ESAIM: PS 23 (2019) 1–36
https://doi.org/10.1051/ps/2018004

ESAIM: Probability and Statistics
www.esaim-ps.org

CONSISTENCY OF A LIKELIHOOD ESTIMATOR FOR STOCHASTIC DAMPING HAMILTONIAN SYSTEMS. TOTALLY OBSERVED DATA☆

José Rafael León[1], Luis-Ángel Rodríguez[2] and Roberto Ruggiero[2,*]

Abstract. In this work we prove the consistency of an estimator for a stochastic damping Hamiltonian system considering that both position and velocity are observed. Next we perform some simulations, including the case when only position is available, to see how the estimators work numerically and then compare the obtained results with those obtained by other authors.

Mathematics Subject Classification. 62M05, 60G10

Received October 27, 2016. Accepted November 21, 2017.

1. Introduction

It is of interest in many researching areas to have a consistent estimator in order to identify a stochastic model which fits discretely observed data. In many cases the data is totally observed (*i.e.* position and velocity are observed) nevertheless, in some applications it is not possible to measure both components and only position is available. In several applications as random mechanics in engineering, neurophysiology, finance, simulations in chemistry, etc. (see for example the works of Robert and Spanos [17], Lindner and Schimansky-Geier [11], Nicolau [13], Lelièvre *et al.* [9]), the stochastic model consists on non-linear stochastic oscillators (or non-linear random vibrations systems) of the form

$$\ddot{x}_t + g_1(x_t; a)\, \dot{x}_t + g_2(x_t; b)\, x_t = \sigma \mathrm{d}W_t, \tag{1.1}$$

where g_1 and g_2 are non-linear functions of the position x_t, depending on the parameters a, b. The term $g_1(x_t; a)$ represents the damping coefficient force of the system and the term $g_2(x_t; b)\, x_t$ represents the drift force which is supposed to be driven by a potential $V(x)$ such that $V'(x) = g_2(x; b)\, x$. The term W_t is a standard Brownian motion and $\sigma > 0$ is the amplitude of the noise. For a further study of this kind of models we refer to the book of Gitterman [8]. Model (1.1) can be represented as a two dimensional hypoelliptic diffusion (or stochastic

☆This work was partially supported by the MathAmSud 16MATH03 SIDRE project.

Keywords and phrases: Two dimensional hypoelliptic diffusion, stochastic damping Hamiltonian systems, likelihood estimator, consistency, discretely observed data.

[1] Escuela de Matemáticas, Facultad de Ciencias, Universidad Central de Venezuela and IMERL, Facultad de Ingeniería, UDELAR, Montevideo, Uruguay.
[2] Dpto. de Matemáticas, FACYT, Universidad de Carabobo, Valence, Venezuela.

[*] Corresponding author: ruggiero@uc.edu.ve

damping Hamiltonian system) of the form

$$\begin{cases} \mathrm{d}x_t = y_t \mathrm{d}t \\ \mathrm{d}y_t = \sigma \mathrm{d}W_t - [g_1(x_t;a)y_t + g_2(x_t;b)x_t]\,\mathrm{d}t. \end{cases} \tag{1.2}$$

Under some assumptions for the potential $V(x)$, if the damping coefficient force $g_1(x;a) = c > 0$ and the random coefficient $\sigma > 0$, it is known from Abbaoui and Bendjeddou [1] that there exists a time stationary solution for system (1.2). In the paper of Wu [21] were established some probabilistic properties of model (1.2) like exponential ergodicity of the invariant distribution, among others, under more general assumptions on the potential $V(x)$ and the damping coefficient force $g_1(x;a)$ than in the work of Abbaoui and Bendjeddou [1]. Statistical inference for classic diffusions is widely studied and in many cases parametric estimation is carried out by finding a contrast function obtained by performing an Euler approximation directly on the model (see *e.g.* Genon-Catalot *et al.* [6] for further information). Nevertheless, for hypoelliptic diffusions Euler's approximation cannot be done directly on the model due to the fact that the variance matrix is degenerate. To avoid this problem, a maximum likelihood based estimator for model (1.1) was introduced by Ozaki [14] performing a local linearization method in order to obtain a discrete model for which the variance matrix is not degenerate. Talay [19] used an implicit Euler scheme in order to demonstrate exponential convergence to the invariant measure. In the paper of Pokern *et al.* [16] they focused on higher order hypoelliptic diffusions and proposed an empirical approximation of the likelihood based on Itô expansions and a hybrid Gibbs sampler to construct a Bayesian estimator. In the work of Samson and Thieullen [18] they used the Euler approximation only in the second equation of the system (1.2) allowing them to construct a consistent estimator.

In the context of non-parametric estimation works are few. In two recent works of Cattiaux *et al.* [3, 4] non-parametric estimation of the invariant density and the drift term was presented using kernel estimators which led them to establish a central limit theorem for such estimators under partial observations.

In this work we prove the consistency of the contrast obtained by the local linearization method in the case of totally observed data. Our work is strongly inspired by the article of Ozaki [14]. In his paper Ozaki pointed out that there is a bias that comes from estimating the parameters using the discrete linearized model from the observations of the continuous model. This paper is organized as follows: In Section 2 we present the model and set some assumptions. Section 3 contains a brief sketch of the local linearization method and are shown two important results relative to the model. In Section 4 we prove the consistency of the estimator, assuming the data to be completely observed. Section 5 is devoted to perform a simulation study in order to see how the estimators work numerically and then we compare the results obtained with those obtained by Ozaki [14], Samson *et al.* [18] and Pokern *et al.* [16]. For comparison purposes we include a case of simulation when only position is available and the velocity is approximated using an Euler discrete scheme. Finally, conclusions are presented.

2. Preliminaires

In this section we will set some notation and assumptions that will be considered throughout this paper. In this paper we consider the system of stochastic differential equations:

$$\begin{cases} \mathrm{d}x(t) = y(t)\mathrm{d}t \\ \mathrm{d}y(t) = \sigma \mathrm{d}W_t - [g_1(x(t);a)y(t) + g_2(x(t);b)x(t)]\,\mathrm{d}t, \end{cases} \tag{2.1}$$

where $z(t) := (x(t), y(t))$, $t \geq 0$ denotes the position and the velocity of a particle at the time t. The parameters are denoted by a and b and each of them represents a set of parameters $a = (a_1, \ldots, a_q)$ and $b = (b_1, \ldots, b_p)$ contained in the expression for g_1 and g_2 respectively. We will assume that

$$a \in \varTheta_1 \subset \mathbb{R}^q \quad \text{and} \quad b \in \varTheta_2 \subset \mathbb{R}^p,$$

where Θ_1 and Θ_2 are compact sets. Besides, we denote by $\theta = (a, b) \in \Theta \subset R^{p+q}$ the parameters, $\Theta = \Theta_1 \times \Theta_2$ as a compact set containing the parameters. The true parameters will be denoted by $\theta_0 = (a_0, b_0) \in \mathring{\Theta}$ (the interior of Θ). We suppose that the strength of the noise, denoted by σ, is a positive constant. For $z = (x, y) \in \mathbb{R}^2$ we define $P : \mathbb{R}^2 \times \Theta \to \mathbb{R}$ as

$$P(z;\theta) := -g_1(x;a)y - g_2(x;b)x. \tag{2.2}$$

The partial derivatives of P will be denoted by the vectors $\nabla_z P(z;\theta) = (P_x(z;\theta), P_y(z;\theta)) \in \mathbb{R}^2$ and $\nabla_\theta P(z;\theta) = (P_a(z;\theta), P_b(z;\theta)) \in \mathbb{R}^{p+q}$. In the sequel we will use generic positive constants $\mathbf{C}$ or $\mathbf{C}_r$ (in this case the constant depends on the subscript), that can be different from line to line unless otherwise specified.

For the system (2.1), throughout this paper, we make the following assumptions

H1 (1) For all $N > 0 : \sup_{|x|\le N} \; |g_1(x;a)| < \infty$.
(2) There exist positive constants c, L such that $g_1(x;a) \ge c$, for all $|x| > L$.
(3) $g_1(x;a) \in \mathcal{C}^2(\mathbb{R} \times \mathring{\Theta}_1)$.
(4) There exists a polynomial $q_1(|x|) = \mathbf{C}_\theta |x|^{\gamma_1}$, $\mathbf{C}_\theta > 0$ such that

$$g_1(x;a) \le \mathbf{C}_\theta |x|^{\gamma_1} \quad \text{and} \quad g_1^{'}(x;a) \le \mathbf{C}_\theta |x|^{\gamma_1},$$

for all $x \in \mathbb{R}$.

H2 (1) We have a potential $V(x;b)$ such that $V(x;b)$ is bounded from below and $V(x;b) \in \mathcal{C}^2(\mathbb{R} \times \mathring{\Theta}_2)$.
(2) Its derivative has the form $V^{'}(x;b) = g_2(x;b)x$.
(3) We will assume that the damping condition

$$+\infty \ge \liminf_{|x|\to+\infty} \frac{x^2 \cdot g_2(x;b)}{|x|} \ge v > 0,$$

holds for some constant v.
(4) The potential has polynomial growth at infinity. There exists a polynomial $q_2(|x|) = \mathbf{C}_\theta |x|^{\gamma_2}$, $\mathbf{C}_\theta > 0$ such that

$$V(x;b) \le \mathbf{C}_\theta |x|^{\gamma_2} \quad \text{and} \quad V^{'}(x;b) \le \mathbf{C}_\theta |x|^{\gamma_2},$$

for all $x \in \mathbb{R}$.

H3 (1) $P : \mathbb{R}^2 \times \Theta \to \mathbb{R}^2$ is continuous.
(2) $P(z;\theta) \in \mathcal{C}^2(\mathbb{R}^2 \times \mathring{\Theta})$; for each θ, there exist positive constants $\mathbf{C}_\theta$ and γ such that $P(z;\theta) \le \mathbf{C}_\theta(1 + ||z||)^\gamma$, for all $z \in \mathbb{R}^2$.
(3) For all $\theta \ne \theta^{'}$, $P(z;\theta) \ne P(z;\theta^{'})$.
(4) There exists a positive constant γ such that

$$\sup_{\theta' \in \Theta} \|\nabla_\theta P(z;\theta')\| \le \sup_{\theta' \in \Theta} \{\mathbf{C}_{\theta'}\} \|z\|^\gamma = \mathbf{C}\|z\|^\gamma.$$

For each $\theta \in \Theta$, it can be inferred from hypotheses H1 and H2 that the followings bounds hold

$$|P_x(z;\theta)| \le \mathbf{C}_\theta \, \|z\|^{\gamma_3} \quad \text{and} \quad |P_y(z;\theta)| \le \mathbf{C}_\theta \, \|z\|^{\gamma_1}, \quad \forall z \in \mathbb{R}^2. \tag{2.3}$$

According to the work of Wu [21] assumptions H1–H2 imply that

- there exists a Foster-Lyapunov function $\Psi \geq 1$ which satisfies $\mathcal{L}\Psi \leq -\alpha\Psi + D\,1_K$, for some constants $\alpha, D > 0$ and some compact subset K. Here $\mathcal{L}$ denotes the infinitesimal generator

$$\mathcal{L} = \frac{\sigma^2}{2}\frac{\mathrm{d}^2}{\mathrm{d}y^2} + y\frac{\mathrm{d}}{\mathrm{d}x} - (g_1(x;a)y + g_2(x;b)x)\,\frac{\mathrm{d}}{\mathrm{d}y}.$$

- Also Ψ satisfies

$$\log\left(\Psi(x,y)\right) \geq D\left(|y|^2 + |x|\right),$$

when $\|z\|$ goes to infinity, for some well chosen constant $D > 0$.

It is inferred from the observations above that the process $z(t) = (x(t), y(t))$ is positive recurrent and has a unique invariant measure $\pi_c(z) = \pi_c(x, y)$. Notice that the invariant measure $\pi_c(z(0))\mathrm{d}z(0)$ depends on the parameters θ which defines the model. For example in the case of the Duffing oscillator, considering $g_1(x;a) = g_1(a)$ a constant, the expression for the invariant measure (up to a numerical constant factor), is given by

$$\pi_c(x,y)\mathrm{d}x\mathrm{d}y = \exp\left(-\frac{2g_1(a)}{\sigma^2}\mathbf{H}(x,y)\right)\mathrm{d}x\mathrm{d}y,$$

where $\mathbf{H}(x,y) = \frac{1}{2}\left|y\right|^2 + V(x)$ is the Hamiltonian. In general a closed expression for the invariant measure is unknown. For example when $g_1(x;a) = a_1x^2 - a_2$ with $a_1, a_2 > 0$, $V(x) = \frac{1}{2}b^2x^2$ (Van der Pol oscillator) its invariant measure is unknown.

These quoted facts entail also some properties related to asymptotic independence. In fact the process $z(t)$ becomes α-mixing and all the polynomial moments of the measure π_c are finite. See Cattiaux *et al.* [3]. In this work we assume that the distribution of the process $z(0) = (x(0), y(0))$ is the invariant measure π_c.

3. Locally linearized model

In this section we start by presenting the local linearization method introduced by Ozaki [14, 15] and we present explicitly its joint density.

In Ozaki's article [14] a statistical identification method for nonlinear random vibration system models, based on the maximum likelihood method for discrete time models, was obtained. He considered the continuous time stochastic nonlinear differential equation model for the nonlinear random vibration system (1.1). The model (1.1) is expressed as a bivariate stochastic dynamical system model

$$\dot{z}(t) = f(z(t);\theta) + \mathrm{d}\mathbf{W}(t), \tag{3.1}$$

where

- $z(t) = (x(t), y(t))'$,
- $f(z(t);\theta) = (y(t), -g_1(x(t);a)y(t) - g_2(x(t);b)x(t))' = (y(t), P(z(t);\theta))'$ and
- $\mathrm{d}\mathbf{W}(t) = (0, \sigma\mathrm{d}W_t)'$.
- The process $\{z(t)\}$ is observed at the equally spaced discrete times $t = ih$, $i = 0, \ldots, N$.

Notation A' is used for the transpose of a vector or a matrix A. The variance-covariance matrix for the bivariate white noise $\mathrm{d}\mathbf{W}(t)$ is given by

$$\mathbb{E}\left[\mathrm{d}\mathbf{W}(t)\,\mathrm{d}\mathbf{W}(s)'\right] = \begin{bmatrix} 0 & 0 \\ 0 & \sigma^2\,\delta(t-s) \end{bmatrix},$$

where $\delta(t)$ is the Dirac delta function. We follow the work of Ozaki [15] in order to find a discrete scheme which approximates the continuous model (3.1) in the interval $[ih, (i+1)h)$ for $h > 0$ small enough. First we consider the homogeneous equation

$$\dot{z}(t) = f(z(t); \theta). \tag{3.2}$$

By taking the derivatives in (3.2) we get

$$\ddot{z}(t) = J(z(t); \theta)\dot{z}(t), \tag{3.3}$$

where

$$J(z(t); \theta) = \left[\frac{\partial f(z; \theta)}{\partial z}\right]_{z=z(t)} = \begin{bmatrix} 0 & 1 \\ g_1'(x; a)y + g_2'(x; b)x + g_2(x; b) & g_1(x; a) \end{bmatrix}.$$

At this point, we will add the following hypothesis

H4 The matrix $\left[\frac{\partial f(z;\theta)}{\partial z}\right]$ is not singular for all $z \in \mathbb{R}^2$ and $\theta \in \mathring{\Theta}$, that is

$$g_1'(x; a)y + g_2'(x; b)x + g_2(x; b) \neq 0,$$

for $(x, y) \in \mathbb{R}^2$ and $\theta \in \mathring{\Theta}$.

Now, assuming that f is linear in the interval $[ih, (i+1)h)$ the Jacobian matrix becomes constant in this interval, therefore $J(z(s); \theta) = J(z(t); \theta)$ for $s \in [ih, (i+1)h)$ and integrating $\ddot{z}(s) = J(z(t); \theta)\dot{z}(s)$ two times we obtain

$$\begin{aligned} z(t+h) &= z(t) + J(z(t); \theta)^{-1}(e^{J(z(t);\theta)h} - I)f(z(t); \theta) \\ &= [I + J(z(t); \theta)^{-1}(e^{J(z(t);\theta)h} - I)F(z(t); \theta)]z(t), \end{aligned}$$

where

$$F(z(t); \theta) := \begin{bmatrix} 0 & 1 \\ -g_2(x(t); a) & -g_1(x(t); b) \end{bmatrix}$$

is a matrix such that $F(z(t); \theta)z(t) = f(z(t); \theta)$. We define $A(z(t), h; \theta) = I + J^{-1}(z(t); \theta)(e^{J(z(t);\theta)h} - I)F(z(t); \theta)$ and the matrix $L(z(t); \theta)h$ by $e^{L(z(t);\theta)h} = A(z(t), h; \theta)$. Now we consider the bivariate stochastic equation

$$\dot{z}(s) = L(z(t); \theta)z(s) + \mathrm{d}\mathbf{W}(s). \tag{3.4}$$

Using the parameters variation formula, the solution of the equation (3.4) is given by

$$z(t+h) = e^{L(z(t);\theta)h}z(t) + \int_t^{t+h} e^{L(z(t);\theta)(t+h-s)}\mathrm{d}\mathbf{W}(s). \tag{3.5}$$

Remark 3.1. We must point out that the above step is where the noise is propagated to the two coordinates. Thus, although it is a first order discretization, this procedure allows to take advantage over other first order methods.

