE with  $h \equiv 0$ , followed by correcting for the discrepancy by weighting according to their likelihood ratio. Clearly, the form of the drift in the preceding display is restrictive, but necessary for absolute continuity.

Whereas lifting the assumption of uniform ellipticity seems hard, lifting the assumption that  $L = I$  is possible. Indeed, it was shown by Marchand (2012) in a clever way how this can be done by using the guiding term

$$v(t, x) := a(t, x)L'(La(t, x)L')^{-1}\frac{v - Lx}{T - t} \quad (11)$$

to be superimposed on the drift of the original diffusion. Hence, the proposal satisfies the SDE

$$dX_t^\Delta = b(t, X_t^\Delta) dt + v(t, X_t^\Delta) dt + \sigma(t, X_t^\Delta) dW_t, \quad X_0^\Delta = x_0.$$

By applying Ito's lemma to  $(T - t)^{-1}(LX^\Delta(t) - v)$ , followed by the law of the iterated logarithm for Brownian motion, the rate at which  $LX^\Delta(t)$  converges to  $v$  can be derived. Interestingly, the same guiding term as in (11) was used in a specific setting by Arnaudon et al. (2019), where the guiding term was rewritten as  $\sigma(t, x)(L\sigma(t, x))^+(v - Lx)/(T - t)$ , assuming that  $L\sigma$  has linearly independent rows. Here  $A^+$  denotes the Moore-Penrose inverse of the matrix  $A$ . The form of the guiding term in (11) suggests that invertibility of  $La(t, x)L'$  suffices, which, depending on the precise form of  $L$ , would allow for some forms of hypo-ellipticity. However, we believe there are fundamental problems when one wants to include for example integrated diffusions. We return to this in the discussion in section 7.

1.3.2. *Extending Schauer et al. (2017)*. In case  $L$  is not the identity matrix, the conditioned diffusion also satisfies the SDE (4), albeit with an adjusted definition of  $r(t, x)$ . To find the right form of  $r(t, x)$ , assume without loss of generality that  $\text{rank } L = m < d$ . Let  $(f_1, \dots, f_m)$  denote an orthonormal basis of  $\text{Col}(L')$ , and let  $(f_{m+1}, \dots, f_d)$  denote an orthonormal basis of  $\ker L$ . Then for  $A \subset \mathbb{R}^m$

$$\begin{aligned} \mathbb{P}^{(t,x)} \left( X_T \in A \times \mathbb{R}^{d-m} \right) = \\ \int_A \left( \int_{\mathbb{R}^{d-m}} p \left( t, x; T, \sum_{i=1}^d \xi_i f_i \right) d\xi_{m+1}, \dots, d\xi_d \right) d\xi_1, \dots, d\xi_m. \end{aligned}$$

Suppose  $x = \sum_{i=1}^d \xi_i f_i$  is such that  $Lx = v$ . This is equivalent to

$$\sum_{i=1}^m \xi_i Lf_i = v, \quad (12)$$

since  $f_{m+1}, \dots, f_d \in \ker L$ . Hence if  $\xi_1, \dots, \xi_m$  are determined by (12) and if we define

$$\rho(t, x) = \int_{\mathbb{R}^{d-m}} p \left( t, x; T, \sum_{i=1}^d \xi_i f_i \right) d\xi_{m+1}, \dots, d\xi_d,$$

then this is the density of  $X_T | X_t$ , concentrated on the subspace  $LX_T = v$ .

In case  $\text{rank } L = d$ , we can assume without loss of generality that  $L = I$  which is the situation of fully observing  $X_T$ . Summarising, we define

$$\rho(t, x) = \begin{cases} p(t, x; T, v) & \text{if } m = d \\ \int_{\mathbb{R}^{d-m}} p \left( t, x; T, \sum_{i=1}^d \xi_i f_i \right) d\xi_{m+1}, \dots, d\xi_d & \text{if } m < d \end{cases} \quad (13)$$

and let  $r(t, x) = \nabla_x \rho(t, x)$ . The definition of guided proposals in the partially observed hypo-elliptic case is then just as in the uniformly elliptic case with a full observation: replace the intractable transition density  $p$  appearing in the definition of  $\rho$  by  $\tilde{p}$  to yield  $\tilde{\rho}$ . Then define

$$\tilde{r}(t, x) = \nabla_x \log \tilde{\rho}(t, x)$$

and let the process  $X^\circ$  be defined by equation (5) with  $b^\circ(t, x) = b(t, x) + a(t, x)\tilde{r}(t, x)$ . For  $t < T$ , it is conceivable that  $\mathbb{P}_t^\star$  is absolutely continuous with respect to  $\mathbb{P}_t^\circ$  because clearly equation (9) is solved by  $\eta(s, x) = \sigma(s, x)(r(s, x) - \tilde{r}(s, x))$ . Contrary to the hypo-elliptic setting in [Delyon and Hu \(2006\)](#), no specific form of the drift needs to be imposed here. However, it is not clear whether

- $\|LX_t^\circ - v\|$  tends to zero as  $t \uparrow T$ ;
- $\mathbb{P}_T^\star \ll \mathbb{P}_T^\circ$ .

The two main results of this paper (Proposition 2.8 and Theorem 2.14) provide conditions such that this is indeed the case. Interestingly, in the hypo-elliptic case the necessary “matching condition” on the parameters of the auxiliary process  $\tilde{X}$  not only involves its diffusion coefficient  $\tilde{\sigma}(t)$ , but its drift  $\tilde{b}(t, x)$  as well. In particular, simply equating  $\tilde{b}$  to zero turns the measures  $\mathbb{P}_T^\star$  and  $\mathbb{P}_T^\circ$  mutually singular. For deriving the rate at which  $\|LX_t^\circ - v\|$  decays we employ a completely different method of proof compared to the analogous result in [Schauer et al. \(2017\)](#), using techniques detailed in [Mao \(1997\)](#). While the proof of the absolute continuity result is along the lines of that in [Schauer et al. \(2017\)](#), having a partial observation and hypo-ellipticity requires nontrivial adaptations of that proof.

*Put shortly, our results show that guided proposals can be defined for partially observed hypo-elliptic diffusions exactly as in [Schauer et al. \(2017\)](#), if an extra restriction on the drift  $\tilde{b}$  of the auxiliary process  $\tilde{X}$  is taken into account.*

Whereas most of the results are derived for  $\sigma$  depending on the state  $x$ , the applicability of our methods is mostly confined to the case where  $\sigma$  is only allowed to depend on  $t$ . The difficulty lies in checking the fourth inequality of Assumption 2.7 appearing in Section 2. On the other hand, numerical experiments can give insight whether the law of a particular proposal process and the law of the conditional process are equivalent.

Examples of hypo-elliptic diffusion processes that fall into our setup include

- (1) integrated diffusions, when either the rough, smooth, or both components are observed;

- (2) higher order integrated diffusions;
- (3) NLCAR models;
- (4) the class of hypo-elliptic diffusions considered in [Hairer et al. \(2011\)](#).

These examples are listed here for illustration purpose. We stress that the derived results are more general.

Whereas some examples that we discuss can be treated by the approach of [Delyon and Hu \(2006\)](#) (which is restricted to SDEs of the form (10)), our approach extends well beyond this class of models (see for instance Example 3.8). Moreover, the hypo-elliptic bridges proposed by [Delyon and Hu \(2006\)](#) are bridges of a linear process, whereas the bridges we propose *only* use a linear process to derive the guiding term that is superimposed on the true drift. This means that only our approach is able to incorporate nonlinearity in the drift of the proposal.

**1.4. A toy problem.** Here we first consider a two-dimensional uniformly elliptic diffusion with unit diffusion coefficient, which is fully observed. Upon taking  $\tilde{b} \equiv 0$  and  $\tilde{\sigma} = \sigma$ , we have

$$dX_t^\circ = b(t, X_t^\circ) dt + \frac{v - X_t^\circ}{T - t} dt + dW_t.$$

The guiding term can be viewed as the distance left to be covered,  $v - X_t^\circ$ , divided by the remaining time  $T - t$ . This simple expression is to be contrasted to a hypo-elliptic diffusion, the simplest example perhaps being an integrated diffusion, with both components observed, i.e. a diffusion with

$$b(t, x) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x =: Bx \quad \text{and} \quad \sigma = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

It follows from the results in this paper that using guided proposals we obtain an “exact” proposal, i.e.  $X_t^\circ = X_t^*$  upon taking  $\tilde{B} = B$ ,  $\tilde{\beta} \equiv 0$  and  $\tilde{\sigma} = \sigma$ . The SDE for  $X^\circ$  takes the form

$$dX_t^\circ = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} X_t^\circ dt + \begin{bmatrix} 0 \\ 18 \frac{v_1 - X_{t,1}^\circ}{(T-t)^2} + \frac{10v_2 - 28X_{t,2}^\circ}{T-t} \end{bmatrix} dt + \begin{bmatrix} 0 \\ 1 \end{bmatrix} dW_t,$$

(where  $X_{t,i}$  and  $v_i$  denote the  $i$ -th component of  $X_t$  and  $v$  respectively). This is an elementary consequence of the process being Gaussian and follows for example directly as a special case of either Lemma 2.5 or Equation (10).

Even for this relatively simple case the guiding term behaves radically different compared to the uniformly elliptic case. The pulling term only acts on the rough coordinate and is not any longer inversely proportional to the remaining time. This illustrates the inherent difficulty of the problem and explains the centring and scaling of  $X^\circ$  that we will introduce for studying its behaviour.

**1.5. Outline.** In Section 2 we present the main results of the paper. We illustrate the main theorems by applying it to various forms of partially conditioned hypo-elliptic diffusions in Section 3. In Section 4 we illustrate our work with simulation examples for the FitzHugh-Nagumo model and a twice integrated diffusion model. The proof of the proposition on the behaviour of  $X^\circ$  near the end-point is given in Section 5 and the proof



of the theorem on absolute continuity is given in Section 6. We end with a discussion in Section 7. Some technical and additional results are gathered in the Appendix.

### 1.6. Frequently used notation.

1.6.1. *Inequalities.* We use the following notation to compare two sequences  $\{a_n\}$  and  $\{b_n\}$  of positive real numbers:  $a_n \lesssim b_n$  (or  $b_n \gtrsim a_n$ ) means that there exists a constant  $C > 0$  that is independent of  $n$  and is such that  $a_n \leq Cb_n$ . As a combination of the two we write  $a_n \asymp b_n$  if both  $a_n \lesssim b_n$  and  $a_n \gtrsim b_n$ . We will also write  $a_n \gg b_n$  to indicate that  $a_n/b_n \rightarrow \infty$  as  $n \rightarrow \infty$ . By  $a \vee b$  and  $a \wedge b$  we denote the maximum and minimum of two numbers  $a$  and  $b$  respectively.

1.6.2. *Linear algebra.* We denote the smallest and largest eigenvalue of a square matrix  $A$  by  $\lambda_{\min}(A)$  and  $\lambda_{\max}(A)$  respectively. The  $p \times p$  identity matrix is denoted by  $I_p$ . The  $p \times q$  matrix with all entries equal to zero is denoted by  $0_{p \times q}$ . For matrices we use the spectral norm, which equals the largest singular value of the matrix. The determinant of the matrix  $A$  is denoted by  $|A|$  and the trace by  $\text{tr}(A)$ .

1.6.3. *Stochastic processes.* For easy reference, the following table summaries the various processes around. The rightmost three columns give the drift, diffusion coefficient and measure on  $C([0, t], \mathbb{R}^d)$  respectively.

|             |  |             |                  |                        |
|-------------|--|-------------|------------------|------------------------|
| $X$         | original, unconditioned diffusion process, defined by (1)  | $b$         | $\sigma$         | $\mathbb{P}_t$         |
| $X^*$       | corresponding bridge, conditioned on $v$ , defined by (14)   | $b^*$       | $\sigma$         | $\mathbb{P}_t^*$       |
| $X^\circ$   | proposal process defined by (5)  | $b^\circ$   | $\sigma$         | $\mathbb{P}_t^\circ$   |
| $\tilde{X}$ | linear process defined by (8) whose transition densities $\tilde{p}$ appear in the definition of $X^\circ$ | $\tilde{b}$ | $\tilde{\sigma}$ | $\tilde{\mathbb{P}}_t$ |

We write

$$a(t, x) = \sigma(t, x)\sigma(t, x)' \quad \text{and} \quad \tilde{a}(t) = \tilde{\sigma}(t)\tilde{\sigma}(t)'.$$

The state-space of  $X$ ,  $X^*$  and  $X^\circ$  is  $\mathbb{R}^d$ . The Wiener process lives on  $\mathbb{R}^d$ . The observation is determined by the  $m \times d$  matrix  $L$ . Finally, the orthonormal basis  $\{f_1, \dots, f_d\}$  for  $\mathbb{R}^d$  defined in Section 1.3.2 is fixed throughout, as are the numbers  $\xi_1, \dots, \xi_m$  defined via Equation (12).

## 2. MAIN RESULTS

Throughout, we assume

**Assumption 2.1.** Both  $b$  and  $\sigma$  are globally Lipschitz continuous in both arguments.

This ensures that a strong solution to the SDE (1) exists. We define the conditioned process, denoted by  $X^*$  to be a diffusion process satisfying the SDE

$$dX_t^* = b(t, X_t^*) dt + a(t, X_t^*)r(t, X_t^*) dt + \sigma(t, X_t^*) dW_t, \quad X_0^* = x_0. \quad (14)$$

Here  $r(t, x) = \nabla_x \log \rho(t, x)$ . A derivation is given in section D.

**Assumption 2.2.** The process  $X$  has transition densities such that the mapping  $\rho: \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}$  is  $C^{\infty, \infty}$  and strictly positive for all  $s < T$  and  $x \in \mathbb{R}^d$ .

For fixed  $x \in \mathbb{R}^d$ ,  $s$  and  $t > s + \varepsilon$ , the mapping  $(t, y) \rightarrow p(s, x; t, y)$  is continuous and bounded.

In general Assumption 2.2 is established by verifying Hörmander's hypoellipticity conditions; see Williams (1981). The assumption is satisfied in particular under suitable conditions for the diffusion as described by equations (2) and (3). Note that the results in this paper are not limited to this special case.

**Proposition 2.3.** Suppose that the matrix valued function  $t, x \mapsto \underline{\sigma}$  in the hypo-elliptic model given by (2) and (3) has rank  $k'$  for all  $(t, x)$ . Furthermore suppose that  $\underline{\sigma}$  and  $\underline{\beta}$  are infinitely often differentiable with respect to  $(t, x)$ . Then the process  $(X_t)$  admits a smooth (i.e.  $C^\infty$ ) density which is also smooth with respect to the initial condition.

*Proof.* This is a special case of Proposition C.1 in Appendix C.  $\square$

**2.1. Existence of guided proposals.** The guiding term of  $X^\circ$  involves  $\tilde{r}: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}$ . In the uniformly elliptic case it is easily verified that this mapping is well defined. This need not be the case in the hypo-elliptic setting.

Let  $\Phi(t)$  denote the fundamental matrix solution of the ODE  $d\Phi(t) = \tilde{B}(t)\Phi(t) dt$ ,  $\Phi(0) = I$  and set  $\Phi(t, s) = \Phi(t)\Phi(s)^{-1}$ . Define

$$L(t) := L\Phi(T, t). \tag{15}$$

**Assumption 2.4.** The matrix

$$\int_t^T \Phi(T, s)\tilde{a}(s)\Phi(T, s)' ds$$

is strictly positive definite for  $t < T$ .

In the uniformly elliptic setting, this assumption is always satisfied. Under this assumption, the matrix

$$M^\dagger(t) := \int_t^T L(s)\tilde{a}(s)L(s)' ds$$

is also strictly positive definite for all  $t \in [0, T)$  and, in particular, invertible.

**Lemma 2.5.** If Assumption 2.4 holds then,

$$\tilde{r}(t, x) = L(t)'M(t)(v - \mu(t) - L(t)x), \quad t \in [0, T], \tag{16}$$

where

$$\mu(t) = \int_t^T L(s)\tilde{\beta}(s) ds$$

and

$$M(t) = [M^\dagger(t)]^{-1}. \tag{17}$$

*Proof.* The solution to the SDE for  $\tilde{X}_u$  is given by

$$\tilde{X}_u = \Phi(u, t)x + \int_t^u \Phi(u, s)\tilde{\beta}(s) ds + \int_t^u \Phi(u, s)\tilde{\sigma}(s) dW_s, \quad u \geq t, \quad \tilde{X}_t = x.$$

Cf. [Liptser and Shiryaev \(2001\)](#), Theorem 4.10. The result now follows directly upon taking  $u = T$ , multiplying both sides with  $L$  and using the definition of  $L(t)$ .  $\square$

In [Appendix A](#) easily verifiable conditions for the existence of  $\tilde{p}$  are given for the case  $L = I$ .

Since  $t \mapsto \mu(t)$  and  $t \mapsto M(t)$  are continuous and  $x \mapsto \tilde{r}(t, x)$  is linear in  $x$  for fixed  $t$ , the process  $X^\circ$  is well defined on intervals bounded away from  $T$ .

