



Parametric inference for multidimensional hypoelliptic ergodic diffusion with full observations

Anna Melnykova

► To cite this version:

Anna Melnykova. Parametric inference for multidimensional hypoelliptic ergodic diffusion with full observations. 2019. hal-01704010v2

HAL Id: hal-01704010

<https://hal.science/hal-01704010v2>

Preprint submitted on 11 Jan 2019 (v2), last revised 15 Jul 2020 (v4)

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 multidimensional hypoelliptic ergodic diffusion with full observations

Anna Melnykova *

Abstract

Multidimensional hypoelliptic diffusions arise naturally as models of neuronal activity. Estimation in those models is complex because of the degenerate structure of the diffusion coefficient. We build a consistent estimator of the drift and variance parameters with the help of a discretized log-likelihood of the continuous process when discrete time observations of both coordinates are available on an interval $T = N\Delta$, with Δ the time step between the observations. We discuss the difficulties generated by the hypoellipticity and provide a proof of the consistency and the asymptotic normality of the estimator in the asymptotic setting $T \rightarrow \infty$ as $\Delta \rightarrow 0$. We test our approach numerically on the hypoelliptic FitzHugh-Nagumo model, which describes the firing mechanism of a neuron.

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

1 Introduction

Hypoelliptic diffusions naturally occur in various applications, most notably in neuroscience, molecular physics and in mathematical finance. In particular, models of neuronal activity — either on the scale of one single neuron (Höpfner et al., 2016, Leon and Samson, 2017), or on a scale of 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 a hypoelliptic diffusion.

* Anna Melnykova [†]

The main difference to the classical (or elliptic) setting is that in the hypoelliptic case the dimensionality of the noise is lower than the dimensionality of the system of stochastic differential equations (SDE), which describes the process. Hypoellipticity can be intuitively explained in the following way: though the covariance matrix of noise is singular due to a degenerate diffusion coefficient, smooth transition density with respect to the Lebesgue measure still exists. That is the case when the noise is propagated to all the coordinates through the drift term.

Properties of hypoelliptic diffusions significantly differ from those of elliptic ones, when all coordinates are driven by a Brownian motion. Thus they are more difficult to study. The first problem is that each coordinate has a variance of different order. It is the main cause why classical numerical approximation methods do not work well with hypoelliptic diffusions. In particular, it is proven that for hypoelliptic systems the classical Euler-Maruyama scheme does not preserve ergodic properties of the true process (Mattingly et al., 2002). The second problem is the degenerate diffusion coefficient. As the explicit form of the transition density is often unknown, parametric inference is usually based on its discrete approximation with the 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.

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, (dW_t) is a standard Brownian motion defined on some probability space $(\Omega, \mathcal{F}_t, \mathbb{P})$, $(\theta^{(1)}, \theta^{(2)}, \sigma)$ 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 random variable.

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 preserving the ergodic property. Then we propose an estimation technique which maximizes the likelihood function of the discrete approximate model in the asymptotic setting $T = N\Delta \rightarrow \infty$ and $\Delta = \Delta \rightarrow 0$ as $N \rightarrow \infty$. Let us discuss the solutions proposed by other authors for hypoelliptic systems of different types.

Several works treat the parametric inference problem for a particular case of system (1). It is natural to introduce first the class of stochastic Damping Hamiltonian systems, also known as Langevin equations (Gardiner and Collett, 1985). These hypoelliptic models arise as a 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 system of SDE with a non-constant drift and a constant variance, its solution can be interval-wise approximated by a system with a linear drift, and the original covariance matrix being expanded by adding higher-order terms. It allows to construct a quasi Maximum Likelihood Estimator. Pokern et al. (2007) attempt to solve the problem of the non-invertibility of the covariance matrix for the particular case of system (2) with a constant variance with the help of Itô-Taylor expansion of the transition density. The parameters are then estimated with a Gibbs sampler based on the discretized model with the noise propagated into the first coordinate with order $\Delta^{\frac{3}{2}}$. 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 the system (2). This method works reasonably well in practice even for more general models when it is possible to convert a system (1) to a simpler form (2). However, the transformation of the observations sampled from the continuous model (1) requires the prior knowledge of the parameters involved in the first equation (see Samson and Thieullen (2012)). The other 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 available continuous-time observations, and then it is used for computing the parameters of the drift term. Resulting estimator is strongly consistent. Few other works are also

devoted to a 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. But the estimation is conducted separately for each coordinate, so it requires partial knowledge of the parameters of the system.

In this paper we propose 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, approximating the transition density and proposing a contrast estimator based on the discretized log-likelihood. While we attempt to estimate 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 the first in the literature. We finish with numerical experiments, testing the proposed approach on the hypoelliptic FitzHugh-Nagumo model and compare it to 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

2.1 Notations

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

$$dZ_t = A(Z_t; \theta)dt + B(Z_t; \sigma)d\tilde{W}_t, \quad Z_0 = \omega_0, \quad t \in [0, T] \quad (3)$$

where $Z_t = (X_t, Y_t)^T$, \tilde{W}_t is a standard two-dimensional Brownian motion defined on the filtered probability space, ω_0 is a \mathcal{F}_0 -measurable 2-dimensional random vector, and $\theta = (\theta^{(1)}, \theta^{(2)})$ is the vector of parameters. Matrices A and B represent, respectively, the drift and the diffusion coefficient, that is $A(Z_t; \theta) = (a_1(X_t, Y_t; \theta^{(1)}), a_2(X_t, Y_t; \theta^{(1)}))^T$ 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 partial derivatives (unless the arguments need to be specified): $\partial_{x_i} f \equiv \frac{\partial f}{\partial x_i}(x_1, \dots, x_p)$, $\partial_{x_i, x_j}^2 \equiv \frac{\partial^2 f}{\partial x_i \partial x_j}(x_1, \dots, x_p) \forall i, j \in [1, p]$. We suppress the dependency on the parameters, when their values are clear from context, otherwise additional indices are introduced. True values of the parameters are denoted by $\theta_0^{(1)}, \theta_0^{(2)}, \sigma_0$, and, further, by P_0 we will denote the probability $P_{\theta_0^{(1)}, \theta_0^{(2)}, \sigma_0}$. We also adopt the notations from Pokern et al. (2007) and refer to the variable Y_t which is directly driven by Gaussian noise as "rough", and to X_t as "smooth".

2.2 Assumptions

Further, we are working under the following assumptions:

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

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

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

where $\|\cdot\|$ is the standard Euclidean norm. Further, denote by Z_0 the initial value of the process Z_t , we assume that $\mathbb{E}\|Z_0\|^2 < \infty$.

A3 Process Z_t is ergodic and there exists a unique invariant probability measure ν_0 with finite moments of any order.

A4 Both functions $a_1(Z_t; \theta^{(1)})$ and $a_2(Z_t; \theta^{(2)})$ are identifiable, that is $a_i(Z_t; \theta^{(i)}) \equiv a_i(Z_t; \theta_0^{(i)}) \Leftrightarrow \theta^{(i)} = \theta_0^{(i)}$.

Assumption (A1) ensures that the weak Hörmander condition is satisfied, thus the system is hypoelliptic in the sense of stochastic calculus of variations

(Nualart, 2006, Malliavin and Thalmaier, 2006). In order to prove it we first write the coefficients of system (3) as two vector fields, converting (3) from 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 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 the system (3), moreover this solution is Feller (Revuz and Yor, 2013). (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 with 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 a 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 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 is ergodic, then its sampling $\{Z_i\}$, $i \in [0, N]$ is also ergodic (Genon-Catalot et al., 2000).

3 Discrete model

This section is organized as follows: first, we introduce the linear model, and explain how the maximum likelihood estimator can be constructed in this case. Then, we propose a discrete scheme for a general model (3), where we use a piece-wise linear approximation of the underlying process, which results in a piece-wise Gaussian density.

3.1 Linear SDE

Let us start with a particular case of (3). We consider a linear homogeneous equation with a drift vector and diffusion matrix given, respectively, by

$$A(Z_t; \theta) = \begin{pmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{pmatrix} \begin{pmatrix} X_t \\ Y_t \end{pmatrix} =: A_\theta Z_t \quad (5)$$

and

$$B(Z_t; \theta) = \begin{pmatrix} 0 & 0 \\ 0 & \sigma \end{pmatrix} =: B_\sigma. \quad (6)$$

The equation is written as

$$dZ_t = A_\theta Z_t dt + B_\sigma d\tilde{W}_t \quad (7)$$

In this case we can also verify an assumption (A4) by checking the stability of the matrix A_θ (Le-Breton and Musiela, 1985). That is, if eigenvalues of A_θ have negative real parts, then A_θ is stable and the solution of equation (7) is ergodic. Furthermore, we know that this solution is a Gaussian process given by

$$Z_t = e^{A_\theta T} Z_0 + \int_0^T e^{A_\theta(T-t)} B_\sigma d\tilde{W}_t \quad (8)$$

with a mean vector and a covariance matrix defined, respectively, as

$$\mathbb{E}[Z_T] = e^{A_\theta T} Z_0 \quad (9)$$

$$\Sigma[Z_T; \theta, \sigma^2] = \mathbb{E} \left[\left(\int_0^T e^{A_\theta(T-t)} B_\sigma d\tilde{W}_t \right) \left(\int_0^T e^{A_\theta(T-t)} B_\sigma d\tilde{W}_t \right)^T \right]. \quad (10)$$

Remark. Note that the matrix (10) is of full rank, while $B_\sigma B_\sigma^T$ is not.

3.2 General SDE and discrete linearization

Now let us explain how the above mentioned results can be applied to a general system of type (3). The idea is to use a piece-wise approximation of the solution of (3) by the solutions of a linear homogeneous equation. It can be considered as an improved Euler scheme, where the solution is piece-wise approximated by the solution of a system with constant coefficients. In that case instead of a constant matrix A_θ we can use the Jacobian of the drift vector $A(Z_t; \theta)$, which we assume to be constant on each small interval of size Δ :

$$\begin{pmatrix} \partial_x a_1(x, y; \theta^{(1)}) & \partial_y a_1(x, y; \theta^{(1)}) \\ \partial_x a_2(x, y; \theta^{(2)}) & \partial_y a_2(x, y; \theta^{(2)}) \end{pmatrix} =: J(z; \theta) \quad (11)$$

Let us start from the diffusion term (10). After passing to a discrete time step, it is written as follows:

$$\Sigma [Z_{i+1}; \theta, \sigma^2 | \mathcal{F}_i] = \mathbb{E} \left[\left(\int_{i\Delta}^{\Delta(i+1)} e^{J(Z_i; \theta)(\Delta-t)} B(Z_i; \sigma) d\tilde{W}_t \right) \left(\int_{i\Delta}^{\Delta(i+1)} e^{J(Z_i; \theta)(\Delta-t)} B(Z_i; \sigma) d\tilde{W}_t \right)^T \right], \quad (12)$$

Representation (12) is not convenient for computational purposes because of matrix exponents. However, it can be approximated as follows, thanks to the discrete time setting with an arbitrary small step Δ :

