Stat Inference Stoch Process
DOI 10.1007/s11203-016-9137-1

On maximum likelihood estimation of the drift matrix of a degenerated O–U process

Ana Prior[1] · **Marina Kleptsyna**[2] ·
Paula Milheiro-Oliveira[3]

Received: 28 August 2014 / Accepted: 19 May 2016

Abstract In this work, we consider a $2n$-dimension Ornstein–Uhlenbeck (O–U) process with a singular diffusion matrix. This process represents a currently used model for mechanical systems subject to random vibrations. We study the problem of estimating the drift parameters of the stochastic differential equation that governs the O–U process. The maximum likelihood estimator proposed and explored in Koncz (J Anal Math 13(1):75–91, 1987) is revisited and applied to our model. We prove the local asymptotic normality property and the convergence of moments of the estimator. Simulation studies based on representative examples taken from the literature illustrate the obtained theoretical results.

Keywords Maximum likelihood estimator · Ornstein–Uhlenbeck process · Local asymptotic normality property · Singular diffusion matrix

Mathematics Subject Classification 62F12 · 60H10

1 Introduction

Many real-life mechanical and structural systems respond dynamically to random environmental loads such as wind, wave or earthquake forces sometimes leading to stochastic estimation problems (see for instance Lutes and Sarkani 1997; Peeters and Roeck 2001). Electrical circuits and power systems are also examples of systems exhibiting stochastically disturbed dynamics. The $2n$-dimension Ornstein–Uhlenbeck (O–U) process appears in the

✉ Ana Prior
afpontes@adm.isel.pt

1 Instituto Superior de Engenharia de Lisboa and CMUP, Instituto Politécnico de Lisboa, Rua Conselheiro Emídio Navarro, 1, 1959–007 Lisboa, Portugal

2 Laboratoire Manceau de Mathématiques, Université du Maine, Avenue Olivier Messiaen, 72085 Le Mans Cedex 9, France

3 Faculdade de Engenharia and CMUP, Universidade do Porto, Rua Dr. Roberto Frias, 4200–465 Porto, Portugal

Published online: 28 May 2016

engineering literature as a model for mechanical or electrical systems subject to random excitations. The associated problem of estimating the drift matrix, based on observations of the stochastic response process, has been investigated in e.g. Perninge et al. (2011). The general multidimensional O–U model and the parameter estimation problem are studied in detail in Le Breton (1977), Arató (1982) and Rao (1999). However, most of the results obtained in those studies, regarding the Maximum Likelihood Estimator (MLE) of the drift matrix, require the invertibility of the diffusion matrix. This is the case of the asymptotic normality of the estimator. The MLE for the drift matrix of a general multidimensional O–U model, with no restrictions on the invertibility of the diffusion matrix, has later been proposed and explored in Koncz (1987), where the properties of the Laplace transform and the moments of the MLE are also investigated. Irrespectively of the stability of the process, the MLE is known to be unbiased and consistent (cf. Basak and Lee 2008 and the related work of Lin and Lototsky 2011), under some rank condition on the matrices governing the dynamics of the O–U process.

While extensively studied in the past, asymptotic properties of the MLE remain unknown in a number of cases. In particular, proofs of asymptotic normality in Arató (1982), Rao (1999) and Kutoyants (2004) require certain regularity and ergodicity conditions, under which the estimator is asymptotically normal with the rate of convergence $\sqrt{T}$. In Koncz (1987) Novikov's approach is generalised and applied to the general estimation problem of a stable drift matrix, for the O–U process in any dimension, and without any restrictions on the diffusion matrix. However Koncz (1987) does not address the large sample asymptotic behaviour of the MLE nor the convergence of moments and its efficiency. The results in Koncz (1987) are based on the computation of the matrix Laplace transform. Recently, Lin and Lototsky (2014) identified and investigated all possible modes of asymptotic distributional behaviour of the MLE for a class of Stochastic Differential Equations (SDEs) in dimension two with the emphasis on the non-ergodic case. In particular, local asymptotic properties and the convergence rates were found in all stability regions. The problem in a higher dimension still remains open.

In this paper, we investigate the asymptotic properties of the MLE for the drift matrix of a particular 2n-dimension O–U process with a singular diffusion matrix. This problem corresponds to an extension of the work carried out in Lin and Lototsky (2014) to higher dimensions in the ergodic case. It is also related to the study presented in Brockwell et al. (2007), where the estimation of the continuous time auto-regressive process of order p (CAR(p)) is analysed and the asymptotic normality of the MLE is proved. The paper Brockwell et al. (2007) considers a particular case of the problem from this paper, corresponding to $2n = p = 2$. The rank condition mentioned in Basak and Lee (2008) trivially verifies in our case, implying consistency and asymptotic efficiency of the MLE. Therefore, we will be interested in the Local Asymptotic Normality (LAN) property and in the convergence of moments. The results from Koncz (1987) are not helpful in this context since they require the computation of the Laplace transform matrix and we are interested in the Laplace transform of a quadratic form.

This paper implements the program proposed in Ibragimov and Khasminskii (1981) [Theorem I.10.1] to derive asymptotic properties of the MLE. Our approach uses the representation of the estimation error as the linear transformation $\hat{\theta}_T - \theta = \langle M_T \rangle^{-1} \cdot M_T$ of a martingale M_T and is based on its Laplace transform. The main idea is to deduce all the properties of the MLE, including the convergence of moments, from the weak convergence of the scaled likelihoods. The asymptotic covariance matrix of the estimator is given in terms of the solution of a Lyapunov equation. In addition, we establish the convergence of all the moments of $\sqrt{T}(\hat{\theta}_T - \theta)$, which does not come as a direct consequence of the ergodicity. The proposed method will work also for partially observed models and it can be extended to more general classes of noises.

In a number of special cases the covariance matrix of the asymptotic estimation error has a certain useful structure. For example, for the linear harmonic stochastic oscillator with n degrees of freedom it has a block-diagonal form. Furthermore, in the cases where this matrix has certain commutative properties, the covariance matrix can be found explicitly. Simulation studies illustrate the asymptotic behaviour of the estimator and show a good empirical fit with the obtained theoretical results in the non-asymptotic setup.

The paper is organized as follows: in Sect. 2 we describe the estimation problem under study, we introduce the notations used throughout this work (Sect. 2.1) and we formulate the main results (Sect. 2.2). Section 3 contains some auxiliary material and Sect. 4 gives the proofs of the main results. The computation of the Fisher information matrix is explored in Sect. 5. The results obtained by numerical simulations are presented in Sect. 6. Section 7 summarises the conclusions of our study.

2 Statement of the problem and the main results

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_t, P)$ be a filtered probability space. We consider the O–U process with a particular structure

$$dX_t = AX_t dt + B^{\frac{1}{2}} dW_t,\ t \geq 0, \quad X_0 = 0, \tag{1}$$

where $\{X_t\}, \{W_t\} \in \mathbb{R}^{2n}$ and $\{W_t\}$ a standard Wiener process. The matrices A and $B^{\frac{1}{2}}$ are given by

$$A = \begin{pmatrix} 0 & Id_{n\times n} \\ -\theta_1 & -\theta_2 \end{pmatrix} \qquad \text{and} \qquad B^{\frac{1}{2}} = \begin{pmatrix} 0 & 0 \\ 0 & \Sigma^{\frac{1}{2}} \end{pmatrix}, \tag{2}$$

where θ_1 and θ_2 are real $n \times n$ matrices and Σ is a real $n \times n$ symmetric positive definite matrix.

We will assume that

(A) A is a stable matrix, that is, all eigenvalues of A have negative real parts.

Notice that if, for instance,

(A′) θ_1 and θ_2 are commutative symmetric positive definite matrices

then (A) holds.

The matrices θ_1 and θ_2 are unknown and our problem is to estimate these matrices, based on the observation of $\{X_t\}$ on the interval $[0, T]$. Since the matrix Σ can be computed with probability one through the quadratic variation of X_t on any finite interval, we consider Σ known.

Notations Throughout the paper we will use the following notations:

U^* denotes the transpose of a matrix U;

$tr(U)$ denotes the trace of a square matrix U;

The symbol $\otimes$ stands for the Kronecker product of two matrices;

$\|\cdot\|$ is the Euclidean norm;

If $U = (U^1 \quad U^2 \cdots U^p)$ is an $m \times p$ matrix then $vec(U)$ is an mp column-vector built from the columns of the matrix U, stacked one by one. More precisely

$$vec(U) = \begin{pmatrix} U^1 \\ U^2 \\ \cdots \\ U^p \end{pmatrix};$$

If η is a $p \times n$ random matrix, we will write that $\eta \sim \mathbf{M}\mathcal{N}_{p\times n}(0, \Sigma_{p\times p}, \mathcal{I}(A))$ (matrix normal distribution) when $vec(\eta) \sim \mathcal{N}_{pn}(0, \mathcal{I}(A)_{n\times n} \otimes \Sigma_{p\times p})$.

