



**HAL**  
open science

# Hypoelliptic diffusions: filtering and inference from complete and partial observations

Susanne Ditlevsen, Adeline Samson

► **To cite this version:**

Susanne Ditlevsen, Adeline Samson. Hypoelliptic diffusions: filtering and inference from complete and partial observations. *Journal of the Royal Statistical Society: Series B*, 2019, 81 (2), pp.361-384. 10.1111/rssb.12307 . hal-01627616v3

**HAL Id: hal-01627616**

**<https://hal.science/hal-01627616v3>**

Submitted on 26 Dec 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Hypoelliptic diffusions: discretization, filtering and inference from complete and partial observations

Susanne Ditlevsen

*Department of Mathematical Sciences, University of Copenhagen.*

E-mail: susanne@math.ku.dk

Adeline Samson

*Univ. Grenoble Alpes, LJK, F-38000 Grenoble, France*

*CNRS, LJK, F-38000 Grenoble, France.*

E-mail: adeline.leclercq-samson@imag.fr

**Summary.** The statistical problem of parameter estimation in partially observed hypoelliptic diffusion processes is naturally occurring in many applications. However, due to the noise structure, where the noise components of the different coordinates of the multidimensional process operate on different time scales, standard inference tools are ill conditioned. In this paper, we propose to use a higher order scheme to discretize the process and approximate the likelihood, such that the different time scales are appropriately accounted for. We show consistency and asymptotic normality with non-typical convergence rates. When only partial observations are available, we embed the approximation into a filtering algorithm for the unobserved coordinates, and use this as a building block in a Stochastic Approximation Expectation Maximization algorithm. We illustrate on simulated data from three models; the Harmonic Oscillator, the FitzHugh-Nagumo model used to model the membrane potential evolution in neuroscience, and the Synaptic Inhibition and Excitation model used for determination of neuronal synaptic input.

## 1. Introduction

Hypoelliptic diffusion processes appear naturally in a variety of applications, but most parameter estimation procedures are ill conditioned, especially when only partial observations are available. Hypoellipticity means that the diffusion matrix of the stochastic differential equation (SDE) defining the multidimensional diffusion process is not of full rank, but its solutions admit a smooth density. In this paper we consider parametric estimation for hypoelliptic diffusions defined as solutions to an SDE of the following form:

$$\begin{cases} dV_t &= a(V_t, U_t)dt \\ dU_t &= A(V_t, U_t)dt + \Gamma(V_t, U_t)dB_t \end{cases} \quad (1)$$

where  $V_t \in \mathcal{X}_V \subset \mathbb{R}$ ,  $U_t \in \mathcal{X}_U \subset \mathbb{R}^p$  with  $p \geq 1$  and  $B_t$  is an  $m$ -dimensional Brownian motion, from discrete observations of the full system  $(V_t, U_t^T)^T$ , or from discrete observations of  $V_t$  only (partial observations), the latter being the most realistic in applications. Here,  $T$  denotes transposition. Thus, the noise acts directly on  $U_t$ , so these are the *rough* components, whereas  $V_t$  is only indirectly affected by the noise, which is propagated

through  $a(\cdot)$ , and has to depend on  $U_t$  for the model to be hypoelliptic, and thus,  $V_t$  is the *smooth* component.

A prominent example is the large class of stochastic damping Hamiltonian systems, also called Langevin equations, describing the motion of a particle subject to potential, dissipative and random forces (Wu, 2001; Cattiaux et al., 2014b,a, 2015; Comte et al., 2017). In this case  $a(\cdot) = U_t$  and  $A(\cdot) = -c(V_t, U_t)U_t - \nabla P(V_t)$ , for some function  $c(\cdot)$  and where  $P(\cdot)$  is the potential. They typically arise from a second order differential equation, which develops into a higher dimensional system with some coordinates representing positions, and some coordinates representing velocities. The noise is degenerate because it acts directly on the coordinates of the momentum only, and not on the positions. These models have many applications, such as molecular dynamics (Leimkuhler & Matthews, 2015, eqs. (6.30)-(6.31)), stochastic volatility models, paleoclimate research (Ditlevsen et al., 2002), neural mass models (Ableidinger et al., 2017), random mechanics or classical physics. Specific examples are the harmonic oscillator (HO), where  $A(\cdot) = -DV_t - \gamma U_t$ , which will be our first example, the van der Pol oscillator where  $A(\cdot) = \mu(1 - V_t^2)U_t - V_t$  and the Duffing oscillator where  $A(\cdot) = -\delta U_t - \beta V_t - \alpha V_t^3 + \gamma \cos \omega t$ . In this setting, parametric estimation has been considered before, taking advantage of the special structure of  $a(V_t, U_t) = U_t$ . Samson & Thieullen (2012) propose contrast estimators based on the fully observed system, by approximating the unobserved coordinate  $U_t$  by the differences of the observed coordinate  $V_t$ . Pokern et al. (2009) propose a Gibbs algorithm in a Bayesian framework, still relying on the simple form of  $a$ . The particular case of integrated diffusions, where the dynamics of  $U_t$  do not depend on  $V_t$ , has been investigated by Genon-Catalot et al. (2000); Ditlevsen & Sørensen (2004); Gloter (2006).

However, many applications need to allow for a more flexible formulation of function  $a(\cdot)$ . For example, it can be convenient to model parts of a large deterministic system exhibiting multiple time scales by a low dimensional stochastic model, leading to a hypoelliptic structure on the reduced model (Pavliotis & Stuart, 2008). An important field of application is neuronal models of membrane potential evolution, where the noise only acts on the input, or on the ion channel dynamics, leading to hypoelliptic SDEs. Examples are the FitzHugh-Nagumo (FHN) model (DeVille et al., 2005; Leon & Samson, 2017), which is our second example, the Hodgkin-Huxley model (Goldwyn & Shea-Brown, 2011; Tuckwell & Ditlevsen, 2016), or conductance based models with stochastic channel dynamics (Ditlevsen & Greenwood, 2013). Also neural field models are often hypoelliptic (Coombes & Byrne, 2017; Ditlevsen & Löcherbach, 2017). It is therefore important to develop reliable estimation methods for this class of models. A particular sub-class are hypoelliptic homogeneous Gaussian diffusions, where the drift is linear and the diffusion is constant, which were considered by Le Breton & Musiela (1985), and where the transition density is explicitly known. A simple example is the HO mentioned above.

Ergodicity of these models have been studied, based on the hypoellipticity of the system (Mattingly et al., 2002). But even if the model is ergodic, the degenerate noise structure complicates the statistical analysis and many standard tools break down. The main difficulty with hypoelliptic models compared to the elliptic case is the transition density for time  $\Delta$ , which converges towards a point measure when  $\Delta \rightarrow 0$  at a faster rate,  $1/\Delta^2$  (Cattiaux et al., 2014a; Comte et al., 2017), compared to the elliptic case of  $1/\Delta$ .

In general, the transition density is unknown, and the estimation fails if approximating by the Euler-Maruyama scheme, since the scheme can fail to be ergodic for any choice of time step, even if the underlying SDE is (Mattingly et al., 2002). Intuitively, the problem arises because the diffusion matrix is not of full rank, and lower order schemes will have a degenerate variance matrix, even if the underlying model does not, thanks to the hypoellipticity. Pokern et al. (2009) suggested to circumvent this problem by adding the first non-zero noise terms arising in the smooth components of the Itô-Taylor expansion of the process. However, this only remedies the estimation of the diffusion parameters, whereas drift parameters are poorly estimated, see their Fig. 4 and our Table 1. Instead we suggest to approximate the unknown transition density with a higher order scheme, namely the order 1.5 strong Taylor scheme (Kloeden & Platen, 1992), which leads to the same approximation of the variance up to leading order as in Pokern et al. (2009), but also approximates the mean correctly. We propose a contrast based on this scheme, and prove consistency under the standard asymptotics of  $\Delta \rightarrow 0$  and  $n\Delta \rightarrow \infty$ . The proof relies on the higher order approximation of the mean, and thus, provides an explanation of why the consistency failed for the estimator proposed by Pokern et al. (2009). To our surprise, we also obtain asymptotic normality, but with faster convergence rates for parameters of the smooth components than the usual rates of the rough components.

When only partial observations are available, i.e., only some coordinates are observed, the statistical difficulties increase. One particular problem is that the system is coupled, such that the unobserved coordinates are not autonomous, and the hidden Markov model framework is degenerate. We therefore embed the approximation into a filtering algorithm for the unobserved path and a Stochastic Approximation Expectation Maximization (SAEM) algorithm, as suggested in Ditlevsen & Samson (2014) for the elliptic case. This framework furthermore extends the class we can handle considerably by allowing for general drift functions also for the smooth components, as well as for state dependent diffusion matrices.

The running examples throughout the paper are the HO model, where we compare with the estimators proposed in Pokern et al. (2009) and Samson & Thieullen (2012), the FHN model, where we allow for a general  $a(\cdot)$  in the drift of the smooth component, and the Synaptic Inhibition and Excitation (SIE) model, where  $p > 1$  and the diffusion matrix is state dependent. In Section 2 we introduce the general model, the likelihood and notation, we discuss conditions for hypoellipticity, give formulas for moments and introduce the three example models. In Section 3 we give the discretization scheme and present some theoretical results of the scheme needed to show consistency of the estimators. In Section 4 we present contrast estimators for the completely observed case, which will serve as a basis for the partially observed case, where the unobserved components have to be imputed before employing the contrast estimator. In Section 5 we introduce the particle filter to impute the hidden path and the SAEM algorithm to estimate by alternating between imputation and estimation from the fully observed system, and we give indications of how to choose the initial parameter values for the algorithm. In Section 6 we conduct a simulation study on the three example models, and we compare with other estimators. In Appendix, proofs are gathered.

## 2. Models

In this paper we consider parametric estimation for hypoelliptic diffusions defined as solutions to an Itô SDE of the following form:

$$\begin{cases} dV_t &= a(V_t, U_t; \psi)dt \\ dU_t &= A(V_t, U_t; \varphi)dt + \Gamma(V_t, U_t; \sigma)dB_t \end{cases} \quad (2)$$

where  $V_t \in \mathcal{X}_V \subset \mathbb{R}$ ,  $U_t \in \mathcal{X}_U \subset \mathbb{R}^p$  with  $p \geq 1$  and  $B_t$  is an  $m$ -dimensional Brownian motion. Denote the full state space by  $(V_t, U_t^T)^T \in \mathcal{X} \subset \mathbb{R}^{p+1}$ . Functions  $a : \mathcal{X} \mapsto \mathbb{R}$  and  $A : \mathcal{X} \mapsto \mathbb{R}^p$  are drift functions depending on an unknown parameter vector  $\beta = (\psi, \varphi)$ . Denote the full drift vector by  $b = (a, A^T)^T$ . Furthermore,  $\Gamma : \mathcal{X} \mapsto \mathbb{R}^{p \times m}$  is a partial diffusion coefficient matrix depending on an unknown parameter vector  $\sigma$ , the full diffusion matrix being

$$\Sigma(v, u; \sigma) = \begin{bmatrix} \mathbf{0}_m \\ \Gamma(v, u; \sigma) \end{bmatrix} \quad (3)$$

where  $\mathbf{0}_m$  is the  $m$ -dimensional row vector of zeros. Equation (2) is assumed to have a weak solution, and the coefficient functions  $a$ ,  $A$  and  $\Gamma$  are assumed to be smooth enough to ensure the uniqueness in law of the solution, for every  $\beta$  and  $\sigma$ . Most importantly, the process is assumed to be hypoelliptic, meaning that it admits a smooth density with respect to the Lebesgue measure, see Section 2.3. We will restrict attention to diagonal noise in the examples below, such that  $m = p$  and

$$\Gamma(v, u; \sigma) = \begin{bmatrix} \sigma_1(v, u; \sigma) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_p(v, u; \sigma) \end{bmatrix} \quad (4)$$

where  $\sigma_j(v, u; \sigma) > 0$  for  $(v, u^T)^T \in \mathcal{X}$  and  $j = 1, \dots, p$ . In the applications below  $p = 1$  or 2.

### 2.1. Likelihood and objectives

In model (2), the parameters  $\psi$ ,  $\varphi$  and  $\sigma$  are unknown. The objective of this paper is to estimate these from observations of the first coordinate  $V_t$  at discrete times  $t_0, t_1, \dots, t_n$ , with equidistant time steps  $\Delta = t_{j+1} - t_j$ . The ideal would be to maximize the likelihood  $p(V_{0:n}; \beta, \sigma)$  of the data  $V_{0:n} = (V_0, \dots, V_n)$ , where we write  $V_j := V_{t_j}$  for  $j = 0, 1, \dots, n$ . However, the likelihood is intractable, not only because the transition density of model (2) is generally unknown, but also because  $V_{0:n}$  is not Markovian, only  $(V_t, U_t)$  is Markovian. Even if there is no noise on the first coordinate, the hypoellipticity condition implies that the transition density of model (2) exists. Denote the unknown transition density by  $p(V_{t+\Delta}, U_{t+\Delta} | V_t, U_t; \beta, \sigma)$ , then the complete likelihood, assuming all coordinates observed and using the Markov property of  $(V_t, U_t)$ , is given by

$$p(V_{0:n}, U_{0:n}; \beta, \sigma) = \prod_{i=0}^{n-1} p(V_{i+1}, U_{i+1} | V_i, U_i; \beta, \sigma). \quad (5)$$

The marginal likelihood of  $V_{0:n}$ , when only the first coordinate is observed, is a high-dimensional integral,

$$p(V_{0:n}; \beta, \sigma) = \int \prod_{i=0}^{n-1} p(V_{i+1}, U_{i+1} | V_i, U_i; \beta, \sigma) dU_{0:n}, \quad (6)$$

which is difficult to handle.

A standard approximation to the unknown transition density is given by the Euler-Maruyama scheme, where the transition density is normal with mean and variance given by the drift and diffusion coefficients multiplied by  $\Delta$ . However, since the diffusion coefficient on the first coordinate is zero, the normal distribution of the scheme is singular, and the estimation breaks down. The same happens for the slightly improved Milstein scheme. We suggest instead to approximate with a higher order scheme with strong order 1.5, where, as we shall see, a stochastic term of order  $\Delta^{3/2}$  appears in the first coordinate, which is a filtered version of the stochasticity from the other coordinates. This stochasticity is enough to ensure that the estimation procedure works, as long as drift terms of the same order is maintained in the approximation. Denote by

$$p_\Delta(V_{i+1}, U_{i+1} | V_i, U_i; \beta, \sigma) \quad (7)$$

the approximated transition density from this scheme.

In Section 4, we assume all coordinates  $(V_t, U_t^T)^T$  observed at discrete time points, and explain how we can estimate the parameters in that case. In Section 5 we assume only  $V_t$  observed, and suggest to impute the hidden coordinates  $U_t$  and discuss how to maximize the likelihood  $p_\Delta(V_{0:n}; \beta, \sigma)$ . Before detailing the estimation approaches, we give further details on hypoellipticity and some moment properties of the process. Section 3 is devoted to the discretization scheme of order 1.5.

## 2.2. Notation

We write  $\partial_u a(V_i, U_i)$  for the row vector of partial derivatives evaluated at time  $t_i$ ,  $(\partial_{u_1} a(v, u), \dots, \partial_{u_p} a(v, u))|_{(v,u)=(V_i, U_i)}$ , and likewise for the Jacobian matrix of  $A$  and  $\sigma$ . Let  $\nabla_{\bar{\Gamma}}^2(\cdot) = \sum_{j=1}^p \sigma_j^2(v, u) \frac{\partial^2}{\partial u_j^2}(\cdot)$  denote a weighted Laplace type operator, where we choose this notation not to confuse with the time step  $\Delta$ . Operated on vectors it is applied componentwise. Write  $\bar{\Gamma}(v, u) = (\sigma_1(v, u), \dots, \sigma_p(v, u))$  for the vector of entries in the diagonal of matrix (4). Let  $\partial_x f^i$  denote the  $n$ -dimensional row vector of partial derivatives of the  $i$ th component of a generic function  $f : \mathcal{X} \rightarrow \mathbb{R}^n$  with respect to the elements of  $x$ , or just  $\partial_x f$  if  $n = 1$ . We will sometimes use notation  $b$  for the drift, and sometimes  $a, A$ , depending on what is most notational convenient. Note that  $b_1 = a$  and  $b_{j+1} = A_j$  for  $j = 1, \dots, p$ . We sometimes write  $X_t = (V_t, U_t^T)^T$  for the process, but use  $V_t$  and  $U_t$  when we need to distinguish between the smooth and the rough parts of the process. Let  $I_p$  denote the identity matrix of dimension  $p$  and  $\mathbf{1}_p$  the  $p$ -column vector of ones.

## 2.3. Hypoellipticity

An SDE is hypoelliptic if the diffusion matrix is not of full rank, but its solutions admit a smooth density with respect to the Lebesgue measure. Hörmander's theorem asserts

that this is the case if the SDE in its Stratonovich form satisfies the weak Hörmander condition. We write  $\sigma^j : \mathbb{R}^p \rightarrow \mathbb{R}^p$  for the  $m$  column vectors of the diffusion matrix  $\Gamma$ , and  $\tilde{\sigma}^j : \mathbb{R}^{p+1} \rightarrow \mathbb{R}^{p+1}$  for the  $m$  column vectors of the diffusion matrix (3), such that  $\tilde{\sigma}^j = (0, (\sigma^j)^T)^T$ .