Proposition 1. *The second-order Taylor approximation of matrix $\Sigma(Z_{i+1}; \theta, \sigma^2 | Z_i)$ defined in (12) has the following form:*

$$b^2(Z_i; \sigma) \begin{pmatrix} (\partial_y a_1)^2 \frac{\Delta^3}{3} & (\partial_y a_1) \frac{\Delta^2}{2} + (\partial_y a_1)(\partial_y a_2) \frac{\Delta^3}{3} \\ (\partial_y a_1) \frac{\Delta^2}{2} + (\partial_y a_1)(\partial_y a_2) \frac{\Delta^3}{3} & \Delta + (\partial_y a_2) \frac{\Delta^2}{2} + (\partial_y a_2)^2 \frac{\Delta^3}{3} \end{pmatrix} + O(\Delta^4), \quad (13)$$

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

In practice it is enough to use only the higher-order terms of (13):

$$\Sigma_\Delta(Z_{i+1}; \theta, \sigma^2 | Z_i) := b^2(Z_i; \sigma) \begin{pmatrix} (\partial_y a_1)^2 \frac{\Delta^3}{3} & (\partial_y a_1) \frac{\Delta^2}{2} \\ (\partial_y a_1) \frac{\Delta^2}{2} & \Delta \end{pmatrix} \quad (14)$$

The corresponding inverse matrix of (14) is defined by:

$$\Sigma_\Delta^{-1}(Z_{i+1}; \theta, \sigma^2 | Z_i) = \frac{1}{b^2(Z_i; \sigma)} \begin{pmatrix} \frac{12}{(\partial_y a_1)^2 \Delta^3} & \frac{6}{(\partial_y a_1) \Delta^2} \\ \frac{6}{(\partial_y a_1) \Delta^2} & \frac{4}{\Delta} \end{pmatrix} \quad (15)$$

It is also natural to approximate the drift term by applying the definition of the matrix exponent to (9):

$$\mathbb{E}[Z_{i+1} | Z_i] = Z_i + \Delta J(Z_i; \theta) Z_i + \frac{\Delta^2}{2} J^2(Z_i; \theta) Z_i + O(\Delta^3) \quad (16)$$

where $J(Z_i; \theta)$ is the Jacobian of the drift vector $A(Z_t; \theta)$ given by (11). However, in order to avoid the accumulation of errors, we recall again that the expression $J(Z_i; \theta) Z_i$ serves only as an approximation of the non-constant

matrix $A(Z_t; \theta)$, so that we can element-wise rewrite the approximation of the drift as:

$$\begin{aligned}\bar{A}_1(Z_i; \theta^{(1)}) &:= X_i + \Delta a_1(X_i, Y_i; \theta^{(1)}) + \\ &\quad \frac{\Delta^2}{2} \left(\partial_x a_1(X_i, Y_i; \theta^{(1)}) + \partial_y a_1(X_i, Y_i; \theta^{(1)}) \right) a_1(X_i, Y_i; \theta^{(1)}) \\ \bar{A}_2(Z_i; \theta^{(2)}) &:= Y_i + \Delta a_2(X_i, Y_i; \theta^{(2)}) + \\ &\quad \frac{\Delta^2}{2} \left(\partial_x a_2(X_i, Y_i; \theta^{(2)}) + \partial_y a_2(X_i, Y_i; \theta^{(2)}) \right) a_2(X_i, Y_i; \theta^{(2)}).\end{aligned}\tag{17}$$

Then the element-wise approximation of Z_{i+1} conditionally on Z_i is written as:

$$\begin{aligned}X_{i+1} &= \bar{A}_1(Z_i; \theta^{(1)}) + b(Z_i; \sigma) \partial_y a_1(Z_i; \theta^{(1)}) \xi_1 \\ Y_{i+1} &= \bar{A}_2(Z_i; \theta^{(2)}) + b(Z_i; \sigma) \xi_2,\end{aligned}\tag{18}$$

where ξ_1 and ξ_2 are normal random variables, such that $\mathbb{E}[\xi_1] = \mathbb{E}[\xi_2] = 0$, $\text{Var}[\xi_1] = \frac{\Delta^3}{3}$, $\text{Var}[\xi_2] = \Delta$ and $\text{Cov}[\xi_1 \xi_2] = \frac{\Delta^2}{2}$. Numerically they can be simulated by decomposing the matrix (14) with the help of, for example, LU or Cholesky decomposition and multiplying it by independent standard normal variables. Method of the decomposition does not affect the theoretical properties of the scheme. About the diffusion term, as an approximation we may take any matrix $\bar{B}(Z_i; \theta, \sigma^2)$ such that $\bar{B} \bar{B}^T = \Sigma(Z_i; \theta, \sigma^2)$. It can be obtained, for example, with Cholesky or LU decomposition. Note that the approximated diffusion term now depends on parameters of the drift term.

Now we want to study the rate of weak convergence of the scheme. We will rely on the result of the following Proposition (recall that the true value of vector of parameters is denoted by θ_0):

Proposition 2 (Weak convergence of the local linearization scheme). *The following holds:*

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

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

$$\mathbb{E}(f(Z_{t+\Delta})|Z_t = z) = \sum_{i=0}^j \frac{\Delta^i}{i!} L^i f(z) + O(\Delta^{j+1}), \quad (19)$$

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

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

where $\nabla_B^2(\cdot) = \sigma^2 \frac{\partial^2}{\partial y^2}(\cdot)$ is a weighted Laplace type operator. Since the process is approximated by (16), it coincides with (19) up to the terms of order Δ^2 . \square

Remark. Note that it is possible to add more terms to expansion (16). However, we will only take terms up to Δ^2 : on the one hand, it allows us to have an approximation precise enough so that Proposition 2 holds. In particular, bounds obtained for the first equation are crucial for proving the consistency of the estimator introduced in Section 4. On the other hand, it keeps expressions simple enough and spares computational cost. Also note that the approximation of X_{i+1} depends only on $\theta^{(1)}$ (for Y_{i+1} — respectively, on $\theta^{(2)}$).

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 some already known results for the linear homogeneous system and propose an analogous, separate estimation of the drift and the variance term.

4.1 Contrast estimator

Let us introduce a contrast function for system (3). In the elliptic case this function is defined as -2 times the log-likelihood of the discretized model (Florens-Zmirou, 1989, Kessler, 1997). In hypoelliptic case, however, we must modify this criterion taking into account the specific structure of the

covariance matrix. Most notably, the contrast is obtained by dividing the first part by 2:

$$\begin{aligned} \mathcal{L}_{N,\Delta}(\theta, \sigma^2; Z_{0:N}) = & \frac{1}{2} \sum_{i=0}^{N-1} (Z_{i+1} - \bar{A}(Z_i; \theta))^T \Sigma_{\Delta}^{-1}(Z_i; \theta, \sigma^2) (Z_{i+1} - \bar{A}(Z_i; \theta)) \\ & + \sum_{i=0}^{N-1} \log \det(\Sigma_{\Delta}(Z_i; \theta, \sigma^2)), \end{aligned} \quad (20)$$

where the inverse matrix Σ_{Δ}^{-1} is given by (15). Then the estimator is defined as:

$$(\hat{\theta}_{N,\Delta}, \hat{\sigma}_{N,\Delta}^2) = \arg \min_{\theta, \sigma^2} \mathcal{L}_{N,\Delta}(\theta, \sigma^2; Z_{0:N}) \quad (21)$$

For system (2) this correction is not proposed nor in Ozaki (1989), nor in Pokern et al. (2007). We justify this bias theoretically in Lemma 3. However, the 2-dimensional criterion (21) is tricky to analyse because of the different orders of variance for the first and the second coordinate. When both equations are driven by the same parameters, the task is simplified as the parameters can be then estimated from a one-dimensional criteria, which involves only one of the two equations.

When it is not the case, the estimation procedure depends heavily on the first function of the drift vector. That means that the estimation of $\theta^{(1)}$ is asymptotically independent from the estimation of $\theta^{(2)}$ and σ , while the inverse is, in general, not true. It is, again, caused by the specific structure of (1) and different order of variance of coordinates. On the contrary, in the case of elliptic or Hamiltonian system (2) estimation of drift and diffusion parameters is independent (see in particular Kessler (1997), Samson and Thieullen (2012) etc.). We begin the study from the following Lemma, on which the consistency of $\hat{\theta}^{(1)}$ crucially relies:

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

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

Proof is postponed to Appendix. On the next step, we obtain the asymptotic normality of (21) with respect to $\theta^{(1)}$:

Theorem 1. *Under the assumptions (A1)-(A4), $\Delta \rightarrow 0$, $N\Delta \rightarrow \infty$ and $N\Delta^2 \rightarrow 0$ the following holds:*

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

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

The idea of the proof of consistency for the diffusion and the rough term parameters follows Gloter and Sørensen (2009). To start with, we introduce an additional assumption on the first function in the drift term, namely, the linearity with respect to the vector of parameters $\theta^{(1)}$:

$$\begin{aligned} a_1(z; \theta_i^{(1)}) - a_1(z; \theta_j^{(1)}) &= a_1(z; \theta_i^{(1)} - \theta_j^{(1)}) \quad \forall \theta_i^{(1)}, \theta_j^{(1)} \in \Theta_1 \\ C a_1(z; \theta^{(1)}) &= a_1(z; C \theta^{(1)}) \quad \forall C \in \mathbb{R} \end{aligned} \quad (\text{H1})$$

Then, we have to rely on the estimated value of the first parameter to study the consistency of $\theta^{(2)}$ and σ . The idea of the proofs is the following: since we are working in a compact set, we can always find a sequence of estimators $\hat{\theta}_{N,\Delta}^{(1)}$ such that a sequence of $(\hat{\theta}_{N,\Delta}^{(1)} - \theta_0^{(1)})$ is tight. Then we use it 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 (see, for example, Kessler (1997)).

Remark. *Assumption (H1) is rather restrictive, but the idea of the proof can be reused for a more general case (for example, under the condition of Lipschitz continuity with respect to parameter $\theta^{(1)}$) at the cost of additional technicalities, which are omitted in this paper.*

Consistency follows from the following Lemmas (proofs are postponed to Appendix):

Lemma 2. *Under assumptions (A1)-(A4), $\Delta \rightarrow 0$ and $N\Delta \rightarrow \infty$ the following holds:*

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

Lemma 3. *Under assumptions (A1)-(A4), (H1) and when $\Delta \rightarrow 0$ while $N\Delta \rightarrow \infty$ the following holds:*

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

Theorem 2. *Under assumptions (A1)-(A4), (H1) and the condition $\Delta \rightarrow 0$, $N\Delta \rightarrow \infty$ and $N\Delta^2 \rightarrow 0$ the following holds:*

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

and

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

4.2 Least squares estimator

We recall that for the linear model (7) the estimators for the unknown constant matrices A_θ and B_σ , given by (5) and (6), are defined Le-Breton and Musiela (1985). It is proposed to compute B_σ by the sample covariance matrix, and for the drift term the maximum likelihood estimator can be constructed, assuming that the term σ is known. Both estimators are defined as follows:

$$\hat{A}_\theta = \left[\int_0^T dZ_t Z_t^T \right] \left[\int_0^T Z_t Z_t^T dt \right]^{-1}, \quad \hat{B}_\sigma = \frac{1}{T} [Z_t]. \quad (22)$$

It is proven in Le-Breton and Musiela (1985) that (22) is strongly consistent when the observed time interval is fixed. However, estimators given by (22) can not be used directly for a system of a general form, since they rely on

the continuous observations of the linear process. On the other hand, we can still adapt the approach of separate estimation of the drift and variance terms with some limitations, which will be discussed further.

Idea of the estimator for the drift term basically comes from the natural bound on the quadratic form (20). We know that for a matrix A the upper bound on its vector product with a vector x written as $x^T A x$ is given by $\|A\| \|x\|^2$. Translating it back to the contrast minimization problem, we may use this bound to estimate the drift parameters independently from the diffusion coefficient by minimizing the norm of the drift vector (which is a quadratic variance). Thus, consider

$$\hat{\theta}_{N,\Delta}^d = \arg \min_{\theta} \sum_{i=0}^{N-1} \|Z_{i+1} - \bar{A}(Z_i; \theta)\|^2 := \arg \min_{\theta} \mathcal{L}_{N,\Delta}^d(\theta; Z_{0:N}). \quad (23)$$

Basically, minimization of this form corresponds to a classical least square estimator. Of course, this estimator does not take into account the specific structure of the noise, so that we may expect it to be less performing. Using the same reasoning as for a linearized contrast we may prove the next Theorem (the proof is postponed to appendix):

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

$$\hat{\theta}_{N,\Delta}^d \xrightarrow{\mathbb{P}_0} \theta_0,$$

and

$$\begin{pmatrix} \sqrt{N\Delta^3} \\ \sqrt{N\Delta} \end{pmatrix}^T (\hat{\theta}_{N,\Delta}^d - \theta_0^d) \xrightarrow{\mathcal{D}} \begin{bmatrix} \mathcal{N}\left(0, \frac{1}{3}C_1^{-1} \int b^2(z; \sigma_0) (\partial_y a_1(z; \theta_0^{(1)})^2 (\partial_{\theta^{(1)}} a_1(z; \theta_0^{(1)})) (\partial_{\theta^{(1)}} a_1(z; \theta_0^{(1)})^T \nu_0(dz)\right) \\ \mathcal{N}\left(0, C_2^{-1} \int b^2(z; \sigma_0) (\partial_{\theta^{(2)}} a_2(z; \theta_0^{(2)}) (\partial_{\theta^{(2)}} a_2(z; \theta_0^{(2)})^T \nu_0(dz)\right) \end{bmatrix},$$

where

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

Remark. *Note the difference in the scaling between least squares ("drift-only") contrast and the results obtained in Theorems 1 and 3. That means that the estimator defined by (23) has a bigger variance than the one defined by (21). It is a price to pay for the asymptotic independence of both parameters.*

What about the diffusion term, in the case when $b(x, y; \sigma) \equiv \sigma f(x, y)$, parameter σ can be computed explicitly with the help of the sample covariance matrix. Good properties of this approach for the elliptic case are proven in Florens-Zmirou (1989). When it comes to the hypoelliptic systems, this approach must be modified, as the discretization step of order Δ does not allow to compute the terms of order Δ^3 , which represent the propagated noise. However, the value of σ can still be inferred from the observations of the rough coordinate by computing

$$\tilde{\sigma}_{\Delta, N}^2 = \frac{1}{N\Delta} \sum_{i=0}^{N-1} \frac{(Y_{i+1} - Y_i)^2}{f(X_i, Y_i)}. \quad (24)$$

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.

5 Simulation study

5.1 The model

The two estimators $(\hat{\theta}_{\Delta, N}, \hat{\sigma}_{\Delta, N}^2)$ and $(\hat{\theta}_{\Delta, N}^d, \tilde{\sigma}_{\Delta, N}^2)$ 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 it was improved by adding two sources of noise to the both coordinates, what results in an elliptic SDE. 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 version with noise only in the second coordinate as studied in Leon and Samson (2017). More precisely, the behaviour of the neuron is defined through the solution of the system

$$\begin{cases} dX_t = \frac{1}{\varepsilon}(X_t - X_t^3 - Y_t - s)dt \\ dY_t = (\gamma X_t - Y_t + \beta)dt + \sigma dW_t, \end{cases} \quad (25)$$

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 channel kinetic.

Parameter s is the magnitude of the stimulus current and is often known in experiments, ε 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 $\theta = (\gamma, \beta, \varepsilon, \sigma)$.

Hypoellipticity and ergodicity of (25) are proven in Leon and Samson (2017). In Jensen et al. (2012) it is proven that s is unidentifiable when only one coordinate is observed. Parametric inference for elliptic FitzHugh-Nagumo model both in fully- and partially observed case is investigated in Jensen (2014). The same problem, but for the hypoelliptic setting is studied in Ditlevsen and Samson (2017).

5.2 Experimental design

In order to make our experiments more representative, 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. Sample trajectories for both settings are shown on Figure 1.

We organize the trials as follows: first, we generate 100 trajectories using formula (17) for each set of parameters with $\Delta = 0.001$ and $N = 500000$. Then we downsample the sequence and work only with each 10-th value of the process, so that $\Delta = 0.01$ and $N = 50000$. We estimate the parameters by minimizing the contrast to the contrast given by (20). We refer to this method as linearized contrast. For the least square estimator (LSE) we do the following: we estimate the parameter σ explicitly from the observations of the second variable by (24), and then compute the parameters of the drift by minimizing (23). 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.

The minimization of the criterions is conducted with the `optim` function in **R** with the Conjugate Gradient method. In Table 1 we present the mean value of the estimated parameters and the standard deviation (in brackets). Figure 2 illustrates the estimation densities. Linearized contrast is depicted in blue, least square estimator —in red, 1.5 scheme in green.

The first thing is that the estimation of the diffusion coefficient σ with the 2-dimensional linearized estimator is biased in both sets of data, even though the contrast is corrected with respect to the hypoellipticity of the system. This bias does not appear in the one-dimensional criteria and when

	γ	β	ε	σ
Set 1:	1.5	0.3	0.1	0.6
Lin. contrast	1.477 (1.056)	0.289 (0.428)	0.100 (0.561)	0.672 (0.291)
LSE	1.460 (1.059)	0.311 (0.403)	0.100 (0.562)	0.611 (0.287)
1.5 scheme	1.497 (1.055)	0.299 (0.393)	0.099 (0.563)	0.597 (0.288)
Set 2:	1.2	1.3	0.1	0.4
Lin. contrast	1.199 (0.531)	1.315 (0.621)	0.102 (0.683)	0.472 (0.340)
LSE	1.170 (0.423)	1.268 (0.598)	0.100 (0.678)	0.400 (0.381)
1.5 scheme	1.221 (0.645)	1.324 (0.777)	0.088 (0.575)	0.398 (0.338)

Table 1: Comparison between different schemes

the value is directly computed from the observations. Thus its origin may be explained by the dimensionality of the system. Parameters of the second coordinate γ and β are estimated efficiently with all three methods, though the 1.5 scheme provides a more accurate estimation. It is expected, since one of the parameters is fixed to its real value. However, in the case of ε , 1-dimensional criteria does not score better than the linearized and least squares based estimators. This parameter seems to be underestimated in the case of 1.5 scheme, and a bit overestimated with the linearization scheme, as well as the diffusion coefficient.

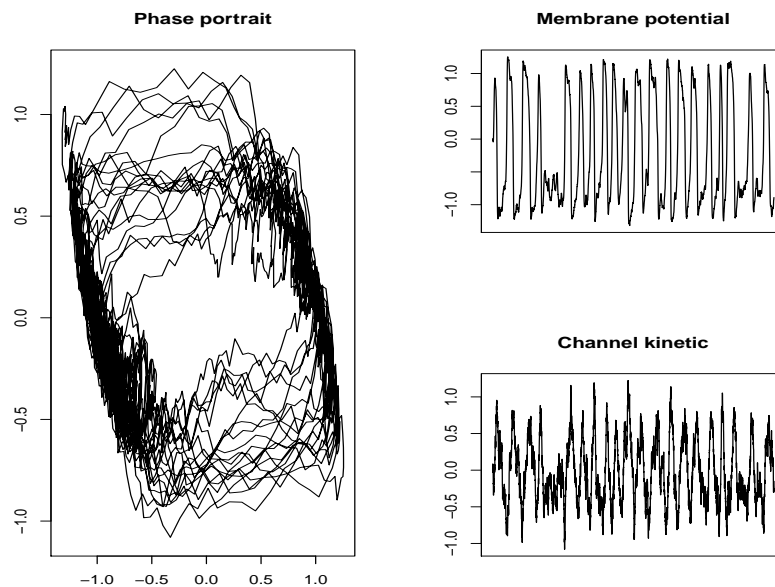
Also note that the experiments illustrate the speed of convergence proven in Theorems 1, 2 and 3. In particular, we may see that the quadratic-variance based contrast gives the same performance as the classical contrast for the parameters of rough coordinate, but parameter ε has a much bigger variance.

During the simulation study it is observed that ε 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.

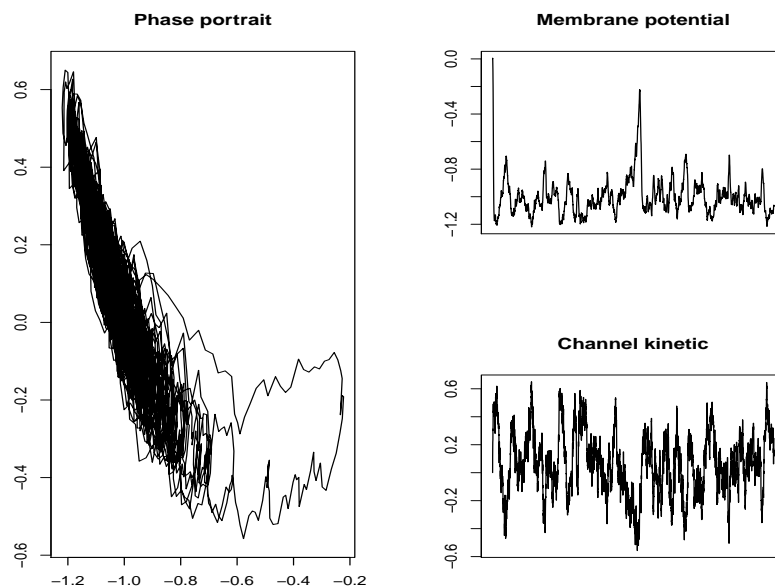
6 Conclusions

The proposed contrast estimator successfully generalizes parametric inference methods developed for models of type (2) for more general class (1). Numerical study shows that it can be used with no prior knowledge of the parameters. It is the most prominent advantage of our method over the analogous works.

From the theoretical point of view, our estimators also reveal good properties. Both linearized contrast and the least square estimators are consis-

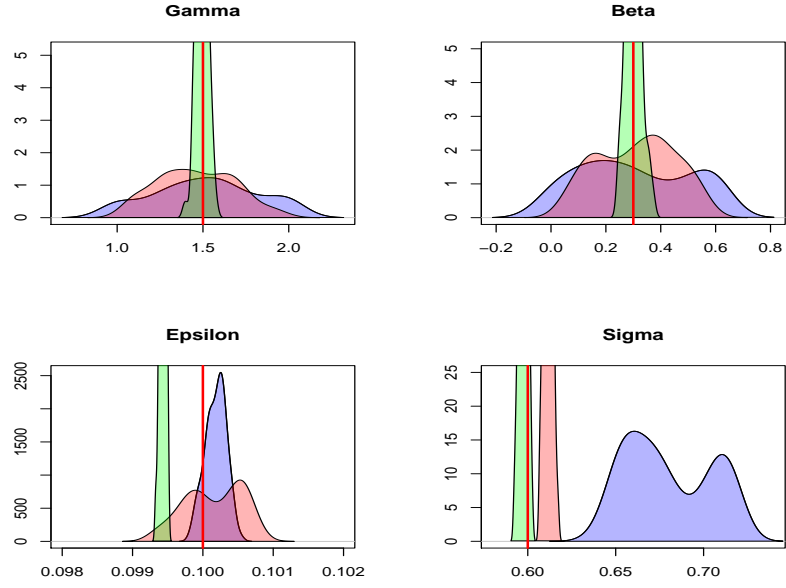


(a) Excitatory set

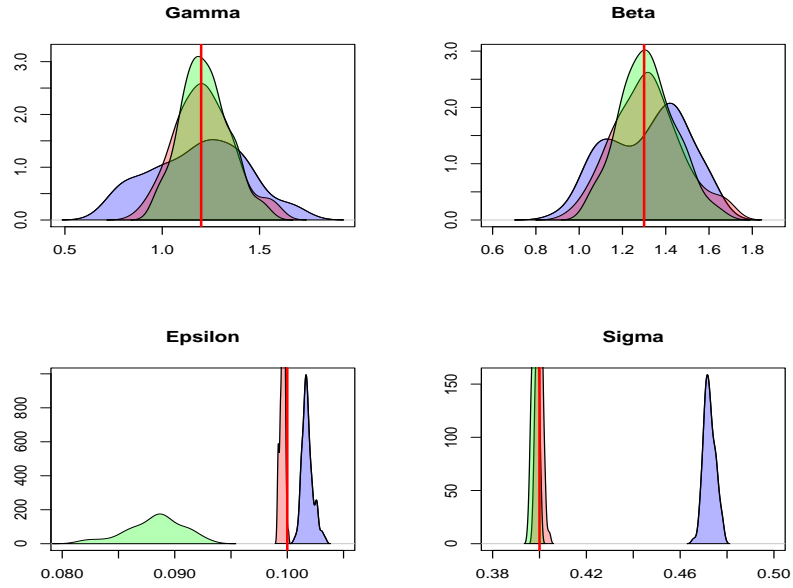


(b) Oscillatory set

Figure 1: Trajectories for two sets of parameters



(a) Excitatory set



(b) Oscillatory set

Figure 2: Estimation density for the linearized contrast (blue), the LSE (red) and 1.5 scheme (green) estimators.

tent. In the case of the contrast, estimator of the rough coordinate asymptotically depends on the estimator of the smooth coordinate, and its performance is sensitive to the form of the drift term. Question of asymptotic normality is even more intricate. We prove the asymptotic normality under rather restrictive assumptions of the drift term even if the method can be applied to more general models. The normality of the LSE is studied under no additional assumptions on the drift term. It is noted that the variance of least square estimator is bigger than the variance of the contrast estimator.

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). It would allow to face real experimental data.

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, and the general rule which describes the behaviour of the contrast in that case is not yet derived. Finally, it is crucial to pair the method with a robust optimization procedure, since, especially in higher dimensions, 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.

References

- Ableidinger, M., Buckwar, E., and Hinterleitner, H. (2017). A stochastic version of the jansen and rit neural mass model: Analysis and numerics. *The Journal of Mathematical Neuroscience*, 7(1):1.
- Arnold, L. and Kliemann, W. (1987). On unique ergodicity for degenerate diffusions. *Stochastics: an international journal of probability and stochastic processes*, 21(1):41–61.

- Cattiaux, P., León, J. R., and Prieur, C. (2014). Estimation for stochastic damping hamiltonian systems under partial observation. ii drift term. *ALEA*, 11(1):p–359.
- Cattiaux, P., León, J. R., Prieur, C., et al. (2016). Estimation for stochastic damping hamiltonian systems under partial observation. iii. diffusion term. *The Annals of Applied Probability*, 26(3):1581–1619.
- Ditlevsen, S. and Löcherbach, E. (2017). Multi-class oscillating systems of interacting neurons. *SPA*, 127:1840–1869.
- Ditlevsen, S. and Samson, A. (2017). Hypoelliptic diffusions: discretization, filtering and inference from complete and partial observations. *submitted*.
- Fitzhugh, R. (1961). Impulses and physiological states in theoretical models of nerve membrane. *Biophysical Journal*, 1(6):445–466.
- Florens-Zmirou, D. (1989). Approximate discrete-time schemes for statistics of diffusion processes. *Statistics: A Journal of Theoretical and Applied Statistics*, 20(4):547–557.
- Gardiner, C. W. and Collett, M. J. (1985). Input and output in damped quantum systems: Quantum stochastic differential equations and the master equation. *Phys. Rev. A*, 31:3761–3774.
- Genon-Catalot, V. and Jacod, J. (1993). 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.
- Genon-Catalot, V., Jeantheau, T., and Laredo, C. (1999). Parameter estimation for discretely observed stochastic volatility models. *Bernoulli*, 5(5):855–872.
- Genon-Catalot, V., Jeantheau, T., and Larédo, C. (2000). Stochastic volatility models as hidden markov models and statistical applications. *bernoulli* 6 1051–1079. *Mathematical Reviews (MathSciNet)*, 10:3318471.
- Gloter, A. and Sørensen, M. (2009). Estimation for stochastic differential equations with a small diffusion coefficient. *Stochastic Processes and their Applications*, 119(3):679–699.
- Hall, P. and Heyde, C. (1980). *Martingale limit theory and its application*. Probability and mathematical statistics. Academic Press.

- Hodgkin, A. L. and Huxley, A. F. (1952). A quantitative description of membrane currents and its application to conduction and excitation in nerve. *Journal of Physiology-London*, 117(4):500–544.
- Höpfner, R., Löcherbach, E., Thieullen, M., et al. (2016). 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.
- Jensen, A. C. (2014). *Statistical Inference for Partially Observed Diffusion Processes*. Phd thesis, University of Copenhagen.
- Jensen, A. C., Ditlevsen, S., Kessler, M., and Papaspiliopoulos, O. (2012). Markov chain monte carlo approach to parameter estimation in the fitzhugh-nagumo model. *Physical Review E*, 86(4):041114.
- Karatzas, I. and Shreve, S. E. (1987). *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics. Springer, 1 edition.
- Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scandinavian Journal of Statistics*, 24(2):211–229.
- Kloeden, P. E., Platen, E., and Schurz, H. (2003). *Numerical solution of SDE through computer experiments*. Universitext. Springer.
- Le-Breton, A. and Musiela, M. (1985). Some parameter estimation problems for hypoelliptic homogeneous gaussian diffusions. *Banach Center Publications*, 16(1):337–356.
- Leon, J. R. and Samson, A. (2017). Hypoelliptic stochastic FitzHugh-Nagumo neuronal model: mixing, up-crossing and estimation of the spike rate. *Annals of Applied Probability*. to appear.
- Malliavin, P. and Thalmaier, A. (2006). *Stochastic calculus of variations in mathematical finance*. Springer finance. Springer, 1 edition.
- Mattingly, J. C., Stuart, A. M., and Higham, D. J. (2002). Ergodicity for sdes and approximations: locally lipschitz vector fields and degenerate noise. *Stochastic processes and their applications*, 101(2):185–232.
- Nualart, D. (2006). *Malliavin Calculus and Related Topics*. Springer. New York.
- Ozaki, T. (1989). Statistical identification of nonlinear random vibration systems. *Journal of Applied Mechanics*, 56:186–191.

- Pokern, Y., Stuart, A. M., and Wiberg, P. (2007). Parameter estimation for partially observed hypoelliptic diffusions. *J. Roy. Stat. Soc.*, 71(1):49–73.
- Revuz, D. and Yor, M. (2013). *Continuous martingales and Brownian motion*, volume 293. Springer Science & Business Media.
- Roynette, B. (1975). Sur les processus de diffusion de dimension 2. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 32(1-2):95–110.
- Samson, A. and Thieullen, M. (2012). Contrast estimator for completely or partially observed hypoelliptic diffusion. *Stochastic Processes and their Applications*, 122:2521–2552.
- Van der Pol, B. (1920). A theory of the amplitude of free and forced triode vibrations. *Radio Review*, 1(1920):701–710.
- Wu, L. (2001). Large and moderate deviations and exponential convergence for stochastic damping hamiltonian systems. *Stochastic processes and their applications*, 91(2):205–238.

7 Appendix

7.1 Properties of the scheme

Proposition 1. Let us consider each integral of (13) separately. Denote:

$$\mathcal{W}_{t+\Delta} = \int_t^{t+\Delta} e^{J(t+\Delta-s)} B(Z_t; \sigma) d\tilde{W}_s,$$

where we suppress the dependency of the Jacobian of the starting point on the interval in order to keep notations simple. Recalling the Jacobian of system (3) and the definition of the matrix exponent, we have:

$$\begin{aligned} \mathcal{W}_{t+\Delta} &= \int_t^{t+\Delta} (I + J(t + \Delta - s) + O(\Delta^2)) B(Z_t; \sigma) d\tilde{W}_s = \\ &= \int_t^{t+\Delta} \left[\begin{pmatrix} 1 + \partial_x a_1(t + \Delta - s) & \partial_y a_1(t + \Delta - s) \\ \partial_x a_2(t + \Delta - s) & 1 + \partial_y a_2(t + \Delta - s) \end{pmatrix} + O(\Delta^2) \right] \begin{pmatrix} 0 \\ 1 \end{pmatrix} b(Z_t; \sigma) dW_s \\ &= b(Z_t; \sigma) \left[\begin{aligned} &\partial_y a_1 \int_t^{t+\Delta} (t + \Delta - s) dW_s + O(\Delta^2) \\ &\int_t^{t+\Delta} dW_s + \partial_y a_2 \int_t^{t+\Delta} (t + \Delta - s) dW_s + O(\Delta^2) \end{aligned} \right] \end{aligned}$$

Then we can calculate $\mathbb{E} [\mathcal{W}_{t+\Delta} \mathcal{W}'_{t+\Delta}]$:

$$\mathbb{E} [\mathcal{W}_{t+\Delta} \mathcal{W}'_{t+\Delta}] = b^2(Z_t; \sigma) \mathbb{E} \begin{pmatrix} \Sigma_{\Delta}^{(1)} & \Sigma_{\Delta}^{(12)} \\ \Sigma_{\Delta}^{(12)} & \Sigma_{\Delta}^{(2)} \end{pmatrix} + O(\Delta^4),$$

where entries are given by:

$$\begin{aligned} \Sigma_{\Delta}^{(1)} &= (\partial_y a_1)^2 \left[\int_t^{t+\Delta} (t + \Delta - s) dW_s \right]^2 \\ \Sigma_{\Delta}^{(12)} &= \left(\partial_y a_1 \int_t^{t+\Delta} (t + \Delta - s) dW_s \right) \left(\int_t^{t+\Delta} dW_s + \partial_y a_2 \int_t^{t+\Delta} (t + \Delta - s) dW_s \right) \\ \Sigma_{\Delta}^{(2)} &= \left(\int_t^{t+\Delta} dW_s + \partial_y a_2 \int_t^{t+\Delta} (t + \Delta - s) dW_s \right)^2 \end{aligned}$$

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

$$\begin{aligned} \mathbb{E}[\Sigma_{\Delta}^{(1)}] &= (\partial_y a_1)^2 \mathbb{E} \left[\int_t^{t+\Delta} (t + \Delta - s) dW_s \right]^2 = \\ &= (\partial_y a_1)^2 \int_t^{t+\Delta} (t + \Delta - s) ds = (\partial_y a_1)^2 \frac{\Delta^3}{3} \end{aligned}$$

Now consider the product of two stochastic integrals in the remaining terms. Assume for simplicity that $t = 0$. From the properties of the stochastic integrals (Karatzas and Shreve, 1987), it is straightforward to see that:

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

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 $\Delta \rightarrow 0$ and $N\Delta \rightarrow \infty$, let $f \in \mathbb{R} \times \Theta \rightarrow \mathbb{R}$ be such that f is differentiable with respect to z and θ , with derivatives of polynomial growth in z uniformly in θ . 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. However, as its proof is based only on ergodicity of the process and the assumptions analogous to ours, and not on the discretization scheme or dimensionality, we take it for granted without giving a formal generalization for a multi-dimensional case. Then proposition 2 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 θ . Assume $\Delta \rightarrow 0$ and $N\Delta \rightarrow \infty$. Then:*

1. $\frac{1}{N\Delta^3} \sum_{i=0}^{N-1} \frac{f(Z_i; \theta)}{(\partial_y a_1(Z_i; \theta_0^{(1)}))^2} \left(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}) \right)^2 \xrightarrow{\mathbb{P}_0} \frac{1}{3} \int f(z; \theta) b^2(z; \sigma_0) \nu_0(dz)$
2. $\frac{1}{N\Delta} \sum_{i=0}^{N-1} f(Z_i; \theta) \left(Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(2)}) \right)^2 \xrightarrow{\mathbb{P}_0} \int f(z; \theta) b^2(z; \sigma_0) \nu_0(dz)$
3. $\frac{1}{N\Delta^2} \sum_{i=0}^{N-1} \frac{f(Z_i; \theta)}{(\partial_y a_1(Z_i; \theta_0^{(1)}))} \left(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}) \right) \left(Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(2)}) \right) \xrightarrow{\mathbb{P}_0} \frac{1}{2} \int f(z; \theta) b^2(z; \sigma_0) \nu_0(dz)$

Proof. Let us denote:

$$\begin{aligned} \zeta_i^{(1)} &= \frac{1}{N\Delta^3} \frac{f(Z_i; \theta)}{(\partial_y a_1(Z_i; \theta_0^{(1)}))^2} \left(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}) \right)^2 \\ \zeta_i^{(2)} &= \frac{1}{N\Delta} f(Z_i; \theta) \left(Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(2)}) \right)^2 \\ \zeta_i^{(1,2)} &= \frac{1}{N\Delta^2} \frac{f(Z_i; \theta)}{\partial_y a_1(Z_i; \theta_0^{(1)})} \left(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}) \right) \left(Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(2)}) \right) \end{aligned}$$

Thanks to Proposition 2 we know that:

$$\mathbb{E} \left[\zeta_i^{(1)} | \mathcal{F}_i \right] = \frac{1}{3N} f(Z_i; \theta) b^2(Z_i; \sigma_0) + O(\Delta).$$

Then from Lemma 4 it follows that for $N \rightarrow \infty$ uniformly in θ :

$$\sum_{i=0}^{N-1} \mathbb{E} \left[\zeta_i^{(1)} | \mathcal{F}_i \right] \xrightarrow{\mathbb{P}_0} \frac{1}{3} \int f(z; \theta) b^2(z; \sigma_0) \nu_0(dz).$$

The same applies for terms $\zeta_i^{(2)}$ and $\zeta_i^{(1,2)}$. \square

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 θ . Assume $\Delta \rightarrow 0$ and $N\Delta \rightarrow \infty$. Then the following convergence results hold:*

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

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

uniformly in θ .

Proof. Consider (i). Expectation of the sum tends to zero for $\Delta \rightarrow 0$ and $N\Delta \rightarrow \infty$ due to Proposition 2. 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. \square

Remark. *Note that in our experiment design from the fact that*

$$\frac{1}{N\Delta^2} \sum_{i=0}^{N-1} f(Z_i; \theta) (X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)})) \xrightarrow{\mathbb{P}_0} 0 \text{ it follows that}$$

$$\frac{1}{N\Delta} \sum_{i=0}^{N-1} f(Z_i; \theta) (X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)})) \xrightarrow{\mathbb{P}_0} 0, \text{ and also } \frac{1}{N} \sum_{i=0}^{N-1} f(Z_i; \theta) (X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)})) \xrightarrow{\mathbb{P}_0} 0, \text{ since } \Delta \rightarrow 0.$$

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

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

$$(i) \quad \frac{1}{\sqrt{N\Delta^3}} \sum_{i=0}^{N-1} f(Z_i; \theta) (X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)})) \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, \frac{1}{3} \nu_0 \left(b^2(z; \sigma_0) (\partial_y a_1(z; \theta_0^{(1)}))^2 f^2(z; \theta) \right) \right)$$

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

*Proof. **Proof of (i).*** Recall that the expression for X_{i+1} conditionally on X_i can be written as (18). That means, we can rewrite the left part of the equation as:

$$\begin{aligned}
& \frac{1}{\sqrt{N}\Delta^3} \sum_{i=0}^{N-1} f(Z_i; \theta) (X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)})) = \\
& \frac{1}{\sqrt{N}\Delta^3} \sum_{i=0}^{N-1} f(Z_i; \theta) b(Z_i; \sigma_0) \partial_y a_{1, \theta_0} \xi_1 + \frac{1}{\sqrt{N}\Delta^3} \sum_{i=0}^{N-1} f(Z_i; \theta) \delta_i,
\end{aligned}$$

where δ_i is an error term, such that $\mathbb{E}[\delta_i | \mathcal{F}_i] = O(\Delta^3)$ and $\mathbb{E}[\delta_i^2 | \mathcal{F}_i] = O(\Delta^4)$ (see Proposition 2). Recall that ξ_1 is normally distributed with $\mathbb{E}[\xi_1] = 0$ and $\mathbb{E}[\xi_1^2] = \frac{\Delta^3}{3}$, then $\mathbb{E}[f(Z_i; \theta) b(Z_i; \sigma_0) \partial_y a_{1, \theta_0} \xi_1 | \mathcal{F}_i] = 0$ and $\mathbb{E}[(f(Z_i; \theta) b(Z_i; \sigma_0) \partial_y a_{1, \theta_0} \xi_1)^2 | \mathcal{F}_i] = \frac{\Delta^3}{3} (f(Z_i; \theta) b(Z_i; \sigma_0) \partial_y a_{1, \theta_0})^2$. As a consequence,

$$\frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} \mathbb{E} [(f(Z_i; \theta) b(Z_i; \sigma_0) \partial_y a_{1, \theta_0} \xi_1)^2 | \mathcal{F}_i] \rightarrow \frac{1}{3} \nu_0 \left((f(z; \theta) b(z; \sigma_0) \partial_y a_{1, \theta_0})^2 \right).$$

Furthermore, since the function f is bounded, that means that $\frac{1}{N} \sum_{i=0}^{N-1} \mathbb{E} [(f(Z_i; \theta) b(Z_i; \sigma_0) \partial_y a_{1, \theta_0} \xi_1)^4 | \mathcal{F}_i]$ is also bounded and converges to 0 pointwise. Then, the Theorem 3.2 in Hall and Heyde (1980) implies that:

$$\frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} f(Z_i; \theta) b(Z_i; \sigma_0) \partial_y a_{1, \theta_0} \xi_1 \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, \frac{1}{3} \nu_0 \left(b^2(z; \sigma_0) (\partial_y a_1(z; \theta_0^{(1)}))^2 f^2(z; \theta) \right) \right)$$

It remains to consider the error term $\frac{1}{\sqrt{N\Delta^3}} \sum_{i=0}^{N-1} f(Z_i; \theta) \delta_i$. First, note that $\frac{1}{\sqrt{N\Delta^3}} \sum_{i=0}^{N-1} \mathbb{E} [f(Z_i; \theta) \delta_i | \mathcal{F}_i] = \sqrt{N} O(\sqrt{\Delta^3})$. It converges to 0 since by our assumption that $N\Delta^2$ (and, as a consequence, $N\Delta^3$) tends to 0. Also, $\frac{1}{N\Delta^3} \sum_{i=0}^{N-1} \mathbb{E} [(f(Z_i; \theta) \delta_i)^2 | \mathcal{F}_i] = O(\Delta)$. It completes the proof for (i).

Proof of (ii). We again rewrite the approximation in the same way as we did in the preceding part:

$$\begin{aligned} & \frac{1}{\sqrt{N\Delta^3}} \sum_{i=0}^{N-1} f(Z_i; \theta) (X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}))^2 - \\ & \quad \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} f(Z_i; \theta) \frac{1}{3} b^2(Z_i; \sigma_0) (\partial_y a_1(Z_i; \theta_0^{(1)}))^2 = \\ & \quad \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} f(Z_i; \theta) b^2(Z_i; \sigma_0) (\partial_y a_1(Z_i; \theta_0^{(1)}))^2 \left(\frac{1}{\Delta^3} \xi_1^2 - \frac{1}{3} \right) + \\ & \quad \frac{2}{\sqrt{N\Delta^3}} \sum_{i=0}^{N-1} f(Z_i; \theta) b(Z_i; \sigma_0) (\partial_y a_1(Z_i; \theta_0^{(1)})) \delta_i \xi_1 + \\ & \quad \frac{1}{\sqrt{N\Delta^3}} \sum_{i=0}^{N-1} f(Z_i; \theta) (\partial_y a_1(Z_i; \theta_0^{(1)})) \delta_i^2 \end{aligned}$$

It is easy to see that the expectation of the last two terms of this expression (and its square) converge to 0 under condition that $N\Delta^2 \rightarrow 0$, since $\mathbb{E}[(\xi_1 \delta_i)^2 | \mathcal{F}_i] = O(\Delta_N^7)$, and $\mathbb{E}[(\delta_i)^2 | \mathcal{F}_i] = O(\Delta^4)$.

Simple computation shows that $\mathbb{E} \left[\frac{1}{\Delta^3} \xi_1^2 | \mathcal{F}_i \right] = 0$, $\mathbb{E} \left[\left(\frac{1}{\Delta^3} \xi_1^2 - \frac{1}{3} \right)^2 | \mathcal{F}_i \right] = \frac{2}{9}$.

Then

$$\frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} \mathbb{E} \left[\left(b^2(Z_i, \sigma_0) (\partial_y a_1(Z_i; \theta_0^{(1)}))^2 \left(\frac{1}{\Delta^3} \xi_1^2 - \frac{1}{3} \right) \right)^2 | \mathcal{F}_i \right] \rightarrow \frac{2}{9} \nu_0 \left(f^2(z; \theta) b^4(z; \sigma_0) (\partial_y a_1(z; \theta_0^{(1)}))^4 \right)$$

Then, since $\mathbb{E} \left[\left(b^2(Z_i, \sigma_0) (\partial_y a_1(Z_i; \theta_0^{(1)}))^2 \left(\frac{1}{\Delta^3} \xi_1^2 - \frac{1}{3} \right) \right)^2 | \mathcal{F}_i \right]$ is bounded, we may state that $\frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} \mathbb{E} \left[\left(b^2(Z_i, \sigma_0) (\partial_y a_1(Z_i; \theta_0^{(1)}))^2 \left(\frac{1}{\Delta^3} \xi_1^2 - \frac{1}{3} \right) \right)^2 | \mathcal{F}_i \right] \rightarrow 0$ and apply again the Theorem 3.2 in Hall and Heyde (1980), which results in statement (ii).

Proof of (iii-iv). The proofs are again based on representation (18) and are absolutely analogous to the proof of (i-ii). The main difference is that in the rough coordinate the noise term (of order $\sqrt{\Delta}$) dominates the drift term (or order Δ), so that as $\Delta \rightarrow 0$, the increments $Y_{i+1} - Y_i$ follow the same distribution as $Y_{i+1} - Y_i - \Delta a_2(Z_i; \theta_0^{(2)})$ up to higher-order terms.

Proof of (v). To simplify the proof for the cross-term, we recall that the representation (18) 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):

$$X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}) = b(Z_i; \sigma_0) (\partial_y a_1) \left(\frac{\Delta^{\frac{3}{2}}}{\sqrt{12}} \eta_1 + \frac{\Delta^{\frac{3}{2}}}{2} \eta_2 \right) + \delta_i^1$$

$$Y_{i+1} - Y_i = \Delta a_2(Z_i; \theta_0^{(2)}) + b(Z_i; \sigma_0) \Delta^{\frac{1}{2}} \eta_2 + \delta_i^2,$$

where δ^1 and δ^2 are error terms such that $\mathbb{E}[\delta_i^k | \mathcal{F}_i] = O(\Delta^3)$ and $\mathbb{E}[(\delta_i^k)^2 | \mathcal{F}_i] = O(\Delta^4)$ (see Proposition 2), and η_1 and η_2 are standard independent normal variables.

It is easy to see that the matrix of noise used for this representation multiplied by its transpose is equivalent to (14), so that the approximation is equivalent to (18). It is possible to repeat the proof for any type of approximation, but we will restrict ourselves to one specific case. Then we

obtain:

$$\begin{aligned}
& \frac{1}{\sqrt{N}\Delta^2} \sum_{i=0}^{N-1} f(Z_i; \theta) \left(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}) \right) (Y_{i+1} - Y_i) - \\
& \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} f(Z_i; \theta) \frac{1}{2} b(Z_i; \sigma_0) (\partial_y a_1(Z_i; \theta^{(1)})) = \frac{1}{\sqrt{N}\Delta^2} \sum_{i=0}^{N-1} f(Z_i; \theta) \cdot \\
& \left(b(Z_i; \sigma_0) (\partial_y a_1) \left(\frac{\Delta^{\frac{3}{2}}}{\sqrt{12}} \eta_1 + \frac{\Delta^{\frac{3}{2}}}{2} \eta_2 \right) + \delta_i^1 \right) \left(\Delta a_2(Z_i; \theta^{(2)}) + b(Z_i; \sigma_0) \Delta^{\frac{1}{2}} \eta_2 + \delta_i^2 \right) - \\
& \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} f(Z_i; \theta) \frac{1}{2} b(Z_i; \sigma_0) (\partial_y a_1(Z_i; \theta^{(1)}))
\end{aligned}$$

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

$$\begin{aligned}
\mathbb{E} \left[\left(\left(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}) \right) (Y_{i+1} - Y_i) - \frac{\Delta^2}{2} b^2(Z_i; \sigma_0) (\partial_y a_1) \right)^2 | \mathcal{F}_i \right] = \\
\frac{4\Delta^4}{3} b^4(Z_i; \sigma_0) (\partial_y a_1)^2 + O(\Delta^5)
\end{aligned}$$

Since $\frac{\Delta}{N} \rightarrow 0$ by design we see that

$$\begin{aligned}
& \frac{1}{\sqrt{N}\Delta^2} \sum_{i=0}^{N-1} f(Z_i; \theta) \left(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}) \right) (Y_{i+1} - Y_i) - \\
& \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} f(Z_i; \theta) \frac{1}{2} b(Z_i; \sigma_0) (\partial_y a_1(Z_i; \theta_0^{(1)})) \rightarrow \\
& \frac{4}{3} \nu_0 \left(b^4(z; \sigma_0) (\partial_y a_1(z; \theta_0^{(1)}))^2 \right)
\end{aligned}$$

Further, since

$$\mathbb{E} \left[f^4(Z_i; \theta) \left(\left(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}) \right) (Y_{i+1} - Y_i) - \frac{\Delta^2}{2} b^2(Z_i; \sigma_0) (\partial_y a_1) \right)^4 | \mathcal{F}_i \right]$$

is bounded, and consequently

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

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 convergence in distribution for the increments of the second coordinate hold without any assumption on the parameters of the function $a_2(z; \theta)$. 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 $\sum_{i=0}^{N-1} f(Z_i; \theta) (Y_{i+1} - \bar{A}_2(Z_i; \theta^{(2)}))$ holds, with a proper scaling, for any value of $\theta^{(2)}$.

7.3 Consistency and asymptotic normality of the linearized contrast estimator

Lemma 1. Consider

$$\frac{\Delta}{N} \left[\mathcal{L}_{N,\Delta}(\theta^{(1)}, \theta^{(2)}, \sigma^2; Z_{0:N}) - \mathcal{L}_{N,\Delta}(\theta_0^{(1)}, \theta_0^{(2)}, \sigma_0^2; Z_{0:N}) \right] = T_1 + T_2 + T_3 + T_4,$$

where the terms are given as follows:

$$\begin{aligned} T_1 &= \frac{6\Delta}{N\Delta^3} \sum_{i=0}^{N-1} \left[\frac{(X_{i+1} - \bar{A}_1(Z_i; \theta^{(1)}))^2}{b^2(Z_i; \sigma) (\partial_y a_1(Z_i; \theta^{(1)}))^2} - \frac{(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}))^2}{b^2(Z_i; \sigma_0) (\partial_y a_1(Z_i; \theta_0^{(1)}))^2} \right] \\ T_2 &= -\frac{6\Delta}{N\Delta^2} \sum_{i=0}^{N-1} \left[\frac{(X_{i+1} - \bar{A}_1(Z_i; \theta^{(1)})) (Y_{i+1} - \bar{A}_2(Z_i; \theta^{(2)}))}{\partial_y a_1(Z_i; \theta^{(1)}) b^2(Z_i; \sigma)} \right. \\ &\quad \left. - \frac{(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)})) (Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(2)}))}{\partial_y a_1(Z_i; \theta_0^{(1)}) b^2(Z_i; \sigma_0)} \right] \\ T_3 &= \frac{2\Delta}{N\Delta} \sum_{i=0}^{N-1} \left[\frac{(Y_{i+1} - \bar{A}_2(Z_i; \theta^{(2)}))^2}{b^2(Z_i; \sigma)} - \frac{(Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(2)}))^2}{b^2(Z_i; \sigma_0)} \right] \\ T_4 &= \frac{\Delta}{N} \sum_{i=0}^{N-1} \log \left(\frac{\partial_y a_1(Z_i; \theta^{(1)})}{\partial_y a_1(Z_i; \theta_0^{(1)})} \right) \end{aligned}$$

Consider term T_1 :

$$\begin{aligned}
T_1 &= \frac{6\Delta}{N\Delta^3} \sum_{i=0}^{N-1} \left[\frac{\left(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}) + \bar{A}_1(Z_i; \theta_0^{(1)}) - \bar{A}_1(Z_i; \theta^{(1)}) \right)^2}{b^2(Z_i; \sigma) (\partial_y a_1(Z_i; \theta^{(1)}))^2} - \right. \\
&\quad \left. \frac{\left(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}) \right)^2}{b^2(Z_i; \sigma_0) (\partial_y a_1(Z_i; \theta_0^{(1)}))^2} \right] = \frac{6\Delta}{N\Delta^3} \sum_{i=0}^{N-1} \left[\left(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}) \right)^2 \left[\frac{1}{b^2(Z_i; \sigma) (\partial_y a_1(Z_i; \theta^{(1)}))^2} - \right. \right. \\
&\quad \left. \left. \frac{1}{b^2(Z_i; \sigma_0) (\partial_y a_1(Z_i; \theta_0^{(1)}))^2} \right] + \frac{\Delta}{b^2(Z_i; \sigma) (\partial_y a_1(Z_i; \theta^{(1)}))^2} \left(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}) \right) \right. \\
&\quad \left. \left(a_1(Z_i; \theta_0^{(1)}) - a_1(Z_i; \theta^{(1)}) \right) + \frac{\Delta^2}{b^2(Z_i; \sigma) (\partial_y a_1(Z_i; \theta^{(1)}))^2} \left(a_1(Z_i; \theta_0^{(1)}) - a_1(Z_i; \theta^{(1)}) \right)^2 \right].
\end{aligned}$$

Recalling Lemmas 4, 6 and 5 we have that:

$$\begin{aligned}
&\frac{6}{N\Delta^2} \sum_{i=0}^{N-1} \left(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}) \right)^2 \left[\frac{1}{b^2(Z_i; \sigma) (\partial_y a_1(Z_i; \theta^{(1)}))^2} - \frac{1}{b^2(Z_i; \sigma_0) (\partial_y a_1(Z_i; \theta_0^{(1)}))^2} \right] \xrightarrow{\mathbb{P}_0} 0 \\
&\frac{6}{N\Delta} \sum_{i=0}^{N-1} \frac{1}{b^2(Z_i; \sigma) (\partial_y a_1(Z_i; \theta^{(1)}))^2} \left(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}) \right) (a_1(Z_i; \theta_0^{(1)}) - a_1(Z_i; \theta^{(1)})) \xrightarrow{\mathbb{P}_0} 0 \\
&\frac{6}{N} \sum_{i=0}^{N-1} \frac{(a_1(Z_i; \theta_0^{(1)}) - a_1(Z_i; \theta^{(1)}))^2}{b^2(Z_i; \sigma) (\partial_y a_1(Z_i; \theta^{(1)}))^2} \xrightarrow{\mathbb{P}_0} 6 \int \frac{(a_1(z; \theta_0) - a_1(z; \theta))^2}{b^2(z; \sigma) (\partial_y a_1(z; \theta))^2} \nu_0(dz).
\end{aligned}$$

Now consider T_2 :

$$\begin{aligned}
& -\frac{6}{N\Delta} \sum_{i=0}^{N-1} \left[\left(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}) + \bar{A}_1(Z_i; \theta_0^{(1)}) - \bar{A}_1(Z_i; \theta^{(1)}) \right) \right. \\
& \left. \frac{\left(Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(2)}) + \bar{A}_2(Z_i; \theta_0^{(2)}) - \bar{A}_2(Z_i; \theta^{(2)}) \right)}{b^2(Z_i; \sigma) \partial_y a_1(Z_i; \theta^{(1)})} - \frac{\left(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}) \right) \left(Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(2)}) \right)}{b^2(Z_i; \sigma_0) \partial_y a_1(Z_i; \theta_0^{(1)})} \right] = \\
& -\frac{6}{N\Delta} \sum_{i=0}^{N-1} \left[\left(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}) \right) \left(Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(2)}) \right) \right. \\
& \left. \left[\frac{1}{b^2(Z_i; \sigma) (\partial_y a_1(Z_i; \theta^{(1)}))} - \frac{1}{b^2(Z_i; \sigma_0) (\partial_y a_1(Z_i; \theta_0^{(1)}))} \right] \right. \\
& + \frac{\Delta}{b^2(Z_i; \sigma) (\partial_y a_1(Z_i; \theta^{(1)}))} \left[\left(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}) \right) (a_2(Z_i; \theta_0^{(2)}) - a_2(Z_i; \theta^{(2)})) + \right. \\
& \left. \left(Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(2)}) \right) (a_1(Z_i; \theta_0^{(1)}) - a_1(Z_i; \theta^{(1)})) \right] + \\
& \left. \frac{\Delta^2}{b^2(Z_i; \sigma) (\partial_y a_1(Z_i; \theta^{(1)}))} (a_1(Z_i; \theta_0^{(1)}) - a_1(Z_i; \theta^{(1)})) (a_2(Z_i; \theta_0^{(2)}) - a_2(Z_i; \theta^{(2)})) \right].
\end{aligned}$$

Thanks to Lemma 6 we know that:

$$\begin{aligned}
& -\frac{6}{N} \sum_{i=0}^{N-1} \frac{1}{b^2(Z_i; \sigma) (\partial_y a_1(Z_i; \theta^{(1)}))} \left[\left(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}) \right) (a_2(Z_i; \theta_0^{(2)}) - a_2(Z_i; \theta^{(2)})) + \right. \\
& \left. \left(Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(2)}) \right) (a_1(Z_i; \theta_0^{(1)}) - a_1(Z_i; \theta^{(1)})) \right] \xrightarrow{\mathbb{P}_0} 0
\end{aligned}$$

and, respectively, for the cross-term we have:

$$-\frac{6\Delta}{N} \sum_{i=0}^{N-1} \frac{(a_1(Z_i; \theta_0^{(1)}) - a_1(Z_i; \theta^{(1)})) (a_2(Z_i; \theta_0^{(2)}) - a_2(Z_i; \theta^{(2)}))}{b^2(Z_i; \sigma) (\partial_y a_1(Z_i; \theta^{(1)}))} \xrightarrow{\mathbb{P}_0} 0.$$

Then we use the fact that the variance of the cross-term is of order Δ^2 and by Lemma 5 we obtain:

$$-\frac{6}{N\Delta} \sum_{i=0}^{N-1} \frac{\left(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}) \right) \left(Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(2)}) \right)}{b^2(Z_i; \sigma) (\partial_y a_1(Z_i; \theta_0^{(1)}))} \left[\frac{(\partial_y a_1(Z_i; \theta_0^{(1)}))}{(\partial_y a_1(Z_i; \theta^{(1)}))} - 1 \right] \xrightarrow{\mathbb{P}_0} 0.$$

Finally, consider the term T_3 :

$$\begin{aligned}
T_3 &= \frac{2}{N} \sum_{i=0}^{N-1} \frac{1}{b^2(Z_i; \sigma)} \left[(Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(2)}))^2 + \right. \\
&\quad (Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(2)}))(\bar{A}_2(Z_i; \theta_0^{(2)}) - \bar{A}_2(Z_i; \theta^{(2)})) + \\
&\quad \left. (\bar{A}_2(Z_i; \theta_0^{(2)}) - \bar{A}_2(Z_i; \theta^{(2)}))^2 - (Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(2)}))^2 \right] = \\
&\quad \frac{2}{N} \sum_{i=0}^{N-1} \frac{1}{b^2(Z_i; \sigma_0)} \left[\Delta(Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(2)}))(a_2(Z_i; \theta_0^{(2)}) - a_2(Z_i; \theta^{(2)})) + \right. \\
&\quad \left. \Delta^2(a_2(Z_i; \theta_0^{(2)}) - a_2(Z_i; \theta^{(2)}))^2 \right]
\end{aligned}$$

Lemma 6 implies that $T_3 \xrightarrow{\mathbb{P}_0} 0$. The same holds for T_4 . Thus, we indeed have

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

□

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 $\hat{\theta}_{N, \Delta}^{(1)}$ 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 the identifiability of the drift function we have the consistency, that is $\hat{\theta}_{N, \Delta}^{(1)} \rightarrow \theta_0^{(1)}$.

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 (20). Then we have:

$$\begin{aligned}
&\int \frac{\Delta}{N} \frac{\partial^2}{\partial \theta^{(1)} \partial \theta^{(1)}} \mathcal{L}_{N, \Delta} \left(\theta_0^{(1)} + u(\hat{\theta}_{N, \Delta}^{(1)} - \theta_0), \theta^{(2)}, \sigma; z \right) du \cdot \sqrt{\frac{N}{\Delta}} (\hat{\theta}_{N, \Delta}^{(1)} - \theta_0^{(1)}) = \\
&\quad - \sqrt{\frac{\Delta}{N}} \frac{\partial}{\partial \theta^{(1)}} \mathcal{L}_{N, \Delta}(\theta_0^{(1)}, \theta^{(2)}, \sigma; z)
\end{aligned}$$

Note that the values of $\theta^{(2)}$ and σ may be taken arbitrary. Now we have to compute the first and the second order derivatives of (20). 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:

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

Lemma 7 gives us the following convergence to zero in probability:

$$\begin{aligned} \sqrt{\frac{\Delta}{N}} \sum_{i=1}^{N-1} \left[\frac{2\partial_{y,\theta^{(1)}}^2 a_1}{\partial_y a_1} - \frac{12\partial_{y,\theta^{(1)}}^2 a_1}{b^2(Z_i; \sigma) (\partial_y a_1)^3} \frac{(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}))^2}{\Delta^3 (\partial_y a_1)^2} \right] &\xrightarrow{\mathbb{P}_0} 0 \\ \sqrt{\frac{\Delta}{N}} \sum_{i=1}^{N-1} \left[-6 \frac{(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}))(Y_{i+1} - \bar{A}_2(Z_i; \theta^{(2)}))(\partial_{y,\theta^{(1)}}^2 a_1)}{\Delta^2 b^2(Z_i; \sigma) (\partial_y a_1)^2} \right] &\xrightarrow{\mathbb{P}_0} 0. \end{aligned}$$

The result holds since both terms consist of the variable which converges in distribution to a normal variable due to Lemma 7, multiplied by $\sqrt{\Delta}$, which converges to 0 by design. Then, by Slutsky's theorem, they converge to 0 in probability. Finally, applying again Lemma 7, we get:

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

Thus, we have the following convergence in law:

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

For the second order derivative we split again the expression (27) in several

parts and study their convergence:

$$\begin{aligned}
T_1 &:= \frac{\Delta}{N} \sum_{i=1}^{N-1} \frac{\partial^2}{\partial \theta^{(1)} \partial \theta^{(1)}} \left[\frac{2\partial_{y,\theta^{(1)}}^2 a_1}{\partial_y a_1} - \frac{12\partial_{y,\theta^{(1)}}^2 a_1}{b^2(Z_i; \sigma)(\partial_y a_1)^3} \frac{(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}))^2}{\Delta^3(\partial_y a_1)^2} \right] = \\
&\frac{\Delta}{N} \sum_{i=1}^{N-1} \left[\frac{2(\partial_{y\theta^{(1)\theta^{(1)}}}^3) - (\partial_{y,\theta^{(1)}}^2 a_1)^2}{(\partial_y a_1)} + \frac{12(\partial_{y,\theta^{(1)}}^2 a_1)^2}{\Delta^3 b^2(Z_i; \sigma)(\partial_y a_1)^4} (X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}))^2 \right. \\
&\left. + \frac{12}{\Delta^3 b^2(Z_i; \sigma)(\partial_y a_1)^5} \left[(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}))^2 \left((\partial_y a_1)^2 \partial_{y\theta^{(1)\theta^{(1)}}}^3 a_1 - 2(\partial_{y\theta^{(1)}}^2)^2 \partial_y a_1 \right) \right. \right. \\
&\left. \left. + 2\Delta(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}))(\partial_y a_1)^3 (\partial_{y\theta^{(1)}}^2 a_1) \right] \right]
\end{aligned}$$

$$\begin{aligned}
T_2 &:= \frac{\Delta}{N} \sum_{i=1}^{N-1} -\frac{12\Delta}{\Delta^2 b^2(Z_i; \sigma)(\partial_y a_1)^2} \left[(\partial_{\theta^{(1)}} a_1)^2 + \right. \\
&\left. (X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)})) \frac{(\partial_{\theta^{(1)\theta^{(1)}}}^2 a_1)(\partial_y a_1) - (\partial_{y\theta^{(1)}}^2 a_1)(\partial_{\theta^{(1)}} a_1)}{(\partial_y a_1)^2} \right]
\end{aligned}$$

$$T_3 := \frac{\Delta}{N} \sum_{i=1}^{N-1} \frac{6(Y_{i+1} - \bar{A}_2(Z_i; \theta^{(1)}))}{\Delta b^2(Z_i; \sigma)} \frac{(\partial_y a_1)^2 \partial_{\theta^{(1)\theta^{(1)}}}^2 a_1 - 2(\partial_{\theta^{(1)}} a_1)(\partial_{y\theta^{(1)}}^2 a_1)}{(\partial_y a_1)^4}$$

$$\begin{aligned}
T_4 &:= \frac{\Delta}{N} \sum_{i=1}^{N-1} \frac{6(Y_{i+1} - \bar{A}_2(Z_i; \theta^{(1)}))}{\Delta^2 b^2(Z_i; \sigma)} \left[\frac{(\partial_{y\theta^{(1)}}^2 a_1)^2 - (\partial_{y\theta^{(1)\theta^{(1)}}}^3 a_1)(\partial_y a_1)}{(\partial_y a_1)^2} \right. \\
&\left. (X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)})) + \frac{\Delta(\partial_{\theta^{(1)}} a_1)(\partial_{y\theta^{(1)}}^2 a_1)}{(\partial_y a_1)} \right]
\end{aligned}$$

Easy to see that the terms T_1, T_3 and T_4 converge to 0 by Lemmas 6 and 5.