An equivalent representation of the process $\{X_t\}$ is the solution $X_t = (X_t^1, X_t^2)^*$ of the system

$$\begin{cases} dX_t^1 = X_t^2 dt \\ dX_t^2 = -\Theta X_t dt + \Sigma^{\frac{1}{2}} dW_t^2 \end{cases}, \tag{3}$$

with an unknown $n \times 2n$ matrix $\Theta = (\theta_1, \theta_2)$ and the standard Wiener process $\{W_t^2\}$ in $\mathbb{R}^n$.

The estimation problem is, obviously, equivalent to estimation of the column vector $\theta = (\theta_i) \in \mathbb{R}^{n(2n)\times 1}$, which represents row-by-row concatenation of the n last rows of matrix A, based on observation of $\{X_t\}$ over the interval $[0, T]$.

In real world applications arising in mechanical, structural or electrical systems, the drift matrix appears in the particular form

$$A = \begin{pmatrix} 0 & Id_{n\times n} \\ -M^{-1}K & -M^{-1}C \end{pmatrix}, \tag{4}$$

where M is a known invertible matrix, and K, C are unknown matrices that have to be estimated.

2.1 The maximum likelihood estimator

For a fixed value of the parameter Θ, let P_Θ denote the probability measure induced by X on the function space $C[0, T]$ and let F_T^X be the natural filtration of X. By the Girsanov theorem the likelihood, i.e. the Radon-Nikodym derivative of P_Θ with respect to the reference measure on $C[0, T]$, corresponding to $\Theta = 0$, restricted to F_T^X, is given by the conditional expectation

$$\mathcal{L}_T(\Theta, X) = \tilde{E}\left(\frac{dP_\Theta}{d\tilde{P}_0} | \mathcal{F}_T^X\right),$$

where

$$\frac{d\tilde{P}_0}{dP_\Theta} = \exp\left(\int_0^T (\Sigma^{-\frac{1}{2}}\Theta X_t)^* dW_t^2 - \frac{1}{2}\int_0^T (\Sigma^{-\frac{1}{2}}\Theta X_t)^* \Sigma^{-\frac{1}{2}}\Theta X_t \, dt\right).$$

Or, equivalently,

$$\mathcal{L}_T(\Theta, X) = \exp\left(-\int_0^T X_t^* \Theta^* \Sigma^{-1} \, dX_t^2 - \frac{1}{2}\int_0^T X_t^* \Theta^* \Sigma^{-1} \Theta X_t \, dt\right). \tag{5}$$

The MLE $\hat{\Theta}_T = \underset{\Theta}{\operatorname{argmax}} \, \mathcal{L}_T(\Theta, X)$, associated with model (3), is given by

$$\hat{\Theta}_T = -\int_0^T dX_t^2 \cdot X_t^* \cdot \left(\int_0^T X_t X_t^* \, dt\right)^{-1}. \tag{6}$$

Hence,

$$\hat{\Theta}_T - \Theta = -\Sigma^{\frac{1}{2}} \left(\int_0^T X_t \, dW_t^{2*}\right)^* \cdot \left(\int_0^T X_t X_t^* \, dt\right)^{-1}. \tag{7}$$

Equivalently, $\hat{\Theta}_T$ can be written as

$$\hat{\Theta}_T = Diag\left(\left(\int_0^T X_t X_t^* \, dt\right)^{-1}_{n\,blocks}\right)\begin{pmatrix} \int_0^T X_t dX_t^{(n+1)} \\ \int_0^T X_t dX_t^{(n+2)} \\ \vdots \\ \int_0^T X_t dX_t^{(2n)} \end{pmatrix}$$
$$= \left(Id_{n\times n} \otimes \left(\int_0^T X_t X_t^* \, dt\right)^{-1}\right) \cdot \begin{pmatrix} \int_0^T X_t dX_t^{(n+1)} \\ \int_0^T X_t dX_t^{(n+2)} \\ \vdots \\ \int_0^T X_t dX_t^{(2n)} \end{pmatrix}. \quad (8)$$

The consistency and efficiency for θ are established in Basak and Lee (2008). Since matrix B is singular, asymptotic normality does not literally follow from Rao (1999), nevertheless, ergodicity of the process $\{X_t\}$ is sufficient to ensure the LAN property (cf. Arató (1982)[Theorem 4.6-2]). However, efficiency of the MLE and control of the moments

$$E\left(\int_0^T \|X_t X_t^*\| dt\right)^{-k}$$

are not directly implied by the ergodicity and therefore are not straightforward. One possible approach to these asymptotic properties, based on the Ibragimov–Khasminskii program, is presented in the next section.

Regarding estimation of the matrices K and C in (4), note that $\hat{K}$ and $\hat{C}$ can be obtained by a linear transformation of the estimators of Θ or θ:

$$(\hat{K} \quad \hat{C}) = -M(vec^{-1}(\hat{\theta}))^*,$$

where vec^{-1} is the inversion of the vectorization operation vec, defined above.

2.2 Main results

In this section we show that our model is regular in Ibragimov–Khasminskii's sense, which ensures that the LAN property holds, as well as convergence of all moments (Ibragimov and Khasminskii 1981). The verification of conditions (C1)–(C3) is based on computation of the Laplace transform.

The main result of this paper is the following:

Theorem 1 *The MLE given by* (6) *satisfies*

$$\sqrt{T}(\hat{\Theta}_T - \Theta) \xrightarrow[T\to+\infty]{\mathcal{L}} \eta,$$

where η *is a matrix valued random variable,* $\eta \sim \mathbf{M}\mathcal{N}_{n\times 2n}\left(0, \mathcal{I}^{-1}(A), \Sigma\right)$*, and* $\mathcal{I}(A)$ *is the unique solution of the Lyapunov equation*

$$A\mathcal{I}(A) + \mathcal{I}(A)A^* + B = 0. \quad (9)$$

Moreover, the following convergence of moments holds true

$$E\|\sqrt{T}(\hat{\theta}_T - \theta)\|^k \underset{T\to+\infty}{\longrightarrow} E\|\eta\|^k, \quad k \geq 1.$$

The next result is of independent interest, and it gives the exponential moments convergence.

Proposition 1 *For any $y \in \mathbb{R}$ and any symmetric positive definite $2n \times 2n$ matrix Q there exists T_0 such that for any $T \geq T_0$ we have*

$$E \exp\left(\frac{y}{2T}\int_0^T X_t^* Q X_t dt\right) = \exp\left(\frac{y}{2} tr(\mathcal{I}(A)Q)\right)\left(1 + \frac{c|y|}{T}\|Q\|\right),$$

where $\mathcal{I}(A)$ is the unique solution of the Lyapunov equation (9) and the term c is uniformly bounded with respect to y, Q and T.

The invertibility of $\mathcal{I}(A)$ is ensured by the controllability of the pair $(A, B^{\frac{1}{2}})$ while the stability of A (assumption (A)) ensures that the solution of (9) is unique. The controllability of the pair $(A, B^{\frac{1}{2}})$ is a consequence of the rank condition $rank(B^{\frac{1}{2}}|AB^{\frac{1}{2}}|\cdots|A^{2n-1}B^{\frac{1}{2}}) = 2n$, which, in our case, is verified by elementary calculations.

Proposition 1 is proved for all $y \in \mathbb{R}$. For $y < 0$ this property can be proved in a very simple manner by using the dominated convergence theorem. The case $y > 0$ implies convergence of the exponential moments.

3 Auxiliary results

3.1 Algebraic and differential Riccati equations

This subsection gathers general results on algebraic Riccati equations and small perturbations of a matrix Riccati equation, which will be useful when proving the main results.

Let us recall the explicit formula for the positive solution D_μ of the algebraic Riccati equation, with symmetric positive definite matrix Q, (see Kucera 1973)

$$D_\mu A + A^* D_\mu - 2D_\mu B D_\mu = \frac{\mu}{2}Q. \tag{10}$$

When $\mu < 0$, we have

$$D_\mu = FG^{-1},$$

where $F = [f_1, f_2, \ldots, f_m]$, $G = [g_1, g_2, \ldots, g_m]$ and $\begin{pmatrix} f_i \\ g_i \end{pmatrix}$, $i = 1, \ldots m$ are the eigenvectors of $M = \begin{pmatrix} A & -B \\ \frac{\mu}{2}Q & -A^* \end{pmatrix}$, in such an order that G is invertible.