For smooth vector fields  $f(x)$  and  $g(x) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , the  $i$ th component of the Lie bracket  $[f, g]$  is defined by  $[f, g]^i = (\partial_x g^i)f - (\partial_x f^i)g$ ,  $i = 1, \dots, n$ . Define the set  $\mathcal{L}$  of vector fields by the initial members  $\tilde{\sigma}^j \in \mathcal{L}$ ,  $j = 1, \dots, m$  and recursively by

$$L \in \mathcal{L} \implies [b, L], [\tilde{\sigma}^1, L], \dots, [\tilde{\sigma}^m, L] \in \mathcal{L}. \quad (8)$$

The weak Hörmander condition is fulfilled if the vectors of  $\mathcal{L}$  span  $\mathbb{R}^{p+1}$ . Assume that the initial members span  $\mathbb{R}^p$ , which in particular is the case if  $\Gamma(v, u)$  is given by (4). Therefore, we only need to check if there exists some  $L \in \mathcal{L}$  which has the first element different from zero. The first iteration of (8) for system (2) yields

$$\begin{aligned} [b, \tilde{\sigma}^j]^1 &= -\partial_u a(v, u)\sigma^j(v, u) \\ [\tilde{\sigma}^i, \tilde{\sigma}^j]^1 &= 0 \end{aligned}$$

for  $i, j = 1, \dots, p$ . This leads us to the following sufficient and necessary condition for system (2) to be hypoelliptic.

**(C1)**  $\forall (v, u^T)^T \in \mathcal{X}$ ,  $\partial_u a(v, u)\sigma^j(v, u) \neq 0$  for at least one  $j = 1, \dots, p$ .

This is a natural assumption; the noise on some of the components of  $u$  should be propagated to the first coordinate, which can only happen if  $a(v, u)$  depends on at least one component of  $u$ . Note that the system has to be in its Stratonovich form, whereas we assume model (2) in its Itô form. However, the condition still holds, since it only involves the drift of the first component. If  $\Gamma(v, u)$  in (4) does not depend on  $(v, u^T)^T$ , the Itô and the Stratonovich forms coincide. If it is state dependent, a conversion from Itô to Stratonovich form will change the drift functions of the  $U_t$  coordinates, but not of  $V_t$ .

#### 2.4. Moments

The distribution of  $X_t = (V_t, U_t^T)^T$  in eq. (2) is in general unknown, but moments can be approximated when  $X_t$  is ergodic. For sufficiently smooth and integrable functions  $f : \mathcal{X} \mapsto \mathbb{R}$ , then

$$\mathbb{E}(f(X_{t+\Delta})|X_t = x) = \sum_{i=0}^k \frac{\Delta^i}{i!} L^i f(x) + \mathcal{O}(\Delta^{k+1}) \quad (9)$$

where  $L$  is the generator of model (2)-(4),

$$Lf(x) = (\partial_x f(x))b(x) + \frac{1}{2} \nabla_{\Gamma}^2 f(x),$$

and  $L^i f$  means  $i$  times iterated application of the generator (Sørensen, 2012). This yields the first conditional moment of the  $j$ 'th component of  $X_t$ ,

$$\mathbb{E}(X_{t+\Delta}^{(j)}|X_t = x) = x^{(j)} + \Delta b_j(x) + \frac{\Delta^2}{2} Lb_j(x) + \mathcal{O}(\Delta^3). \quad (10)$$

In particular, for model (2) we have

$$\mathbb{E}(V_{t+\Delta}|X_t = x) = v + \Delta a(x) + \frac{\Delta^2}{2} \partial_x a(x) b(x) + \frac{\Delta^2}{4} \nabla_{\Gamma}^2 a(x) + \mathcal{O}(\Delta^3), \quad (11)$$

$$\mathbb{E}(U_{t+\Delta}|X_t = x) = u + \Delta A(x) + \frac{\Delta^2}{2} \partial_x A(x) b(x) + \frac{\Delta^2}{4} \nabla_{\Gamma}^2 A(x) + \mathcal{O}(\Delta^3). \quad (12)$$

Furthermore,

$$\text{Var}(V_{t+\Delta}|X_t = x) = \frac{\Delta^3}{3} \partial_u a \Gamma \Gamma^T (\partial_u a)^T + \mathcal{O}(\Delta^4) \quad (13)$$

$$\begin{aligned} \text{Var}(U_{t+\Delta}^j|X_t = x) &= \Delta \sigma_j^2(x) + \quad (14) \\ &\frac{\Delta^2}{2} \left( A_j \partial_{u_j} \sigma_j^2(x) + 2\sigma_j^2(x) \partial_{u_j} A_j(x) + \frac{1}{2} \sigma_j^2(x) \partial_{u_j^2}^2 \sigma_j^2(x) \right) + \mathcal{O}(\Delta^3) \end{aligned}$$

Note how the order of the variance of the first coordinate is  $\Delta^3$ , whereas the mean is of order  $\Delta$ . This is the cause of the statistical difficulties of estimating the parameters.

## 2.5. Three examples

### 2.5.1. Harmonic Oscillator

Harmonic oscillators are common in nature, and the model is central in classical mechanics. Consider the damped harmonic oscillator driven by a white noise forcing (Pokern et al., 2009),

$$\begin{cases} dV_t &= U_t dt \\ dU_t &= (-DV_t - \gamma U_t) dt + \sigma dB_t \end{cases} \quad (15)$$

with  $\gamma, D, \sigma > 0$ . Here,  $p = 1$ . The process has no parameter  $\psi$ , which makes parameter estimation much easier, and thus  $\beta = \varphi = (D, \gamma)$ . The drift function and diffusion coefficient are Lipschitz. For this linear model we know the true distribution. The process is an ergodic Ornstein-Uhlenbeck process, i.e., a Gaussian process. Define

$$X_t = \begin{pmatrix} V_t \\ U_t \end{pmatrix}; \quad M = \begin{pmatrix} 0 & 1 \\ -D & -\gamma \end{pmatrix}; \quad \Sigma = \begin{pmatrix} 0 \\ \sigma \end{pmatrix}$$

Then

$$dX_t = MX_t dt + \Sigma dB_t$$

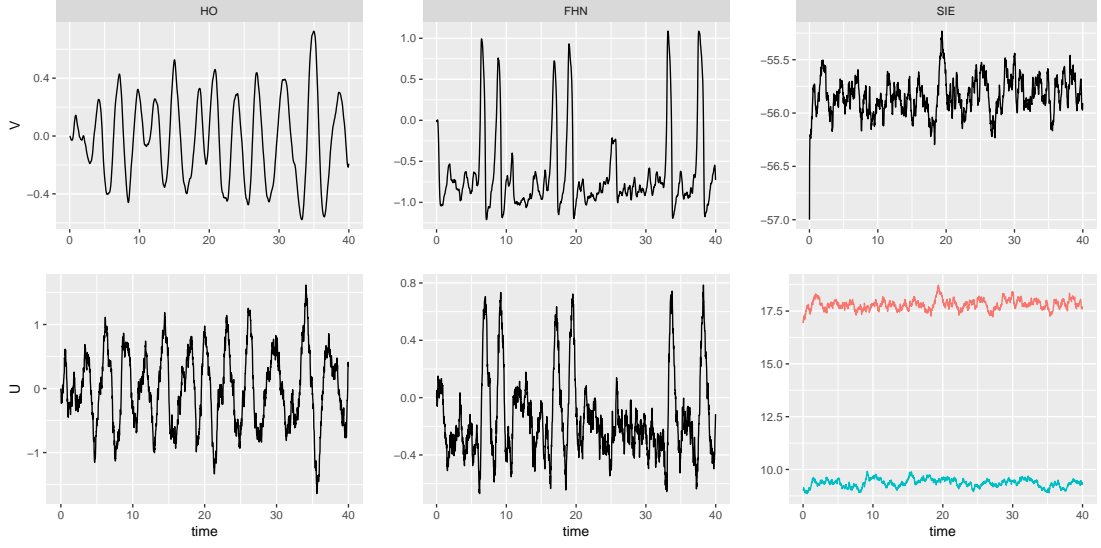
and the conditional distribution is

$$(X_{t+\Delta}|X_t = x) \sim \mathcal{N} \left( e^{\Delta M} x, \int_0^{\Delta} e^{sM} \Sigma \Sigma^T e^{sM^T} ds \right). \quad (16)$$

Let  $d = \frac{1}{2} \sqrt{\gamma^2 - 4D}$ , then

$$\mathbb{E}(X_{t+\Delta}|X_t = x) = e^{-\frac{1}{2}\gamma\Delta} \begin{pmatrix} (\cosh(d\Delta) + \frac{\gamma}{2d} \sinh(d\Delta)) x_1 + \left(\frac{1}{d} \sinh(d\Delta)\right) x_2 \\ \left(-\frac{D}{d} \sinh(d\Delta)\right) x_1 + \left(\cosh(d\Delta) - \frac{\gamma}{2d} \sinh(d\Delta)\right) x_2 \end{pmatrix}. \quad (17)$$





**Fig. 1.** Simulated traces of the three example models. Left: Harmonic Oscillator. Middle: FitzHugh-Nagumo model. Right: Synaptic Inhibition and Excitation model. Upper plots: The smooth coordinate  $V$ . Lower plots: The rough coordinates  $U$ . Parameter values are given in Section 6.

To compare with the analysis of the other models, we make a Taylor expansion in  $\Delta$  up to order 2 obtaining

$$\mathbb{E}(X_{t+\Delta}|X_t = x) = x + \Delta B_{\text{HO}}(x) + \mathcal{O}(\Delta^3) \quad (18)$$

where

$$\Delta B_{\text{HO}}(x) = \Delta \begin{pmatrix} x_2 - (Dx_1 + \gamma x_2) \frac{\Delta}{2} \\ -(Dx_1 + \gamma x_2) + (\gamma(Dx_1 + \gamma x_2) - Dx_2) \frac{\Delta}{2} \end{pmatrix}. \quad (19)$$

Furthermore,

$$\begin{aligned} \text{Var}(X_{t+\Delta}|X_t = x) &= \frac{\sigma^2}{2\gamma D} \begin{bmatrix} 1 & 0 \\ 0 & D \end{bmatrix} + \quad (20) \\ &\frac{\sigma^2 e^{-\gamma\Delta}}{4d^2} \begin{bmatrix} \frac{2}{\gamma} - \frac{d}{D} \sinh(2d\Delta) - \frac{\gamma}{2D} \cosh(2d\Delta) & \cosh(2d\Delta) - 1 \\ \cosh(2d\Delta) - 1 & \frac{2D}{\gamma} + d \sinh(2d\Delta) - \frac{\gamma}{2} \cosh(2d\Delta) \end{bmatrix} \end{aligned}$$

with Taylor expansion up to order 3 in  $\Delta$

$$\text{Var}(X_{t+\Delta}|X_t = x) = \sigma^2 \begin{bmatrix} \frac{1}{3}\Delta^3 & \frac{1}{2}\Delta^2 - \frac{1}{2}\Delta^3\gamma \\ \frac{1}{2}\Delta^2 - \frac{1}{2}\Delta^3\gamma & \Delta - \gamma\Delta^2 + \frac{1}{3}\Delta^3(2\gamma^2 - D) \end{bmatrix} + \mathcal{O}(\Delta^4) \quad (21)$$

where we need a higher order for the variance for later convergence results, since otherwise the variance of the first coordinate is zero.

The invariant distribution is Gaussian,

$$X_\infty \sim \mathcal{N}\left(0, \frac{\sigma^2}{2\gamma D} \begin{bmatrix} 1 & 0 \\ 0 & D \end{bmatrix}\right).$$

The solution of this system has thus moments of any order. An example trace can be found in Figure 1.

### 2.5.2. FitzHugh-Nagumo

A prototype of a model of a spiking neuron is the FitzHugh-Nagumo model, which is a minimal representation of more realistic neuron models, such as the Hodgkin-Huxley model, modeling the neuronal firing mechanisms (FitzHugh, 1961; Nagumo et al., 1962; Hodgkin & Huxley, 1952).

Consider the stochastic hypoelliptic FitzHugh-Nagumo model, defined as the solution to the system

$$\begin{cases} dV_t &= \frac{1}{\varepsilon}(V_t - V_t^3 - U_t + s)dt, \\ dU_t &= (\gamma V_t - U_t + \beta) dt + \sigma dB_t, \end{cases} \quad (22)$$

where the variable  $V_t$  represents the membrane potential of a neuron at time  $t$ , and  $U_t$  is a recovery variable, which could represent channel kinetics. Also in this model is  $p = 1$ .

Parameter  $s$  is the magnitude of the stimulus current. When only  $V_t$  is observed,  $s$  is not identifiable (Jensen et al., 2012). Often  $s$  represents injected current and is thus controlled in a given experiment, and it is therefore reasonable to assume it known, so that  $\psi = \varepsilon$ . Thus, parameters to be estimated are  $\sigma, \varepsilon$  and  $\varphi = (\gamma, \beta)$ .

The distribution of  $X_t = (V_t, U_t)^T$  is unknown, but moments can be approximated by using (9), where the generator of model (22) is

$$Lf(x) = \frac{1}{\varepsilon}(x_1 - x_1^3 - x_2 + s) \frac{\partial f}{\partial x_1} + (\gamma x_1 - x_2 + \beta) \frac{\partial f}{\partial x_2} + \frac{1}{2} \sigma^2 \frac{\partial^2 f}{\partial x_2^2}.$$

We obtain

$$\mathbb{E}(X_{t+\Delta} | X_t = x) = x + \Delta B_{\text{FHN}}(x) + \mathcal{O}(\Delta^3)$$

where

$$\begin{aligned} \Delta B_{\text{FHN}}(x) = & \quad (23) \\ \Delta & \left( \begin{array}{l} \frac{1}{\varepsilon}(x_1 - x_1^3 - x_2 + s) + \frac{\Delta}{2} \frac{1}{\varepsilon} \left( \frac{1}{\varepsilon}(1 - 3x_1^2)(x_1 - x_1^3 - x_2 + s) - (\gamma x_1 - x_2 - \beta) \right) \\ (\gamma x_1 - x_2 + \beta) + \frac{\Delta}{2} \left( \frac{\gamma}{\varepsilon}(x_1 - x_1^3 - x_2 + s) - (\gamma x_1 - x_2 + \beta) \right) \end{array} \right) \end{aligned}$$

and

$$\text{Var}(X_{t+\Delta} | X_t = x) = \sigma^2 \begin{bmatrix} \frac{1}{3} \Delta^3 \frac{1}{\varepsilon^2} & -\frac{1}{2} \Delta^2 \frac{1}{\varepsilon} \\ -\frac{1}{2} \Delta^2 \frac{1}{\varepsilon} & \Delta - \Delta^2 \end{bmatrix} + \mathcal{O}(\Delta^3). \quad (24)$$

An example trace can be found in Figure 1.

### 2.5.3. Synaptic-conductance model

A neuron, which reliably can be characterized as a single electrical compartment, and which receives excitatory and inhibitory synaptic bombardment, has a voltage dynamics across the membrane that can be described by this conductance-based model with

diffusion synaptic input (Dayan & Abbott, 2001; Berg & Ditlevsen, 2013)

$$\begin{cases} CdV_t &= (-G_L(V_t - V_L) - G_{E,t}(V_t - V_E) - G_{I,t}(V_t - V_I) + I_{inj})dt \\ dG_{E,t} &= -\frac{1}{\tau_E}(G_{E,t} - \bar{g}_E)dt + \sigma_E\sqrt{G_{E,t}}dB_{E,t} \\ dG_{I,t} &= -\frac{1}{\tau_I}(G_{I,t} - \bar{g}_I)dt + \sigma_I\sqrt{G_{I,t}}dB_{E,t} \end{cases} \quad (25)$$

where  $C$  is the total capacitance,  $G_L$ ,  $G_E$  and  $G_I$  are the leak, excitation, and inhibition conductances,  $V_L$ ,  $V_E$  and  $V_I$  are their respective reversal potentials, and  $I_{inj}$  is the injected current. The conductances  $G_{e,t}$  and  $G_{i,t}$  are assumed to be stochastic functions of time, where  $(B_{E,t})$  and  $(B_{I,t})$  are two independent Brownian motions. The square roots in the diffusion coefficient ensures that the conductances stay positive. Parameters  $\tau_E, \tau_I$  are time constants,  $\bar{g}_E, \bar{g}_I$  the mean conductances, and  $\sigma_E, \sigma_I$  the diffusion coefficients, scaling the variability of these two processes. Here,  $U_t = (G_{e,t}, G_{i,t})^T$  and  $p = 2$ . We assume to know the capacitance and the reversal potentials, which are easily determined in independent experiments (Berg & Ditlevsen, 2013), as well as  $I_{inj}$ , which is controlled by the experimenter. Thus, there is no parameter  $\psi$ ,  $\varphi = (\bar{g}_E, \bar{g}_I, \tau_E, \tau_I)$ , and  $\sigma = (\sigma_E, \sigma_I)$ .