**Lemma 2.6.** Under [Assumptions 2.1](#) and [2.4](#) we have that for any  $t < T$ , the SDE for  $X^\circ$  has a unique strong solution on  $[0, t]$ .

*Throughout, without explicitly stating it in lemmas and theorems, we will assume that [Assumptions 2.1](#), [2.2](#) and [2.4](#) hold true.*

**2.2. Behaviour of guided proposals near the endpoint.** Let  $\Delta(t)$  be an invertible  $m \times m$  diagonal matrix-valued measurable function on  $[0, T)$ . Define

$$Z_{\Delta, t} = \Delta(t)(v - \mu(t) - L(t)X_t^\circ) \quad (18)$$

and

$$L_\Delta(t) = \Delta(t)L(t) \quad M_\Delta(t) = \Delta(t)^{-1}M(t)\Delta(t)^{-1}. \quad (19)$$

Note that for  $t \approx T$ , we have  $\Phi(T, t) \approx I$  and hence  $Z_{\Delta, t} \approx \Delta(t)(v - LX_t^\circ)$ . The matrix  $\Delta(t)$  is a scaling matrix which in the hypo-elliptic case incorporates the difference in rate of convergence for smooth and rough components of  $LX_t^\circ$  to  $v$ , when  $t \uparrow T$ . In the uniformly elliptic case, we can always take  $\Delta(t) = I_m$ .

The following assumption is of key importance.

**Assumption 2.7.** There exists an invertible  $m \times m$  diagonal matrix-valued function  $\Delta(t)$ , which is measurable on  $[0, T)$ , a  $t_0 < T$ ,  $\alpha \in (0, 1]$  and positive constants  $\underline{c}$ ,  $\bar{c}$ ,  $c_1$ ,  $c_2$  and  $c_3$  such that for all  $t \in [t_0, T)$

$$\begin{aligned} \underline{c}(T-t)^{-1} &\leq \lambda_{\min}(M_\Delta(t)) \leq \lambda_{\max}(M_\Delta(t)) \leq \bar{c}(T-t)^{-1}, \\ \left\| L_\Delta(t) \left( \tilde{b}(t, x) - b(t, x) \right) \right\| &\leq c_1 \\ \text{tr} \left( L_\Delta(t) a(t, x) L_\Delta(t)' \right) &\leq c_2 \\ \|L_\Delta(t)(\tilde{a}(t) - a(t, x))L_\Delta(t)'\| &\leq c_3(T-t)^\alpha. \end{aligned} \quad (20)$$

**Proposition 2.8.** Under [Assumption 2.7](#), there exists a positive number  $C$  such that

$$\limsup_{t \uparrow T} \frac{\|Z_{\Delta, t}\|}{\sqrt{(T-t) \log(1/(T-t))}} \leq C \quad \text{a.s.}$$

**Remark 2.9.** If  $\sigma$  is state-dependent, it is particularly difficult to ensure that the fourth inequality in [\(20\)](#) is satisfied. There is at least one non-trivial example where this inequality can be assured (see [Example 2.12](#)). In [Section 7](#) we further discuss the case of state dependent diffusivity. In the simpler case where  $\sigma$  only depends on  $t$ , we can

always take  $\tilde{\sigma}(t) = \sigma(t)$  and then the fourth inequality is trivially satisfied. In Section 3 we verify (20) for a wide range of examples. As a prelude: for the SDE system specified by (2) and (3) one takes  $\tilde{B} = B$  and  $\tilde{\sigma} = \sigma$ . Then  $\Delta(t)$  can be chosen such that the first inequality is satisfied. The second condition of (20) encapsulates a matching condition on the drift which induces some restrictions on  $\tilde{\beta}$  and  $\underline{\beta}$ . The third inequality is then usually satisfied automatically.

The uniformly elliptic case is particularly simple:

**Corollary 2.10** (Uniformly elliptic case). Assume that either (i) the diffusivity  $\sigma$  is constant and  $\tilde{\sigma} = \sigma$  or (ii)  $\sigma$  depends on  $t$  and  $\tilde{\sigma}(t) = \sigma(t)$ . Assume  $a$  is strictly positive definite and that  $b(t, x) - \tilde{b}(t, x)$  is bounded on  $[0, T] \times \mathbb{R}^d$ . Then the conclusion of Proposition 2.8 holds true with  $\Delta(t) = I_m$ .

**Remark 2.11.** The behaviour of  $\|Z_{\Delta, t}\|$  that we obtain agrees with the results of Schauer et al. (2017). That paper is confined to  $L = I$  and the uniformly elliptic case, but includes the case of state-dependent diffusion coefficient. Under this assumption, it suffices that  $\tilde{\sigma}(T)' \tilde{\sigma}(T) = a(T, v)$ , a condition that can always be ensured to be satisfied.

The proofs of Theorem 2.8 and Corollary 2.10 are given in Section 5.

In Section 3 we give a set of tractable hypo-elliptic models for which the conclusion of Theorem 2.8 is valid. The appropriate choice of the scaling matrix  $\Delta(t)$  is really problem specific. Moreover, the assumptions on the auxiliary process depend on the choice of  $L$ .

In most cases it will not be possible to satisfy the fourth inequality of Assumption 2.7 when the diffusion coefficient is state-dependent. The following example shows an exception.

**Example 2.12.** Suppose the diffusion is uniformly elliptic and  $L = \begin{bmatrix} \underline{L} & 0_{m \times k'} \end{bmatrix}$ , where  $L \in \mathbb{R}^{m \times k}$  and  $d = k + k'$ . Now suppose  $a(t, x)$  is of block form:

$$a(t, x) = \begin{bmatrix} a_{11}(t) & 0_{k \times k'} \\ 0_{k' \times k} & a_{22}(t, x) \end{bmatrix}$$

and that we take  $\tilde{a}$  to be of the same block form. Upon taking  $\tilde{B} = 0_{d \times d}$  and  $\Delta(t) = I_d$ , we see that  $L_{\Delta}(t) = L$  and hence

$$L_{\Delta}(t)(\tilde{a}(t) - a(t, x))L_{\Delta}(t)' = \underline{L}(\tilde{a}_{11}(t) - a_{11}(t))\underline{L}'.$$

Therefore, if we choose  $\tilde{a}_{11}(t)$  to be equal to  $a_{11}(t)$  the fourth inequality in Assumption 2.7 is trivially satisfied.

Empirically however, it appears that Assumption 2.7 is stronger than needed for valid guided proposals, see Example 4.5.

**2.3. Absolute continuity.** The following theorem gives sufficient conditions for absolute continuity of  $\mathbb{P}_T^*$  with respect to  $\mathbb{P}_T^{\circ}$ . First we introduce an assumption.

**Assumption 2.13.** There exists a constant  $C$  such that

$$p(s, x; t, y) \leq C\tilde{p}(s, x; t, y) \quad 0 \leq s < t < T \quad (21)$$

for all  $x, y \in \mathbb{R}^d$ .

**Theorem 2.14.** Assume there exists a positive  $\delta$  such that  $|\Delta(t)| \lesssim (T-t)^{-\delta}$ . If Assumptions 2.7 and 2.13 hold true, then

$$\frac{d\mathbb{P}_T^*}{d\mathbb{P}_T^\circ}(X^\circ) = \frac{\tilde{\rho}(0, x_0)}{\rho(0, x_0)} \Psi_T(X^\circ),$$

where

$$\Psi_t(X^\circ) = \exp\left(\int_0^t \mathcal{G}(s, X_s^\circ) ds\right), \quad (22)$$

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

and  $\tilde{H}(s) = L(s)'M(s)L(s)$ .

The proof is given in Section 6.

**Remark 2.15.** The expression for the Radon-Nikodym derivative does depend on the intractable transition densities  $p$  via the term  $\rho(0, x_0)$ . This is a multiplicative term that only shows up in the denominator and therefore cancels in the acceptance probability for sampling diffusion bridges using the Metropolis-Hastings algorithm.

The following lemma is useful for verifying Assumption 2.13. Its proof is located in Section 6.

**Lemma 2.16.** Assume  $\eta(s, x)$  satisfies the equation

$$\sigma(s, x)\eta(s, x) = b(s, x) - \tilde{b}(s, x)$$

and that  $\eta$  is bounded. Then there exists a constant  $C$  such that

$$p(s, x; t, y) \leq C\tilde{p}(s, x; t, y)$$

for all  $x, y \in \mathbb{R}^d$  for all  $0 \leq s < t \leq T$ .

### 3. TRACTABLE HYPO-ELLIPTIC MODELS

In this section we give several examples of hypo-elliptic models that satisfy Assumption 2.7. In the following we write  $X_t = [X_{t,1} \ \cdots \ X_{t,d}]'$ .

In each of the examples we choose an appropriate scaling matrix  $\Delta(t)$  and verify the conditions of Assumption 2.7. For this, we need to evaluate  $L_\Delta(t)$  and  $M_\Delta(t)$ . The computations are somewhat tedious by hand (though straightforward), and for that reason we used the computer algebra system Maple for this. Ideally, instead of the conditions appearing in Assumption 2.7, one would like to have conditions only containing  $b$ ,  $\tilde{b}$ ,  $\sigma$  and  $\tilde{\sigma}$ . This however seems hard to obtain and maybe a bit too much to ask for, given the wide diversity in behaviour of hypo-elliptic diffusions and the generality of the matrix  $L$ . In each of the examples, we state the model and the conditions on  $\tilde{b}$  and  $\tilde{\sigma}$  such that Assumption 2.7 is satisfied.

**Example 3.1. Integrated diffusion, fully observed.** Suppose

$$\begin{aligned} dX_{t,1} &= X_{t,2} dt \\ dX_{t,2} &= \underline{\beta}(t, X_t) dt + \gamma dW_t, \end{aligned} \tag{23}$$

where  $\underline{\beta}: [0, T] \times \mathbb{R}^2 \rightarrow \mathbb{R}$  is bounded and globally Lipschitz in both arguments. If  $L = I_2$ , and the coefficients of the auxiliary process  $\tilde{X}$  satisfy

$$\tilde{B}(t) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \tilde{\beta}_1(t) = 0, \quad \tilde{\sigma} = \begin{bmatrix} 0 \\ \gamma \end{bmatrix}$$

then Assumption 2.7 is satisfied.

*Proof:* As we expect the rate of the first component, which is smooth, to converge to the endpoint one order faster than the second component, which is rough, we take

$$\Delta(t) = \begin{bmatrix} (T-t)^{-1} & 0 \\ 0 & 1 \end{bmatrix}.$$

We have

$$b(t, x) - \tilde{b}(t, x) = \begin{bmatrix} 0 \\ \underline{\beta}(t, x) \end{bmatrix}.$$

By choice of  $\tilde{B}$  and  $\Delta$  we get

$$L_\Delta(t) = \Delta(t)L\Phi(T, t) = \Delta(t)\Phi(T, t) = \begin{bmatrix} 1/(T-t) & 1 \\ 0 & 1 \end{bmatrix}.$$

and

$$M(t) = \frac{1}{\gamma^2} \begin{bmatrix} 12/(T-t)^3 & 6/(T-t)^2 \\ 6/(T-t)^2 & 4/(T-t) \end{bmatrix} \implies M_\Delta(t) = \frac{1}{\gamma^2} (T-t)^{-1} \begin{bmatrix} 12 & 6 \\ 6 & 4 \end{bmatrix}.$$

Now it is trivial to verify that Assumption 2.7 is satisfied.

**Example 3.2. Integrated diffusion, smooth component observed.** Consider the same setting as in the previous example, but now with  $L = \begin{bmatrix} 1 & 0 \end{bmatrix}$ . That is, only the smooth (integrated) component is observed. Then Assumption 2.7 is satisfied.

*Proof:* Upon taking  $\Delta(t) = (T-t)^{-1}$  we get

$$M_\Delta(t) = 3\gamma^{-2}(T-t)^{-1} \quad \text{and} \quad L_\Delta(t) = \begin{bmatrix} 1/(T-t) & 1 \end{bmatrix} \tag{24}$$

from which the claim easily follows

**Example 3.3. Integrated diffusion, rough component observed.** Consider the same setting as in Example 3.1, but now with  $L = \begin{bmatrix} 0 & 1 \end{bmatrix}$ . That is, only the rough component is observed. Then Assumption 2.7 is satisfied.

*Proof:* Taking  $\Delta(t) = 1$  we get  $M_\Delta(t) = \gamma^{-2}(T-t)^{-1}$  and  $L_\Delta(t) = \begin{bmatrix} 0 & 1 \end{bmatrix}$ , from which the claim easily follows.

The guiding term is completely independent of the first component. This is not surprising, as this example is equivalent to fully observing a one-dimensional uniformly elliptic diffusion (described by the second component).

**Example 3.4. NLCAR( $p$ )-model.** The integrated diffusion model is a special case of the class of continuous-time nonlinear autoregressive models (Cf. [Stramer and Roberts \(2007\)](#)). The real valued process  $Y$  is called a  $p$ -th order NLCAR model if it solves the  $p$ -th order SDE

$$dY_t^{(p-1)} = \underline{\beta}(t, Y_t) dt + \gamma dW_t.$$

We assume  $\underline{\beta}: [0, T] \times \mathbb{R}^2 \rightarrow \mathbb{R}$  is bounded and globally Lipschitz in both arguments. This example corresponds to the model specified by (2)–(3) with  $d = p$ ,  $d' = 1$  and  $k = p - 1$ . Integrated diffusions correspond to  $p = 2$ . Observing only the smoothest component means that we have  $L = [1 \ 0_{1 \times d-1}]$ . This class of models includes in particular continuous-time autoregressive and continuous-time threshold autoregressive models, as defined in [Brockwell \(1994\)](#).

We consider the NCLAR(3)-model more specifically, which can be written explicitly as a diffusion in  $\mathbb{R}^3$  with

$$b(t, x) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ 0 \\ \underline{\beta}(t, x) \end{bmatrix} \quad \sigma = \begin{bmatrix} 0 \\ 0 \\ \gamma \end{bmatrix}. \quad (25)$$

If either  $L = I_3$  or  $L = [1 \ 0 \ 0]$ , Assumption 2.7 is satisfied if the coefficients of the auxiliary process  $\tilde{X}$  satisfy

$$\tilde{B}(t) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad \tilde{\beta}_1(t) = \tilde{\beta}_2(t) = 0, \quad \tilde{\sigma} = \begin{bmatrix} 0 \\ 0 \\ \gamma \end{bmatrix}. \quad (26)$$

*Proof:* If  $L = I_3$  we take the scaling matrix

$$\Delta(t) = \begin{bmatrix} (T-t)^{-2} & 0 & 0 \\ 0 & (T-t)^{-1} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

to account for the different degrees of smoothness of the paths of the diffusion. We then obtain, defining  $w(t) = (T-t)^{-1}$

$$M_\Delta(t) = \frac{3w(t)}{\gamma^2} \begin{bmatrix} 240 & -120 & 20 \\ -120 & 64 & -12 \\ 20 & -12 & 3 \end{bmatrix} \quad \text{and} \quad L_\Delta(t) = \begin{bmatrix} w(t)^2 & w(t) & 1/2 \\ 0 & w(t) & 1 \\ 0 & 0 & 1 \end{bmatrix},$$

from which the claim is easily verified.

In case  $L = [1 \ 0 \ 0]$ , we take  $\Delta(t) = (T-t)^{-2}$  and since

$$M_\Delta(t) = \frac{20}{\gamma^2} (T-t)^{-1} \quad \text{and} \quad L_\Delta(t) = [(T-t)^{-2} \ (T-t)^{-1} \ 1/2].$$

Assumption 2.7 is satisfied. See Example 4.1 for a numerical illustration of this example.

**Example 3.5.** Assume the following model for FM-demodulation:

$$d \begin{bmatrix} X_{t,1} \\ X_{t,2} \\ X_{t,2} \end{bmatrix} = \begin{bmatrix} X_{t,2} \\ -\alpha X_{t,2} \\ \sqrt{2\gamma} \sin(\omega t + X_{t,1}) \end{bmatrix} dt + \begin{bmatrix} 0 & 0 \\ \sqrt{2\gamma\alpha} & 0 \\ 0 & \psi \end{bmatrix} d \begin{bmatrix} W_{t,1} \\ W_{t,2} \end{bmatrix}.$$

Here, the observation is determined by  $L = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$ . Motivated by this example, we check our results for a diffusion with coefficients

$$b(t, x) = Bx + \begin{bmatrix} 0 \\ \underline{\beta}_2(t, x) \\ \underline{\beta}_3(t, x) \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 & 0 \\ 0 & -\alpha & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \sigma = \begin{bmatrix} 0 & 0 \\ \gamma_1 & \gamma_2 \\ \gamma_3 & \gamma_4 \end{bmatrix}.$$

Note that this is a slight generalisation of the FM-demodulation model. We will assume that  $\gamma_3^2 + \gamma_4^2 \neq 0$  and  $\underline{\beta}_3$  to be bounded. If  $\tilde{B}(t) = B$  and  $\tilde{\sigma} = \sigma$ , then Assumption 2.7 is satisfied.