The solution exists for a small *positive* μ due to the Cayley–Hamilton theorem, due to the analytical properties of the eigenvectors of M, and

$$\|D_\mu\| \leq c|\mu|\|Q\|$$

with a constant c that does not depend on μ and Q. The positive solution D_ν of

$$D_\nu A + A^* D_\nu - 2\nu D_\nu B D_\nu = \frac{\mu}{2}Q$$

is such that $\|D_\nu - D_0\| \leq c|\nu|$ where $D_0 A + A^* D_0 - \frac{\mu}{2}Q = 0$.

Now we consider the differential equation and the initial condition

$$\begin{cases} \dot{\gamma}_t = A\gamma_t + \gamma_t A^* + \mu\gamma_t Q\gamma_t + B \\ \gamma_0 = 0 \end{cases}. \tag{11}$$

Its solution exists for every $\mu < 0$ and for $\mu > 0$ small enough. Moreover, $\gamma_t(\mu)$ is an analytic function of μ as long as $|\mu| \le \mu_0$ for some $\mu_0 > 0$.

This equation can be linearised and its solution can be written as

$$\gamma_t = \Psi_1^{-1}(t)\Psi_2(t),$$

where $\begin{pmatrix}\dot{\Psi}_1 & \dot{\Psi}_2\end{pmatrix} = (\Psi_1 \quad \Psi_2)\,\Lambda_\mu$, $\Psi_1(0) = Id_{2n\times 2n}$, $\Psi_2(0) = 0$, with $\Lambda_\mu = \begin{pmatrix}-A & B\\ 0 & A^*\end{pmatrix} + \mu H$ and $H = \begin{pmatrix}0 & 0\\ Q & 0\end{pmatrix}$.

The following property holds: $\int_0^T tr(\mu\gamma_t Q)dt = -T tr(A) + \det \Psi_1(T)$.

For all $T > 0$ we have

$$\Psi_1(T) = (Id_{2n\times 2n} \quad 0)\, G_\mu \mathcal{D}(e^{T\lambda_k(\mu)})\, G_\mu^{-1} \begin{pmatrix} Id_{2n\times 2n}\\ 0\end{pmatrix},$$

where $G_\mu = G_0 + P_\mu$, G_0 is an upper triangular block matrix, P_μ is such that $\|P_\mu\| \le c\mu\|Q\|$, for some constant c, and $\mathcal{D}(\cdot)$ denotes a diagonal matrix of order $4n$.

3.2 Computation of the Laplace transform

For the solution $\{X_t\}_{0\le t\le T}$ of (1) we compute the Laplace transform

$$L_T(\mu, Q, X) = E\exp\left(\mu\int_0^T X_t^* Q X_t dt\right),$$

where Q is a symmetric positive definite matrix and $\mu \le \mu_0$ for some $\mu_0 > 0$.

The Novikov approach, generalised in Koncz (1987), gives

$$L_T(\mu, Q, X) = \exp\left(-T tr(BD_\mu)\det(Id_{2n\times 2n} - 2D_\mu\Gamma(T)\right)^{-\frac{1}{2}},$$

with Γ, $\tilde{a}$ and D_μ satisfying

$$\begin{aligned}\dot{\Gamma} &= \tilde{a}\Gamma + \Gamma\tilde{a}^* + B, \quad \Gamma(0) = 0,\\ \tilde{a} &= A - 2BD_\mu\end{aligned}$$

and the Eq. (10) in Sect. 3.1. For a discussion on small perturbation properties of the solution D_μ see also Sect. 3.1.

In the paper Kleptsyna et al. (2008), the filtering approach is proposed, which in a partially observed diffusion setting gives

$$L_T(\mu, Q, X) = \exp\left(\frac{\mu}{2}\int_0^T tr(Q\gamma_s)ds\right),$$

where γ_s is a solution of Eq. (11).

The substance of the two approaches is different, but of course they lead to the same result. The connection between them can be established in the same way as in Kleptsyna et al. (2001).

4 Proof of the main results

First, we will prove Proposition 1, using a small perturbation argument (see auxiliary material in Sect. 3). In order to prove Proposition 1 we will state and prove two lemmas. To this end we define $\Lambda_\mu = \begin{pmatrix} -A & B \\ \mu Q & A^* \end{pmatrix}$ and denote by λ_j the eigenvalues of Λ_μ such that $\Re e(\lambda_j(\mu)) > 0$.

Lemma 1 *For $\mu = \frac{y}{T}$, $y \in R$,*

$$E \exp\left(\frac{y}{2T}\int_0^T X_t^* Q X_t dt\right) = \exp\left(\frac{y}{2}\sum_{j=1}^{2n} \lambda'_j(0)\right)\left(1 + \frac{y}{T}C(||Q||)\right)$$

where $\lambda'_j(0)$ is the derivative with respect to μ of $\lambda_j(\mu)$ at $\mu = 0$.

Note that, for small values of μ, the spectrum of Λ_μ, $sp\left(\Lambda_\mu\right)$, contains the eingenvalues $\lambda_j(\mu)$ such that $\Re e(\lambda_j(\mu)) > 0$ and the eigenvalues $\lambda_i(\mu)$ such that $\Re e(\lambda_i(\mu)) < 0$, therefore $sp\left(\Lambda_\mu\right)$ is approximated by $sp(\Lambda_0) = \left\{-\lambda_j(A)\right\} \cup \left\{\lambda_j(A)\right\}$.

Proof of Lemma 1 For $\mu \leq \mu_0$, define the Laplace transform

$$L_T(\mu, Q, X) = E \exp\left(\frac{\mu}{2}\int_0^T X_t^* Q X_t dt\right).$$

From Kleptsyna et al. (2008) and the standard method of linearisation of the matrix Riccati differential equations (see Sect. 3), we obtain

$$L_T(\mu, Q, X) = \exp\left(\frac{\mu}{2} T\, tr(A)\right)(\det \Psi_1(T))^{-\frac{1}{2}},$$

where $(\Psi_1 \quad \Psi_2)$ is the solution of

$$\begin{cases} \left(\dot{\Psi}_1 \quad \dot{\Psi}_2\right) = \left(\Psi_1 \quad \Psi_2\right)\Lambda_\mu \\ \left(\Psi_1(0) \quad \Psi_2(0)\right) = (Id_{2n\times 2n} \quad 0) \end{cases},$$

with

$$\Lambda_\mu = \Lambda_0 + \mu H, \quad \Lambda_0 = \begin{pmatrix} -A & B \\ 0 & A^* \end{pmatrix} \quad \text{and} \quad H = \begin{pmatrix} 0 & 0 \\ Q & 0 \end{pmatrix}.$$

This means that

$$\Psi_1(T) = (Id_{2n\times 2n} \quad 0)\, G_\mu \mathcal{D}(e^{T\lambda_k(\mu)})\, G_\mu^{-1}\begin{pmatrix} Id_{2n\times 2n} \\ 0 \end{pmatrix},$$

where G_μ is, asymptotically, an upper triangular block matrix, and $\mathcal{D}(\cdot)$ is a diagonal matrix of order $4n$, in the sense discussed in Sect. 3. As a consequence, using algebraic properties of the determinant, we have

$$\det \Psi_1(T) = \exp\left(T\sum_{j=1}^{2n}\lambda_j(\mu)\right)(1 + \mu C(||Q||)).$$

Thus,

$$L_T(\mu, Q, X) = \exp\left(\frac{\mu}{2} T \sum_{j=1}^{2n} (\lambda_j(\mu) - \lambda_j(0))\right)(1 + \mu C(||Q||))$$

and by the Taylor's expansion on the eigenvalues $\lambda_j(\mu)$ of Λ_μ such that $\Re e(\lambda_j(\mu)) > 0$ with $\mu = \frac{y}{T}$, we get

$$L_T\left(\frac{y}{T}, Q, X\right) = \exp\left(\frac{y}{2} \sum_{j=1}^{2n} \lambda'_j(0)\right)\left(1 + \frac{y}{T} C(||Q||)\right).$$

□

Lemma 2 *With the notation established in Lemma 1, the following equality holds*

$$\sum_{j=1}^{2n} \lambda'_j(0) = tr(\mathcal{I}(A)Q),$$

where $\mathcal{I}(A)$ is the unique solution of (9).

Proof of Lemma 2 Define

$$P(\lambda, \mu) = \det(\Lambda_0 + \mu H - \lambda Id_{4n\times 4n}).$$

Applying the Implicit Function Theorem to the characteristic equation $P(\lambda, \mu) = 0$ we compute the derivative of λ with respect to μ in 0

$$\lambda'_\mu(0) = -\frac{P'_\mu(\lambda, 0)}{P'_\lambda(\lambda, 0)}, \tag{12}$$

where

$$P'_\mu(\lambda, 0) = \det(\Lambda_0 - \lambda Id_{4n\times 4n})\, tr((\Lambda_0 - \lambda Id_{4n\times 4n})^{-1} H),$$
$$P'_\lambda(\lambda, 0) = -\det(\Lambda_0 - \lambda Id_{4n\times 4n})\, tr((\Lambda_0 - \lambda Id_{4n\times 4n})^{-1}).$$

By some algebra we find that

$$tr((\Lambda_0 - \lambda Id_{4n\times 4n})^{-1} H) = tr(\Phi Q)$$

where $\Phi = (A + \lambda Id_{2n\times 2n})^{-1} B (A^* - \lambda Id_{2n\times 2n})^{-1}$.

We apply (12) to the derivative λ'_j of the eigenvalue λ_j with respect to μ and obtain

$$\sum_{j:\, \lambda_j(0)\in Sp(-A)} \lambda'_j(0) = \sum_{j:\, \lambda_j\in Sp(-A)} \frac{tr(\Phi(\lambda_j)Q)\det\left(\Lambda_0 - \lambda_j Id_{4n\times 4n}\right)}{\det\left(\Lambda_0 - \lambda_j Id_{4n\times 4n}\right) tr\left(\Lambda_0 - \lambda_j Id_{4n\times 4n}\right)^{-1}}.$$

In order to complete the proof of the lemma, we must check that

$$\mathcal{I}(A) = \sum_{j:\, \lambda_j\in Sp(-A)} \frac{\Phi(\lambda_j)\det\left(\Lambda_0 - \lambda_j Id_{4n\times 4n}\right)}{\det\left(\Lambda_0 - \lambda_j Id_{4n\times 4n}\right) tr\left(\Lambda_0 - \lambda_j Id_{4n\times 4n}\right)^{-1}} \tag{13}$$

satisfies the Lyapunov equation (9).

On one hand,

$$\Phi = -\sum_{k\geq 0} \left(A + \lambda_j Id_{2n\times 2n}\right)^{-1} B \frac{(A^*)^k}{\lambda_j^{k+1}}$$

with

$$(A + \lambda_j Id_{2n\times 2n})^{-1} = GD\left(\frac{1}{\lambda_i(0) + \lambda_j}\right) G^{-1},$$

for some orthogonal matrix G. On the other hand,

$$\det(\Lambda_0 - \lambda_j Id_{4n\times 4n}) = \prod_{k=1}^{4n} (\lambda_k(0) - \lambda_j)$$

and

$$tr(\Lambda_0 - \lambda_j Id_{4n\times 4n})^{-1} = \sum_{m=1}^{4n} \frac{1}{\lambda_m(0) - \lambda_j}.$$

Thus,

$$\det(\Lambda_0 - \lambda_j Id_{4n\times 4n})^{-1} tr(\Lambda_0 - \lambda_j Id_{4n\times 4n})^{-1} = \sum_{m=1}^{4n} \prod_{k\neq m} (\lambda_k(0) - \lambda_j).$$

Hence,

$$\frac{(A + \lambda_j Id_{4n\times 4n})^{-1} \det\left(\Lambda_0 - \lambda_j Id_{4n\times 4n}\right)}{\det\left(\Lambda_0 - \lambda_j Id_{4n\times 4n}\right) tr\left(\Lambda_0 - \lambda_j Id_{4n\times 4n}\right)^{-1}} = G\bar{D}G^{-1},$$

with the diagonal matrix $\bar{D}$ with entries

$$\frac{\frac{1}{\lambda_i(0) + \lambda_j} \prod_{k=1}^{4n} \left(\lambda_k(0) - \lambda_j\right)}{\sum_m^{4n} \prod_{k\neq m} \left(\lambda_k(0) - \lambda_j\right)}.$$

Notice that these diagonal entries equal 1 if $i = j$ and equal 0 otherwise. Hence expression (13) can be simplified and it can be rewritten as

$$\mathcal{I}(A) = \sum_{k\geq 0} GD\left(\left(-\frac{1}{\lambda_i(0)}\right)^{k+1}\right) G^{-1} B A^{*k} = \sum_{k\geq 0} (-1)^{k+1} A^{-(k+1)} B A^{*k}.$$

Elementary linear algebra shows that this matrix satisfies Eq. (9).

□

The assertion of Proposition 1 follows from Lemmas 1 and 2.
Now, we return to the proof of the main result.

Proof of Theorem 1 To prove Theorem 1 we will check that Ibragimov–Khasminskii's conditions (Ibragimov and Khasminskii (1981) [Theorem I.10.1]) hold. Suppose that $\Theta \in \Xi$, where Ξ is a bounded open convex subset of the space of $n \times 2n$ matrices verifying assumption (A′) in Sect. 2.

First, we need to write the expression for the likelihood ratio

$$Z_T(U) = \frac{\mathcal{L}_T\left(\Theta + \frac{1}{\sqrt{T}}U, X\right)}{\mathcal{L}_T(\Theta, X)},$$

where $\mathcal{L}_T$ is the likelihood function given by (5) and U is an arbitrary $n \times 2n$ matrix such that $\Theta + \frac{1}{\sqrt{T}}U \in \Xi$.

Notice that

$$Z_T(U) = \exp\left(-\frac{1}{\sqrt{T}}\int_0^T X_t^* U^* \Sigma^{-\frac{1}{2}} dW_t^2 - \frac{1}{2T}\int_0^T X_t^* U^* \Sigma^{-1} U X_t\, dt\right). \tag{14}$$

Let us prove that conditions (C1) to (C3) hold:

(C1) $Z_T(U) \xrightarrow[T\to+\infty]{\mathcal{L}} \exp\left(vec^*(U^*)vec(\zeta) - \frac{1}{2}vec^*(U^*)\cdot(\Sigma^{-1}\otimes\mathcal{I}(A))\cdot vec(U^*)\right)$, for some matrix-valued random variable $\zeta \sim \mathbf{M}\mathcal{N}_{n\times 2n}\left(0, \mathcal{I}(A), \Sigma^{-1}\right)$.
Indeed, (14) can be written as

$$Z_T(U) = \exp\left(-\frac{1}{\sqrt{T}}N_T - \frac{1}{2}\frac{\langle N\rangle_T}{T}\right), \tag{15}$$

where $N_t = \int_0^t X_s^* U^* \Sigma^{-\frac{1}{2}} dW_s^2$ is a $\mathcal{F}_t^X$ – scalar martingale.
Proposition 1 implies that

$$\frac{1}{T}\langle N\rangle_T \xrightarrow[T\to+\infty]{P} tr\left(\mathcal{I}(A)U^*\Sigma^{-1}U\right)$$

hence $\frac{1}{\sqrt{T}}N_T \xrightarrow[T\to+\infty]{\mathcal{L}} \xi$, where ξ is a centered Gaussian variable with variance $E\xi^2 = tr\left(\mathcal{I}(A)U^*\Sigma^{-1}U\right)$. In fact, $\xi = vec^*(U^*)vec(\zeta)$ for some $\zeta \sim \mathbf{M}\mathcal{N}_{n\times 2n}\left(0, \mathcal{I}(A), \Sigma^{-1}\right)$.

(C2) for some $\chi > 0$ and $C > 0$ we have

$$E_\Theta Z_T^{\frac{1}{2}}(U) \le C\exp(-\chi\|U\|^2), \quad \forall\, U \text{ such that } \Theta + \frac{1}{\sqrt{T}}U \in \Xi.$$

Indeed, from (15), considering $\Theta_1^T = \Theta + \frac{1}{2}\frac{U}{\sqrt{T}}$ the actual value of Θ, for $A(\Theta)$ as defined in (1), we have

$$E_\Theta Z_T^{\frac{1}{2}}(U) = E_{\Theta_1^T}\exp\left(-\frac{1}{8T}\int_0^T X_t^* U^* \Sigma^{-1} U X_t\, dt\right).$$

Applying the same kind of computations as in Koncz (1987) (See Statement 1 and Remark 2) we obtain

$$E_\Theta Z_T^{\frac{1}{2}}(U) = \exp\left(-T\, tr(BD)\right)\cdot\det\left(Id_{2n\times 2n} - 2D\Gamma(T)\right)^{-\frac{1}{2}},$$

where the matrix D is the solution of the Riccati equation

$$DA\left(\Theta + \frac{U}{\sqrt{T}}\right) + A^*\left(\Theta + \frac{U}{\sqrt{T}}\right)D - 2DBD = -\frac{1}{4}\frac{U^*}{\sqrt{T}}\Sigma^{-1}\frac{U}{\sqrt{T}}.$$

Notice that D depends only on $V=\frac{U}{\sqrt{T}}$ and $D=0$ *if and only if* $U=0$. The matrix $D_\nu=\frac{D}{\|V\|^2}$, with $\nu=\|V\|$, satisfies the equation

$$D_\nu A\left(\Theta+\nu\bar{V}\right)+A^*\left(\Theta+\nu\bar{V}\right)D_\nu-2\nu D_\nu BD_\nu=-\frac{1}{4}\bar{V}^*\Sigma^{-1}\bar{V},$$

with $\bar{V}=\frac{V}{\|V\|}$. Now, let

$$\chi=\inf_{V\neq 0,\|V\|\leq diam(\Xi)}\frac{tr\,(BD)}{\|V\|^2}.$$

On one hand,

$$\chi\geq\inf_{\bar{V}:\,\|\bar{V}\|=1,\quad 0<\nu\leq diam(\Xi)}tr\,(BD_\nu)\geq\inf_{\bar{V}:\,\|\bar{V}\|=1,\quad 0\leq\nu\leq diam(\Xi)}tr\,(BD_\nu)>0,$$

since D_ν is a continuous and positive definite matrix function on a compact set and

$$tr\,(BD_0)=\frac{1}{4}tr\left(\bar{V}^*\mathcal{I}(A)\Sigma^{-1}\bar{V}\right)>0.$$

Thus, $tr\,(BD)\geq\chi\|V\|^2$ holds for this constant $\chi>0$, that is

$$\exp\left(-Ttr(BD)\right)\leq\exp(-\chi\|U\|^2). \tag{16}$$

On the other hand, for all $T>0$ we have

$$\inf_{V\neq 0,\|V\|\leq diam(\Xi)}\det\left(Id_{2n\times 2n}-2D\Gamma(T)\right)>0.$$

Let $\Gamma_\infty=\lim_{T\to+\infty}\Gamma(T)$ be the solution of the Lyapunov equation

$$\tilde{a}\Gamma_\infty+\Gamma_\infty\tilde{a}^*+B=0,$$

with $\tilde{a}=A-2BD$. The matrix $D^{-1}-2\Gamma_\infty$ is positive definite, since it is a solution of the equation

$$\tilde{a}(D^{-1}-2\Gamma_\infty)+(D^{-1}-2\Gamma_\infty)\tilde{a}^*+\frac{1}{4}D^{-1}V^*\Sigma^{-1}V=0.$$

Therefore, for all V such that $\|V\|\leq diam(\Xi)$, we have $\det\left(Id_{2n\times 2n}-2D\Gamma_\infty\right)>0$. As a consequence, for all T

$$\det\left(Id_{2n\times 2n}-2D\Gamma(T)\right)\geq C_1>0.$$

This bound and (16) verifies (C2).

(C3) for any compact $K\subset\Xi$ there exists $C>0,\ \beta>0$ such that for $m>n^2$ we have

$$\sup_G\|U_1-U_2\|^{-2m}E_\Theta\left|Z_T^{\frac{1}{2m}}(U_1)-Z_T^{\frac{1}{2m}}(U_2)\right|^{2m}\leq C(1+R^\beta),$$

where $G=\left\{U_i:\ \|U_i\|\leq R,\ \Theta\in K,\ \Theta+\frac{U_i}{\sqrt{T}}\in\Xi\right\}$.

We follow the ideas developed in Kutoyants (2004). Let $\Delta U=U_2-U_1$. In order to verify this condition, first notice that

$$\begin{aligned}\sup_G \|\Delta U\|^{-2m} E_\Theta \left| Z_T^{\frac{1}{2m}}(U_1) - Z_T^{\frac{1}{2m}}(U_2)\right|^{2m} &= \sup_G E_\Theta\, Z_T(U_1) \left| \frac{1-\left(\frac{Z_T(U_2)}{Z_T(U_1)}\right)^{\frac{1}{2m}}}{\|\Delta U\|}\right|^{2m} \\ &= \sup_G E_{\Theta_1^T} \left|\frac{1-\Upsilon_T(\nu)}{\nu}\right|^{2m},\end{aligned}$$

where $\Theta_1^T = \Theta + \frac{U_1}{\sqrt{T}}$, $\nu = \|\Delta U\|$ and $\Upsilon_T(\nu)$ is given by

$$\Upsilon_T(\nu) = \exp\left(-\frac{\nu}{2m\sqrt{T}}\int_0^T X_t^* \bar{V}^* \Sigma^{-\frac{1}{2}} d\bar{W}_t - \frac{\nu^2}{2mT}\int_0^T X_t^* \bar{V}^*\Sigma^{-1}\bar{V} X_t\, dt\right),$$

with $\bar{V} = \frac{\Delta U}{\|\Delta U\|}$ and a Wiener process $\{\bar{W}_t\}$ with respect to the measure with density $Z_T(U_1)$.
In this context, we can write

$$\begin{aligned}\frac{\Upsilon_T(\nu)-1}{\nu} = \int_0^1 \Upsilon_T(\alpha\nu)\Bigg[&-\frac{1}{2m\sqrt{T}}\int_0^T X_t^*\bar{V}^*\Sigma^{-\frac{1}{2}} d\bar{W}_t \\ &-\frac{\alpha\nu}{mT}\int_0^T X_t^*\bar{V}^*\Sigma^{-1}\bar{V}X_t\, dt\Bigg] d\alpha.\end{aligned}$$

Hence, for some constants $C_m,\ \tilde{C}_m > 0$ depending on m we have the upper bounds

$$\begin{aligned}E_{\Theta_1^T}\left|\frac{1-\Upsilon_T(\nu)}{\nu}\right|^{2m} &\leq C_m E_{\Theta_1^T}\int_0^1 \Upsilon_T^{2m}(\alpha\nu)\Bigg[\left(\frac{1}{\sqrt{T}}\int_0^T X_t^*\bar{V}^*\Sigma^{-\frac{1}{2}}d\bar{W}_t\right)^{2m} \\ &\quad + \left(\frac{1}{T}\int_0^T X_t^*\bar{V}^*\Sigma^{-1}\bar{V}X_t\, dt\right)^{2m}\Bigg] d\alpha \\ &\leq \tilde{C}_m \int_0^1 E_\Theta \left(\frac{1}{T}\int_0^T \tilde{X}_t^*\bar{V}^*\Sigma^{-1}\bar{V}\tilde{X}_t\, dt\right)^{2m} d\alpha,\end{aligned}$$

where $\tilde{X}_t$ satisfies the equation

$$d\tilde{X}_t = A\left(\Theta + \frac{U_1}{\sqrt{T}} + \alpha\frac{\Delta U}{\sqrt{T}}\right)\tilde{X}_t dt + B^{\frac{1}{2}} dW_t.$$

So, the remaining problem is the uniform estimation of the polynomial moments of $\frac{1}{T}\int_0^T \tilde{X}_t^* Q\tilde{X}_t dt$, with $Q = \bar{V}^*\Sigma^{-1}\bar{V}$. Again, let us use the Laplace transform, since for any $k \geq 1$

$$E_\Theta\left(\frac{1}{T}\int_0^T \tilde{X}_t^* Q\tilde{X}_t dt\right)^k = \frac{\partial^k}{\partial\mu^k} L_T\left(\mu, Q, \tilde{X}\right)|_{\mu=0}.$$

It is sufficient to estimate $\frac{\partial^k}{\partial \mu^k} D_\mu|_{\mu=0}$, where the matrix D_μ satisfies

$$D_\mu A\left(\Theta_1^T + \alpha\frac{\Delta U}{\sqrt{T}}\right) + A^*\left(\Theta_1^T + \alpha\frac{\Delta U}{\sqrt{T}}\right) D_\mu - 2D_\mu B D_\mu = -\frac{\mu}{T} Q.$$

Notice that $D_0 = 0$ and that D_μ is an analytical function in a neighborhood of 0 (see Sect. 3.1). The derivative $H_k = \frac{\partial^k}{\partial \mu^k} D_\mu|_{\mu=0}$ satisfies the Lyapunov equation

$$H_k A\left(\Theta + \alpha\frac{\Delta U}{\sqrt{T}}\right) + A^*\left(\Theta + \alpha\frac{\Delta U}{\sqrt{T}}\right) H_k = P(H_i),$$

where P is a linear combination of H_i and the explicit formula for the solution (see Sect. 3.1) gives

$$\|H_k\| \le C\varepsilon^{-1}(1 + \|U_i\|^{2kn^2}) \le C\varepsilon^{-1}(1 + R^{2kn^2}),$$

where $\varepsilon = (\min_{\Theta\in K,\, j\le 2n} |\lambda_j(A)|)^{2n^2} > 0$ and $\lambda_j(A)$ are the eigenvalues of $A(\Theta)$.

□

5 Computation of the Fisher information matrix

From the Lyapunov equation (9) we can obtain properties and, in some cases, the explicit form of the Fisher information matrix and its inverse, the asymptotic covariance matrix of the MLE $\hat{\theta} = \widehat{vec(\Theta)}$. Arató et al. (2002) and, more recently, Brockwell et al. (2007) considered the case $n = 1$. The generalisation in Brockwell et al. (2007) to higher dimensional models does not include the model investigated in this paper, given by (1)–(2). In the next paragraphs, we study the two cases $n = 1$ and $n \ge 2$ separately.

5.1 The case $n = 1$

In this case, we can explicitly compute the asymptotic covariance matrix of the MLE $\hat{\theta} = \widehat{vec(\Theta)}$.

The matrices A and $B^{\frac{1}{2}}$ are

$$A = \begin{pmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{c}{m} \end{pmatrix} \qquad \text{and} \qquad B^{\frac{1}{2}} = \begin{pmatrix} 0 & 0 \\ 0 & \sigma \end{pmatrix},$$

with m, k and $c > 0$.

The estimator in (6) can be written as

$$\hat{\theta}_T = \left(\int_0^T X_t X_t^*\, dt\right)^{-1} \cdot \int_0^T X_t\, dX_t^{(2)}.$$

Solving (9), we can compute the Fisher information matrix, $\mathcal{I}(A)$, and by Theorem 1 we can conclude that

$$\sqrt{T}(\hat{\theta}_T - \theta) \underset{\mathcal{L}}{\longrightarrow} \mathcal{N}(0, Q_\infty),$$

with $Q_\infty = \begin{pmatrix} \frac{2kc}{m^2} & 0 \\ 0 & \frac{2c}{m} \end{pmatrix}$.

Arató et al. (2002) also calculated the asymptotic covariance matrix. This asymptotic convergence is a particular case of the result obtained by Lin and Lototsky (2014) in a more general context, see [Theorem 3.2].

5.2 The case $n \geq 2$

In this case, we can also explicitly compute the asymptotic covariance matrix of the MLE $\hat{\theta} = \widehat{vec(\Theta)}$, provided that the product of the matrices θ_1, θ_2 and Σ in model (1)–(2) is commutative. The result is obtained by splitting the Lyapunov equation (9) into blocks of dimension $n \times n$

$$\begin{pmatrix} 0 & Id_{n\times n} \\ -\theta_1 & -\theta_2 \end{pmatrix}\begin{pmatrix} \mathcal{I}_1 & \mathcal{I}_2 \\ \mathcal{I}_2^* & \mathcal{I}_4 \end{pmatrix} + \begin{pmatrix} \mathcal{I}_1 & \mathcal{I}_2 \\ \mathcal{I}_2^* & \mathcal{I}_4 \end{pmatrix}\begin{pmatrix} 0 & -\theta_1^* \\ Id_{n\times n} & -\theta_2^* \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & \Sigma \end{pmatrix} = 0,$$

where $\mathcal{I}_1, \mathcal{I}_2$ and $\mathcal{I}_4$ are square $n \times n$ matrices representing the blocks in the upper triangular symmetric matrix $\mathcal{I}(A)$. Hence, we have the following system

$$\begin{cases} \mathcal{I}_2^* + \mathcal{I}_2 = 0 \\ \mathcal{I}_4 - \mathcal{I}_1\theta_1^* - \mathcal{I}_2\theta_2^* = 0 \\ -\theta_1\mathcal{I}_1 - \theta_2\mathcal{I}_2^* + \mathcal{I}_4 = 0 \\ -\theta_1\mathcal{I}_2 - \theta_2\mathcal{I}_4 - \mathcal{I}_2^*\theta_1^* - \mathcal{I}_4\theta_2^* = -\Sigma \end{cases}.$$

If the matrices θ_1, θ_2 and Σ commute we can write the solution of the Lyapunov equation (9) as the following block-diagonal matrix

$$\mathcal{I}(A) = \begin{pmatrix} \frac{1}{2}\theta_1^{-1}\theta_2^{-1}\Sigma & 0 \\ 0 & \frac{1}{2}\theta_2^{-1}\Sigma \end{pmatrix}.$$

Since we assumed that all eigenvalues of A have negative real parts, this is the unique solution of the Lyapunov equation. Thus, the Fisher information matrix is

$$\Sigma^{-1} \otimes \mathcal{I}(A).$$

Notice that if Σ is diagonal then the Fisher information matrix (and its inverse) is block-diagonal.

6 Simulation results

We present examples of 2-dimension and 4-dimension O–U processes. For all the examples we generated 200 sample paths from model (1)–(4) on the time interval [0, 2000] seconds (s). We computed statistics for the estimates on different time intervals.

The following discretisations of the integrals, involved in expression (6),

– $\int_{i\Delta t}^{j\Delta t} X_{r,t}^2 dt \cong \sum_{l=i}^{j-1} \Delta t X_{r,t_l}^2, \qquad r = 1, \ldots, 2n$

– $\int_{i\Delta t}^{j\Delta t} X_{r,t} dX_{s,t} \cong \sum_{l=i}^{j-1} X_{r,t_l}\left(X_{s,t_l} - X_{s,t_{l-1}}\right), \quad r = 1, \cdots, 2n,\ s = n+1, \ldots, 2n,$
$r \neq s$

$$- \int_{i\Delta t}^{j\Delta t} X_{s,t} dX_{s,t} \cong \frac{1}{2} \sum_{l=i}^{j-1} \left(X_{s,t_l}^2 - X_{s,t_{l-1}}^2 - \sigma \Delta t \right), \qquad s = n+1, \ldots, 2n$$

$$- \int_{i\Delta t}^{j\Delta t} X_{r,t} X_{s,t} dt \cong \sum_{l=i}^{j-1} \Delta t X_{r,t_l} X_{s,t_l}, \qquad r, s = 1, \ldots, 2n,$$

were used in our computations. The simulations are based on this discretisation of the solution of the linear SDE.

6.1 The case $n = 1$

We study two examples. In Example 1, we consider a scenario encountered in structural engineering (cf. Clough and Penzien 1982). Example 2 is the one considered in Samson and Thieullen (2012), but with continuous time observations.

Figures 1, 2 and 3 show the results for the estimator in (6) on the simulated paths of Example 1 for different values of T. Tables 1 and 4 summarise the results for each example. The Gaussian curves exhibited in Figs. 2 and 3 (dotted lines), as well as the p-values for the Kolmogorov-Smirnov test in Table 2 (and also in Table 5), where determined using the marginal distribution parameters derived from Theorem 1. The solid lines represent the curves

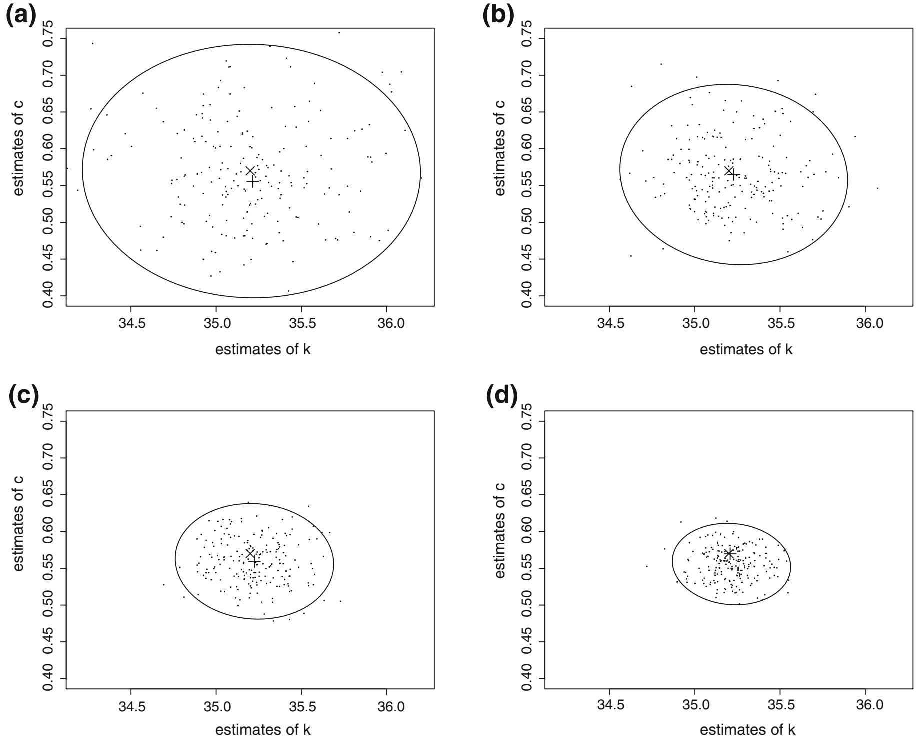

Fig. 1 95 % probability ellipsis for $(\hat{k}, \hat{c})$ and for different values of T (Example 1): **a** $T = 200\,s$, **b** $T = 500\,s$, **c** $T = 1000\,s$, **d** $T = 2000\,s$. The mean point is indicated by a "+" sign and a "×" sign indicates the true value of the parameters. The ellipses are built upon the asymptotic distribution set by Theorem 1

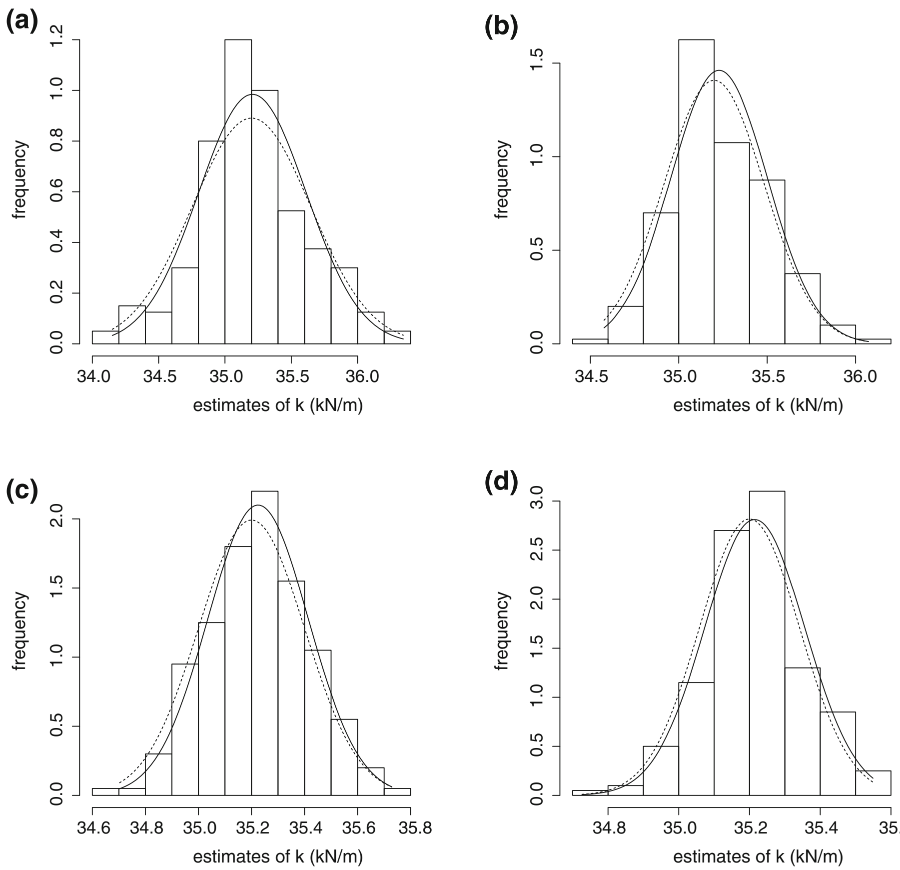

Fig. 2 Histograms for $\hat{k}$ for different values of T (Example 1): **a** $T = 200\,s$, **b** $T = 500\,s$, **c** $T = 1000\,s$, **d** $T = 2000\,s$. The *dotted line* represents the Gaussian curve derived from Theorem 1 while the *solid line* represents the Gaussian fit

obtained by fitting Gaussian distributions to the estimates that were obtained for k and c. Results on the correlation between estimates are given in Table 3 (and also in Table 6).

Example 1 $k = 35.2\ kN/m, c = 0.57\ kNs/m, m = 0.933\ ton, \sigma = 1.$

Example 2 $k = 4\ kN/m, c = 0.5\ kNs/m, m = 1\ ton, \sigma = 1.$

We can see that the estimates of parameters k and c exhibit a considerable bias, while time T is not large enough, but this bias vanishes with time. The parameters can be quite accurately estimated, with the asymptotic marginal distributions showing no evidence of non-Gaussianity. A larger bias was observed in Example 2 of Samson and Thieullen (2012) but considering discrete time observations. The correlation of the estimators of k and c was studied. The null hypotheses of non-correlation was not rejected for all cases (p-values are shown in Tables 3 and 6).

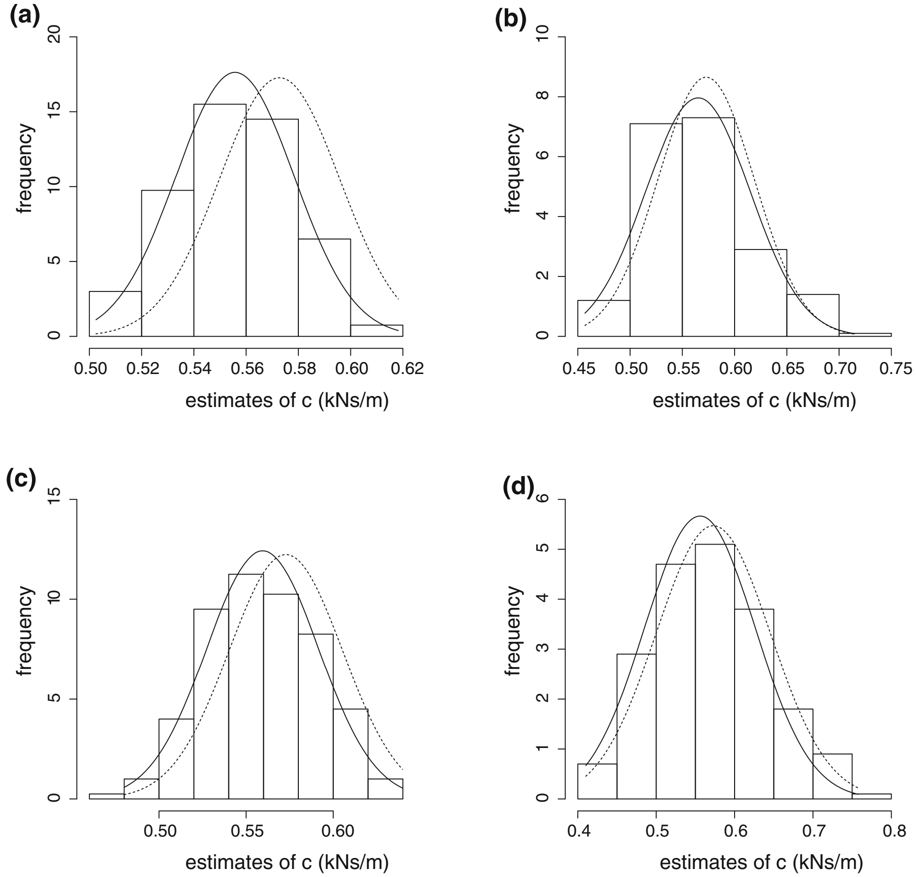

Fig. 3 Histograms for $\hat{c}$ for different values of T (Example 1): **a** $T = 200\,s$, **b** $T = 500\,s$, **c** $T = 1000\,s$, **d** $T = 2000\,s$. The *dotted line* represents the Gaussian curve derived from Theorem 1 while the *solid line* represents the Gaussian fit

6.2 The case $n = 2$

We analyse two examples, Examples 3 and 4, from Magalhães (2010) [Chap. 2, p. 63]. They differ from each other in the matrices M and C. We considered the units: kN/m for K, ton for M and kNs/m for C. In both examples, we considered $K = \begin{pmatrix} 100 & 0 \\ 0 & 100 \end{pmatrix}$ and $\Sigma = Id_{2\times 2}$. The results obtained in the simulations are summarised in Table 7 (Example 3) and Table 8 (Example 4).

Example 3 $M = \begin{pmatrix} 1.25 & 0.25 \\ 0.25 & 1.25 \end{pmatrix}$ and $C = \begin{pmatrix} 0.3275 & -0.0725 \\ -0.0725 & 0.3275 \end{pmatrix}$.

Table 1 Mean value and standard deviation computed over the estimates of parameters k and c for different values of T (Example 1)

T	200s	500s	1000s	2000s
$mean(k)$ $(st.dev.(k))$	35.21 (0.405)	35.22 (0.273)	35.22 (0.189)	35.21 (0.142)
$mean(c)$ $(st.dev.(c))$	0.556 (0.070)	0.565 (0.050)	0.559 (0.032)	0.569 (0.023)

Table 2 Results of the Kolmogorov-Smirnov normality test applied to the estimates of parameters k and c for different values of T (Example 1)

T	200s	500s	1000s	2000s
p-value for $\hat{k}$	0.5336	0.4966	0.6691	0.9900
p-value for $\hat{c}$	0.7447	0.4852	0.8844	0.9747

Table 3 Results of the non-correlation test for $\hat{k}$ and $\hat{c}$ (Example 1)

T	200s	500s	1000s	2000s
p-value	0.3101	0.3878	0.4895	0.8627

Table 4 Mean value and standard deviation computed over the estimates of parameters k and c for different values of T (Example 2)

T	200s	500s	1000s	2000s
$mean(k)$ $(st.dev.(k))$	3.997 (0.148)	4.001 (0.087)	3.999 (0.061)	3.999 (0.044)
$mean(c)$ $(st.dev.(c))$	0.509 (0.072)	0.497 (0.043)	0.495 (0.032)	0.495 (0.023)

Table 5 Results of the Kolmogorov-Smirnov normality test applied to the estimates of parameters k and c for different values of T (Example 2)

T	200s	500s	1000s	2000s
p-value for $\hat{k}$	0.8500	0.8732	0.9002	0.9398
p-value for $\hat{c}$	0.3724	0.6371	0.7979	0.8730

Table 6 Results of the non-correlation test for $\hat{k}$ and $\hat{c}$ (Example 2)

T	200s	500s	1000s	2000s
p-value	0.1073	0.4417	0.6748	0.7885

The solution of the Lyapunov equation (9) is (see Sect. 5)

$$\mathcal{I}(A) = \begin{pmatrix} \begin{matrix} 0.0283 & 0.0158 \\ 0.0158 & 0.0283 \end{matrix} & O \\ O & \begin{matrix} 2.0956 & 0.8456 \\ 0.8456 & 2.0956 \end{matrix} \end{pmatrix}$$

Table 7 Mean value and standard deviation computed over the estimates of matrix parameters K and C for different values of T (Example 3)

T	200s	500s	1000s	2000s
$mean(k_{11})\ (st.dev.(k_{11}))$	100.61 (2.49)	100.11 (1.24)	100.04 (0.86)	100.05 (0.57)
$mean(k_{12})\ (st.dev.(k_{12}))$	−0.32 (2.40)	−0.11 (1.28)	−0.05 (0.83)	−0.05 (0.55)
$mean(k_{21})\ (st.dev.(k_{21}))$	−0.37 (2.46)	−0.14 (1.28)	−0.11 (0.86)	−0.06 (0.55)
$mean(k_{22})\ (st.dev.(k_{22}))$	100.84 (2.91)	100.19 (1.29)	100.16 (0.86)	100.09 (0.58)
$mean(c_{11})\ (st.dev.(c_{11}))$	0.4861 (0.499)	0.3899 (0.323)	0.3223 (0.152)	0.2958 (0.103)
$mean(c_{12})\ (st.dev.(c_{12}))$	−0.0900 (0.976)	−0.1651 (0.518)	−0.1001 (0.291)	−0.0809 (0.206)
$mean(c_{21})\ (st.dev.(c_{21}))$	0.0088 (1.010)	0.0211 (0.486)	−0.0323 (0.294)	−0.0600 (0.213)
$mean(c_{22})\ (st.dev.(c_{22}))$	0.4207 (0.539)	0.2844 (0.291)	0.2973 (0.153)	0.2912 (0.106)

Table 8 Mean value and standard deviation computed over the estimates of matrix parameters K and C for different values of T (Example 4)

T	200s	500s	1000s	2000s
$mean(k_{11})\ (st.dev.(k_{11}))$	121.51 (37.58)	105.80 (8.80)	102.91 (4.67)	101.70 (2.49)
$mean(k_{12})\ (st.dev.(k_{12}))$	−0.93 (3.27)	−0.17 (1.28)	−0.13 (0.82)	−0.04 (0.54)
$mean(k_{21})\ (st.dev.(k_{21}))$	−0.17 (3.45)	−0.19 (1.18)	−0.10 (0.77)	−0.05 (0.48)
$mean(k_{22})\ (st.dev.(k_{22}))$	121.16 (37.88)	106.01 (8.63)	102.98 (4.43)	101.77 (2.41)
$mean(c_{11})\ (st.dev.(c_{11}))$	0.664 (3.36)	0.259 (1.37)	0.275 (1.02)	0.294 (0.75)
$mean(c_{12})\ (st.dev.(c_{12}))$	0.453 (8.81)	0.061 (3.04)	−0.032 (2.06)	−0.147 (1.53)
$mean(c_{21})\ (st.dev.(c_{21}))$	−0.785 (7.35)	−0.229 (3.09)	−0.144 (2.03)	−0.045 (1.48)
$mean(c_{22})\ (st.dev.(c_{22}))$	0.422 (3.35)	0.447 (1.37)	0.334 (1.02)	0.260 (0.76)

and the Fisher information matrix of θ is

$$\Sigma^{-1} \otimes \mathcal{I}(A) = \begin{pmatrix} \begin{matrix} 0.0283 & 0.0158 & 0 & 0 \\ 0.0158 & 0.0283 & 0 & 0 \\ 0 & 0 & 2.0956 & 0.8456 \\ 0 & 0 & 0.8456 & 2.0956 \end{matrix} & O \\ O & \begin{matrix} 0.0283 & 0.0158 & 0 & 0 \\ 0.0158 & 0.0283 & 0 & 0 \\ 0 & 0 & 2.0956 & 0.8456 \\ 0 & 0 & 0.8456 & 2.0956 \end{matrix} \end{pmatrix}.$$

Example 4 $M = \begin{pmatrix} 1.0125 & 0.0125 \\ 0.0125 & 1.0125 \end{pmatrix}$ and $C = \begin{pmatrix} 0.303 & -0.098 \\ -0.098 & 0.303 \end{pmatrix}$.

The solution of the Lyapunov equation (9) is

$$\mathcal{I}(A) = \begin{pmatrix} \begin{matrix} 0.0190 & 0.0066 \\ 0.0066 & 0.0190 \end{matrix} & O \\ O & \begin{matrix} 1.8734 & 0.6266 \\ 0.6266 & 1.8734 \end{matrix} \end{pmatrix}$$

and the Fisher information matrix of θ is

$$\Sigma^{-1} \otimes \mathcal{I}(A) = \begin{pmatrix} \begin{matrix} 0.0190 & 0.0066 & 0 & 0 \\ 0.0066 & 0.0190 & 0 & 0 \\ 0 & 0 & 1.8734 & 0.6266 \\ 0 & 0 & 0.6266 & 1.8734 \end{matrix} & O \\ O & \begin{matrix} 0.0190 & 0.0066 & 0 & 0 \\ 0.0066 & 0.0190 & 0 & 0 \\ 0 & 0 & 1.8734 & 0.6266 \\ 0 & 0 & 0.6266 & 1.8734 \end{matrix} \end{pmatrix}.$$

Notice that the estimates of the matrices K and C are reasonably close to the true values, exhibiting deviations that are compatible with the asymptotic distribution obtained in Theorem 1 and further addressed in Sect. 5.2. According to Magalhães (2010), in terms of estimation error, matrices in Example 3 should be easier to estimate than those of Example 4, because the latter is characterised by closely spaced vibration modes, while the vibration modes that characterise Example 3 are well separated modes. Our computational experiments are in agreement with the results in Magalhães (2010).

7 Conclusions

Although the $2n$-dimension model investigated in this paper fails to verify the restrictions on the invertibility of the diffusion matrix, it is still possible to obtain asymptotic properties of the MLE by studying the Laplace transform of a martingale associated with the observation process. The convergence of all the moments of the estimation error $\sqrt{T}(\hat{\theta}_T - \theta)$ is proved. The asymptotic covariance matrix of the estimator is given in terms of the solution of a Lyapunov equation. If the lower blocks in the considered drift matrix (that is, matrices $M^{-1}K$ and $M^{-1}C$, where M, K and C represent, respectively, the mass, the stiffness and the damping matrices in the SDE modelling a one degree of freedom mechanical system) and Σ happen to verify the commutativity property of multiplication of matrices then an explicit expression of the covariance matrix of the MLE can be found. Furthermore, if the matrix Σ is diagonal, the Fisher information matrix is a block diagonal matrix. Simulation studies are presented considering 2-dimension and 4-dimension O–U processes examples. The simulation results illustrate the asymptotic behaviour of the MLE and are in agreement with the theoretical results.

We believe that the proposed method will also work for the partially observed models and that it can be extended to more general classes of noises.

Acknowledgments Research partially funded by the European Structural Fund – programme FEDER – and by the Portuguese Government through the FCT – Fundação para a Ciência e Tecnologia under contracts UID/MAT/00144/2013 and PROTEC/IPL-ISEL. We are very grateful to the referees for their constructive criticism when reviewing this paper.

References

Arató M (1982) Linear stochastic systems with constant coefficients. A statistical approach. Lecture notes in control and information sciences, vol 45. Springer, Berlin

Arató M, Pap G, Varga K (2002) Estimation of the mean of multivariate AR processes. