



**HAL**  
open science

# Parametric inference for hypoelliptic ergodic diffusions with full observations

Anna Melnykova

► **To cite this version:**

Anna Melnykova. Parametric inference for hypoelliptic ergodic diffusions with full observations. *Statistical Inference for Stochastic Processes*, 2020, 23, pp.595-635. 10.1007/s11203-020-09222-4. hal-01704010v4

**HAL Id: hal-01704010**

**<https://hal.science/hal-01704010v4>**

Submitted on 15 Jul 2020

**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.

# Parametric inference for hypoelliptic diffusions with full observations

Anna Melnykova \*

## Abstract

Multidimensional hypoelliptic diffusions arise naturally in different fields, for example to model neuronal activity. Estimation in those models is complex because of the degenerate structure of the diffusion coefficient. In this paper we consider hypoelliptic diffusions, given as a solution of two-dimensional stochastic differential equations (SDEs), with the discrete time observations of both coordinates being available on an interval  $T = n\Delta_n$ , with  $\Delta_n$  the time step between the observations. The estimation is studied in the asymptotic setting, with  $T \rightarrow \infty$  as  $\Delta_n \rightarrow 0$ . We build a consistent estimator of the drift and variance parameters with the help of a discretized log-likelihood of the continuous process. We discuss the difficulties generated by the hypoellipticity and provide a proof of the consistency and the asymptotic normality of the estimator. We test our approach numerically on the hypoelliptic FitzHugh-Nagumo model, which describes the firing mechanism of a neuron.

*Keywords:* parametric inference, hypoelliptic diffusions, FitzHugh-Nagumo model, contrast estimator

## 1 Introduction

Hypoelliptic diffusions naturally occur in various applications, most notably in neuroscience, molecular physics and mathematical finance. In particular, neuronal activity of one single neuron (Höpfner et al., 2016, Leon and Samson, 2018), or a large population of neurons (Ditlevsen and Löcherbach, 2017, Ableidinger et al., 2017), or exotic models of option pricing (Malliavin and Thalmaier, 2006) are described by hypoelliptic diffusions.

The main difference between classical (or *elliptic*) and hypoelliptic systems of stochastic differential equations (SDE) is that in the latter case the rank of the diffusion matrix is lower than the dimension of the system itself. More formally, hypoellipticity can be explained in the following way: though the covariance matrix is singular, smooth transition densities with respect

---

\* [anna.melnykova@u-cergy.fr](mailto:anna.melnykova@u-cergy.fr), Université de Cergy-Pontoise, AGM UMR-CNRS 8088 and Université de Grenoble Alpes, LJK UMR-CNRS 5224

to the Lebesgue measure still exist. That is the case when the noise is propagated to all the coordinates through the drift term.

Hypoelliptic SDEs present a number of extra challenges in comparison to elliptic systems. The most important problem is the degenerate diffusion coefficient. As the explicit form of the transition densities of a SDE is often unknown, parametric inference is usually based on discrete approximation with a piece-wise Gaussian processes (see, for example Kessler (1997)). But in the hypoelliptic case this approach cannot be applied directly because the covariance matrix of the approximated transition density is not invertible, since its rank is smaller than the dimension of the system. The second problem is that each coordinate has a variance of different order. It needs to be taken into account when constructing the discretization scheme for approximating the density.

Now let us be more specific. Consider a two-dimensional system of stochastic differential equations of the form:

$$\begin{cases} dX_t = a_1(X_t, Y_t; \theta^{(1)})dt \\ dY_t = a_2(X_t, Y_t; \theta^{(2)})dt + b(X_t, Y_t; \sigma)dW_t, \end{cases} \quad (1)$$

where  $(X_t, Y_t)^T \in \mathbb{R} \times \mathbb{R}$ ,  $(a_1(X_t, Y_t; \theta^{(1)}), a_2(X_t, Y_t; \theta^{(2)}))^T$  is the drift term,  $(0, b(X_t, Y_t; \sigma))^T$  is the diffusion coefficient,  $W$  is a standard Brownian motion defined on some probability space  $(\Omega, \mathcal{F}_t, \mathbb{P})$ , where  $\mathcal{F}_t$  contains the information about all states of the process until time  $t$ .  $(\theta^{(1)}, \theta^{(2)}, \sigma)$  is the vector of the unknown parameters, taken from some compact set  $\Theta_1 \times \Theta_2 \times \Xi$ , and  $(x_0, y_0)$  is a bounded  $\mathcal{F}_0$ -measurable random variable, thus independent on  $(X_t, Y_t)$ .

The goal of this paper is to estimate the parameters of (1) from discrete observations of both coordinates  $X$  and  $Y$ . It is achieved in two steps: first, we consider a discretization scheme in order to approximate the transition density of the continuous process. Then we propose an estimation technique which maximizes the likelihood function of the discrete approximate model in the asymptotic setting  $T = n\Delta_n \rightarrow \infty$  and  $\Delta_n \rightarrow 0$  as  $n \rightarrow \infty$ . Let us discuss the solutions proposed by other authors for hypoelliptic systems.

Several works treat the parametric inference problem for a particular case of system (1), the class of stochastic Damping Hamiltonian systems, also known as Langevin equations (Gardiner and Collett, 1985). These hypoelliptic models arise as the stochastic expansion of 2-dimensional deterministic dynamical systems — for example, the Van der Pol oscillator (Van der Pol, 1920) perturbed by noise. They are defined as the solution of the following SDE:

$$\begin{cases} dX_t = Y_t dt \\ dY_t = a_2(X_t, Y_t; \theta)dt + b(X_t, Y_t; \sigma)dW_t. \end{cases} \quad (2)$$

The particular case of Hamiltonian systems with  $b(X_t, Y_t; \sigma) \equiv \sigma$  and  $a_2(X_t, Y_t; \theta) = g_1(X_t; \theta)X_t + g_2(X_t; \theta)Y_t$  is considered in Ozaki (1989), where the link between the continuous-time solution of (2) and the corresponding discrete model is obtained with the so-called local linearization scheme. The idea of this scheme is the following: for a SDE with a non-constant drift and a

constant variance, its solution can be interval-wise approximated by the solution of a system with a linear drift (see Biscay et al. (1996), Ozaki (2012), Jimenez and Carbonell (2015)). This scheme allows to construct a quasi Maximum Likelihood Estimator. Consistency of the estimator based on the local linearization scheme for Hamiltonian SDEs is proven in León et al. (2018). Pokern et al. (2007) attempt to solve the problem of the non-invertibility of the covariance matrix for the particular case of (2) with a constant variance with the help of Itô-Taylor expansion of the transition density. The parameters are estimated with the Gibbs sampler based on the discretized model with the noise propagated to the first coordinate with order  $\Delta_n^{\frac{3}{2}}$ . This approach allows to estimate the variance coefficient, but it is not suitable for estimating the parameters of the drift term. In Samson and Thieullen (2012) it is shown that a consistent estimator for fully and partially observed data can be constructed using only the discrete approximation of the second equation of system (2). This method can be used in practice even for more general models, on condition that the system (1) can be converted to the simpler form (2). However, this transformation of the observations sampled from the continuous model (1) often requires the prior knowledge of the parameters involved in the first equation which is often unrealistic. The particular case of (1), when  $b(X_t, Y_t; \sigma) \equiv \sigma$  and the drift term is linear and thus the transition density is known explicitly, is treated in Le-Breton and Musiela (1985). A consistent maximum likelihood estimator is then constructed in two steps — first, a covariance matrix of the process is estimated from the available continuous-time observations, and then it is used for computing the parameters of the drift term. The resulting estimator is strongly consistent as  $T \rightarrow \infty$ . Few works are devoted to the non-parametric estimation of the drift and the variance terms (Cattiaux et al., 2014, 2016).

To the best of our knowledge for systems (1) the only reference is Ditlevsen and Samson (2017). They construct a consistent estimator using a discretization scheme based on a Itô-Taylor expansion. To take into account different variance orders in each variable they construct two separate estimators for the rough and the smooth variables. However, this approach has several limitations. The first problem consists in minimizing two different criteria simultaneously, which is not very natural from a numerical point of view. The second problem is that in order to prove the convergence of the estimator for each variable, the parameters in the other variable need to be fixed to their true values.

In this paper, we want to avoid these limitations by proposing a single estimation criteria, able to estimate simultaneously all the parameters. This allows to prove the theoretical convergence of the vector of estimators, without any assumption on the knowledge of a set of parameters. Moreover, we illustrate that from a numerical point of view, the estimation of the first coordinate parameters are less biased than those obtained with approach Ditlevsen and Samson (2017). More precisely, we develop a new estimation method, adjusting the local linearization scheme described in Ozaki (1989) developed for the models of type (2) to the more general class of SDEs (1). Under the hypoellipticity assumption this scheme propagates the noise to both coordinates of the system and allows to obtain an invertible covariance matrix. We start with describing the discretization scheme, proving the rate of convergence even when only one part of the parameters is fixed at

the true value. Then we propose a contrast estimator based on the discretized log-likelihood, estimating the parameters included in the drift and diffusion coefficient simultaneously. Then we study the convergence of the scheme and prove the consistency and the asymptotic normality of the proposed estimator based on the 2-dimensional contrast. To the best of our knowledge, the proof of this consistency is new in the literature. We finish with numerical experiments, testing the proposed approach on the hypoelliptic FitzHugh-Nagumo model and compare it to the other estimators.

This paper is organized as follows: Section 2 presents the model and assumptions. Discrete model is introduced in Section 3. The estimators are studied in Section 4 and illustrated numerically in Section 5. We close with Section 6, devoted to conclusions and discussions. Formal proofs are gathered in Appendix.

## 2 Models and assumptions

We assume that both variables of (1) are discretely observed at equally spaced periods of time  $\Delta_n$  on the time interval  $[0, T]$ . The vector of observations at time  $i\Delta_n$  is denoted by  $Z_i = (X_i, Y_i)^T$ , where  $Z_i$  is the value of the process at time  $i\Delta_n$ ,  $i \in \{0, \dots, n\}$ , where  $n = \frac{T}{\Delta_n}$ . We further assume that it is possible to draw a sufficiently large and accurate sample of data, i.e. that  $T = n\Delta_n \rightarrow \infty$ , with the partition size  $\Delta_n \rightarrow 0$  as  $n \rightarrow \infty$ . Let us also introduce the vector notations:

$$dZ_t = A(Z_t; \theta)dt + B(Z_t; \sigma)dW_t, \quad Z_0 = z_0, \quad t \in [0, T] \quad (3)$$

where  $Z_t = (X_t, Y_t)^T$ ,  $W_t$  is a one-dimensional Brownian motion defined on the filtered probability space,  $z_0 = (x_0, y_0)$ , and  $\theta = (\theta^{(1)}, \theta^{(2)})$  is the vector of drift parameters. Matrices  $A$  and  $B$  represent, respectively, the drift and the diffusion coefficient, that is  $A(Z_t; \theta) = (a_1(X_t, Y_t; \theta^{(1)}), a_2(X_t, Y_t; \theta^{(2)}))^T$  and

$$B(Z_t; \sigma) = \begin{pmatrix} 0 & 0 \\ 0 & b(Z_t; \sigma) \end{pmatrix}. \quad (4)$$

Throughout the paper we use the following abbreviations for the partial derivatives (unless the arguments need to be specified):  $\partial_{x_i} f \equiv \frac{\partial f}{\partial x_i}(x_1, \dots, x_p)$ ,  $\partial_{x_i, x_j}^2 \equiv \frac{\partial^2 f}{\partial x_i \partial x_j}(x_1, \dots, x_p) \forall i, j \in \{1, \dots, p\}$ . We suppress the dependency on the parameters, when their values are clear from the context, otherwise additional indexes are introduced. True values of the parameters are denoted by  $\theta_0^{(1)}, \theta_0^{(2)}, \sigma_0$  and  $P_0$  is the probability  $P_{\theta_0^{(1)}, \theta_0^{(2)}, \sigma_0}$ . We also refer to the variable  $Y_t$  which is directly driven by the Gaussian noise as "rough", and to  $X_t$  as "smooth".

We are working under the following set of assumptions:

- A1** The functions  $a_1(Z_t; \theta^{(1)})$ ,  $a_2(Z_t; \theta^{(2)})$  and  $b(Z_t; \sigma)$  have bounded partial derivatives of every order, uniformly in  $\theta$  and  $\sigma$  respectively. Furthermore  $\partial_y a_1 \neq 0 \quad \forall (x, y) \in \mathbb{R}^2$ .

**A2** *Global Lipschitz and linear growth conditions.*  $\forall t, s \in [0, \infty) \exists K_\theta$  s.t.:

$$\begin{aligned} \|A(Z_t; \theta) - A(Z_s; \theta)\| + \|B(Z_t; \sigma) - B(Z_s; \sigma)\| &\leq K_\theta \|Z_t - Z_s\| \\ \|A(Z_t; \theta)\|^2 + \|B(Z_t; \sigma)\|^2 &\leq K_\theta^2 (1 + \|Z_t\|^2), \end{aligned}$$

where  $\|\cdot\|$  is the standard Euclidean norm.

**A3** The process  $(Z_t)_{t \geq 0}$  is ergodic and there exists a unique invariant probability measure  $\nu_0$  with finite moments of any order.

**A4** The functions  $a_1(Z_t; \theta^{(1)})$ ,  $a_2(Z_t; \theta^{(2)})$  and  $b(Z_t; \sigma)$  are identifiable. By the identifiability we mean that  $u(Z_t; \theta) \equiv u(Z_t; \theta_0) \Leftrightarrow \theta = \theta_0$ . The diffusion coefficient is assumed to be strictly positive with a non-zero derivative with respect to  $\sigma$ , that is  $b(Z_t; \sigma) > 0$ ,  $\partial_\sigma b(Z_t; \sigma) \neq 0 \quad \forall t$ .

Further, we introduce a rather restrictive assumption, which is required for the study of the consistency and asymptotic normality of the estimator for the parameters of the rough variable.

**A5** The function  $a_1(Z_t; \theta^{(1)})$  can be represented in the following form:

$$a_1(z; \theta^{(1)}) = f(z) + (\theta^{(1)})^T g(x), \quad (5)$$

where  $g(x)$  is a vector-valued function of the same dimension as vector  $\theta^{(1)}$ ,  $f(z)$  is a continuous function. Functions  $f(z)$  and  $g(x)$  are such that the assumptions (A1)-(A4) hold everywhere.

The full force of the last assumption will be explained in Section 4.1, when the estimator is introduced. Representation (5) implies that the derivative  $\partial_y a_1$  does not depend on the parameter  $\theta^{(1)}$ . It will be shown in Section 3 that the marginal variance of variable  $X$  depends on  $\partial_y a_1$ . However, condition (A5) ensures that  $\theta^{(1)}$  appears only in the drift, which simplifies the analysis. We note however that in practice the estimator shows good results even when (A5) does not hold, as it will be shown in the simulation study.

Assumption (A1) ensures that the system is hypoelliptic in the sense of the stochastic calculus of variations (Nualart, 2006, Malliavin and Thalmaier, 2006). In order to prove it we first write the coefficients of the system (3) as two vector fields, converting (3) from the Itô to the Stratonovich form:

$$A_0(x, y) = \begin{pmatrix} a_1(x, y; \theta^{(1)}) \\ a_2(x, y; \theta^{(2)}) - \frac{1}{2} b(x, y; \sigma) \partial_y b(x, y; \sigma) \end{pmatrix} \quad A_1(x, y) = \begin{pmatrix} 0 \\ b(x, y; \sigma) \end{pmatrix}.$$

Then their Lie bracket is equal to

$$[A_0, A_1] = \begin{pmatrix} \partial_y a_1(x, y; \theta^{(1)}) \\ \partial_x a_2(x, y; \theta^{(2)}) - \frac{1}{2} \partial_x b(x, y; \sigma) \partial_{xy}^2 b(x, y; \sigma) \end{pmatrix}.$$

By (A1) the first element of this vector is not equal to 0, thus we conclude that  $A_1$  and  $[A_0, A_1]$  generate  $\mathbb{R}^2$ . That means that the weak Hörmander condition is satisfied and as a result the transition density for the system (3) exists, though not necessarily has an explicit form. (A2) is a sufficient condition to ensure the existence and uniqueness in law of the strong solution of system (3), moreover this solution is Feller (Revuz and Yor, 2013). However, Feller property and the existence of the strong solution often hold under milder assumptions, thus (A2) can be relaxed. (A4) is a standard condition which is needed to prove the consistency of the estimator. (A3) ensures that we can apply the weak ergodic theorem. That is, for any continuous function  $f$  of polynomial growth at infinity:

$$\frac{1}{T} \int_0^T f(Z_s) ds \xrightarrow{T \rightarrow \infty} \nu_0(f) \quad \text{a.s.},$$

where  $\nu_0(\cdot)$  is the stationary density of model (3). By choosing this notation we highlight that  $\nu_0(\cdot) := \nu_{\theta_0^{(1)}, \theta_0^{(2)}, \sigma_0}(\cdot)$ .

We do not investigate the conditions under which the process  $(Z_t)_{t \geq 0}$  is ergodic as it is not the main focus of this work. Ergodicity of the stochastic damping Hamiltonian system (2) is studied in Wu (2001). Conditions for a wider class of hypoelliptic SDEs can be found in Roynette (1975), Mattingly et al. (2002), Arnold and Kliemann (1987). It is also important to know that if the process  $(Z_t)_{t \geq 0}$  is ergodic then its sampling  $\{Z_i\}$ ,  $i \in \{0, \dots, n\}$  is also ergodic (Genon-Catalot et al., 2000).

### 3 Discrete model

In this section we introduce a Local Linearization scheme, which approximates the solution  $Z_t$  of (3) by the solution of a piece-wise linear autonomous equation. This solution has a piece-wise Gaussian density. We use the approximated solution to construct a discretization scheme and study its properties.

#### 3.1 Approximation with the Local Linearization scheme

Local Linearization refers to the family of approximation schemes studied by different authors (Biscay et al., 1996, Ozaki, 2012, Jimenez and Carbonell, 2015). The idea consists in approximating the solution of a general SDE by the solution of an autonomous linear SDE, which can be solved explicitly. Before we proceed to the derivation of the scheme, let us introduce additional notations. The Jacobian of the drift vector  $A(z; \theta)$  is given by

$$\begin{pmatrix} \partial_x a_1(x, y; \theta^{(1)}) & \partial_y a_1(x, y; \theta^{(1)}) \\ \partial_x a_2(x, y; \theta^{(2)}) & \partial_y a_2(x, y; \theta^{(2)}) \end{pmatrix} =: J(z; \theta). \quad (6)$$

We also define the Hessian matrix of the  $j$ -th coordinate ( $j = 1, 2$ ) in the drift vector  $A(Z_t; \theta)$  as:

$$\begin{pmatrix} \partial_{xx}^2 a_j(x, y; \theta^{(j)}) & \partial_{xy}^2 a_j(x, y; \theta^{(j)}) \\ \partial_{yx}^2 a_j(x, y; \theta^{(j)}) & \partial_{yy}^2 a_j(x, y; \theta^{(j)}) \end{pmatrix} =: H_{a_j}(z; \theta^{(j)}). \quad (7)$$

For further use we also compute the following operator, which corresponds to the cross-term between the diffusion and drift in Itô-Taylor-expansion for each coordinate:

$$\text{Tr} \left[ B^T(z; \sigma) H_{a_j}(z; \theta^{(j)}) B(z; \sigma) \right] = b^2(z; \sigma) \partial_{yy}^2 a_j(z; \theta^{(j)}).$$

We now consider the Itô-Taylor expansion of the drift term on the interval  $t \in [i\Delta_n, (i+1)\Delta_n]$ :

$$A(z_t; \theta) \approx A(z_i; \theta) + J(z_i; \theta)(z_t - z_i) + \frac{(t - i\Delta_n)}{2} b^2(z_i; \sigma) \partial_{yy}^2 A(z_i; \theta),$$

where  $\partial_{yy}^2 A(z; \theta) := (\partial_{yy}^2 a_1(z; \theta^{(1)}), \partial_{yy}^2 a_2(z; \theta^{(2)}))^T$ .