*Proof:* As the observation is on the rough component, we choose  $\Delta(t) = 1$ . We have

$$M_\Delta(t) = (T - t)^{-1}(\gamma_3^2 + \gamma_4^2)^{-1}$$

and  $L_\Delta(t) = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$ . Hence  $L_\Delta(t) \left( \tilde{b}(t) - b(t, x) \right) = \tilde{\beta}_3(t) - \underline{\beta}_3(t, x)$  and the other conditions are easily verified.

**Example 3.6.** Assume  $[X_{t,1} \ X_{t,2}]'$  gives the position in the plane of a particle at time  $t$ . Suppose the velocity vector of the particle at time  $t$ , denoted by  $[X_{t,3} \ X_{t,4}]'$  satisfies a SDE driven by a 2-dimensional Wiener process. The evolution of  $X_t = [X_{t,1} \ X_{t,2} \ X_{t,3} \ X_{t,4}]'$  is then described by the SDE

$$\begin{aligned} dX_{t,1} &= X_{t,3} dt \\ dX_{t,2} &= X_{t,4} dt \\ d \begin{bmatrix} X_{t,3} \\ X_{t,4} \end{bmatrix} &= \begin{bmatrix} \underline{\beta}_3(t, X_t) \\ \underline{\beta}_4(t, X_t) \end{bmatrix} dt + \gamma dW_t, \end{aligned}$$

where  $W_t \in \mathbb{R}^2$ . This example corresponds to the case  $d = 4$ ,  $d' = 2$  and  $k = 2$  in the model specified by (2)–(3). Observing only the location corresponds to

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

In matrix-vector notation the drift of the diffusion is given by  $b(t, x) = Bx + \beta(t, x)$ , where

$$B = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \beta(t, x) = \begin{bmatrix} 0 \\ 0 \\ \underline{\beta}_3(t, X_t) \\ \underline{\beta}_4(t, X_t) \end{bmatrix}.$$

We will assume diffusion coefficient  $\sigma = \begin{bmatrix} 0 & 0 & \gamma_1 & \gamma_3 \\ 0 & 0 & \gamma_2 & \gamma_4 \end{bmatrix}'$ , where  $\gamma_1\gamma_4 - \gamma_2\gamma_3 \neq 0$ . If  $\tilde{\beta}_1(t) = \tilde{\beta}_2(t) = 0$ ,  $\tilde{B}(t) = B$  and  $\tilde{\sigma} = \sigma$ , then Assumption 2.7 is satisfied.

*Proof:* As we observed the first two coordinates, which are both smooth, we take  $\Delta(t) = (T - t)^{-1}I_2$ . The claim now follows from

$$M_\Delta(t) = (T - t)^{-1} \frac{3}{(\gamma_1\gamma_4 - \gamma_2\gamma_3)^2} \begin{bmatrix} -\gamma_3^2 - \gamma_4^2 & \gamma_1\gamma_3 + \gamma_2\gamma_4 \\ \gamma_1\gamma_3 + \gamma_2\gamma_4 & -\gamma_1^2 - \gamma_2^2 \end{bmatrix}$$



and  $L_\Delta(t) = \begin{bmatrix} (T-t)^{-1} & 0 & 1 & 0 \\ 0 & (T-t)^{-1} & 0 & 1 \end{bmatrix}$ .

**Example 3.7.** [Hairer et al. \(2011\)](#) consider SDEs of the form

$$dX_t = \begin{bmatrix} 0 & 1 \\ 0 & \theta \end{bmatrix} X_t dt + \begin{bmatrix} 0 \\ \underline{\beta}(t, X_t) \end{bmatrix} dt + \begin{bmatrix} 0 \\ \gamma \end{bmatrix} dW_t,$$

where  $X_t = [X_{t,1} \ X_{t,2}]'$ ,  $\theta > 0$  and the conditioning is specified by  $L = [1 \ 0]$ . As explained in [Hairer et al. \(2011\)](#) the solution to this SDE can be viewed as the time evolution of the state of a mechanical system with friction under the influence of noise. Assume  $(t, x) \mapsto \underline{\beta}(t, x)$  is bounded and Lipschitz in both arguments. Note that this hypo-elliptic SDE is not of the form given in (2) and (3). However, if

$$\tilde{B}(t) = B_\theta = \begin{bmatrix} 0 & 1 \\ 0 & \theta \end{bmatrix}, \quad \tilde{\beta}_1(t) = 0 \quad \text{and} \quad \tilde{\sigma} = \begin{bmatrix} 0 \\ \gamma \end{bmatrix},$$

then Assumption 2.7 is satisfied.

*Proof:* Upon taking  $\Delta(t) = (T-t)^{-1}$ , we find that

$$\lim_{t \uparrow T} (T-t)M_\Delta(t) = 3\gamma^{-2} \quad \text{and} \quad L_\Delta(t) = \begin{bmatrix} \frac{1}{T-t} & \frac{e^{\theta(T-t)} - 1}{\theta(T-t)} \end{bmatrix}.$$

This is to be compared with the expressions in (24). We conclude similarly as in Example 3.2 that the conditions in Assumption 2.7 are satisfied.

**Example 3.8.** This is an example to illustrate that our approach applies beyond equations of the form (10). We assume

$$dX_t = BX_t dt + \begin{bmatrix} 0 \\ \underline{\beta}(t, X_t) \end{bmatrix} dt + \begin{bmatrix} 1 \\ 1 \end{bmatrix} dW_t$$

with  $\underline{\beta}: [0, T] \times \mathbb{R}^2 \rightarrow \mathbb{R}$  bounded and globally Lipschitz in both arguments. Suppose  $L = [1 \ 0]$ . If

$$\tilde{B} = B := \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \tilde{\beta} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \tilde{\sigma} = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

then 2.7 holds.

*Proof:* Using  $\Delta(t) = 1$ , we have  $L_\Delta(t) = [1 \ T-t]$ ,  $\lim_{t \uparrow T} (T-t)M_\Delta(t) = 1$  and the claim follows as in the previous examples.

#### 4. NUMERICAL ILLUSTRATIONS

In this section we will discuss implementational aspects of our sampling method, and we will illustrate the method by some representative numerical examples. We implemented the examples as parts of the authors' software package [Bridge](#) ([Schauer et al. \(2018\)](#)), written in the programming language [Julia](#), ([Bezanson et al. \(2012\)](#).) The corresponding code is available, [van der Meulen and Schauer \(2019\)](#).

For computing the guiding term and likelihood ratio, we have the following backwards ordinary differential equations

$$dL(t) = -L(t)\tilde{B}(t) dt, \quad L(T) = L \quad (27)$$

$$dM^\dagger(t) = -L(t)\tilde{a}(t)L(t)' dt, \quad M^\dagger(T) = 0_{m \times m} \quad (28)$$

$$d\mu(t) = -L(t)\tilde{\beta}(t) dt, \quad \mu(T) = 0_{m \times 1}, \quad (29)$$

where  $t \in [0, T]$ . These are easily derived, cf. lemma 2.4 in [van der Meulen and Schauer \(2017b\)](#). These backward differential equations need only be solved once. Next, Algorithm 1 from [van der Meulen and Schauer \(2017a\)](#) can be applied. This algorithm describes a Metropolis-Hastings sampler for simulating diffusion bridges using guided proposals. We briefly recap the steps of this algorithm, more details can be found in [van der Meulen and Schauer \(2017a\)](#). As we assume  $X^\circ$  to be a strong solution to the SDE specified by Equations (5) and (7), there is a measurable mapping  $\mathcal{GP}$  such that  $X^\circ = \mathcal{GP}(x_0, W)$ , where  $x_0$  is the starting point and  $W$  a  $\mathbb{R}^d$ -valued Wiener process ( $\mathcal{GP}$  abbreviates *Guided Proposal*). As  $x_0$  is fixed, we will write, with slight abuse of notation,  $X^\circ = \mathcal{GP}(W)$ . The algorithm requires to choose a tuning parameter  $\rho \in [0, 1)$  and proceeds according to the following steps.

- (1) Draw a Wiener process  $Z$  on  $[0, T]$ , Set  $X = g(Z)$ .
- (2) Propose a Wiener process  $W$  on  $[0, T]$ . Set

$$Z^\circ = \rho Z + \sqrt{1 - \rho^2} W \quad (30)$$

and  $X^\circ = \mathcal{GP}(Z^\circ)$ .

- (3) Compute  $A = \Psi_T(X^\circ)/\Psi_T(X)$  (where  $\Psi_T$  is as defined in (22)). Sample  $U \sim \text{Uniform}(0, 1)$ . If  $U < A$  then set  $Z = Z^\circ$  and  $X = X^\circ$ .
- (4) Repeat steps (2) and (3).

The invariant distribution of this chain is precisely  $\mathbb{P}_T^*$ . If the guided proposal is good, then we may use  $\rho = 0$ , which yields an independence sampler. However, for difficult bridge simulation schemes, possibly caused by a large value of  $T$  or strong nonlinearity in the drift or diffusion coefficient, a value of  $\rho$  close to 1 may be required. The proposal in Equation (30) is precisely the pCN method, see e.g. [Cotter et al. \(2013\)](#).

In the implementation we use a fine equidistant grid, which is transformed by the mapping  $\tau: [0, T] \rightarrow [0, T]$  given by  $\tau(s) = s(2 - s/T)$ . Motivation for this choice is given in Section 5 of [van der Meulen and Schauer \(2017a\)](#). Intuitively, the guiding term gets stronger near  $T$  and therefore we use a finer grid the closer we get to  $T$ . The guided proposal is simulated on this grid, and using the values obtained,  $\Psi_T(X^\circ)$  is approximated by Riemann approximation. Furthermore, for numerical stability we solve the equation for  $M^\dagger(t)$  using  $M^\dagger(T) = 10^{-10}I_{m \times m}$  instead of  $M^\dagger(T) = 0_{m \times m}$ .

**Example 4.1.** Assume the NCLAR(3)-model, as described in Example 3.4 with  $\underline{\beta}(t, x) = -6 \sin(2\pi x)$  and  $x_0 = [0 \ 0 \ 0]'$ . We first condition the process on hitting

$$v = [1/32 \ 1/4 \ 1]'$$

at time  $T = 0.5$ , assuming  $L = I_3$  (full observation at time  $T$ ). The idea of this example is that sample paths of the rough component are mean-reverting at levels  $k \in \mathbb{Z}$ , with

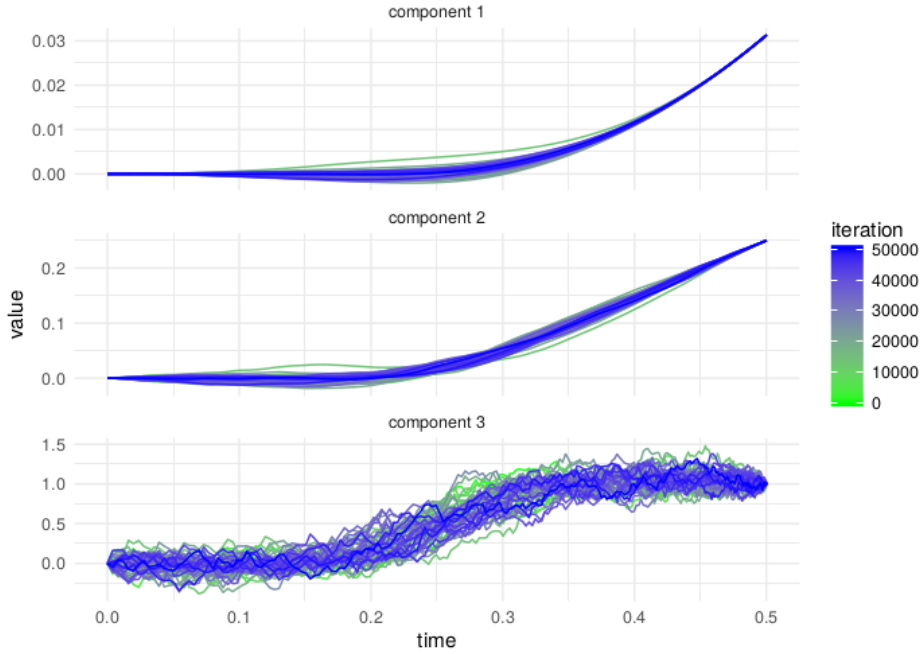


FIGURE 1. Sampled guided diffusion bridges when conditioning on  $X_T = [1/32 \ 1/4 \ 1]'$  in Example 4.1.

occasional noise-driven shifts from one level to another. The given conditioning then forces the process to move halfway the interval (at about time 0.25) from level 0 to level 1, remaining at level approximately level 1 up till time  $T$ . Such paths are rare events and obtaining these by forward simulation is computationally extremely intensive.

We construct guided proposals according to (26) with  $\tilde{\beta}_3(t) = 0$ . Iterates of the sampler using  $\rho = 0.85$  are shown in Figure 1. The average Metropolis-Hastings acceptance percentage was 43%. We need a value of  $\rho$  close to 1 as we cannot easily incorporate the strong nonlinearity into the guiding term of the guided proposal. We repeated the simulation, this time only conditioning on  $LX_T = 1/32$ , where  $L = [1 \ 0 \ 0]$ . We again took  $\rho = 0.95$ , leading to an average Metropolis-Hastings acceptance percentage of 24%. The results are in Figure 2. The distribution of bridges turns out to be bimodal. The latter is confirmed by extensive forward simulation and only keeping those paths which approximately satisfy the conditioning.

**Example 4.2.** Ditlevsen and Samson (2017) consider the stochastic hypo-elliptic FitzHugh-Nagumo model, which is specified by the SDE

$$dX_t = \begin{bmatrix} 1/\varepsilon & -1/\varepsilon \\ \gamma & -1 \end{bmatrix} X_t dt + \begin{bmatrix} -X_{t,1}^3/\varepsilon + s/\varepsilon \\ \beta \end{bmatrix} dt + \begin{bmatrix} 0 \\ \sigma \end{bmatrix} dW_t, \quad X_0 = x(0). \quad (31)$$

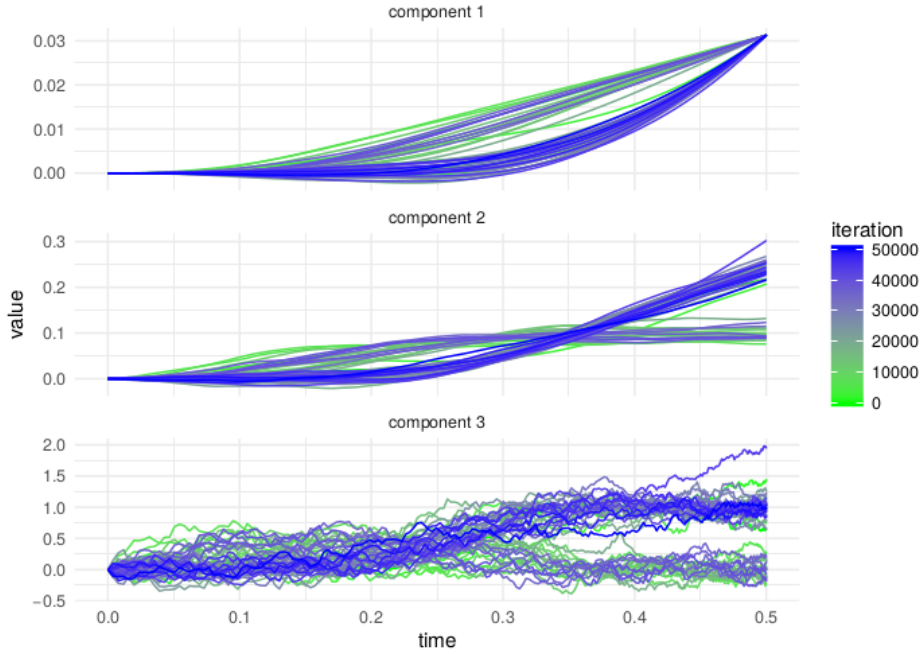


FIGURE 2. Sampled guided diffusion bridges when conditioning on  $LX_T = 1/32$  with  $L = [1 \ 0 \ 0]'$  in Example 4.1.

Only the first component is observed, hence  $L = [1 \ 0]$ . We consider the same parameter values as in [Ditlevsen and Samson \(2017\)](#):

$$[\varepsilon \ s \ \gamma \ \beta \ \sigma] = [0.1 \ 0 \ 1.5 \ 0.8 \ 0.3]. \quad (32)$$

A realisation of a sample path on  $[0, 10]$  is given in Figure 3.

While this example formally does not fall into our setup, the conditions of Assumption 2.7 strongly suggest that the component of the drift with smooth path, i.e. the first component of  $b$ , certainly needs to match at the observed endpoint. We construct guided proposals by linearising the drift term  $X_{t,1}$  at the observed endpoint  $v$ . Hence, using that  $-x^3 \approx 2a^3 - 3a^2x$  for  $x$  near  $a$ , we take

$$\tilde{B}(t) = \begin{bmatrix} 1/\varepsilon - 3v^2/\varepsilon & -1/\varepsilon \\ \gamma & -1 \end{bmatrix}, \quad \tilde{\beta}(t) = \begin{bmatrix} 2v^3/\varepsilon + s/\varepsilon \\ \beta \end{bmatrix} \quad \tilde{\sigma} = \begin{bmatrix} 0 \\ \sigma \end{bmatrix}.$$

To illustrate the performance of our method, we take a rather challenging, strongly nonlinear problem. We consider bridges over the time-interval  $[0, T]$  with  $T = 2$ , starting at  $x(0) = [-0.5 \ -0.6]$ . In Figure 4 we forward simulated 100 paths, to access the behaviour of the process. Next, we consider two cases:

- (a) Conditioning on the first coordinate at the endpoint of a “typical” path; we took  $v = -1$ .
- (b) Conditioning on the first coordinate at the endpoint of an “extreme” path; we took  $v = 1.1$ .