The distribution of  $X_t = (V_t, U_t^T)^T$  is also unknown for this model, but as before, moments can be approximated by using the generator of model (25),

$$\begin{aligned} Lf(x) &= \frac{1}{C}(-G_L(x_1 - V_L) - x_2(x_1 - V_E) - x_3(x_1 - V_I) + I_{inj})\frac{\partial f}{\partial x_1} \\ &\quad - \frac{1}{\tau_E}(x_2 - \bar{g}_E)\frac{\partial f}{\partial x_2} - \frac{1}{\tau_I}(x_3 - \bar{g}_I)\frac{\partial f}{\partial x_3} + \frac{1}{2}\sigma_E^2x_2\frac{\partial^2 f}{\partial x_2^2} + \frac{1}{2}\sigma_I^2x_3\frac{\partial^2 f}{\partial x_3^2} \end{aligned}$$

and equation (9). We obtain

$$\mathbb{E}(X_{t+\Delta}|X_t = x) = x + \Delta B_{\text{SIE}}(x) + \mathcal{O}(\Delta^3) \quad (26)$$

where

$$\Delta B_{\text{SIE}}(x) = \Delta \begin{pmatrix} b_1(x) - \frac{\Delta}{2C}(b_1(x)(G_L + x_2 + x_3) + b_2(x)(x_1 - V_E) + b_3(x)(x_1 - V_I)) \\ b_2(x) - \frac{\Delta}{2}\left(b_2(x)\frac{1}{\tau_E}\right) \\ b_3(x) - \frac{\Delta}{2}\left(b_3(x)\frac{1}{\tau_I}\right) \end{pmatrix} \quad (27)$$

and

$$\begin{aligned} \text{Var}(X_{t+\Delta}|X_t = x) &= \quad (28) \\ &\begin{bmatrix} 0 & -\frac{\Delta^2}{4}\sigma_E^2(x_1 - V_E) & -\frac{\Delta^2}{4}\sigma_I^2(x_1 - V_I) \\ -\frac{\Delta^2}{4}\sigma_E^2(x_1 - V_E) & \sigma_E^2\left(\Delta x_2 - \frac{\Delta^2}{2\tau_E}(3x_2 - \bar{g}_E)\right) & 0 \\ -\frac{\Delta^2}{4}\sigma_I^2(x_1 - V_I) & 0 & \sigma_I^2\left(\Delta x_3 - \frac{\Delta^2}{2\tau_I}(3x_3 - \bar{g}_I)\right) \end{bmatrix} + \mathcal{O}(\Delta^3). \end{aligned}$$

An example trace can be found in Figure 1. The red trace is the excitatory conductance, the green trace is the inhibitory conductance.

### 3. Discretization scheme

The transition density for model (2) is generally unknown, and a possible approximation to the likelihood function is the likelihood for some approximating scheme of the discretized process  $X_{0:n}$ . We will write  $\tilde{X}_i$  for the approximated process, or  $\tilde{V}_i$  and  $\tilde{U}_i$  where relevant.

The most commonly applied scheme to approximate the likelihood in SDEs, especially for high-frequency data, is the Euler-Maruyama approximation of model (2), which leads to a discretized model defined as follows

$$\begin{aligned}\tilde{V}_{i+1} &= \tilde{V}_i + \Delta a(\tilde{V}_i, \tilde{U}_i), \\ \tilde{U}_{i+1} &= \tilde{U}_i + \Delta A(\tilde{V}_i, \tilde{U}_i) + \Gamma(\tilde{V}_i, \tilde{U}_i)\eta_i,\end{aligned}\quad (29)$$

where  $(\eta_i)$  are centered Gaussian vectors with variance  $\Delta I_p$ . Thus, the transition density of the approximate discretized scheme is a degenerate Gaussian distribution, since there is no stochastic term on the first coordinate. The same happens for the Milstein-scheme with strong order of convergence equal to 1. This causes the estimation to break down, as we will explain later.

#### 3.1. Discretization with 1.5 scheme

We propose to use a higher order scheme, namely the 1.5 strong order scheme (Kloeden & Platen, 1992), using the hypoellipticity of (2) to propagate the noise into the first coordinate. For a diagonal diffusion matrix as in (4) the scheme is as follows, where we for readability have suppressed the dependence on  $(V_i, U_i)$ ,

$$\begin{aligned}\tilde{V}_{i+1} &= \tilde{V}_i + \Delta a + \frac{\Delta^2}{2} \partial_x a b + \frac{\Delta^2}{4} \nabla_{\bar{\Gamma}}^2 a + \partial_u a \Gamma \xi_i \\ \tilde{U}_{i+1} &= \tilde{U}_i + \Delta A + \frac{\Delta^2}{2} \partial_x A b + \frac{\Delta^2}{4} \nabla_{\bar{\Gamma}}^2 A + \Gamma \eta_i + \partial_u A \Gamma \xi_i \\ &\quad + \frac{1}{2} \partial_u \bar{\Gamma} \Gamma (\eta_i^{*2} - \mathbf{1}_p) + \partial_u \bar{\Gamma} A (\Delta \eta_i - \xi_i) + \frac{1}{2} \nabla_{\bar{\Gamma}}^2 \bar{\Gamma} (\Delta \eta_i - \xi_i) \\ &\quad + \frac{1}{2} \partial_u \bar{\Gamma} (\partial_u \sigma \Gamma) \Gamma (\frac{1}{3} \eta_i \eta_i^T - \Delta I_p) \eta_i\end{aligned}\quad (30)$$

where  $(\eta_i)$  are centered Gaussian vectors with variance  $\Delta I_p$ ,  $(\xi_i)$  are centered Gaussian vectors with variance  $\Delta^3/3I_p$ ,  $\text{Cov}(\eta_i, \xi_i) = \Delta^2/2I_p$  and  $\text{Cov}(\eta_i, \xi_j) = 0$  for  $i \neq j$ . Furthermore,  $\eta_i^{*2}$  denotes the vector with the squared entries of  $\eta_i$ . Notice how noise of order  $\Delta^{3/2}$  is now propagated into the first equation, since the last term on the right hand side of (30) is non-zero if condition (C1) is fulfilled. If  $\Gamma$  is independent of the process (additive noise) then the last two lines in (31) are zero.

To simplify the notation later on, we rewrite equations (30)-(31) as

$$\begin{pmatrix} \tilde{V}_{i+1} \\ \tilde{U}_{i+1} \end{pmatrix} = \begin{pmatrix} \tilde{V}_i \\ \tilde{U}_i \end{pmatrix} + \Delta B(\tilde{V}_i, \tilde{U}_i) + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}_{p+1}(0, \Sigma(\tilde{V}_i, \tilde{U}_i)) \quad (32)$$

where  $\Delta B(v, u)_j = \Delta b_j + \frac{\Delta^2}{2} \partial_x b_j b + \frac{\Delta^2}{4} \nabla_{\bar{\Gamma}}^2 b_j$  is the scheme for the drift and  $\Sigma(v, u)$  is

the variance matrix of the scheme. Up to leading order, the variance matrix is given by

$$\Sigma(v, u) = \begin{pmatrix} \partial_u a \Gamma \Gamma^T (\partial_u a)^T \frac{\Delta^3}{3} & \partial_u a \Gamma \Gamma^T \frac{\Delta^2}{2} \\ \Gamma \Gamma^T (\partial_u a)^T \frac{\Delta^2}{2} & \Gamma \Gamma^T \Delta \end{pmatrix} \quad (33)$$

Since the mean term coincides with the true mean up to order  $\Delta^2$ , see eqs. (11) and (12), the functions  $B(v, u)$  for models (15), (22), and (25) are given in (19), (23) and (27), respectively, where  $x = (v, u^T)^T$ . The variance matrix  $\Sigma(\tilde{V}_i, \tilde{U}_i)$  of the above scheme for the three models (15), (22), and (25) are

$$\Sigma_{\text{HO}} = \sigma^2 \begin{pmatrix} \frac{1}{2}\Delta^2 - \frac{1}{3}\Delta^3\gamma & \frac{1}{2}\Delta^2 - \frac{1}{3}\Delta^3\gamma \\ \frac{1}{2}\Delta^2 - \frac{1}{3}\Delta^3\gamma & \Delta - \Delta^2\gamma + \frac{1}{3}\Delta^3\gamma^2 \end{pmatrix} \quad (34)$$

$$\Sigma_{\text{FHN}} = \sigma^2 \begin{pmatrix} \frac{1}{3}\Delta^3\epsilon^{-2} & (-\frac{1}{2}\Delta^2 + \frac{1}{3}\Delta^3)\epsilon^{-1} \\ (-\frac{1}{2}\Delta^2 + \frac{1}{3}\Delta^3)\epsilon^{-1} & \Delta - \Delta^2 + \frac{1}{3}\Delta^3 \end{pmatrix} \quad (35)$$

$$\Sigma_{\text{SIE}}(\tilde{V}_i, \tilde{U}_i) = \begin{pmatrix} \frac{\Delta^3}{3}\tilde{V}_i \left( \sigma_E^2 \tilde{G}_{E,i} + \sigma_I^2 \tilde{G}_{I,i} \right) & -\sigma_E^2 \tilde{V}_i \tilde{G}_{E,i} \left( \frac{\Delta^2}{2} + \frac{\Delta^3}{6\tau_E} \right) & -\sigma_I^2 \tilde{V}_i \tilde{G}_{I,i} \left( \frac{\Delta^2}{2} + \frac{\Delta^3}{6\tau_I} \right) \\ -\sigma_E^2 \tilde{V}_i \tilde{G}_{E,i} \left( \frac{\Delta^2}{2} + \frac{\Delta^3}{6\tau_E} \right) & \sigma_E^2 \tilde{G}_{E,i} \left( \Delta - \frac{\Delta^2}{2\tau_E} + \frac{\Delta^3}{12\tau_E^2} \right) & 0 \\ -\sigma_I^2 \tilde{V}_i \tilde{G}_{I,i} \left( \frac{\Delta^2}{2} + \frac{\Delta^3}{6\tau_I} \right) & 0 & \sigma_I^2 \tilde{G}_{I,i} \left( \Delta - \frac{\Delta^2}{2\tau_I} + \frac{\Delta^3}{12\tau_I^2} \right) \end{pmatrix} \quad (36)$$

For comparison, we recall the variance matrix for the HO model suggested by Pokern et al. (2009),

$$\Sigma_{\text{HO, Pokern}} = \sigma^2 \begin{pmatrix} \frac{1}{3}\Delta^3 & \frac{1}{2}\Delta^2 \\ \frac{1}{2}\Delta^2 & \Delta \end{pmatrix} \quad (37)$$

which coincides with (34) up to lowest order at each matrix entry. Furthermore, it coincides with (33) when  $p = 1, a(v, u) = u$  and  $\Gamma(v, u) = \sigma$ . This is sufficient to obtain reasonable estimates of variance parameters, but their drift approximation is not of sufficiently high order to ensure reasonable estimates of drift parameters.

## 3.2. Theoretical results

### 3.2.1. Strong and weak convergence

The scheme (30)–(31) has a strong order 1.5 and a weak order 2 convergence (Kloeden & Platen, 1992). The following bounds follow by comparing eqs. (9)–(14) with eqs. (30)–(31). These bounds are needed later to prove consistency.

*Proposition 1. [Weak convergence]*

$$\begin{aligned} \mathbb{E}(V_{i+1} - V_i - \Delta B(X_i)_1 | X_i) &= \mathcal{O}(\Delta^3) \\ \mathbb{E}(U_{i+1} - U_i - \Delta B(X_i)_{(-1)} | X_i) &= \mathcal{O}(\Delta^3) \\ \mathbb{E}((V_{i+1} - V_i - \Delta B(X_i)_1)^2 | X_i) &= \frac{\Delta^3}{3} \partial_u a \Gamma \Gamma^T (\partial_u a)^T + \mathcal{O}(\Delta^4) \\ \mathbb{E}((U_{i+1} - U_i - \Delta B(X_i)_{(-1)})(U_{i+1} - U_i - \Delta B(X_i)_{(-1)})^T | X_i) &= \Delta \Gamma \Gamma^T + \mathcal{O}(\Delta^2) \\ \mathbb{E}((V_{i+1} - V_i - \Delta B(X_i)_1)^4 | X_i) &= \mathcal{O}(\Delta^4) \\ \mathbb{E}(((U_{i+1} - U_i - \Delta B(X_i)_{(-1)})(U_{i+1} - U_i - \Delta B(X_i)_{(-1)})^T)^2 | X_i) &= \mathcal{O}(\Delta^2) \end{aligned}$$

### 3.2.2. Convergence of the transition density

To prove the convergence of the particle filter to impute the unobserved path (Section 5), we need a stronger convergence rate of the transition density of the scheme than the result which [Bally & Talay \(1996\)](#) prove on the convergence rate of the density of the Euler-Maruyama scheme when the SDE is elliptic, which we repeat here for convenience.

*Proposition 2 (BALLY & TALAY (1996)' RESULT).* Assume that the drift functions and the diffusion coefficients are 2 times differentiable with bounded derivatives of all orders up to 2. Let  $p(x|x_0)$  denote the exact transition density at  $x$  at time  $\Delta$  given it was at  $x_0$  at time 0 under an elliptic SDE, and let  $p_\Delta(x|x_0)$  denote the transition density of the corresponding Euler approximation. Then

$$\forall x, \quad |p(x|x_0) - p_\Delta(x|x_0)| \leq C \Delta e^{-C'|x-x_0|^2} \quad (38)$$

where  $C$  and  $C'$  are constants not depending on  $\Delta$ .

This result is much more difficult to obtain for a hypoelliptic SDE such as system (2), in particular because the Euler approximation does not propagate noise into the first coordinate. [Bally & Talay \(1996\)](#) propose a perturbed version of the Euler scheme, introducing an independent uncorrelated additive noise at each step (similar to introducing a measurement noise in the model). However, we use a higher order scheme instead to handle the hypoellipticity, which should be more precise. As discussed by [Del Moral et al. \(2001\)](#), so far an estimate like (38) is not available for more sophisticated discretization schemes. In this paper, we conjecture that the result from [Bally & Talay \(1996\)](#) can be extended to a higher order scheme with corresponding improved bounds, under the following assumption

**(H1)** Functions  $a$ ,  $A$ ,  $\Gamma$  are 2 times differentiable with bounded derivatives with respect to  $u$  and  $v$  of all orders up to 2.

*Conjecture 1.* Consider a discretization scheme of model (2), with weak order 2. Let  $p_\Delta(x|x_0)$  denote the corresponding density at  $x$  at time  $\Delta$  with initial condition  $x_0$ . Under H1, there exist constants  $C$  and  $C'$  such that

$$|p(x|x_0) - p_\Delta(x|x_0)| \leq C \Delta^2 e^{-C'|x-x_0|^2} \quad (39)$$

Even if we cannot prove this in general, it can be checked for the HO model, where the exact transition density is known. In fact, for this model we obtain a better bound of  $\Delta^3$ . Assumption **(H1)** is very restrictive, and probably not needed for most models, and in fact not fulfilled for the FHN and the SIE models. However, we conjecture that these models are still sufficiently well-behaved for the above to hold.

#### 4. Complete observations

In this Section we investigate parameter estimation when all coordinates are discretely observed. Later, we extend to the situation where only the first coordinate is observed.

##### 4.1. Contrast estimator

The goal is to estimate the parameter  $\theta = (\psi, \varphi, \sigma)$  by maximum likelihood of the approximate model, with complete likelihood

$$p_{\Delta}(V_{0:n}, U_{0:n}; \theta) = p(V_0, U_0; \theta) \prod_{i=1}^n p_{\Delta}(V_i, U_i | V_{i-1}, U_{i-1}; \theta). \quad (40)$$

It corresponds to a pseudo-likelihood for the exact diffusion, with exact complete likelihood given in (5). The estimator is then the minimizer of minus 2 times the log complete likelihood:

$$\arg \min_{\theta} \sum_{i=0}^{n-1} \left( (X_{i+1} - X_i - \Delta B(X_i; \theta))^T \Sigma_i^{-1} (X_{i+1} - X_i - \Delta B(X_i; \theta)) + \log \det(\Sigma_i) \right) \quad (41)$$

This criteria is ill behaved because the system is hypoelliptic, so the order of the variance for  $V$  is  $\Delta^3$  and of order  $\Delta$  for  $U$ . Therefore, we propose to separate the estimation of parameter  $\psi$  of the first coordinate from parameters  $(\phi, \sigma)$  of the second coordinate.

We thus introduce two new contrasts and their corresponding estimators.

**Definition 1.** *The estimator of the parameters of the first coordinate is given by*

$$\hat{\psi}_n = \arg \min_{\psi} \left( \frac{3}{\Delta^3} \sum_{i=0}^{n-1} \frac{(V_{i+1} - V_i - \Delta B(X_i; \theta)_1)^2}{(\partial_u a(X_i; \psi)) \Gamma \Gamma^T(X_i; \sigma) (\partial_u a(X_i; \psi))^T} \right. \\ \left. + \sum_{i=0}^{n-1} \log((\partial_u a(X_i; \psi)) \Gamma \Gamma^T(X_i; \sigma) (\partial_u a(X_i; \psi))^T) \right) \quad (42)$$

where the parameters  $\varphi$  and  $\sigma$  are fixed to the true values.