When the process $\{z(t)\}$ is observed at the equally spaced discrete times $t = ih$, $i = 1, \ldots, N$ we consider the discrete process $\{z(ih)\} = \{z_{ih}\}$ defined by means of the equation

$$z_{(i+1)h} = A_{ih} z_{ih} + \int_{ih}^{(i+1)h} e^{L(z_{ih};\theta)[(i+1)h-s]} \mathrm{d}\mathbf{W}(s), \tag{3.6}$$

We consider the observation equation to be

$$z((i+1)h) = z(ih) + \int_{ih}^{(i+1)h} f(z(s);\theta)\mathrm{d}s + \int_{ih}^{(i+1)h} \mathrm{d}\mathbf{W}(s)\mathrm{d}s. \tag{3.7}$$

The variance-covariance matrix $\Sigma_{ih}(h;\theta)$ for the discrete time white noise $\int_{ih}^{(i+1)h} e^{L(z_{ih};\theta)[(i+1)h-s]}\mathrm{d}\mathbf{W}(s)$ is given by

$$\begin{aligned} \Sigma_{ih}(h;\theta) &= \frac{\sigma^2}{(\mu_1-\mu_2)^2}\begin{bmatrix} \frac{1}{2\mu_1}E_1 - \frac{2}{\mu_1+\mu_2}E_{1,2} + \frac{1}{2\mu_2}E_2 & \frac{1}{2}E_1 - E_{1,2} + \frac{1}{2}E_2 \\ \frac{1}{2}E_1 - E_{1,2} + \frac{1}{2}E_2 & \frac{\mu_1}{2}E_1 - \frac{2\mu_1\mu_2}{\mu_1+\mu_2}E_{1,2} + \frac{\mu_2}{2}E_2 \end{bmatrix} \\ &:= \frac{\sigma^2}{(\mu_1-\mu_2)^2}\begin{bmatrix} a'_{11}(z_{ih},h;\theta) & a'_{12}(z_{ih},h;\theta) \\ a'_{12}(z_{ih},h;\theta) & a'_{22}(z_{ih},h;\theta)\end{bmatrix} = \sigma^2\begin{bmatrix} a_{11}(z_{ih},h;\theta) & a_{12}(z_{ih},h;\theta) \\ a_{12}(z_{ih},h;\theta) & a_{22}(z_{ih},h;\theta)\end{bmatrix}, \end{aligned} \tag{3.8}$$

where

- $\mu_1 = \mu_1(z_{ih},h;\theta)$ and $\mu_2 = \mu_2(z_{ih},h;\theta)$ are the eigenvalues of the matrix $J(z_{ih};\theta)$,
- $E_1 = e^{2\mu_1 h} - 1$,
- $E_2 = e^{2\mu_2 h} - 1$ and
- $E_{1,2} = e^{(\mu_1+\mu_2)h} - 1$.

Remark 3.2. Notice that the variance-covariance matrix for the continuous noise $\mathrm{d}\mathbf{W}(t)$ is singular, while the variance-covariance of the discrete noise is symmetric and, for a fixed $h > 0$, it has full rank equal to 2. Moreover, the matrix $\Sigma_{ih}(h;\theta)$ is the true covariance matrix for the discrete time white noise $\int_{ih}^{(i+1)h} e^{L(z_{ih};\theta)[(i+1)h-s]}\mathrm{d}\mathbf{W}(s)$.

In the next lemma an important result regarding the asymptotic behavior for the elements of the matrices $\Sigma_{ih}(h;\theta)$ and $\Sigma_{ih}^{-1}(h;\theta)$ is stated.

Lemma 3.3. *Consider the matrices $\Sigma_{ih}(h;\theta)$ and $\Sigma_{ih}^{-1}(h;\theta)$. Under the hypotheses $H1$–$H3$ and $H5''$ we have*

$$\frac{1}{h^3}a_{11}(z_{ih},h;\theta) \to \frac{1}{3}, \quad \frac{1}{h}a_{22}(z_{ih},h;\theta) \to 1 \quad \text{and} \quad \frac{1}{h^2}a_{12}(z_{ih},h;\theta) \to \frac{1}{2}, \tag{3.9}$$

and

$$h^3 b_{11}(z_{ih},h;\theta) \to 12, \quad hb_{22}(z_{ih},h;\theta) \to 4 \quad \text{and} \quad h^2 b_{12}(z_{ih},h;\theta) \to -6, \tag{3.10}$$

in probability. Moreover

$$\frac{h^3}{2}||\Sigma_{ih}^{-1}(h;\theta)|| = O_{\mathbb{P}}(1), \tag{3.11}$$

uniformly in θ.

Its proof is left to Appendix A..

Model (3.6) is then rewritten as

$$z_{(i+1)h} = A_{ih}(h;\theta) z_{ih} + B_{ih}(h;\theta) \Xi_{(i+1)h}, \tag{3.12}$$

where

- $A_{ih}(h;\theta) = A(z_{ih}, h; \theta)$,
- the initial condition is $z_{ih} = z(ih)$, *i.e.* the model starts with the invariant measure of the continuous model.
- $B_{ih}(h;\theta) = U_{ih}(h;\theta) \begin{bmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{bmatrix}$. Here $\lambda_1 = \lambda_1(z_{ih}, h; \theta)$ and $\lambda_2 = \lambda_2(z_{ih}, h; \theta)$ stands for the eigenvalues of the matrix $\frac{1}{\sigma^2} \Sigma_{ih}(h;\theta)$, and the columns of the matrix $U_{ih}(h;\theta)$ are the normalized eigenvectors associated to each eigenvalues.
- $\Xi_{(i+1)h}$ is a bivariate Gaussian noise with variance-covariance matrix $\sigma^2 I$ and mean zero.

In the following lemma we compute the joint density of the discrete model (3.12)

Lemma 3.4. *The joint density $p(z_h, \dots, z_{hN}; \theta, \sigma^2)$ of the discrete model (3.12) is given by*

$$\frac{p(z_h)}{(2\pi)^{\frac{N-1}{2}} \sigma^{2(N-1)}} \prod_{i=1}^{N-1} \frac{1}{|\det B_{(i-1)h}(h;\theta)|}$$
$$\times \exp\left[-\frac{1}{2\sigma^2} \left[z_{ih} - A_{(i-1)h}(h;\theta) z_{(i-1)h} \right]' \left[B^{-1}_{(i-1)h}(h;\theta) \right]' \left[B^{-1}_{(i-1)h}(h;\theta) \right] \left[z_{ih} - A_{(i-1)h}(h;\theta) z_{(i-1)h} \right] \right].$$

The proof of this lemma is postponed to Appendix A..

As a consequence of the lemma (3.4), the log likelihood of the model (3.12) is given by

$$l(\theta) = \log p\left(z_h, \dots, z_{Nh}; \theta, \sigma^2 \right). \tag{3.13}$$

Finally, keeping only the terms that depend on θ we define the approximate the likelihood contrast for the discrete model (3.12) as

$$M_{N,h}(z_h^{Nh};\theta) = \frac{1}{2N} \sum_{i=0}^{N-1} \left\| B^{-1}_{ih}(h;\theta) \left[z_{(i+1)h} - A_{ih}(h;\theta)\, z_{ih} \right] \right\|^2. \tag{3.14}$$

4. Asymptotic consistency of the estimator

In this section we establish the main result of this paper. We use the notation $z_h^{Nh} = (z_h, \dots, z_{Nh})$. As $h = h_N$, for $h \to 0$, we need to assume that $hN \to \infty$, then $N \to \infty$.

Remark 4.1. We will consider below as a contrast the logarithm of the quotients of likelihoods. This statistics can be written as

$$\left[M_{N,h}(z_h^{Nh};\theta) - M_{N,h}(z_h^{Nh};\theta_0) \right].$$

The estimator obtained *via* the minimization of this latter function is the same to the one gotten by minimization of the function $M_{N,h}(z_h^{Nh};\theta)$.

First we set some notations

(1) $\hat{\theta}_{N,h} = \arg\min_{\theta \in \Theta} M_{N,h}(z_h^{Nh};\theta)$.

(2) As usual, $\theta_0 = (a_0, b_0)$ stands for the true parameters of the continuous model (1.1). An important feature in the work of Ozaki [14] is the fact that the parameters of the discrete model (3.12) are exactly the same as the parameters of the continuous model (1.1). It allows us to regard θ_0 in two equivalent ways
   - $\theta_0 = \arg\min_{\theta\in\Theta} \mathbb{E}_{Q_{d,h}}\left[M_{N,h}(z_h^{Nh};\theta)\right]$. Here $Q_{d,h}$ is the stationary measure of the pair $(z_{ih}, z_{(i+1)h})$ given by $p(z_{(i+1)h} \mid z_{ih})\pi_d(z_{ih})$, where $p(\cdot \mid \cdot)$ is the density of the transition kernel of the discrete model (3.12) and π_d its invariant measure.
   - $\theta_0 = \arg\min_{\theta\in\Theta} \mathbb{E}_{\pi_c}\left[M_{N,h}(z_h^{Nh};\theta)\right]$. Here π_c is the stationary measure of the continuous process (1.1).

(3) For the continuous process (1.1) let $Q_{c,h}$ denote the stationary measure of the pair $(z(t), z(t+h))$ given by $p(z(t+h) \mid z(t))\pi_c(z(t))$, where $p(\cdot \mid \cdot)$ is the density of the transition kernel of the continuous model (1.1).

In order to proof the consistency of the likelihood estimator we need to set the following technical hypotheses: for h_0 small enough

H5 $\mathbb{E}\left[e^{(2h_0\|J(z_0;\theta)\|)}\,\|J(z_0;\theta)\|^4\,\|f(z_0;\theta)\|^2\right] < \infty$,

H5′ $\mathbb{E}\left[\sup_{\theta\in\Theta}\left\{e^{(2h_0\|J(z_0;\theta)\|)}\,\|J(z_0;\theta)\|^4\,\|f(z_0;\theta)\|^2\right\}\right] < \infty$, and

H5″ $\mathbb{E}\left[\sup_{\theta\in\Theta}\left\{||J(z_0;\theta)||e^{h_0||J(z_0;\theta)||}\right\}\right] < \infty$

Remark 4.2. In the general case for g_1 and g_2 we do not dare to assert that these conditions hold. Nevertheless, in the case of the Linear, Kramer's, Van der Pol and FitzHug-Nagumo oscillators these hypotheses are fulfilled and the proof can be found in Appendix A.

The idea is to prove that the estimator $\hat{\theta}_{N,\,h}$ is consistent in the following sense: if $h \to 0$ and $hN \to \infty$, then

$$\lim_{N\to\infty} \hat{\theta}_{N,\,h} \to \theta_0 \qquad \text{in probability.} \tag{4.1}$$

In order to prove the above statement we will use a similar technique as the used in Wald [20] for proving the consistency of an estimator. First we need to prove that $\lim_{h\to 0, hN\to\infty} \frac{1}{h}[M_{N,h}(z_h^{Nh};\theta) - M_{N,h}(z_h^{Nh};\theta_0)]$ exists and it is a deterministic function. We need that $h \to 0$ and $hN \to \infty$ simultaneously, so h depends on N. Next we prove (4.1). An antecedent of our proof in the case of ordinary diffusions can be found in the work of Genon-Catalot [5].

Before we begin with the proof of the main result we establish the following result whose proof is left to Appendix A. Besides we will review the expression of the contrast (3.14) to rewrite it conveniently.

Lemma 4.3. *Let $z(s) = (x(s), y(s))'$ satisfies the equation (3.1). Then*

$$\mathbb{E}\left[|x(s) - x(s')|^{2k}\right] = O(|s-s'|^{3k}) \quad \textit{and} \quad \mathbb{E}\left[|y(s) - y(s')|^{2k}\right] = O(|s-s'|^{k}), \tag{4.2}$$

for all positive integer k.

The contrast (3.14) can be rewritten as

$$\begin{aligned} M_{N,h}(z_h^{Nh};\theta) &= \frac{1}{2N}\sum_{i=0}^{N-1}\left\|B_{ih}^{-1}(h;\theta)\left[z_{(i+1)h} - A_{ih}(h;\theta)\, z_{ih}\right]\right\|^2 \\ &= \frac{1}{2N}\sum_{i=0}^{N-1}\left\langle \Sigma_{ih}^{-1}(h;\theta)\left(z_{(i+1)h} - A_{ih}(h;\theta))z_{ih}\right), z_{(i+1)h} - A_{ih}(h;\theta)z_{ih}\right\rangle. \end{aligned} \tag{4.3}$$

Regarding the term $z_{(i+1)h} - A_{ih}(h;\theta)z_{ih}$ in (4.3), as $z_{(i+1)h}$ satisfies the observation equation (3.7) with the true parameters θ_0, we have

$$z_{(i+1)h} - A_{ih}(h;\theta)z_{ih} = \int_{ih}^{(i+1)h} f(z(s);\theta_0)\mathrm{d}s + \begin{bmatrix} 0 \\ \sigma(W_{(i+1)h} - W_{ih}) \end{bmatrix} + (I - A_{ih}(h;\theta))z_{ih}. \tag{4.4}$$

By using the Taylor expansion of $(I - A_{ih}(h;\theta))z_{ih}$ it follows from (4.4)

$$\begin{aligned} z_{(i+1)h} - A_{ih}(h;\theta)z_{ih} &= \int_{ih}^{(i+1)h} f(z(s);\theta_0)\mathrm{d}s + \begin{bmatrix} 0 \\ \sigma(W_{(i+1)h} - W_{ih}) \end{bmatrix} - hf(z_{ih};\theta) \\ &\quad -\frac{h^2}{2}J(z_{ih},\theta)f(z_{ih};\theta) - \sum_{k=3}^{\infty}\frac{J^{k-1}(z_{ih},\theta)}{k!}h^{k-3}f(z_{ih};\theta) \\ &= \int_{ih}^{(i+1)h} [f(z(s);\theta_0) - f(z_{ih};\theta)]\mathrm{d}s + \begin{bmatrix} 0 \\ \sigma(W_{(i+1)h} - W_{ih}) \end{bmatrix} \\ &\quad -\frac{h^2}{2}J(z_{ih},\theta)f(z_{ih};\theta) - \sum_{k=3}^{\infty}\frac{J^{k-1}(z_{ih},\theta)}{k!}h^{k-3}f(z_{ih};\theta). \end{aligned} \tag{4.5}$$

Moreover, as

$$J(z_{ih};\theta) = \begin{bmatrix} 0 & 1 \\ -g_1'(x_{ih};a)y_{ih} - g_2'(x_{ih};b)x_{ih} - g_2(x_{ih};b) & -g_1(x_{ih};a) \end{bmatrix} := \begin{bmatrix} 0 & 1 \\ P_x(z_{ih};\theta) & P_y(z_{ih};\theta) \end{bmatrix}$$

we obtain

$$-\frac{h^2}{2}J(z_{ih},\theta)f(z_{ih};\theta) = \frac{h^2}{2}\begin{bmatrix} -P(z_{ih};\theta) \\ \ell_1(z_{ih};\theta) \end{bmatrix}, \tag{4.6}$$

where $\ell_1(z_{ih};\theta) = g_1(x_{ih};\theta)P(z_{ih};\theta) - y_{ih}P_x(z_{ih};\theta)$. It can be easily checked from hypotheses H1 and H2 that

$$|\ell_1(z_{ih};\theta)| \leq \mathbf{C}_\theta\left(\|z_{ih}\|^{2\gamma_1+1} + \|z_{ih}\|^{\gamma_1+\gamma_2+1} + \|z_{ih}\|^{\gamma_1+2} + \|z_{ih}\|^{\gamma_2+2}\right) := \mathbf{C}_\theta\|z_{ih}\|^{\gamma_4}, \tag{4.7}$$

for $z \in K$. Moreover, as the polynomial moments of the process z_{ih} are bounded we have

$$\mathbb{E}\left[|\ell_1(z_{ih};\theta)|\right] \leq \mathbf{C}_\theta\mathbb{E}\left[\|z_{ih}\|^{\gamma_4}\right] < \infty. \tag{4.8}$$

Now, recalling that $\frac{h^2}{2} = \int_{ih}^{(i+1)h}[(i+1)h - s]\mathrm{d}s$ and considering the observation equation (3.7) we obtain

$$\begin{aligned} z_{(i+1)h} - A_{ih}(h;\theta)z_{ih} &= \begin{bmatrix} \sigma\int_{ih}^{(i+1)h}[(i+1)h-u]\mathrm{d}W_u - \int_{ih}^{(i+1)h}[(i+1)h-u]\left(P(z(u);\theta_0) - P(z_{ih};\theta)\right)\mathrm{d}u \\ \int_{ih}^{(+1)h}\left(P(z(s);\theta_0) - P(z_{ih};\theta)\right)\mathrm{d}s + \sigma\int_{ih}^{(i+1)h}\mathrm{d}W_s + \frac{h^2}{2}\ell_1(z_{ih};\theta) \end{bmatrix} \\ &\quad -\sum_{k=3}^{\infty}\frac{J^{k-1}(z_{ih},\theta)}{k!}h^k f(z_{ih};\theta) \\ &:= \mathcal{W}_i(\theta) + \mathcal{K}_i(\theta) - \sum_{k=3}^{\infty}\frac{J^{k-1}(z_{ih},\theta)}{k!}h^k f(z_{ih};\theta), \end{aligned} \tag{4.9}$$

where

$$\mathcal{W}_i(\theta) = \begin{bmatrix} \sigma \int_{ih}^{(i+1)h} [(i+1)h-u] \mathrm{d}W_u - \int_{ih}^{(i+1)h} [(i+1)h-u] \left(P(z(u);\theta_0) - P(z_{ih};\theta)\right) \mathrm{d}u \\ \int_{ih}^{(+1)h} \left(P(z(s);\theta_0) - P(z_{ih};\theta)\right) \mathrm{d}s + \sigma \int_{ih}^{(i+1)h} \mathrm{d}W_s \end{bmatrix}$$
$$:= \begin{bmatrix} \mathcal{W}_{i1}(\theta) \\ \mathcal{W}_{i2}(\theta) \end{bmatrix}.$$

and

$$\mathcal{K}_i(\theta) = \begin{bmatrix} 0 \\ \frac{h^2}{2}\ell_1(z_{ih};\theta) \end{bmatrix}.$$

The calculations above lead to the following expression for the contrast (4.3)

$$\begin{aligned} & M_{N,h}(z_h^{Nh};\theta) \\ &= \frac{1}{2N}\sum_{i=0}^{N-1}\left\langle \varSigma_{ih}^{-1}(h;\theta)\left[\mathcal{W}_i(\theta)+\mathcal{K}_i(\theta)-\sum_{k=3}^{\infty}\frac{J^{k-1}(z_{ih},\theta)}{k!}h^k f(z_{ih};\theta)\right], \mathcal{W}_i(\theta)+\mathcal{K}_i(\theta)-\sum_{k=3}^{\infty}\frac{J^{k-1}(z_{ih},\theta)}{k!}h^k f(z_{ih};\theta)\right\rangle \\ &= \frac{1}{2N}\sum_{i=0}^{N-1}\left\langle \varSigma_{ih}^{-1}(h;\theta)(\mathcal{W}_i(\theta)+\mathcal{K}_i(\theta)), \mathcal{W}_i(\theta)+\mathcal{K}_i(\theta)\right\rangle \\ &\quad + \frac{1}{2N}\sum_{i=0}^{N-1}\left\langle \varSigma_{ih}^{-1}(h;\theta)\sum_{k=3}^{\infty}\frac{J^{k-1}(z_{ih},\theta)}{k!}h^k f(z_{ih};\theta), \sum_{k=3}^{\infty}\frac{J^{k-1}(z_{ih},\theta)}{k!}h^k f(z_{ih};\theta)\right\rangle \\ &\quad - \frac{1}{N}\sum_{i=0}^{N-1}\left\langle \varSigma_{ih}^{-1}(h;\theta)(\mathcal{W}_i(\theta)+\mathcal{K}_i(\theta)), \sum_{k=3}^{\infty}\frac{J^{k-1}(z_{ih},\theta)}{k!}h^k f(z_{ih};\theta)\right\rangle \\ &:= \mathcal{D}_1(\theta)+\mathcal{D}_2(\theta)-\mathcal{D}_3(\theta). \end{aligned} \tag{4.10}$$

From the calculations above we set the following

Lemma 4.4. *Under the hypotheses $H1$, $H2$, $H3$ and $H5$ we have*

$$\frac{1}{h}(\mathcal{D}_2(\theta)-\mathcal{D}_3(\theta)) = o_{\mathbb{P}}(1)$$

Proof. Working with the expression $\mathcal{D}_2(\theta)$ it yields

$$\begin{aligned} |\mathcal{D}_2(\theta)| &\le h^6 ||\varSigma_{ih}^{-1}(h;\theta)|| \frac{1}{2N}\sum_{i=0}^{N-1}\left(\sum_{k=3}^{\infty}\frac{||J(z_{ih};\theta)||^{k-1}h^{k-3}}{k!}||f(z_{ih};\theta)||\right)^2 \\ &\le h^6 ||\varSigma_{ih}^{-1}(h;\theta)|| \frac{1}{2N}\sum_{i=0}^{N-1}\left(\sum_{k=3}^{\infty}\frac{||J(z_{ih};\theta)||^{k-3}h^{k-3}}{k!}||J(z_{ih};\theta)||^2||f(z_{ih};\theta)||\right)^2 \\ &\le h^6 ||\varSigma_{ih}^{-1}(h;\theta)|| \frac{1}{2N}\sum_{i=0}^{N-1} e^{(2h||J(z_{ih};\theta)||)}||J(z_{ih})||^4||f(z_{ih};\theta)||^2. \end{aligned} \tag{4.11}$$

By hypothesis H5 we get

$$\begin{aligned}\mathbb{E}\left[\frac{1}{N}\sum_{i=0}^{N-1} e^{(2h||J(z_{ih};\theta)||)}||J(z_{ih};\theta)||^4||f(z_{ih};\theta)||^2\right] &= \mathbb{E}[e^{(2h||J(z_0;\theta)||)}||J(z_0;\theta)||^4||f(z_0;\theta)||^2] \\ &\leq \mathbb{E}[e^{(2h_0||J(z_0;\theta)||)}||J(z_0;\theta)||^4||f(z_0;\theta)||^2] < \infty,\end{aligned}$$

and from the Lemma 3.3 we have $\frac{h^3}{2}||\Sigma_{ih}^{-1}(h;\theta)|| = O(1)$. This two facts yield first, using the Chebyshev inequality, that

$$\frac{1}{N}\sum_{i=0}^{N-1} e^{(2h||J(z_{ih};\theta)||)}||J(z_{ih};\theta)||^4||f(z_{ih};\theta)||^2 = O_{\mathbb{P}}(1),$$

this is, the term is bounded in probability. Moreover we obtain

$$|\mathcal{D}_2(\theta)| = O(h^3), \tag{4.12}$$

and therefore

$$\frac{|\mathcal{D}_2(\theta)|}{h} = O(h^2).$$

On the other hand, let us consider the term $\mathcal{D}_3(\theta)$. In Remark 4.6 below we prove that

$$\frac{1}{N}\sum_{i=0}^{N-1}\left\langle \Sigma_{ih}^{-1}(h;\theta)(\mathcal{W}_i(\theta)+\mathcal{K}_i(\theta)), \mathcal{W}_i(\theta)+\mathcal{K}_i(\theta)\right\rangle = O_{\mathbb{P}}(1).$$

Now by using Cauchy-Schwarz inequality twice (remember that $\left\langle \Sigma^{-1}w, w\right\rangle = \left\langle \Sigma^{-\frac{1}{2}}w, \Sigma^{-\frac{1}{2}}w\right\rangle = ||\Sigma^{-\frac{1}{2}}w||^2$) we get

$$\begin{aligned}|\mathcal{D}_3(\theta)| \leq &\left(\frac{1}{N}\sum_{i=0}^{N-1}\left\langle \Sigma_{ih}^{-1}(h;\theta)(\mathcal{W}_i(\theta)+\mathcal{K}_i(\theta)), \mathcal{W}_i(\theta)+\mathcal{K}_i(\theta)\right\rangle\right)^{\frac{1}{2}} \\ &\times\left(\frac{1}{N}\sum_{i=0}^{N-1}\left\langle \Sigma_{ih}^{-1}(h;\theta)\sum_{k=3}^{\infty}\frac{J^{k-1}}{k!}h^k f(z_{ih};\theta), \sum_{k=3}^{\infty}\frac{J^{k-1}(z_{ih};\theta)}{k!}h^k f(z_{ih};\theta)\right\rangle\right)^{\frac{1}{2}}.\end{aligned}$$

But using (A.13) we obtain $\frac{|\mathcal{D}_3(\theta)|}{h} = O_{\mathbb{P}}(1)O(h^{\frac{1}{2}})$, that tends to zero. In this way the following approximation is obtained

$$M_{N,h}(z_h^{Nh};\theta) = \frac{1}{2N}\sum_{i=0}^{N-1}\left\langle \Sigma_{ih}^{-1}(h;\theta)(\mathcal{W}_i(\theta)+\mathcal{K}_i(\theta)), \mathcal{W}_i(\theta)+\mathcal{K}_i(\theta)\right\rangle + O_{\mathbb{P}}(h^{\frac{3}{2}}).$$

□

In view of the observations above and Lemmas 4.4 and 4.3 the following result holds

Theorem 4.5. *Let $\theta \in \Theta$. Then, under the hypotheses H1–H5, for $h \to 0$, $N \to \infty$, $hN \to \infty$*

$$\frac{1}{h}\left(M_{N,h}(z_h^{Nh};\theta) - M_{N,h}(z_h^{Nh};\theta_0)\right) \to \frac{13}{2}\mathbb{E}[(P(z_0,\theta) - P(z_0;\theta_0))^2] := M(\theta),$$

in probability.

Proof. In virtue of the results in Lemma 4.4, for proving the theorem it is enough to study

$$\frac{1}{2hN}\left(\sum_{i=0}^{N-1}\left\langle \Sigma_{ih}^{-1}(h;\theta)(\mathcal{W}_i(\theta)+\mathcal{K}_i(\theta)), \mathcal{W}_i(\theta)+\mathcal{K}_i(\theta)\right\rangle - \sum_{i=0}^{N-1}\left\langle \Sigma_{ih}^{-1}(h;\theta_0)(\mathcal{W}_i(\theta)+\mathcal{K}_i(\theta)), \mathcal{W}_i(\theta)+\mathcal{K}_i(\theta)\right\rangle\right).$$

For each $\theta \in \Theta$ we develop each addend of the first sum obtaining

$$\frac{1}{2hN}\sum_{i=0}^{N-1}\left\langle \Sigma_{ih}^{-1}(h;\theta)\mathcal{W}_i(\theta), \mathcal{W}_i(\theta)\right\rangle + \frac{1}{2hN}\sum_{i=0}^{N-1}\left\langle \Sigma_{ih}^{-1}(h;\theta)\mathcal{K}_i(\theta), \mathcal{K}_i(\theta)\right\rangle + \frac{1}{hN}\sum_{i=0}^{N-1}\left\langle \Sigma_{ih}^{-1}(h;\theta)\mathcal{W}_i(\theta), \mathcal{K}_i(\theta)\right\rangle. \tag{4.13}$$

We begin with the second term in (4.13)

$$\begin{aligned}\frac{1}{2hN}\sum_{i=0}^{N-1}\left\langle \Sigma_{ih}^{-1}(h;\theta)\mathcal{K}_i(\theta), \mathcal{K}_i(\theta)\right\rangle &= \frac{1}{2hN}\sum_{i=0}^{N-1}\left\langle \begin{bmatrix}\frac{h^2}{2}b_{12}(h)\ell_1(z_{ih};\theta)\\ \frac{h^2}{2}b_{22}(h)\ell_1(z_{ih};\theta)\end{bmatrix}, \begin{bmatrix}0\\ \frac{h^2}{2}\ell_1(z_{ih};\theta)\end{bmatrix}\right\rangle\\ &= h^3 b_{22}\frac{1}{8N}\sum_{i=0}^{N-1}\ell_1^2(z_{ih};\theta).\end{aligned} \tag{4.14}$$

By taking the expectation it follows

$$\frac{1}{2hN}\sum_{i=0}^{N-1}\mathbb{E}\left[\left\langle \Sigma_{ih}^{-1}(h;\theta)\mathcal{K}_i(\theta), \mathcal{K}_i(\theta)\right\rangle\right] = h^3 b_{22}\frac{1}{8N}\sum_{i=0}^{N-1}\mathbb{E}\left[\ell_1^2(z_{ih};\theta)\right]. \tag{4.15}$$

By the stationarity of the process $\{z_{ih}\}$ it follows from (4.15)

$$h^3 b_{22}\frac{1}{8N}\sum_{i=0}^{N-1}\mathbb{E}\left[\ell_1^2(z_{ih};\theta)\right] = h^2 h b_{22}\frac{1}{8}\mathbb{E}(\ell_1^2(z_0;\theta)) \to 0,$$

for $h \to 0$, in L^1 and therefore in probability.

The third term in (4.13) can be written as

$$\frac{1}{hN}\sum_{i=0}^{N-1}\left\langle \Sigma_{ih}^{-1}(h;\theta)\mathcal{W}_i(\theta), \mathcal{K}_i(\theta)\right\rangle = \frac{h}{N}\sum_{i=0}^{N-1}\left(b_{12}\mathcal{W}_{i1}(\theta)\ell_1(z_{ih};\theta) + b_{22}\mathcal{W}_{i2}(\theta)\ell_1(z_{ih};\theta)\right). \tag{4.16}$$

For the first addend in (4.16) we have

$$\frac{h}{N}\sum_{i=0}^{N-1} b_{12}\mathcal{W}_{i1}(\theta)\ell_1(z_{ih};\theta) = \frac{h}{N}\sum_{i=0}^{N-1} b_{12}\int_{ih}^{(i+1)h}[(i+1)h-u]\mathrm{d}W_u\ell_1(z_{ih};\theta)$$
$$-\frac{h}{N}\sum_{i=0}^{N-1} b_{12}\int_{ih}^{(i+1)h}[(i+1)h-u][P(z(u);\theta_0)-P(z_{ih};\theta)]\mathrm{d}u\ell_1(z_{ih};\theta).$$

By taking expectation and using Cauchy-Schwarz inequality, stationarity, finiteness of the polynomial moments of the process z_0, the estimates in (2.3) (4.8) and from Lemma 3.3 we obtain

$$\mathbb{E}\left[\left|\frac{h}{N}\sum_{i=0}^{N-1} b_{12}\mathcal{W}_{i1}(\theta)\ell_1(z_{ih};\theta)\right|\right] \leq \mathbf{C}\left\{|b_{12}|h^2\left[\mathbb{E}\ell_1^2(z_0;\theta)\right]^{\frac{1}{2}}h^{\frac{1}{2}} + |b_{12}|h^2[\mathbb{E}P^2(z_0;\theta_0)]^{\frac{1}{2}}\right\}h$$
$$= o(1). \tag{4.17}$$

Proceeding in the same way as above we get

$$\mathbb{E}\left[\left|\frac{h}{N}\sum_{i=0}^{N-1} b_{22}\mathcal{W}_{i2}(\theta)\ell_1(z_{ih};\theta)\right|\right] \leq \mathbf{C}\left\{|b_{22}|h[\mathbb{E}\ell_1^2(z_0;\theta)]^{\frac{1}{2}}h^{\frac{1}{4}} + |b_{22}|h[\mathbb{E}P^2(z_0;\theta_0)]^{\frac{1}{2}}\right\}h$$
$$= o(1). \tag{4.18}$$

From (4.17) and (4.18) we conclude

$$\mathbb{E}\left[\left|\frac{1}{hN}\sum_{i=0}^{N-1}\left\langle \varSigma_{ih}^{-1}(h;\theta)\mathcal{W}_i(\theta),\mathcal{K}_i(\theta)\right\rangle\right|\right] = o(1). \tag{4.19}$$

In this way we can write

$$\frac{1}{h}\left(M_{N,h}(z_h^{Nh};\theta) - M_{N,h}(z_h^{Nh};\theta_0)\right) = \frac{1}{2hN}\sum_{i=0}^{N-1}\left\langle \varSigma_{ih}^{-1}(h;\theta)\mathcal{W}_i(\theta),\mathcal{W}_i(\theta)\right\rangle$$
$$-\frac{1}{2hN}\sum_{i=0}^{N-1}\left\langle \varSigma_{ih}^{-1}(h;\theta_0)\mathcal{W}_i(\theta_0),\mathcal{W}_i(\theta_0)\right\rangle + o_{\mathbb{P}}(1).$$

Let us develop the term

$$\frac{1}{2hN}\sum_{i=0}^{N-1}\left\langle \varSigma_{ih}^{-1}(h;\theta)\mathcal{W}_i(\theta),\mathcal{W}_i(\theta)\right\rangle.$$

We have

$$\frac{1}{2hN}\sum_{i=0}^{N-1}\left\langle \varSigma_{ih}^{-1}(h;\theta)\mathcal{W}_i(\theta),\mathcal{W}_i(\theta)\right\rangle = \frac{b_{11}(h)}{2hN}\sum_{i=0}^{N-1}\mathcal{W}_{i1}^2(\theta) + \frac{b_{12}(h)}{hN}\sum_{i=0}^{N-1}\mathcal{W}_{i1}(\theta)\mathcal{W}_{i2}(\theta) + \frac{b_{22}(h)}{2hN}\sum_{i=0}^{N-1}\mathcal{W}_{12}^2(\theta).$$

Let us denote by $\mathcal{G}_{jN}(\theta)$, $j = 1, \ldots, 10$ each term in the developing of the expression above. Then

$$\mathcal{G}_{1N}(\theta) = \frac{b_{11}(h)}{2hN} \sum_{i=0}^{N-1} \left(\int_{ih}^{(i+1)h} [(i+1)h - u][P(z(u); \theta_0) - P(z_{ih}; \theta)] \mathrm{d}u \right)^2.$$

$$\mathcal{G}_{2N}(\theta) = -\frac{\sigma b_{11}(h)}{hN} \sum_{i=0}^{N-1} \int_{ih}^{(i+1)h} \int_{ih}^{(i+1)h} [(i+1)h - u_1][(i+1)h - u_2][P(z(u_2); \theta_0) - P(z_{ih}; \theta)] \mathrm{d}W_{u_1} \mathrm{d}u_2.$$

$$\mathcal{G}_{3N}(\theta) = \frac{\sigma b_{12}(h)}{hN} \sum_{i=0}^{N-1} \int_{ih}^{(i+1)h} [(i+1)h - u] \mathrm{d}W_u \int_{ih}^{(i+1)h} [P(z(s); \theta_0) - P(z_{ih}; \theta)] \mathrm{d}s.$$

$$\mathcal{G}_{4N}(\theta) = -\frac{b_{12}(h)}{hN} \sum_{i=0}^{N-1} \int_{ih}^{(i+1)h} [P(z(s); \theta_0) - P(z_{ih}; \theta)] \mathrm{d}s \int_{ih}^{(i+1)h} [(i+1)h - u][P(z(u); \theta_0) - P(z_{ih}; \theta)] \mathrm{d}u.$$

$$\mathcal{G}_{5N}(\theta) = -\frac{b_{12}(h)}{hN} \sum_{i=0}^{N-1} \int_{ih}^{(i+1)h} \mathrm{d}W_s \int_{ih}^{(i+1)h} [(i+1)h - u][P(z(u); \theta_0) - P(z_{ih}; \theta)] \mathrm{d}u.$$

$$\mathcal{G}_{6N}(\theta) = \frac{b_{22}(h)}{2hN} \sum_{i=0}^{N-1} \left(\int_{ih}^{(i+1)h} [P(z(s); \theta_0) - P(z_{ih}; \theta)] \mathrm{d}s \right)^2.$$

$$\mathcal{G}_{7N}(\theta) = \frac{b_{22}(h)}{hN} \sum_{i=0}^{N-1} \int_{ih}^{(i+1)h} \mathrm{d}W_s \int_{ih}^{(i+1)h} [P(z(s); \theta_0) - P(z_{ih}; \theta)] \mathrm{d}s.$$

$$\mathcal{G}_{8N}(\theta) = \frac{\sigma^2 b_{11}(h)}{2hN} \sum_{i=0}^{N-1} \left(\int_{ih}^{(i+1)h} [(i+1)h - u] \mathrm{d}W_u \right)^2.$$

$$\mathcal{G}_{9N}(\theta) = \frac{\sigma b_{12}(h)}{hN} \sum_{i=0}^{N-1} \int_{ih}^{(i+1)h} [(i+1)h - u] \mathrm{d}W_u \int_{ih}^{(i+1)h} \mathrm{d}W_s.$$

$$\mathcal{G}_{10N}(\theta) = \frac{b_{22}(h)}{hN} \sum_{i=0}^{N-1} \left(\int_{ih}^{(i+1)h} \mathrm{d}W_s \right)^2.$$

It is plain to see that $\frac{1}{h}\left(M_{N,h}(z_h^{Nh}; \theta) - M_{N,h}(z_h^{Nh}; \theta_0)\right) = \sum_{j=1}^{7} \mathcal{G}_{jN}(\theta) - \sum_{j=1}^{7} \mathcal{G}_{jN}(\theta_0) + o_{\mathbb{P}}(1)$, since the terms that only depend on the noise become canceled. Henceforth we only consider the remaining terms. Let us begin with the term $\mathcal{G}_{1N}$. It can be written as

$$\begin{aligned}
\mathcal{G}_{1N}(\theta) = &\frac{h^3 b_{11}(h)}{8N} \sum_{i=0}^{N-1} [P(z_{ih}; \theta_0) - P(z_{ih}; \theta)]^2 \\
&+ \frac{b_{11}(h)}{2hN} \sum_{i=0}^{N-1} \big(\int_{ih}^{(i+1)h} [(i+1)h - u]\{P_x(\tilde{z}(u); \theta_0)[x(u) - x_{ih}] + P_y(\tilde{z}(u); \theta_0)[y(u) - y_{ih}]\} \mathrm{d}u \big)^2 \\
&+ \frac{h^2 b_{11}(h)}{2hN} \sum_{i=0}^{N-1} [P(z_{ih}; \theta_0) - P(z_{ih}; \theta)] \\
&\times \int_{ih}^{(i+1)h} [(i+1)h - u]\{P_x(\tilde{z}(u); \theta_0)[x(u) - x_{ih}] + P_y(\tilde{z}(u); \theta_0)[y(u) - y_{ih}]\} \mathrm{d}u \\
= &\mathcal{G}_{1N,1}(\theta) + \mathcal{G}_{1N,2}(\theta) + \mathcal{G}_{1N,3}(\theta),
\end{aligned}$$

where $\tilde{z}(u) = z_{ih} + \lambda(z(u) - z_{ih})$ for some $0 \le \lambda \le 1$. For the first term, using the Ergodic Theorem, the asymptotic behavior for $b_{11}(h)$ in Lemma 3.3 and the stationarity of the process $\{z_{ih}\}$ we have

$$\begin{aligned}\mathcal{G}_{1N,1}(\theta) &= \frac{b_{11}(h)}{2hN}\sum_{i=0}^{N-1}[P(z_{ih};\theta_0) - P(z_{ih};\theta)]^2\left(\int_{ih}^{(i+1)h}[(i+1)h - u]\mathrm{d}u\right)^2 \\ &= \frac{h^3 b_{11}(h)}{8N}\sum_{i=0}^{N-1}[P(z_{ih};\theta_0) - P(z_{ih};\theta)]^2 \to \frac{3}{2}E\left\{[P(z_0;\theta_0) - P(z_0;\theta)]^2\right\}, \end{aligned} \tag{4.20}$$

for $h \to 0$ and $N \to \infty$. For the second term, using $(a+b)^2 \le 2(a^2+b^2)$ and the Cauchy-Schwarz inequality we have

$$\begin{aligned}\mathcal{G}_{1N,2}(\theta) &= \frac{b_{11}(h)}{2hN}\sum_{i=0}^{N-1}\left(\int_{ih}^{(i+1)h}[(i+1)h-u]\left\{P_x(\tilde{z}(u);\theta_0)[x(u)-x_{ih}] + P_y(\tilde{z}(u);\theta_0)[y(u)-y_{ih}]\right\}\mathrm{d}u\right)^2 \\ &\le \frac{\mathbf{C}_\theta h^3 b_{11}(h)}{hN}\sum_{i=0}^{N-1}\int_{ih}^{(i+1)h}\left(P_x^2(\tilde{z}(u);\theta_0)[x(u)-x_{ih}]^2 + P_y^2(\tilde{z}(u);\theta_0)[y(u)-y_{ih}]^2\right)\mathrm{d}u.\end{aligned}$$

Recalling the bounds for P_x and P_y in (2.3) it follows

$$\begin{aligned}\mathcal{G}_{1N,2}(\theta) \le \frac{\mathbf{C}_\theta h^3 b_{11}(h)}{hN}\sum_{i=0}^{N-1}\int_{ih}^{(i+1)h} &\Big\{\|\tilde{z}(u)\|^{2\gamma_3}[x(u)-x_{ih}]^2 \\ &+ \|\tilde{z}(u)\|^{2\gamma_1}[y(u)-y_{ih}]^2\Big\}\mathrm{d}u.\end{aligned}$$

Since $\tilde{z}(u) = \lambda z(u) + (1-\lambda)z_{ih}$, then $\|\tilde{z}(u)\| = \|\lambda z(u) + (1-\lambda)z_{ih}\|$. We define the polynomial

$$Q(||z(u)||, ||z_{ih}||) = \|\tilde{z}(u)\|^{2\gamma_3} + \|\tilde{z}(u)\|^{2\gamma_1}.$$

Besides, as $|x(u) - x_{ih}|^2 \le \|z(u) - z_{ih}\|^2$ and $|x(u) - x_{ih}|^2 \le \|z(u) - z_{ih}\|^2$, it results

$$|\mathcal{G}_{1N,2}(\theta)| \le \frac{\mathbf{C}_\theta h^3 |b_{11}(h)|}{hN}\sum_{i=0}^{N-1}\int_{ih}^{(i+1)h}|Q(||z(u)||, ||z_{ih}||)|\,\|z(u)-z_{ih}\|^2\mathrm{d}u.$$

Taking expectation and using Cauchy-Schwarz inequality we obtain

$$\begin{aligned}\mathbb{E}\left[|\mathcal{G}_{1N,2}(\theta)|\right] \le \frac{\mathbf{C}_\theta h^3 |b_{11}(h)|}{hN}\sum_{i=0}^{N-1}&\left\{\int_{ih}^{(i+1)h}\mathbb{E}\left(|Q(||z(u)||, ||z_{ih}||)|\right)^2\mathrm{d}u\right\}^{\frac{1}{2}} \\ &\times\left\{\int_{ih}^{(i+1)h}\mathbb{E}\|z(u)-z_{ih}\|^2\mathrm{d}u\right\}^{\frac{1}{2}}.\end{aligned}$$

Finally, as the polynomial moments of the processes $z(u)$ and z_{ih} are finite, using the bound $\mathbb{E}\|z(u) - z_{ih}\|^2 = O(h)$, the stationarity and the relations in Lemma 3.3, we conclude

$$\mathbb{E}\left[|\mathcal{G}_{1N,2}(\theta)|\right] \le \mathbf{C}_\theta h^3 |b_{11}(h)|\left[\mathbb{E}\left(\left|Q'(||z_0||)\right|\right)^2\right]^{\frac{1}{2}} h^{\frac{1}{2}} \to 0, \tag{4.21}$$

for $h \to 0$. In the case of the third term $\mathcal{G}_{1N,3}(\theta)$, due to its form the convergence to zero can be inferred from the facts that $\mathcal{G}_{1N,1}(\theta)$ is bounded and $\mathcal{G}_{1N,2}(\theta)$ converges to zero in probability, that is

$$\mathcal{G}_{1N,3}(\theta) \to 0, \tag{4.22}$$

in probability. In view of the results in (4.20)–(4.22) we have, for $h \to 0$ and $N \to \infty$

$$\mathcal{G}_{1N} \to \frac{3}{2} E\left\{[P(z_0;\theta_0) - P(z_0;\theta)]^2\right\}, \tag{4.23}$$

in probability.

Next we study the term $\mathcal{G}_{2N}$. We define the approximation of $\mathcal{G}_{2N}$ by

$$\tilde{\mathcal{G}}_{2N,1} = -\frac{\sigma b_{11} h}{2N} \sum_{i=0}^{N-1} [P(z_{ih};\theta_0) - P(z_{ih};\theta)] \int_{ih}^{(i+1)h} ((i+1)h - u_1) dW_{u_1}.$$

Conditioning with respect to the σ–algebra $\mathfrak{F}_{u_1} := \sigma(z_t : t \leq u_1)$ it is easy to show that $\mathbb{E}[\tilde{\mathcal{G}}_{2N}] = 0$. Moreover

$$\mathbb{E}[\tilde{\mathcal{G}}_{2N,1}^2] = b_{11}^2 h^6 \frac{\sigma^2}{8hN} \mathbb{E}[(P(z_0;\theta_0) - P(z_0;\theta))^2] \to 0.$$

This last assertion is because always we have $hN \to \infty$. In order to study the original random variable $\mathcal{G}_{2N}$, consider the first order Taylor approximation

$$P(z(u_2);\theta_0) = P(z_{ih};\theta_0) + P_x(\tilde{z}(u_2);\theta_0)[x(u_2) - x_{ih}] + P_y(\tilde{z}(u_2);\theta_0)[y(u_2) - y_{ih}],$$

where $\tilde{z}(u_2) = z_{ih} + \lambda_1(z(u_2) - z_{ih})$ for some $0 \leq \lambda_1 \leq 1$. In this way, we have

$$\mathcal{G}_{2N} = \tilde{\mathcal{G}}_{2N,1} + \tilde{\mathcal{G}}_{2N,2} + \tilde{\mathcal{G}}_{2N,3},$$

where

$$\tilde{\mathcal{G}}_{2N,2} = -b_{11} \frac{\sigma}{hN} \sum_{i=0}^{N-1} \int_{ih}^{(i+1)h} \int_{ih}^{(i+1)h} ((i+1)h - u_1)((i+1)h - u_2) P_x(\tilde{z}(u_2);\theta_0)(x(u_2) - x_{ih}) \mathrm{d}W_{u_1} \mathrm{d}u_2,$$

and

$$\tilde{\mathcal{G}}_{2N,3} = b_{11} \frac{\sigma}{hN} \sum_{i=0}^{N-1} \int_{ih}^{(i+1)h} \int_{ih}^{(i+1)h} ((i+1)h - u_1)((i+1)h - u_2) g_1(\tilde{x}(u_2);\theta_0)(y(u_2) - y_{ih}) \mathrm{d}W_{u_1} \mathrm{d}u_2.$$

Using the Lemma 4.3, the polynomial bounds of g_1, g_2 and their derivatives, we can show that $\mathbb{E}|\tilde{\mathcal{G}}_{2N,2}| = O(h)$. For the term $\tilde{\mathcal{G}}_{2N,3}$, if we define

$$\tilde{\mathcal{G}}_{2N,4} = b_{11} \frac{\sigma}{hN} \sum_{i=0}^{N-1} \int_{ih}^{(i+1)h} \int_{ih}^{(i+1)h} ((i+1)h - u_1)((i+1)h - u_2) g_1(x_{ih};\theta_0)(y(u_2) - y_{ih}) \mathrm{d}W_{u_1} \mathrm{d}u_2,$$

the same procedure that for $\tilde{\mathcal{G}}_{2N,2}$ gives $\mathbb{E}\left|\tilde{\mathcal{G}}_{2N,3} - \tilde{\mathcal{G}}_{2N,4}\right| = o_{\mathbb{P}}(1)$. By using the definition we get

$$\tilde{\mathcal{G}}_{2N,4} = b_{11}\frac{\sigma}{hN}\sum_{i=0}^{N-1} g_1(x_{ih};\theta_0)\int_{ih}^{(i+1)h}\int_{ih}^{(i+1)h}((i+1)h-u_1)((i+1)h-u_2)(y(u_2)-y_{ih})\mathrm{d}W_{u_1}\mathrm{d}u_2 + o_{\mathbb{P}}(1),$$

and taking expectation we obtain

$$\mathbb{E}[\tilde{\mathcal{G}}_{2N,4}] \to 3\sigma\mathbb{E}[g_1(x_0;\theta_0)].$$

Besides it is easy to show that the variance tends to zero, thus

$$\tilde{\mathcal{G}}_{2N,4} \to 3\sigma\mathbb{E}[g_1(x_0;\theta_0)],$$

in probability. In view of the results above we can conclude

$$\mathcal{G}_{2N} - 3\sigma\mathbb{E}[g_1(x_0;\theta_0)] = \tilde{\mathcal{G}}_{2N,1} + \tilde{\mathcal{G}}_{2N,2} + (\tilde{\mathcal{G}}_{2N,3} - \tilde{\mathcal{G}}_{2N,4}) + \tilde{\mathcal{G}}_{2N,4} - 3\sigma\mathbb{E}[g_1(x_0;\theta_0)],$$

each term tending to zero in probability. Hence for $h \to 0$ and $hN \to \infty$

$$\mathcal{G}_{2N} \to 3\sigma\mathbb{E}[g_1(x_0;\theta_0)], \tag{4.24}$$

in probability.

In the studying of the remaining terms we will only sketch the proofs because of their similarity to the previous ones. Let us focus now on the term $\mathcal{G}_{3N}$. Asymptotically this sequence is equivalent to

$$\tilde{\mathcal{G}}_{3N} = -\frac{\sigma^2 b_{12}(h)}{hN}\sum_{i=0}^{N-1} g_1(x_{ih};\theta_0)\int_{ih}^{(i+1)h}[(i+1)h-u]\mathrm{d}W_u\int_{ih}^{(i+1)h}[(i+1)h-s]\mathrm{d}W_s.$$

From Lemma 3.3 we have

$$\mathbb{E}[\tilde{\mathcal{G}}_{3N}] = -\frac{\sigma^2 h^2 b_{12}(h)}{3}\mathbb{E}[g_1(x_0;\theta_0)] \to 2\sigma^2\mathbb{E}[g_1(x_0;\theta_0)].$$

and its variance tend to 0. Thus for $h \to 0$ and $hN \to \infty$

$$\mathcal{G}_{3N} \to 2\sigma\mathbb{E}[g_1(x_0;\theta_0)], \tag{4.25}$$

in probability.

For the term $\mathcal{G}_{4N}$, in a similar procedure that with the term $\mathcal{G}_{1N}$ we obtain, for $h \to 0$ and $hN \to \infty$

$$\mathcal{G}_{4N} \to 3\mathbb{E}[(P(z_0;\theta_0) - P(z_0;\theta))^2], \tag{4.26}$$

in probability.

For the term

$$\mathcal{G}_{5N} = -\frac{b_{12}(h)}{hN}\sum_{i=0}^{N-1}\int_{ih}^{(i+1)h}\mathrm{d}W_s\int_{ih}^{(i+1)h}[(i+1)h-u][P(z(u);\theta_0) - P(z_{ih};\theta)]\mathrm{d}u,$$

by using the approximation

$$\tilde{\mathcal{G}}_{5N} = -\frac{b_{12}(h)}{hN}\sum_{i=0}^{N-1} g_1(x_{ih};\theta_0)\int_{ih}^{(i+1)h}[(i+1)h-u]\mathrm{d}W_u\int_{ih}^{(i+1)h}[(i+1)h-s]\mathrm{d}W_s,$$

it can be inferred as before that for $h \to 0$ and $hN \to \infty$

$$\mathcal{G}_{5N} \to -2\sigma^2\mathbb{E}[g_1(x_0;\theta_0)], \tag{4.27}$$

in probability.

For the term $\mathcal{G}_{6N}$, in the same fashion as in the term $\mathcal{G}_{1N}$ we have for $h \to 0$ and $hN \to \infty$

$$\mathcal{G}_{6N} \to 2\mathbb{E}[(P(z_0;\theta_0) - P(z_0;\theta))^2], \tag{4.28}$$

in probability.

The term $\mathcal{G}_{7N}$ gives the same limit as

$$\tilde{\mathcal{G}}_{7N} = -\frac{\sigma^2 b_{22}(h)}{hN}\sum_{i=0}^{N-1} g_1(x_{ih};\theta_0)\int_{ih}^{(i+1)h}\mathrm{d}W_{s_1}\int_{ih}^{(i+1)h}[(i+1)h-s_2]\mathrm{d}W_{s_2}.$$

Thus, proceeding in a similar way as with the term $\mathcal{G}_{3N}$ we obtain for $h \to 0$ and $hN \to \infty$

$$\mathcal{G}_{7N} \to -2\sigma^2\mathbb{E}[g_1(x_0;\theta_0)], \tag{4.29}$$

in probability.

Finally, adding the results of equations (4.23)–(4.29) we obtain

$$\frac{1}{h}(M_{N,h}(z_h^{Nh};\theta) - M_{N,h}(z_h^{Nh};\theta_0)) \to \frac{13}{2}\mathbb{E}[(P(z_0,\theta) - P(z_0;\theta_0))^2].$$

□

Remark 4.6. The above procedure give as a bonus that for all θ it holds

$$\frac{1}{N}\sum_{i=0}^{N-1} < \Sigma_{ih}^{-1}(h;\theta)(\mathcal{W}_i(\theta) + \mathcal{K}_i(\theta)), \mathcal{W}_i(\theta) + \mathcal{K}_i(\theta) >= O_{\mathbb{P}}(1).$$

In fact the only terms that their limit do not vanish are the corresponding to the noise for instance

$$b_{11}\frac{\sigma^2}{2N}\sum_{i=0}^{N-1}(\int_{ih}^{(i+1)h}((i+1)h-u)dW_u)^2 \to 2\sigma^2,$$

by using the LLN. The other terms of this type can be treated similarly.

Remark 4.7. Notice that $M(\theta) = \frac{13}{2}\mathbb{E}[(P(z_0;\theta) - P(z_0;\theta_0))^2] \geq 0$ for all $\theta \in \Theta$ and $M(\theta_0) = 0$, so it is inferred that $M(\theta)$ attains a global minimum at the point $\theta = \theta_0$ whose value is 0. We will assume that there exists an open neigbourhood U_{θ_0} of θ_0 such that $U_{\theta_0} \subset \Theta$ and $\nabla_\theta P(z_0;\breve{\theta}) \neq (0,\ldots,0)$ for all $\breve{\theta} \in U_{\theta_0}$, $\breve{\theta} \neq \theta_0$. This condition guarantees that θ_0 is the unique point of U_{θ_0} where M attains the minimum. In the four models we mentioned in Remark 4.2 this condition is fulfilled as we will show in Appendix A.. This identifiability condition will be used in the theorem bellow.

In the following result it is proved the consistency of the estimator $\hat{\theta}_{N,h}$.

Theorem 4.8. *Let $\theta \in \Theta$, where Θ is a compact set. Let $\{h_N\}_{N\geq 1}$ a sequence such that $h_N \to 0$ and $h_N N \to \infty$ for $N \to \infty$. Under the hypotheses of Theorem 4.5 we have*

$$\lim_{N\to\infty} \hat{\theta}_{N,h_N} = \theta_0,$$

in probability.

Before we start the proof of the theorem, a needed result is established. A brief sketch of the proof can be found in Appendix A.

Lemma 4.9. *Let $\Theta \subset \mathbb{R}^{p+q}$ a compact set and suppose that the hypotheses $H1$–$H5$ are fulfilled. In addition, suppose that*

$$\mathbb{E}[\sup_{\theta\in\Theta} \ell_1^2(z_{ih};\theta)] < \infty.$$

Then for a fixed h, the sequence $\{\frac{1}{h_N}\left[M_{N,h_N}(z_{h_N}^{Nh_N};\theta)\right]\}$ converges uniformly with respect to θ, in probability.

Remark 4.10. The additional hypothesis $\mathbb{E}[\sup_{\theta\in\Theta} \ell_1^2(z_{ih};\theta)] < \infty$ can be achieved if the constant $\mathbf{C}_\theta$ in equations (4.7) and (4.8) can be bounded uniformly for $\theta \in \Theta$ by a constant $\mathbf{C}$. In the four models mentioned above this hypothesis is fulfilled.

Proof of theorem 4.8. Let us define $\tilde{M}_N(\theta) = \frac{1}{h_N}\left[M_{N,h_N}(z_{h_N}^{Nh_N};\theta) - M_{N,h_N}(z_{h_N}^{Nh_N};\theta_0)\right]$. Notice that $\tilde{M}_N(\theta)$ and $M_{N,h_N}(\theta)$ attains the minimum at the same value $\hat{\theta}_{N,h_N}$. We have

$$0 \leq M(\hat{\theta}_{N,h_N}) = M(\hat{\theta}_{N,h_N}) - \tilde{M}_N(\hat{\theta}_{N,h_N}) + \tilde{M}_N(\hat{\theta}_{N,h_N}).$$

As $\hat{\theta}_{N,h_N} = \arg\min_{\theta\in\Theta} \tilde{M}_N(\theta)$ it holds

$$\tilde{M}_N(\hat{\theta}_{N,h_N}) \leq \tilde{M}_N(\theta_0).$$

This implies

$$0 \leq M(\hat{\theta}_{N,h_N}) \leq M(\hat{\theta}_{N,h_N}) - \tilde{M}_N(\hat{\theta}_{N,h_N}) + \tilde{M}_N(\theta_0).$$

From the compactness hypothesis, Theorem 4.5 and Lemma (4.9), by taking absolute value we get

$$\begin{aligned} 0 &\leq |M(\hat{\theta}_{N,h_N})| \leq |M(\hat{\theta}_{N,h_N}) - M_N(\hat{\theta}_{N,h_N})| + |M_N(\theta_0)| \\ &\leq 2\sup_{\theta\in\Theta} |M(\theta) - M_N(\theta)| \to 0, \end{aligned}$$

in probability. It follows that

$$\lim_{N\to\infty} M(\hat{\theta}_{N,h_N}) = 0 \ (= M(\theta_0)) \text{ in probability.} \tag{4.30}$$

Let θ_* be a limit point of the sequence $\{\hat{\theta}_{N,h_N}\}$. As Θ is a compact set then there exists a subsequence $\{\hat{\theta}_{N_j,h_{N_j}}\} \subseteq \{\hat{\theta}_{N,h_N}\}$ such that

$$\lim_{j\to\infty} \hat{\theta}_{N_j,h_{N_j}} = \theta_*.$$

Taking in account that $M(\theta)$ is continuous and by (4.30) we deduce that

$$\lim_{j\to\infty} M(\hat{\theta}_{N_j,h_{N_j}}) = M(\theta_*).$$

In virtue of the assumption of unicity for θ_0 we made in Remark 4.7 we obtain $\theta_* = \theta_0$ which allow us to conclude that

$$\lim_{N\to\infty} \hat{\theta}_{N,h_N} = \theta_0,$$

in probability. □

Remark 4.11. We have just shown consistency of the estimator in the case when the data is totally available. We think that a similar result holds, when only position is available, by approximating the velocity with an Euler scheme. Defining the contrast

$$M_{N,h}(x_1,\ldots,x_N,\theta) = \frac{1}{2(N-1)}\sum_{i=2}^{N}\left\| B\left(\begin{bmatrix}\frac{x(i)-x(i-1)}{h}\\ x(i)\end{bmatrix};\theta\right)^{-1}\left[\begin{bmatrix}\frac{x(i+1)-x(i)}{h}\\ x(i+1)\end{bmatrix}\right.\right.$$
$$\left.\left. -A\left(\begin{bmatrix}\frac{x(i)-x(i-1)}{h}\\ x(i)\end{bmatrix};\theta\right)\begin{bmatrix}\frac{x(i)-x(i-1)}{h}\\ x(i)\end{bmatrix}\right]\right\|^2,$$

and the estimator

$$\hat{\theta}_{N,h} := \arg\min_{\theta\in\Theta} M_{N,h}(x_1,\ldots,x_N,\theta),$$

we dare to affirm that it is possible to prove that $\hat{\theta}_{N,h} \to \theta_0$ when $h \to 0$ and $Nh \to \infty$ in some sense. While it is not the aim of this work to develop the theoretical proofs, we consider important to perform computational essays with this approach in order to compare with the estimator $\hat{\theta}$ and the estimations obtained by other authors. In the next section we show the results.

5. Simulation study

In this section we perform several simulations to check how the Ozaki's local linearization method works numerically and then we check how our estimation for the derivative estimation works. Comparison between the results obtained for the two cases are made. Also, comparison with the results of Ozaki [14], Pokern *et al.* [16] and Samson *et al.* [18] are carried out. We simulate three data sets from locally linearized models obtained from each of two models: the Van der Pol stochastic oscillator

$$\ddot{x} + (ax^2 + bx + c)\dot{x} + \mathrm{d}x = \sigma \mathrm{d}W_t, \tag{5.1}$$

and the Duffin stochastic oscillator

$$\ddot{x} + \mathrm{d}\dot{x} + (ax^2 + bx + c)x = \sigma \mathrm{d}W_t. \tag{5.2}$$

Here a, b, c and d are the parameters to be estimated. As the variance σ^2 need to be estimated also, then the sets of parameters has the form $\theta = (a, b, c, d, \sigma^2)$. We use for the simulations the following sets of parameters

Van der Pol model	Duffin model
$\theta_{vdP1} = (1, 0, -1, 14.8, 1)$	$\theta_{D1} = (4, 0, 14.8, 0.25, 0.25)$
$\theta_{vdP2} = (2, 0, 2, 10, 1)$	$\theta_{D2} = (1, 0, -1, 1, 0.25)$
$\theta_{vdP3} = (0, 0, 0.5, 4, 1)$	$\theta_{D3} = (4, 3, -4, 1, 0.25)$

1. The sets of parameters were chosen in order to obtain functions $g_1(x(t); a)$ and $g_2(x(t); a)$ which fulfill the assumptions $H1$ and $H2$.
2. In order to compare our results with those obtained by Ozaki we chose the sets of parameters θ_{vdP1} and θ_{D1} to be the same as in the work of Ozaki [14], except the variance in the Duffin model: in this work we consider $\sigma^2 = 0.5$ and Ozaki considers $\sigma^2 = 4$. The set of parameters θ_{D1} yields the single-well potential $V(x) = x^4 + 7.4x$, that is, a potential which has one minimum.
3. The set of parameters θ_{D2} yields the so called Kramers oscillator. In this case the potential $V(x) = \frac{x^4}{4} - \frac{x^2}{2}$ is known to be a symmetric double-well potential, that is, a potential which has a minimum attained at two different points. This kind of potentials are considered in the literature as the overlap of two single-well potentials having the same minimum.
4. The set of parameters θ_{D3} yields a non-symmetric double-well potential, that is, a potential which has two relatives minima: an absolute minimum and a relative minimum. In this case the potential is given by $V(x) = x^4 + \frac{3x^3}{2} - 2x^2$.
5. The set of parameters θ_{vdP3} yields the same model as in the works of Pokern *et al.* [16] and Samson *et al.* [18].

Once we have established the parameters to be used, first we consider that both position and velocity are observed. In this case six data sets having the form $z(t_i) = \begin{bmatrix} y(i) \\ x(i) \end{bmatrix}$ are simulated from a locally linearized model corresponding with each set of parameters. Here $y(i)$ stands for the velocity and $x(i)$ for the position of the system in the time t_i. Each data set is obtained from the discrete model (3.12) following the work of Ozaki [14] and this is done in the following way: for each set of parameters θ we calculate

1. The matrix $A(z(t); \theta) = I + J^{-1}(z(t); \theta)[e^{J(z(t);\theta)} - I]F(z(t); \theta)$. In the implementation of our algorithm the matrix $e^{J(z(t);\theta)\,h}$ is approximated by the matrix $I + J(z(t); \theta)h + \frac{(J(z(t);\theta)h)^2}{2}$.
2. The matrix $B(z(t); \theta)$. In our algorithm it was replaced by the Cholesky decomposition of the matrix $\frac{1}{\sigma^2}\Sigma_{ih}(h; \theta)$.
3. The data is obtained from the equality

$$z(t_i) = A(z(t_i); \theta) \cdot z(t_i) + B(z(t_i); \theta) \cdot randn(2, 1).$$

In the simulations for all the sets of parameters, except for θ_{vdP3}, the step size is $h = 0.1$. In order to observe the behavior of the algorithms when the sample size increases, four different sizes for each data set are considered $N = 100$, $N = 1000$, $N = 5000$ and $N = 10000$. For comparison reasons in the simulations corresponding the set of parameters θ_{vdP3} we consider $h = 0.1, n = 100$, $h = 0.1, n = 1000$ and $h = 0.01, n = 1000$. For the parameter sets θ_{vdP1} we set the initial conditions $z_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. For the parameter set θ_{vdP3} we assume the initial conditions $z_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ as in the works of Pokern *et al.* [16] and Samson *et al.* [18]. For the parameter set θ_{vdP2} we assume the initial conditions to be $z_0 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. In the case when only position is available another six data set will be obtained having the form $z(t_i) = \begin{bmatrix} \hat{y}(i) \\ x(i) \end{bmatrix}$. Here $\widehat{y}(i) = \frac{x(i+1) - x(i)}{h}$ is the approximation of the velocity using a

Table 1. Estimated parameters for Van der Pol model 1. The stepsize considered is $h = 0.1$. Mean and standard error of the parameters estimator are computed on 100 simulated data sets. Four values of the sample size N were taken in account $N = 100, 1000, 5000, 10000$. Three estimators are compared: completely observed data, completely observed data obtained by Ozaki and partially observed data.

E	θ_{vdP1}	$N = 100$	std	$N = 1000$	std	$N = 5000$	std	$N = 10\,000$	std
$\widehat{a^C}$		1.021	0.0906	0.8831	0.0450	1.0139	0.0389	1.0151	0.0754
$\widehat{a^C}_{Oz}$	1			0.9878					
$\widehat{a^P}$		1.0050	0.0536	1.0445	0.0921	1.0304	0.0430	1.0118	0.1572
$\widehat{b^C}$		-0.0122	0.1076	0.0130	0.0845	-0.0101	0.0920	-0.0260	0.1018
$\widehat{b^C}_{Oz}$	0			-0.0112					
$\widehat{b^P}$		-0.0060	0.1005	0.0133	0.0827	0.0048	0.0992	0.0190	0.1010
$\widehat{c^C}$		-0.9816	0.0939	-0.8447	0.0453	-1.0252	0.1192	-1.0466	0.1894
$\widehat{c^C}_{Oz}$	-1			-0.9742					
$\widehat{c^P}$		-0.9830	0.0917	-0.9983	0.1566	-1.0031	0.0681	-0.9621	0.1468
$\widehat{d^C}$		14.6709	0.2747	14.8869	0.1764	14.6319	0.3578	14.9214	0.3247
$\widehat{d^C}_{Oz}$	14.8			14.8520					
$\widehat{d^P}$		14.6936	0.2851	14.7911	0.4355	14.7890	0.3356	14.8515	0.3634
$\widehat{\sigma^2}^C$		0.9966	0.1929	1.1337	0.0302	1.0125	0.1372	1.0819	0.1738
$\widehat{\sigma^2}^C_{Oz}$	1			1.0141					
$\widehat{\sigma^2}^P$		1.0984	0.1750	1.0389	0.1306	1.0812	0.1597	0.9687	0.2749

standard Euler scheme. As in the case of totally observed data we use the same step size and the same four different sizes for each data set.

Once we have the twelve data sets we are ready for the estimation stage. In order to avoid minimization algorithms which use the gradient of the function to be minimized, we use a slight modification of the simplex based Nelder-Mead algorithm. For each data set, except for the corresponding to the parameters θ_{vdP3}, we run the algorithms a hundred times and compute the mean and the standard deviation of the estimations. For the data sets corresponding to the parameters θ_{vdP3} we run the algorithms a thousand times and compute the mean and the standard deviation of the estimations. The estimated parameters in the case of completely and partially observed data are denoted by $\widehat{\theta^C}$ and $\widehat{\theta^P}$ respectively. The estimated parameters in the papers of Ozaki, Pokern *et al.* and Samson *et al.* are identified with the lower cases Oz, Po and Sa respectively. The E in the first column of each table stands for the estimator. In each table the second column contains the real values of the parameters. For each parameter set the results obtained are presented in the tables below.

In Table 1 it can be observed a good performance when comparing the cases of totally and partially observed data. The results reported by Ozaki in [14] correspond with the case $N = 1000$. In this case Ozaki's estimator works better than ours, nevertheless when the sample size is increased, the estimation is improved and the results are similar with those obtained by Ozaki. Another feature arising in Table 1 is the fact that the Euler scheme works considerably good even in the case when the sample size is $N = 100$. In Table 2 the results are good for the two estimators.

Section 3 in the work of Pokern *et al.* [16] is devoted to carry out a simulation study by considering three models of stochastic oscillators. Two of them are linear stochastic oscillators: stochastic growth and harmonic

TABLE 2. Estimated parameters for Van der Pol model 2. Mean and standard error of the parameters estimator are computed on 100 simulated data sets. Four values of the sample size N were taken in account $N = 100, 1000, 5000, 10\,000$. Two estimators are compared: completely observed data and partially observed data.

E	θ_{vdP2}	$N = 100$	std	$N = 1000$	std	$N = 5000$	std	$N = 10\,000$	std
$\widehat{a^C}$		1.6312	0.1953	1.7804	0.3194	2.0858	0.3507	2.2434	0.3919
$\widehat{a^P}$	2	1.8357	0.1993	2.2059	0.3098	1.7038	0.4114	2.1447	0.3257
$\widehat{b^C}$		0.0037	0.0202	0.0024	0.0434	−0.0018	0.0275	−0.0027	0.0341
$\widehat{b^P}$	0	−0.0020	0.0381	0.0010	0.0374	0.0009	0.0267	0.0007	0.0211
$\widehat{c^C}$		2.4118	0.1054	2.1037	0.6030	1.8546	0.2834	2.2821	0.1562
$\widehat{c^P}$	2	1.8926	0.0419	2.2159	0.0804	2.0421	0.3175	1.9150	0.2966
$\widehat{d^C}$		9.8338	0.3999	9.8214	0.6702	9.5326	0.3745	10.4202	0.6696
$\widehat{d^P}$	10	9.2858	0.4746	10.5078	0.6587	9.9742	0.4193	9.4332	0.5790
$\widehat{\sigma^2}^C$		1.0251	0.0051	0.7892	0.0160	0.9284	0.0074	0.8958	0.0249
$\widehat{\sigma^2}^P$	1	0.8941	0.0033	0.9283	0.0073	1.1351	0.0128	1.1284	0.0119

TABLE 3. Estimated parameters for Van der Pol model 3. Mean and standard error of the parameters estimator are computed on 1000 simulated data sets. Three cases were taken in account $h = 0.1, N = 100$, $h = 0.1, N = 1000$ and $h = 0.01, N = 1000$. Four estimators are compared: completely observed data and partially observed data obtained by us, completely observed data and partially observed data obtained by Samson *et al.* and the variance estimation obtained by Pokern *et al.* for partially observed data.

E	θ_{vdP3}	$N = 100$		$N = 1000$		$N = 1000$	
		$h = 0.1$	std	$h = 0.1$	std	$h = 0.01$	std
$\widehat{c^C}$		0.3849	0.0785	0.4866	0.0693	0.5057	0.0876
$\widehat{c^C}_{Sa}$		1.022	0.098	1.086	0.271	0.678	0.326
$\widehat{c^P}$	0.5	0.3795	0.0619	0.5142	0.0673	0.4284	0.0729
$\widehat{c^P}_{Sa}$		1.285	0.275	1.215	0.096	0.699	0.330
$\widehat{d^C}$		3.8508	0.1000	3.9965	0.0864	3.8189	0.1255
$\widehat{d^C}_{Sa}$		3.567	0.489	3.488	0.187	4.034	0.642
$\widehat{d^P}$	4	3.7767	0.1352	4.0210	0.0741	3.9715	0.1184
$\widehat{d^P}_{Sa}$		3.588	0.494	3.501	0.188	4.032	0.644
$\widehat{\sigma^2}^C$		0.8033	0.0114	0.9954	0.0033	1.0151	0.0008
$\widehat{\sigma^2}^C_{Sa}$		0.980	0.069	0.974	0.021	0.996	0.021
$\widehat{\sigma^2}^P$	1	0.9377	0.0076	1.0176	0.0029	0.9685	0.0005
$\widehat{\sigma^2}^P_{Sa}$		0.946	0.074	0.956	0.021	0.994	0.023
$\widehat{\sigma^2}^P_{Po}$		1.154	0.074	1.114	0.025	1.016	0.013

Table 4. Estimated parameters for Duffin model 1. Mean and standard error of the parameters estimator are computed on 100 simulated data sets. Four values of the sample size N were taken in account $N = 100, 1000, 5000, 10000$. Three estimators are compared: completely observed data, partially observed data and completely observed data obtained by Ozaki.

E	θ_{D1}	$N = 100$	std	$N = 1000$	std	$N = 5000$	std	$N = 10\ 000$	std
$\widehat{a^C}$		8.1745	5.5976	3.7635	1.8398	4.1978	0.7339	4.2733	0.6285
$\widehat{a^C}_{Oz}$	4			3.7954					
$\widehat{a^P}$		7.2542	4.6135	4.4201	1.0071	4.3048	1.0117	4.4406	1.6694
$\widehat{b^C}$		0.0018	0.0476	−0.0001	0.0216	−0.0015	0.0261	0.0020	0.0280
$\widehat{b^C}_{Oz}$	0			0.3551					
$\widehat{b^P}$		0.0062	0.0562	0.0023	0.0205	−0.0004	0.0155	−0.0018	0.0226
$\widehat{c^C}$		14.5822	0.4943	14.8029	0.2010	14.7909	0.1550	14.8392	0.1446
$\widehat{c^C}_{Oz}$	14.8			14.9385					
$\widehat{c^P}$		13.5058	0.3861	14.8850	0.1470	14.7139	0.1352	14.8116	0.1895
$\widehat{d^C}$		0.2037	0.0282	0.3042	0.0322	0.2328	0.0196	0.2152	0.0164
$\widehat{d^C}_{Oz}$	0.25			0.1589					
$\widehat{d^P}$		0.3175	0.0312	0.2337	0.0266	0.1990	0.0276	0.3578	0.0238
$\widehat{\sigma^2}^C$	0.25	0.2253	0.0011	0.2625	0.0004	0.2494	0.0005	0.2343	0.0002
$\widehat{\sigma^2}^C_{Oz}$	4			4.0270					
$\widehat{\sigma^2}^P$	0.25	0.2016	0.0009	0.2664	0.0003	0.2457	0.0001	0.2564	0.0003

oscillator. The third is a trigonometric stochastic oscillator. The same three models were considered for simulating in the work of Samson *et al.* [18]. Our framework is a little more general in the sense that it allows nonlinear damping coefficients. Nevertheless, in order to compare our estimators with those of them we consider in the Van der Pol model (5.1) the parameters values to be $a = b = 0$. It yields the second model considered in the previously mentioned works of Pokern *et al.* and Samson *et al.* In Table 3 we show the results.

In Table 3 we can observe that all estimators works relatively good. Our estimators happen to improve considerably the other results when considering $h = 0.1, N = 1000$. However, when the sample sizes increases and the stepsize decreases our estimators does not improve themselves.

For the Duffin oscillator the following results are showed in Table 4. We observe a good behavior of our estimators in the both cases of totally and partially observed data. Also we observe that our results are similar of those of Ozaki. In the case of the Kramers oscillator corresponding to the Duffin model 2, the results presented in Table 5 show us that the estimator does not works properly for the short sample size $N = 100$. However when the sample size increases it works correctly.

In the case of the Duffin model 3, as showed in Table 6 both of the two estimators works relatively good when the sample data is $N = 100, 1000$ and $N = 5000$. Nevertheless, when $N = 10000$ the estimator improves the results. We recall that, in this case, the potential V is a non-symmetric double well potential and we suspect that is due to this fact that a greater sample size is needed.

6. Conclusions

Consistency in the sense of probability convergence of the maximum likelihood based estimator for completely observed data proposed by Ozaki in [14] is proved. Several simulations were carried out for the cases of completely observed data. Also some simulations were made of partially observed data approximating the velocity by an

Table 5. Estimated parameters for Duffin model 2. Mean and standard error of the parameters estimator are computed on 100 simulated data sets. Four values of the sample size N were taken in account $N = 100, 1000, 5000, 10000$. Two estimators are compared: completely observed data and partially observed data.

E	θ_{D2}	$N = 100$	std	$N = 1000$	std	$N = 5000$	std	$N = 10000$	std
$\widehat{a^C}$		0.9080	0.1621	0.9612	0.1105	1.0133	0.2119	1.0906	0.2086
$\widehat{a^P}$	1	0.7809	0.1360	1.0864	0.1215	1.0213	0.0409	1.0123	0.1838
$\widehat{b^C}$		0.0024	0.0264	−0.0062	0.0375	0.0094	0.0493	−0.0023	0.0358
$\widehat{b^P}$	0	0.0054	0.0384	−0.0030	0.0404	−0.0045	0.0350	−0.0017	0.0385
$\widehat{c^C}$		−0.6701	0.1591	−0.9799	0.1461	−0.9673	0.2161	−1.1705	0.2769
$\widehat{c^P}$	−1	−0.8271	0.2117	−1.1448	0.1629	−1.0228	0.0527	−1.0428	0.2730
$\widehat{d^C}$		0.9873	0.1405	0.9477	0.0468	0.9600	0.1052	0.9238	0.2817
$\widehat{d^P}$	1	0.7615	0.1284	1.0387	0.1466	0.9545	0.0443	0.9389	0.1893
$\widehat{\sigma^2}^C$		0.1964	0.0014	0.2437	0.0033	0.2515	0.0068	0.2483	0.0178
$\widehat{\sigma^2}^P$	0.25	0.2499	0.0035	0.2556	0.0038	0.2588	0.0004	0.2434	0.0053

Table 6. Estimated parameters for Duffin model 3. Mean and standard error of the parameters estimator are computed on 100 simulated data sets. Four values of the sample size N were taken in account $N = 100, 1000, 5000, 10000$. Two estimators are compared: completely observed data and partially observed data.

E	θ_{D3}	$N = 100$	std	$N = 1000$	std	$N = 5000$	std	$N = 10\,000$	std
$\widehat{a^C}$		3.4716	1.9433	3.6678	3.0803	3.4254	3.3269	3.9382	3.2953
$\widehat{a^P}$	4	3.4407	2.3827	3.7521	3.1696	3.4967	2.4346	4.0809	3.9604
$\widehat{b^C}$		1.8289	5.1476	2.2959	8.7992	1.7797	9.2544	2.7084	9.7698
$\widehat{b^P}$	3	2.0885	6.5927	2.4166	8.7129	1.4950	6.9542	3.1291	11.2185
$\widehat{c^C}$		−4.5941	3.3701	−4.3126	6.2528	−4.5677	6.5279	−4.3060	7.2521
$\widehat{c^P}$	−4	−4.0059	4.5588	−4.3184	6.0181	−5.0932	5.0833	−3.9774	7.9922
$\widehat{d^C}$		0.8143	0.3162	0.8946	0.2225	0.8671	0.6516	0.9832	0.3476
$\widehat{d^P}$	1	0.9104	0.3277	0.8920	0.3492	0.8999	0.4105	0.9610	0.6087
$\widehat{\sigma^2}^C$		0.2674	0.0417	0.2606	0.0161	0.2856	0.0747	0.2597	0.0357
$\widehat{\sigma^2}^P$	0.25	0.3093	0.1138	0.2737	0.0820	0.2594	0.0476	0.2513	0.0439

Euler scheme. It allowed us to notice some features of the estimator and to compare our results with previous results obtained by Ozaki [14], Pokern *et al.* [16] and Samson *et al.* [18]. The comparison yields that the Euler scheme based estimator for the derivative works considerably as good as those proposed by the other authors. However, some considerations could be taken in account in order to extend the scope of this work. We think that the local linearization framework can be extended for higher order systems. Also, higher order schemes for the derivative can be proved to work both theoretically and numerically.

Appendix A

Proof of Lemma 3.4. Consider the matrix $B_{ih} := B_{ih}(h;\theta)$ as defined below the equation (3.12) and consider the transformation

$$\begin{aligned} \Xi_h &= B_h^{-1}\left(z_{2h} - A_h\, z_h\right) \\ \Xi_{2h} &= B_{2h}^{-1}\left(z_{3h} - A_{2h}\, z_{2h}\right) \\ &\vdots\ \vdots\ \vdots \\ \Xi_{(N-1)h} &= B_{Nh}^{-1}\left(z_{Nh} - A_{(N-1)h}\, z_{(N-1)h}\right), \end{aligned}$$

where $B_{ih}^{-1} = B_{ih}^{-1}(h;\theta)$ and $A_{ih} = A_{ih}(h;\theta)$, for all $i = 1,\dots,N$. The random variables $\Xi_{2h},\dots,\Xi_{(N-1)h}$ are independent and distributed according to the Normal law, with parameters 0 and $\sigma^2 I$. The Jacobian of this transformation is

$$J(z,\Xi) = \begin{bmatrix} B_1^{-1} & 0 & 0 & 0 & \cdots & 0 & 0 \\ \frac{\partial \Xi_3}{\partial z_2} & B_2^{-1} & 0 & 0 & \cdots & 0 & 0 \\ 0 & \frac{\partial \Xi_4}{\partial z_3} & B_3^{-1} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & \frac{\partial \Xi_m}{\partial z_{m-1}} & B_{m-1}^{-1} \end{bmatrix}$$

which is a triangular matrix and so its determinant is given by $\prod_{i=1}^{m-1} \frac{1}{\det B_i}$. Then, by the bivariate Normal distribution formula we have

$$p(\Xi_i) = \frac{1}{2\pi\, \det(\Sigma)^{\frac{1}{2}}}\, \exp\left[-\frac{1}{2}\, \Xi_i'\, \Sigma^{-1}\, \Xi_i\right], \qquad i = 2,\dots,m.$$

In our case, $\Sigma = \sigma^2 I$, so $\det \Sigma = \sigma^4$, $\Sigma^{-1} = \frac{1}{\sigma^2} I$, and it follows that

$$\begin{aligned} p\left(\Xi_i\right) &= \frac{1}{2\pi\,\sigma^2}\, \exp\left[-\frac{1}{2\sigma^2}\, \Xi_i'\, \Xi_i\right] \\ &= \frac{1}{2\pi\,\sigma^2}\, \exp\left[-\frac{1}{2\sigma^2}\, \left[z_i - A_{i-1} z_{i-1}\right]' \left[B_{i-1}^{-1}\right]' \left[B_{i-1}^{-1}\right] \left[z_i - A_{i-1} z_{i-1}\right]\right], \end{aligned}$$

so we can conclude that

$$\begin{aligned} p\left(z_1, z_2, \dots, z_m\right) = p\left(\Xi_2 \mid z_1\right) \dots p\left(\Xi_m \mid z_1\right) p(z_1) &= \frac{p(z_1)}{(2\pi)^{m-1}\sigma^{2(m-1)}} \prod_{i=2}^{m} \frac{1}{|\det B_{i-1}|} \\ &\times \exp\left[-\frac{1}{2\sigma^2}\left[z_i - A_{i-1} z_{i-1}\right]' \left[B_{i-1}^{-1}\right]' \left[B_{i-1}^{-1}\right] \left[z_i - A_{i-1} z_{i-1}\right]\right]. \end{aligned}$$

□

Proof of Lemma 4.3. From the observation equation (3.7) the equations for $x(s)$ and $y(s)$ are given by

$$x(s) - x(s') = y(s)(s - s') + \int_{s'}^{s} y(t)\mathrm{d}t = \int_{s'}^{s} (y(t) - y(s))\mathrm{d}t \tag{A.1}$$

and

$$y(t) - y(s') = \sigma \left(W_t - W_{s'}\right) - \int_{s'}^{t} P(z(u);\theta)\mathrm{d}u, \tag{A.2}$$

so

$$x(s) - x(s') = \sigma \int_{s'}^{s} \left(W_t - W_{s'}\right)\mathrm{d}t - \int_{s'}^{s}\int_{s'}^{t} P(z(u);\theta)\mathrm{d}u\mathrm{d}t, \tag{A.3}$$

and therefore using Jensen's inequality we obtain

$$\begin{aligned}\mathbb{E}\left[|x(s) - x(s')|^{2k}\right] &= \mathbb{E}\left[\left|\sigma \int_{s'}^{s} \left(W_t - W_{s'}\right)\mathrm{d}t - \int_{s'}^{s}\int_{s'}^{t} P(z(u);\theta)\mathrm{d}u\mathrm{d}t\right|^{2k}\right] \\ &\leq \mathbf{C}_\theta \left\{\sigma^{2k}\mathbb{E}\left[\left|\int_{s'}^{s} \left(W_t - W_{s'}\right)\mathrm{d}t\right|^{2k}\right] + \mathbb{E}\left[\left|\int_{s'}^{s}\int_{s'}^{t} P(z(u);\theta)\mathrm{d}u\mathrm{d}t\right|^{2k}\right]\right\}. \end{aligned}\tag{A.4}$$

Using Jensen's inequality and an appropriate change of variables, for the first summand in (A.4) we have

$$\begin{aligned}\mathbb{E}\left[\left|\int_{s'}^{s} \left(W_t - W_{s'}\right)\mathrm{d}t\right|^{2k}\right] &= \mathbb{E}\left[(s-s')^{2k}\left|\int_{\frac{s'}{s-s'}}^{\frac{s}{s-s'}} \left(W_{r(s-s')} - W_{s'}\right)\mathrm{d}r\right|^{2k}\right] \\ &\leq (s-s')^{2k}\int_{\frac{s'}{s-s'}}^{\frac{s}{s-s'}} \mathbb{E}\left[\left|W_{r(s-s')} - W_{s'}\right|^{2k}\right]\mathrm{d}r \\ &= \frac{(2k)!}{2^k(k+1)!}(s-s')^{3k} = O\left([s-s']^{3k}\right), \end{aligned}\tag{A.5}$$

and for the second summand in (A.4) it results

$$\begin{aligned}\mathbb{E}\left[\left|\int_{s'}^{s}\int_{s'}^{t} P(z(u);\theta)\mathrm{d}u\mathrm{d}t\right|^{2k}\right] &= \mathbb{E}\left[(s-s')^{2k}\left|\int_{s'}^{s}\int_{s'}^{t} P(z(u);\theta)\mathrm{d}u\frac{\mathrm{d}t}{s-s'}\right|^{2k}\right] \\ &\leq \frac{(s-s')^{4k}}{2k+1}\mathbb{E}\left[\left|P(z(r(s-s')));\theta)\right|^{2k}\right]\mathrm{d}r. \end{aligned}\tag{A.6}$$

From inequality (2.3) we have

$$\left|P(z(r(s-s')));\theta)\right|^{2k} \leq \mathbf{C}_\theta \left\|z(r(s-s'))\right\|^{2k\gamma_3},$$

therefore

$$\mathbb{E}\left[\left|P(z(r(s-s')));\theta)\right|^{2k}\right] \leq \mathbf{C}_\theta\mathbb{E}\left[\mathbf{C}_\theta \left\|z(r(s-s'))\right\|^{2k\gamma_3}\right] < \infty,$$

where the last inequality is due to the fact that the polynomial moments of the process $z(t)$ are finite. By the stationarity of the process $z(t)$, it follows from (A.6)

$$\mathbb{E}\left[\left|\int_{s'}^{s}\int_{s'}^{t} P(z(u);\theta)\mathrm{d}u\mathrm{d}t\right|^{2k}\right] \leq \mathbf{C}_\theta\mathbb{E}\left[\mathbf{C}_\theta\left\|z_0\right\|^{2k\gamma_3}\right](s-s')^{4k} = O\left([s-s']^{4k}\right). \tag{A.7}$$

By (A.5) and (A.7) it follows from (A.4)

$$\mathbb{E}\left[|x(s)-x(s')|^{2k}\right] \leq \mathbf{C}_\theta\left\{\sigma^{2k}O\left([s-s']^{3k}\right)+O\left([s-s']^{4k}\right)\right\} = O\left([s-s']^{3k}\right).$$

By performing similar computations, from equation (A.2) we have

$$\mathbb{E}\left[|y(s)-y(s')|^{2k}\right] = O\left([s-s']^{k}\right).$$

□

Proof of Lemma 3.3. In order to simplify the notation, in this proof we will consider

1. $J(z_{ih};\theta) = J_{ih}(\theta) = \begin{bmatrix} 0 & 1 \\ d_{12}(z_{ih};\theta) & d_{22}(z_{ih};\theta)\end{bmatrix}$.
2. The eigenvalues of $J_{ih}(\theta)$ are denoted by

$$\mu_1(z_{ih},h;\theta) = \mu_1 = \frac{-d_{22}(z_{ih};\theta)+\sqrt{d_{22}^2(z_{ih};\theta)+4d_{12}(z_{ih};\theta)}}{2}$$
$$\mu_2(z_{ih},h;\theta) = \mu_2 = \frac{-d_{22}(z_{ih};\theta)-\sqrt{d_{22}^2(z_{ih};\theta)+4d_{12}(z_{ih};\theta)}}{2}.$$

3. $\Sigma_{ih}(h;\theta) = \Sigma_{ih}(\theta) = \begin{bmatrix} a_{11}(\theta) & a_{12}(\theta) \\ a_{12}(\theta) & a_{22}(\theta)\end{bmatrix}$.
4. $\Sigma_{ih}^{-1}(h;\theta) = \Sigma_{ih}^{-1}(\theta) = \begin{bmatrix} b_{11}(\theta) & b_{12}(\theta) \\ b_{12}(\theta) & b_{22}(\theta)\end{bmatrix}$.
5. $A_{ih}(h;\theta) = A_{ih}(\theta)$.
6. For other expressions which depends on h we will omit the h in their arguments.

First we will prove that $\Sigma_{ih}(\theta)$ converges uniformly with respect to θ in probability. From hypotheses H1–H3 it is easy to check that there exists a positive constant $\mathbf{C}_\theta$ such that for $j=1,2$ we have $|\mu_j| \leq \mathbf{C}_\theta||z_{ih}||^\gamma$, for certain exponent γ. Hence the eigenvalues of J are bounded in probability. We will demonstrate first the uniform convergence in probability w.r.t θ of the matrix $\Sigma_{ih}(h;\theta)$. It will be done for the element $a_{11}(\theta)$ since the other elements can be treated in the same fashion. We have

$$\begin{aligned}\frac{a_{11}(\theta)}{h^3} &= \frac{1}{3}+\frac{1}{(\mu_1-\mu_2)^2}\sum_{k=4}^{\infty}\frac{(2\mu_1)^{k-1}-2(\mu_1+\mu_2)^{k-1}+(2\mu_2)^{k-1}}{k!}h^{k-3} \\ &= \frac{1}{3}+\frac{1}{(\mu_1-\mu_2)^2}\sum_{k=4}^{\infty}\frac{(\mu_1)^{k-1}[2^{k-1}-2(1+\frac{\mu_2}{\mu_1})^{k-1}+2^{k-1}(\frac{\mu_2}{\mu_1})^{k-1}]}{k!}h^{k-3}. \end{aligned} \quad \text{(A.8)}$$

Denoting $x = \frac{\mu_2}{\mu_1}$ and defining the polynomial

$$Q_{k-1}(x) = 2^{k-1}-2(1+x)^{k-1}+2^{k-1}x^{k-1},$$

it can be readily seen that 1 is a double root of Q and therefore

$$Q_{k-1}(x) = (x-1)^2\tilde{Q}_{k-3}(x).$$

It follows from (A.8)

$$\frac{a_{11}(\theta)}{h^3}=\frac{1}{3}+\sum_{k=4}^{\infty}\frac{(\mu_1)^{k-3}\tilde{Q}_{k-3}(\frac{\mu_2}{\mu_1})}{k!}h^{k-3}=\frac{1}{3}+h\sum_{k=1}^{\infty}\frac{(\mu_1)^{k}\tilde{Q}_{k}(\frac{\mu_2}{\mu_1})}{(k+3)!}h^{k-1}.$$

Let us find a bound for the above expression. On one hand, for all $\varepsilon>0$, there exists a $\delta>0$ such that $\left||\tilde{Q}_k(x)|-|\tilde{Q}_k(1)|\right|<\varepsilon$ for all $x\in(1-\delta,1+\delta)$. By choosing $\varepsilon=|\tilde{Q}_k(1)|$ we get $|\tilde{Q}_k(x)|\le 2|\tilde{Q}_k(1)|$, for all $x\in(1-\delta,1+\delta)$. On the other hand, for x outside $(1-\delta,1+\delta)$ we have $|\tilde{Q}_k(x)|\le\frac{1}{\delta^2}|Q_k(x)|$. Hence $|\tilde{Q}_k(x)|\le 2|\tilde{Q}_k(1)|+\frac{1}{\delta^2}|Q_k(x)|$ for all x. In addition $|Q_k(x)|\le 32^k(1+|x|^k)$, that is $|Q_k(\frac{\mu_2}{\mu_1})|\le 32^k\frac{\mu_2^k+\mu_1^k}{\mu_1^k}$. These computations together with the inequality $|\mu_j|\le\|J_{ih}(\theta)\|$ yields

$$\begin{aligned}\left|\sum_{k=1}^{\infty}\frac{(\mu_1)^{k}\tilde{Q}_{k}(\frac{\mu_2}{\mu_1})}{(k+3)!}h^{k-1}\right| &\le \mathbf{C}\left[2\sum_{k=0}^{\infty}\frac{|\mu_1|^{k+1}}{(k+4)!}h^k+3\sum_{k=0}^{\infty}\frac{|\mu_1|^{k+1}+|\mu_2|^{k+1}}{(k+4)!}h^k\right]\\ &\le\frac{\mathbf{C}}{4}||J_{ih}(\theta)||e^{h||J_{ih}(\theta)||}.\end{aligned}$$

Under the hypothesis H5$''$ it holds

$$\left|\frac{1}{h^3}a_{11}(\theta)-\frac{1}{3}\right|=o_{\mathbb{P}}(1),$$

uniformly over θ.

Now we prove the uniform convergence of the matrix $\varSigma_{ih}^{-1}(\theta)$. As quoted before, we will prove the uniform convergence only for the element $b_{22}(\theta)$ since in a similar way it can be obtained de result for the other entries. To obtain the result for the entry $b_{22}(\theta)$ we need to consider only the convergence of the determinant $\Delta(\theta)=a_{11}(\theta)a_{22}(\theta)-a_{12}^2(\theta)$. Indeed

$$|hb_{22}(\theta)-4|=\left|\frac{\frac{a_{11}(\theta)}{h^3}}{\frac{\Delta(\theta)}{h^4}}-4\right|\le\frac{1}{\frac{\Delta(\theta)}{h^4}}\left|\frac{a_{11}(\theta)}{h^3}-\frac{1}{3}\right|+\frac{4}{\frac{\Delta(\theta)}{h^4}}+\left|\frac{\Delta(\theta)}{h^4}-\frac{1}{12}\right|. \tag{A.9}$$

The two terms inside of the absolute values converge uniform in probability due to the result we obtained for $\varSigma_{ih}(h;\theta)$. Hence only remain to study the fraction $(\dfrac{\Delta(\theta)}{h^4})^{-1}$. By using the uniform convergence in probability for $\varSigma_{ih}(\theta)$, we know that for all $\varepsilon>0$ there exists a $\delta>0$ such that

$$\mathbb{P}\left\{\sup_{\theta\in\varTheta}\left|\frac{\Delta(\theta)}{h^4}-\frac{1}{12}\right|<\delta\right\}\ge 1-\varepsilon.$$

Since for all $\omega\in\{\sup_{\theta\in\varTheta}|\frac{\Delta(\theta)(\omega)}{h^4}-\frac{1}{12}|<\delta\}$ we have $\frac{1}{12}-\delta<\frac{\Delta(\theta)(\omega)}{h^4}<\frac{1}{12}+\delta$ then we get

$$\frac{1}{12}-\delta<\inf_{\theta\in\varTheta}\frac{\Delta(\theta)(\omega)}{h^4}<\frac{1}{12}+\delta.$$

Thus

$$\mathbb{P}\left\{\sup_{\theta\in\varTheta}(\frac{\Delta(\theta)}{h^4})^{-1}<\frac{1}{\frac{1}{12}-\delta}\right\}=\mathbb{P}\left\{\frac{1}{\inf_{\theta\in\varTheta}\frac{\Delta(\theta)(\omega)}{h^4}}<\frac{1}{\frac{1}{12}-\delta}\right\}$$

$$\geq \mathbb{P}\left\{\frac{1}{12}-\delta<\inf_{\theta\in\Theta}\frac{\Delta(\theta)(\omega)}{h^4}<\frac{1}{12}+\delta\right\}$$
$$\geq \mathbb{P}\left\{\forall\theta,\ \frac{1}{12}-\delta<\frac{\Delta(\theta)(\omega)}{h^4}<\frac{1}{12}+\delta\right\}$$
$$= \mathbb{P}\left\{\sup_{\theta\in\Theta}\left|\frac{\Delta(\theta)}{h^4}-\frac{1}{12}\right|<\delta\right\}\geq 1-\varepsilon.$$

Hence the sequence $(\frac{\Delta(\theta)}{h^4})^{-1}$ is uniformly bounded in probability and this implies that (A.9) converges uniformly in probability, this is

$$\sup_{\theta\in\Theta}|hb_{22}(\theta)-4|\to 0,$$

in probability. From the results we conclude that

$$\frac{h^3}{2}||\Sigma_{ih}^{-1}(\theta)||=O_{\mathbb{P}}(1), \tag{A.10}$$

uniformly in θ. □

Sketch of the proof of Lemma 4.9. Let us recall from (4.10) the expression for the contrast

$$\frac{1}{h}M_{N,h}(z_h^{Nh};\theta)=\frac{1}{2hN}\sum_{i=0}^{N-1}\left\langle\Sigma_{ih}^{-1}(\theta)\left[z_{(i+1)h}-A_{ih}(\theta)\,z_{ih}\right],z_{(i+1)h}-A_{ih}(\theta)\,z_{ih}\right\rangle$$
$$=\frac{1}{h}\left[\mathcal{D}_1(\theta)+\mathcal{D}_2(\theta)-\mathcal{D}_3(\theta)\right]. \tag{A.11}$$

Also we recall that we are interested in the difference

$$\frac{1}{h}\left[M_{N,h}(z_h^{Nh};\theta)-M_{N,h}(z_h^{Nh};\theta_0)\right]=\frac{1}{h}\left[\mathcal{D}_1(\theta)-\mathcal{D}_1(\theta_0)+\mathcal{D}_2(\theta)-\mathcal{D}_2(\theta_0)-\mathcal{D}_3(\theta)+\mathcal{D}_3(\theta_0)\right]. \tag{A.12}$$

Consider the term $\mathcal{D}_2(\theta)$ in (A.11). We have

$$\sup_{\theta\in\Theta}|\mathcal{D}_2(\theta)|\leq\sup_{\theta\in\Theta}\left\{h^6||\Sigma_{ih}^{-1}(\theta)||\right\}\frac{1}{N}\sum_{i=0}^{N-1}\sup_{\theta\in\Theta}\left\{\sum_{k=3}^{\infty}\frac{||J_{ih}(\theta)||^{k-1}h^{k-3}}{k!}||f(z_{ih};\theta)||\right\}^2$$
$$\leq\sup_{\theta\in\Theta}\left\{h^6||\Sigma_{ih}^{-1}(\theta)||\right\}\frac{1}{N}\sum_{i=0}^{N-1}e^{(2h||J_{ih}(\theta)||)}||J_{ih}(\theta)||^4||f(z_{ih};\theta)||^2.$$

From hypothesis H5′, taking expectation and using the result in Lemma 3.3 we get

$$|\mathcal{D}_2(\theta)|=O_{\mathbb{P}}(h^3),$$

uniformly in θ. Then

$$\frac{|\mathcal{D}_2(\theta)|}{h}=O_{\mathbb{P}}(h^2), \tag{A.13}$$

uniformly in θ.

For studying the term $\mathcal{D}_3(\theta)$ we will suppose that the analysis of the term $\mathcal{D}_1(\theta)$ is already done, that is: we will suppose that

$$\frac{1}{2N}\sum_{i=0}^{N-1}\left\langle \Sigma_{ih}^{-1}(\theta)(\mathcal{W}_i(\theta)+\mathcal{K}_i(\theta)),\mathcal{W}_i(\theta)+\mathcal{K}_i(\theta)\right\rangle = O_{\mathbb{P}}(1), \tag{A.14}$$

uniformly in θ. Now, using the Cauchy-Schwarz inequality twice, for the term $\mathcal{D}_3(\theta)$ we have

$$\begin{aligned}|\mathcal{D}_3(\theta)| \leq &\left(\frac{1}{N}\sum_{i=0}^{N-1}\left\langle \Sigma_{ih}^{-1}(\theta)(\mathcal{W}_i(\theta)+\mathcal{K}_i(\theta)),\mathcal{W}_i(\theta)+\mathcal{K}_i(\theta)\right\rangle\right)^{\frac{1}{2}}\\ &\times\left(\frac{1}{N}\sum_{i=0}^{N-1}\left\langle \Sigma_{ih}^{-1}(\theta)\sum_{k=3}^{\infty}\frac{J_{ih}^{k-1}(\theta)}{k!}h^{k-3}f(z_{ih};\theta),\sum_{k=3}^{\infty}\frac{J_{ih}^{k-1}(\theta)}{k!}h^{k-3}f(z_{ih};\theta)\right\rangle\right)^{\frac{1}{2}}.\end{aligned}$$

From (A.13) and (A.14) it follows

$$\frac{|\mathcal{D}_3(\theta)|}{h} = O_{\mathbb{P}}(1)O(h^{\frac{1}{2}}), \tag{A.15}$$

which converges to zero uniformly in θ. Only remains to prove the uniform convergence for the term $\mathcal{D}_1(\theta)$. According (4.13) $\mathcal{D}_1(\theta)$ is given by

$$\frac{\mathcal{D}_1(\theta)}{h} = \frac{1}{2hN}\sum_{i=0}^{N-1}\left\langle \Sigma_{ih}^{-1}(\theta)\mathcal{W}_i(\theta),\mathcal{W}_i(\theta)\right\rangle + \frac{1}{2hN}\sum_{i=0}^{N-1}\left\langle \Sigma_{ih}^{-1}(\theta)\mathcal{K}_i(\theta),\mathcal{K}_i(\theta)\right\rangle + \frac{1}{hN}\sum_{i=0}^{N-1}\left\langle \Sigma_{ih}^{-1}(\theta)\mathcal{W}_i(\theta),\mathcal{K}_i(\theta)\right\rangle.$$

For the second addend we have

$$\frac{1}{2hN}\sum_{i=0}^{N-1}\left\langle \Sigma_{ih}^{-1}(\theta)\mathcal{K}_i(\theta),\mathcal{K}_i(\theta)\right\rangle = \frac{h^3 b_{22}(\theta)}{8N}\sum_{i=0}^{N-1}\ell^2(z_{ih};\theta).$$

By the hypothesis in Lemma 4.9 and the result in Lemma 3.3, taking expectation we obtain

$$\frac{1}{8N}\sum_{i=0}^{N-1}\mathbb{E}[\sup_{\theta\in\Theta}\ell^2(z_{ih};\theta)] = \frac{1}{8}\mathbb{E}[\sup_{\theta\in\Theta}\ell^2(z_0;\theta)] < \infty.$$

Hence

$$\frac{1}{2hN}\sum_{i=0}^{N-1}\sup_{\theta\in\Theta}\{\left\langle \Sigma_{ih}^{-1}(\theta)\mathcal{K}_i(\theta),\mathcal{K}_i(\theta)\right\rangle\} = o_{\mathbb{P}}(h^2).$$

Furthermore the third term in (4.13) can be treated in the same fashion in order to obtain

$$\frac{1}{2hN}\sum_{i=0}^{N-1}\sup_{\theta\in\Theta}\{\left\langle \Sigma_{ih}^{-1}(\theta)\mathcal{W}_i(\theta),\mathcal{K}_i(\theta)\right\rangle\} \to 0,$$

in probability. To finish the proof we will consider the difference (A.12) and we will demonstrate that the sequence

$$\frac{1}{2hN}\left[\sum_{i=0}^{N-1}\left\langle \Sigma_{ih}^{-1}(\theta)\mathcal{W}_i(\theta),\mathcal{W}_i(\theta)\right\rangle-\sum_{i=0}^{N-1}\left\langle \Sigma_{ih}^{-1}(\theta_0)\mathcal{W}_i(\theta_0),\mathcal{W}_i(\theta_0)\right\rangle\right]$$

converges in probability uniformly on θ to a continuous function of θ. Assuming that we have proven the above result let us prove that (A.14) holds. In view of the precedents results the only term that remains to study is

$$\frac{1}{2N}\sum_{i=0}^{N-1}<\Sigma_d^{-1}\mathcal{W}_i(\theta),\mathcal{W}_i(\theta)>=h\sum_{i=1}^{10}\mathcal{G}_{iN}.$$

Let us begin with the three last terms in the above sum. They are the product of a sequence bounded in probability uniformly in θ times a sum $\frac{1}{N}\sum_{i=0}^{N-1}\mathcal{Z}_i$ for certain i.i.d. random variables $\mathcal{Z}_i$ that do not depend on θ. Then the LLN gives that these sums converge a.s. to they expectation. Yielding $h\sum_{i=8}^{10}\mathcal{G}_{iN}$ is bounded in probability uniformly in θ. Moreover, we will show that $\sum_{i=1}^{7}\mathcal{G}_{iN}$ tends in probability uniformly in θ to a continuous function of θ. Thus after multiplication by h it tends toward zero uniformly in θ. Summarizing we get that (A.14) holds. Let us consider now the sum $\sum_{i=1}^{7}\mathcal{G}_{iN}$. For the first term we have

$$\begin{aligned}\mathcal{G}_{1N}=&\frac{h^3b_{11}(\theta)}{8N}\sum_{i=0}^{N-1}[P(z_{ih};\theta_0)-P(z_{ih};\theta)]^2\\&+\frac{b_{11}(\theta)}{2hN}\sum_{i=0}^{N-1}\Big(\int_{ih}^{(i+1)h}[(i+1)h-u]\{P_x(\tilde z(u);\theta_0)[x(u)-x_{ih}]+P_y(\tilde z(u);\theta_0)[y(u)-y_{ih}]\}\mathrm{d}u\Big)^2\\&+\frac{h^2b_{11}(\theta)}{2hN}\sum_{i=0}^{N-1}[P(z_{ih};\theta_0)-P(z_{ih};\theta)]\\&\times\int_{ih}^{(i+1)h}[(i+1)h-u]\{P_x(\tilde z(u);\theta_0)[x(u)-x_{ih}]+P_y(\tilde z(u);\theta_0)[y(u)-y_{ih}]\}\mathrm{d}u.\end{aligned}$$

In the second addend, the only term which depends on θ is $b_{11}(\theta)$ and its uniform convergence in θ was already proved. The third one converges to zero in probability uniformly in θ as a consequence of the convergence of the second term and the convergence in probability uniform in θ of the first term (to be proved) and the Schwarz inequality. In consequence only remain to study the convergence of the first term. But

$$\mathcal{G}_{1N}=\left(\frac{h^3b_{11}(\theta)}{8}-\frac{3}{2}\right)\frac{1}{N}\sum_{i=0}^{N-1}[P(z_{ih};\theta_0)-P(z_{ih};\theta)]^2+\frac{3}{2}\frac{1}{N}\sum_{i=0}^{N-1}[P(z_{ih};\theta_0)-P(z_{ih};\theta)]^2.$$

Let us consider the first summand. By using the hypothesis H3-(4) we get

$$\sup_{\theta\in\Theta}\left(\frac{1}{N}\sum_{i=0}^{N-1}[P(z_{ih};\theta_0)-P(z_{ih};\theta)]^2\right)\le\sup_{\theta\in\Theta}\left(\frac{1}{N}\sum_{i=0}^{N-1}||\nabla_\theta P(z_{ih};\lambda_1\theta_0+\lambda_2\theta)||^2||\theta_0-\theta||^2\right).$$

Furthermore

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\frac{1}{N}\sum_{i=0}^{N-1}[P(z_{ih};\theta_0)-P(z_{ih};\theta)]^2\right]\leq \mathbf{C}\mathbb{E}[||z_0||^\gamma].$$

Therefore this term is bounded in probability uniformly in θ. This results in the convergence to zero in probability uniformly in θ of the first summand. Let now consider the second summand.

If we define $G(z;\theta) := [P(z;\theta_0) - P(z;\theta)]^2$, it holds

$$\sup_{\theta\in\Theta}|G(z;\theta)| \leq \sup_{\theta\in\Theta}(||\nabla_\theta P(z;\lambda_1\theta_0+\lambda_2\theta)||^2||\theta_0-\theta||^2)\leq \mathbf{C}||z||^\gamma,$$

this domination and the fact that z_{ih} is mixing give the LLN uniform with respect to θ. In consequence the second term converges to $\frac{3}{2}\mathbb{E}[(P(z_0;\theta_0)-P(z_0;\theta))^2]$ uniformly in θ. The remaining terms $\mathcal{G}_{iN}$ can be treated similarly. Indeed the decomposition made for $\mathcal{G}_{2N}$ into three terms gives that the first term tends to zero in probability uniformly in θ, the other two terms in the decomposition do not depend on θ. Hence we get that the original term $\mathcal{G}_{2N}$ converges in probability towards $3\sigma\mathbb{E}[g_1(x_0;\theta_0)]$ uniformly with respect to θ. The terms $\mathcal{G}_{iN}$ $i=3,4,5,7$ can be treated similarly and finally the term $\mathcal{G}_{6N}$ can be handled as was $\mathcal{G}_{1N}$. □

Proof of the assertion in Remark 4.2.

- **Linear oscillator.** In this case $g_1(x;a)=a$ and $g_2(x;b)=b^2$ the matrix J is equal to

$$J=\begin{pmatrix}0 & 1\\ -b^2 & -a\end{pmatrix}.$$

  Thus the exponential of this matrix is a constant.
- **Kramer's oscillator.** In this case again $g_1(x;a)=a$ and by taking for instance the quadratic potential $V(x)=\frac{b_4}{4}x^4+\frac{b_2}{2}x^2$ we have that $g_2(x;\mathbf{b})=b_4x^2+b_2$ where we have set $\mathbf{b}=(b_4,b_2)$. In this form

$$J=\begin{pmatrix}0 & 1\\ -ay-3b_4x^2-b_2 & -a\end{pmatrix}.$$

  Hence $||J(z)||=1+|ay+2b_4x^3+(b_4x^2+b_2)|+|a|$. Obtaining

$$\mathbb{E}[e^{||J(z_0)||}]\leq \mathbf{C}\int_{\mathbb{R}^2}e^{2ay}e^{2b_4|x|^3+b_4x^2}e^{-\frac{2a}{\sigma^2}(\frac{1}{2}y^2+(\frac{b_4}{4}x^4+\frac{b_2}{2}x^2))}\mathrm{d}x\mathrm{d}y<\infty.$$

  Here we have used the explicit expression for the invariant measure for this model given in Wu [21]. Other polynomial potentials provides the same result.
- **Van der Pol oscillator.** Now $g_1(x;\mathbf{a})=a_1x^2-a_2$ where both constants are greater than zero. And again $g_2(x;b)=b^2$. Then

$$J=\begin{pmatrix}0 & 1\\ -2a_1xy-b^2 & -a_1x^2-a^2\end{pmatrix}.$$

  However in the formula (5.16) of Wu it is shown that there exists two positive arbitrary constants $a>0$, and $0<\varepsilon<g_1(x;\mathbf{a})\frac{b^2}{2\sigma^2}$ such that

$$\int e^{(aH(x,y)+(g_1(x;\mathbf{a})\frac{b^2}{2\sigma^2}-\varepsilon)x^4)}\mathrm{d}\mu(z)<\infty, \tag{A.16}$$

where $H(x,y) = \frac{1}{2}y^2 + \frac{b^2}{2}x^2$ and μ is the invariant measure of the system. This last result implies by a simple computation that in this example also $\mathbb{E}[e^{2h||J(z_0)||}] < \infty$. In fact the relation (A.16) implies that $\mathbb{E}[e^{\alpha x_0^4 + \frac{a}{2}y_0^2}] < \infty$, for a certain constant β. Thus
$\mathbb{E}[e^{2h||J(z_0)||}] \leq \mathbf{C}\int_{\mathbb{R}^2} e^{2h(2a_1|xy|)}e^{2ha_1x^2}\mathrm{d}\mu(z) \leq \mathbf{C}\int_{\mathbb{R}^2} e^{2ha_1(x^2+y^2)}e^{2ha_1x^2}\mathrm{d}\mu(z) < \infty$,as long as $h < h_0$, for a certain h_0 small enough.

- **Hypoelliptic FitzHugh-Nagumo model.** In León-Samson [10] the following modified FitzHugh-Nagumo system is considered

$$\begin{aligned}\mathrm{d}X_t &= Y_t\mathrm{d}t \\ \mathrm{d}Y_t &= \frac{1}{\varepsilon}(Y_t(1-\varepsilon-3X_t^2) - X_t(\gamma-1) - X_t^3 - (s+\beta))\mathrm{d}t + \sigma\mathrm{d}W_t\end{aligned}$$

Thus $g_1(x) = \frac{1}{\varepsilon}(\varepsilon + 3x^2 - 1)$ and $g_2(x)x = \frac{1}{\varepsilon}\left[x(\gamma-1) + x_t^3 + (s+\beta)\right]$ and so

$$f(z) = \begin{bmatrix} y \\ -\frac{1}{\varepsilon}(\varepsilon+3x^2-1)y - \frac{x(\gamma-1)}{\varepsilon} - \frac{x^3}{\varepsilon} - \frac{(s+\beta)}{\varepsilon}\end{bmatrix}.$$

In this manner we get

$$J(z) = \begin{pmatrix} 0 & 1 \\ -\frac{1}{\varepsilon}\left[6xy + (\gamma-1) + 3x^2\right] & -\frac{1}{\varepsilon}(\varepsilon+3x^2-1)\end{pmatrix}.$$

Computing the norm of this last matrix we have

$$||J(z)|| \leq 1 + \frac{1}{\varepsilon}\left[6|xy| + (\gamma-1) + 3x^2 + (1+\varepsilon) + 3x^2\right].$$

Hence

$$e^{2h||J(z)||} \leq \mathbf{C}e^{12h(|xy|+x^2)} \leq \mathbf{C}e^{2h(3y^2+9x^2)}.$$

In León-Samson [10] a Liapunov function $\Psi(z)$ for this system was found that have the following asymptotic behavior

$$\Psi(z) \sim e^{aH(z)}e^{by}e^{d|x|^3}, \text{ when } |x| \to \infty.$$

for certain constants a, b, d and $H(z) = \frac{y^2}{2} + \frac{1}{\varepsilon}(\frac{x^4}{4} + \frac{\gamma-1}{2}x^2 + (s+\beta)x)$ is the Hamiltonian of the system. We know from Wu [21] that denoting by μ the invariant measure of the system it holds $\int \Psi(z)\mathrm{d}\mu(z) < \infty$. In consequence we have

$$\mathbb{E}[e^{2h||J(z_0)||}] \leq \mathbf{C}\int e^{2h(9x^2+3y^2)}\mathrm{d}\mu(z) < \infty,$$

as long as $h < h_0$, for a certain h_0 small enough.

□

Remark A.1. With an analogous procedure it can be proved that

$$\mathbb{E}[e^{(2h||J(z_0||)}||J(z_0)||^4||f(z_0)||^2] < \infty$$

holds for the all the precedent models.

Proof of the assertion in Remark 4.7.

- **Linear oscillator.** In this case $P(z_0;\theta) = -ay_0 - b^2x_0$ and $\theta = (a,b)$. So

$$\nabla_\theta P(z_0;\theta) = (-y_0, -2bx_0).$$

- **Kramer's oscillator.** In this case $g_1(x;a) = a$ and by taking for instance the quartic potential $V(x) = \frac{b_4}{4}x^4 + \frac{b_2}{2}x^2$ we have $P(z_0;\theta) = -ay_0 - \frac{b_2}{2}x_0 - \frac{b_4}{4}x^3$ and $\theta = (a, b_2, b_4)$. Hence

$$\nabla_\theta P(z_0;\theta) = \left(-y_0, \frac{-x_0}{2}, \frac{-3b_4x_0^2}{2}\right)$$

- **Van der Pol oscillator.** Now $g_1(x;\mathbf{a}) = a_1x^2 - a_2$, with $a_1, a_2 > 0$ and $g_2(x;b) = b^2$. In this case $P(z_0;\theta) = -(a_1x_0^2 - a_2)y_0 - b^2x_0$ and $\theta = (a_1, a_2, b)$. Then

$$\nabla_\theta P(z_0;\theta) = \left(-x_0^2y_0, y_0, -2bx_0\right)$$

- **Hypoelliptic FitzHugh-Nagumo model.** In this case

$$g_1(x)y = \frac{1}{\varepsilon}(\varepsilon + 3x^2 - 1)y \quad \text{and} \quad g_2(x)x = \frac{1}{\varepsilon}\left[x(\gamma - 1) + x_t^3 + (s + \beta)\right].$$

Therefore

$$P(z_0;\theta) = -\frac{1}{\varepsilon}(\varepsilon + 3x_0^2 - 1)y_0 - \frac{1}{\varepsilon}\left[x_0(\gamma - 1) + x_0^3 + (s + \beta)\right],$$

and

$$\theta = (\varepsilon, \gamma, s, \beta).$$

We have

$$\nabla_\theta P(z_0;\theta) = \frac{1}{\varepsilon}\left(\frac{(3x_0^2y_0 - 1)y_0 + x_0(\gamma - 1) + x_0^3 + s + \beta}{\varepsilon}, -x_0, -1, -1\right).$$

□

Acknowledgements. The authors want to thank the referees who did an excellent job in reviewing this article meticulously. Their observations and suggestions improved greatly this work.

References

[1] L. Abbaoui and A. Bendjeddou, Exact stationary solutions for a class of nonlinear oscillators. *Adv. Appl. Stat.* **1** (2001) 99–106.
[2] P. Billingsley, Convergence of Probability Measures. Wiley, New York (1968).
[3] P. Cattiaux, J. León and C. Prieur, Estimation for stochastic damping hamiltonian systems under partial observations – I. Invariant density. *Stoch. Process Appl.* **124** (2014) 1236–1260.

[4] P. Cattiaux, J. León and C. Prieur, Estimation for stochastic damping hamiltonian systems under partial observations – II Drift Term. *ALEA* **11** (2014) 1236–1260.

[5] V. Genon-Catalot, Maximum contrast estimation for diffusion processes from discrete observations. *Statistics* **21** (1990) 99–116.

[6] V. Genon-Catalot, T. Jeantheau and C. Laredo, Parameter estimation for discretely observed stochastic volatility models. *Bernoulli* **5** (1999) 855–872.

[7] V. Genon-Catalot, T. Jeantheau and C. Laredo, Stochastic volatility models as hidden markov models and statistical applications. *Bernoulli* **6** (2000) 1051–1079.

[8] M. Gitterman, The Noisy Oscillator: The First Hundred Years, From Einstein Until Now, 1st edn. World Scientific Publishing Co. Pte. Ltd. (2005).

[9] T. Lelièvre, M. Rousset and G. Stoltz, Free Energy Computations a Mathematical Perspective. World Scientific (2010).

[10] J.R. León and A. Samson, Hypoelliptic Stochastic Fitzhugh-Nagumo Neuronal Model: Mixing, Up-Crossing and Estimation of the Spike Rate. *Ann. Appl. Probab.* **28** (2018) 2243–2274.

[11] B. Lindner and L. Schimansky-Geier, Analytical approach to the stochastic FitzHugh-Nagumo system and coherence resonance. *Phys. Rev. E* **60** (1999) 7270–7276.

[12] J. Mathews and K. Fink, Numerical Methods Using MATLAB, 1st edn. Prentice Hall (1999).

[13] J. Nicolau, Nonparametric Estimation of second-order stochastic differential equations. *Econ. Theory* **23** (2007) 880–898.

[14] T. Ozaki, Statistical identification of nonlinear random vibration systems. *J. Appl. Mech.* **111** (1989) 186–191.

[15] T. Ozaki, A bridge between nonlinear time series models and nonlinear stochastic dynamical systems: a local linearization approach. *Stat. Sin.* **2** (1992) 113–135.

[16] Y. Pokern, A. Stuart and P. Wiberg, Parameter estimation for partially observed hypoelliptic diffusions. *J. R. Stat. Soc.* **71** (2009) 49–73.

[17] J.B. Roberts and P.D. Spanos, Random Vibration and Statistical Linearization. Dover (2003).

[18] A. Samson and M. Thieullen, Contrast estimator for completely or partially observed hypoelliptic diffusion. *Stoch. Process. Appl.* **122** (2012) 2521–2552.

[19] D. Talay, Stochastic hamiltonian systems:exponential convergence to the invariant measure, and discretization by the implicit euler scheme. *Markov Process. Relat. Fields* **8** (2002) 1–36.

[20] A. Wald, Note on the consistency of the maximum likelihood estimate. *Ann. Math. Stat.* **20** (1949) 595–601.

[21] L. Wu, Large and moderate deviations and exponential convergence for stochastic damping hamiltonian systems. *Stoch. Process. Appl.* **91** (2001) 205–238.