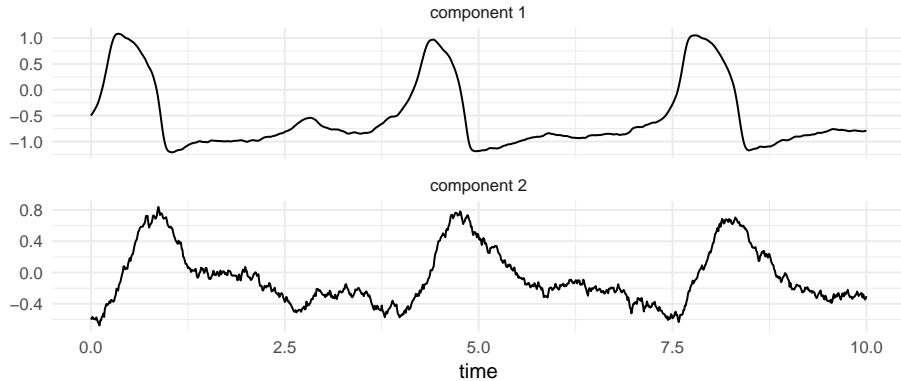


FIGURE 3. A realisation of a sample path of the FitzHugh-Nagumo model as specified in Equation (31), with parameter values as in (32).

We ran the sampler for 50.000 iterations, using  $\rho = 0$  and  $\rho = 0.9$  in cases (a) and (b) respectively. The percentage of accepted proposals in the Metropolis-Hastings step equals 64% and 21% respectively. In Figures 5 and 6 we plotted every 1000-th sampled path out of the 50.000 iterations for the “typical” and “extreme” cases respectively. Figure 5 immediately demonstrates that for a typical path, guided proposals very closely resemble true bridges (using Figure 4 as comparison). To assess whether in the “extreme” case the sampled bridges resemble true bridges, we also forward simulated the process, only keeping those paths for which  $|Lx_T - v| < 0.01$ . The resulting paths are shown in Figure 7 and resemble those in Figure 6 quite well.

This example is extremely challenging in the sense that we take a rather long time horizon ( $T = 2$ ), the noise-level on the second coordinate is small and the drift of the diffusion is highly nonlinear. As a result, the true distribution of bridges is multimodal. Even in much simpler settings, sampling from a multimodal distribution using MCMC constitutes a difficult problem. Here, the multimodality is recovered remarkably well by our method as can be seen from Figure 6.

**Remark 4.3.** We have chosen for 50.000 iterations in the chosen examples. However, qualitatively the same figures of simulating bridges can be obtained by reducing the number of iterations to approximately 10.000.

**4.1. Numerical checks on the validity of guided proposals.** In this section we first investigate the quality of guided proposals over long time spans. Next, we empirically demonstrate that the conditions of our main theorem, especially Assumption 2.7, is stronger than actually needed. In each numerical experiment we compare two histogram estimators for  $v \mapsto \rho(0, x_0; T, v)$ . The first estimator is obtained by making a histogram of a large number of forward simulations of the unconditioned diffusion process. Denote by  $\{A_k\}$  the bins of this histogram. A second estimator is obtained by using the equality

$$\rho(0, x_0; T, v) = \tilde{\rho}(0, x_0; T, v) \mathbb{E} [\Psi_T(X^{\circ, T, v})]$$

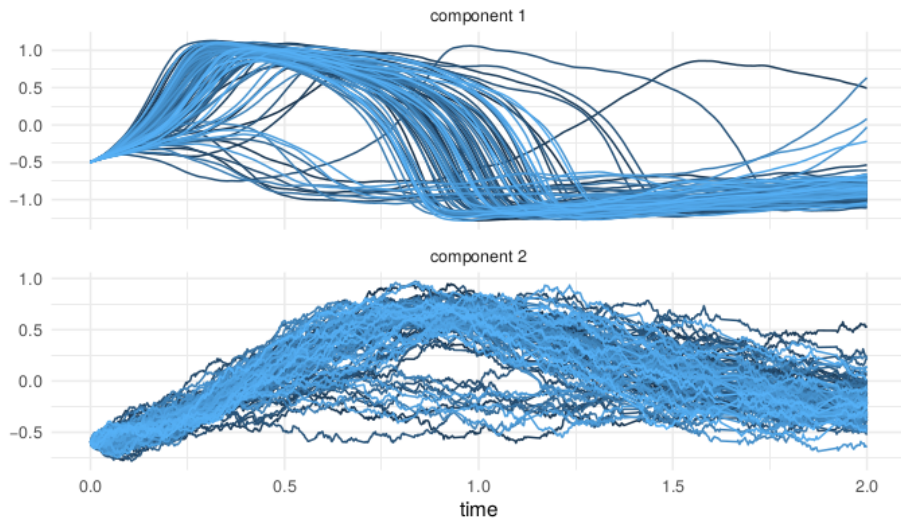


FIGURE 4. Realisations of 100 forward sampled paths for the FitzHugh-Nagumo model as specified in Equation (31), with parameter values as in (32).

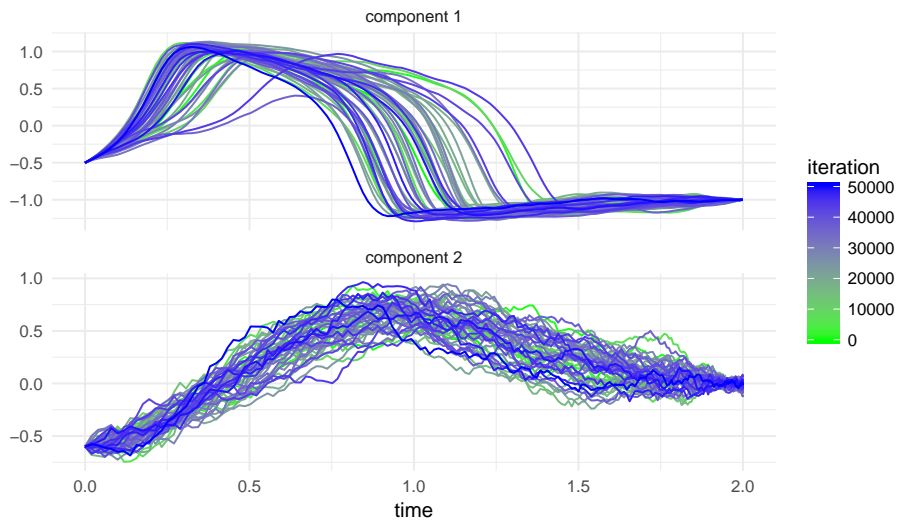


FIGURE 5. Sampled guided diffusion bridges when conditioning on  $v = -1$  (typical case).

which is a direct consequence of Theorem 2.14. Note that we extended the notation to highlight that  $\rho$ ,  $\tilde{\rho}$  and  $X^\circ$  depend on  $T$  and  $v$ . We use the relation in the previous

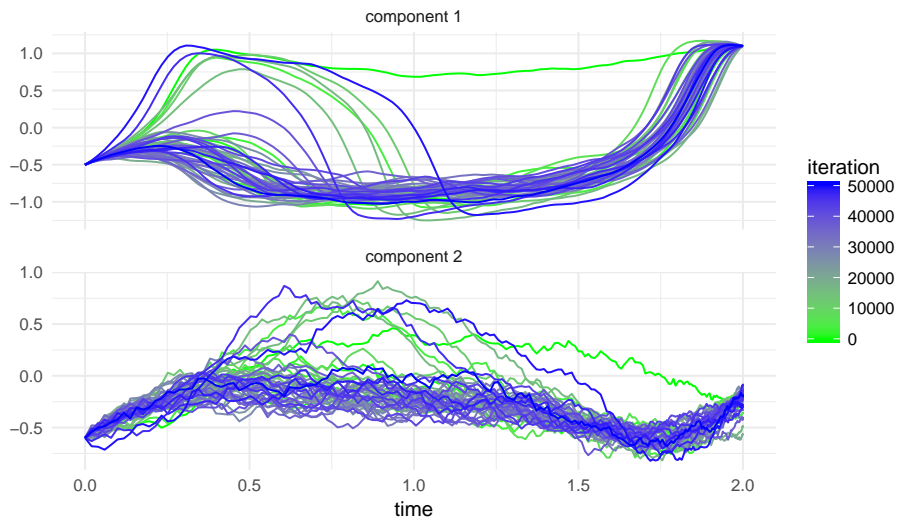


FIGURE 6. Sampled guided diffusion bridges when conditioning on  $v = 1.1$  (extreme case). The “outlying” green curve corresponds to the initialisation of the algorithm.

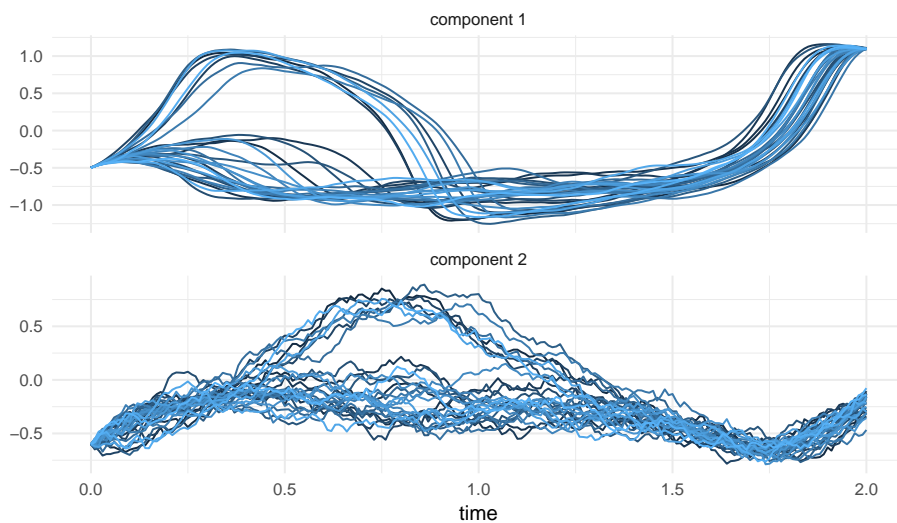


FIGURE 7. Realisations of 30 forward sampled paths for the FitzHugh-Nagumo model as specified in Equation (31), with parameter values as in (32). Only those paths are kept for which  $|Lx_T - v| < 0.01$ , where  $v = 1.1$  (the conditioning for the “extreme” case).



display as follows: for each bin  $A_k$

$$\int_{A_k} \rho(0, x_0; T, v) dv = \mathbb{E} \left[ \mathbf{1}_{A_k}(\tilde{V}) \frac{\tilde{\rho}(0, x_0; T, \tilde{V})}{q(\tilde{V})} \Psi_T(X^{T, \tilde{V}, \circ}) \right]$$

where  $\tilde{V}$  is sampled from the density  $q$ . Hence,  $\int_{A_k} \rho(0, x_0; T, v) dv$  can be approximated using importance sampling where repeatedly first the endpoint  $v$  is sampled from  $q$  and subsequently a guided proposal is simulated that is conditioned to hit  $v$  at time  $T$ . In our experiments we took the importance sampling density  $q$  to be the Gaussian density with mean and covariance obtained from the unconditioned forward simulated endpoint values.

Note that the setup is such that this is feasible, at least when estimating the entire histogram, but of course it would be prohibitively expensive to use forward sampling to compute the density in a single small bin or at a single point.

**Example 4.4.** Consider the non-linear hypo-elliptic 2d system determined by drift  $b(t, x) = Bx + \beta(t) + [0; \frac{1}{2} \sin(x_2)]$  with  $B = \frac{1}{10} [-1 \ 1; 0 \ -1]$ ,  $\beta(t) = [0 \ \frac{1}{2} \sin(t/4)]$ , and dispersion  $\sigma \equiv [0; 2]$  (with a semicolon separating matrix rows). Starting at  $X_0 = [0; -\pi/2]$ , we assume to observe  $V = LX_T + Z$  with  $L = [1 \ 1]$ ,  $Z \sim N(0, 10^{-6})$ . We consider both  $T = 4\pi$  and  $T = 40\pi$ , the latter to check how guided proposals perform over a very long time span. We take guided proposals derived from  $\tilde{b}(t, x) = Bx + \beta(t)$  and  $\tilde{\sigma} = \sigma$ .

In Figure 8 the two histograms are contrasted. Interestingly, the results show no degradation in performance when increasing  $T$  by an order. For the simulations we took  $K = 70$  bins and 100 000 samples of  $V$  respective draws from  $\tilde{V}$  (thus on average approximately 1 500 draws per bin) and time grid  $t_i = s_i(2 - s_i/T)$  with  $s_i = h i$ ,  $h = 0.01$ , therefore decreasing step-size towards  $T$  while keeping the number of grid points equal to  $T/h$ , as suggested in [van der Meulen and Schauer \(2017a\)](#). The implementation is based on our Julia package [Mider and Schauer \(2019\)](#) with package co-author Marcin Mider. The figures also serve to verify the correctness of the implementation.

**Example 4.5.** It is interesting to ask if – numerically speaking – the change of measure is successful in cases where  $\sigma$  depends on  $x$  and the fourth inequality of Assumption 2.4 cannot be verified. For that purpose, we slightly adjust the setting of the previous example by now taking  $L = [1 \ 0]$  and  $\sigma(t, x) = [0; 2 + \frac{1}{2} \cos(x_2)]$  and repeating the experiment. In this case we chose  $\tilde{\sigma} = \sigma(0, [0; 0])$ . As the problem is more difficult, we took less bins ( $K = 50$ ) and set  $h = 0.005$  (otherwise keeping our previous choices.) The resulting Figure 9 shows no indication of lack of absolute continuity or loss of probability mass. This strongly indicates that guided proposals can perform perfectly fine for the present complex setting that includes state-dependent diffusion coefficient and hypo-ellipticity.

However, care is needed, in Figure 10 we show the result for the same experiment, but with  $L$  changed to  $L = [1 \ 1]$ . Here, the loss of probability mass indicates violation of absolute continuity. We conjecture that  $La(T, v)L' = L\tilde{a}(T)L'$  may be the “right” restriction on choosing  $\tilde{a}(T)$ . To obtain empirical evidence, we redid the experiment with  $L = [1 \ 1]$  but now  $\sigma(t, x) = [0; 2 + \frac{1}{2} \cos(Lx)]$ . In this case one can match the



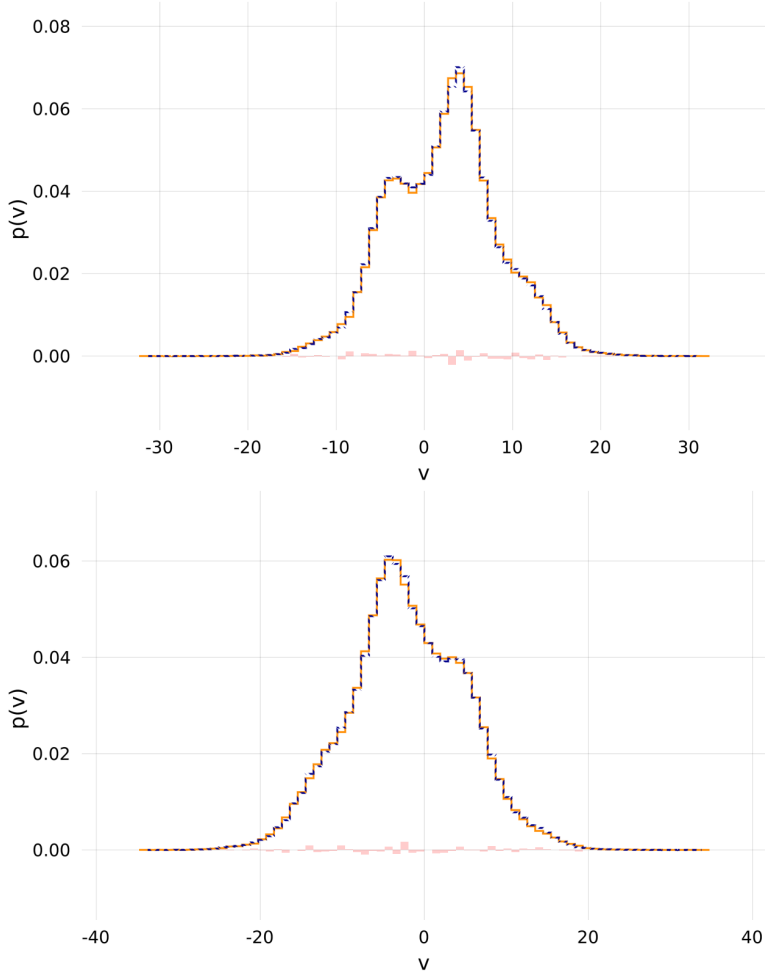


FIGURE 8. Dark orange: Histogram baseline estimate of the density of observation  $V = LX_T + Z$ ,  $Z \sim N(0, 10^{-6})$  from forward simulation. Dashed blue: observation density estimate using weighted histogram of points  $\tilde{V}$  sampled from Gaussian distribution weighted with importance weights from guided proposals steered towards those points. Top:  $T = 4\pi$ . Bottom:  $T = 40\pi$ . Pink: difference between histograms.

diffusivity at time  $T$  by taking  $\tilde{\sigma} = [0; 2 + \frac{1}{2} \cos(v)]$ . The resulting figure (Figure 11) indicates no loss of absolute continuity, supporting the conjecture.

## 5. PROOFS OF PROPOSITION 2.8 AND COROLLARY 2.10

In this section we give proofs of the results from Section 2.2 on the behaviour of guided proposals near the conditioning point. For clarity, the proof of Proposition 2.8 is split up over subsections 5.1, 5.2 and 5.3. The proof of Corollary 2.10 is in section 5.4.

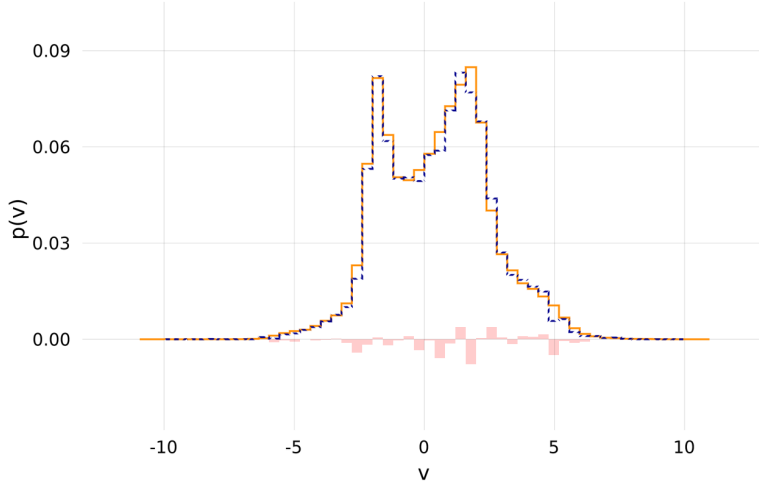


FIGURE 9. As Figure 8, but estimates for the model with observation operator  $L = [1 \ 0]$  and  $\sigma(t, x) = [0; 2 + \frac{1}{2} \cos(x_2)]$ , at  $T = 4\pi$ .

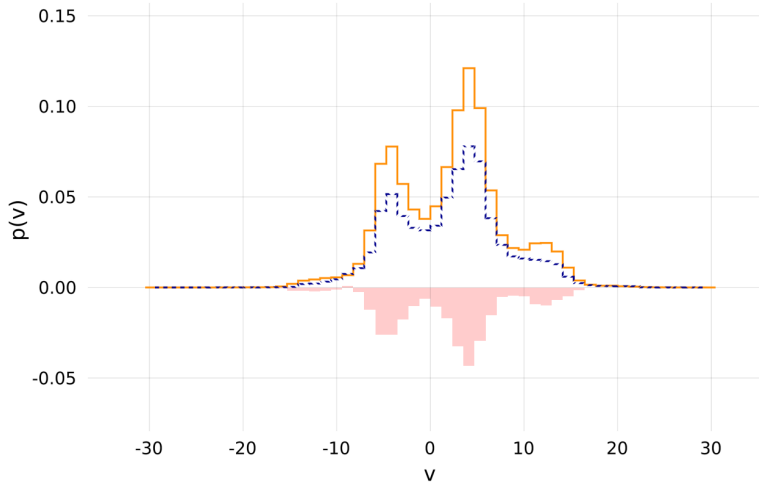


FIGURE 10. As Figure 8, but estimates for the model with observation operator  $L = [1 \ 1]$  and  $\sigma(t, x) = [0; 2 + \frac{1}{2} \cos(x_2)]$ , at  $T = 4\pi$ . Note the loss of probability mass indicating lack of absolute continuity.

**5.1. Centring and scaling of the guided proposal.** To reduce notational overhead, we write  $a_t \equiv a(t, X_t^\circ)$ . Then  $\tilde{b}_t$ ,  $b_t$  and  $\sigma_t$  are defined similarly. Our starting point is the expression for  $\tilde{r}$  in (16).

**Lemma 5.1.** If we define

$$Z_t = v - \mu(t) - L(t)X_t^\circ,$$

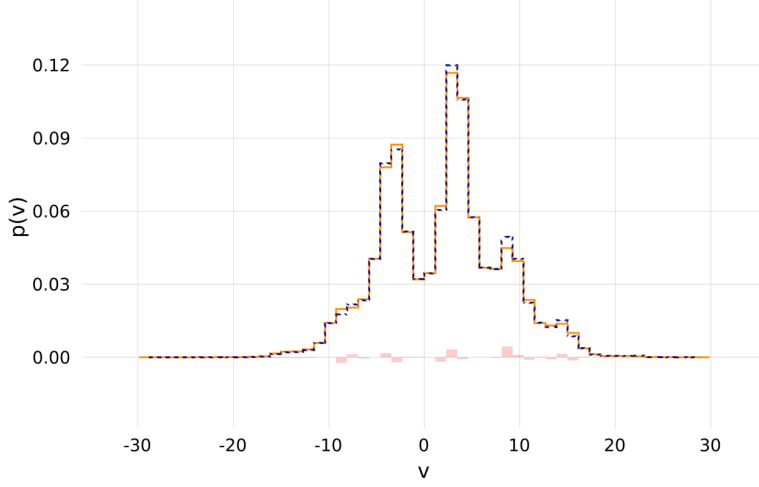


FIGURE 11. As Figure 8, but estimates for the model with observation operator  $L = [1 \ 1]$  and  $\sigma(t, x) = [0; 2 + \frac{1}{2} \cos(x_1 + x_2)]$ , at  $T = 4\pi$ .

then

$$dZ_t = L(t) \left( \tilde{b}_t - b_t \right) dt + L(t)\sigma_t dW_t - L(t)a_t L(t)' M(t) Z_t dt.$$

*Proof.* We have

$$dZ_t = - \left( \frac{d}{dt} L(t) \right) X_t^\circ dt - \frac{d}{dt} \mu(t) - L(t) dX_t^\circ.$$

The results now follows because the first two terms on the right-hand-side together equal  $L(t)\tilde{b}(t, X_t^\circ)$ .  $\square$

**Lemma 5.2.** We have

$$\begin{aligned} \frac{1}{2} d \left( Z_t' M(t) Z_t \right) &= \frac{1}{2} Z_t' M(t) L(t) \left( \tilde{a}(t) - a_t \right) L(t)' M(t) Z_t dt \\ &\quad + Z_t' M(t) L(t) \left( \tilde{b}_t - b_t \right) dt \\ &\quad + Z_t' M(t) L(t) \sigma_t dW_t - \frac{1}{2} Z_t' M(t) L(t) a_t L(t)' M(t) Z_t dt \\ &\quad + \text{tr} \left( L(t) a_t L(t)' M(t) \right) dt. \end{aligned}$$

*Proof.* By Ito's lemma

$$\frac{1}{2} d \left( Z_t' M(t) Z_t \right) = \frac{1}{2} Z_t' \frac{dM(t)}{dt} Z_t dt + Z_t' M(t) dZ_t + \text{tr} \left( L(t) a_t L(t)' M(t) \right) dt.$$

Next, substitute the SDE for  $Z_t$  from lemma 5.1 and use that

$$\frac{dM(t)}{dt} = -M(t) \frac{dM(t)^{-1}}{dt} M(t) = M(t) L(t) \tilde{a}(t) L(t)' M(t).$$

The final equality follows from the fact that  $M^\dagger(t) = M(t)^{-1}$  satisfies the ordinary differential equation  $dM^\dagger(t) = -L(t)\tilde{a}(t)L(t)' dt$ . The result follows upon reorganising terms.  $\square$

Whereas in the uniformly elliptic case all elements of  $Z_t$  and  $M(t)$  behave in the same way as a function of  $T - t$ , this is not the case in the hypo-elliptic case. For this reason, we introduce a diagonal scaling matrix  $\Delta(t)$ .

**Lemma 5.3.** Let  $\Delta(t)$  be an invertible  $m \times m$  diagonal matrix. If  $Z_{\Delta,t}$ ,  $L_\Delta(t)$  and  $M_\Delta(t)$  are as defined in Equations (18) and (19) then

$$\begin{aligned} \frac{1}{2} d \left( Z'_{\Delta,t} M_\Delta(t) Z_{\Delta,t} \right) &= \frac{1}{2} Z'_{\Delta,t} M_\Delta(t) L_\Delta(t) (\tilde{a}(t) - a_t) L_\Delta(t)' M_\Delta(t) Z_{\Delta,t} dt \\ &\quad + Z'_{\Delta,t} M_\Delta(t) L_\Delta(t) \left( \tilde{b}_t - b_t \right) dt \\ &\quad + Z'_{\Delta,t} M_\Delta(t) L_\Delta(t) \sigma_t dW_t - \frac{1}{2} Z'_{\Delta,t} M_\Delta(t) L_\Delta(t) a_t L_\Delta(t)' M_\Delta(t) Z_{\Delta,t} dt \\ &\quad + \text{tr} \left( L_\Delta(t) a_t L_\Delta(t)' M_\Delta(t) \right) dt. \end{aligned} \tag{33}$$

Moreover,

$$\tilde{r}(t, X_t^\circ) = L_\Delta(t)' M_\Delta(t) Z_{\Delta,t} \tag{34}$$

*Proof.* This is a straightforward consequence of Lemma 5.2. The expression for  $\tilde{r}$  follows from equation (16).  $\square$

**5.2. Recap on notation and results.** For clarity we summarise our notation, some of which was already defined in Section 1.6. The auxiliary process is defined by the SDE  $d\tilde{X}_t = (\tilde{B}(t)\tilde{X}_t + \tilde{\beta}(t)) dt + \tilde{\sigma}(t) dW_t$ . The matrix  $\Phi(t)$  satisfies the ODE  $d\Phi(t) = \tilde{B}(t)\Phi(t) dt$  and we set  $\Phi(T, t) = \Phi(T)\Phi(t)^{-1}$ . A realisation  $v$  of  $V = LX_T$  is observed. The scaled process is defined by

$$Z_{\Delta,t} = \Delta(t) \left( v - \int_t^T L(\tau) \tilde{\beta}(\tau) d\tau - L(t) X_t^\circ \right),$$

where  $L(t) = L\Phi(T, t)$  and  $L_\Delta(t) = \Delta(t)L(t)$ . Furthermore, we defined  $M(t) = \left( \int_t^T L(\tau) \tilde{a}(\tau) L(\tau)' d\tau \right)^{-1}$  and  $M_\Delta(t) = \Delta(t)^{-1} M(t) \Delta(t)^{-1}$ , where  $\tilde{a}(t) = \tilde{\sigma}(t) \tilde{\sigma}(t)'$ . Finally, the guiding term in the SDE for the guided proposal  $X_t^\circ$  is given by  $a(t, X_t^\circ) L_\Delta(t)' M_\Delta(t) Z_{\Delta,t}$ . The process  $t \mapsto Z_{\Delta,t}$  is the key object to be studied in this section.

**5.3. Proof of Proposition 2.8.** The line of proof is exactly as suggested in Mao (1992) (page 341):

- (1) Start with the Lyapunov function  $V(t, Z_{\Delta,t}) = \frac{1}{2} Z_{\Delta,t}' M_\Delta(t) Z_{\Delta,t}$ .
- (2) Apply Ito's lemma to  $V(t, Z_{\Delta,t})$ .
- (3) Use martingale inequalities to bound the stochastic integral.
- (4) Apply a Gronwall type inequality.

We bound all terms appearing in equation (33). Note that the first term on the right-hand-side vanishes. We start with the Wiener integral term. To this end, fix  $t_0 \in [0, T)$  and let

$$N_t = \int_{t_0}^t Z'_{\Delta,s} M_{\Delta}(s) L_{\Delta}(s) \sigma \, dW_s.$$

Then

$$\int_{t_0}^t Z'_{\Delta,s} M_{\Delta}(s) L_{\Delta}(s) \sigma_s \, dW_s - \frac{1}{2} \int_{t_0}^t Z'_{\Delta,s} M_{\Delta}(s) L_{\Delta}(s) a_s L_{\Delta}(s)' M_{\Delta}(s) Z_{\Delta,s} \, ds = N_t - \frac{1}{2} [N]_t.$$

Now  $N_t$  can be bounded using an exponential martingale inequality. Let  $\{\gamma_n\}$  be a sequence of positive numbers. Define for  $n \in \mathbb{N}$ ,  $t_n = T - 1/n$  and

$$E_n = \left\{ \sup_{0 \leq t \leq t_{n+1}} \left( N_t - \frac{1}{2} [N]_t \right) > \gamma_n \right\}.$$

By the exponential martingale inequality of Theorem 1.7.4 in Mao (1997), we obtain that  $P(E_n) \leq e^{-\gamma_n}$ . If we assume  $\sum_{n=1}^{\infty} e^{-\gamma_n} < \infty$ , then by the Borel-Cantelli lemma  $P(\limsup_{n \rightarrow \infty} E_n) = 0$ . Hence, for almost all  $\omega$ ,  $\exists n_0(\omega)$  such that for all  $n \geq n_0(\omega)$

$$\sup_{t_0 \leq t \leq t_{n+1}} \left( N_t - \frac{1}{2} \int_0^t [N]_t \right) \leq \gamma_n. \quad (35)$$

Let  $\varepsilon > 0$ . Upon taking  $\gamma_n = (1 + 2\varepsilon) \log n$  we get  $\sum_{n=1}^{\infty} e^{-\gamma_n} = \sum_{n=1}^{\infty} n^{-1-2\varepsilon} < \infty$ . Since  $M_{\Delta}(t)$  is strictly positive definite

$$\lambda_{\min}(M_{\Delta}(t)) \|Z_{\Delta,t}\|^2 \leq Z'_{\Delta,t} M_{\Delta}(t) Z_{\Delta,t}.$$

Assume  $t_0 < t < t_{n+1}$ . Combining the inequality of the preceding display with Lemma 5.3 and substituting the bound in (35), we obtain that for any  $\varepsilon > 0$

$$\begin{aligned} \frac{1}{2} \lambda_{\min}(M_{\Delta}(t)) \|Z_{\Delta,t}\|^2 &\leq \frac{1}{2} Z_{\Delta,t_0} M_{\Delta}(t_0) Z_{\Delta,t_0} \\ &+ \int_{t_0}^t \|Z_{\Delta,s}\| \|M_{\Delta}(s)\| \|L_{\Delta}(s) (\tilde{b}_s - b_s)\| \, ds \\ &+ \gamma_n + \int_{t_0}^t \text{tr} (L_{\Delta}(s) a_s L_{\Delta}(s)' M_{\Delta}(s)) \, ds \\ &+ \frac{1}{2} \int_{t_0}^t Z'_{\Delta,s} M_{\Delta}(s) L_{\Delta}(s) (\tilde{a}(s) - a_s) L_{\Delta}(s)' M_{\Delta}(s) Z_{\Delta,s} \, ds \end{aligned}$$

Recall that for positive semidefinite matrices  $A$  and  $C$  we have  $|\text{tr}(AC)| \leq \text{tr}(A) \text{tr}(C) \leq \text{tr}(A) p \lambda_{\max}(C)$ , if  $C \in \mathbb{R}^{p \times p}$ . Hence,

$$\text{tr} (L_{\Delta}(s) a_s L_{\Delta}(s)' M_{\Delta}(s)) \leq \text{tr} (L_{\Delta}(s) a_s L_{\Delta}(s)') m \lambda_{\max}(M_{\Delta}(s)) \quad (36)$$

Furthermore, as

$$\|M_{\Delta}(s)\| = \sqrt{\lambda_{\max}(M_{\Delta}(s)^2)} = \lambda_{\max}(M_{\Delta}(s)) \quad (37)$$

we can combine the preceding three inequalities to obtain

$$\begin{aligned}
\frac{1}{2}\lambda_{\min}(M_{\Delta}(t))\|Z_{\Delta,t}\|^2 &\leq \frac{1}{2}Z_{\Delta,t_0}M_{\Delta}(t_0)Z_{\Delta,t_0} \\
&+ \int_{t_0}^t \|Z_{\Delta,s}\|\lambda_{\max}(M_{\Delta}(s))\|L_{\Delta}(s)(\tilde{b}_s - b_s)\| ds \\
&+ \gamma_n + m \int_{t_0}^t \text{tr}(L_{\Delta}(s)a_s L_{\Delta}(s)')\lambda_{\max}(M_{\Delta}(s)) ds \\
&+ \frac{1}{2} \int_{t_0}^t Z'_{\Delta,s}M_{\Delta}(s)L_{\Delta}(s)(\tilde{a}(s) - a_s)L_{\Delta}(s)'M_{\Delta}(s)Z_{\Delta,s} ds.
\end{aligned} \tag{38}$$

Upon substituting the bounds in (20) we get, for certain positive constants  $C_0, C_1, C_2, C_3$  and  $C_4$  that

$$\begin{aligned}
(T-t)^{-1}\|Z_{\Delta,t}\|^2 &\leq C_0 + C_1 \int_{t_0}^t \|Z_{\Delta,s}\|(T-s)^{-1} ds \\
&+ C_2\gamma_n + C_3 \int_{t_0}^t (T-s)^{-1} d + C_4 \int_{t_0}^t \|Z_{\Delta,s}\|^2(T-s)^{\alpha-2}.
\end{aligned} \tag{39}$$

If we define  $\xi_t = (T-t)^{-1}\|Z_{\Delta,t}\|^2$ , then this inequality can be rewritten as

$$\begin{aligned}
\xi_t &\leq C_0 + C_2\gamma_n + C_3 \log\left(\frac{T-t_0}{T-t}\right) \\
&+ C_1 \int_{t_0}^t (T-s)^{-1/2}\sqrt{\xi_s} ds + C_4 \int_{t_0}^t (T-s)^{\alpha-1}\xi_s ds.
\end{aligned} \tag{40}$$

By Lemma B.1 in the appendix this implies

$$\begin{aligned}
\xi_t &\leq \left( \sqrt{C_0 + C_2\gamma_n + C_3 \log\left(\frac{T-t_0}{T-t}\right)} + \frac{1}{2}C_1 \left( \sqrt{T-t_0} - \sqrt{T-t} \right) \right)^2 \\
&\quad \times \exp\left( C_4 \int_{t_0}^t (T-s)^{\alpha-1} \right).
\end{aligned}$$

Now divide both sides of this inequality by  $\log(1/(T-t))$  and consider  $t_n < t < t_{n+1}$ . Then  $\log n \leq \log(1/(T-t))$ . It then follows that

$$\limsup_{t \uparrow T} \frac{\|Z_{\Delta,t}\|^2}{(T-t)\log(1/(T-t))} \leq C_2(1+2\varepsilon) + C_3.$$

Now let  $\varepsilon \downarrow 0$ .

**5.4. Proof of Corollary 2.10.** As  $\Delta(t) = I_m$  it is easy to see that  $M(t) = \mathcal{O}(1/(T-t))$  and  $L_{\Delta}(t) = \mathcal{O}(1)$ . This behaviour of  $M(t)$  is also contained in the first inequality of Lemma 8 in Schauer et al. (2017) (note that in that paper,  $\tilde{H}$  corresponds to  $M$  as defined in this paper). Now it is easy to see that the conditions of theorem 2.8 are satisfied.

## 6. ABSOLUTE CONTINUITY WITH RESPECT TO THE GUIDED PROPOSAL DISTRIBUTION

6.1. **Proof of Theorem 2.14.** We start with a result that gives the Radon-Nikodym derivative of  $\mathbb{P}_t^*$  relative to  $\mathbb{P}_t^\circ$  for  $t < T$ .

**Proposition 6.1.** For  $t < T$  we have

$$\frac{d\mathbb{P}_t^*}{d\mathbb{P}_t^\circ}(X^\circ) = \frac{\tilde{\rho}(0, x_0) \rho(t, X_t^\circ)}{\rho(0, x_0) \tilde{\rho}(t, X_t^\circ)} \Psi_t(X^\circ), \quad (41)$$

where  $\Psi_t$  is defined in (22).

*Proof.* Although this result is not a special case of proposition 1 in Schauer et al. (2017) (where it is assumed that  $L = I$  and that the diffusion is uniformly elliptic), the arguments for deriving the likelihood ratio of  $\mathbb{P}_t^*$  with respect to  $\mathbb{P}_t^\circ$  are the same and therefore omitted. The only thing that needs to be checked is that  $\tilde{\rho}(t, x)$  satisfies the Kolmogorov backward equation associated to  $\tilde{X}$ . This can be proved along the lines of Lemma 3.4 and Corollary 3.5 of van der Meulen and Schauer (2018). Let  $\tilde{\mathcal{F}}_t = \sigma(\tilde{X}_s, 0 \leq s \leq t)$  and set  $\tilde{Y}_t = \tilde{\rho}(t, \tilde{X}_t)$ . Now

$$\begin{aligned} \mathbb{E}[\tilde{Y}_t | \tilde{\mathcal{F}}_s] &= \int_{\mathbb{R}^d} \tilde{\rho}(t, x) \tilde{p}(s, \tilde{X}_s; t, x) dx \\ &= \int_{\mathbb{R}^d} \tilde{p}(s, \tilde{X}_s; t, x) \int_{\mathbb{R}^{d-m}} \left( \tilde{p}(t, x; T, \sum_{j=1}^d \xi_j f_j \right) d\xi_{m+1}, \dots, d\xi_d \\ &= \int_{\mathbb{R}^{d-m}} \tilde{p} \left( s, \tilde{X}_s; T, \sum_{j=1}^d \xi_j f_j \right) d\xi_{m+1}, \dots, d\xi_d = \tilde{\rho}(s, \tilde{X}_s) = \tilde{Y}_s. \end{aligned}$$

That is,  $(\tilde{Y}_t, \tilde{\mathcal{F}}_t)$  is a martingale. If  $\tilde{\mathcal{L}}$  denotes the infinitesimal generator of  $\tilde{X}_t$ , then  $\mathcal{K} = \partial/(\partial t) + \tilde{\mathcal{L}}$  is the infinitesimal generator of the space time process  $(t, \tilde{Y}_t)$ . Since  $\tilde{Y}_t$  is a martingale, the mapping  $(t, x) \mapsto \tilde{\rho}(t, x)$  is space-time harmonic. Then by Proposition 1.7 in chapter VII of Revuz and Yor (1991)  $\mathcal{K}\tilde{\rho}(t, x) = 0$ . That is,  $\tilde{\rho}(t, x)$  satisfies Kolmogorov's backward equation.  $\square$

This absolute continuity result is only useful for simulating conditioned diffusions if it can be shown to hold in the limit  $t \uparrow T$  as well. The main line of proof is the same as in the proof of Theorem 1 in Schauer et al. (2017), where at various places  $p$  and  $\tilde{p}$  need to be replaced with  $\rho$  and  $\tilde{\rho}$ . However, some of the auxiliary results that are used require new arguments in the present setting. Moreover, the assumed Aronson type bounds are not suitable for hypo-elliptic diffusions.

6.2. **Proof of Theorem 2.14.** We start with introducing some notation. Define the mapping  $g_\Delta: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^m$  by

$$g_\Delta(t, x) = \Delta(t) (v - \mu(t) - L(t)x)$$

and note that  $Z_{\Delta,t} = g_{\Delta}(t, X_t^{\circ})$ . For a diffusion process  $Y$  we define the stopping time

$$\sigma_k(Y) = T \wedge \inf_{t \in [0, T]} \left\{ \|g_{\Delta}(t, Y_t)\| \geq k \sqrt{(T-t) \log(1/(T-t))} \right\},$$

where  $k \in \mathbb{N}$ . We write

$$\sigma_k^{\circ} = \sigma_k(X^{\circ}) \quad \sigma_k = \sigma_k(X) \quad \sigma_k^{\star} = \sigma_k(X^{\star}).$$

Define  $\bar{\rho} = \tilde{\rho}(0, x_0)/\rho(0, x_0)$ . By Proposition 6.1, for any  $t < T$  and bounded,  $\mathcal{F}_t$ -measurable  $f$ , we have

$$\mathbb{E} \left[ f(X^{\star}) \frac{\tilde{\rho}(t, X_t^{\star})}{\rho(t, X_t^{\star})} \right] = \mathbb{E} [f(X^{\circ}) \bar{\rho} \Psi_t(X^{\circ})]. \quad (42)$$

By taking  $f_t(x) = \mathbf{1}\{t \leq \sigma_k(x)\}$ , we get

$$\bar{\rho} \mathbb{E} [\Psi_t(X^{\circ}) \mathbf{1}\{t \leq \sigma_k^{\circ}\}] = \mathbb{E} \left[ \frac{\tilde{\rho}(t, X_t^{\star})}{\rho(t, X_t^{\star})} \mathbf{1}\{t \leq \sigma_k^{\star}\} \right]. \quad (43)$$

Next, we take  $\lim_{k \rightarrow \infty} \lim_{t \uparrow T}$  on both sides. We start with the left-hand-side. By Lemma 6.2, for each  $k \in \mathbb{N}$ ,  $\sup_{0 \leq t \leq T} \Psi_t(X^{\circ})$  is uniformly bounded on the event  $\{T = \sigma_k^{\circ}\}$ . Hence, by the dominated convergence theorem we obtain

$$\lim_{k \rightarrow \infty} \lim_{t \uparrow T} \mathbb{E} [\Psi_t(X^{\circ}) \mathbf{1}\{t \leq \sigma_k^{\circ}\}] = \lim_{k \rightarrow \infty} \mathbb{E} [\Psi_T(X^{\circ}) \mathbf{1}\{T \leq \sigma_k^{\circ}\}].$$

Since by definition  $\sigma_k^{\circ} \leq T$ , we have  $\{T \leq \sigma_k^{\circ}\} = \{T = \sigma_k^{\circ}\}$ . Furthermore,

$$\mathbf{1}\{T = \sigma_k^{\circ}\} = \mathbf{1} \left\{ \|Z_{\Delta,t}\| \leq k \sqrt{(T-t) \log(1/(T-t))} \right\} \uparrow 1 \quad \text{as } k \rightarrow \infty,$$

by Proposition 2.8. Therefore, by monotone convergence

$$\lim_{k \rightarrow \infty} \lim_{t \uparrow T} \mathbb{E} [\Psi_t(X^{\circ}) \mathbf{1}\{t \leq \sigma_k^{\circ}\}] = \mathbb{E} [\Psi_T(X^{\circ})].$$

It remains to show that the right-hand-side of (43) tends to 1. We write

$$\begin{aligned} \rho(0, x_0) \mathbb{E} \left[ \frac{\tilde{\rho}(t, X_t^{\star})}{\rho(t, X_t^{\star})} \mathbf{1}\{t \leq \sigma_k^{\star}\} \right] &= \mathbb{E} [\tilde{\rho}(t, X_t) \mathbf{1}\{t \leq \sigma_k\}] \\ &= \mathbb{E} [\tilde{\rho}(t, X_t)] - \mathbb{E} [\tilde{\rho}(t, X_t) \mathbf{1}\{t > \sigma_k\}] \end{aligned}$$

By Lemma 6.4 the first of the terms on the right-hand-side tends to  $\rho(0, x_0)$  when  $t \uparrow T$ . The second term tends to zero by Lemma 6.5.

To complete the proof we note that by equation (42) and Lemma 6.4 we have  $\bar{\rho} \mathbb{E} [\Psi_t(X^{\circ})] \rightarrow 1$  as  $t \uparrow T$ . In view of the preceding and Scheffé's Lemma this implies that  $\Psi_t(X^{\circ}) \rightarrow \Psi_T(X^{\circ})$  in  $L^1$ -sense as  $t \uparrow T$ . Hence for  $s < T$  and a bounded,  $\mathcal{F}_s$ -measurable, continuous functional  $g$ ,

$$\mathbb{E} [g(X^{\circ}) \bar{\rho} \Psi_T(X^{\circ})] = \lim_{t \uparrow T} \mathbb{E} \left[ g(X^{\star}) \frac{\tilde{\rho}(t, X_t^{\star})}{\rho(t, X_t^{\star})} \right].$$

By Lemma 6.4 this converges to  $\mathbb{E} g(X^{\star})$  as  $t \uparrow T$  and we find that  $\mathbb{E} g(X^{\circ}) \bar{\rho} \Psi_T(X^{\circ}) = \mathbb{E} g(X^{\star})$ .



**Lemma 6.2.** Under Assumption 2.7 there exists a positive constant  $K$  (not depending on  $k$ ) such that

$$\Psi_t(X^\circ) \mathbf{1}_{t \leq \sigma_m^\circ} \leq \exp(Kk^2).$$

*Proof.* To bound  $\Psi_t(X^\circ)$ , we will first rewrite  $\mathcal{G}(s, X^\circ)$  in terms of  $Z_{\Delta,t}$ ,  $L_\Delta(t)$  and  $M_\Delta(t)$ , as defined in (18) and (19). By display (34), we have

$$\tilde{r}(t, X_t^\circ) = L_\Delta(t)' M_\Delta(t) Z_{\Delta,t} \quad \text{and} \quad \tilde{H}(t) = L_\Delta(t)' M_\Delta(t) L_\Delta(t).$$

Here, the expression for  $\tilde{H}(t)$  was obtained from

$$\tilde{H}(t) = -DL(t)'M(t)(v - \mu(t) - L(t)x) = D(L(t)'M(t)L(t)x) = L(t)'M(t)L(t).$$

Hence,

$$\begin{aligned} \mathcal{G}(s, X_s^\circ) &= (b(s, X_s^\circ) - \tilde{b}(s, X_s^\circ))' L_\Delta(s)' M_\Delta(s) Z_{\Delta,s} \\ &\quad - \frac{1}{2} \text{tr}([a_s - \tilde{a}(s)] L_\Delta(s)' M_\Delta(s) L_\Delta(s)) \\ &\quad + \frac{1}{2} Z_{\Delta,s}' M_\Delta(s) L_\Delta(s) [a_s - \tilde{a}(s)] L_\Delta(s)' M_\Delta(s) Z_{\Delta,s}. \end{aligned} \tag{44}$$

On the event  $\{t \leq \sigma_k^\circ\}$  we have

$$\|Z_{\Delta,t}\| \leq k \sqrt{(T-t) \log(1/(T-t))}.$$

The absolute value of the first term of  $\mathcal{G}$  can be bounded by

$$\begin{aligned} &\|M_\Delta(s)\| \left\| L_\Delta(s) \left( \tilde{b}(s, X_s^\circ) - b(s, X_s^\circ) \right) \right\| \|Z_{\Delta,s}\| \\ &\leq (T-s)^{-1} \|Z_{\Delta,s}\| \\ &\leq c_1 m (T-s)^{-1/2} \sqrt{\log(1/(T-s))}. \end{aligned}$$

Here we bounded  $\|M_\Delta(t)\| \leq \lambda_{\max}(M_\Delta(t))$ , as in (37). The absolute value of twice the second term of  $\mathcal{G}$  can be bounded by

$$\text{tr}(L_\Delta(s)(a_s - \tilde{a}(s))L_\Delta(s)') k \lambda_{\max}(M_\Delta(s)),$$

just as in (36). As for a  $p \times p$  matrix  $A$  we have  $\text{tr}(A) \leq p \lambda_{\max}(A) = p \|A\|^2$  (recall we assume the spectral norm on matrices throughout), this can be bounded by

$$m \|L_\Delta(s)(a_s - \tilde{a}(s))L_\Delta(s)'\|^2 m \lambda_{\max}(M_\Delta(s)) \leq m^2 c_3 \bar{c} (T-s)^{2\alpha-1}$$

The absolute value of twice the third term of  $\mathcal{G}$  can be bounded by

$$\begin{aligned} &\|Z_{\Delta,s}\|^2 \|M_\Delta(s)\|^2 \|L_\Delta(s)(a(s) - \tilde{a}(s))L_\Delta(s)'\| \\ &\leq k^2 (T-s) \log(1/(T-s)) \bar{c}^2 (T-s)^{-2} c_3 (T-s)^\alpha \\ &\leq k^2 \bar{c}^2 c_3 (T-s)^{\alpha-1} \log(1/(T-s)). \end{aligned}$$

We conclude that all three terms in  $\mathcal{G}$  are integrable on  $[0, T]$ . □

**Lemma 6.3.** For all bounded, continuous  $f: [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$

$$\lim_{t \uparrow T} \int f(t, x) \tilde{p}(t, x; T, v) dx = f(T, v).$$

*Proof.* The proof is just as in Lemma 7 of Schauer et al. (2017).  $\square$

**Lemma 6.4.** If Assumption 2.13 holds true,  $0 < t_1 < t_2 < \dots < t_N < t < T$  and  $g \in C_b(\mathbb{R}^{Nd})$ , then

$$\lim_{t \uparrow T} \mathbb{E} \left[ g(X_{t_1}^*, \dots, X_{t_N}^*) \frac{\tilde{\rho}(t, X_t^*)}{\rho(t, X_t^*)} \right] = \mathbb{E} [g(X_{t_1}^*, \dots, X_{t_N}^*)].$$

*Proof.* The joint density  $q$  of  $(X_{t_1}, \dots, X_{t_N})$ , conditional on  $X_{t_0} = x_0$  is given by  $q(x_1, \dots, x_N) = \prod_{i=1}^N p(t_{i-1}, x_{i-1}; t_i, x_i)$ . Hence,

$$\begin{aligned} & \mathbb{E} \left[ g(X_{t_1}^*, \dots, X_{t_N}^*) \frac{\tilde{\rho}(t, X_t^*)}{\rho(t, X_t^*)} \right] \\ &= \int g(x_1, \dots, x_N) \frac{\tilde{\rho}(t, x)}{\rho(t, x)} q(x_1, \dots, x_N) \frac{p(t_N, x_N; t, x) \rho(t, x)}{\rho(0, x_0)} dx_1 \dots dx_N dx \quad (45) \\ &= \frac{1}{\rho(0, x_0)} \int g(x_1, \dots, x_N) q(x_1, \dots, x_N) F(t; t_N, x_N) dx_1 \dots dx_N, \end{aligned}$$

where for  $t_N < t < T$

$$F(t; t_N, x_N) = \int p(t_N, x_N; t, x) \tilde{\rho}(t, x) dx.$$

We can assume  $t \geq (T + t_N)/2$ . For fixed  $t_N$  and  $x_N$ , the mapping  $(t, x) \mapsto p(t_N, x_N; t, x)$  is continuous and bounded, for  $t$  bounded away from  $t_N$ . By Lemma 6.3 it follows that  $F(t; t_N, x_N) \rightarrow \rho(t_N, x_N)$  when  $t \uparrow T$ . The argument is finished by taking the limit  $t \uparrow T$  on both sides of equation (45), interchanging limit and integral on the right-hand-side and noting that the limit on the right-hand-side coincides with  $\mathbb{E} [g(X_{t_1}^*, \dots, X_{t_N}^*)]$ .

The interchange is permitted by dominated convergence. To see this, first note that  $g$  is assumed to be bounded. Next,

$$\int \left( \prod_{i=1}^n p(t_{i-1}, x_{i-1}; t_i, x_i) \right) p(t_N, x_N; t, x) \tilde{\rho}(t, x) dx dx_1 \dots dx_N \leq C^{N+1} \tilde{\rho}(t_0, x_0),$$

which follows from repeated application of Assumption 2.13.  $\square$

**Lemma 6.5.** Assume that there exists a positive  $\delta$  such that  $|\Delta(t)| \lesssim (T - t)^{-\delta}$ . If Assumption 2.13 holds true, then

$$\lim_{k \rightarrow \infty} \lim_{t \uparrow T} \mathbb{E} [\tilde{\rho}(t, X_t) \mathbf{1}\{t > \sigma_k\}].$$

*Proof.* As in the proof of Lemma 5 in Schauer et al. (2017), it suffices to show that

$$\lim_{k \rightarrow \infty} \lim_{t \uparrow T} \mathbb{E} \left[ \mathbf{1}\{t > \sigma_k\} \int p(\sigma_k, X_{\sigma_k}; t, z) \tilde{\rho}(t, z) dz \right] = 0. \quad (46)$$

Applying Assumption 2.13 and using the Chapman-Kolmogorov relations, we obtain

$$\int p(\sigma_k, X_{\sigma_k}; t, z) \tilde{\rho}(t, z) dz \leq C \tilde{\rho}(\sigma_k, X_{\sigma_k}).$$

Define  $\tilde{Z}_t = v - \mu(t) - L(t)\tilde{X}_t$ . If we denote its transition density by  $\tilde{q}$ , then

$$\tilde{\rho}(t, y) = \tilde{q}(t, v - \mu(t) - L(t)y; T, 0), \quad y \in \mathbb{R}^d \quad \text{and} \quad t \in [0, T),$$

since  $\tilde{r}(t, x)$  depends on  $x$  only via  $L(t)x$ . Define the set

$$\mathcal{A}_k = \{(t, y) \in [0, T) \times \mathbb{R}^d : \|\Delta(t)(v - \mu(t) - L(t)y)\| = k\eta(t)\},$$

where  $\eta(t) = \sqrt{(T-t) \log(1/(T-t))}$ . Then

$$\mathbb{E} \left[ \mathbf{1}_{\{t > \sigma_k\}} \int p(\sigma_k, X_{\sigma_k}; t, z) \tilde{\rho}(t, z) dz \right] \leq \mathbb{E} \left[ \sup_{(t,y) \in \mathcal{A}_k} \tilde{\rho}(t, y) \right],$$

since by definition of  $\sigma_k$ ,  $\|\Delta(\sigma_k)(v - \mu(\sigma_k) - L(\sigma_k)X_{\sigma_k})\| = k\eta(\sigma_k)$ . The expectation on the right-hand-side is now superfluous. It is easily derived that  $\tilde{Z}_t$  satisfies the SDE

$$d\tilde{Z}_t = L(t)\tilde{\sigma}(t) dW_t$$

and hence for  $x \in \mathbb{R}^m$

$$\tilde{q}(t, x; T, 0) = \varphi_m(0; x, M^\dagger(t)),$$

where we denote the density of the multivariate normal distribution in  $\mathbb{R}^m$  with mean vector  $\nu$  and covariance matrix  $\Upsilon$ , evaluated in  $u$  by  $\varphi_m(u; \nu, \Upsilon)$ . Hence, stitching the previous derivations together we obtain

$$\mathbb{E} \left[ \mathbf{1}_{\{t > \sigma_k\}} \int p(\sigma_k, X_{\sigma_k}; t, z) \tilde{\rho}(t, z) dz \right] \leq \sup_{(t,y) \in \mathcal{A}_k} \varphi_m(0; x, M^\dagger(t)).$$

The right-hand-side, multiplied with  $(2\pi)^{m/2}$  equals

$$|M(t)|^{1/2} \exp \left( -\frac{1}{2}(v - \mu(t) - L(t)y)' M(t)(v - \mu(t) - L(t)y) \right)$$

which can be further bounded by

$$\begin{aligned} & \sup_{(t,y) \in \mathcal{A}_k} |\Delta(t)| |M_\Delta(t)|^{1/2} \exp \left( -\frac{1}{2}(v - \mu(t) - L(t)y)' \Delta(t) M_\Delta(t) \Delta(t)(v - \mu(t) - L(t)y) \right) \\ & \leq \sup_{(t,y) \in \mathcal{A}_k} |\Delta(t)| (\lambda_{\max}(M_\Delta(t)))^{m/2} \exp \left( -\frac{1}{2} \|\Delta(t)(v - \mu(t) - L(t)y)\|^2 \lambda_{\min}(M_\Delta(t)) \right) \\ & \leq \sup_{t \in [0, T)} |\Delta(t)| \left( \frac{\bar{c}}{T-t} \right)^{m/2} \exp \left( -\frac{ck^2\eta(t)^2}{2(T-t)} \right) \\ & \lesssim \sup_{t \in [0, T)} (T-t)^{-\delta-m/2} \exp \left( -\frac{ck^2\eta(t)^2}{2(T-t)} \right). \end{aligned}$$

Next, the maximum can be bounded, followed by taking the limit  $k \rightarrow \infty$ , to see that this tends to zero. This is exactly as in the proof of Lemma 5 in [Schauer et al. \(2017\)](#).  $\square$

**6.3. Proof of Lemma 2.16.** By absolute continuity of the laws of  $\tilde{X}$  and  $X$  and the abstract Bayes' formula, for bounded  $\mathcal{F}_T$ -measurable  $f$  we have

$$\mathbb{E}[f(X) \mid X_T = v] = \frac{\tilde{p}(0, x_0; T, v)}{p(0, x_0; T, v)} \mathbb{E} \left[ f(\tilde{X}) \frac{d\mathbb{P}_T}{d\tilde{\mathbb{P}}_T}(\tilde{X}) \mid \tilde{X}_T = v \right].$$

Hence, upon taking  $f \equiv 1$  and applying Girsanov's theorem we get

$$p(0, x_0; T, v) = \tilde{p}(0, x_0; T, v) \mathbb{E} \left[ \exp \left( \int_0^T \eta(\tilde{X}_s)' dW_s - \frac{1}{2} \int_0^T \|\eta(\tilde{X}_s)\|^2 ds \right) \mid \tilde{X}_T = v \right].$$

Since  $\eta$  is bounded this implies

$$p(0, x_0; T, v) \propto \tilde{p}(0, x_0; T, v) \mathbb{E} \left[ \exp \left( \int_0^T \eta(\tilde{X}_s)' dW_s \right) \mid \tilde{X}_T = v \right].$$

Upon defining  $\tau_j = \int_0^T \eta_j(\tilde{X}_s)^2 ds$ , the Dambis-Dubins-Schwarz theorem implies that the expectation on the right-hand-side equals

$$\mathbb{E} \left[ e^{-\sum_{j=1}^{d'} \int_0^T \eta_j(\tilde{X}_s) dW_s^j} \mid \tilde{X}_T = v \right] = \mathbb{E} \left[ e^{\sum_{j=1}^{d'} W_{\tau_j}^j} \mid \tilde{X}_T = v \right].$$

By boundedness of  $\eta$  there exists constants  $\{K_j\}_{j=1}^{d'}$  such that  $\tau_j \leq K_j$ . Hence the right-hand-side of the preceding display can be bounded by

$$\mathbb{E} \left[ e^{\sum_{j=1}^{d'} \sup_{0 \leq s \leq K_j} W_s^j} \mid \tilde{X}_T = v \right] = \mathbb{E} \left[ e^{\sum_{j=1}^{d'} \sup_{0 \leq s \leq K_j} W_s^j} \right] = \prod_{j=1}^{d'} \mathbb{E} \left[ e^{\sup_{0 \leq s \leq K_j} W_s^j} \right],$$

where the final equality follows from the components of  $\bar{W}$  being independent. The expectation on the right-hand-side is finite, the constant only depending on  $T$ . To see this: if  $B_t$  is a one-dimensional Brownian motion, then  $\bar{B}_t = \sup_{0 \leq s \leq t} B_s$  has density  $f_{\bar{B}_t}(x) = \sqrt{2/(\pi t)} e^{-x^2/(2t)} \mathbf{1}_{[0, \infty)}(x)$ , which implies that  $\mathbb{E}[\exp(\bar{B}_t)] < \infty$ .

The statement of the theorem now follows by considering the processes  $X$  and  $\tilde{X}$  started in  $x$  at time  $s$  and noting that the derived constant only depends on  $T$ .

## 7. DISCUSSION

### 7.1. Extending the approach in Marchand (2012) to hypo-elliptic diffusions.

A potential advantage of the approach in Marchand (2012) is that, at least in the uniformly elliptic case, there is no matching condition for the diffusion coefficient to be satisfied. Inspecting the guiding term in (11), it can be seen that it is also well defined when  $\ker(\sigma(t, x)'L') = \{0\}$ , since this assures that the inverse of  $La(t, x)L'$  exists for all  $t \geq 0$  and  $x \in \mathbb{R}^d$ . Unfortunately, this excludes for example the case where the smooth component of an integrated diffusion process is observed (Example 3.2). Here, the guiding term is given by

$$\text{guid}_1(t, x) := aL(t)' \left( \int_t^T L(\tau) \tilde{a} L(\tau)' d\tau \right)^{-1} (v - L(t)x).$$

Now it is tempting to adjust the proposals by [Marchand \(2012\)](#) in Equation (11) in the same way as was done for guided proposals, by replacing  $L$  by  $L(t)$ . This leads to the guiding term

$$\text{guid}_2(t, x) := aL(t)'(L(t)aL(t)')^{-1} \frac{v - L(t)x}{T - t}.$$

This guiding term will *not* give correct bridges though. To see this, if  $\underline{\beta} \equiv 0$  then  $X^\circ = X^*$  but

$$\text{guid}_2(t, x) = \frac{1}{3} \text{guid}_1(t, x) \quad \text{with} \quad \text{guid}_1(t, x) = \begin{bmatrix} 0 \\ 3 \frac{v - x_1 - (T-t)x_2}{(T-t)^2} \end{bmatrix} \quad (47)$$

(here  $x_i$  denotes the  $i$ -th component of the vector  $x$ ). We stress that  $\text{guid}_2(t, x)$  has never been proposed in [Marchand \(2012\)](#) and that the guiding term in Equation (11) is perfectly valid in the uniformly elliptic case. The point we make here, is that it is far from straightforward to generalise [Marchand \(2012\)](#)'s work to the hypo-elliptic setting. Possibly, the correct generalisation of [Marchand \(2012\)](#)'s work to the hypo-elliptic case is to take the guiding term of the form

$$a(t, X_t^\circ)L(t)' \left( \int_t^T L(\tau)a(\tau, X_\tau^\circ)L(\tau)' d\tau \right)^{-1} (v - L(t)X_t^\circ).$$

This term is however unattractive from a computational point of view.

**7.2. State-dependent diffusion coefficient.** We have formulated our results for state-dependent diffusion coefficients  $\sigma$ . The main difficulty however resides in checking the fourth inequality of Assumption 2.7. We conjecture that the “right” way to deal with this term is to bound

$$\|L_\Delta(t)(\tilde{a}(t) - a(t, X_t^\circ))L_\Delta(t)'\| \lesssim \|Z_{\Delta,t}\|.$$

Then the final term in inequality (39) would be replaced with  $C_4 \int_{t_0}^t (T-s)^{-2} \|Z_{\Delta,s}\|^3 ds$ . The conjecture is motivated by the proof of Theorem 2 in [Schauer et al. \(2017\)](#). Obtaining such an inequality is not straightforward, as is the corresponding Gronwall type argument. We postpone such investigations to future research.

#### ACKNOWLEDGEMENT

We thank O. Papaspiliopoulos (Universitat Pompeu Fabra Barcelona), S. Sommer (University of Copenhagen) and M. Mider (University of Warwick) for stimulating discussions on diffusion bridge simulation.

#### APPENDIX A. EXISTENCE OF $\tilde{r}$ IF $L = I$

In case  $L = I$ , the existence problem of transition densities has been studied in control theory as well.

**Definition A.1.** The pair  $(\tilde{B}, \tilde{\sigma})$  is called *completely controllable* at  $s$  if, for any  $t > s$  and  $x, y \in \mathbb{R}^d$ , there exists a function  $v \in L^2[s, t]$  and corresponding solution  $Y$  of

$$dY(u) = \left( \tilde{B}(u)Y(u) + \tilde{\sigma}(u)v(u) \right) du, \quad Y(s) = x$$

such that  $Y(t) = y$ .

The following lemma is proved in [Hermes and LaSalle \(1969\)](#).

**Lemma A.2.** The following are equivalent:

- (1)  $(\tilde{B}, \tilde{\sigma})$  is completely controllable at  $s$ ;
- (2) Non degenerate gaussian transition densities  $\tilde{p}(s, x; t, y)$  exist;
- (3) For arbitrary gaussian initial data  $\tilde{X}_s$ , the random vector  $\tilde{X}_t$  is non degenerate Gaussian for  $t > s$ .

If  $\tilde{B}, \tilde{\sigma}$  in (8) are constant matrices, complete controllability is equivalent to  $\text{rank}(C) = d$ , where the controllability matrix  $C$  is defined by

$$C := [\tilde{\sigma}, \tilde{B}\tilde{\sigma}, \dots, \tilde{B}^{d-1}\tilde{\sigma}]$$

(see page 74 in [Hermes and LaSalle \(1969\)](#) or proposition 6.5 in [Karatzas and Shreve \(1991\)](#)). This provides an easily verifiable condition for complete controllability.

## APPENDIX B. GRONWALL TYPE INEQUALITY

In the proof of [Theorem 2.8](#) we used the following Gronwall type inequality:

**Lemma B.1.** Assume  $t \mapsto \zeta(t)$  is continuously differentiable and nonnegative on  $[t_0, t_1]$ . Assume  $t \mapsto f_1(t)$  and  $t \mapsto f_2(t)$  are continuous and nonnegative on  $[t_0, t_1]$ . Suppose  $t \mapsto u(t)$  is a continuous and nonnegative function on  $[t_0, t_1]$  satisfying the inequality

$$u(t) \leq \zeta(t) + \int_{t_0}^t f_1(s) \sqrt{u(s)} ds + \int_{t_0}^t f_2(s) u(s) ds, \quad t \in [t_0, t_1].$$

Then

$$u(t) \leq \left( \sqrt{\zeta(t_0) + \int_{t_0}^t |\zeta'(s)| ds} + \frac{1}{2} \int_{t_0}^t f_1(s) ds \right)^2 \exp \left( \int_{t_0}^t f_2(s) ds \right).$$

*Proof.* This is a special case of [theorem 2.1 in Agarwal et al. \(2005\)](#). In their notation, we have  $n = 2$ ,  $w_1(x) = \sqrt{x}$ ,  $W_1(x) = 2\sqrt{x}$  (taking  $u_1 = 0$ ),  $w_2(x) = x$ ,  $W_2(x) = \log x$  (taking  $u_2 = 1$ ).  $\square$

## APPENDIX C. HYPOELLIPTICITY

**Proposition C.1.** Consider the diffusion (1) with  $b(t, x) = Bx + \beta(t, x)$  for  $B \in \mathbb{R}^{d \times d}$  and  $\beta \in C^\infty([0, T] \times \mathbb{R}^d, \mathbb{R}^d)$ , and with  $\sigma \in C^\infty([0, T] \times \mathbb{R}^d, \mathbb{R}^{d \times d'})$ . Suppose that, for all  $(t, x) \in (0, T) \times \mathbb{R}^d$ , the pair  $(B, \sigma(t, x))$  is controllable, i.e. the rank of the matrix concatenation

$$[\sigma(t, x) \quad B\sigma(t, x) \quad \dots \quad B^{d-1}\sigma(t, x)]$$

is equal to  $d$ . Furthermore suppose that for all  $(t, x) \in (0, T) \times \mathbb{R}^d$  and all tuples  $(n_0, n_1, \dots, n_d) \in \{0, 1, \dots, d-1\}^{d+1}$ ,

$$(\partial_t)^{n_0} \prod_{i=1}^d (\partial_{x_i})^{n_i} \beta(t, x) \in \text{Col } \sigma(t, x)$$

and

$$\text{Col} \left( (\partial_t)^{n_0} \prod_{i=1}^d (\partial_{x_i})^{n_i} \sigma(t, x) \right) \subset \text{Col} \sigma(t, x),$$

i.e. the column spaces of all partial derivatives of  $\beta(t, x)$  and  $\sigma(t, x)$ , including  $\beta(t, x)$  itself, belong to the column space of  $\sigma(t, x)$ . Finally suppose there exists at most one strong solution to (1) (which is the case if e.g.  $\beta$  and  $\sigma$  satisfy a linear growth condition). Then for all initial conditions  $x_0$  and all  $t \geq 0$ , the distribution of  $X_t$  admits a density function  $p(t, x, y)$ :

$$\mathbb{E}_{x_0}[f(X_t)] = \int_{\mathbb{R}^d} p(t, x_0, y) f(y) dy, \quad f \in C_0(\mathbb{R}^d),$$

and  $p$  is a smooth (infinitely often continuously differentiable) function on  $(0, \infty) \times \mathbb{R}^d \times \mathbb{R}^d$ .

*Proof.* Write  $(\sigma_j)_{j=1}^{d'}$  for the columns of  $\sigma$  so that

$$\sigma(t, x) = [\sigma_1(t, x) \quad \dots \quad \sigma_{d'}(t, x)].$$

The Stratonovich form of (1) is given by

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

where  $\tilde{b}(t, x) = Bx + \tilde{\beta}(t, x)$  with coordinates of  $\tilde{\beta}$  given by

$$\tilde{\beta}^i(t, x) = \beta^i(t, x) - \frac{1}{2} \sum_{j=1}^{d'} \sum_{l=1}^d \sigma_j^l (\partial_l \sigma_j^i)(t, x).$$

Observe that  $\tilde{\beta}(t, x) \in \text{Col} \sigma(t, x)$ , just like  $\beta(t, x)$ .

In particular, the generator of the diffusion (1) can be given in terms of the first order differential operators

$$\mathcal{A}_0 f(t, x) = \partial_t f(t, x) + \langle \tilde{b}(t, x), \nabla_x f(t, x) \rangle, \quad \mathcal{A}_j f(t, x) = \langle \sigma_j(t, x), \nabla_x f(t, x) \rangle, \quad j = 1, \dots, d',$$

as  $\mathcal{L} = \mathcal{A}_0 + \frac{1}{2} \sum_{j=1}^{d'} \mathcal{A}_j^2$ . In this proof we will use without further comment (i) Einstein's summation convention and (ii) the canonical identification of first order partial differential operators  $\mathcal{A} = a^i \partial_i = a^0(t, x) \partial_t + \sum_{i=1}^d a^i(x) \partial_{x_i}$  (acting on functions  $f : [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}$ ) with vector fields  $[a^0(t, x) \quad a^1(t, x) \quad \dots \quad a^d(t, x)]^T \in C^\infty([0, \infty) \times \mathbb{R}^d; \mathbb{R} \times \mathbb{R}^d)$  without further comment. The commutator  $[\mathcal{U}_1, \mathcal{U}_2]$  of two vector fields  $\mathcal{U}_1, \mathcal{U}_2$  is as usual defined by

$$[\mathcal{U}_1, \mathcal{U}_2] f(t, x) = \mathcal{U}_1 \mathcal{U}_2 f(t, x) - \mathcal{U}_2 \mathcal{U}_1 f(t, x).$$

For  $l = 0, \dots, d-1$ , write

$$V^l := \text{Col} [\sigma(t, x) \quad B\sigma(t, x) \quad \dots \quad B^l \sigma(t, x)].$$

Write  $[\cdot, \mathcal{A}_0]^l$  for taking the Lie bracket with  $\mathcal{A}_0$  repeatedly, i.e. recursively we define

$$[\mathcal{U}, \mathcal{A}_0]^0 f = \mathcal{U} f \quad \text{and} \quad [\mathcal{U}, \mathcal{A}_0]^{l+1} = [[\mathcal{U}, \mathcal{A}_0]^l, \mathcal{A}_0], \quad l = 0, 1, 2, \dots$$

We first compute

$$[\mathcal{A}_j, \mathcal{A}_0]f = \sigma_j^k B_k^i \partial_i f - (\partial_t \sigma_j^k) \partial_k f - B_l^i x^l (\partial_i \sigma_j^k) \partial_k f - \tilde{\beta}^i (\partial_i \sigma_j^k) \partial_k f + \sigma_j^k (\partial_k \tilde{\beta}^i) \partial_i f.$$

Observe that the first term represents the operator  $\langle B\sigma_j, \nabla f \rangle$ , and the remaining terms assume values in  $V^0 = \text{Col } \sigma(t, x)$ . By iterating we obtain  $[\mathcal{A}_j, \mathcal{A}_0]^l = B^l \sigma + \mathcal{U}$ , where  $\mathcal{U}(t, x) \in V^{l-1}$  for all  $(t, x)$ . By the controllability assumption on  $B$  and  $\sigma$ , the vectors

$$\{[\mathcal{A}_j, \mathcal{A}_0]^l(t, x) : l = 1, \dots, d-1, j = 1, \dots, d'\}$$

span  $\mathbb{R}^d$  for all  $(t, x)$ . Adding  $\mathcal{A}_0$  to the collection of vectors gives that

$$\text{span}\{\mathcal{A}_0, [\mathcal{A}_j, \mathcal{A}_0]^l(t, x) : l = 1, \dots, d-1, j = 1, \dots, d'\}$$

has dimension  $d+1$ , for all  $(t, x) \in (0, T) \times \mathbb{R}^d$ . The result now follows from Hörmander's theorem lifted to  $(0, T) \times \mathbb{R}^d$ , e.g. (Williams, 1981, Corollary 5.8).  $\square$

#### APPENDIX D. DERIVATION OF THE CONDITIONED PROCESS

The SDE for the conditioned process, given in (14), can be derived using Doob's  $h$ -transform.

**Assumption D.1.** The mapping  $\rho: \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}$  is  $C^{1,2}$  and strictly positive.

Suppose  $0 \leq s < t < T$ . By the Chapman-Kolmogorov equations, for a compactly supported  $C^\infty$ -function  $f: \mathbb{R}^d \rightarrow \mathbb{R}$  we have

$$\mathbb{E}[f(X_t) \mid X_s = x, LX_T = v] = \int f(y) p(s, x; t, y) \frac{\rho(t, y)}{\rho(s, x)} dy.$$

Define  $g(t, x) = f(x)\rho(t, x)$ . Using the preceding display we find that the infinitesimal generator of the conditioned process, say  $\mathcal{L}^*$ , equals

$$\begin{aligned} \mathcal{L}^* f(x) &= \lim_{\Delta \downarrow 0} \Delta^{-1} (\mathbb{E}[f(X_{s+\Delta}) \mid X_s = x, LX_T = v] - f(x)) \\ &= \frac{1}{\rho(s, x)} \lim_{\Delta \downarrow 0} \Delta^{-1} \left( \int g(s + \Delta, y) p(s, x; s + \Delta, y) dy - g(s, x) \right) \\ &= \frac{1}{\rho(s, x)} \lim_{\Delta \downarrow 0} \Delta^{-1} (\mathbb{E}[g(s + \Delta, X_{s+\Delta}) \mid X_s = x] - g(s, x)). \end{aligned}$$

By Assumption D.1,  $g$  is a compactly supported  $C^\infty$ -function in the domain of the infinitesimal generator  $\mathcal{K}$  of the space-time process  $(t, X_t)$ . Therefore,

$$\mathcal{L}^* f(x) = \frac{1}{\rho(s, x)} (\mathcal{K}g)(s, x),$$

where

$$\mathcal{K}\varphi(s, x) = \frac{\partial}{\partial s} \varphi(s, x) + \sum_i b_i(s, x) D_i \varphi(s, x) + \frac{1}{2} \sum_{i,j} a_{ij}(s, x) D_{ij}^2 \varphi(s, x).$$



Here (and in the following) all summations run over  $1, \dots, d$ ,  $D_i = \frac{\partial}{\partial x_i}$  and  $D_{ij}^2 = \frac{\partial^2}{\partial x_i \partial x_j}$ . Using the definition of  $g$  we get

$$\begin{aligned} \mathcal{L}^* f(x) &= \sum_i \left( b_i(s, x) + \sum_j a_{ij}(s, x) \frac{D_j \rho(s, x)}{\rho(s, x)} \right) D_i f(x) \\ &\quad + \frac{1}{2} \sum_{i,j} a_{ij}(s, x) D_{ij}^2 f(x) + \frac{f(x)}{\rho(s, x)} (\mathcal{K}\rho)(s, x). \end{aligned}$$

We claim  $(\mathcal{K}\rho)(s, x) = 0$  (that is,  $\rho(t, x)$  satisfies Kolmogorov's backward equation). The drift and diffusion coefficients of the conditioned process can then be identified from the infinitesimal generator  $\mathcal{L}^*$ . To verify the claim, first note that  $Z_t = \rho(t, X_t)$  defines a martingale: if  $\mathcal{F}_s$  is the natural filtration of  $X$ , then

$$\mathbb{E}[Z_t | \mathcal{F}_s] = \int p(s, X_s; t, x) \rho(t, x) dx = Z_s$$

where we used the Chapman-Kolmogorov equations. Therefore,  $(t, x) \mapsto \rho(t, x)$  is space-time harmonic and then the claim follows from proposition 1.7 of Chapter VII in [Revuz and Yor \(1991\)](#).

#### REFERENCES

- Agarwal, R. P., Deng, S. and Zhang, W. (2005). Generalization of a retarded Gronwall-like inequality and its applications. *Applied Mathematics and Computation* **165**(3), 599 – 612.
- Arnaudon, A., Holm, D. D. and Sommer, S. (2019). A geometric framework for stochastic shape analysis. *Foundations of Computational Mathematics* **19**(3), 653–701.
- Bayer, C. and Schoenmakers, J. (2013). Simulation of forward-reverse stochastic representations for conditional diffusions. *Ann. Appl. Probab.* To appear.
- 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.
- Beskos, A., Roberts, G., Stuart, A. and Voss, J. (2008). MCMC methods for diffusion bridges. *Stochastics and Dynamics* **08**(03), 319–350.
- 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., Finch, S. and Sørensen, M. (2016). Simulation of multivariate diffusion bridges. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78**(2), 343–369.
- Brockwell, P. (1994). On continuous-time threshold ARMA processes. *Journal of Statistical Planning and Inference* **39**(2), 291 – 303.
- Clairon, Q. and Samson, A. (2017). Optimal control for estimation in partially observed elliptic and hypoelliptic stochastic differential equations. Working paper or preprint.
- 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.

- Comte, F., Prieur, C. and Samson, A. (2017). Adaptive estimation for stochastic damping Hamiltonian systems under partial observation. *Stochastic Processes and their Applications* **127**(11), 3689 – 3718.
- 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**(3), 424–446.
- Delyon, B. and Hu, Y. (2006). Simulation of conditioned diffusion and application to parameter estimation. *Stochastic Processes and their Applications* **116**(11), 1660 – 1675.
- Ditlevsen, S. and Samson, A. (2017). Hypocoelliptic diffusions: discretization, filtering and inference from complete and partial observations. arXiv:1707.04235.
- 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.
- Golightly, A. and Wilkinson, D. J. (2006). Bayesian sequential inference for nonlinear multivariate diffusions. *Stat. Comput.* **16**(4), 323–338.
- Hairer, M., Stuart, A. M. and Voss, J. (2009). Sampling conditioned diffusions. In *Trends in Stochastic Analysis*, volume 353 of *London Mathematical Society Lecture Note Series*, pp. 159–186. Cambridge University Press.
- Hairer, M., Stuart, A. M. and Voss, J. (2011). Sampling conditioned hypoelliptic diffusions. *Ann. Appl. Probab.* **21**(2), 669–698.
- Hermes, H. and LaSalle, J. P. (1969). *Functional analysis and time optimal control*. Academic Press, New York-London. Mathematics in Science and Engineering, Vol. 56.
- Karatzas, I. and Shreve, S. E. (1991). *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition.
- 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.
- Lindström, E. (2012). A regularized bridge sampler for sparsely sampled diffusions. *Stat. Comput.* **22**(2), 615–623.
- Liptser, R. S. and Shiryaev, A. N. (2001). *Statistics of random processes. I*, volume 5 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, expanded edition. General theory, Translated from the 1974 Russian original by A. B. Aries, Stochastic Modelling and Applied Probability.
- Lu, F., Lin, K. and Chorin, A. (2016). Comparison of continuous and discrete-time data-based modeling for hypoelliptic systems. *Commun. Appl. Math. Comput. Sci.* **11**(2), 187–216.
- Mao, X. (1992). Almost sure polynomial stability for a class of stochastic differential equations. *Quart. J. Math. Oxford Ser. (2)* **43**(171), 339–348.
- Mao, X. (1997). *Stochastic differential equations and their applications*. Horwood Publishing Series in Mathematics & Applications. Horwood Publishing Limited, Chichester.
- Marchand, J.-L. (2012). *Conditionnement de processus markoviens*. IRMAR, Ph.d. Thesis Université de Rennes 1.

- Melnykova, A. (2018). Parametric inference for multidimensional hypoelliptic diffusion with full observations. *arXiv:1802.02943* .
- Mider, M. and Schauer, M. (2019). BridgeSDEInference 0.1.1. doi:10.5281/zenodo.3446185.
- Papaspiliopoulos, O., Roberts, G. O. and Stramer, O. (2013). Data Augmentation for Diffusions. *J. Comput. Graph. Statist.* **22**(3), 665–688.
- Pokern, Y., Stuart, A. M. and Wiberg, P. (2009). Parameter estimation for partially observed hypoelliptic diffusions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71**(1), 49–73.
- Revuz, D. and Yor, M. (1991). *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin.
- Samson, A. and Thieullen, M. (2012). A contrast estimator for completely or partially observed hypoelliptic diffusion. *Stochastic Processes and their Applications* **122**(7), 2521 – 2552.
- Schauer, M., van der Meulen, F. and van Zanten, H. (2017). Guided proposals for simulating multi-dimensional diffusion bridges. *Bernoulli* **23**(4A), 2917–2950.
- Schauer, M. et al. (2018). Bridge 0.9.0. doi:10.5281/zenodo.1406163.
- Sørensen, M. (2012). Estimating functions for diffusion-type processes. In *Statistical methods for stochastic differential equations*, volume 124 of *Monogr. Statist. Appl. Probab.*, pp. 1–107. CRC Press, Boca Raton, FL.
- Stramer, O. and Roberts, G. O. (2007). On Bayesian analysis of nonlinear continuous-time autoregression models. *J. Time Ser. Anal.* **28**(5), 744–762.
- van der Meulen, F. and Schauer, M. (2017a). Bayesian estimation of discretely observed multi-dimensional diffusion processes using guided proposals. *Electron. J. Stat.* **11**(1), 2358–2396.
- van der Meulen, F. and Schauer, M. (2017b). Continuous-discrete smoothing of diffusions. arXiv:1712.03807.
- van der Meulen, F. and Schauer, M. (2018). Bayesian estimation of incompletely observed diffusions. *Stochastics* **90**(5), 641–662.
- van der Meulen, F. and Schauer, M. (2019). Code examples hypoelliptic diffusions 0.1.0. doi:10.5281/zenodo.3457570, <https://github.com/mschauer/code-examples-hypoelliptic-diffusions>.
- Whitaker, G. A., Golightly, A., Boys, R. J. and Sherlock, C. (2017a). Improved bridge constructs for stochastic differential equations. *Statistics and Computing* **27**(4), 885–900.
- Williams, D. (1981). To begin at the beginning: ... In *Stochastic integrals (Proc. Sympos., Univ. Durham, Durham, 1980)*, volume 851 of *Lecture Notes in Math.*, pp. 1–55. Springer, Berlin-New York.

DELFT INSTITUTE OF APPLIED MATHEMATICS, FACULTY OF ELECTRICAL ENGINEERING, MATHEMATICS AND COMPUTER SCIENCE, DELFT UNIVERSITY OF TECHNOLOGY, VAN MOURIK BROEKMANWEG 6, 2628 XE DELFT, THE NETHERLANDS

*E-mail address:* `joris.bierkens@tudelft.nl`

DELFT INSTITUTE OF APPLIED MATHEMATICS, FACULTY OF ELECTRICAL ENGINEERING, MATHEMATICS AND COMPUTER SCIENCE, DELFT UNIVERSITY OF TECHNOLOGY, VAN MOURIK BROEKMANWEG 6, 2628 XE DELFT, THE NETHERLANDS

*E-mail address:* `f.h.vandermeulen@tudelft.nl`

DEPARTMENT OF MATHEMATICAL SCIENCES, CHALMERS UNIVERSITY OF TECHNOLOGY AND UNIVERSITY OF GOTHENBURG, SE-412 96 GÖTEBORG, SWEDEN

*E-mail address:* `smoritz@chalmers.se`