The estimator of the parameters of the second coordinate is given by

$$(\hat{\varphi}_n, \hat{\sigma}_n^2) = \arg \min_{\varphi, \sigma^2} \left( \sum_{i=0}^{n-1} \log(\det(\Gamma \Gamma^T(X_i; \sigma))) \right. \\ \left. + \sum_{i=0}^{n-1} (U_{i+1} - U_i - \Delta B(X_i; \theta)_{(-1)})^T (\Delta \Gamma \Gamma^T(X_i; \sigma))^{-1} (U_{i+1} - U_i - \Delta B(X_i; \theta)_{(-1)}) \right) \quad (43)$$

where the parameter  $\psi$  is fixed to its true value.

The first contrast corresponds to the pseudo-likelihood of the marginal distribution of the first coordinate. The second contrast is a simplification of the pseudo-likelihood of the marginal of the coordinates with direct noise: the variance appearing in the pseudo-likelihood is  $\Delta\Gamma\Gamma^T(X_i, \sigma)(1 + o(\Delta))$  and is simplified to  $\Delta\Gamma\Gamma^T(X_i, \sigma)$  in the contrast (43), since the variance is dominated by the lowest order term. The contrasts (42) and (43) require the other parameters to be fixed to their true values. The numerical optimization of the criteria is not sensitive to those fixed values as they appear in higher order term.

#### 4.2. Theoretical properties of the contrast estimators

We start by proving the consistency of the contrast estimators. The asymptotics are in number of observations  $n$  and length of time step between observations  $\Delta_n$ , where we have introduced an index  $n$  to clarify the relevant asymptotics.

*Proposition 3.* Denote by  $\psi_0$  the true values of the parameter, and assume  $(\varphi, \sigma^2)$  known. If  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$  then

$$\hat{\psi}_n \xrightarrow{P} \psi_0$$

*Proposition 4.* Denote by  $(\varphi_0, \sigma_0^2)$  the true values of the parameters, and assume  $\psi$  known. If  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$  then

$$(\hat{\varphi}_n, \hat{\sigma}_n^2) \xrightarrow{P} (\varphi_0, \sigma_0^2)$$

The proofs are given in Appendix 7.

The convergence conditions are standard: the length of the observation interval has to increase for consistency of drift parameters. For consistency of the variance parameter, it can be proven that only  $\Delta_n \rightarrow 0$  and  $n \rightarrow \infty$  are needed, but we will not pursue that here. It is possible to study directly the consistency of the estimator minimizing the pseudo-likelihood (41) using the same arguments as in the proof of Propositions 3 and 4, but at the price of heavy matrix calculations, which are skipped in this paper.

The estimators are asymptotically normal. We prove the result for  $(\hat{\varphi}_n, \hat{\sigma}_n^2)$  and give some partial proofs for  $\hat{\psi}_n$ .

**Theorem 1.** Let  $\nu(\cdot)$  denote the stationary density of model (2). If  $\Delta_n \rightarrow 0$ ,  $n\Delta_n \rightarrow \infty$  and  $n\Delta_n^2 \rightarrow 0$ , then

$$\begin{aligned} \sqrt{n\Delta_n}(\hat{\varphi}_n - \varphi_0) &\xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \left(\nu\left((\partial_\varphi A(\cdot, \varphi_0))^T (\Gamma\Gamma^T(\cdot, \sigma_0))^{-1} (\partial_\varphi A(\cdot, \varphi_0))\right)\right)^{-1}\right) \\ \sqrt{n}(\hat{\sigma}_n - \sigma_0) &\xrightarrow{\mathcal{D}} \mathcal{N}\left(0, 2\left(\nu\left((\partial_\sigma \Gamma\Gamma^T(\cdot, \varphi_0))^T (\Gamma\Gamma^T(\cdot, \sigma_0))^{-1} (\partial_\sigma \Gamma\Gamma^T(\cdot, \varphi_0))\right)\right)^2\right) \end{aligned}$$

**Theorem 2.** Let  $\nu(\cdot)$  denote the stationary density of model (2). Assume the drift function  $a$  can be decomposed as either:  $a(x; \psi) = a_v(v, \psi) + a_u(u)$  or  $a(x; \psi) = a_v(v) + \psi a_u(x)$ . If  $\Delta_n \rightarrow 0$ ,  $n\Delta_n \rightarrow \infty$  and  $n\Delta_n^2 \rightarrow 0$ , then

$$\sqrt{\frac{n}{\Delta_n}}(\hat{\psi}_n - \psi_0) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{3}\left(\nu\left((\partial_\psi a(\cdot, \psi_0))^T (\partial_u a(\cdot, \psi_0)\Gamma\Gamma^T(\cdot, \sigma_0)(\partial_u a(\cdot, \psi_0))^T)^{-1} (\partial_\psi a(\cdot, \varphi_0))\right)\right)^{-1}\right)$$

The proofs are given in Appendix 7.



## 5. Partial observations

In this Section we assume that we do not observe the coordinates  $U_t$ , which is the most relevant case for applications. The likelihood to maximize is therefore not the complete approximate likelihood, but the approximate likelihood  $p_\Delta(V_{0:n}; \theta)$  defined as the integral of the complete approximate likelihood (40) with respect to the hidden components.

$$p_\Delta(V_{0:n}; \theta) = \int p(X_0; \theta) \prod_{i=1}^n p_\Delta(X_i | X_{i-1}; \theta) dU_{0:n}. \quad (44)$$

It corresponds to a discretization of the exact likelihood (6).

The multiple integrals of equation (44) are difficult to handle and it is not possible to maximize the pseudo-likelihood directly. As explained in Section 4, it is easier to maximize the complete approximate likelihood, after imputing the hidden coordinates.

For models where  $a(v, u) = f(v) + g(v)u$  for some functions  $f$  and  $g$  that do not depend on the parameter, such as in the HO model, the imputation is intuitive: the unobserved coordinate  $U_t$  can be approximated by the differences of the observed coordinate  $V_t$ ,  $U_i \approx ((V_{i+1} - V_i)/\Delta - f(V_i))/g(V_i)$ . However, this induces a bias in the estimation of  $\sigma$  (see Samson & Thieullen, 2012, for more details), and is moreover only applicable for drift functions of the observed coordinate such that  $u$  can be isolated. We will take advantage of that when initializing the estimation algorithm in Section 5.3.

In this paper we propose to use a particle filter, also known as Sequential Monte Carlo (SMC), to impute the hidden coordinates. Then, this imputed path is plugged into a stochastic SAEM algorithm (Delyon et al., 1999), as done in Ditlevsen & Samson (2014) for the elliptic case. The SMC proposed by Ditlevsen & Samson (2014) allows to filter a hidden coordinate that is not autonomous in the sense that the equation for  $U_t$  depends on the first coordinate  $V_t$ . Here, we extend the algorithm to the case of  $p$  hidden coordinates, to deal with a  $p + 1$ -dimensional SDE.

More precisely, the observable vector  $V_{0:n}$  is then part of a so-called complete vector  $(V_{0:n}, U_{0:n})$ , where  $U_{0:n}$  has to be imputed. At each iteration of the SAEM algorithm, the unobserved data are filtered under the smoothing distribution  $p_\Delta(U_{0:n} | V_{0:n}; \theta)$  with an SMC. Then the parameters are updated using the pseudo-likelihood proposed in Section 4. Details on the filtering are given in Section 5.1, and the SAEM algorithm is presented in Section 5.2.

### 5.1. Particle filter

The SMC proposed in Ditlevsen & Samson (2014) is designed for a  $p = 1$ -dimensional hidden coordinate. Here we extend to the general case. For notational simplicity,  $\theta$  is omitted in the rest of this Section.

The SMC algorithm provides  $K$  particles  $(U_{0:n}^{(k)})_{k=1, \dots, K}$  and weights  $(W_{0:n}^{(k)})_{k=1, \dots, K}$  such that the empirical measure  $\Psi_n^K = \sum_{k=1}^K W_n(U_{0:n}^{(k)}) \mathbf{1}_{U_{0:n}^{(k)}}$  approximates the conditional smoothing distribution  $p_\Delta(U_{0:n} | V_{0:n})$  (Doucet et al., 2001). The SMC method relies on proposal distributions  $q(U_i | V_i, V_{i-1}, U_{i-1})$  to sample the particles from these distributions. We write  $V_{0:i} = (V_0, \dots, V_i)$  and likewise for  $U_{0:i}$ .

*Algorithm 1 (SMC algorithm).*

- At time  $i = 0$ :  $\forall k = 1, \dots, K$

(a) sample  $U_0^{(k)}$  from  $p(U_0|V_0)$

(b) compute and normalize the weights:

$$w_0(U_0^{(k)}) = p(V_0, U_0^{(k)}), \quad W_0(U_0^{(k)}) = \frac{w_0(U_0^{(k)})}{\sum_{k=1}^K w_0(U_0^{(k)})}$$

- At time  $i = 1, \dots, n$ :  $\forall k = 1, \dots, K$

(a) sample indices  $A_{i-1}^{(k)} \sim r(\cdot | W_{i-1}(U_{0:i-1}^{(1)}), \dots, W_{i-1}(U_{0:i-1}^{(K)}))$  and set

$$U'_{0:i-1} = U_{0:i-1}^{(A_{i-1}^{(k)})}$$

(b) sample  $U_i^{(k)} \sim q(\cdot | V_{i-1:i}, U'_{i-1})$  and set  $U_{0:i}^{(k)} = (U'_{0:i-1}, U_i^{(k)})$

(c) compute and normalize the weights  $W_i(U_{0:i}^{(k)}) = \frac{w_i(U_{0:i}^{(k)})}{\sum_{k=1}^K w_i(U_{0:i}^{(k)})}$  with

$$w_i(U_{0:i}^{(k)}) = \frac{p_\Delta(V_{0:i}, U_{0:i}^{(k)})}{p_\Delta(V_{0:i-1}, U'_{0:i-1}) q(U_i^{(k)} | V_{i-1:i}, U'_{0:i-1})}$$

Natural choices for the proposal  $q$  are either the transition density  $q(U_i | V_{i-1:i}, U_{i-1}) = p_\Delta(U_i | V_{i-1}, U_{i-1})$  or the conditional distribution  $q(U_i | V_{i-1:i}, U_{i-1}) = p_\Delta(U_i | V_{i-1:i}, U_{i-1})$ , following [Ditlevsen & Samson \(2014\)](#). The two choices are not equivalent in the hypoelliptic case because the covariance matrix of the approximate scheme is not diagonal. The conditional distribution gives better results in practice and is used in the simulations. This is due to the extra information provided by also conditioning on  $V_i$ .

In the following, we present some asymptotic convergence results on the SMC algorithm. The assumptions can be found in [Appendix 8.2](#). For a bounded Borel function  $f$ , denote  $\Psi_n^K f = \sum_{k=1}^K f(U_n^{(k)}) W_n(U_{0:n}^{(k)})$ , the conditional expectation of  $f$  under the empirical measure  $\Psi_n^K$ . We also denote  $\pi_{n,\Delta} f = \mathbb{E}_\Delta(f(U_n) | V_{0:n})$  and  $\pi_n f = \mathbb{E}(f(U_n) | V_{0:n})$ , the conditional expectation under the smoothing distribution  $p_\Delta(U_{0:n} | V_{0:n})$  of the approximate model, and under the smoothing distribution  $p(U_{0:n} | V_{0:n})$  of the exact model, respectively.

*Proposition 5.* Under assumption (SMC3), for any  $\varepsilon > 0$ , and for any bounded Borel function  $f$  on  $\mathbb{R}$ , there exist constants  $C_1$  and  $C_2$ , such that

$$\mathbb{P}(|\Psi_n^K f - \pi_{n,\Delta} f| \geq \varepsilon) \leq C_{1,\Delta} \exp\left(-K \frac{\varepsilon^2}{C_{2,\Delta} \|f\|^2}\right) \quad (45)$$

where  $\|f\|$  is the sup-norm of  $f$  and  $C_{1,\Delta}$ ,  $C_{2,\Delta}$  are constants detailed in [Ditlevsen & Samson \(2014\)](#).

The proof is the same as in [Ditlevsen & Samson \(2014\)](#). The hypoellipticity of the process is not a problem as the filter is applied on the discretized process where the noise has been propagated to the first coordinate.

We also study the convergence of the particle filter to the exact filtering distribution, as was done in [Del Moral et al. \(2001\)](#) for an Euler scheme to filter an autonomous hidden coordinate. Here the hypoellipticity plays a role and we have to work with the 1.5 scheme. The proof uses eq. (39) of [Conjecture 1](#).

*Proposition 6.* *Under assumption (SMC3), consider a number of particles  $K$  and a step size  $\Delta$  of the discretisation scheme of weak order 2 such that*

$$\Delta^2 \leq 1/\sqrt{K}. \quad (46)$$

*For any  $\varepsilon > 0$ , and for any bounded Borel function  $f$  on  $\mathbb{R}$ , there exist constants  $C_1$  and  $C_2$ , such that*

$$\mathbb{P}(|\Psi_n^K f - \pi_n f| \geq \varepsilon) \leq C_1 \exp\left(-K \frac{\varepsilon^2}{C_2 \|f\|^2}\right) \quad (47)$$

where  $\|f\|$  is the sup-norm of  $f$ .

*Remark 1.* *The condition (46) linking the step size  $\Delta$  of the discretization scheme and the number of particles  $K$  is a less restrictive condition than the one in [Del Moral et al. \(2001\)](#) ( $\Delta \leq 1/\sqrt{K}$ ) when  $\Delta$  is less than 1. This was expected because the weak order of the scheme is 2, whereas it is 1 for the Euler scheme.*

The proof is given in [Appendix 8.1](#).

## 5.2. SAEM

The estimation method is based on a stochastic version of the EM algorithm ([Dempster et al., 1977](#)), namely the SAEM algorithm ([Delyon et al., 1999](#)) coupled to the SMC algorithm, as already proposed by [Ditlevsen & Samson \(2014\)](#) in the elliptic case. To fulfill convergence conditions of the algorithm, we consider the particular case of a distribution from an exponential family. More precisely, we assume:

- (M1) The parameter space  $\Theta$  is an open subset of  $\mathbb{R}^p$ . The complete likelihood belongs to a curved exponential family, i.e.,  $\log p_\Delta(V_{0:n}, U_{0:n}; \theta) = -\psi(\theta) + \langle S(V_{0:n}, U_{0:n}), \nu(\theta) \rangle$ , where  $\psi$  and  $\nu$  are two functions of  $\theta$ ,  $S(V_{0:n}, U_{0:n})$  is known as the minimal sufficient statistic of the complete model, taking its value in a subset  $\mathcal{S}$  of  $\mathbb{R}^d$ , and  $\langle \cdot, \cdot \rangle$  is the scalar product on  $\mathbb{R}^d$ .

The three models considered in this paper satisfy this assumption. Details of the sufficient statistic  $S$  for the HO model are given in [Appendix 8.3](#).

Under assumption (M1), introducing a sequence of positive numbers  $(a_m)_{m \in \mathbb{N}}$  decreasing to zero, the SAEM-SMC algorithm is defined as follows.

*Algorithm 2 (SAEM-SMC ALGORITHM).*

- Iteration 0: initialization of  $\hat{\theta}_0$  and set  $s_0 = 0$ .
- Iteration  $m \geq 1$ :

S-Step: simulation of the non-observed data  $(U_{0:n}^{(m)})$  with SMC targeting the distribution  $p_{\Delta}(U_{0:n}|V_{0:n}; \hat{\theta}_{m-1})$ .

SA-Step: update  $s_{m-1}$  using the stochastic approximation:

$$s_m = s_{m-1} + a_{m-1} \left[ S(V_{0:n}, U_{0:n}^{(m)}) - s_{m-1} \right] \quad (48)$$

M-Step: update of  $\hat{\theta}_m$  by  $\hat{\theta}_m = \arg \max_{\theta \in \Theta} (-\psi(\theta) + \langle s_m, \nu(\theta) \rangle)$ .

Following [Ditlevsen & Samson \(2014\)](#), we can prove the convergence of the SAEM-SMC algorithm, under standard assumptions that are recalled in the Appendix.

**Theorem 3.** *Assume that (M1)-(M5), (SAEM1)-(SAEM3), and (SMC1)-(SMC3) hold. Then, with probability 1,  $\lim_{m \rightarrow \infty} d(\hat{\theta}_m, \mathcal{L}) = 0$  where  $\mathcal{L} = \{\theta \in \Theta, \partial_{\theta} \ell_{\Delta}(\theta) = 0\}$  is the set of stationary points of the log-likelihood  $\ell_{\Delta}(\theta) = \log p_{\Delta}(V_{0:n}; \theta)$ .*

*Moreover, under assumptions (LOC1)-(LOC3) given in [Delyon et al. \(1999\)](#) on the regularity of the log-likelihood, the sequence  $\hat{\theta}_m$  converges with probability 1 to a (local) maximum of the likelihood  $p_{\Delta}(V_{0:n}; \theta)$ .*

The classical assumptions (M1)-(M5) are usually satisfied. Assumption (SAEM1) is easily satisfied by choosing properly the sequence  $(a_m)$ . Assumptions (SAEM2) and (SAEM3) depend on the regularity of the model. They are satisfied for the 3 approximate models.

The SAEM algorithm converges to a maximum of the approximate likelihood (44). Assuming [Conjecture 1](#), the distance between this pseudo-likelihood and the exact likelihood is bounded.

*Proposition 7.* *Under H1, there exists a constant  $C$ , independent of  $\theta$ , such that for any  $\theta \in \Theta$ , and any vector  $V_{0:n}$*

$$|p(V_{0:n}; \theta) - p_{\Delta}(V_{0:n}; \theta)| \leq Cn\Delta^2$$

The proof follows the proof of [Theorem 2](#) from [Ditlevsen & Samson \(2014\)](#) with a 2 weak order scheme. The result is improved: for a time interval  $T = n\Delta$ , the bound decreases with  $\Delta$ .

### 5.3. Initializing the algorithm

The SAEM algorithm requires initial values of  $\theta$  to start. We detail our strategy to find initial values for the two first models. The SIE model is arbitrarily initialized with unknown parameters fixed to 1.

For the HO model, we run the two-dimensional contrast based on complete observations of the two coordinates. As the  $U$  coordinate is not observed, we replace it by the increments of  $V$ :  $\tilde{U}_i = (V_{i+1} - V_i)/\Delta$ . Then the two dimensional criteria is minimized and initial values  $\hat{D}_0, \hat{\gamma}_0, \hat{\sigma}_0$  are obtained. The value  $\hat{\sigma}_0$  is biased due to the approximation of  $U_i$ , as shown by [Samson & Thieullen \(2012\)](#). Therefore, we apply the bias correction suggested by [Samson & Thieullen \(2012\)](#) and use  $\hat{\sigma}_0 = \sqrt{\frac{3}{2}}\tilde{\sigma}_0$  as initial value.

**Table 1.** Harmonic Oscillator, mean and standard deviation (in parentheses) of estimators calculated from 100 trajectories with  $\Delta = 0.02$  and  $n = 1000$ . Five estimation methods. Complete observations: new contrast estimator given in eq. (43) and Euler contrast from Samson & Thieullen (2012). Partial observations: SAEM, Euler contrast from Samson & Thieullen (2012) and estimator from Pokern et al. (2009) obtained with  $n = 10000$  and  $\Delta = 0.01$  (only the mean values for  $D$  and  $\gamma$  are given in their paper).

		Observations				
		Complete		Partial		
	True	New Contrast	Euler Contrast	SAEM	Euler Contrast	Pokern
$D$	4.0	3.712 (0.634)	3.969 (0.540)	4.081 (0.503)	3.969 (0.540)	1.099 (-)
$\gamma$	0.5	0.701 (0.287)	0.716 (0.273)	0.663 (0.273)	0.754 (0.278)	0.139 (-)
$\sigma$	0.5	0.496 (0.014)	0.496 (0.011)	0.509 (0.012)	0.503 (0.011)	- (-)

For the FHN model, the problem is more difficult because the unknown parameter  $\varepsilon$  appears in the equation of the observed coordinate. We fix an arbitrary value for  $\hat{\varepsilon}_0$ . Then we replace the hidden coordinate  $U_i$  by  $\tilde{U}_i = V_{i+1} - V_i^3 + s - \hat{\varepsilon}_0 \frac{V_{i+1} - V_i}{\Delta}$ . Using  $(V_i, \tilde{U}_i)$ , we minimize the two-dimensional contrast to obtain initial values  $\hat{\gamma}_0, \hat{\beta}_0, \hat{\varepsilon}_0$ .

## 6. Simulation study

### 6.1. Harmonic Oscillator

Parameter values of the Harmonic Oscillator used in the simulations are the same as those of Pokern et al. (2009); Samson & Thieullen (2012). The values are:  $D = 4$ ,  $\gamma = 0.5$ ,  $\sigma = 0.5$ . Trajectories are simulated with the exact distribution eqs. (16)–(17)–(20) with time step  $\Delta = 0.02$  and  $n = 1000$  points. Then  $\theta$  is estimated on each simulated trajectory. A hundred repetitions are used to evaluate the performance of the estimators.

The Particle filter aims at filtering the hidden process ( $U_t$ ) from the observed process ( $V_t$ ). We illustrate its performance on a simulated trajectory, with  $\theta$  fixed at its true value. The SMC Particle filter algorithm is implemented with  $K = 100$  particles and the conditional transition density as proposal.

The performance of the SAEM-SMC algorithm is illustrated on 100 simulated trajectories. The SAEM algorithm is implemented with  $m = 80$  iterations and a sequence  $(a_m)$  equal to 1 during the 30 first iterations and equal to  $a_m = 1/(m - 30)^{0.9}$  for  $m > 30$ . The SMC algorithm is implemented with  $K(m) = 100$  particles at each iteration of the SAEM algorithm. The SAEM algorithm is initialized automatically by maximizing the log likelihood of the complete data, replacing the hidden ( $U_{i\Delta}$ ) by the differences  $((V_{(i+1)\Delta} - V_{i\Delta})/\Delta)$ .

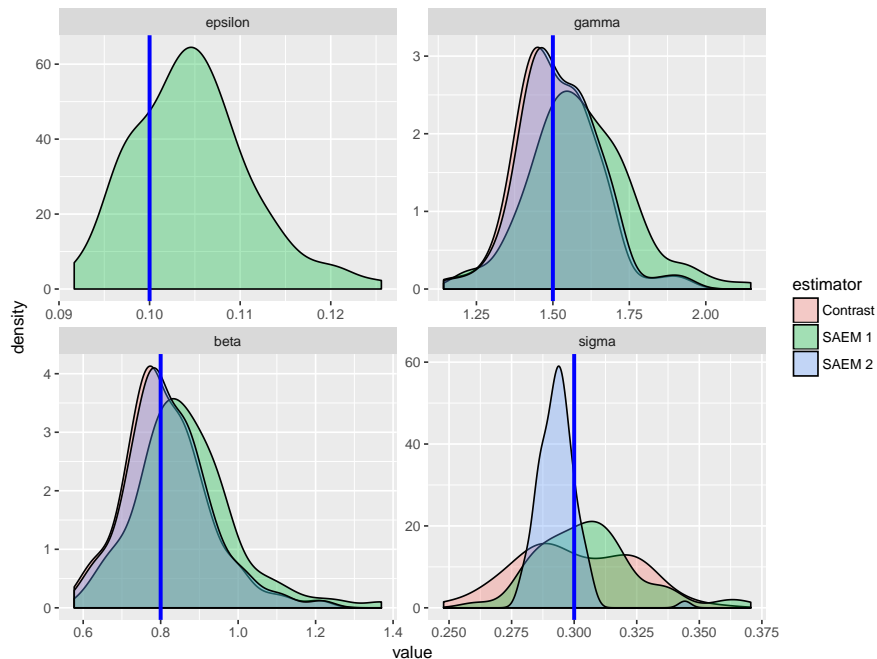
Several estimators are compared. The complete observation case is illustrated with the new contrast estimator (numerical optimisation of contrast (43)) and the Euler contrast from Samson & Thieullen (2012) (explicit estimators). The partial observation case is illustrated with the SAEM estimator and the Euler contrast from Samson & Thieullen (2012). Bayesian results from Pokern et al. (2009) are also recalled, even if they are obtained with a different sampling ( $n = 10000$  and  $\Delta = 0.01$ ). Results are given in Table 1.

The first four estimators give overall acceptable results, while the estimator of Pokern et al. (2009) is seriously biased. The best results are obtained with SAEM. It might seem surprising that the SAEM performs even better than the estimators based on complete observations. This is due to the sensitivity of the numerical optimisation of the contrast (43) to the initial conditions, that were set to  $(3, 1, 1)$ . The stochasticity of the SAEM algorithm helps to avoid local optimization points, while the numerical optimizer might get stuck in some local minimum. The optimization of the Euler contrast is explicit for the HO model, and there is thus no dependence on initial conditions. It therefore outperforms the new contrast for  $D$ .

Comparing SAEM and the Euler contrast for the partial observation case, they give results of the same order, even if slightly better for the SAEM. However, SAEM is much more time consuming. Note also that the SAEM algorithm provides confidence intervals easily, which is not possible with the contrast estimators.

## 6.2. FitzHugh-Nagumo

Parameter values of the FitzHugh-Nagumo used in the simulations are :  $\varepsilon = 0.1$ ,  $s = 0$ ,  $\gamma = 1.5$ ,  $\beta = 0.8$ ,  $\sigma = 0.3$ . Trajectories are simulated with time step  $\delta = 0.002$  and  $n = 1000$  points are subsampled with observation time step  $\Delta = 10\delta$ . Then  $\theta$  is estimated on each simulated trajectory. A hundred repetitions are used to evaluate the performance of the estimators.



**Fig. 2.** FHN estimation results for partial observations. Densities of estimated parameters over 100 repetitions for the new contrast method assuming  $\varepsilon$  known (red), SAEM assuming  $\varepsilon$  known (blue), SAEM estimating  $\varepsilon$  (green). The blue vertical lines are the true values.

**Table 2.** FitzHugh-Nagumo model. Mean and standard deviation (in parentheses) of estimators calculated from 100 trajectories with  $\Delta = 0.02$  and  $n = 1000$ . Seven estimation methods. Complete observations,  $\varepsilon$  fixed: new contrast estimator and Euler contrast from [Samson & Thieullen \(2012\)](#). Complete observations,  $\varepsilon$  estimated: new contrast estimator. Partial observations,  $\varepsilon$  fixed: SAEM, new contrast and Euler contrast from [Samson & Thieullen \(2012\)](#)  $\varepsilon$  fixed. Partial observations,  $\varepsilon$  estimated. SAEM.

True		Complete observations			
		$\varepsilon$ fixed New Contrast	$\varepsilon$ fixed Euler Contrast	$\varepsilon$ estimated New Contrast	
$\varepsilon$	0.1	–	–	0.101 (0.0005)	
$\gamma$	1.5	1.412 (0.221)	1.363 (0.201)	1.516 (0.149)	
$\beta$	0.8	0.826 (0.146)	0.756 (0.131)	0.822 (0.131)	
$\sigma$	0.3	0.303 (0.014)	0.338 (0.024)	0.299 (0.007)	
True		Partial observations			
		$\varepsilon$ fixed SAEM	$\varepsilon$ fixed New Contrast	$\varepsilon$ fixed Euler Contrast	$\varepsilon$ estimated SAEM
$\varepsilon$	0.1	–	–	–	0.105 (0.006)
$\gamma$	1.5	1.523 (0.130)	1.512 (0.129)	1.499 (0.196)	1.592 (0.165)
$\beta$	0.8	0.822 (0.110)	0.815 (0.110)	0.779 (0.107)	0.865 (0.129)
$\sigma$	0.3	0.293 (0.008)	0.300 (0.023)	0.381 (0.038)	0.306 (0.021)

Several estimators are compared. First note that  $\varepsilon$  is difficult to estimate because it appears in the first coordinate. Therefore, we first fix it at its true value. This allows to transform the system into a Langevin equation with  $dV_t = Z_t dt$ , and to apply the Euler contrast proposed by [Samson & Thieullen \(2012\)](#). With  $\varepsilon$  fixed, we compare in the complete observation case the contrast estimator (numerical optimisation of contrast (43)) and the Euler contrast from [Samson & Thieullen \(2012\)](#) (explicit estimators). We also include the estimation of the full parameter vector by the new contrast given in eqs. (42) and (43). In the partial observation case we compare the SAEM estimator, the new contrast and the Euler contrast from [Samson & Thieullen \(2012\)](#). We also run the SAEM algorithm where  $\varepsilon$  is not fixed but estimated.

The SAEM algorithm is implemented with  $m = 350$  iterations and a sequence  $(a_m)$  equal to 1 during the 250 first iterations and equal to  $a_m = 1/(m - 250)^{0.9}$  for  $m > 250$ . The SMC algorithm is implemented with  $K = 100$  particles at each SAEM iteration. The SAEM algorithm is initialized automatically by maximizing the log likelihood of the complete data, replacing the hidden  $(U_{i\Delta})$  by the differences  $(V_{i\Delta} - V_{i\Delta}^3 s - \varepsilon(V_{(i+1)\Delta} - V_{i\Delta})) / \Delta$ ,  $\varepsilon$  being initialized at 0.12. Results are given in Table 2, and densities of estimates in the partially observed case are presented in Figure 2.

The results are acceptable overall. In the complete observation case, the new contrast gives better results than the Euler. This is expected because the new contrast has a higher order of convergence. For the partial observation case, when  $\varepsilon$  is fixed, the performance of SAEM and the contrast are close. The Euler contrast gives better results with partial observations than complete observations (except for  $\sigma$ ). This might be due to the sensitivity of the numerical optimization used to minimize the criteria. Finally, SAEM gives good results when  $\varepsilon$  is estimated, and this is the only method that can



**Table 3.** Synaptic conductance hypoelliptic model, estimation results obtained from 100 repeated trajectories with SAEM, from partial observations.

	Parameters					
	$\tau_E$	$\tau_I$	$\bar{g}_E$	$\bar{g}_I$	$\sigma_E$	$\sigma_I$
true	0.500	1.000	17.800	9.400	0.100	0.100
mean	0.580	1.105	17.696	9.167	0.080	0.021
SE	0.036	0.138	0.124	0.266	0.002	0.003

estimate it.

### 6.3. Synaptic-conductance model

Parameter values of the SIE model used in the simulations are :  $G_L = 50$ ,  $V_L = -70$ ,  $V_E = 0$ ,  $V_I = -80$ ,  $I = -60$ ,  $\tau_E = 0.5$ ,  $\tau_I = 1$ ,  $\bar{g}_E = 17.8$ ,  $\bar{g}_I = 9.4$ ,  $\sigma_E = 0.1$ ,  $\sigma_I = 0.1$ . Initial conditions of the system are  $V_0 = -60$ ,  $G_{e,0} = 10$ ,  $G_{i,0} = 1$ .

Trajectories are simulated with time step  $\delta = 0.002$  and  $n = 1000$  points are subsampled with observation time step  $\Delta = 10\delta$ . Then  $\theta = (\tau_E, \tau_I, \bar{g}_E, \bar{g}_I, \sigma_E, \sigma_I)$  is estimated on each simulated trajectory. A hundred repetitions are used to evaluate the performance of the estimators.

The SAEM algorithm is implemented with  $m = 100$  iterations and a sequence  $(a_m)$  equal to 1 during the 80 first iterations and equal to  $a_m = 1/(m - 80)^{0.9}$  for  $m > 80$ . The SMC algorithm is implemented with  $K(m) = 100$  particles at each iteration of the SAEM algorithm. The SAEM algorithm is initialized with all unknown parameters fixed to 1.

Results are given in Table 3. Excitatory parameters  $(\tau_E, \bar{g}_E, \sigma_E)$  are accurately estimated. Results are less accurate for the inhibitory parameters, especially  $\sigma_I$  which is difficult to identify and estimate. Inhibitory conductances are generally more difficult to estimate, as also observed in Berg & Ditlevsen (2013), where analytic expressions for approximations of the variance of the estimators of the conductances in a similar model were derived from the Fisher Information matrix. This is because the dynamics of  $V_t$  are close to the inhibitory reversal potential  $V_I$ , whereas it is far from the excitatory reversal potential  $V_E$ , and thus, the synaptic drive is higher for excitation.

## Acknowledgements

Adeline Samson has been supported by the LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01). The work is part of the Dynamical Systems Interdisciplinary Network, University of Copenhagen. Villum Visiting Professor Programme funded a longer stay of A. Samson at University of Copenhagen.

## References

M. Ableidinger, et al. (2017). ‘A stochastic version of the Jansen and Rit neural mass model: Analysis and numerics’. *Journal of Mathematical Neuroscience* .



- V. Bally & D. Talay (1996). ‘The law of the Euler scheme for stochastic differential equations I. Convergence rate of the distribution function’. *Monte Carlo Methods and Applications* **2**:93–128.
- R. W. Berg & S. Ditlevsen (2013). ‘Synaptic inhibition and excitation estimated via the time constant of membrane potential fluctuations’. *J Neurophys* **110**:1021–1034.
- P. Cattiaux, et al. (2014a). ‘Estimation for Stochastic Damping Hamiltonian Systems under Partial Observation. I. Invariant density’. *Stochastic Processes and their Applications* **124**:1236–1260.
- P. Cattiaux, et al. (2014b). ‘Estimation for Stochastic Damping Hamiltonian Systems under Partial Observation. II. Drift term’. *ALEA* **11**:359–384.
- P. Cattiaux, et al. (2015). ‘Estimation for Stochastic Damping Hamiltonian Systems under Partial Observation. III. Diffusion term’. *Annals of Applied Probability* .
- F. Comte, et al. (2017). ‘Adaptive estimation for stochastic damping Hamiltonian systems under partial observation’. *Stochastic Processes and Their Applications* To appear.
- S. Coombes & A. Byrne (2017). *Lecture Notes in Nonlinear Dynamics in Computational Neuroscience: from Physics and Biology to ICT*, chap. Next generation neural mass models. Springer. In press.
- P. Dayan & L. Abbott (2001). *Theoretical Neuroscience*. MIT Press.
- P. Del Moral, et al. (2001). ‘The Monte-Carlo method for filtering with discrete-time observations’. *Probab. Theory Related Fields* **120**(3):346–368.
- B. Delyon, et al. (1999). ‘Convergence of a stochastic approximation version of the EM algorithm’. *Ann. Statist.* **27**:94–128.
- A. Dempster, et al. (1977). ‘Maximum likelihood from incomplete data via the EM algorithm’. *Jr. R. Stat. Soc. B* **39**:1–38.
- R. DeVillie, et al. (2005). ‘Two distinct mechanisms of coherence in randomly perturbed dynamical systems’. *Physical Review E* **72**(3, 1).
- P. Ditlevsen, et al. (2002). ‘The fast climate fluctuations during the stadial and interstadial climate states’. *Annals of Glaciology* **35**:457–462.
- S. Ditlevsen & P. Greenwood (2013). ‘The Morris-Lecar neuron model embeds a leaky integrate-and-fire model’. *Journal of Mathematical Biology* **67**(2):239–259.
- S. Ditlevsen & E. Löcherbach (2017). ‘Multi-class oscillating systems of interacting neurons’. *Stochastic Processes and Their Applications* **127**:1840–1869.
- S. Ditlevsen & A. Samson (2014). ‘Estimation in the partially observed stochastic Morris-Lecar neuronal model with particle filter and stochastic approximation methods.’. *Annals of Applied Statistics* **2**:674–702.

- S. Ditlevsen & M. Sørensen (2004). ‘Inference for observations of integrated diffusion processes’. *Scand. J. Statist.* **31**(3):417–429.
- A. Doucet, et al. (2001). ‘An introduction to sequential Monte Carlo methods’. In *Sequential Monte Carlo methods in practice*, Stat. Eng. Inf. Sci., pp. 3–14. Springer, New York.
- R. FitzHugh (1961). ‘Impulses and Physiological States in Theoretical Models of Nerve Membrane’. *Biophysical Journal* **1**(6):445–466.
- V. Genon-Catalot & J. Jacod (1993). ‘On the estimation of the diffusion coefficient for multi-dimensional diffusion processes’. *Ann. Inst. H. Poincaré Probab. Statist.* **29**(1):119–151.
- V. Genon-Catalot, et al. (2000). ‘Stochastic volatility models as hidden Markov models and statistical applications’. *Bernoulli* **6**(6):1051–1079.
- A. Gloter (2006). ‘Parameter estimation for a discretely observed integrated diffusion process’. *Scand. J. Statist.* **33**(1):83–104.
- J. H. Goldwyn & E. Shea-Brown (2011). ‘The What and Where of Adding Channel Noise to the Hodgkin-Huxley Equations’. *PLOS Computational Biology* **7**(11).
- P. Hall & C. C. Heyde (1980). *Martingale limit theory and its application*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York.
- A. Hodgkin & A. Huxley (1952). ‘A quantitative description of membrane current and its application to conduction and excitation in nerve’. *Journal of Physiology-London* **117**(4):500–544.
- A. Jensen, et al. (2012). ‘A Markov Chain Monte Carlo approach to parameter estimation in the FitzHugh-Nagumo model.’. *Physical Review E* **86**:041114.
- M. Kessler (1997). ‘Estimation of an ergodic diffusion from discrete observations’. *Scand. J. Statist.* **24**(2):211–229.
- P. E. Kloeden & E. Platen (1992). *Numerical Solution of Stochastic Differential Equations*. Springer-Verlag Berlin.
- A. Le Breton & M. Musiela (1985). ‘Some parameter estimation problems for hypoelliptic homogeneous Gaussian diffusions’. *Banach Center Publications* **16**(1):337–356.
- B. Leimkuhler & C. Matthews (2015). *Molecular Dynamics with deterministic and stochastic numerical methods*, vol. 39 of *Interdisciplinary Applied Mathematics*. Springer International Publishing Switzerland.
- J. Leon & A. Samson (2017). ‘Hypoelliptic stochastic FitzHugh-Nagumo neuronal model: mixing, up-crossing and estimation of the spike rate’. *Submitted* .
- J. Mattingly, et al. (2002). ‘Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise’. *Stochastic Process. Appl.* **101**:185–232.

- J. Nagumo, et al. (1962). ‘An active pulse transmission line simulating nerve axon’. *Proc. Inst. Radio Eng.* **50**:2061–2070.
- G. Pavliotis & A. Stuart (2008). *Multiscale Methods. Averaging and Homogenization*. Springer.
- Y. Pokern, et al. (2009). ‘Parameter estimation for partially observed hypoelliptic diffusions’. *J. Roy. Stat. Soc. B* **71**(1):49–73.
- A. Samson & M. Thieullen (2012). ‘Contrast estimator for completely or partially observed hypoelliptic diffusion’. *Stochastic Processes and Their Applications* **122**:2521–2552.
- M. Sørensen (2012). *Statistical methods for stochastic differential equations*, chap. Estimating functions for diffusion-type processes, pp. 1–107. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Chapman and Hall/CRC.
- H. C. Tuckwell & S. Ditlevsen (2016). ‘The Space-Clamped Hodgkin-Huxley System with Random Synaptic Input: Inhibition of Spiking by Weak Noise and Analysis with Moment Equations’. *Neural Computation* **28**(10):2129–2161.
- L. Wu (2001). ‘Large and moderate deviations and exponential Convergence for Stochastic damping Hamiltonian Systems’. *Stochastic Process. Appl.* **91**:205–238.

## 7. Appendix: Proofs of Propositions 3 and 4 and Theorems 1 and 2

To ease the notation, we assume that  $p = 1$  throughout this Section. Furthermore, let  $B_i(\theta) := B(X_i; \theta)$  and  $\Gamma_i(\sigma) := \Gamma(X_i; \sigma)$ , and note that  $\Gamma(\cdot)$  is a scalar. Let  $\nu(\cdot)$  denote the stationary density of model (2). We write  $\mathcal{G}_i$  for the filtration endorsed by  $(X_t, t \leq t_i)$ .

### 7.1. Technical lemmas

We first present the equivalent of Lemma 8-10 of Kessler (1997) that are essential for the proofs of consistency. Equivalent of Lemma 7 is presented in Proposition 1.

**Lemma 1.** *Let  $f : \mathbb{R}^{p+1} \times \Theta \rightarrow \mathbb{R}$  be a function with derivatives of polynomial growth in  $x$ , uniformly in  $\theta$ . Assume  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$ . Then*

$$\nu_n(f) := \frac{1}{n} \sum_{i=1}^n f(X_i, \theta) \xrightarrow{P_{\theta_0}} \int f(x, \theta) \nu_0(dx)$$

*uniformly in  $\theta$ .*

The proof is the same as the proof of Lemma 8 in Kessler (1997), since it does not depend on the scheme, only on the ergodicity of the process.

**Lemma 2.** *Let  $f : \mathbb{R}^{p+1} \times \Theta \rightarrow \mathbb{R}$  be a function with derivatives of polynomial growth in  $x$ , uniformly in  $\theta$ .*

(a) Assume  $\Delta_n \rightarrow 0$  and  $n \rightarrow \infty$ . Then

$$Q_{1,n}(f) := \frac{1}{n\Delta_n^2} \sum_{i=0}^{n-1} f(X_i, \theta)(V_{i+1} - V_i - \Delta_n B_i(\theta)_1)^2 \xrightarrow{P_{\theta_0}} 0,$$

uniformly in  $\theta$ .

(b) Assume  $\Delta_n \rightarrow 0$  and  $n \rightarrow \infty$ . Then

$$Q_{2,n}(f) := \frac{1}{n\Delta_n} \sum_{i=0}^{n-1} f(X_i, \theta)(U_{i+1} - U_i - \Delta_n B_i(\theta)_2)^2 \xrightarrow{P_{\theta_0}} \int f(x, \theta) \Gamma^2(x; \sigma_0) \nu_0(dx),$$

uniformly in  $\theta$ .

**Proof of Lemma 2** To prove the first assertion (first coordinate), let

$$\xi_{i+1}(\theta) = \frac{1}{n\Delta_n^2} f(X_i, \theta)(V_{i+1} - V_i - \Delta_n B_i(\theta)_1)^2$$

Due to Proposition 1 and the ergodic theorem, Lemma 1, we have

$$\begin{aligned} \sum_{i=0}^{n-1} \mathbb{E}_\theta(\xi_i(\theta) | \mathcal{G}_{i-1}) &= \mathcal{O}(\Delta_n) \rightarrow 0 \text{ for } \Delta_n \rightarrow 0 \\ \sum_{i=0}^{n-1} \mathbb{E}_\theta(\xi_i(\theta)^2 | \mathcal{G}_{i-1}) &= \frac{1}{n} \mathcal{O}(1) \rightarrow 0 \text{ for } n \rightarrow \infty \end{aligned}$$

Hence, Lemma 9 from [Genon-Catalot & Jacod \(1993\)](#) proves the convergence for all  $\theta$ . Uniformity in  $\theta$  follows as for Lemma 1. The proof of the second assertion is the same. The scaling is different because the variance of the scheme is of order  $\Delta_n$  instead of order  $\Delta_n^3$  (Proposition 1).  $\square$

**Lemma 3.** Let  $f : \mathbb{R}^{p+1} \times \Theta \rightarrow \mathbb{R}$  be a function with derivatives of polynomial growth in  $x$ , uniformly in  $\theta$ .

(a) Assume  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$ . Then

$$I_{1,f} := \frac{1}{n\Delta_n^2} \sum_{i=0}^{n-1} f(X_i, \theta)(V_{i+1} - V_i - \Delta_n B_i(\theta)_1) \xrightarrow{P_{\theta_0}} 0,$$

uniformly in  $\theta$

(b) Assume  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$ . Then

$$I_{2,f} := \frac{1}{n\Delta_n} \sum_{i=0}^{n-1} f(X_i, \theta)(U_{i+1} - U_i - \Delta_n B_i(\theta)_2) \xrightarrow{P_{\theta_0}} 0,$$

uniformly in  $\theta$

(c) Assume  $\Delta_n \rightarrow 0$  and  $n \rightarrow \infty$ . Then

$$I_{3,f} := \frac{1}{n} \sum_{i=0}^{n-1} f(X_i, \theta)(U_{i+1} - U_i - \Delta_n B_i(\theta)_2) \xrightarrow{P_{\theta_0}} 0,$$

uniformly in  $\theta$

**Proof of Lemma 3** To prove the first assertion (first coordinate), let

$$\xi_{i+1}(\theta) = \frac{1}{n\Delta_n^2} f(X_i, \theta)(V_{i+1} - V_i - \Delta_n B_i(\theta)_1)$$

Due to Proposition 1 and intermediate calculations (not shown), we have

$$\begin{aligned} \sum_{i=0}^{n-1} \mathbb{E}_{\theta_0}(\xi_i(\theta) | \mathcal{G}_{i-1}) &= \mathcal{O}(\Delta_n) \rightarrow 0 \text{ for } \Delta_n \rightarrow 0 \\ \sum_{i=0}^{n-1} \mathbb{E}_{\theta_0}(\xi_i(\theta)^2 | \mathcal{G}_{i-1}) &= \frac{1}{n\Delta_n} \mathcal{O}(1) \rightarrow 0 \text{ for } n\Delta_n \rightarrow \infty \end{aligned}$$

Hence, Lemma 9 from [Genon-Catalot & Jacod \(1993\)](#) proves the convergence for all  $\theta$ . The proof of uniformity in  $\theta$  is the same as for Lemma 10 of [Kessler \(1997\)](#).

The proofs of the second and third assertions are the same, only the scalings are different due to Proposition 1.  $\square$

Next we present some Lemmas which are needed to prove asymptotic normality.

**Lemma 4.** (a) Assume that  $n\Delta_n^2 \rightarrow 0$ . Then

$$\frac{1}{\sqrt{n\Delta_n}} \sum_{i=0}^{n-1} f(X_i)(U_{i+1} - U_i - \Delta_n B_i(\theta_0)_2) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \nu_0(\Gamma^2 f^2(\cdot)))$$

(b) Assume that  $n\Delta_n^2 \rightarrow 0$ . Then

$$\frac{1}{\sqrt{n\Delta_n}} \sum_{i=0}^{n-1} f(X_i)(U_{i+1} - U_i)^2 - \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} f(X_i) \Gamma_i^2(\sigma_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 2\nu_0((\Gamma(\cdot))^4 f^2(\cdot)))$$

*Proof of Lemma 4.* Recall that  $U_{i+1} - U_i - \Delta_n B_i(\theta_0)_2 = \sqrt{\Delta_n} \tilde{\xi}_i^U \Gamma_i(\sigma_0) + \epsilon_i^U$ , where  $\sqrt{\Delta_n} \tilde{\xi}_i^U = \eta_i + \partial_u A \xi_i$  and  $\epsilon_i^U$  is the difference between the true process and the scheme. Thus,  $\mathbb{E}(\tilde{\xi}_i^U) = 0$ ,  $\text{Var}(\tilde{\xi}_i^U) = 1 + \mathcal{O}(\Delta_n)$ ,  $\text{Cov}(\xi_i^U, \xi_{i+1}^U) = 0$ , and from Proposition 1 follows that  $\mathbb{E}(\epsilon_i^U) = \mathcal{O}(\Delta_n^3)$  and  $\text{Var}(\epsilon_i^U) = \mathcal{O}(\Delta_n^2)$ . To prove assertion a), rewrite

$$\begin{aligned} \frac{1}{\sqrt{n\Delta_n}} \sum_{i=0}^{n-1} f(X_i)(U_{i+1} - U_i - \Delta_n B_i(\theta_0)_2) &= \frac{\sqrt{\Delta_n}}{\sqrt{n\Delta_n}} \sum_{i=0}^{n-1} \tilde{\xi}_i^U \Gamma_i(\theta) f(X_i) + \frac{1}{\sqrt{n\Delta_n}} \sum_{i=0}^{n-1} \epsilon_i^U f(X_i) \\ &= T_1 + T_2 \end{aligned}$$

Since  $\mathbb{E}(\tilde{\xi}_i^U \Gamma_i(\theta) f(X_i) | \mathcal{G}_i) = 0$  and  $\mathbb{E}((\tilde{\xi}_i^U \Gamma_i(\theta) f(X_i))^2 | \mathcal{G}_i) = (\Gamma_i(\theta))^2 f(X_i)^2 (1 + \mathcal{O}(\Delta_n))$ , then  $\frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E} \left( (\tilde{\xi}_i^U \Gamma_i(\theta) f(X_i))^2 | \mathcal{G}_i \right) \rightarrow \nu_0(\Gamma^2(\cdot, \theta) f(\cdot)^2)$ . Since  $\mathbb{E}((\tilde{\xi}_i^U)^4 (\Gamma_i(\theta))^4 f(X_i)^4 | \mathcal{G}_i)$  is bounded it follows that  $\frac{1}{n^2} \sum_{i=0}^{n-1} \mathbb{E}((\tilde{\xi}_i^U)^4 (\Gamma_i(\theta))^4 f(X_i)^4 | \mathcal{G}_i) \rightarrow 0$ . Using theorem 3.2 in [Hall & Heyde \(1980\)](#), these two conditions are sufficient to imply

$$T_1 = \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} \tilde{\xi}_i^U \Gamma_i(\sigma) f(X_i) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \nu_0(f^2 \Gamma^2)).$$

Then we study  $T_2$ . We have  $\frac{1}{\sqrt{n\Delta_n}} \sum_{i=0}^{n-1} \mathbb{E}(\epsilon_i^U | \mathcal{G}_i) = \sqrt{n} \mathcal{O}(\sqrt{\Delta_n^5})$  and  $\frac{1}{n\Delta_n} \sum_{i=0}^{n-1} \mathbb{E}((\epsilon_i^U)^2 | \mathcal{G}_i) = \mathcal{O}(\Delta_n)$ . The condition  $n\Delta_n^2 \rightarrow 0$  implies  $n\Delta_n^5 \rightarrow 0$  and  $T_2 \rightarrow 0$ . This gives the proof of 1.