This transformation allows us to find an approximate solution of (3). We introduce a new process  $(\tilde{Z}_t)_{t \in [i\Delta_n, (i+1)\Delta_n]}$  which is the solution of the following linear equation (see Section 5.6 in Karatzas and Shreve (1987)):

$$d\tilde{Z}_t = \left( A(\tilde{Z}_i; \theta) + J(\tilde{Z}_i; \theta)(\tilde{Z}_t - \tilde{Z}_i) + \frac{1}{2} b^2(\tilde{Z}_i; \sigma) \partial_{yy}^2 A(\tilde{Z}_i; \theta)(t - i\Delta_n) \right) dt + B(\tilde{Z}_i; \theta) dW_t.$$

The solution for the above equation is given for  $t \in [i\Delta_n, (i+1)\Delta_n]$  by

$$\begin{aligned} \tilde{Z}_t = \tilde{Z}_i + \int_{i\Delta_n}^t e^{J(\tilde{Z}_i; \theta)(t-s)} \left( A(\tilde{Z}_i; \theta) - J(\tilde{Z}_i; \theta)\tilde{Z}_i + \frac{1}{2} b^2(\tilde{Z}_i; \sigma) \partial_{yy}^2 A(\tilde{Z}_i; \theta)(s - i\Delta_n) \right) ds + \\ \int_{i\Delta_n}^t e^{J(\tilde{Z}_i; \theta)(s-i\Delta_n)} B(\tilde{Z}_i; \theta) dW_s. \end{aligned} \quad (8)$$

Note that conditionally on  $\tilde{Z}_i$ ,  $\tilde{Z}_{i+1}$  is a normal variable, whose expectation and variance are given, respectively, by:

$$\mathbb{E} \left[ \tilde{Z}_{i+1} | \tilde{Z}_i \right] = \tilde{Z}_i + \int_{i\Delta_n}^{(i+1)\Delta_n} e^{J(\tilde{Z}_i; \theta)((i+1)\Delta_n - s)} \left( A(\tilde{Z}_i; \theta) - J(\tilde{Z}_i; \theta)\tilde{Z}_i + \frac{1}{2} b^2(\tilde{Z}_i; \sigma) \partial_{yy}^2 A(\tilde{Z}_i; \theta)(s - i\Delta_n) \right) ds, \quad (9)$$

$$\Sigma \left[ \tilde{Z}_{i+1} | \tilde{Z}_i \right] = \mathbb{E} \left[ \left( \int_{i\Delta_n}^{(i+1)\Delta_n} e^{J(\tilde{Z}_i; \theta)((i+1)\Delta_n - s)} B(\tilde{Z}_i; \sigma) dW_s \right) \left( \int_{i\Delta_n}^{(i+1)\Delta_n} e^{J(\tilde{Z}_i; \theta)((i+1)\Delta_n - s)} B(\tilde{Z}_i; \sigma) dW_s \right)^T \right]. \quad (10)$$

The approximation of the solution of (3)  $(\tilde{Z}_i)_{i \geq 0}$  is then defined recursively as a sum of random variables with the mean and variance given by (9) and (10). However, these expressions are not



convenient for the numerical implementation because of the integrals and the matrix exponents. One possible solution is to rely on numerical integration algorithms when implementing the scheme. But we propose to simplify (9)-(10) in order to obtain the final scheme which is easier to implement and analyze from the theoretical point of view. We use the following propositions, whose proofs are postponed to appendix:

**Proposition 1.** *The component-wise drift approximation for (9) is given by:*

$$\mathbb{E} \left[ \tilde{Z}_{i+1} | \tilde{Z}_i \right] = \begin{pmatrix} \bar{A}_1 \\ \bar{A}_2 \end{pmatrix} + O(\Delta_n^3),$$

where  $\bar{A}_1$  and  $\bar{A}_2$  are given as follows:

$$\begin{aligned} \bar{A}_1(\tilde{Z}_i; \theta^{(1)}, \theta^{(2)}, \sigma) &:= \tilde{X}_i + \Delta_n a_1(\tilde{Z}_i; \theta^{(1)}) + \\ &\quad \frac{\Delta_n^2}{2} \left( \frac{\partial a_1(\tilde{Z}_i; \theta^{(1)})}{\partial x} a_1(\tilde{Z}_i; \theta^{(1)}) + \frac{\partial a_1(\tilde{Z}_i; \theta^{(1)})}{\partial y} a_2(\tilde{Z}_i; \theta^{(2)}) \right) + \\ &\quad \frac{\Delta_n^2}{4} b^2(\tilde{Z}_i; \sigma) \partial_{yy}^2 a_1(\tilde{Z}_i; \theta^{(1)}) \\ \bar{A}_2(\tilde{Z}_i; \theta^{(1)}, \theta^{(2)}, \sigma) &:= \tilde{Y}_i + \Delta_n a_2(\tilde{Z}_i; \theta^{(2)}) + \\ &\quad \frac{\Delta_n^2}{2} \left( \frac{\partial a_2(\tilde{Z}_i; \theta^{(2)})}{\partial x} a_1(\tilde{Z}_i; \theta^{(1)}) + \frac{\partial a_2(\tilde{Z}_i; \theta^{(2)})}{\partial y} a_2(\tilde{Z}_i; \theta^{(2)}) \right) + \\ &\quad \frac{\Delta_n^2}{4} b^2(\tilde{Z}_i; \sigma) \partial_{yy}^2 a_2(\tilde{Z}_i; \theta^{(2)}). \end{aligned} \tag{11}$$

**Proposition 2.** *The matrix  $\Sigma[\tilde{Z}_{i+1} | \tilde{Z}_i]$  defined in (10) is approximated by:*

$$b^2(\tilde{Z}_i; \sigma) \begin{pmatrix} (\partial_y a_1)^2 \frac{\Delta_n^3}{3} & (\partial_y a_1) \frac{\Delta_n^2}{2} + (\partial_y a_1)(\partial_y a_2) \frac{\Delta_n^3}{3} \\ (\partial_y a_1) \frac{\Delta_n^2}{2} + (\partial_y a_1)(\partial_y a_2) \frac{\Delta_n^3}{3} & \Delta_n + (\partial_y a_2) \frac{\Delta_n^2}{2} + (\partial_y a_2)^2 \frac{\Delta_n^3}{3} \end{pmatrix} + O(\Delta_n^4), \tag{12}$$

where the derivatives are computed at time  $i\Delta_n$ .

Both from the theoretical and computational points of view, it is enough to use only the lower-order terms of (12). Thus, we define:

$$\Sigma_{\Delta_n}(\tilde{Z}_{i+1}; \theta, \sigma^2) := b^2(\tilde{Z}_i; \sigma) \begin{pmatrix} \left( (\partial_y a_1(\tilde{Z}_i; \theta^{(1)}))^2 \frac{\Delta_n^3}{3} & \partial_y a_1(\tilde{Z}_i; \theta^{(1)}) \frac{\Delta_n^2}{2} \right) \\ \partial_y a_1(\tilde{Z}_i; \theta^{(1)}) \frac{\Delta_n^2}{2} & \Delta_n \end{pmatrix}. \tag{13}$$

The inverse of (13) is defined by:

$$\Sigma_{\Delta_n}^{-1}(\tilde{Z}_{i+1}; \theta, \sigma^2) = \frac{1}{b^2(\tilde{Z}_i; \sigma)} \begin{pmatrix} \frac{12}{(\partial_y a_1(\tilde{Z}_i; \theta^{(1)}))^2 \Delta_n^3} & -\frac{6}{\partial_y a_1(\tilde{Z}_i; \theta^{(1)}) \Delta_n^2} \\ -\frac{6}{\partial_y a_1(\tilde{Z}_i; \theta^{(1)}) \Delta_n^2} & \frac{4}{\Delta_n} \end{pmatrix}. \tag{14}$$

Finally, the element-wise approximation of  $\tilde{Z}_{i+1}$  conditionally on  $\tilde{Z}_i$  is written as:

$$\begin{aligned}\tilde{X}_{i+1} &= \bar{A}_1(\tilde{Z}_i; \theta^{(1)}, \theta^{(2)}, \sigma) + \xi_{1,i} \\ \tilde{Y}_{i+1} &= \bar{A}_2(\tilde{Z}_i; \theta^{(1)}, \theta^{(2)}, \sigma) + \xi_{2,i},\end{aligned}\tag{15}$$

where  $(\xi_{1,i})$  and  $(\xi_{2,i})$  are normal random sequences with zero means, independent in  $i$ , such that the covariance matrix of vector  $(\xi_{1,i}, \xi_{2,i})$  is given by (13). Numerically they can be simulated by decomposing the matrix (13) with the help of the LU or Cholesky decomposition, i.e. any matrix  $\bar{B}(Z_i; \theta, \sigma^2)$  such that  $\bar{B}\bar{B}^T = \Sigma(Z_i; \theta, \sigma^2)$ , and multiply it by a 2-dimensional vector whose entries are independent standard normal variables. The chosen method of the decomposition does not affect the theoretical properties of the scheme. Note that the approximated diffusion term now depends on the parameters of the drift term. It is proven that the approximated solution  $\tilde{Z}$  converges weakly to the true solution  $Z$  with order 2 (see Theorem 2 in Jimenez and Carbonell (2015)).

Now we want to study component-wise the moments of the obtained discretization, build on the observations of the process  $(Z_t)_{t \geq 0}$ . We will rely on the result of the following Proposition (recall that the true value of the vector of parameters is denoted by  $\theta_0$ ):

**Proposition 3** (Moments of the discretized process). *The following holds:*

$$\begin{aligned}\mathbb{E} \left[ X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta_0^{(2)}, \sigma_0) | Z_i \right] &= O(\Delta_n^3) \\ \mathbb{E} \left[ Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(1)}, \theta_0^{(2)}, \sigma_0) | Z_i \right] &= O(\Delta_n^3) \\ \mathbb{E} \left[ \left( X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta_0^{(2)}, \sigma_0) \right)^2 | Z_i \right] &= (\partial_y a_1)_{\theta_0^{(1)}}^2 \frac{\Delta_n^3}{3} b^2(Z_i; \sigma_0) + O(\Delta_n^4) \\ \mathbb{E} \left[ \left( Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(1)}, \theta_0^{(2)}, \sigma_0) \right)^2 | Z_i \right] &= \Delta_n b^2(Z_i; \sigma_0) + O(\Delta_n^2) \\ \mathbb{E} \left[ \left( X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta_0^{(2)}, \sigma_0) \right) \left( Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(1)}, \theta_0^{(2)}, \sigma_0) \right) | Z_i \right] &= (\partial_y a_1)_{\theta_0^{(1)}} \frac{\Delta_n^2}{2} b^2(Z_i; \sigma_0) + O(\Delta_n^3),\end{aligned}$$

where  $\mathbb{E}$  is taken under  $\mathbb{P}_0$  and the derivatives  $\partial_y a_1$  are computed at time  $i\Delta_n$ .

*Proof.* The moments of the Feller process can be approximated by its generator (Kloeden et al., 2003). That is, for a sufficiently smooth and integrable function  $f : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ :

$$\mathbb{E}(f(Z_{t+\Delta_n}) | Z_t = z) = \sum_{i=0}^j \frac{\Delta_n^i}{i!} L^i f(z) + O(\Delta_n^{j+1}),\tag{16}$$

where  $L^i f(z)$  is the  $i$  times iterated generator of model (3) given by

$$L f(z) = (\partial_z f(z)) A(z) + \frac{1}{2} \nabla_B^2 f(z),$$

where  $\nabla_B^2(\cdot) = b^2(z; \sigma) \frac{\partial^2}{\partial y^2}(\cdot)$  is a weighted Laplace type operator. Since the process is approximated by (11), it coincides with (16) up to the terms of order  $\Delta_n^2$ .  $\square$

Further, we need an extension of Proposition 3 which gives the order of moments of the increments of the discrete process when parameters are fixed to their true values only partly. By doing that, we loose one order of accuracy in first two bounds, but the results for the variance remain unchanged. Note however that we cannot obtain the last three terms unless  $\theta^{(1)} = \theta_0^{(1)}$ . This is the main technical challenge to overcome when constructing an estimator.

**Proposition 4.** *The following holds:*

- (i)  $\mathbb{E} \left[ X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma) | Z_i \right] = O(\Delta_n^2)$
- (ii)  $\mathbb{E} \left[ Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(1)}, \theta_0^{(2)}, \sigma) | Z_i \right] = O(\Delta_n^2)$
- (iii)  $\mathbb{E} \left[ \left( X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma) \right)^2 | Z_i \right] = (\partial_y a_1)_{\theta_0^{(1)}}^2 \frac{\Delta_n^3}{3} b^2(Z_i; \sigma) + O(\Delta_n^4)$
- (iv)  $\mathbb{E} \left[ (Y_{i+1} - Y_i)^2 | Z_i \right] = \Delta_n b^2(Z_i; \sigma) + O(\Delta_n^2)$
- (v)  $\mathbb{E} \left[ \left( X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma) \right) (Y_{i+1} - Y_i) | Z_i \right] = (\partial_y a_1)_{\theta_0^{(1)}} \frac{\Delta_n^2}{2} b^2(Z_i; \sigma) + O(\Delta_n^3),$

where  $\mathbb{E}$  is taken under  $\mathbb{P}_0$  and the derivatives  $\partial_y a_1$  are computed at time  $i\Delta_n$ .

*Proof.* We show the result for (i) and (iii). Start with (i):

$$X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma) = X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta_0^{(2)}, \sigma_0) + \bar{A}_1(Z_i; \theta_0^{(1)}, \theta_0^{(2)}, \sigma_0) - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma).$$

The difference  $\mathbb{E} \left[ X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta_0^{(2)}, \sigma_0) | Z_i \right] = O(\Delta_n^3)$  by Proposition 3 and the assumption (A2). It remains to consider the second part:

$$\begin{aligned} & \bar{A}_1(Z_i; \theta_0^{(1)}, \theta_0^{(2)}, \sigma_0) - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma) = \\ & \frac{\Delta_n^2}{2} \left( \frac{\partial a_1(Z_i; \theta_0^{(1)})}{\partial y} \left( a_2(Z_i; \theta_0^{(2)}) - a_2(Z_i; \theta^{(2)}) \right) + \frac{\partial_{yy}^2 a_1(Z_i; \theta_0^{(1)})}{2} (b^2(Z_i; \sigma_0) - b^2(Z_i; \sigma)) \right). \end{aligned}$$

Thus,  $\mathbb{E} \left[ X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta_0^{(2)}, \sigma_0) | Z_i \right] = O(\Delta_n^2)$ . Let us now consider (iii):

$$\begin{aligned} \left( X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma) \right)^2 &= \left( X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta_0^{(2)}, \sigma_0) \right)^2 + \\ & \quad \left( \bar{A}_1(Z_i; \theta_0^{(1)}, \theta_0^{(2)}, \sigma_0) - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma) \right)^2 + \\ & \quad 2 \left( X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta_0^{(2)}, \sigma_0) \right) \left( \bar{A}_1(Z_i; \theta_0^{(1)}, \theta_0^{(2)}, \sigma_0) - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma) \right). \end{aligned}$$

Again, by Proposition 3 and previous computations we have the following:

$$\begin{aligned} \mathbb{E} \left[ \left( X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta_0^{(2)}, \sigma_0) \right)^2 \middle| Z_i \right] &= (\partial_y a_1)_{\theta_0}^2 \frac{\Delta_n^3}{3} b^2(Z_i; \sigma) + O(\Delta_n^4) \\ \mathbb{E} \left[ \left( \bar{A}_1(Z_i; \theta_0^{(1)}, \theta_0^{(2)}, \sigma_0) - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma) \right)^2 \middle| Z_i \right] &= O(\Delta_n^4) \\ \mathbb{E} \left[ 2 \left( X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta_0^{(2)}, \sigma_0) \right) \left( \bar{A}_1(Z_i; \theta_0^{(1)}, \theta_0^{(2)}, \sigma_0) - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma) \right) \middle| Z_i \right] &= O(\Delta_n^5). \end{aligned}$$

The rest of proofs follows the same pattern.  $\square$

## 4 Parameter estimation

In this section we propose a contrast estimator based on the pseudo-likelihood function and prove its consistency and asymptotic normality. Then we discuss other already known results for the linear homogeneous SDEs (least squares estimator in particular) and show how it works in the general case.

### 4.1 Contrast estimator

Let us introduce the so-called contrast function for the system (3). This function is defined as  $-2$  times the log-likelihood of the discretized model (Florens-Zmirou, 1989, Kessler, 1997):

$$\begin{aligned} \mathcal{L}_{n, \Delta_n}(\theta, \sigma^2; Z_{0:n}) &= \frac{1}{2} \sum_{i=0}^{n-1} (Z_{i+1} - \bar{A}(Z_i; \theta))^T \Sigma_{\Delta_n}^{-1}(Z_i; \theta, \sigma^2) (Z_{i+1} - \bar{A}(Z_i; \theta)) \\ &\quad + \sum_{i=0}^{n-1} \log \det(\Sigma_{\Delta_n}(Z_i; \theta, \sigma^2)), \end{aligned} \quad (17)$$

where the inverse matrix  $\Sigma_{\Delta_n}^{-1}$  is given by (14). Then the local linearization (LL) estimator is defined as:

$$(\hat{\theta}_{n, \Delta_n}, \hat{\sigma}_{n, \Delta_n}^2) = \arg \min_{\theta, \sigma^2} \mathcal{L}_{n, \Delta_n}(\theta, \sigma^2; Z_{0:n}), \quad (18)$$

where  $\hat{\theta}_{n, \Delta_n} = (\hat{\theta}_{n, \Delta_n}^{(1)}, \hat{\theta}_{n, \Delta_n}^{(2)})$ .

Before proceeding to the proofs, let us explain how the contrast estimator works in the classical elliptic setting and give a roadmap for the proofs of consistency and asymptotic normality of the estimator (18), following Kessler (1997). The first notable difference between the estimator (18) and the elliptic case is that in the elliptic case the estimation of the drift and the variance parameters can be separated. For example, the contrast estimator for a 1-dimensional SDE discretized with the Euler-Maruyama scheme is defined as follows:

$$\left( \hat{\theta}, \hat{\sigma}^2 \right) = \arg \min_{\theta, \sigma^2} \sum_{i=1}^n \left( \frac{1}{2} \frac{(X_{i+1} - X_i - \Delta_n a(X_i; \theta))^2}{\Delta_n b^2(X_i; \sigma)} + \log b^2(X_i; \sigma) \right), \quad (19)$$

where  $a(x; \theta)$  is a drift term. Here the estimation of the parameter  $\theta$  is independent of the value of  $\sigma$ , because the minimization of the criteria boils down to minimizing the expression  $(X_{i+1} - X_i - \Delta_n a(X_i; \theta))^2$  and  $\hat{\theta}$  converges to  $\theta_0$  with a rate  $\sqrt{n\Delta_n}$ . For the variance term, the estimator of  $\sigma$  converges independently of the value of  $\theta$ , because  $(X_{i+1} - X_i - \Delta_n a(X_i; \theta))^2$  is of order  $\Delta_n$  for any  $\theta$  and it is enough to ensure the convergence of the variance parameter. The convergence rate for the variance is  $\sqrt{n}$ . This property is also shared by the estimator for the Hamiltonian SDE proposed by Samson and Thieullen (2012).

In a general hypoelliptic setting the parametric inference is more complicated. First, the drift parameter  $\theta$  is contained in the covariance matrix  $\Sigma_{\Delta_n}$ . Second, the variance of the first variable is of order  $\Delta_n^3$ , while for an arbitrary chosen vector of parameters  $\theta^{(1)}$  the expression  $(X_{i+1} - \bar{A}_1(Z_i; \theta^{(1)}, \theta_0^{(2)}, \sigma))^2$  is of order  $\Delta_n^2$ . It is not enough to show the convergence of the diffusion parameter in a standard way. From the practical point of view it means that if we launch the minimization algorithm on (18) only with respect to  $\theta^{(2)}$  and  $\sigma^2$ , it will not converge to the true value. The inverse, however, is possible: using the Proposition 4 the consistency result for  $\hat{\theta}^{(1)}$  can be obtained without fixing  $\theta^{(2)}$  and  $\sigma$ .

Ditlevsen and Samson (2017) propose to overcome the problem of dependency between the estimators by separating the estimation of the rough and smooth variables. They introduce two separate contrasts, based on the approximate marginal distribution on each variable:

$$\hat{\theta}^{(1)} = \arg \min_{\theta^{(1)}} \sum_{i=1}^{n-1} \left( \frac{3 \left( X_{i+1} - \bar{A}_1(Z_i; \theta^{(1)}, \theta_0^{(2)}, \sigma_0) \right)^2}{2 \Delta_n^3 \left( \partial_y a_1(Z_i; \theta^{(1)}) \right)^2 b^2(Z_i; \sigma_0)} + \log \left( \partial_y a_1(Z_i; \theta^{(1)}) b(Z_i; \sigma_0) \right)^2 \right), \quad (20)$$

$$\left( \hat{\theta}^{(2)}, \hat{\sigma}^2 \right) = \arg \min_{\theta^{(2)}, \sigma^2} \sum_{i=1}^{n-1} \frac{1}{2} \left( \frac{\left( Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma) \right)^2}{\Delta_n b^2(Z_i; \sigma)} + \log b^2(Z_i; \sigma) \right). \quad (21)$$

The estimation is then conducted as follows: first, the parameters of the first equation are estimated from (20). Estimator (20) is shown to converge with rate  $\sqrt{\frac{n}{\Delta_n}}$ . Since the parameters of the second equation are contained only in higher order terms, they are shown to have no impact on the convergence of the estimator. We are able to show that thanks to the Proposition 4. Then, the obtained value  $\hat{\theta}^{(1)}$  is plugged in (21). The contrast is minimized with respect to  $\sigma$  and  $\theta^{(2)}$ . It is proven that the estimator  $\hat{\theta}^{(2)}$  converges with the rate  $\sqrt{n\Delta_n}$  and  $\hat{\sigma}^2$  — with a rate  $\sqrt{n}$ . The rates are identical to the rates of convergence obtained in Kessler (1997) for elliptic systems. The weak point of the scheme is that in order to prove the convergence of the estimator (21) the value of  $\theta^{(1)}$  needs to be fixed to  $\theta_0^{(1)}$ .

We choose a different approach and focus on the 2-dimensional contrast without splitting the numerical procedure in two parts. We still need to take into account the different rates of convergence and the eventual dependencies between the parameters. Thus, the proof of the consistency and asymptotic normality is splitted in two principal steps. The first step is a proof

of the consistency and the convergence rate for  $\hat{\theta}_{n,\Delta_n}^{(1)}$  (Theorem 1). Except for the unusual convergence rate, the proof repeats the standard techniques from Kessler (1997) and Ditlevsen and Samson (2017), adapted to the unknown value of  $\theta^{(2)}$ . The second step, however, is more intricate. As in Ditlevsen and Samson (2017), the estimators for  $\theta^{(2)}$  and  $\sigma^2$  do not converge for an arbitrary  $\theta^{(1)}$ . However, we prove that the consistency and the asymptotic normality still hold for  $\hat{\theta}_{n,\Delta_n}^{(2)}$  and  $\hat{\sigma}_{n,\Delta_n}^2$ , because the sequence of estimators  $\hat{\theta}_{n,\Delta_n}^{(1)}$  is tight and converges with rates proven in Theorem 1. It is proven at the cost of an additional assumption (A5) on the function  $a_1(Z_t; \theta^{(1)})$  in the drift term.

We begin the study from the following Lemma, on which the consistency of  $\hat{\theta}^{(1)}$  crucially relies:

**Lemma 1.** *Under the assumptions (A1)-(A4), and assuming  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$ , the following holds:*

$$\lim_{n \rightarrow \infty, \Delta_n \rightarrow 0} \frac{\Delta_n}{n} \left[ \mathcal{L}_{n,\Delta_n}(\theta^{(1)}, \theta^{(2)}, \sigma^2; Z_{0:n}) - \mathcal{L}_{n,\Delta_n}(\theta_0^{(1)}, \theta^{(2)}, \sigma^2; Z_{0:n}) \right] \xrightarrow{\mathbb{P}_0} 6 \int \frac{(a_1(z; \theta_0^{(1)}) - a_1(z; \theta^{(1)}))^2}{b^2(z; \sigma)(\partial_y a_1)_\theta^2} \nu_0(dz).$$

Proof is postponed to Appendix. On the next step we obtain the consistency and the asymptotic normality of (18) with respect to  $\theta^{(1)}$ :

**Theorem 1.** *Under the assumptions (A1)-(A5), and assuming  $\Delta_n \rightarrow 0$ ,  $n\Delta_n \rightarrow \infty$  and  $n\Delta_n^2 \rightarrow 0$ , the following holds:*

$$\hat{\theta}_{n,\Delta_n}^{(1)} \xrightarrow{\mathbb{P}_0} \theta_0^{(1)},$$

$$\sqrt{\frac{n}{\Delta_n}} (\hat{\theta}_{n,\Delta_n}^{(1)} - \theta_0^{(1)}) \xrightarrow{\mathcal{D}} \mathcal{N} \left( 0, 3 \left( \int \frac{(\partial_{\theta^{(1)}} a_1)(\partial_{\theta^{(1)}} a_1)^T}{b^2(z; \sigma)(\partial_y a_1)} \nu_0(dz) \right)^{-1} \left( \int \frac{b^2(z; \sigma_0)}{b^4(z; \sigma)} (\partial_{\theta^{(1)}} a_1)(\partial_{\theta^{(1)}} a_1)^T \left( 1 + \frac{1}{(\partial_y a_1)^2} \right) \nu_0(dz) \right) \right),$$

where  $\partial_x a_1$  is a simplified notation for  $\partial_x a_1(z; \theta_0^{(1)})$

The asymptotic variance of the estimator slightly differs from the one obtained in Ditlevsen and Samson (2017). It is because the 2-dimensional estimator contains the cross-terms of type  $(X_{i+1} - \bar{A}_1(Z_i; \theta^{(1)}, \theta^{(2)}, \sigma))(Y_{i+1} - \bar{A}_2(Z_i; \theta^{(1)}, \theta^{(2)}, \sigma))$ , not taken into account if the estimator is splitted in two separate contrasts for rough and smooth variables. The speed of convergence, however, stays the same. Notice also that the assumption (A5) is not used for Lemma 1, on which proof of consistency relies. However, it is needed for the asymptotic normality. However, we do not need  $\theta^{(2)}$  and  $\sigma^2$  to be known, on the contrary to Ditlevsen and Samson (2017).

The idea of the proof of consistency for the diffusion and the rough term parameters follows Gloter and Sørensen (2009). Since we are working in a compact set, we can always find a sequence of estimators  $\hat{\theta}_{n,\Delta_n}^{(1)}$  such that the sequence  $(\hat{\theta}_{n,\Delta_n}^{(1)} - \theta_0^{(1)})$  is tight. Then we use the tightness in combination with the rate of convergence obtained in Theorem 1 and the continuous mapping theorem for proving the consistency of the remaining terms in a standard way. On this stage we need an additional assumption (A5). The reason for that is when the parameter  $\theta^{(1)}$  is included in the derivative  $\partial_y a_1$ , this parameter is present both in the drift and the variance term, which substantially complicates the study. Also, assuming the linear shape of  $a_1$  with respect to  $\theta^{(1)}$ , one can fully use the speed of convergence for  $\hat{\theta}_{n,\Delta_n}^{(1)}$  obtained in Theorem 1. It is rather restrictive, but the idea of the proof can be reused for a more general case. For example, consistency for the parameters of the rough variable can be obtained under the condition of Lipschitz continuity with respect to parameter  $\theta^{(1)}$ , at the cost of additional technicalities, which are omitted in this paper. The consistency follows from the following Lemmas, on which Theorem 2 is based (proofs of both Lemmas and the Theorem are postponed to Appendix):

**Lemma 2.** *Under assumptions (A1)-(A5), and assuming  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$ , the following holds:*

$$\lim_{n \rightarrow \infty, \Delta_n \rightarrow 0} \frac{1}{n\Delta_n} \left[ \mathcal{L}_{n,\Delta_n}(\hat{\theta}_{n,\Delta_n}^{(1)}, \theta^{(2)}, \sigma^2; Z_{0:n}) - \mathcal{L}_{n,\Delta_n}(\hat{\theta}_{n,\Delta_n}^{(1)}, \theta_0^{(2)}, \sigma^2; Z_{0:n}) \right] \xrightarrow{\mathbb{P}_0} 2 \int \frac{(a_2(z; \theta^{(2)}) - a_2(z; \theta_0^{(2)}))^2}{b^2(z; \sigma)} \nu_0(dz)$$

**Lemma 3.** *Under assumptions (A1)-(A5), and assuming  $\Delta_n \rightarrow 0$  while  $n\Delta_n \rightarrow \infty$ , the following holds:*

$$\frac{1}{n} \mathcal{L}_{n,\Delta_n}(\hat{\theta}_{n,\Delta_n}^{(1)}, \theta^{(2)}, \sigma^2; Z_{0:n}) \xrightarrow{\mathbb{P}_0} \int \left( \frac{b^2(z; \sigma_0)}{b^2(z; \sigma)} + \log b^2(z; \sigma) \right) \nu_0(dz)$$

**Theorem 2.** *Under assumptions (A1)-(A5), and assuming  $\Delta_n \rightarrow 0$ ,  $n\Delta_n \rightarrow \infty$  and  $n\Delta_n^2 \rightarrow 0$  the following holds:*

$$\hat{\theta}_{n,\Delta_n}^{(2)} \xrightarrow{\mathbb{P}_0} \theta_0^{(2)}, \quad \hat{\sigma}_{n,\Delta_n} \xrightarrow{\mathbb{P}_0} \sigma_0$$

and

$$\begin{aligned} \sqrt{n\Delta_n}(\hat{\theta}_{n,\Delta_n}^{(2)} - \theta_0^{(2)}) &\xrightarrow{\mathcal{D}} \mathcal{N} \left( 0, \left( \int \frac{(\partial_{\theta^{(2)}} a_2(z; \theta_0^{(2)}))(\partial_{\theta^{(2)}} a_2(z; \theta_0^{(2)}))^T}{b^2(z, \sigma)} \nu_0(dz) \right)^{-1} \right) \\ \sqrt{n}(\hat{\sigma}_{n,\Delta_n} - \sigma_0) &\xrightarrow{\mathcal{D}} \mathcal{N} \left( 0, 2 \left( \int \frac{(\partial_{\sigma} b(z, \sigma_0))(\partial_{\sigma} b(z, \sigma_0))^T}{b^2(z, \sigma_0)} \nu_0(dz) \right)^{-1} \right). \end{aligned}$$

The obtained rates coincide with the rates in Ditlevsen and Samson (2017), but with the advantage that we avoid fixing any of the parameters to their true value, instead we work with the estimated sequence  $\hat{\theta}_{n,\Delta_n}^{(1)}$ .

## 4.2 Conditional least squares estimator

For certain applications it is natural to split the estimation of the parameters in the diffusion coefficient and the drift term (see, for example, Le-Breton and Musiela (1985)). First, it reduces the dimension of the optimization problem, and thus spares the computational cost. Second, it is easier to generalize the drift-based least square estimator to the high-dimensional hypoelliptic systems, when the approximation of the diffusion matrix is difficult to compute. The idea is to compute the least square estimator of the differences between the discrete observations of  $(Z_t)_{t \geq 0}$  and the expectation of this process computed with the LL scheme. For system (3) however we should still be careful about the order of each difference. In order for the estimator to converge properly we need to renormalize the expression. We do that as follows:

$$\hat{\theta}_{n, \Delta_n}^{LSE} := \left( \hat{\theta}^{(1), LSE}, \hat{\theta}^{(2), LSE} \right)^T \quad (22)$$

where

$$\begin{aligned} \hat{\theta}_{n, \Delta_n}^{LSE, (1)} &:= \arg \min_{\theta^{(1)}} \sum_{i=0}^{n-1} \frac{(X_{i+1} - \bar{A}_1(Z_i; \theta^{(1)}, \theta^{(2)}, \sigma))^2}{\Delta_n^3} \\ \hat{\theta}_{n, \Delta_n}^{LSE, (2)} &:= \arg \min_{\theta^{(2)}} \sum_{i=0}^{n-1} \frac{(Y_{i+1} - \bar{A}_2(Z_i; \theta^{(1)}, \theta^{(2)}, \sigma))^2}{\Delta_n}, \end{aligned}$$

where  $\bar{A}_j(Z_i; \theta^{(1)}, \theta^{(2)}, \sigma), j = 1, 2$  are defined in (11). Using the same reasoning as for the LL contrast we prove the next Theorem (the proof is postponed to appendix):

**Theorem 3.** *Under the assumptions (A1)-(A4) and the conditions  $\Delta_n \rightarrow 0$ ,  $n\Delta \rightarrow \infty$  and  $n\Delta_n^2 \rightarrow 0$  the following holds:*

$$\hat{\theta}_{n, \Delta_n}^{LSE} \xrightarrow{\mathbb{P}_0} \theta_0,$$

and

$$\begin{pmatrix} \sqrt{\frac{n}{\Delta_n}} (\hat{\theta}_{n, \Delta_n}^{LSE, (1)} - \theta_0) \\ \sqrt{n\Delta_n} (\hat{\theta}_{n, \Delta_n}^{LSE, (2)} - \theta_0) \end{pmatrix} \xrightarrow{\mathcal{D}} 2\mathcal{N} \left( 0, I_2 \cdot \begin{bmatrix} \frac{1}{3} \int b^2(z; \sigma_0) (\partial_y a_1(z; \theta_0^{(1)}))^2 (\partial_{\theta^{(1)}} a_1(z; \theta_0^{(1)}))^2 \nu_0(dz) \\ \int b^2(z; \sigma_0) (\partial_{\theta^{(2)}} a_2(z; \theta_0^{(2)}))^2 \nu_0(dz) \end{bmatrix} \right),$$

where  $I_2$  is a  $2 \times 2$  identity matrix,  $\hat{\theta}_{n, \Delta_n}^{LSE, (j)}$  denote the  $j$ -th element of the vector  $\mathcal{L}_{n, \Delta_n}^{LSE}(\theta; Z_{0:n})$  and

$$C_i = \int (\partial_{\theta^{(i)}} a_i(z; \theta_0^{(i)})) (\partial_{\theta^{(i)}} a_i(z; \theta_0^{(i)}))^T \nu_0(dz)$$

The advantage of this estimator over the LL contrast is that due to the absence of the cross-terms, the estimation of both parameters is independent. For instance, in Theorem 3 we prove the consistency of the estimator with respect to  $\theta^{(2)}$  without assumption (A5) and fixing  $\theta^{(1)}$  to



the estimated sequence  $\hat{\theta}_{n,\Delta_n}^{(1)}$ . Also, since the term  $(\partial_y a_1(z; \theta_0^{(1)}))$  is not present in the variance, we do not need (A5) to obtain the asymptotic normality for the estimator of  $\theta^{(1)}$ . The asymptotic variance differs from that obtained in Theorems 1-2. Since the terms  $\partial_y a_1(z; \theta_0^{(1)})$  and  $b(x, y; \sigma)$  were not included in the normalization, they appear in the covariance matrix and influence the performance of the estimator. Thus, in comparison to LL estimator defined by (18), conditional least square estimator may perform worse and be prone to outliers when the diffusion coefficient and the value of  $(\partial_y a_1(z; \theta_0^{(1)}))^2$  are large.

Note that the result of Theorem 3 holds for any  $\sigma$ . However, the diffusion coefficient cannot be estimated from criteria (22). One possible way to estimate it would be to plug in the obtained drift parameters in the 2-dimensional criteria (17) or in the 1-dimensional criteria from Ditlevsen and Samson (2017), given by (21). Analogously, when the noise in SDE (1) is additive (i.e.,  $b \equiv \text{const}$ ), or in a special case when  $b(x, y; \sigma) \equiv \sigma f(x, y)$ , the parameter  $\sigma$  can be estimated explicitly with the help of the sample covariance matrix. The properties of this approach for the elliptic case are proven in Kessler (1997), Jacod and Protter (2011). For hypoelliptic systems, this approach must be modified, as the discretization of order  $\Delta_n$  does not allow to compute the terms of order  $\Delta_n^3$ , which represent the propagated noise. However, the value of  $\sigma$  can still be inferred from the observations of the rough coordinate by computing

$$\hat{\sigma}_{n,\Delta_n}^2 = \frac{1}{n\Delta_n} \sum_{i=0}^{n-1} \frac{(Y_{i+1} - Y_i)^2}{f^2(X_i, Y_i)}. \quad (23)$$

It can be shown that this estimator is consistent and asymptotically normal. In fact, it is a straightforward consequence of point (iv) of Lemma 7 (see Appendix), but we do not aim to provide the details here as it only concerns the particular case of model (1), that is, when the diffusion term depends linearly on only one unknown parameter. However, we test the performance of the estimator (23) in Section 5, devoted to the numerical experiments.

## 5 Simulation study

### 5.1 The model

The two estimators  $(\hat{\theta}_{n,\Delta_n}, \hat{\sigma}_{n,\Delta_n}^2)$  and  $(\hat{\theta}_{n,\Delta_n}^{LSE}, \hat{\sigma}_{n,\Delta_n}^2)$  are evaluated on the simulation study with a hypoelliptic stochastic neuronal model called FitzHugh-Nagumo model (Fitzhugh, 1961). It is a simplified version of the Hodgkin-Huxley model (Hodgkin and Huxley, 1952), which describes in a detailed manner activation and deactivation dynamics of a spiking neuron. First it was studied in the deterministic case, then in the stochastic elliptic setting with two sources of noise in both coordinates. However, it is often argued that only ion channels are perturbed by noise, while the membrane potential depends on them in a deterministic way. This idea leads to a 2-dimensional hypoelliptic diffusion. In this paper we consider a hypoelliptic SDE with noise only in the second coordinate as studied in Leon and Samson (2018). More precisely, the behaviour

of the neuron is defined through the solution of the system

$$\begin{cases} dX_t = \frac{1}{\varepsilon}(X_t - X_t^3 - Y_t - s)dt \\ dY_t = (\gamma X_t - Y_t + \beta)dt + \sigma dW_t, \end{cases} \quad (24)$$

where the variable  $X_t$  represents the membrane potential of the neuron at time  $t$ , and  $Y_t$  is a recovery variable, which could represent the channel kinetic. The parameter  $s$  is the magnitude of the stimulus current and is often known in experiments,  $\varepsilon$  is a time scale parameter and is typically significantly smaller than 1, since  $X_t$  moves "faster" than  $Y_t$ . Parameters to be estimated are  $\theta = (\gamma, \beta, \varepsilon, \sigma)$ . For system (24) we obtain the following expressions for  $\bar{A}$  and  $\Sigma_{\Delta_n}$ , which we plug in (17):

$$\bar{A}(Z_i; \theta) = \begin{pmatrix} X_i + \frac{\Delta_n}{\varepsilon}(X_i - X_i^3 - Y_i + s) + \frac{\Delta_n^2}{2\varepsilon} \left( \frac{(1-3X_i^2)}{\varepsilon}(X_i - X_i^3 - Y_i + s) - (\gamma X_i - Y_i + \beta) \right) \\ Y_i + \Delta_n(\gamma X_i - Y_i + \beta) + \frac{\Delta_n^2}{2} \left( \frac{\gamma}{\varepsilon}(X_i - X_i^3 - Y_i + s) - (\gamma X_i - Y_i + \beta) \right) \end{pmatrix}$$

$$\Sigma_{\Delta_n}(Z_i; \theta, \sigma) = \sigma^2 \begin{pmatrix} \frac{\Delta_n^3}{3\varepsilon^2} & \frac{\Delta_n^2}{2\varepsilon} \\ \frac{\Delta_n^2}{2\varepsilon} & \Delta_n \end{pmatrix}$$

Hypoellipticity and ergodicity of (24) are proven in Leon and Samson (2018). The same problem, but for the hypoelliptic setting is studied in Jensen (2014), Ditlevsen and Samson (2017).

## 5.2 Experimental design

We consider two different settings: an excitatory and an oscillatory behaviour. For the first regime, the drift parameters are set to  $\gamma = 1.5$ ,  $\beta = 0.3$ ,  $\varepsilon = 0.1$ ,  $s = 0.01$  and the diffusion coefficient  $\sigma = 0.6$ , and for the second  $\gamma = 1.2$ ,  $\beta = 1.3$ ,  $\varepsilon = 0.1$ ,  $s = 0.01$  and  $\sigma = 0.4$ . The diffusion coefficient does not change the behaviour pattern, only the "noisiness" of the observations. The starting point is  $(X_0, Y_0) = (0, 0)$ . Sample trajectories for both settings are shown on Figure 1.

We organize the trials as follows: first, we generate 100 trajectories using recursive formula (15) for each set of parameters with  $\Delta_n = 0.0001$  and  $n = 500000$ . The observed time interval is thus equal to 50. Then we subsample the sequence so that we can vary the discretization step  $\Delta_n$  and eventually truncate the observed time interval. We estimate the parameters by minimizing the contrast (17). We refer to this method as LL contrast. For the least square estimator (LSE) we do the following: we estimate the parameter  $\sigma$  explicitly from the observations of the second variable by (23), and then compute the parameters of the drift by minimizing (22). In addition, we compare both methods to the 1.5 strong order scheme (Ditlevsen and Samson, 2017), based on two separate estimators for each coordinate, which are defined in (20) and (21).

The minimization of the criterions is conducted with the `optim` function in **R** with the Conjugate Gradient method. As the initial value of parameters we take  $\theta_0 \pm U([0, 1])$ , where  $U$  stays for the uniform probabilistic law. In Tables 1-2 we present the mean value of the estimated

parameters and their standard deviation (in brackets), computed over 100 trajectories for each set of parameters. The reported value of  $\sigma$  is obtained as  $\sqrt{\sigma^2}$ , since only  $\sigma^2$  is identifiable. Figures 2-3 illustrate the estimation densities for  $\Delta_n = 0.01$  and the interval of observations being fixed to  $T = 5$  or  $T = 50$ . The LL contrast is depicted in blue, the least square estimator — in red, the 1.5 scheme in green.

The estimation of the diffusion coefficient  $\sigma$  with the LL estimator is slightly biased in both sets of data. This bias does not appear in the one-dimensional criteria and when the value is directly computed from the observations as a mean empirical variance. The performance of the LL contrast improves when we reduce the step size and increase the observed time interval. However, when  $\Delta_n$  becomes too small the performance of LL contrast with respect to  $\sigma$  is worse than the one-dimensional estimators for  $\sigma$  given by (21) and (23). It is slightly biased and its variance is bigger than that of LSE and 1.5 estimator. One possible explanation is that the estimation of  $\sigma$  with the LL contrast, as it is shown in Theorem 2, depends heavily on the convergence of the parameters of the first coordinate. Minor inaccuracies in the estimation of the drift parameters lead to non-negligible errors in  $\hat{\sigma}$ . Note, for example, that the LL scheme scores better on interval  $T = 50$  for  $\Delta_n = 0.01$  than for  $\Delta_n = 0.001$  (see Table 1), while for the other schemes it is not the case. Thus, it is important to ensure that  $n \rightarrow \infty$  faster than  $\Delta_n \rightarrow 0$ , as required by Theorem 2.

Parameters of the second coordinate  $\gamma$  and  $\beta$  are estimated accurately with all three methods once the time interval  $T$  is big enough (see the bottom pictures on Figures 2-3 for  $T = 50$ ). However, when  $T = 5$ , 1.5 scheme scores considerably worse than the LL and LSE estimator. Also when estimating  $\varepsilon$ , the 1-dimensional criteria (20) does not score better than the LL and LSE estimators. This parameter seems to be underestimated in the case of the 1.5 scheme, and this bias is bigger in the case of the inhibitory setting for  $\Delta_n = 0.01$ . The problems in the inhibitory setting are anticipated, since the trajectory is more erratic than in the excitatory case. Drift parameters are thus more difficult to estimate: the variance of the estimators is bigger in average. Also, during the simulation study it is observed that  $\varepsilon$  is the most sensitive to the initial value with which the `optim` function is initialized, since it directly regulates the amount of noise which is propagated to the first coordinate. However, as predicted by Theorems 1-2, estimators for  $\varepsilon$  converge indeed faster than for the rest of the parameters.

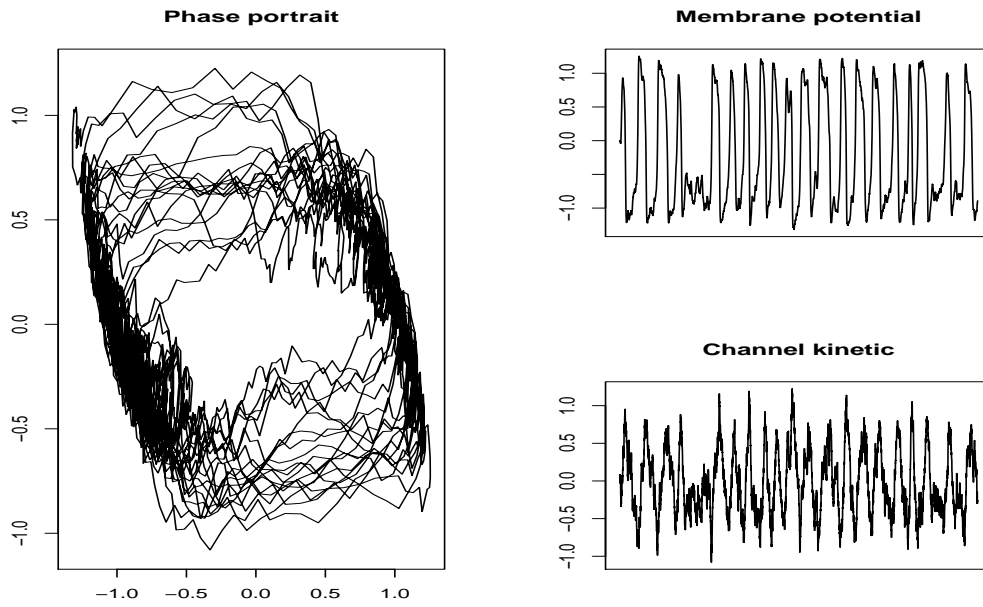
## 6 Conclusions

The proposed contrast estimator generalizes parametric inference methods developed for models of type (2) to more general class (1). Numerical study shows that it can be used with no prior knowledge of the parameters. It is the main advantage of our method over the analogous works, in particular Ditlevsen and Samson (2017), where the convergence of the estimator is proven with the parameters being partly fixed to their true values.

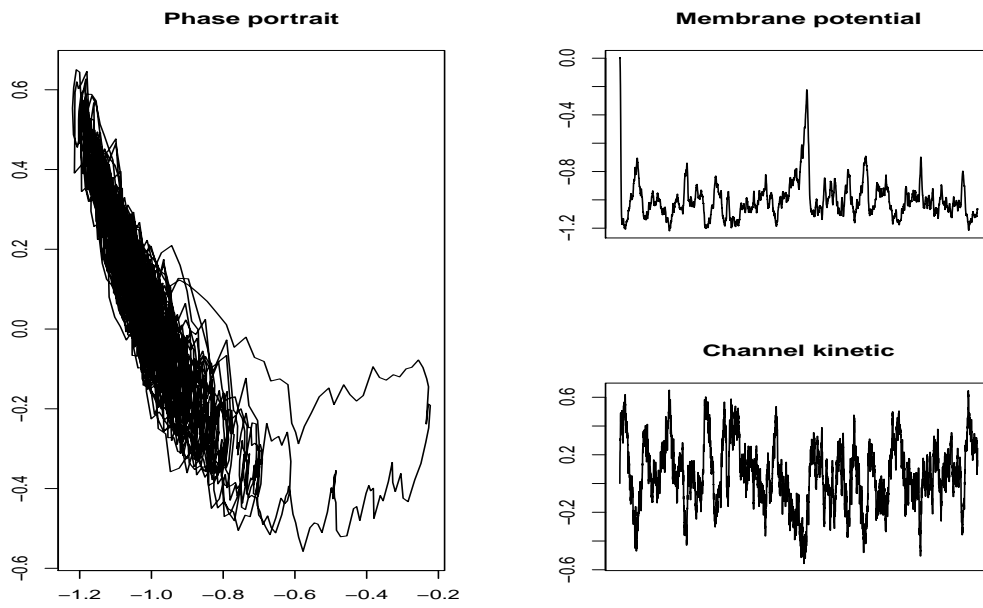
From the theoretical point of view, our estimators reveal good properties. Both the contrast based on the local linearization scheme and the least square estimators are consistent. In the case

$\Delta_n = 0.01, T = 5$	$\gamma$	$\beta$	$\varepsilon$	$\sigma$
LC	1.501 (0.053)	0.302 (0.055)	0.101 (0.001)	0.592 (0.056)
LSE	1.488 (0.108)	0.311 (0.149)	0.100 (0.000)	0.612 (0.020)
1.5 scheme	1.561 (0.362)	0.324 (0.295)	0.099 (0.000)	0.598 (0.019)
$\Delta_n = 0.01, T = 10$	$\gamma$	$\beta$	$\varepsilon$	$\sigma$
LC	1.504 (0.055)	0.306 (0.053)	0.100 (0.001)	0.562 (0.026)
LSE	1.503 (0.069)	0.299 (0.176)	0.100 (0.000)	0.610 (0.014)
1.5 scheme	1.540 (0.237)	0.301 (0.212)	0.099 (0.000)	0.596 (0.013)
$\Delta_n = 0.01, T = 50$	$\gamma$	$\beta$	$\varepsilon$	$\sigma$
LC	1.500 (0.050)	0.297 (0.052)	0.100 (0.000)	0.560 (0.018)
LSE	1.513 (0.072)	0.302 (0.068)	0.100 (0.000)	0.610 (0.007)
1.5 scheme	1.495 (0.095)	0.301 (0.093)	0.099 (0.000)	0.596 (0.007)
$\Delta_n = 0.001, T = 5$	$\gamma$	$\beta$	$\varepsilon$	$\sigma$
LC	1.505 (0.054)	0.306 (0.051)	0.100 (0.000)	0.699 (0.090)
LSE	1.498 (0.062)	0.290 (0.072)	-47.86 (477.2)	0.599 (0.005)
1.5 scheme	1.497 (0.183)	0.304 (0.169)	0.100 (0.000)	0.598 (0.005)
$\Delta_n = 0.001, T = 10$	$\gamma$	$\beta$	$\varepsilon$	$\sigma$
LC	1.513 (0.049)	0.302 (0.054)	0.100 (0.000)	0.662 (0.096)
LSE	1.501 (0.051)	0.299 (0.052)	0.100 (0.000)	0.600 (0.004)
1.5 scheme	1.513 (0.159)	0.288 (0.161)	0.100 (0.000)	0.599 (0.004)
$\Delta_n = 0.001, T = 50$	$\gamma$	$\beta$	$\varepsilon$	$\sigma$
LC	1.487 (0.054)	0.303 (0.050)	0.100 (0.000)	0.628 (0.098)
LSE	1.493 (0.056)	0.303 (0.052)	0.100 (0.000)	0.601 (0.002)
1.5 scheme	1.488 (0.066)	0.302 (0.068)	0.100 (0.000)	0.600 (0.002)

Table 1: Set 1,  $\gamma_0 = 1.5, \beta_0 = 0.3, \varepsilon_0 = 0.1, \sigma_0 = 0.6$ .. Value without brackets: mean, value in parentheses: standard deviation.



(a) Excitatory set

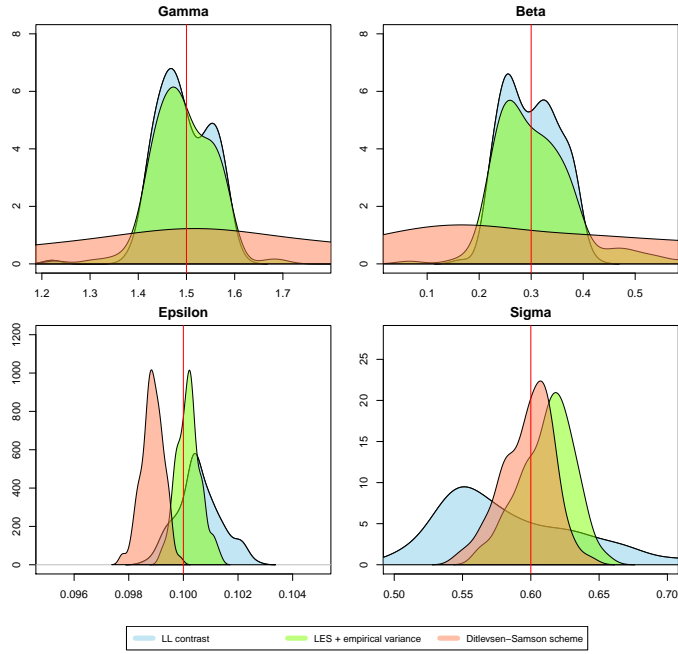


(b) Oscillatory set

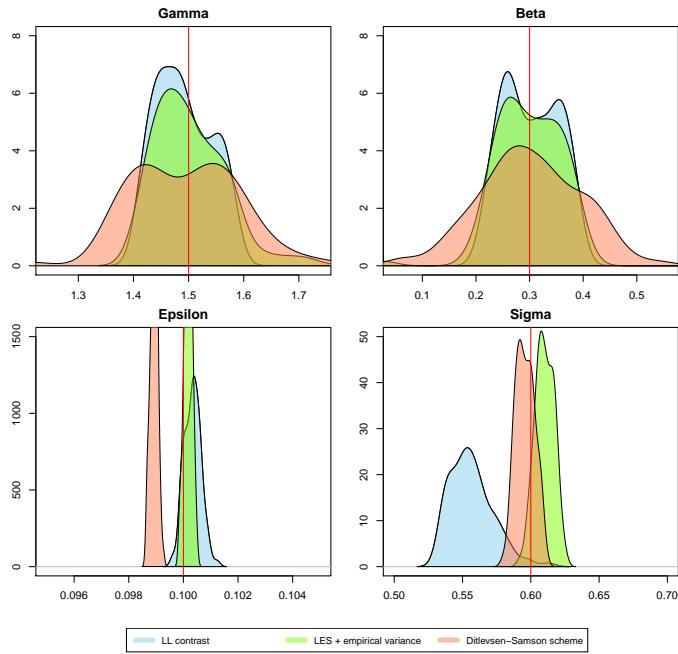
Figure 1: Trajectories for two sets of parameters

$\Delta_n = 0.01, T = 5$	$\gamma$	$\beta$	$\varepsilon$	$\sigma$
LC	1.205 (0.046)	1.311 (0.053)	0.100 (0.001)	0.357 (0.013)
LSE	1.243 (0.771)	1.592 (0.887)	0.101 (0.002)	0.400 (0.014)
1.5 scheme	1.324 (0.357)	1.415 (0.365)	0.095 (0.002)	0.397 (0.014)
$\Delta_n = 0.01, T = 10$	$\gamma$	$\beta$	$\varepsilon$	$\sigma$
LC	1.201 (0.053)	1.303 (0.053)	0.100 (0.001)	0.356 (0.008)
LSE	1.251 (0.367)	1.507 (0.521)	0.100 (0.001)	0.399 (0.009)
1.5 scheme	1.260 (0.187)	1.354 (0.188)	0.091 (0.003)	0.396 (0.009)
$\Delta_n = 0.01, T = 50$	$\gamma$	$\beta$	$\varepsilon$	$\sigma$
LC	1.200 (0.046)	1.302 (0.048)	0.101 (0.001)	0.357 (0.004)
LSE	1.207 (0.208)	1.374 (0.288)	0.100 (0.001)	0.400 (0.004)
1.5 scheme	1.217 (0.073)	1.304 (0.075)	0.083 (0.009)	0.398 (0.004)
$\Delta_n = 0.001, T = 5$	$\gamma$	$\beta$	$\varepsilon$	$\sigma$
LC	1.206 (0.052)	1.302 (0.050)	0.100 (0.000)	0.370 (0.052)
LSE	1.183 (0.074)	1.330 (0.126)	0.100 (0.000)	0.400 (0.004)
1.5 scheme	1.239 (0.170)	1.327 (0.177)	0.100 (0.000)	0.400 (0.004)
$\Delta_n = 0.001, T = 10$	$\gamma$	$\beta$	$\varepsilon$	$\sigma$
LC	1.193 (0.050)	1.303 (0.050)	0.100 (0.000)	0.345 (0.013)
LSE	1.183 (0.069)	1.328 (0.101)	0.100 (0.000)	0.400 (0.003)
1.5 scheme	1.231 (0.126)	1.328 (0.114)	0.099 (0.000)	0.400 (0.003)
$\Delta_n = 0.001, T = 50$	$\gamma$	$\beta$	$\varepsilon$	$\sigma$
LC	1.201 (0.052)	1.301 (0.053)	0.100 (0.000)	0.344 (0.009)
LSE	1.207 (0.208)	1.374 (0.288)	0.100 (0.001)	0.400 (0.004)
1.5 scheme	1.206 (0.088)	1.295 (0.084)	0.099 (0.000)	0.400 (0.001)

Table 2: Set 2:  $\gamma_0 = 1.2, \beta_0 = 1.3, \varepsilon_0 = 0.1, \sigma_0 = 0.4$ . Value without brackets: mean, value in parentheses: standard deviation.

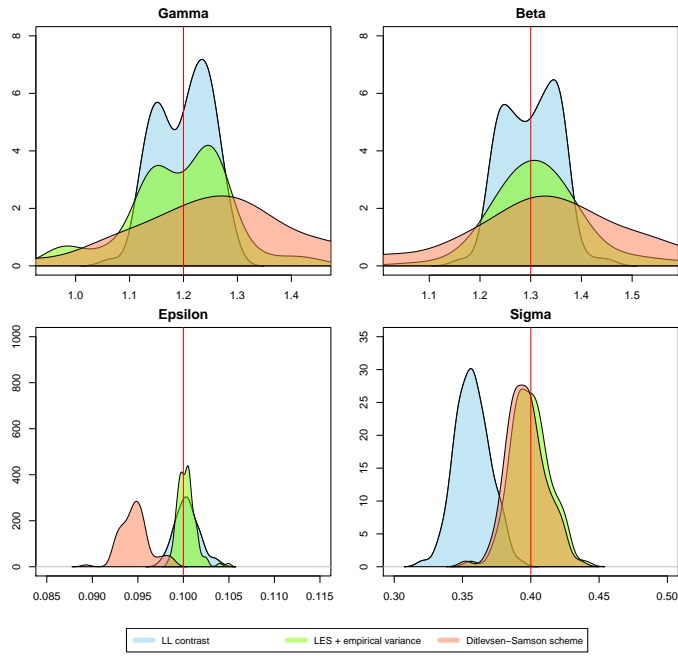


(a)  $T = 5$

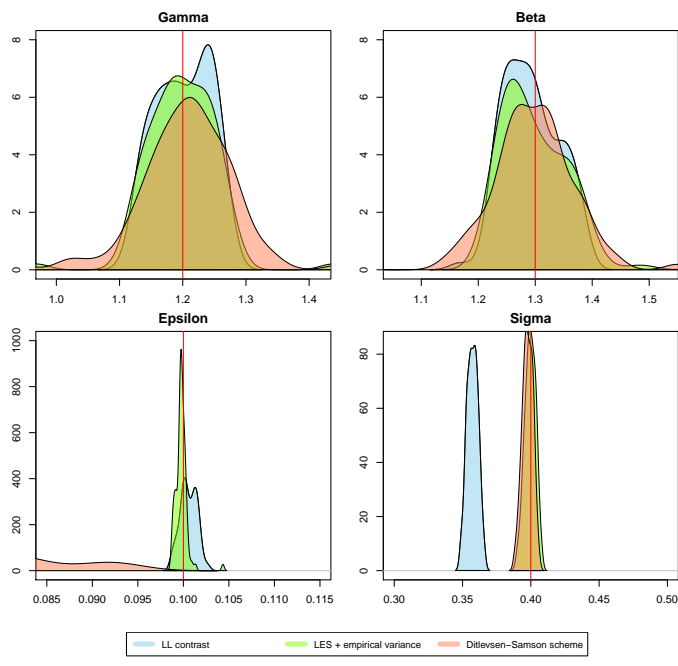


(b)  $T = 50$

Figure 2: Estimation density for the LL contrast (blue), the LSE (red) and 1.5 scheme (green) estimators for the excitatory set.  $\Delta_n = 0.01$



(a)  $T = 5$



(b)  $T = 50$

Figure 3: Estimation density for the LL contrast (blue), the LSE (red) and 1.5 scheme (green) estimators for the inhibitory set.  $\Delta_n = 0.01$



interval is not big enough.

The most important direction of the prospective work is the adaptation of the estimation method to the case when only the observations of the first coordinate are available. Under proper conditions it must be possible to couple the contrast minimization with one of the existing filtering methods and estimate the parameters of the system (at least, partially).

Another point is the generalization of the contrast to systems of higher dimension. In practice we often deal with high-dimensional systems with arbitrary number of rough and smooth variables. The general rule which describes gives the contrast function in that case is not yet established. The most important step here would be to establish the condition of hypoellipticity for such a system. Finally, it is crucial to pair the method with a robust optimization procedure, since the minimization of the contrast is sensitive to choice of the discretization step and initial conditions.

## Acknowledgements

Author's work was financially supported by LabEx PERSYVAL-Lab and LABEX MME-DII. Sincere gratitude is expressed to Adeline Leclercq-Samson and Eva Löcherbach for numerous discussions, helpful remarks and careful proof-reading of the paper draft, as well to the referee, whose comments helped to substantially improve this paper.

This is a post-peer-review, pre-copyedit version of an article published in *Statistical Inference for Stochastic Processes*. The final authenticated version is available online at: <http://dx.doi.org/10.1007/s11203-020-09222-4>

## References

- Markus Ableidinger, Evelyn Buckwar, and Harald Hinterleitner. A stochastic version of the jansen and rit neural mass model: Analysis and numerics. *The Journal of Mathematical Neuroscience*, 7(1):1, August 2017. ISSN 2190-8567. doi: 10.1186/s13408-017-0046-4. URL <http://dx.doi.org/10.1186/s13408-017-0046-4>.
- Ludwig Arnold and Wolfgang Kliemann. On unique ergodicity for degenerate diffusions. *Stochastics: an international journal of probability and stochastic processes*, 21(1):41–61, 1987.
- R. Biscay, J. C. Jimenez, J. J. Riera, and P. A. Valdes. Local linearization method for the numerical solution of stochastic differential equations. *Annals of the Institute of Statistical Mathematics*, 48(4):631–644, Dec 1996. ISSN 1572-9052. doi: 10.1007/BF00052324. URL <https://doi.org/10.1007/BF00052324>.
- Patrick Cattiaux, José R León, and Clémentine Prieur. Estimation for stochastic damping hamiltonian systems under partial observation. ii drift term. *ALEA*, 11(1):p–359, 2014.

- Patrick Cattiaux, Jose R León, Clémentine Prieur, et al. Estimation for stochastic damping hamiltonian systems under partial observation. iii. diffusion term. *The Annals of Applied Probability*, 26(3):1581–1619, 2016.
- S. Ditlevsen and A. Samson. Hypoelliptic diffusions: discretization, filtering and inference from complete and partial observations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 07 2017. doi: 10.1111/rssb.12307.
- Susanne Ditlevsen and Eva Löcherbach. Multi-class oscillating systems of interacting neurons. *SPA*, 127:1840–1869, 2017.
- Richard Fitzhugh. Impulses and physiological states in theoretical models of nerve membrane. *Biophysical Journal*, 1(6):445–466, 1961.
- Danielle Florens-Zmirou. Approximate discrete-time schemes for statistics of diffusion processes. *Statistics: A Journal of Theoretical and Applied Statistics*, 20(4):547–557, 1989.
- C. W. Gardiner and M. J. Collett. Input and output in damped quantum systems: Quantum stochastic differential equations and the master equation. *Phys. Rev. A*, 31:3761–3774, 1985. doi: 10.1103/PhysRevA.31.3761. URL <https://link.aps.org/doi/10.1103/PhysRevA.31.3761>.
- V Genon-Catalot, T Jeantheau, and C Larédo. Stochastic volatility models as hidden markov models and statistical applications. *bernoulli* 6 1051–1079. *Mathematical Reviews (MathSciNet)*, 10:3318471, 2000.
- Valentine Genon-Catalot and Jean Jacod. On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. In *Annales de l’IHP Probabilités et statistiques*, volume 29, pages 119–151, 1993.
- Valentine Genon-Catalot, Thierry Jeantheau, and Catherine Laredo. Parameter estimation for discretely observed stochastic volatility models. *Bernoulli*, 5(5):855–872, 10 1999. URL <http://projecteuclid.org/euclid.bj/1171290402>.
- Arnaud Gloter and Michael Sørensen. Estimation for stochastic differential equations with a small diffusion coefficient. *Stochastic Processes and their Applications*, 119(3):679–699, 2009. URL <https://EconPapers.repec.org/RePEc:eee:spapps:v:119:y:2009:i:3:p:679-699>.
- Peter Hall and Heyde. *Martingale limit theory and its application*. Probability and mathematical statistics. Academic Press, 1980. ISBN 0123193508,9780123193506. URL <http://gen.lib.rus.ec/book/index.php?md5=4A150FD028C59EC118AA1E84CDC93370>.
- A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane currents and its application to conduction and excitation in nerve. *Journal of Physiology-London*, 117(4): 500–544, 1952.

- R Höpfner, E Löcherbach, M Thieullen, et al. Ergodicity for a stochastic hodgkin–huxley model driven by ornstein–uhlenbeck type input. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 52, pages 483–501. Institut Henri Poincaré, 2016.
- Jean Jacod and Philip Protter. *Discretization of processes*, volume 67. Springer Science & Business Media, 2011.
- Anders Christian Jensen. *Statistical Inference for Partially Observed Diffusion Processes*. PhD thesis, University of Copenhagen, 2014.
- J.C. Jimenez and F. Carbonell. Convergence rate of weak local linearization schemes for stochastic differential equations with additive noise. *Journal of Computational and Applied Mathematics*, 279:106 – 122, 2015. ISSN 0377-0427. doi: <https://doi.org/10.1016/j.cam.2014.10.021>. URL <http://www.sciencedirect.com/science/article/pii/S0377042714004671>.
- Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics. Springer, 1 edition, 1987. ISBN 9780387965352,0387965351,3540965351. URL <http://gen.lib.rus.ec/book/index.php?md5=274CB8135034DBB04FE43119E28EEA52>.
- Mathieu Kessler. Estimation of an ergodic diffusion from discrete observations. *Scandinavian Journal of Statistics*, 24(2):211–229, 1997.
- Peter Eris Kloeden, Eckhard Platen, and Henri Schurz. *Numerical solution of SDE through computer experiments*. Universitext. Springer, 2003. ISBN 3540570748,9783540570745. URL <http://gen.lib.rus.ec/book/index.php?md5=FED28E7C31D55C92DA839156DB965F99>.
- A Le-Breton and M Musiela. Some parameter estimation problems for hypoelliptic homogeneous gaussian diffusions. *Banach Center Publications*, 16(1):337–356, 1985.
- Jose R. Leon and Adeline Samson. Hypoelliptic stochastic FitzHugh-Nagumo neuronal model: mixing, up-crossing and estimation of the spike rate. *Annals of Applied Probability*, 28:2243–2274, 2018. URL <https://hal.archives-ouvertes.fr/hal-01492590>.
- José León, Luis Rodriguez, and Roberto Ruggiero. Consistency of a likelihood estimator for stochastic damping hamiltonian systems. totally observed data. 01 2018.
- Paul Malliavin and Anton Thalmaier. *Stochastic calculus of variations in mathematical finance*. Springer finance. Springer, 1 edition, 2006. ISBN 9783540434313,3-540-43431-3. URL <http://gen.lib.rus.ec/book/index.php?md5=A154244D1FB565DD37F6807340E79039>.
- Jonathan C Mattingly, Andrew M Stuart, and Desmond J Higham. Ergodicity for sdes and approximations: locally lipschitz vector fields and degenerate noise. *Stochastic processes and their applications*, 101(2):185–232, 2002.

- D Nualart. *Malliavin Calculus and Related Topics*. Springer. New York. 2006.
- Thoru Ozaki. Statistical identification of nonlinear random vibration systems. *Journal of Applied Mechanics*, 56:186–191, 1989.
- Tohru Ozaki. *Time series modeling of neuroscience data*. Interdisciplinary statistics. Taylor & Francis, 2012. ISBN 9781420094602,1420094602. URL <http://gen.lib.rus.ec/book/index.php?md5=7a5138e691b7b078e99499eee41474b8>.
- Y. Pokern, A. M. Stuart, and P. Wiberg. Parameter estimation for partially observed hypoelliptic diffusions. *J. Roy. Stat. Soc.*, 71(1):49–73, 2007.
- Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293. Springer Science & Business Media, 2013.
- Bernard Roynette. Sur les processus de diffusion de dimension 2. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 32(1-2):95–110, 1975.
- Adeline Samson and Michèle Thieullen. Contrast estimator for completely or partially observed hypoelliptic diffusion. *Stochastic Processes and their Applications*, 122:2521–2552, 2012. doi: 10.1016/j.spa.2012.04.006. URL <https://hal.archives-ouvertes.fr/hal-00598553>.
- Balthasar Van der Pol. A theory of the amplitude of free and forced triode vibrations. *Radio Review*, 1(1920):701–710, 1920.
- Liming Wu. Large and moderate deviations and exponential convergence for stochastic damping hamiltonian systems. *Stochastic processes and their applications*, 91(2):205–238, 2001.

## 7 Appendix

### 7.1 Properties of the scheme

*Proposition 1.* By integrating (8) by parts two times we get the following:

$$\begin{aligned} \mathbb{E} \left[ \tilde{Z}_{i+1} | \tilde{Z}_i \right] &= \tilde{Z}_i + J^{-1}(\tilde{Z}_i; \theta) \left( e^{J(\tilde{Z}_i; \theta) \Delta_n} - I \right) A(\tilde{Z}_i; \theta) + \\ &\quad \frac{1}{2} J^{-2}(\tilde{Z}_i; \theta) \left( e^{J(\tilde{Z}_i; \theta) \Delta_n} - I - J(\tilde{Z}_i; \theta) \Delta_n \right) b^2(\tilde{Z}_i; \sigma) \partial_{yy}^2 A(\tilde{Z}_i; \theta) \quad (25) \end{aligned}$$

Recall that the matrix exponent for some square matrix  $M$  is given by  $e^M = \sum_{l=0}^{\infty} \frac{M^l}{l!}$ . Then (25) can be simplified as:

$$\begin{aligned} \mathbb{E} \left[ \tilde{Z}_{i+1} | \tilde{Z}_i \right] &= \tilde{Z}_i + J^{-1}(\tilde{Z}_i; \theta) \left( I + \Delta_n J(\tilde{Z}_i; \theta) + \frac{\Delta_n^2}{2} J^2(\tilde{Z}_i; \theta) - I + O(\Delta_n^3) \right) A(\tilde{Z}_i; \theta) + \\ &\frac{1}{2} J^{-2}(\tilde{Z}_i; \theta) \left( I + \Delta_n J(\tilde{Z}_i; \theta) + \frac{\Delta_n^2}{2} J^2(\tilde{Z}_i; \theta) - I - \Delta_n J(\tilde{Z}_i; \theta) + O(\Delta_n^3) \right) b^2(\tilde{Z}_i; \sigma) \partial_{yy}^2 A(\tilde{Z}_i; \theta) = \\ &\tilde{Z}_i + \Delta_n A(\tilde{Z}_i; \theta) + \frac{\Delta_n^2}{2} J(\tilde{Z}_i; \theta) A(\tilde{Z}_i; \theta) + \frac{\Delta_n^2}{4} b^2(\tilde{Z}_i; \sigma) \partial_{yy}^2 A(\tilde{Z}_i; \theta) + O(\Delta_n^3) \end{aligned}$$

Writing the above expression component-wise gives the proposition.  $\square$

*Proposition 2.* Let us consider each integral of (10) separately. Denote:

$$\mathcal{W}_{(i+1)\Delta_n} = \int_{i\Delta_n}^{(i+1)\Delta_n} e^{J(\tilde{Z}_i; \theta)((i+1)\Delta_n - s)} B(\tilde{Z}_i; \sigma) dW_s.$$

Recall that the Jacobian of system (3) is given by (6) and the definition of the matrix exponent, we have:

$$\begin{aligned} \mathcal{W}_{(i+1)\Delta_n} &= \int_{i\Delta_n}^{(i+1)\Delta_n} (I + J(\tilde{Z}_i; \theta)((i+1)\Delta_n - s) + O(\Delta_n^2)) B(\tilde{Z}_i; \sigma) dW_s = \\ &= \int_{i\Delta_n}^{(i+1)\Delta_n} \left[ \begin{pmatrix} 1 + \partial_x a_1(\tilde{Z}_i; \theta^{(1)})((i+1)\Delta_n - s) & \partial_y a_1(\tilde{Z}_i; \theta^{(1)})((i+1)\Delta_n - s) \\ \partial_x a_2(\tilde{Z}_i; \theta^{(2)})((i+1)\Delta_n - s) & 1 + \partial_y a_2(\tilde{Z}_i; \theta^{(2)})((i+1)\Delta_n - s) \end{pmatrix} + O(\Delta_n^2) \right] \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} b(\tilde{Z}_i; \sigma) dW_s \\ &= b(\tilde{Z}_i; \sigma) \begin{bmatrix} 0 & \partial_y a_1(\tilde{Z}_i; \theta^{(1)}) \int_{i\Delta_n}^{(i+1)\Delta_n} ((i+1)\Delta_n - s) dW_s + O(\Delta_n^2) \\ 0 & \int_{i\Delta_n}^{(i+1)\Delta_n} dW_s + \partial_y a_2(\tilde{Z}_i; \theta^{(2)}) \int_{i\Delta_n}^{(i+1)\Delta_n} ((i+1)\Delta_n - s) dW_s + O(\Delta_n^2) \end{bmatrix} \end{aligned}$$

Then we can calculate  $\mathbb{E} \left[ \mathcal{W}_{(i+1)\Delta_n} \mathcal{W}'_{(i+1)\Delta_n} \right]$ :

$$\mathbb{E} \left[ \mathcal{W}_{(i+1)\Delta_n} \mathcal{W}'_{(i+1)\Delta_n} \right] = b^2(\tilde{Z}_i; \sigma) \mathbb{E} \begin{pmatrix} \Sigma_{\Delta_n}^{(1)} & \Sigma_{\Delta_n}^{(12)} \\ \Sigma_{\Delta_n}^{(12)} & \Sigma_{\Delta_n}^{(2)} \end{pmatrix} + O(\Delta_n^4),$$

where entries are given by:

$$\begin{aligned} \Sigma_{\Delta_n}^{(1)} &= \left( \partial_y a_1(\tilde{Z}_i; \theta^{(1)}) \right)^2 \left[ \int_{i\Delta_n}^{(i+1)\Delta_n} ((i+1)\Delta_n - s) dW_s \right]^2 \\ \Sigma_{\Delta_n}^{(12)} &= \left( \partial_y a_1(\tilde{Z}_i; \theta^{(1)}) \int_{i\Delta_n}^{(i+1)\Delta_n} ((i+1)\Delta_n - s) dW_s \right) \left( \int_{i\Delta_n}^{(i+1)\Delta_n} dW_s + \partial_y a_2(\tilde{Z}_i; \theta^{(2)}) \int_{i\Delta_n}^{(i+1)\Delta_n} ((i+1)\Delta_n - s) dW_s \right) \\ \Sigma_{\Delta_n}^{(2)} &= \left( \int_{i\Delta_n}^{(i+1)\Delta_n} dW_s + \partial_y a_2(\tilde{Z}_i; \theta^{(2)}) \int_{i\Delta_n}^{(i+1)\Delta_n} ((i+1)\Delta_n - s) dW_s \right)^2 \end{aligned}$$

The first entry can be easily calculated by the Itô isometry:

$$\begin{aligned} \mathbb{E}[\Sigma_{\Delta_n}^{(1)}] &= \left( \partial_y a_1(\tilde{Z}_i; \theta^{(1)}) \right)^2 \mathbb{E} \left[ \int_{i\Delta_n}^{(i+1)\Delta_n} ((i+1)\Delta_n - s) dW_s \right]^2 = \\ &= \left( \partial_y a_1(\tilde{Z}_i; \theta^{(1)}) \right)^2 \int_{i\Delta_n}^{(i+1)\Delta_n} ((i+1)\Delta_n - s)^2 ds = \left( \partial_y a_1(\tilde{Z}_i; \theta^{(1)}) \right)^2 \frac{\Delta_n^3}{3} \end{aligned}$$

Now consider the product of two stochastic integrals in the terms  $\Sigma_{\Delta_n}^{(12)}$  and  $\Sigma_{\Delta_n}^{(2)}$ . Assume for simplicity that  $t = 0$ . From the properties of the stochastic integrals (Karatzas and Shreve, 1987), it is straightforward to see that:

$$\begin{aligned} \mathbb{E} \left[ \lim_{n \rightarrow \infty} \sum_{t_i, t_{i-1} \in [0, \Delta_n]} (\Delta_n - s)(W_{t_i} - W_{t_{i-1}}) \sum_{t_i, t_{i-1} \in [0, \Delta_n]} (W_{t_i} - W_{t_{i-1}}) \right] &= \\ &= \lim_{n \rightarrow \infty} \sum_{t_i, t_{i-1} \in [0, \Delta_n]} (\Delta_n - s) \mathbb{E} [(W_{t_i} - W_{t_{i-1}})^2] = \int_0^{\Delta_n} (\Delta_n - s) ds = \frac{\Delta_n^2}{2} \end{aligned}$$

That gives the proposition.  $\square$

## 7.2 Auxiliary results

We start with an important Lemma which links the sampling and the probabilistic law of the continuous process:

**Lemma 4** (Kessler (1997)). *Let  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$ , let  $f \in \mathbb{R} \times \Theta \rightarrow \mathbb{R}$  be such that  $f$  is differentiable with respect to  $z$  and  $\theta$ , with derivatives of polynomial growth in  $z$  uniformly in  $\theta$ . Then:*

$$\frac{1}{n} \sum_{i=1}^n f(Z_i; \theta) \xrightarrow{\mathbb{P}_0} \int f(z; \theta) \nu_0(dz) \text{ as } n \rightarrow \infty \text{ uniformly in } \theta.$$

Lemma is proven in Kessler (1997) for the one-dimensional case. Its proof is based only on ergodicity of the process and the assumptions analogous to ours, and not on the discretization scheme or dimensionality. So it can be generalized to a multi-dimensional case.

Proposition 4 in combination with the continuous ergodic theorem and Lemma 4 allow us to establish the following important result:

**Lemma 5.** *Let  $f : \mathbb{R}^2 \times \Theta \rightarrow \mathbb{R}$  be a function with the derivatives of polynomial growth in  $x$ , uniformly in  $\theta$ . Assume  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$ . Then:*

$$(i) \frac{1}{n\Delta_n^3} \sum_{i=0}^{n-1} \frac{f(Z_i; \theta)}{(\partial_y a_1(Z_i; \theta_0^{(1)}))^2} \left( X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma) \right)^2 \xrightarrow{\mathbb{P}_0} \frac{1}{3} \int f(z; \theta) b^2(z; \sigma_0) \nu_0(dz)$$

$$(ii) \frac{1}{n\Delta_n} \sum_{i=0}^{n-1} f(Z_i; \theta) (Y_{i+1} - Y_i)^2 \xrightarrow{\mathbb{P}_0} \int f(z; \theta) b^2(z; \sigma_0) \nu_0(dz)$$

$$(iii) \frac{1}{n\Delta_n^2} \sum_{i=0}^{n-1} \frac{f(Z_i; \theta)}{\partial_y a_1(Z_i; \theta_0^{(1)})} \left( X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma) \right) (Y_{i+1} - Y_i) \xrightarrow{\mathbb{P}_0} \frac{1}{2} \int f(z; \theta) b^2(z; \sigma_0) \nu_0(dz)$$

*Proof.* We consider only the cross-term (iii), since the results for the first and the second term are analogous to Ditlevsen and Samson (2017) (upon replacing the bounds from Proposition 3 by 4). Thanks to Proposition 4 we know that:

$$\mathbb{E} \left[ \frac{1}{n\Delta_n^2} \frac{f(Z_i; \theta)}{\partial_y a_1(Z_i; \theta_0^{(1)})} \left( X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma) \right) (Y_{i+1} - Y_i) \mid \mathcal{F}_i \right] = \frac{1}{2n} f(Z_i; \theta) b^2(Z_i; \sigma_0) + O(\Delta_n).$$

Then from Lemma 4 it follows that for  $n \rightarrow \infty$  uniformly in  $\theta$ :

$$\sum_{i=0}^{n-1} \mathbb{E} \left[ \frac{1}{n\Delta_n^2} \frac{f(Z_i; \theta)}{\partial_y a_1(Z_i; \theta_0^{(1)})} \left( X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma) \right) (Y_{i+1} - Y_i) \mid \mathcal{F}_i \right] \xrightarrow{\mathbb{P}_0} \frac{1}{2} \int f(z; \theta) b^2(z; \sigma_0) \nu_0(dz)$$

□

Let us introduce an auxiliary Lemma which establishes the convergence in probability for the first moments:

**Lemma 6.** *Let  $f : \mathbb{R}^2 \times \Theta \rightarrow \mathbb{R}$  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 the following convergence results hold:*

$$(i) \frac{1}{n\Delta_n} \sum_{i=0}^{n-1} f(Z_i; \theta) (X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma)) \xrightarrow{\mathbb{P}_0} 0$$

$$(ii) \frac{1}{n\Delta_n} \sum_{i=0}^{n-1} f(Z_i; \theta) (Y_{i+1} - \bar{A}_2(Z_i; \theta^{(1)}, \theta_0^{(2)}, \sigma)) \xrightarrow{\mathbb{P}_0} 0$$

uniformly in  $\theta$ .

*Proof.* Consider (ii). Expectation of the sum tends to zero for  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$  due to Proposition 4. Convergence for  $\theta^{(1)}$  is due to Lemma 9 in Genon-Catalot and Jacod (1993) and uniformity in  $\theta^{(1)}$  follows the proof of Lemma 10 in Kessler (1997). The second assertion is proven in the same way. For (i) see Lemma 3 in Ditlevsen and Samson (2017). □

We also need the following Lemma for proving the asymptotic normality of the estimators.

**Lemma 7.** *Assume (A1)-(A4) and  $n\Delta_n \rightarrow \infty$  and  $n\Delta_n^2 \rightarrow 0$ . Then for any bounded function  $f(z; \theta) \in \mathbb{R}^2 \times \Theta \rightarrow \mathbb{R}$  the following holds:*

- (i)  $\frac{1}{\sqrt{n\Delta_n^3}} \sum_{i=0}^{n-1} f(Z_i; \theta)(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{3}\nu_0 \left(b^2(z; \sigma_0)(\partial_y a_1(z; \theta_0^{(1)}))^2 f^2(z; \theta)\right)\right)$
- (ii)  $\frac{1}{\sqrt{n\Delta_n^3}} \sum_{i=0}^{n-1} f(Z_i; \theta)(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma))^2 - \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} f(Z_i; \theta) \frac{1}{3} b^2(z; \sigma_0)(\partial_y a_1(z; \theta_0^{(1)}))^2 \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{2}{9}\nu_0 \left(b^4(z; \sigma_0)(\partial_y a_1(z; \theta_0^{(1)}))^4 f^2(z; \theta)\right)\right)$
- (iii)  $\frac{1}{\sqrt{n\Delta_n}} \sum_{i=0}^{n-1} f(Z_i; \theta)(Y_{i+1} - Y_i) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \nu_0 \left(b^2(z; \sigma_0) f^2(z; \theta)\right)\right)$
- (iv)  $\frac{1}{\sqrt{n\Delta_n}} \sum_{i=0}^{n-1} f(Z_i; \theta)(Y_{i+1} - Y_i)^2 - \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} f(Z_i; \theta) b^2(Z_i; \sigma_0) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, 2\nu_0 \left(b^4(z; \sigma_0) f^2(z; \theta)\right)\right)$
- (v)  $\frac{1}{\sqrt{n\Delta_n^2}} \sum_{i=0}^{n-1} f(Z_i; \theta)(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma))(Y_{i+1} - Y_i) - \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} f(Z_i; \theta) \frac{1}{2} b^2(Z_i; \sigma_0) \partial_y a_1(Z_i; \theta_0^{(1)}) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{4}{3}\nu_0 \left(f(z; \theta) b^4(z; \sigma_0)(\partial_y a_1(z; \theta_0^{(1)}))^2\right)\right)$

*Proof.* We focus on the proof of (v), since (i)-(iv) closely follow Lemmas 4-5 in Ditlevsen and Samson (2017). To simplify the proof for the cross-term, we recall that the representation (15) can be transformed so that the two noise terms are independent. For example, we can use an analogue of such a decomposition proposed in Pokern et al. (2007):

$$X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma) = b(Z_i; \sigma_0) \partial_y a_1(Z_i; \theta_0^{(1)}) \left( \frac{\Delta_n^{\frac{3}{2}}}{\sqrt{12}} \eta_i^1 + \frac{\Delta_n^{\frac{3}{2}}}{2} \eta_i^2 \right) + \delta_i^1$$

$$Y_{i+1} - Y_i = \Delta_n a_2(Z_i; \theta^{(2)}) + b(Z_i; \sigma_0) \Delta_n^{\frac{1}{2}} \eta_i^2 + \delta_i^2,$$

where  $\delta_i^1$  and  $\delta_i^2$  are error terms such that  $\mathbb{E}[\delta_i^k | \mathcal{F}_i] = O(\Delta_n^2)$  and  $\mathbb{E}[(\delta_i^k)^2 | \mathcal{F}_i] = O(\Delta_n^4)$  (see Proposition 4), and  $\eta_i^1$  and  $\eta_i^2$  are standard independent normal variables.

Then Proposition 4 gives that  $\mathbb{E}\left[\left(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma)\right)(Y_{i+1} - Y_i) | \mathcal{F}_i\right] = \frac{\Delta_n^2}{2} b(Z_i; \sigma_0) \partial_y a_1(Z_i; \theta_0^{(1)}) + O(\Delta_n^3)$ , and then  $\mathbb{E}\left[f(Z_i; \theta) \left(\left(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma)\right)(Y_{i+1} - Y_i) - \frac{\Delta_n^2}{2} b^2(Z_i; \sigma_0) \partial_y a_1(Z_i; \theta_0^{(1)})\right) | \mathcal{F}_i\right] = 0$ . With slightly more tedious computations (which are omitted) we get also that

$$\mathbb{E}\left[\left(\left(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma)\right)(Y_{i+1} - Y_i) - \frac{\Delta_n^2}{2} b^2(Z_i; \sigma_0) (\partial_y a_1(Z_i; \theta_0^{(1)}))^2\right)^2 | \mathcal{F}_i\right] =$$

$$\frac{4\Delta_n^4}{3} b^4(Z_i; \sigma_0) (\partial_y a_1(Z_i; \theta_0^{(1)}))^2 + O(\Delta_n^5)$$



Then we obtain:

$$\begin{aligned}
& \frac{1}{\sqrt{n}\Delta_n^2} \sum_{i=0}^{n-1} f(Z_i; \theta) \left( X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma) \right) (Y_{i+1} - Y_i) - \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} f(Z_i; \theta) \frac{1}{2} b^2(Z_i; \sigma_0) \partial_y a_1(Z_i; \theta_0^{(1)}) \\
&= \frac{1}{\sqrt{n}\Delta_n^2} \sum_{i=0}^{n-1} f(Z_i; \theta) \left( b(Z_i; \sigma_0) \partial_y a_1(Z_i; \theta_0^{(1)}) \left( \frac{\Delta_n^{\frac{3}{2}}}{\sqrt{12}} \eta_i^1 + \frac{\Delta_n^{\frac{3}{2}}}{2} \eta_i^2 \right) + \delta_i^1 \right) \\
& \quad \left( \Delta_n a_2(Z_i; \theta^{(2)}) + b(Z_i; \sigma_0) \Delta_n^{\frac{1}{2}} \eta_i^2 + \delta_i^2 \right) - \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} f(Z_i; \theta) \frac{1}{2} b^2(Z_i; \sigma_0) \partial_y a_1(Z_i; \theta_0^{(1)})
\end{aligned}$$

Since  $\frac{\Delta_n}{n} \rightarrow 0$  by design we see that

$$\begin{aligned}
& \frac{1}{n\Delta_n^4} \mathbb{E} \left[ \sum_{i=0}^{n-1} f^2(Z_i; \theta) \left( \left( X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma) \right) (Y_{i+1} - Y_i) - \right. \right. \\
& \quad \left. \left. \frac{\Delta_n^2}{2} b(Z_i; \sigma_0) (\partial_y a_1(Z_i; \theta_0^{(1)})) \right)^2 \right] \rightarrow \frac{4}{3} \nu_0 \left( f^2(z; \theta) b^4(z; \sigma_0) (\partial_y a_1(z; \theta_0^{(1)}))^2 \right)
\end{aligned}$$

Further, since  $\mathbb{E} \left[ f^4(Z_i; \theta) \left( \left( X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma) \right) (Y_{i+1} - Y_i) - \frac{\Delta_n^2}{2} b^2(Z_i; \sigma_0) (\partial_y a_1(Z_i; \theta_0^{(1)})) \right)^4 \middle| \mathcal{F}_i \right]$  is bounded by (A2), we have

$$\frac{1}{n^2 \Delta_n^8} \mathbb{E} \left[ \sum_{i=0}^{n-1} f^4(Z_i; \theta) \left( \left( X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma) \right) (Y_{i+1} - Y_i) - \frac{\Delta_n^2}{2} b^2(Z_i; \sigma_0) (\partial_y a_1(Z_i; \theta_0^{(1)})) \right)^4 \middle| \mathcal{F}_i \right] \rightarrow 0.$$

Therefore, we can apply again the Theorem 3.2 from Hall and Heyde (1980) and obtain the statement (v).  $\square$

**Remark.** Note that the results for the convergence in distribution for the increments of the second coordinate hold without any assumption on the parameters of the function  $a_2(z; \theta^{(2)})$ . It is due to the fact that the order of the noise dominates the order of the drift term (which is not the case in first coordinate, where the noise is propagated with the higher order). As a consequence, the convergence of a functional  $\sum_{i=0}^{n-1} f(Z_i; \theta) (Y_{i+1} - \bar{A}_2(Z_i; \theta^{(1)}, \theta^{(2)}, \sigma))$  holds, with a proper scaling, for any value of  $\theta$ .

### 7.3 Consistency and asymptotic normality of the LL contrast estimator

*Lemma 1.* Consider

$$\frac{\Delta_n}{n} \left[ \mathcal{L}_{n, \Delta_n}(\theta^{(1)}, \theta^{(2)}, \sigma^2; Z_{0:n}) - \mathcal{L}_{n, \Delta_n}(\theta_0^{(1)}, \theta^{(2)}, \sigma^2; Z_{0:n}) \right] = T_1 + T_2 + T_3 + T_4,$$

where the terms are given as follows:

$$\begin{aligned}
T_1 &= \frac{6\Delta_n}{n\Delta_n^3} \sum_{i=0}^{n-1} \left[ \frac{(X_{i+1} - \bar{A}_1(Z_i; \theta^{(1)}, \theta^{(2)}, \sigma))^2}{b^2(Z_i; \sigma) (\partial_y a_1(Z_i; \theta^{(1)}))^2} - \frac{(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma))^2}{b^2(Z_i; \sigma) (\partial_y a_1(Z_i; \theta_0^{(1)}))^2} \right] \\
T_2 &= -\frac{6\Delta_n}{n\Delta_n^2} \sum_{i=0}^{n-1} \frac{1}{b^2(Z_i; \sigma)} \left[ \frac{(X_{i+1} - \bar{A}_1(Z_i; \theta^{(1)}, \theta^{(2)}, \sigma)) (Y_{i+1} - \bar{A}_2(Z_i; \theta^{(1)}, \theta^{(2)}, \sigma))}{\partial_y a_1(Z_i; \theta^{(1)})} - \right. \\
&\quad \left. \frac{(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma)) (Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma))}{\partial_y a_1(Z_i; \theta_0^{(1)})} \right] \\
T_3 &= \frac{2\Delta_n}{n\Delta_n} \sum_{i=0}^{n-1} \left[ \frac{(Y_{i+1} - \bar{A}_2(Z_i; \theta^{(1)}, \theta^{(2)}, \sigma))^2}{b^2(Z_i; \sigma)} - \frac{(Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma))^2}{b^2(Z_i; \sigma)} \right] \\
T_4 &= \frac{\Delta_n}{n} \sum_{i=0}^{n-1} \log \left( \frac{\partial_y a_1(Z_i; \theta^{(1)})}{\partial_y a_1(Z_i; \theta_0^{(1)})} \right)
\end{aligned}$$

Consider term  $T_1$ :

$$\begin{aligned}
T_1 &= \frac{6\Delta_n}{n\Delta_n^3} \sum_{i=0}^{n-1} \frac{1}{b^2(Z_i; \sigma)} \left[ \frac{(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma) + \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma) - \bar{A}_1(Z_i; \theta^{(1)}, \theta^{(2)}, \sigma))^2}{(\partial_y a_1(Z_i; \theta^{(1)}))^2} - \right. \\
&\quad \left. \frac{(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma))^2}{(\partial_y a_1(Z_i; \theta_0^{(1)}))^2} \right] = \frac{6\Delta_n}{n\Delta_n^3} \sum_{i=0}^{n-1} \frac{1}{b^2(Z_i; \sigma)} \left[ (X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma))^2 \left[ \frac{1}{(\partial_y a_1(Z_i; \theta^{(1)}))^2} - \right. \right. \\
&\quad \left. \left. \frac{1}{(\partial_y a_1(Z_i; \theta_0^{(1)}))^2} \right] + \frac{2\Delta_n}{(\partial_y a_1(Z_i; \theta^{(1)}))^2} (X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma)) (a_1(Z_i; \theta_0^{(1)}) - a_1(Z_i; \theta^{(1)})) + \right. \\
&\quad \left. \frac{\Delta_n^2}{(\partial_y a_1(Z_i; \theta^{(1)}))^2} (a_1(Z_i; \theta_0^{(1)}) - a_1(Z_i; \theta^{(1)}))^2 \right].
\end{aligned}$$

Recalling Lemmas 4, 6 and 5 we have that:

$$\begin{aligned} & \frac{6}{n\Delta_n^2} \sum_{i=0}^{n-1} \frac{\left(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma)\right)^2}{b^2(Z_i; \sigma)} \left[ \frac{1}{(\partial_y a_1(Z_i; \theta^{(1)}))^2} - \frac{1}{\left(\partial_y a_1(Z_i; \theta_0^{(1)})\right)^2} \right] \xrightarrow{\mathbb{P}_0} 0 \\ & \frac{6}{n\Delta_n} \sum_{i=0}^{n-1} \frac{1}{b^2(Z_i; \sigma)(\partial_y a_1(Z_i; \theta^{(1)}))^2} \left(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma)\right) (a_1(Z_i; \theta_0^{(1)}) - a_1(Z_i; \theta^{(1)})) \xrightarrow{\mathbb{P}_0} 0 \\ & \frac{6}{n} \sum_{i=0}^{n-1} \frac{(a_1(Z_i; \theta_0^{(1)}) - a_1(Z_i; \theta^{(1)}))^2}{b^2(Z_i; \sigma)(\partial_y a_1(Z_i; \theta^{(1)}))^2} \xrightarrow{\mathbb{P}_0} 6 \int \frac{(a_1(z; \theta_0^{(1)}) - a_1(z; \theta^{(1)}))^2}{b^2(z; \sigma)(\partial_y a_1(z; \theta^{(1)}))^2} \nu_0(dz). \end{aligned}$$

Now consider  $T_2$ , which can be rewritten as:

$$\begin{aligned} & -\frac{6}{n\Delta_n} \sum_{i=0}^{n-1} \frac{(Y_{i+1} - Y_i + O(\Delta_n))}{b^2(Z_i; \sigma)} \left[ \frac{\left(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma) + \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma) - \bar{A}_1(Z_i; \theta^{(1)}, \theta^{(2)}, \sigma)\right)}{\partial_y a_1(Z_i; \theta^{(1)})} \right. \\ & \left. \frac{\left(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma)\right)}{\partial_y a_1(Z_i; \theta_0^{(1)})} \right] = -\frac{6}{n\Delta_n} \sum_{i=0}^{n-1} \frac{(Y_{i+1} - Y_i + O(\Delta_n))}{b^2(Z_i; \sigma)} \left[ \left(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma)\right) \right. \\ & \left. \left[ \frac{1}{\partial_y a_1(Z_i; \theta^{(1)})} - \frac{1}{\partial_y a_1(Z_i; \theta_0^{(1)})} \right] + \frac{\Delta_n}{(\partial_y a_1(Z_i; \theta^{(1)}))} (a_1(Z_i; \theta_0^{(1)}) - a_1(Z_i; \theta^{(1)})) \right]. \end{aligned}$$

Then we use the fact that the expectation of  $\left(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma)\right)$  is of order  $\Delta_n^2$  and of increments  $Y_{i+1} - Y_i$  is of  $\Delta_n$ , and by Lemma 5 we obtain:

$$-\frac{6}{n\Delta_n} \sum_{i=0}^{n-1} \frac{\left(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma)\right) (Y_{i+1} - Y_i + O(\Delta_n))}{b^2(Z_i; \sigma)(\partial_y a_1(Z_i; \theta_0^{(1)}))} \left[ \frac{(\partial_y a_1(Z_i; \theta_0^{(1)}))}{(\partial_y a_1(Z_i; \theta^{(1)}))} - 1 \right] \xrightarrow{\mathbb{P}_0} 0.$$

The same holds for  $T_4$ . Consider then  $T_3$ :

$$\begin{aligned} & \frac{2\Delta_n}{n\Delta_n} \sum_{i=0}^{n-1} \left[ 2 \frac{(Y_{i+1} - \bar{A}_2(Z_i; \theta^{(1)}, \theta^{(2)}, \sigma))^2}{b^2(Z_i; \sigma)} - \frac{\left(Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma)\right)^2}{b^2(Z_i; \sigma)} \right] = \\ & \frac{2}{n} \sum_{i=0}^{n-1} \frac{1}{b^2(Z_i; \sigma)} \left[ \left(Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma)\right) \left(\bar{A}_2(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma) - \bar{A}_2(Z_i; \theta^{(1)}, \theta^{(2)}, \sigma)\right) - \right. \\ & \left. \left(\bar{A}_2(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma) - \bar{A}_2(Z_i; \theta^{(1)}, \theta^{(2)}, \sigma)\right)^2 \right] \end{aligned}$$

This term is of order  $O(\Delta_n^3)$  (since  $\theta^{(1)}$  is contained only in terms of order  $\Delta_n^2$ ), thus it converges to zero as  $\Delta_n \rightarrow 0$ . Thus, we indeed have

$$\lim_{n \rightarrow \infty, \Delta_n \rightarrow 0} \frac{\Delta_n}{n} \left[ \mathcal{L}_{n, \Delta_n}(\theta^{(1)}, \theta^{(2)}, \sigma^2; Z_{0:n}) - \mathcal{L}_{n, \Delta_n}(\theta_0^{(1)}, \theta^{(2)}, \sigma^2; Z_{0:n}) \right] \xrightarrow{\mathbb{P}_0} 6 \int \frac{(a_1(z; \theta_0^{(1)}) - a_1(z; \theta^{(1)}))^2}{b^2(z; \sigma)(\partial_y a_1(z; \theta^{(1)}))^2} \nu_0(dz). \quad (26)$$

□

*Theorem 1 (consistency and asymptotic normality of  $\theta^{(1)}$ ).* Throughout the proof we assume that  $\theta^{(1)} \in \mathbb{R}$  in order to simplify the notations.

**Consistency.** It follows essentially from Lemma 1. Indeed, the result of the Lemma (and the fact that the parameter space is compact) implies that we can find a subsequence  $\hat{\theta}_{n, \Delta_n}^{(1)}$  which converges to some value  $\theta_\infty^{(1)}$ . However, the minimum of the expression in Lemma 1 is attained for  $\theta_0^{(1)}$ . Then by identifiability of the drift function we have the consistency, that is  $\hat{\theta}_{n, \Delta_n}^{(1)} \rightarrow \theta_0^{(1)}$ .

**Asymptotic normality.** The proof follows the standard pattern (see Kessler (1997), Genon-Catalot et al. (1999), Ditlevsen and Samson (2017)). First, we write the Taylor expansion of the function (17). Then we have:

$$\int \frac{\Delta_n}{n} \frac{\partial^2}{\partial \theta^{(1)} \partial \theta^{(1)}} \mathcal{L}_{n, \Delta_n} \left( \theta_0^{(1)} + u(\hat{\theta}_{n, \Delta_n}^{(1)} - \theta_0), \theta^{(2)}, \sigma; z \right) du \cdot \sqrt{\frac{n}{\Delta_n}} (\hat{\theta}_{n, \Delta_n}^{(1)} - \theta_0^{(1)}) = - \sqrt{\frac{\Delta_n}{n}} \frac{\partial}{\partial \theta^{(1)}} \mathcal{L}_{n, \Delta_n}(\theta_0^{(1)}, \theta^{(2)}, \sigma; z)$$

Note that the values of  $\theta^{(2)}$  and  $\sigma$  may be taken arbitrary. Now we have to compute the first and the second order derivatives of (17). We omit the dependency on parameters in the expression for partial derivatives to make it readable and study the convergence of the first order derivative:

$$\begin{aligned} \frac{\partial}{\partial \theta^{(1)}} \mathcal{L}_{n, \Delta_n}(\theta_0^{(1)}, \theta^{(2)}, \sigma; z) &= \sum_{i=1}^{n-1} \left[ \frac{2 \partial_{y, \theta^{(1)}}^2 a_1}{\partial_y a_1} - \frac{6}{b^2(Z_i; \sigma) \partial_y a_1} \right. \\ &\quad \left[ \frac{2(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma))^2}{\Delta_n^3 (\partial_y a_1)^2} \partial_{y, \theta^{(1)}}^2 a_1 + \frac{2(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma))(\partial_{\theta^{(1)}} a_1)}{\Delta_n^2 (\partial_y a_1)} - \right. \\ &\quad \left. \frac{(Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma))(\partial_{\theta^{(1)}} a_1)}{\Delta_n} - \right. \\ &\quad \left. \left. \frac{(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma))(Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma))(\partial_{y, \theta^{(1)}}^2 a_1)}{\Delta_n^2 (\partial_y a_1)} \right] \right] \quad (27) \end{aligned}$$

Under assumption (A5) the only non-zero terms are the following:

$$\frac{\partial}{\partial \theta^{(1)}} \mathcal{L}_{n, \Delta_n}(\theta_0^{(1)}, \theta^{(2)}, \sigma; z) = \sum_{i=1}^{n-1} -\frac{6}{b^2(Z_i; \sigma) \partial_y a_1} \left[ \frac{2(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma))(\partial_{\theta^{(1)}} a_1)}{\Delta_n^2 (\partial_y a_1)} - \frac{(Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma))(\partial_{\theta^{(1)}} a_1)}{\Delta_n} \right]$$

Applying Lemma 7, we get:

$$\begin{aligned} & \frac{1}{\sqrt{n \Delta_n^3}} \sum_{i=1}^{n-1} \left[ \frac{12(\partial_{\theta^{(1)}} a_1)}{b^2(Z_i; \sigma) (\partial_y a_1)} (X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma)) \right] \xrightarrow{\mathcal{D}} \mathcal{N} \left( 0, 36\nu_0 \left( \frac{b^2(z; \sigma_0)}{b^4(z; \sigma)} (\partial_{\theta^{(1)}} a_1)^2 \right) \right) \\ & \frac{1}{\sqrt{n \Delta_n}} \sum_{i=1}^{n-1} \left[ \frac{6(\partial_{\theta^{(1)}} a_1)}{b^2(Z_i; \sigma)} \frac{(Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma))}{(\partial_y a_1)} \right] \xrightarrow{\mathcal{D}} \mathcal{N} \left( 0, 36\nu_0 \left( \frac{b^2(z; \sigma_0)}{b^4(z; \sigma)} \frac{(\partial_{\theta^{(1)}} a_1)^2}{(\partial_y a_1)^2} \right) \right) \end{aligned}$$

Thus, we have the following convergence in law:

$$\sqrt{\frac{\Delta_n}{n}} \frac{\partial}{\partial \theta^{(1)}} \mathcal{L}_{n, \Delta_n}(\theta_0^{(1)}, \theta^{(2)}, \sigma; z) \xrightarrow{\mathcal{D}} \mathcal{N} \left( 0, 36\nu_0 \left( \frac{b^2(z; \sigma_0)}{b^4(z; \sigma)} (\partial_{\theta^{(1)}} a_1)^2 \left( 1 + \frac{1}{(\partial_y a_1)^2} \right) \right) \right)$$

For the second order derivative we split again the expression (27) in several parts and study their convergence:

$$\begin{aligned} T_1 &:= \frac{\Delta_n}{n} \sum_{i=1}^{n-1} -\frac{12\Delta_n}{\Delta_n^2 b^2(Z_i; \sigma) (\partial_y a_1)^2} \left[ (\partial_{\theta^{(1)}} a_1)^2 + (X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma)) \frac{(\partial_{\theta^{(1)} \theta^{(1)}}^2 a_1) (\partial_y a_1)}{(\partial_y a_1)^2} \right] \\ T_2 &:= \frac{\Delta_n}{n} \sum_{i=1}^{n-1} \frac{6(Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma)) (\partial_y a_1)^2 \partial_{\theta^{(1)} \theta^{(1)}}^2 a_1}{\Delta_n b^2(Z_i; \sigma) (\partial_y a_1)^4} \end{aligned}$$

It is easy to see that the terms  $T_2$  converges to 0 by Lemmas 6 and 5.  $T_1$ , according to the Lemma 4 and Lemma 6, converges to  $12 \int \frac{(\partial_{\theta^{(1)}} a_1)^2}{b^2(z; \sigma) (\partial_y a_1)} \nu_0(dz)$ . That gives the result.  $\square$

*Lemma 2.* Note that we cannot infer the value of  $\theta^{(2)}$  with the same scaling as the parameter of the smooth coordinate because the estimator for each variable converges with different speed. Thus, we fix the parameter  $\theta^{(1)}$  to its estimated value  $\hat{\theta}_{n, \Delta_n}^{(1)}$  and consider the same sum, but with a different scaling, namely :

$$\lim_{n \rightarrow \infty, \Delta_n \rightarrow 0} \frac{1}{n \Delta_n} \left[ \mathcal{L}_{n, \Delta_n}(\hat{\theta}_{n, \Delta_n}^{(1)}, \theta^{(2)}, \sigma^2; Z_{0:n}) - \mathcal{L}_{n, \Delta_n}(\hat{\theta}_{n, \Delta_n}^{(1)}, \theta_0^{(2)}, \sigma^2; Z_{0:n}) \right] = T_1 + T_2 + T_3$$

where the terms are given as follows:

$$\begin{aligned}
T_1 &= \frac{6}{n\Delta_n^4} \sum_{i=0}^{n-1} \frac{1}{b^2(Z_i; \sigma) \left( \partial_y a_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}) \right)^2} \left[ \left( X_{i+1} - \bar{A}_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta^{(2)}, \sigma) \right)^2 - \right. \\
&\qquad \qquad \qquad \left. \left( X_{i+1} - \bar{A}_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta_0^{(2)}, \sigma) \right)^2 \right] \\
T_2 &= -\frac{6}{n\Delta_n^3} \sum_{i=0}^{n-1} \frac{1}{b^2(Z_i; \sigma)} \left[ \frac{\left( X_{i+1} - \bar{A}_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta^{(2)}, \sigma) \right) \left( Y_{i+1} - \bar{A}_2(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta^{(2)}, \sigma) \right)}{\partial_y a_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)})} - \right. \\
&\qquad \qquad \qquad \left. \frac{\left( X_{i+1} - \bar{A}_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta_0^{(2)}, \sigma) \right) \left( Y_{i+1} - \bar{A}_2(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta_0^{(2)}, \sigma) \right)}{\partial_y a_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)})} \right] \\
T_3 &= \frac{2}{n\Delta_n^2} \sum_{i=0}^{n-1} \frac{\left[ \left( Y_{i+1} - \bar{A}_2(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta^{(2)}, \sigma) \right)^2 - \left( Y_{i+1} - \bar{A}_2(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta_0^{(2)}, \sigma) \right)^2 \right]}{b^2(Z_i; \sigma)}
\end{aligned}$$

We start with  $T_1$ :

$$\begin{aligned}
T_1 &= \frac{6}{n\Delta_n^4} \sum_{i=0}^{n-1} \frac{1}{b^2(Z_i; \sigma) \left( \partial_y a_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}) \right)^2} \left[ \left( \bar{A}_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta_0^{(2)}, \sigma) - \bar{A}_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta^{(2)}, \sigma) \right) \right. \\
&\qquad \qquad \qquad \left. \left( X_{i+1} - \bar{A}_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta_0^{(2)}, \sigma) \right) - \left( \bar{A}_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta_0^{(2)}, \sigma) - \bar{A}_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta^{(2)}, \sigma) \right)^2 \right] = \\
&= \frac{6}{n\Delta_n^4} \sum_{i=0}^{n-1} \frac{1}{b^2(Z_i; \sigma) \left( \partial_y a_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}) \right)^2} \left[ \frac{\Delta_n^2}{2} \left( \partial_y a_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}) \right) \left( a_2(Z_i; \theta_0^{(2)}) - a_2(Z_i; \theta^{(2)}) \right) \right. \\
&\qquad \qquad \qquad \left( X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta_0^{(2)}, \sigma) \right) + \frac{\Delta_n^3}{2} \left( \partial_y a_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}) \right) \left( a_2(Z_i; \theta_0^{(2)}) - a_2(Z_i; \theta^{(2)}) \right) \\
&\qquad \qquad \qquad \left( a_1(Z_i; \theta_0^{(1)}) - a_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}) \right) + \frac{\Delta_n^4}{4} \left( \partial_y a_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}) \right)^2 \left( a_2(Z_i; \theta_0^{(2)}) - a_2(Z_i; \theta^{(2)}) \right)^2 - \\
&\qquad \qquad \qquad \left. \frac{\Delta_n^4}{4} \left( \partial_y a_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}) \right)^2 \left( a_2(Z_i; \theta_0^{(2)}) - a_2(Z_i; \theta^{(2)}) \right)^2 \right]
\end{aligned}$$

Recall that  $\left( X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta_0^{(2)}, \sigma) \right)$  is of  $O(\Delta_n^3)$  by Proposition 3. Thus, the first summand of the  $T_1$  is of order  $\Delta_n^5$  and converges to 0. The second summand, under assumption (A5), can

be rewritten as

$$\begin{aligned} \frac{3}{n\Delta_n^4} \sum_{i=0}^{n-1} \frac{\Delta_n^3 \left( \theta_0^{(1)} - \hat{\theta}_{n,\Delta_n}^{(1)} \right)^T g(X_i) \left( a_2(Z_i; \theta_0^{(2)}) - a_2(Z_i; \theta^{(2)}) \right)}{b^2(Z_i; \sigma) \left( \partial_y a_1(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}) \right)} &= \\ \frac{3}{\sqrt{n\Delta_n n}} \sum_{i=0}^{n-1} \frac{\sqrt{\frac{n}{\Delta_n}} \left( \theta_0^{(1)} - \hat{\theta}_{n,\Delta_n}^{(1)} \right)^T g(X_i) \left( a_2(Z_i; \theta_0^{(2)}) - a_2(Z_i; \theta^{(2)}) \right)}{b^2(Z_i; \sigma) \left( \partial_y a_1(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}) \right)}. \end{aligned}$$

$\sqrt{\frac{n}{\Delta_n}} \left( \theta_0^{(1)} - \hat{\theta}_{n,\Delta_n}^{(1)} \right)^T$  converges to a normal variable with zero mean due to Theorem 1, and the whole expression converges to 0, because  $n\Delta_n \rightarrow \infty$ . Thus  $T_1$  converges to 0. Consider  $T_2$ :

$$T_2 = -\frac{6}{n\Delta_n^3} \sum_{i=0}^{n-1} \left[ \frac{\Delta_n \left( X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma) \right) \left( a_2(Z_i; \theta^{(2)}) - a_2(Z_i; \theta_0^{(2)}) \right)}{\partial_y a_1(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}) b^2(Z_i; \sigma)} + \frac{\Delta_n^2 \left( a_1(Z_i; \theta_0^{(1)}) - a_1(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}) \right) \left( a_2(Z_i; \theta^{(2)}) - a_2(Z_i; \theta_0^{(2)}) \right)}{\partial_y a_1(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}) b^2(Z_i; \sigma)} \right]$$

Then, the first part of the sum converges to zero in probability after applying Lemma 6. The second part of the sum also converges to zero because  $n\Delta_n \rightarrow \infty$  by design, and  $\hat{\theta}_{n,\Delta_n}^{(1)} \xrightarrow{\mathbb{P}_0} \theta_0^{(1)}$ . So that, recalling (A5), and applying the arguments used above for  $T_1$  to  $a_1(Z_i; \theta_0^{(1)}) - a_1(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}) = \left( \theta_0^{(1)} - \hat{\theta}_{n,\Delta_n}^{(1)} \right) g(X_i)$ , we prove that  $T_2$  also converges to 0. So we just have to consider the remaining term  $T_3$ :

$$\begin{aligned} T_3 &= \frac{2}{n\Delta_n^2} \sum_{i=0}^{n-1} \frac{1}{b^2(Z_i; \sigma)} \left[ (Y_{i+1} - \bar{A}_2(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}, \theta_0^{(2)}, \sigma))^2 + \right. \\ &\quad (Y_{i+1} - \bar{A}_2(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}, \theta_0^{(2)}, \sigma)) (\bar{A}_2(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}, \theta_0^{(2)}, \sigma) - \bar{A}_2(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}, \theta^{(2)}, \sigma)) + \\ &\quad \left. (\bar{A}_2(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}, \theta_0^{(2)}, \sigma) - \bar{A}_2(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}, \theta^{(2)}, \sigma))^2 - (Y_{i+1} - \bar{A}_2(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}, \theta_0^{(2)}, \sigma))^2 \right] = \\ &= \frac{2}{n\Delta_n^2} \sum_{i=0}^{n-1} \frac{1}{b^2(Z_i; \sigma)} \left[ \Delta_n (Y_{i+1} - \bar{A}_2(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}, \theta_0^{(2)}, \sigma)) \left( a_2(Z_i; \theta_0^{(2)}) - a_2(Z_i; \theta^{(2)}) \right) + \right. \\ &\quad \left. \Delta_n^2 \left( a_2(Z_i; \theta_0^{(2)}) - a_2(Z_i; \theta^{(2)}) \right)^2 \right] \end{aligned}$$

The first part of the sum converges to 0 due to Lemma 6. Then we apply Lemma 5 and get the

convergence:

$$\lim_{n \rightarrow \infty, \Delta_n \rightarrow 0} \frac{1}{n\Delta_n} \left[ \mathcal{L}_{n,\Delta_n}(\hat{\theta}_{n,\Delta_n}^{(1)}, \theta^{(2)}, \sigma^2; Z_{0:n}) - \mathcal{L}_{n,\Delta_n}(\hat{\theta}_{n,\Delta_n}^{(1)}, \theta_0^{(2)}, \sigma_0^2; Z_{0:n}) \right] \xrightarrow{\mathbb{P}_0} 2 \int \frac{(a_2(z; \theta_0^{(2)}) - a_2(z; \theta^{(2)}))^2}{b^2(z; \sigma)} \nu_0(dz)$$

□

*Lemma 3.* We can split the contrast in the following sum:

$$\lim_{n \rightarrow \infty, \Delta_n \rightarrow 0} \frac{1}{2n} \mathcal{L}_{n,\Delta_n}(\hat{\theta}_{n,\Delta_n}^{(1)}, \theta^{(2)}, \sigma^2; Z_{0:n}) = \lim_{n \rightarrow \infty, \Delta_n \rightarrow 0} [3T_1 - 3T_2 + T_3 + T_4]$$

where terms are given by follows:

$$\begin{aligned} T_1 &= \frac{1}{n} \sum_{i=0}^{n-1} \frac{(X_{i+1} - \bar{A}_1(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}, \theta^{(2)}, \sigma))^2}{\Delta_n^3 b^2(Z_i; \sigma) (\partial_y a_1(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}))^2} \\ T_2 &= \frac{1}{n} \sum_{i=0}^{n-1} \frac{(X_{i+1} - \bar{A}_1(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}, \theta^{(2)}, \sigma))(Y_{i+1} - \bar{A}_2(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}, \theta^{(2)}, \sigma))}{\Delta_n^2 b^2(Z_i; \sigma) (\partial_y a_1(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}))} \\ T_3 &= \frac{1}{n} \sum_{i=0}^{n-1} \frac{(Y_{i+1} - \bar{A}_2(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}, \theta^{(2)}, \sigma))^2}{\Delta_n b^2(Z_i; \sigma)} \\ T_4 &= \frac{1}{n} \sum_{i=0}^{n-1} \log b^2(Z_i; \sigma) \end{aligned}$$

For the term  $T_1$  we have:

$$\begin{aligned} T_1 &= \frac{1}{n\Delta_n^3} \sum_{i=0}^{n-1} \frac{1}{b^2(Z_i; \sigma)} \frac{(X_{i+1} - \bar{A}_1(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}, \theta^{(2)}, \sigma))^2}{(\partial_y a_1(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}))^2} = \\ &= \frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{b^2(Z_i; \sigma)} \frac{(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma) + \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma) - \bar{A}_1(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}, \theta^{(2)}, \sigma))^2}{\Delta_n^3 (\partial_y a_1(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}))^2} = \\ &= \frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{b^2(Z_i; \sigma)} \left[ \frac{(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma))^2}{\Delta_n^3 (\partial_y a_1(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}))^2} \frac{(\partial_y a_1(Z_i; \theta_0^{(1)}))^2}{(\partial_y a_1(Z_i; \theta_0^{(1)}))^2} + \right. \\ &\quad \left. \frac{2\Delta_n (X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma)) (a_1(Z_i; \theta_0^{(1)}) - a_1(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}))}{\Delta_n^3 (\partial_y a_1(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}))^2} + \frac{\Delta_n^2 (a_1(Z_i; \theta_0^{(1)}) - a_1(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}))^2}{\Delta_n^3 b^2(Z_i; \sigma) (\partial_y a_1(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}))^2} \right] \end{aligned}$$



Thanks to the Lemmas 5 and 6, we know that the second term of the sum converges to 0 in probability, and for the first one we have:

$$\frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{b^2(Z_i; \sigma)} \frac{\left( X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma) \right)^2 (\partial_y a_1(Z_i; \theta_0^{(1)}))^2}{\Delta_n^3 (\partial_y a_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}))^2 (\partial_y a_1(Z_i; \theta_0^{(1)}))^2} \xrightarrow{\mathbb{P}_0} \int \frac{b^2(z; \sigma_0)}{b^2(z; \sigma)} \frac{(\partial_y a_1(z; \theta_0^{(1)}))^2}{(\partial_y a_1(z; \hat{\theta}_{n, \Delta_n}^{(1)}))^2} \nu_0(dz)$$

For the third term, we use the assumption (A5), and then obtain the convergence to 0 in probability thanks to Theorem 1, the continuous mapping theorem and Lemma 4:

$$\frac{2}{n} \sum_{i=0}^{n-1} \frac{\Delta_n^2}{\Delta_n^3} \frac{\left( a_1(Z_i; \theta_0^{(1)}) - a_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}) \right)^2}{b^2(Z_i; \sigma) (\partial_y a_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}))^2} = \frac{2}{n^2} \sum_{i=0}^{n-1} \frac{\left( \sqrt{\frac{n}{\Delta_n}} (\theta_0^{(1)} - \hat{\theta}_{n, \Delta_n}^{(1)}) \right)^2 g^2(X_i)}{b^2(Z_i; \sigma) (\partial_y a_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}))^2} \xrightarrow{\mathbb{P}_0} 0.$$

Then,  $T_2$  decomposes as:

$$\begin{aligned} & \frac{1}{n} \sum_{i=0}^{n-1} \frac{(X_{i+1} - \bar{A}_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta^{(2)}, \sigma))(Y_{i+1} - \bar{A}_2(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta^{(2)}, \sigma))}{\Delta_n^2 b^2(Z_i; \sigma) (\partial_y a_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}))} = \\ & \frac{1}{n} \sum_{i=0}^{n-1} \frac{(\partial_y a_1(Z_i; \theta_0^{(1)}))}{b^2(Z_i; \sigma) (\partial_y a_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}))} \left[ \frac{(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma))(Y_{i+1} - \bar{A}_2(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta_0^{(2)}, \sigma))}{\Delta_n^2 (\partial_y a_1(Z_i; \theta_0^{(1)}))} + \right. \\ & \quad \frac{\Delta_n (X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma))(a_2(Z_i; \theta_0^{(2)}) - a_2(Z_i; \theta^{(2)}))}{\Delta_n^2 (\partial_y a_1(Z_i; \theta_0^{(1)}))} + \\ & \quad \frac{\Delta_n (Y_{i+1} - \bar{A}_2(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta_0^{(2)}, \sigma))(a_1(Z_i; \theta_0^{(1)}) - a_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}))}{\Delta_n^2 (\partial_y a_1(Z_i; \theta_0^{(1)}))} + \\ & \quad \left. \frac{\Delta_n^2 (a_1(Z_i; \theta_0^{(1)}) - a_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}))(a_2(Z_i; \theta_0^{(2)}) - a_2(Z_i; \theta^{(2)}))}{\Delta_n^2 (\partial_y a_1(Z_i; \theta_0^{(1)}))} \right] \end{aligned}$$

Again, using Lemma 6, we know that the second and the third terms are converging to 0 in probability. For the first term, thanks to Lemma 5 we have the following convergence:

$$\frac{1}{n} \sum_{i=0}^{n-1} \frac{(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma))(Y_{i+1} - \bar{A}_2(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta_0^{(2)}, \sigma)) (\partial_y a_1(Z_i; \theta_0^{(1)}))}{\Delta_n^2 b^2(Z_i; \sigma) (\partial_y a_1(Z_i; \theta_0^{(1)})) (\partial_y a_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}))} \xrightarrow{\mathbb{P}_0} \int \frac{b^2(z; \sigma_0)}{b^2(z; \sigma)} \frac{\partial_y a_1(z; \theta_0^{(1)})}{\partial_y a_1(z; \hat{\theta}_{n, \Delta_n}^{(1)})} \nu_0(dz)$$

Finally, we treat the last term:

$$\frac{1}{n} \sum_{i=0}^{n-1} \frac{\Delta_n^2 (a_1(Z_i; \theta_0^{(1)}) - a_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)})) (a_2(Z_i; \theta_0^{(2)}) - a_2(Z_i; \theta^{(2)}))}{\Delta_n^2 b^2(Z_i; \sigma) (\partial_y a_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}))}$$

Using again the Lipschitz continuity of  $a_1$ , Theorem 1 and the Slutsky's theorem, we obtain a convergence to zero in probability for this term.  $T_4$  converges in probability to  $\int \log b^2(z; \sigma) \nu_0(dz)$  due to Lemma 4. Consider  $T_3$ :

$$\begin{aligned} T_3 &= \frac{1}{n \Delta_n} \sum_{i=0}^{n-1} \frac{1}{b^2(Z_i; \sigma)} \left[ (Y_{i+1} - \bar{A}_2(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta_0^{(2)}, \sigma))^2 + \right. \\ &\quad \left. 2(Y_{i+1} - \bar{A}_2(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta_0^{(2)}, \sigma)) (\bar{A}_2(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta_0^{(2)}, \sigma) - \bar{A}_2(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta^{(2)}, \sigma)) + \right. \\ &\quad \left. (\bar{A}_2(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta_0^{(2)}, \sigma) - \bar{A}_2(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta^{(2)}, \sigma))^2 \right] = \frac{1}{n \Delta_n} \sum_{i=0}^{n-1} \frac{(Y_{i+1} - \bar{A}_2(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta_0^{(2)}, \sigma))^2}{b^2(Z_i; \sigma)} + \\ &\quad 2 \frac{\Delta_n}{n \Delta_n} \sum_{i=0}^{n-1} \frac{(Y_{i+1} - \bar{A}_2(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta_0^{(2)}, \sigma)) (a_2(Z_i; \theta_0^{(2)}) - a_2(Z_i; \theta^{(2)}))}{b^2(Z_i; \sigma)} + \\ &\quad \frac{\Delta_n^2}{n \Delta_n} \sum_{i=0}^{n-1} \frac{(a_2(Z_i; \theta_0^{(2)}) - a_2(Z_i; \theta^{(2)}))^2}{b^2(Z_i; \sigma)} \end{aligned}$$

Thanks to Lemma 5 and 6 we conclude that

$$T_3 \xrightarrow{\mathbb{P}_0} \int \frac{b^2(z; \sigma_0)}{b^2(z; \sigma)} \nu_0(dz) + 0 + 0$$

Finally, we obtain

$$\begin{aligned} &\frac{1}{n} \mathcal{L}_{n, \Delta_n}(\theta, \sigma^2; Z_{0:n}) \xrightarrow{\mathbb{P}_0} \\ &\int \left( \frac{b^2(z; \sigma_0)}{b^2(z; \sigma)} \left[ 3 \left( \frac{\partial_y a_1(z; \theta_0^{(1)})}{\partial_y a_1(z; \hat{\theta}_{n, \Delta_n}^{(1)})} \right)^2 - 3 \frac{\partial_y a_1(z; \theta_0^{(1)})}{\partial_y a_1(z; \hat{\theta}_{n, \Delta_n}^{(1)})} + 1 \right] + \log b^2(z; \sigma) \right) \nu_0(dz) \end{aligned}$$

By assumption (A5)  $\partial_y a_1(\cdot)$  does not depend on  $\theta^{(1)}$ , thus  $\frac{\partial_y a_1(z; \theta_0^{(1)})}{\partial_y a_1(z; \hat{\theta}_{n, \Delta_n}^{(1)})} = 1$ . It gives the Lemma.  $\square$

**Theorem 2. Consistency.** The consistency of the estimator for the parameter  $\theta^{(2)}$  is based on Lemma 2, with the arguments analogous to the proof of Theorem 1. For the diffusion parameter  $\sigma$ , the result follows from Lemma 3. Denote  $\mathcal{I}(\sigma, \sigma_0) := \frac{b^2(z; \sigma_0)}{b^2(z; \sigma)} + \log b^2(z; \sigma)$ . We can choose some

subsequence  $n_k$  such that  $\hat{\sigma}_{n,\Delta_n}$  converges to some  $\sigma_\infty$ . By the definition of the estimator we know that  $\mathcal{I}(\sigma_\infty, \sigma_0) \leq \mathcal{I}(\sigma_0, \sigma_0)$ . But we also know that  $\frac{b^2(z; \sigma_0)}{b^2(z; \sigma)} + \log b^2(z; \sigma) \geq 1 + \log b^2(z; \sigma_0)$  and thus  $\mathcal{I}(\sigma_\infty, \sigma_0) \geq \mathcal{I}(\sigma_0, \sigma_0)$ , and by the identifiability assumption  $\sigma_\infty \equiv \sigma_0$ . It proves the consistency of  $\hat{\sigma}$ .

**Asymptotic normality.** The proof follows the standard pattern. Throughout the proof we assume that  $\theta^{(2)}$  and  $\sigma \in \mathbb{R}$  in order to simplify the notations. We write the Taylor expansion of the contrast function defined in (17) and apply an appropriate scaling

$$\int C_{n,\Delta_n}(\varphi_0 + u(\hat{\varphi}_{n,\Delta_n} - \varphi_0); z) du E_{n,\Delta_n} = -D_{n,\Delta_n}(\varphi_0),$$

where by  $\varphi$  we now denote  $(\theta^{(2)}, \sigma)$  and the parameter  $\theta^{(1)}$  is fixed to its estimate  $\hat{\theta}_{n,\Delta_n}^{(1)}$  throughout the proof, and

$$C_{n,\Delta_n}(\theta) := \begin{bmatrix} \frac{1}{n\Delta_n} \frac{\partial^2}{\partial \theta^{(2)} \partial \theta^{(2)}} \mathcal{L}_{n,\Delta_n}(\hat{\theta}_{n,\Delta_n}^{(1)}, \theta^{(2)}, \sigma; Z_{0:n}) & \frac{1}{n\sqrt{\Delta_n}} \frac{\partial^2}{\partial \sigma \partial \theta^{(2)}} \mathcal{L}_{n,\Delta_n}(\hat{\theta}_{n,\Delta_n}^{(1)}, \theta^{(2)}, \sigma; Z_{0:n}) \\ \frac{1}{n\sqrt{\Delta_n}} \frac{\partial^2}{\partial \theta^{(2)} \partial \sigma} \mathcal{L}_{n,\Delta_n}(\hat{\theta}_{n,\Delta_n}^{(1)}, \theta^{(2)}, \sigma; Z_{0:n}) & \frac{1}{n} \frac{\partial^2}{\partial \sigma \partial \sigma} \mathcal{L}_{n,\Delta_n}(\hat{\theta}_{n,\Delta_n}^{(1)}, \theta^{(2)}, \sigma; Z_{0:n}) \end{bmatrix},$$

$$E_{n,\Delta_n} := \begin{bmatrix} \sqrt{n\Delta_n}(\hat{\theta}_n^{(2)} - \theta_0^{(2)}) \\ \sqrt{n}(\hat{\sigma}_n - \sigma_0) \end{bmatrix}, \quad D_{n,\Delta_n} = \begin{bmatrix} \frac{1}{\sqrt{n\Delta_n}} \frac{\partial}{\partial \theta^{(2)}} \mathcal{L}_{n,\Delta_n}(\hat{\theta}_{n,\Delta_n}^{(1)}, \theta^{(2)}, \sigma; Z_{0:n}) \\ \frac{1}{\sqrt{n}} \frac{\partial}{\partial \sigma} \mathcal{L}_{n,\Delta_n}(\hat{\theta}_{n,\Delta_n}^{(1)}, \theta^{(2)}, \sigma; Z_{0:n}) \end{bmatrix}.$$

First, we compute the higher-order terms of the partial derivatives of first and second order with respect to  $\theta^{(2)}$  and  $\sigma$ :

$$\frac{\partial}{\partial \theta^{(2)}} \mathcal{L}_{n,\Delta_n}(\cdot) = \sum_{i=1}^{n-1} \left[ -6 \frac{\Delta_n(\partial_{\theta^{(2)}} a_2)(X_{i+1} - \bar{A}_1(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}, \theta^{(2)}, \sigma))}{\Delta_n^2 b^2(Z_i; \sigma)(\partial_{\hat{\theta}_{n,\Delta_n}^{(1)}} a_1)} + \right. \\ \left. 2 \frac{\Delta_n(\partial_{\theta^{(2)}} a_2)(Y_{i+1} - \bar{A}_2(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}, \theta^{(2)}, \sigma))}{\Delta_n b^2(Z_i; \sigma)} \right] =: D_{n,\Delta_n}^1$$

$$\begin{aligned}
\frac{\partial}{\partial \sigma} \mathcal{L}_{n, \Delta_n}(\cdot) &= - \sum_{i=1}^{n-1} \frac{\partial_{\sigma} b}{b^3(Z_i; \sigma)} \left[ 6 \frac{(X_{i+1} - \bar{A}_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta^{(2)}, \sigma))^2}{\Delta_n^3 (\partial_{\hat{\theta}_{n, \Delta_n}^{(1)}} a_1)^2} - \right. \\
&\quad \left. 6 \frac{(X_{i+1} - \bar{A}_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta^{(2)}, \sigma))(Y_{i+1} - \bar{A}_2(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta^{(2)}, \sigma))}{\Delta_n^2 (\partial_{\hat{\theta}_{n, \Delta_n}^{(1)}} a_1)} + \right. \\
&\quad \left. 2 \frac{(Y_{i+1} - \bar{A}_2(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta^{(2)}, \sigma))^2}{\Delta_n} \right] + \frac{\partial_{\sigma} b}{b(Z_i; \sigma)} =: D_{n, \Delta_n}^2 \\
\frac{\partial^2}{\partial \theta^{(2)} \partial \theta^{(2)}} \mathcal{L}_{n, \Delta_n}(\cdot) &= \sum_{i=1}^{n-1} \left[ -6 \frac{\Delta_n (\partial_{\theta^{(2)} \theta^{(2)}}^2 a_2) (X_{i+1} - \bar{A}_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta^{(2)}, \sigma))}{\Delta_n^2 b^2(Z_i; \sigma) (\partial_{\theta^{(1)}} a_1)} + \right. \\
&\quad \left. 2 \frac{\Delta_n (\partial_{\theta^{(2)} \theta^{(2)}}^2 a_2) (Y_{i+1} - \bar{A}_2(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta^{(2)}, \sigma))}{\Delta_n b^2(Z_i; \sigma)} + \frac{\Delta_n^2 (\partial_{\theta^{(2)}} a_2)^2}{\Delta_n b^2(Z_i; \sigma)} \right] =: C_{n, \Delta_n}^{11} \\
\frac{\partial^2}{\partial \theta^{(2)} \partial \sigma} \mathcal{L}_{n, \Delta_n}(\cdot) &= \sum_{i=1}^{n-1} \frac{\partial_{\sigma} b}{b^2(Z_i; \sigma)} \left[ 12 \frac{\Delta_n (\partial_{\theta^{(2)}} a_2) (X_{i+1} - \bar{A}_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta^{(2)}, \sigma))}{\Delta_n^2 b(Z_i; \sigma) (\partial_{\theta^{(1)}} a_1)} + \right. \\
&\quad \left. 4 \frac{\Delta_n (\partial_{\theta^{(2)}} a_2) (Y_{i+1} - \bar{A}_2(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta^{(2)}, \sigma))}{\Delta_n b(Z_i; \sigma)} \right] =: C_{n, \Delta_n}^{12} = C_{n, \Delta_n}^{21} \\
\frac{\partial^2}{\partial \sigma^2} \mathcal{L}_{n, \Delta_n}(\cdot) &= - \sum_{i=1}^{n-1} \frac{6(\partial_{\sigma} b)^2 - 2b(Z_i; \sigma)(\partial_{\sigma\sigma}^2 b)}{b^4(Z_i; \sigma)} \left[ 6 \frac{(X_{i+1} - \bar{A}_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta^{(2)}, \sigma))^2}{\Delta_n^3 (\partial_{\theta^{(1)}} a_1)^2} - \right. \\
&\quad \left. 6 \frac{(X_{i+1} - \bar{A}_1(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta^{(2)}, \sigma))(Y_{i+1} - \bar{A}_2(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta^{(2)}, \sigma))}{\Delta_n^2 (\partial_{\theta^{(1)}} a_1)} + 2 \frac{(Y_{i+1} - \bar{A}_2(Z_i; \hat{\theta}_{n, \Delta_n}^{(1)}, \theta^{(2)}, \sigma))^2}{\Delta_n} \right] + \\
&\quad 2 \frac{b(Z_i; \sigma)(\partial_{\sigma\sigma}^2 b) - (\partial_{\sigma} b)^2}{b^2(Z_i; \sigma)} =: C_{n, \Delta_n}^{22}
\end{aligned}$$

We start with proving the convergence for the terms  $C_{n, \Delta_n}$ . Then we can obtain a convergence

in probability after few technical steps. We start with  $C_{n,\Delta_n}^{11}$ :

$$\begin{aligned} \frac{1}{n\Delta_n} C_{n,\Delta_n}^{11} &= \frac{1}{n\Delta_n} \sum_{i=1}^{n-1} \left[ -6 \frac{\Delta_n (\partial_{\theta^{(2)}\theta^{(2)}}^2 a_2)(X_{i+1} - \bar{A}_1(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}, \theta^{(2)}, \sigma))}{\Delta_n^2 b^2(Z_i; \sigma) (\partial_{\theta^{(1)}} a_1)} + \right. \\ &\quad \left. 2 \frac{\Delta_n (\partial_{\theta^{(2)}\theta^{(2)}}^2 a_2)(Y_{i+1} - \bar{A}_2(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}, \theta^{(2)}, \sigma))}{\Delta_n b^2(Z_i; \sigma)} + \frac{\Delta_n^2 (\partial_{\theta^{(2)}} a_2)^2}{\Delta_n b^2(Z_i; \sigma)} \right] = \\ \frac{1}{n\Delta_n} \sum_{i=1}^{n-1} &\left[ -6 \frac{\Delta_n (\partial_{\theta^{(2)}\theta^{(2)}}^2 a_2)(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma))}{\Delta_n^2 b^2(Z_i; \sigma) (\partial_{\theta^{(1)}} a_1)} - 6 \frac{\Delta_n^2 (\partial_{\theta^{(2)}\theta^{(2)}}^2 a_2)(a_1(Z_i; \theta_0^{(1)}) - a_1(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}))}{\Delta_n^2 b^2(Z_i; \sigma) (\partial_{\theta^{(1)}} a_1)} \right. \\ &\quad \left. 2 \frac{\Delta_n (\partial_{\theta^{(2)}\theta^{(2)}}^2 a_2)(Y_{i+1} - \bar{A}_2(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}, \theta^{(2)}, \sigma))}{\Delta_n b^2(Z_i; \sigma)} + \frac{\Delta_n^2 (\partial_{\theta^{(2)}} a_2)^2}{\Delta_n b^2(Z_i; \sigma)} \right] \end{aligned}$$

Note that thanks to Lemma 6 we know that

$$\begin{aligned} \frac{1}{n\Delta_n} \sum_{i=1}^{n-1} &\left[ -6 \frac{\Delta_n (\partial_{\theta^{(2)}\theta^{(2)}}^2 a_2)(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma))}{\Delta_n^2 b^2(Z_i; \sigma) (\partial_{\theta^{(1)}} a_1)} + \right. \\ &\quad \left. 2 \frac{\Delta_n (\partial_{\theta^{(2)}\theta^{(2)}}^2 a_2)(Y_{i+1} - \bar{A}_2(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}, \theta^{(2)}, \sigma))}{\Delta_n b^2(Z_i; \sigma)} + \frac{\Delta_n^2 (\partial_{\theta^{(2)}} a_2)^2}{\Delta_n b^2(Z_i; \sigma)} \right] \xrightarrow{\mathbb{P}_0} \int \frac{(\partial_{\theta^{(2)}} a_2)^2}{b^2(z; \sigma)} \nu_0(dz) \end{aligned}$$

What about the remaining term, thanks to the assumption (A5) we have:

$$\begin{aligned} - \frac{6}{n\Delta_n} \sum_{i=1}^{n-1} &\frac{(\partial_{\theta^{(2)}\theta^{(2)}}^2 a_2)(a_1(Z_i; \theta_0^{(1)}) - a_1(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}))}{b^2(Z_i; \sigma) (\partial_{\theta^{(1)}} a_1)} = \\ &- \frac{6}{\sqrt{n\Delta_n}} \frac{1}{n} \sum_{i=1}^{n-1} \frac{(\partial_{\theta^{(2)}\theta^{(2)}}^2 a_2) a_1(Z_i; \sqrt{\frac{n}{\Delta_n}} (\theta_0^{(1)} - \hat{\theta}_{n,\Delta_n}^{(1)}))}{b^2(Z_i; \sigma) (\partial_{\theta^{(1)}} a_1)} \end{aligned}$$

We know that  $(\theta_0^{(1)} - \hat{\theta}_{n,\Delta_n}^{(1)}) \sqrt{\frac{n}{\Delta_n}}$  is normally distributed by Theorem 1, and  $\frac{1}{n} \sum_{i=1}^{n-1} \frac{(\partial_{\theta^{(2)}\theta^{(2)}}^2 a_2)}{b^2(Z_i; \sigma) (\partial_{\theta^{(1)}} a_1)}$  converges to its invariant density by Lemma 4. Then by Slutsky's and the continuous mapping theorem the product also converges in distribution to a normal variable, which is, divided by  $\sqrt{n\Delta_n}$  converges to zero since  $n\Delta_n \rightarrow \infty$  by design. However, as  $n\Delta_n \rightarrow \infty$ , this term converges to 0 in probability. As a result,

$$\frac{1}{n\Delta_n} C_{n,\Delta_n}^{11} \xrightarrow{\mathbb{P}_0} \int \frac{(\partial_{\theta^{(2)}} a_2)^2}{b^2(z; \sigma)} \nu_0(dz)$$

With the same arguments we prove that  $\frac{1}{n\sqrt{\Delta_n}}C_{n,\Delta_n}^{12} = \frac{1}{n\sqrt{\Delta_n}}C_{n,\Delta_n}^{21} \xrightarrow{\mathbb{P}_0} 0$  and that

$$\frac{1}{n}C_{n,\Delta_n}^{22} \xrightarrow{\mathbb{P}_0} -4 \int \frac{(\partial_\sigma b)^2}{b^2(z; \sigma_0)} \nu_0(dz)$$

Then we consider the remaining term, recalling the assumption (A5): We start with the term

$$\begin{aligned} \frac{1}{\sqrt{n\Delta_n}}D_{n,\Delta_n}^1 &= \frac{1}{\sqrt{n\Delta_n}} \sum_{i=1}^{n-1} \left[ -6 \frac{\Delta_n(\partial_{\theta^{(2)}}a_2)(X_{i+1} - \bar{A}_1(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}, \theta^{(2)}, \sigma))}{\Delta_n^2 b^2(Z_i; \sigma)(\partial_{\theta^{(1)}}a_1)} + \right. \\ &2 \frac{\Delta_n(\partial_{\theta^{(2)}}a_2)(Y_{i+1} - \bar{A}_2(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}, \theta^{(2)}, \sigma))}{\Delta_n b^2(Z_i; \sigma)} \left. \right] = \frac{1}{\sqrt{n\Delta_n}} \sum_{i=1}^{n-1} \left[ -6 \frac{(\partial_{\theta^{(2)}}a_2)(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma))}{\Delta_n b^2(Z_i; \sigma)(\partial_{\theta^{(1)}}a_1)} - \right. \\ &\left. 6 \frac{(\partial_{\theta^{(2)}}a_2)a_1(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)} - \theta_0^{(1)})}{b^2(Z_i; \sigma)(\partial_{\theta^{(1)}}a_1)} + 2 \frac{(\partial_{\theta^{(2)}}a_2)(Y_{i+1} - \bar{A}_2(Z_i; \hat{\theta}_{n,\Delta_n}^{(1)}, \theta^{(2)}, \sigma))}{b^2(Z_i; \sigma)} \right] \end{aligned}$$

For the first and the third term we simply apply Lemma 7 and obtain convergence in distribution to  $\mathcal{N}\left(0, \nu_0\left(\frac{(\partial_{\theta^{(2)}}a_2)^2}{b^2(z; \sigma_0)}\right)\right)$ . For the second term we apply the result of Theorem 1, as well as the continuous mapping and Slutsky's theorem we may state that:

$$-6 \sum_{i=1}^{n-1} \frac{(\partial_{\theta^{(2)}}a_2)a_1(Z_i; \sqrt{\frac{n}{\Delta_n}}(\hat{\theta}_{n,\Delta_n}^{(1)} - \theta_0^{(1)}))}{b^2(Z_i; \sigma)(\partial_{\theta^{(1)}}a_1)} \xrightarrow{\mathcal{D}} -6 \int \frac{(\partial_{\theta^{(2)}}a_2)}{b^2(z; \sigma)(\partial_{\theta^{(1)}}a_1)} a_1(z; \tilde{\eta}) \nu_0(dz),$$

where  $\tilde{\eta}$  is distributed as stated in Theorem 1. Then as  $n \rightarrow 0$ ,

$$-\frac{6}{n} \sum_{i=1}^{n-1} \frac{(\partial_{\theta^{(2)}}a_2)a_1(Z_i; \sqrt{\frac{n}{\Delta_n}}(\hat{\theta}_{n,\Delta_n}^{(1)} - \theta_0^{(1)}))}{b^2(Z_i; \sigma)(\partial_{\hat{\theta}_{n,\Delta_n}^{(1)}}a_1)} \xrightarrow{\mathbb{P}_0} 0$$

By analogy, we prove the convergence for the term  $D_{n,\Delta_n}^2$ , obtaining:

$$\frac{1}{\sqrt{n}}D_{n,\Delta_n}^2 \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, 32\nu_0\left(\frac{(\partial_\sigma b)^2}{b^2(z; \sigma_0)}\right)\right)$$

That gives the result. □

## 7.4 Consistency and asymptotic normality of the least squares contrast

*Theorem 3.* The proof will follow the one of the classical contrast. First, we define the following quantities:

$$\begin{aligned}\mathcal{L}_{n,\Delta_n}^{(1),LSE}(\theta^{(1)}, \theta^{(2)}, \sigma; Z_{0:n}) &= \frac{1}{n} \sum_{i=0}^{n-1} \frac{(X_{i+1} - \bar{A}_1(Z_i; \theta^{(1)}, \theta^{(2)}, \sigma))^2}{\Delta_n^3} \\ \mathcal{L}_{n,\Delta_n}^{(2),LSE}(\theta^{(1)}, \theta^{(2)}, \sigma; Z_{0:n}) &= \frac{1}{n} \sum_{i=0}^{n-1} \frac{(Y_{i+1} - \bar{A}_2(Z_i; \theta^{(1)}, \theta^{(2)}, \sigma))^2}{\Delta_n}\end{aligned}$$

**Consistency of  $\hat{\theta}^{(1)}$ .** First, consider:

$$\begin{aligned}\Delta_n \left[ \mathcal{L}_{n,\Delta_n}^{LSE}(\theta^{(1)}, \theta^{(2)}; Z_{0:n}) - \mathcal{L}_{n,\Delta_n}^{LSE}(\theta_0^{(1)}, \theta^{(2)}; Z_{0:n}) \right] &= \\ \frac{\Delta_n}{n\Delta_n^3} \sum_{i=0}^{n-1} \left[ (X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma) + \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma) - \right. \\ &\quad \left. \bar{A}_1(Z_i; \theta^{(1)}, \theta^{(2)}, \sigma))^2 - (X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma))^2 \right] = \\ \frac{\Delta_n^2}{n\Delta_n^3} \sum_{i=0}^{n-1} \left[ 2(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma))(a_1(Z_i; \theta_0^{(1)}) - a_1(Z_i; \theta^{(1)})) + \right. \\ &\quad \left. \Delta_n(a_1(Z_i; \theta_0^{(1)}) - a_1(Z_i; \theta^{(1)}))^2 + O(\Delta_n^2) \right]\end{aligned}$$

Then we have from Lemmas 5, 6:

$$\begin{aligned}\frac{2}{n\Delta_n} \sum_{i=0}^{n-1} (X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}, \theta^{(2)}, \sigma))(a_1(Z_i; \theta_0^{(1)}) - a_1(Z_i; \theta^{(1)})) &\xrightarrow{\mathbb{P}_0} 0 \\ \frac{1}{n} \sum_{i=0}^{n-1} (a_1(Z_i; \theta_0^{(1)}) - a_1(Z_i; \theta^{(1)}))^2 &\xrightarrow{\mathbb{P}_0} \int (a_1(z; \theta_0^{(1)}) - a_1(z; \theta^{(1)}))^2 \nu_0(dz)\end{aligned}$$

We conclude that there exists a subsequence  $\hat{\theta}_{n,\Delta_n}^{(1)} = \arg \min_{\theta} \mathcal{L}_{n,\Delta_n}^{LSE}(\theta; Z_{0:n})$  that tends to  $\theta_\infty$ . Since the minimum is attained at the point  $\theta_0$  and from (A4), we conclude that  $\theta_\infty = \theta_0$ . Hence the estimator is consistent.

**Consistency of  $\hat{\theta}^{(2)}$ .** Consider:

$$\begin{aligned} & \frac{1}{\Delta_n} \left[ \mathcal{L}_{n,\Delta_n}^{LSE}(\theta^{(1)}, \theta^{(2)}; Z_{0:n}) - \mathcal{L}_{n,\Delta_n}^{LSE}(\theta_0^{(1)}, \theta_0^{(2)}; Z_{0:n}) \right] = \\ & \left[ \frac{1}{n\Delta_n^2} \sum_{i=0}^{n-1} (Y_{i+1} - \bar{A}_2(Z_i; \theta^{(1)}, \theta_0^{(2)}, \sigma) + \bar{A}_2(Z_i; \theta^{(1)}, \theta_0^{(2)}, \sigma) - \bar{A}_2(Z_i; \theta^{(1)}, \theta^{(2)}, \sigma))^2 - \right. \\ & \left. (Y_{i+1} - \bar{A}_2(Z_i; \theta^{(1)}, \theta_0^{(2)}, \sigma))^2 \right] = \frac{\Delta_n}{n\Delta_n^2} \sum_{i=0}^{n-1} \left[ 2(Y_{i+1} - \bar{A}_2(Z_i; \theta^{(1)}, \theta_0^{(2)}, \sigma))(a_2(Z_i; \theta_0^{(2)}) - a_2(Z_i; \theta^{(2)})) + \right. \\ & \left. \Delta_n(a_2(Z_i; \theta_0^{(2)}) - a_2(Z_i; \theta^{(2)}))^2 + O(\Delta_n^2) \right] \end{aligned}$$

Thanks to Lemmas 4, 5:

$$\begin{aligned} & \frac{2\Delta_n}{n\Delta_n^2} \sum_{i=0}^{n-1} (Y_{i+1} - \bar{A}_2(Z_i; \theta^{(1)}, \theta_0^{(2)}, \sigma))(a_2(Z_i; \theta_0^{(2)}) - a_2(Z_i; \theta^{(2)})) \xrightarrow{\mathbb{P}_0} 0 \\ & \frac{\Delta_n^2}{n\Delta_n^2} \sum_{i=0}^{n-1} (a_2(Z_i; \theta_0^{(2)}) - a_2(Z_i; \theta^{(2)}))^2 \xrightarrow{\mathbb{P}_0} \int (a_2(z; \theta_0^{(2)}) - a_2(z; \theta^{(2)}))^2 \nu_0(dz) \end{aligned}$$

The consistency is concluded following the same arguments as in the case of  $\theta^{(1)}$ .

**Asymptotic normality.** We apply again a Taylor formula for a function (22):

$$\int C_n(\theta_0 + u(\hat{\theta}_n - \theta_0)) du E_n = D_n(\theta_0),$$

where we define

$$\begin{aligned} C_n(\theta) &:= \begin{bmatrix} \frac{\Delta_n}{n} \frac{\partial^2}{\partial \theta^{(1)} \partial \theta^{(1)}} \mathcal{L}_{n,\Delta_n}^{LSE}(\theta; Z_{0:n}) & \frac{1}{n} \frac{\partial^2}{\partial \theta^{(1)} \partial \theta^{(2)}} \mathcal{L}_{n,\Delta_n}^{LSE}(\theta; Z_{0:n}) \\ \frac{1}{n} \frac{\partial^2}{\partial \theta^{(1)} \partial \theta^{(2)}} \mathcal{L}_{n,\Delta_n}^{LSE}(\theta; Z_{0:n}) & \frac{1}{n\Delta_n} \frac{\partial^2}{\partial \theta^{(2)} \partial \theta^{(2)}} \mathcal{L}_{n,\Delta_n}^{LSE}(\theta; Z_{0:n}) \end{bmatrix}, \\ E_n &:= \begin{bmatrix} \sqrt{\frac{n}{\Delta_n}} (\hat{\theta}_{n,\Delta_n}^{(1)} - \theta_0^{(1)}) \\ \sqrt{n\Delta_n} (\hat{\theta}_n^{(2)} - \theta_0^{(2)}) \end{bmatrix}, \quad D_n(\theta) = \begin{bmatrix} \frac{\sqrt{\Delta_n}}{n} \frac{\partial}{\partial \theta^{(1)}} \mathcal{L}_{n,\Delta_n}^{LSE}(\theta; Z_{0:n}) \\ \frac{1}{n\sqrt{\Delta_n}} \frac{\partial}{\partial \theta^{(2)}} \mathcal{L}_{n,\Delta_n}^{LSE}(\theta; Z_{0:n}) \end{bmatrix}. \end{aligned}$$

Using Lemma 7 we get:

$$D_n(\theta_0) \xrightarrow{\mathcal{D}} -2\mathcal{N} \left( 0, I_2 \cdot \begin{bmatrix} \frac{1}{3} \int b^2(z; \sigma_0) (\partial_y a_1(z; \theta_0^{(1)}))^2 (\partial_{\theta^{(1)}} a_1(z; \theta_0^{(1)}))^2 \nu_0(dz) \\ \int b^2(z; \sigma_0) (\partial_{\theta^{(2)}} a_2(z; \theta_0^{(2)}))^2 \nu_0(dz) \end{bmatrix} \right),$$

where  $I_2$  is  $2 \times 2$  identity matrix. And by Lemmas 5, 4 we have the result for  $C_n(\theta)$ :

$$C_n(\theta_0) \xrightarrow{\mathbb{P}_0} -2 \begin{bmatrix} \int (\partial_{\theta^{(1)}} a_1(z; \theta_0^{(1)}))^2 \nu_0(dz) & 0 \\ 0 & \int (\partial_{\theta^{(2)}} a_2(z; \theta_0^{(2)}))^2 \nu_0(dz) \end{bmatrix}.$$

That, in the combination with the consistency result, gives the theorem.  $\square$