T_2 , 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 3. We can split the contrast in the following sum:

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

where terms are given by follows:

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

For the term T_1 we have:

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

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:

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

For the third term, we use linearity for a_1 with respect to the parameter $\theta^{(1)}$, and then obtain the convergence to 0 in probability thanks to Theorem 1, continuous mapping theorem and Lemma 4:

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

Then, T_2 decomposes as:

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

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:

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

Finally, we treat the remaining term:

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

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 :

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

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

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

Then thanks to the tightness property of the estimated sequence $\hat{\theta}_{N,\Delta}^{(1)}$, we obtain that

$$\frac{\partial_y a_1(Z_i; \theta_0^{(1)})}{\partial_y a_1(Z_i; \hat{\theta}_{N,\Delta}^{(1)})} \xrightarrow{\mathbb{P}_0} 1$$

It gives the Lemma. \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 $\hat{\theta}_{N,\Delta}^{(1)}$ and consider the same sum, but with a different scaling, namely :

$$\lim_{N \rightarrow \infty, \Delta \rightarrow 0} \frac{1}{N\Delta} \left[\mathcal{L}_{N,\Delta}(\hat{\theta}_{N,\Delta}^{(1)}, \theta_0^{(2)}, \sigma_0^2; Z_{0:N}) - \mathcal{L}_{N,\Delta}(\hat{\theta}_{N,\Delta}^{(1)}, \theta_0^{(2)}, \sigma_0^2; Z_{0:N}) \right] = T_1 + T_2 + T_3$$

where the terms are given as follows:

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

T_1 is obviously cancelled. Consider T_2 :

$$\begin{aligned} T_2 &= -\frac{6}{N\Delta^3} \sum_{i=0}^{N-1} \left[\frac{\Delta \left(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}) \right) (a_2(Z_i; \theta_0^{(2)}) - a_2(Z_i; \theta_0^{(2)}))}{\partial_y a_1(Z_i; \hat{\theta}_{N,\Delta}^{(1)}) b^2(Z_i; \sigma)} + \right. \\ &\quad \left. \frac{\Delta^2 (a_1(Z_i; \theta_0^{(1)}) - a_1(Z_i; \hat{\theta}_{N,\Delta}^{(1)})) (a_2(Z_i; \theta_0^{(2)}) - a_2(Z_i; \theta_0^{(2)}))}{\partial_y a_1(Z_i; \hat{\theta}_{N,\Delta}^{(1)}) b^2(Z_i; \sigma)} \right] \end{aligned}$$

Then, the first part of the sum converges to zero in probability after applying Lemma 6. Second part of the sum also converges to zero because $N\Delta \rightarrow \infty$ by design, and $\hat{\theta}_{N,\Delta}^{(1)} \xrightarrow{\mathbb{P}_0} \theta_0^{(1)}$ (so that $a_1(Z_i; \theta_0^{(1)}) - a_1(Z_i; \hat{\theta}_{N,\Delta}^{(1)})$ also converges

to 0). So we just have to consider the remaining term T_3 :

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

The first part of the sum is canceled due to Lemma 6. Then we apply Lemma 5 and get the convergence:

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

□

Theorem 2. 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 (20) and apply an appropriate scaling

$$\int C_{N,\Delta}(\theta_0 + u(\hat{\theta}_{N,\Delta} - \theta_0); z) du E_{N,\Delta} = -D_{N,\Delta}(\theta_0),$$

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

$$\begin{aligned}
C_{N,\Delta}(\theta) &:= \begin{bmatrix} \frac{1}{N\Delta} \frac{\partial^2}{\partial \theta^{(2)} \partial \theta^{(2)}} \mathcal{L}_{N,\Delta}(\theta^{(1)}, \theta^{(2)}, \sigma; Z_{0:N}) & \frac{1}{N\sqrt{\Delta}} \frac{\partial^2}{\partial \sigma \partial \theta^{(2)}} \mathcal{L}_{N,\Delta}(\theta^{(1)}, \theta^{(2)}, \sigma; Z_{0:N}) \\ \frac{1}{N\sqrt{\Delta}} \frac{\partial^2}{\partial \theta^{(2)} \partial \sigma} \mathcal{L}_{N,\Delta}(\theta^{(1)}, \theta^{(2)}, \sigma; Z_{0:N}) & \frac{1}{N} \frac{\partial^2}{\partial \sigma \partial \sigma} \mathcal{L}_{N,\Delta}(\theta^{(1)}, \theta^{(2)}, \sigma; Z_{0:N}) \end{bmatrix}, \\
E_{N,\Delta} &:= \begin{bmatrix} \sqrt{N\Delta}(\hat{\theta}_N^{(2)} - \theta_0^{(2)}) \\ \sqrt{N}(\hat{\sigma}_N - \sigma_0) \end{bmatrix}, \quad D_{N,\Delta} = \begin{bmatrix} \frac{1}{\sqrt{N\Delta}} \frac{\partial}{\partial \theta^{(2)}} \mathcal{L}_{N,\Delta}(\theta^{(1)}, \theta^{(2)}, \sigma; Z_{0:N}) \\ \frac{1}{\sqrt{N}} \frac{\partial}{\partial \sigma} \mathcal{L}_{N,\Delta}(\theta^{(1)}, \theta^{(2)}, \sigma; Z_{0:N}) \end{bmatrix}.
\end{aligned}$$

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

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

We start with proving the convergence for the terms $C_{N,\Delta}$. Note that the parameter $\theta^{(1)}$ is not fixed, and thus we plug-in the estimated sequence $\hat{\theta}_{N,\Delta}^{(1)}$. Then we can obtain a convergence in probability after few technical steps.

We start with $C_{N,\Delta}^{11}$:

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

Note that thanks to Lemma 6 we know that

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

What about the remaining term, we can recall the condition (H1):

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

We know that $(\theta_0^{(1)} - \hat{\theta}_{N,\Delta}^{(1)}) \sqrt{\frac{N}{\Delta}}$ 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 $\sqrt{N\Delta}$ converges to zero since $N\Delta \rightarrow \infty$ by design. However, as $N\Delta \rightarrow \infty$, this term converges to 0 in probability. As a result,

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

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

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

Then we consider the remaining term. Recall the assumption (H1), namely that the function $a_1(Z_i; \theta^{(1)})$ is linear with respect to $\theta^{(1)}$. We start with the term

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

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:

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

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

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

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

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

That gives the result. \square

7.4 Consistency and normality of the least squares contrast

Theorem 3. Consistency. The proof follows the one of the classical contrast. Recall that (23) is written as

$$\mathcal{L}_{N,\Delta}^d(\theta; Z_{0:N}) = \sum_{i=0}^{N-1} \left((X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}))^2 + (Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(2)}))^2 \right)$$

Let us now consider the following difference:

$$\begin{aligned} \frac{1}{N\Delta^2} \left[\mathcal{L}_{N,\Delta}^d(\theta; Z_{0:N}) - \mathcal{L}_{N,\Delta}^d(\theta_0; Z_{0:N}) \right] &= \frac{1}{N\Delta^2} \sum_{i=0}^{N-1} \left[(X_{i+1} - \bar{A}_1(Z_i; \theta^{(1)}))^2 \right. \\ &\quad \left. + (Y_{i+1} - \bar{A}_2(Z_i; \theta^{(2)}))^2 - (X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}))^2 - (Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(2)}))^2 \right] \\ &= \frac{1}{N\Delta^2} \sum_{i=0}^{N-1} \left[2(X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}))(\bar{A}_1(Z_i; \theta_0^{(1)}) - \bar{A}_1(Z_i; \theta^{(1)})) + (\bar{A}_1(Z_i; \theta_0^{(1)}) - \bar{A}_1(Z_i; \theta^{(1)}))^2 \right. \\ &\quad \left. + 2(Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(2)}))(\bar{A}_2(Z_i; \theta_0^{(2)}) - \bar{A}_2(Z_i; \theta^{(2)})) + (\bar{A}_2(Z_i; \theta_0^{(2)}) - \bar{A}_2(Z_i; \theta^{(2)}))^2 \right] \end{aligned}$$

Then we have from Lemmas 5, 6:

$$\begin{aligned} \frac{2\Delta}{N\Delta^2} \sum_{i=0}^{N-1} (X_{i+1} - \bar{A}_1(Z_i; \theta_0^{(1)}))(a_1(Z_i; \theta_0^{(1)}) - a_1(Z_i; \theta^{(1)})) &\xrightarrow{\mathbb{P}_0} 0 \\ \frac{2\Delta}{N\Delta^2} \sum_{i=0}^{N-1} (Y_{i+1} - \bar{A}_2(Z_i; \theta_0^{(2)}))(a_2(Z_i; \theta_0^{(2)}) - a_2(Z_i; \theta^{(2)})) &\xrightarrow{\mathbb{P}_0} 0 \\ \frac{\Delta^2}{N\Delta^2} \sum_{i=0}^{N-1} (a_1(Z_i; \theta_0^{(1)}) - a_1(Z_i; \theta^{(1)}))^2 &\xrightarrow{\mathbb{P}_0} \int (a_1(z; \theta_0^{(1)}) - a_1(z; \theta^{(1)}))^2 \nu_0(dz) \\ \frac{\Delta^2}{N\Delta^2} \sum_{i=0}^{N-1} (a_2(Z_i; \theta_0^{(2)}) - a_2(Z_i; \theta^{(2)}))^2 &\xrightarrow{\mathbb{P}_0} \int (a_2(z; \theta_0^{(2)}) - a_2(z; \theta^{(2)}))^2 \nu_0(dz) \end{aligned}$$

From that we can conclude that there exists a subsequence $\hat{\theta}_{N,\Delta} = \arg \min_{\theta} \mathcal{L}_{N,\Delta}^d(\theta; Z_{0:N})$ that tends to θ_∞ . Since the minimum is attained at the point θ_0 , we conclude that $\theta_\infty = \theta_0$. Then the result follows from the identifiability of the drift functions.

Asymptotic normality. We apply again a Taylor formula for a function (23):

$$\int C_N \left(\theta_0 + u(\hat{\theta}_N - \theta_0) \right) du E_N = D_N(\theta_0),$$

where we define

$$C_N(\theta) := \begin{bmatrix} \frac{1}{N\Delta^3} \frac{\partial^2}{\partial \theta^{(1)} \partial \theta^{(1)}} \mathcal{L}_{N,\Delta}^d(\theta; Z_{0:N}) & \frac{1}{N\Delta^2} \frac{\partial^2}{\partial \theta^{(1)} \partial \theta^{(2)}} \mathcal{L}_{N,\Delta}^d(\theta; Z_{0:N}) \\ \frac{1}{N\Delta^2} \frac{\partial^2}{\partial \theta^{(1)} \partial \theta^{(2)}} \mathcal{L}_{N,\Delta}^d(\theta; Z_{0:N}) & \frac{1}{N\Delta} \frac{\partial^2}{\partial \theta^{(2)} \partial \theta^{(2)}} \mathcal{L}_{N,\Delta}^d(\theta; Z_{0:N}) \end{bmatrix},$$

$$E_N := \begin{bmatrix} \sqrt{N\Delta^3}(\hat{\theta}_{N,\Delta}^{(1)} - \theta_0^{(1)}) \\ \sqrt{N\Delta}(\hat{\theta}_N^{(2)} - \theta_0^{(2)}) \end{bmatrix}, \quad D_N(\theta) = \begin{bmatrix} \frac{1}{N\sqrt{\Delta^3}} \frac{\partial}{\partial \theta^{(1)}} \mathcal{L}_{N,\Delta}^d(\theta; Z_{0:N}) \\ \frac{1}{N\sqrt{\Delta}} \frac{\partial}{\partial \theta^{(2)}} \mathcal{L}_{N,\Delta}^d(\theta; Z_{0:N}) \end{bmatrix}.$$

Using Lemma 7 we get:

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

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 combination with the consistency result, gives the theorem. \square