To prove assertion 2, rewrite

$$\begin{aligned} \frac{1}{\sqrt{n\Delta_n}} \sum_{i=0}^{n-1} f(X_i) ((U_{i+1} - U_i)^2 - \Delta_n \Gamma_i^2(\sigma_0)) &= \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} \Gamma_i^2(\sigma_0) ((\tilde{\xi}_i^U)^2 - 1) f(X_i) \\ &+ \frac{2}{\sqrt{n\Delta_n}} \sum_{i=0}^{n-1} (\epsilon_i^U + \Delta_n B_i(\theta_0)_2) \Gamma_i(\sigma_0) \tilde{\xi}_i^U f(X_i) + \frac{1}{\sqrt{n\Delta_n}} \sum_{i=0}^{n-1} (\epsilon_i^U + \Delta_n B_i(\theta_0)_i)^2 f(X_i) \\ &= T_1 + T_2 + T_3 \end{aligned}$$

Note that  $\mathbb{E}((\tilde{\xi}_i^U)^2 - 1 | \mathcal{G}_i) = \mathcal{O}(\Delta_n)$  and  $\mathbb{E}(((\tilde{\xi}_i^U)^2 - 1)^2 | \mathcal{G}_i) = 2 + \mathcal{O}(\Delta_n)$ . Thus,

$$\frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E} \left( \left( \Gamma_i^2(\sigma_0) ((\tilde{\xi}_i^U)^2 - 1) f(X_i) \right)^2 | \mathcal{G}_i \right) \rightarrow 2\nu_0(\Gamma^4(\cdot, \theta) f(\cdot)^2).$$

Since  $\mathbb{E} \left( \left( ((\tilde{\xi}_i^U)^2 - 1) (\Gamma_i^2(\theta) f(X_i))^4 | \mathcal{G}_i \right) \right)$  is bounded it follows that

$\frac{1}{n^2} \sum_{i=0}^{n-1} \mathbb{E} \left( \left( ((\tilde{\xi}_i^U)^2 - 1) \Gamma_i^2(\theta) f(X_i) \right)^4 | \mathcal{G}_i \right) \rightarrow 0$ . Using theorem 3.2 in [Hall & Heyde \(1980\)](#), these two conditions are sufficient to imply

$$T_1 = \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} \Gamma_i^2(\sigma_0) ((\tilde{\xi}_i^U)^2 - 1) f(X_i) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 2\nu_0(f^2 \Gamma^4)).$$

We have  $\frac{1}{n\Delta_n} \sum_{i=0}^{n-1} \mathbb{E}((\epsilon_i^U + \Delta_n B_i(\theta_0)_2)^2 \Gamma_i^2(\sigma_0) (\tilde{\xi}_i^U)^2 f^2(X_i) | \mathcal{G}_i) = \mathcal{O}(\Delta_n^2)$  goes to 0 when  $\Delta_n \rightarrow 0$  since  $\mathbb{E}((\tilde{\xi}_i^U)^2 (\epsilon_i^U + \Delta_n B_i(\theta_0)_2)^2 | \mathcal{G}_i) = \mathcal{O}(\Delta_n^2)$ , which implies  $T_2 \rightarrow 0$ . Furthermore, the condition  $n\Delta_n^2 \rightarrow 0$  and  $\mathbb{E}((\epsilon_i^U + \Delta_n B_i(\theta_0)_2)^2 | \mathcal{G}_i) = \mathcal{O}(\Delta_n^2)$  imply  $\mathbb{E}(T_3) \rightarrow 0$ . We also have  $\mathbb{E}((\epsilon_i^U + \Delta_n B_i(\theta_0)_2)^4 | \mathcal{G}_i) = \mathcal{O}(\Delta_n^3)$ . We can conclude that  $T_3 \rightarrow 0$ . This proves Lemma 4.  $\square$

**Lemma 5.** (a) Assume that  $n\Delta_n^2 \rightarrow 0$ . Then

$$\frac{1}{\sqrt{n\Delta_n^3}} \sum_{i=0}^{n-1} f(X_i) (V_{i+1} - V_i - \Delta_n B_i(\theta_0)_1) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \frac{1}{3} \nu_0((\partial_u a)^2 \Gamma^2 f^2(\cdot)))$$

(b) Assume that  $n\Delta_n^2 \rightarrow 0$ . Then

$$\begin{aligned} & \frac{1}{\sqrt{n\Delta_n^3}} \sum_{i=0}^{n-1} f(X_i)(V_{i+1} - V_i - \Delta_n B_i(\theta)_1)^2 - \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} f(X_i) \frac{1}{3} \Gamma_i^2(\sigma_0) (\partial_u a)^2 \\ & \xrightarrow{\mathcal{D}} \mathcal{N}(0, \frac{2}{9} \nu_0 (\Gamma^4(\partial_u a)^4 f^2(\cdot))) \end{aligned}$$

*Proof of Lemma 5.* Recall that  $V_{i+1} - V_i - \Delta_n B_i(\theta)_1 = \sqrt{\Delta_n^3} \tilde{\xi}_i^V \Gamma_i + \epsilon_i^V$ , where  $\sqrt{\Delta_n^3} \tilde{\xi}_i^V = \partial_u a \xi_i$  and  $\epsilon_i^V$  is the difference between the true process and the scheme. Thus,  $\mathbb{E}(\tilde{\xi}_i^V) = 0$ ,  $\text{Var}(\tilde{\xi}_i^V) = \frac{1}{3} (\partial_u a)^2$ ,  $\text{Cov}(\xi_i^V, \xi_{i+1}^V) = 0$ , and from Proposition 1 follows that  $\mathbb{E}(\epsilon_i^V) = \mathcal{O}(\Delta_n^3)$  and  $\text{Var}(\epsilon_i^V) = \mathcal{O}(\Delta_n^4)$ . To prove assertion a), rewrite

$$\begin{aligned} \frac{1}{\sqrt{n\Delta_n^3}} \sum_{i=0}^{n-1} f(X_i)(V_{i+1} - V_i - \Delta_n B_i(\theta)_1) &= \frac{\sqrt{\Delta_n^3}}{\sqrt{n\Delta_n^3}} \sum_{i=0}^{n-1} \tilde{\xi}_i^V \Gamma_i(\theta) f(X_i) + \frac{1}{\sqrt{n\Delta_n^3}} \sum_{i=0}^{n-1} \epsilon_i^V f(X_i) \\ &= T_1 + T_2 \end{aligned}$$

Note that  $\mathbb{E}(\tilde{\xi}_i^V \Gamma_i(\theta) f(X_i) | \mathcal{G}_i) = 0$  and  $\mathbb{E}((\tilde{\xi}_i^V \Gamma_i(\theta) f(X_i))^2 | \mathcal{G}_i) = \frac{1}{3} (\partial_u a \Gamma_i(\theta) f(X_i))^2$ . Thus,  $\frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E}((\tilde{\xi}_i^V \Gamma_i(\theta) f(X_i))^2 | \mathcal{G}_i) \rightarrow \frac{1}{3} \nu_0 ((\partial_u a)^2 \Gamma^2(\cdot, \theta) f(\cdot)^2)$ . Since  $\mathbb{E}((\tilde{\xi}_i^V \Gamma_i(\theta) f(X_i))^4 | \mathcal{G}_i)$  is bounded it follows that  $\frac{1}{n^2} \sum_{i=0}^{n-1} \mathbb{E}((\tilde{\xi}_i^V \Gamma_i(\theta) f(X_i))^4 | \mathcal{G}_i) \rightarrow 0$ . Using theorem 3.2 in [Hall & Heyde \(1980\)](#), these two conditions are sufficient to imply

$$T_1 = \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} \tilde{\xi}_i^V \Gamma_i(\sigma) f(X_i) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \frac{1}{3} \nu_0 (f^2(\partial_u a)^2 \Gamma^2)).$$

To study  $T_2$ , note that  $\frac{1}{\sqrt{n\Delta_n^3}} \sum_{i=0}^{n-1} \mathbb{E}(\epsilon_i^V | \mathcal{G}_i) = \sqrt{n} \mathcal{O}(\sqrt{\Delta_n^3})$  and  $\frac{1}{n\Delta_n^3} \sum_{i=0}^{n-1} \mathbb{E}((\epsilon_i^V)^2 | \mathcal{G}_i) = \mathcal{O}(\Delta_n)$ . The condition  $n\Delta_n^2 \rightarrow 0$  implies  $n\Delta_n^3 \rightarrow 0$  and  $T_2 \rightarrow 0$ . This gives the proof of a).

To prove assertion b), rewrite

$$\begin{aligned} & \frac{1}{\sqrt{n\Delta_n^3}} \sum_{i=0}^{n-1} f(X_i) \left( (V_{i+1} - V_i - \Delta_n B_i(\theta)_1)^2 - \Delta_n^3 \frac{1}{3} (\partial_u a)^2 \Gamma_i^2(\sigma_0) \right) \\ &= \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} \Gamma_i^2(\sigma_0) ((\tilde{\xi}_i^V)^2 - \frac{1}{3} (\partial_u a)^2) f(X_i) \\ & \quad + \frac{2}{\sqrt{n\Delta_n^3}} \sum_{i=0}^{n-1} \epsilon_i^V \Gamma_i(\sigma_0) \tilde{\xi}_i^V f(X_i) + \frac{1}{\sqrt{n\Delta_n^3}} \sum_{i=0}^{n-1} (\epsilon_i^V)^2 f(X_i) \\ &= T_1 + T_2 + T_3 \end{aligned}$$

Note that  $\mathbb{E}((\tilde{\xi}_i^V)^2 - \frac{1}{3} (\partial_u a)^2 | \mathcal{G}_i) = 0$  and  $\mathbb{E}(((\tilde{\xi}_i^V)^2 - \frac{1}{3} (\partial_u a)^2)^2 | \mathcal{G}_i) = \frac{2}{9} (\partial_u a)^4$ . Thus,  $\frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E}(\Gamma_i^2(\sigma_0) ((\tilde{\xi}_i^V)^2 - \frac{1}{3} (\partial_u a)^2) f(X_i)^2 | \mathcal{G}_i) \rightarrow \frac{2}{9} \nu_0 (\Gamma^4(\cdot, \theta) (\partial_u a)^4 f(\cdot)^2)$ . Moreover, since  $\mathbb{E}(((\tilde{\xi}_i^V)^2 - \frac{1}{3} (\partial_u a)^2)^4 (\Gamma_i^2(\theta))^4 f(X_i)^4 | \mathcal{G}_i)$  is bounded, it follows that

$\frac{1}{n^2} \sum_{i=0}^{n-1} \mathbb{E}(((\tilde{\xi}_i^V)^2 - \frac{1}{3}(\partial_u a)^2))^4 (\Gamma_i^2(\theta))^4 f(X_i)^4 | \mathcal{G}_i) \rightarrow 0$ . Using theorem 3.2 in [Hall & Heyde \(1980\)](#), these two conditions are sufficient to imply

$$T_1 = \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} \Gamma_i^2(\sigma_0) ((\tilde{\xi}_i^V)^2 - \frac{1}{3}(\partial_u a)^2) f(X_i) \xrightarrow{D} \mathcal{N}(0, \frac{2}{9} \nu_0(\Gamma^4(\cdot, \theta) (\partial_u a)^4 f(\cdot)^2)).$$

We have  $\frac{1}{n\Delta_n^3} \sum_{i=0}^{n-1} \mathbb{E}((\epsilon_i^V)^2 \Gamma_i^2(\sigma_0) (\tilde{\xi}_i^V)^2 f^2(X_i) | \mathcal{G}_i) = \mathcal{O}(\Delta_n)$  goes to 0 when  $\Delta_n \rightarrow 0$  since  $\mathbb{E}((\tilde{\xi}_i^V)^2 (\epsilon_i^V)^2 | \mathcal{G}_i) = \mathcal{O}(\Delta_n^4)$ , which implies  $T_2 \rightarrow 0$ . Furthermore, the condition  $n\Delta_n^2 \rightarrow 0$  and  $\mathbb{E}((\epsilon_i^V)^2 | \mathcal{G}_i) = \mathcal{O}(\Delta_n^4)$  imply  $T_3 \rightarrow 0$ . This proves Lemma 5.  $\square$

## 7.2. Proof of consistency of $\hat{\sigma}_n^2$ , Proposition 3

The estimator  $\hat{\sigma}_n^2$  is defined as the minimal argument of (43) which for  $p = 1$  reduces to

$$\ell_n(\beta, \sigma) = \sum_{i=0}^{n-1} \frac{(U_{i+1} - U_i - \Delta_n B_i(\beta)_2)^2}{\Delta_n \Gamma_i^2(\sigma)} + \sum_{i=0}^{n-1} \log(\Gamma_i^2(\sigma)). \quad (49)$$

We follow [Kessler \(1997\)](#) and the aim is to prove the following lemma

**Lemma 6.** *Assume  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$ . Then*

$$\frac{1}{n} \ell_n(\beta, \sigma) \xrightarrow{P_{\beta_0}} \int \left( \frac{\Gamma^2(x; \sigma_0)}{\Gamma^2(x; \sigma)} + \log \Gamma^2(x; \sigma) \right) \nu_0(dx) =: F(\sigma, \sigma_0) \quad (50)$$

uniformly in  $\theta$ .

Then, using Lemma 6, we can prove that there exists a subsequence  $n_k$  such that  $(\hat{\varphi}_{n_k}, \hat{\sigma}_{n_k})$  converges to a limit  $(\varphi_\infty, \sigma_\infty^2)$ . Hence, by continuity of  $\sigma \rightarrow F(\sigma, \sigma_0)$ , we have

$$\frac{1}{n_k} \ell_{n_k}(\beta, \sigma) \xrightarrow{P_{\theta_0}} F(\sigma_\infty, \sigma_0).$$

By definition of  $(\hat{\varphi}_{n_k}, \hat{\sigma}_{n_k})$ ,  $F(\sigma_\infty, \sigma_0) \leq F(\sigma_0, \sigma_0)$ .

On the other hand, for all  $y > 0, y_0 > 0, (y_0/y) + \log y \geq 1 + \log y_0$ . Thus,  $F(\sigma_\infty, \sigma_0) = F(\sigma_0, \sigma_0)$ , and by identifiability assumption  $\sigma_\infty^2 = \sigma_0^2$ . Hence, there exists a subsequence of  $\hat{\sigma}_n^2$  that converges to  $\sigma_0^2$ . That proves the consistency of  $\hat{\sigma}_n^2$ . It remains to prove Lemma 6.



*Proof of Lemma 6* We have  $\frac{1}{n}\ell_n(\beta, \sigma) = T_1 + T_2 + T_3 + T_4$  with

$$\begin{aligned} T_1 &= \frac{1}{n} \sum_{i=0}^{n-1} \frac{(U_{i+1} - U_i - \Delta_n B_i(\beta_0)_2)^2}{\Delta_n \Gamma_i^2(\sigma)} \\ T_2 &= \frac{2}{n} \sum_{i=0}^{n-1} \frac{(U_{i+1} - U_i - \Delta_n B_i(\beta_0)_2)(B_i(\beta_0)_2 - B_i(\beta)_2)}{\Gamma_i^2(\sigma)} \\ T_3 &= \frac{\Delta_n}{n} \sum_{i=0}^{n-1} \frac{(B_i(\beta_0)_2 - B_i(\beta)_2)^2}{\Gamma_i^2(\sigma)} \\ T_4 &= \frac{1}{n} \sum_{i=0}^{n-1} \log \Gamma_i^2(\sigma) \end{aligned}$$

We start with  $T_1$ . Lemma 2 implies

$$\frac{1}{n\Delta_n} \sum_{i=1}^{n-1} (U_{i+1} - U_i - \Delta_n B_i(\beta_0)_2)^2 \xrightarrow{P_{\theta_0}} \int \Gamma^2(x; \sigma_0) \nu_0(dx)$$

and thus,  $T_1 \xrightarrow{P_{\theta_0}} \int \frac{\Gamma^2(x; \sigma_0)}{\Gamma^2(x; \sigma)} \nu_0(dx)$ , uniformly in  $\theta$ . Using Lemma 3, we obtain that  $T_2 \xrightarrow{P_{\theta_0}} 0$ , uniformly in  $\theta$ . From Lemma 1 follows  $T_3 \xrightarrow{P_{\theta_0}} 0$  and  $T_4 \xrightarrow{P_{\theta_0}} \int \log \Gamma^2(x; \sigma) \nu_0(dx)$ , uniformly in  $\theta$ . Finally, we obtain (50).  $\square$

### 7.3. Proof of consistency of $\hat{\varphi}_n$ , Proposition 3

The estimator  $\hat{\varphi}_n$  is defined as the minimal argument of (49). Consistency of  $\hat{\varphi}_n$  is deduced from the following lemma.

**Lemma 7.** *Assume  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$ . Then*

$$\frac{1}{n\Delta_n} \ell_n(\beta, \sigma) - \frac{1}{n\Delta_n} \ell_n(\beta_0, \sigma) \xrightarrow{P_{\theta_0}} \int \frac{(A(x; \varphi) - A(x; \varphi_0))^2}{\Gamma^2(x; \sigma)} \nu_0(dx)$$

uniformly in  $\theta$ .

Using Lemma 7, there exists a subsequence  $\hat{\varphi}_{n_k}$  that tends to  $\varphi_\infty$ . Hence,

$$\frac{1}{n_k \Delta_{n_k}} \ell_{n_k}(\hat{\beta}_{n_k}, \sigma) - \frac{1}{n_k \Delta_{n_k}} \ell_{n_k}(\beta_0, \sigma) \xrightarrow{P_{\theta_0}} \int \frac{(A(x; \varphi_\infty) - A(x; \varphi_0))^2}{\Gamma^2(x; \sigma)} \nu_0(dx)$$

The consistency follows by identifiability of  $A(x; \varphi)$ . It remains to prove Lemma 7.

**Proof of Lemma 7.** We have  $\frac{1}{n\Delta_n} \ell_n(\beta, \sigma) - \frac{1}{n\Delta_n} \ell_n(\beta_0, \sigma) = T_1 + T_2$  with

$$\begin{aligned} T_1 &= \frac{2}{n\Delta_n} \sum_{i=0}^{n-1} \frac{(U_{i+1} - U_i - \Delta_n B_i(\beta_0)_2)}{\Gamma_i^2(\sigma)} (B_i(\beta_0)_2 - B_i(\beta)_2) \\ T_2 &= \frac{1}{n} \sum_{i=0}^{n-1} \frac{(B_i(\beta_0)_2 - B_i(\beta)_2)^2}{\Gamma_i^2(\sigma)} \end{aligned}$$

Lemma 3 implies  $T_1 \xrightarrow{P_{\theta_0}} 0$ , uniformly in  $\theta$ . Recall that  $B_i(\beta_0)_2 - B_i(\beta)_2 = A(X_i; \varphi_0) - A(X_i; \varphi) + \mathcal{O}(\Delta_n)$ . Combined with Lemma 1 we obtain  $T_2 \xrightarrow{P_{\theta_0}} \int \frac{(A(x; \varphi) - A(x; \varphi_0))^2}{\Gamma^2(x; \sigma)} \nu_0(dx)$ , uniformly in  $\theta$ . Note that the parameter of the first coordinate  $\psi$  is not involved in the limit. The result applies for any  $\psi$ . This gives the Lemma.  $\square$

#### 7.4. Proof of consistency of $\hat{\psi}_n$ , Proposition 4

To ease the notation, we assume that the drift function  $a$  can be split into two functions of  $v$  and  $u$ :  $a(x; \psi) = a_v(v, \psi_v) + \psi_u a_u(u)$ . Estimator  $\hat{\psi}_n = (\hat{\psi}_{v_n}, \hat{\psi}_{u_n})$  is defined as the minimal argument of (42) which for  $p = 1$  reduces to

$$\ell_n(\psi, \sigma) = \frac{3}{\Delta_n^3} \sum_{i=0}^{n-1} \frac{(V_{i+1} - V_i - \Delta_n B_i(\beta)_1)^2}{\psi_u^2 \Gamma_i^2(\sigma) (a'_u(U_i))^2} + n \log(\psi_u^2). \quad (51)$$

Consistency of  $\hat{\psi}_n$  is deduced from the following lemma.

**Lemma 8.** *Assume  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$ . Then*

$$\frac{\Delta_n}{n} \ell_n(\psi, \sigma) - \frac{\Delta_n}{n} \ell_n(\psi_0, \sigma) \xrightarrow{P_{\theta_0}} \int \frac{(a(x; \psi) - a(x; \psi_0))^2}{\psi_u^2 \Gamma^2(x; \sigma) (a'_u(u))^2} \nu_0(dx)$$

uniformly in  $\theta$ .

**Proof of Lemma 8.** We have  $\frac{\Delta_n}{n} \ell_n(\psi, \sigma) - \frac{\Delta_n}{n} \ell_n(\psi_0, \sigma) = T_1 + T_2 + T_3 + T_4$  with

$$\begin{aligned} T_1 &= \frac{3\Delta_n}{n\Delta_n^3} \sum_{i=0}^{n-1} \frac{(V_{i+1} - V_i - \Delta_n B_i(\beta_0)_1)^2}{\Gamma_i^2(\sigma) (a'_u(U_i))^2} \left( \frac{1}{\psi_u^2} - \frac{1}{\psi_{u,0}^2} \right) \\ T_2 &= \frac{6\Delta_n^2}{n\Delta_n^3} \sum_{i=0}^{n-1} \frac{(V_{i+1} - V_i - \Delta_n B_i(\beta_0)_1) (B_i(\beta_0)_1 - B_i(\beta)_1)}{\Gamma_i^2(\sigma) (a'_u(U_i))^2 \psi_u^2} \\ T_3 &= \frac{3\Delta_n^3}{n\Delta_n^3} \sum_{i=0}^{n-1} \frac{(B_i(\beta_0)_1 - B_i(\beta)_1)^2}{\psi_u^2 \Gamma_i^2(\sigma) (a'_u(U_i))^2} \\ T_4 &= \Delta_n \log(\psi_u^2 / \psi_{u,0}^2) \end{aligned}$$

Lemma 2 implies  $T_1 \xrightarrow{P_{\theta_0}} 0$  and Lemma 3 implies  $T_2 \xrightarrow{P_{\theta_0}} 0$ , uniformly in  $\theta$ . From Lemma 1 combined with  $B_i(\beta_0)_1 - B_i(\beta)_1 = a(X_i; \psi_0) - a(X_i; \psi) + \mathcal{O}(\Delta_n)$  follows that  $T_3 \xrightarrow{P_{\theta_0}} 3 \int \frac{(a(x; \psi) - a(x; \psi_0))^2}{\psi_u^2 \Gamma^2(x; \sigma) (a'_u(u))^2} \nu_0(dx)$ , uniformly in  $\theta$ . Finally  $T_4 \xrightarrow{P_{\theta_0}} 0$ , uniformly in  $\theta$ . Note that the parameter of the second coordinate  $\varphi$  is not involved in the limit. The result applies for any  $\varphi$ . This gives the Lemma.  $\square$

#### 7.5. Proof of the asymptotic normality of $(\hat{\varphi}_n, \hat{\sigma}_n^2)$ (Theorem 1)

*Proof of Theorem 1.* The proof of the asymptotic normality is standard, see for instance Genon-Catalot & Jacod (1993); Kessler (1997). Denote  $\theta = (\psi, \varphi, \sigma)$  and

$\hat{\theta}_n = (\psi_0, \hat{\varphi}_n, \hat{\sigma}_n)$ . Let  $\mathcal{L}_n(\theta) = \ell_n(\beta, \sigma)$  from (49). By Taylor's formula,

$$\int \mathcal{C}_n(\theta_0 + u(\hat{\theta}_n - \theta_0)) du \quad \mathcal{E}_n = \mathcal{D}_n$$

where

$$\mathcal{C}_n(\theta) = \begin{bmatrix} \frac{1}{n\Delta_n} \frac{\partial^2}{\partial \varphi^2} \mathcal{L}_n(\theta) & \frac{1}{n\sqrt{\Delta_n}} \frac{\partial^2}{\partial \varphi \partial \sigma} \mathcal{L}_n(\theta) \\ \frac{1}{n\sqrt{\Delta_n}} \frac{\partial^2}{\partial \varphi \partial \sigma} \mathcal{L}_n(\theta) & \frac{1}{n} \frac{\partial^2}{\partial \sigma^2} \mathcal{L}_n(\theta) \end{bmatrix},$$

$$\mathcal{E}_n = \begin{bmatrix} \sqrt{n\Delta_n}(\hat{\varphi}_n - \varphi_0) \\ \sqrt{n}(\hat{\sigma}_n - \sigma_0) \end{bmatrix}, \quad \mathcal{D}_n = \begin{bmatrix} -\frac{1}{\sqrt{n\Delta_n}} \frac{\partial}{\partial \varphi} \mathcal{L}_n(\theta_0) \\ -\frac{1}{\sqrt{n}} \frac{\partial}{\partial \sigma} \mathcal{L}_n(\theta_0) \end{bmatrix}$$

Lemmas 1-2-3 and 4 allow to prove that

$$\mathcal{D}_n \xrightarrow{\mathcal{D}} \mathcal{N} \left( 0, \begin{bmatrix} 4 \int \frac{(\partial_\varphi B_2)^2}{\Gamma^2}(\cdot; \theta_0) \nu_0(dx) & 0 \\ 0 & 2 \int (\frac{\partial_\sigma \Gamma^2}{\Gamma^2})^2(\cdot; \theta_0) \nu_0(dx) \end{bmatrix} \right) \quad (52)$$

(see Kessler, 1997, for more details). From Lemmas 1-2 follows

$$\mathcal{C}_n(\theta_0) \rightarrow C := \begin{bmatrix} 2 \int \frac{(\partial_\varphi B_2)^2}{\Gamma^2}(\cdot; \theta_0) \nu_0(dx) & 0 \\ 0 & \int (\frac{\partial_\sigma \Gamma^2}{\Gamma^2})^2(\cdot; \theta_0) \nu_0(dx) \end{bmatrix}$$

Using the consistency of  $\hat{\theta}_n$ , we obtain the result.  $\square$

### 7.6. Proof of the asymptotic normality of $(\hat{\psi})$

*Proof of Theorem 2.* Denote  $\hat{\theta}_n = (\hat{\psi}, \varphi_0, \sigma_0)$ . Let  $\mathcal{L}_n(\theta) = \ell_n(\psi, \sigma)$  from (51). By Taylor's formula,

$$\int \mathcal{C}_n(\theta_0 + u(\hat{\theta}_n - \theta_0)) du \quad \mathcal{E}_n = \mathcal{D}_n$$

where

$$\mathcal{C}_n(\theta_0) = \frac{\Delta_n}{n} \frac{\partial^2}{\partial \psi^2} \mathcal{L}_n(\theta), \quad \mathcal{E}_n = \sqrt{\frac{n}{\Delta_n}} (\hat{\psi}_n - \psi_0), \quad \mathcal{D}_n = -\sqrt{\frac{\Delta_n}{n}} \frac{\partial}{\partial \psi} \mathcal{L}_n(\theta_0).$$

Lemma 5 yields

$$\mathcal{D}_n \xrightarrow{\mathcal{D}} \mathcal{N} \left( 0, 12 \int \frac{(\partial_\psi B_1)^2}{\Gamma^2(\partial_u a)^2}(\cdot; \theta_0) \nu_0(dx) \right) \quad (53)$$

and Lemmas 3 and 1 yield

$$\mathcal{C}_n(\theta_0) \rightarrow C := 6 \int \frac{(\partial_\psi B_1)^2}{\Gamma^2(\partial_u a)^2}(\cdot; \theta_0) \nu_0(dx).$$

Using the consistency of  $\hat{\theta}_n$ , we obtain the result.  $\square$

## 8. Appendix: details on SAEM-SMC algorithm

### 8.1. Proof of the convergence of the particle filter, Proposition 6

We follow the proof from [Del Moral et al. \(2001\)](#). We need to study, for each time  $j = 1, \dots, n-1$ , the distance between the two kernels  $\|H_j^\Delta f - H_j f\|$  defined, for any function  $f$  and any  $u$ , as:

$$\begin{aligned} H_j f(u) &= \int p(u', v_{j+1}|u, v_j) f(u') du' \\ H_j^\Delta f(u) &= \int p_\Delta(u', v_{j+1}|u, v_j) f(u') du' \end{aligned}$$

We have

$$\begin{aligned} |H_j f(u) - H_j^\Delta f(u)| &= \left| \int (p(u', v_{j+1}|u, v_j) - p_\Delta(u', v_{j+1}|u, v_j)) f(u') du' \right| \\ &\leq \int |p(u', v_{j+1}|u, v_j) - p_\Delta(u', v_{j+1}|u, v_j)| |f(u')| du' \\ &\leq C\Delta^2 \int e^{-C'\|u-u'\|^2} |f(u')| du' \leq C\Delta^2 \|f\| \end{aligned}$$

where the last inequality follows from Conjecture 1 and inequality (39). Thus,

$$\|H_j f - H_j^\Delta f\| \leq C\Delta^2 \|f\|.$$

Following Proposition 2.1 from [Del Moral et al. \(2001\)](#), we deduce that for all bounded Borel functions  $f$ , and all  $j = 1, \dots, n$ , we have

$$\|\pi_{n,\theta} f - \pi_{n,\Delta,\theta} f\| \leq \alpha C_0 \Delta^2 \|f\| A_j \quad (54)$$

where  $\alpha, C_0, A_j$  are [Del Moral et al. \(2001\)](#)'s constants.

Now, we prove (47). We already proved Proposition 5 and inequality (45). It remains to apply (54). For  $\varepsilon$  such that  $\Delta^2 \|f\| A_j \leq \varepsilon/2$ , we have under condition (46) that  $K\varepsilon^2/\|f\|^2 \geq (2\alpha C_0 A_j)^2$ . Thus we obtain the following bound:

$$\mathbb{P}(|\Psi_{n,\theta}^K f - \pi_{n,\theta} f| \geq \varepsilon) \leq \mathbb{P}(|\Psi_{n,\theta}^K f - \pi_{n,\Delta,\theta} f| \geq \varepsilon/2) \leq C_{1,\Delta} \exp\left(-K \frac{\varepsilon^2}{4C_{2,\Delta} \|f\|^2}\right).$$

Otherwise, when  $\Delta^2 \|f\| A_j \geq \varepsilon/2$ ,  $\mathbb{P}\left(|\Psi_{n,\theta}^K f - \pi_{n,\theta} f| \geq \varepsilon\right)$  is smaller than 1, and the result holds changing the constants to  $C_1$  and  $C_2$  as proposed by [Del Moral et al. \(2001\)](#).  $\square$

### 8.2. Assumptions for SAEM convergence

(M2) The functions  $\psi(\theta)$  and  $\nu(\theta)$  are twice continuously differentiable on  $\Theta$ .

(M3) The function  $\bar{s} : \Theta \rightarrow \mathcal{S}$  defined by  $\bar{s}(\theta) = \int S(v, u) p_\Delta(u|v; \theta) dv du$  is continuously differentiable on  $\Theta$ .

- (M4) The function  $\ell_\Delta(\theta) = \log p_\Delta(v, u, \theta)$  is continuously differentiable on  $\Theta$  and  $\partial_\theta \int p_\Delta(v, u; \theta) dv du = \int \partial_\theta p_\Delta(v, u; \theta) dv du$ .
- (M5) Define  $L : \mathcal{S} \times \Theta \rightarrow \mathbb{R}$  by  $L(s, \theta) = -\psi(\theta) + \langle s, \nu(\theta) \rangle$ . There exists a function  $\hat{\theta} : \mathcal{S} \rightarrow \Theta$  such that  $\forall \theta \in \Theta, \forall s \in \mathcal{S}, L(s, \hat{\theta}(s)) \geq L(s, \theta)$ .
- (SAEM1) The positive decreasing sequence of the stochastic approximation  $(a_m)_{m \geq 1}$  is such that  $\sum_m a_m = \infty$  and  $\sum_m a_m^2 < \infty$ .
- (SAEM2)  $\ell_\Delta : \Theta \rightarrow \mathbb{R}$  and  $\hat{\theta} : \mathcal{S} \rightarrow \Theta$  are  $d$  times differentiable, where  $d$  is the dimension of  $S(v, u)$ .
- (SAEM3) For all  $\theta \in \Theta$ ,  $\int \|S(v, u)\|^2 p_\Delta(u|v; \theta) du < \infty$  and the function  $\Gamma(\theta) = \text{Cov}_\theta(S(\cdot, U_{0:n}))$  is continuous, where the covariance is under the conditional distribution  $p_\Delta(U_{0:n}|V_{0:n}; \theta)$ .
- (SAEM4) Let  $\{\mathcal{F}_m\}$  be the increasing family of  $\sigma$ -algebras generated by the random variables  $s_0, U_{0:n}^{(1)}, U_{0:n}^{(2)}, \dots, U_{0:n}^{(m)}$ . For any positive Borel function  $f$ ,  $\mathbb{E}_\Delta(f(U_{0:n}^{(m+1)})|\mathcal{F}_m) = \int f(u) p_\Delta(u|v, \hat{\theta}_m) du$ .
- (SMC1) The number of particles  $K$  used at each iteration of the SAEM algorithm varies along the iteration: there exists a function  $g(m) \rightarrow \infty$  when  $m \rightarrow \infty$  such that  $K(m) \geq g(m) \log(m)$ .
- (SMC2) The function  $S$  is bounded uniformly in  $u$ .
- (SMC3) The functions  $p_\Delta(V_i|U_i, V_{i-1}, U_{i-1}; \theta)$  are bounded uniformly in  $\theta$ .

### 8.3. Sufficient statistics of the HO model

We detail the sufficient statistics for the HO model. Let us denote  $Y_i = V_{i+1} - V_i - U_i$ . There are 6 statistics:

$$\begin{aligned}
S_1 &= \frac{1}{\Delta^5} \sum_{i=0}^{n-1} \left( -\frac{\Delta^3}{2} U_i Y_i + \frac{\Delta^3}{6} (U_{i+1} - U_i) V_i + \frac{\Delta^4}{3} (U_{i+1} - U_i) U_i \right) \\
S_2 &= \frac{1}{\Delta^5} \sum_{i=0}^{n-1} \left( -\Delta Y_i^2 + \frac{2}{3} \Delta^2 Y_i (U_{i+1} - U_i) + \frac{\Delta^3}{6} (U_{i+1} - U_i) U_i \right) \\
S_3 &= \frac{2}{\Delta^5} \sum_{i=0}^{n-1} \left( \frac{\Delta^4}{12} V_i^2 + \frac{\Delta^5}{12} U_i V_i + \frac{\Delta^6}{12} U_i^2 \right) \\
S_4 &= \frac{2}{\Delta^5} \sum_{i=0}^{n-1} \left( \frac{\Delta^2}{3} Y_i^2 + \frac{\Delta^4}{12} U_i^2 + \frac{\Delta^3}{6} Y_i U_i \right) \\
S_5 &= \frac{1}{\Delta^5} \sum_{i=0}^{n-1} \left( \frac{\Delta^3}{6} Y_i V_i + \frac{\Delta^4}{6} U_i V_i + \frac{\Delta^4}{3} U_i Y_i + \frac{\Delta^5}{12} U_i^2 \right) \\
S_6 &= \frac{1}{\Delta - \Delta^2 + \Delta^3/3} \sum_{i=0}^{n-1} (U_{i+1} - U_i - \Delta(-DV_i - \gamma U_i))^2
\end{aligned}$$

Then the maximisation step and the updates of the parameters are as follows:

$$\begin{aligned}\hat{D}_m &= \frac{S_2 S_5 - S_1 S_4}{S_3 S_4 - S_5^2} \\ \hat{\gamma}_m &= \frac{S_1 S_5 - S_2 S_3}{S_3 S_4 - S_5^2} \\ \hat{\sigma}_m^2 &= \frac{S_6}{n\Delta}\end{aligned}$$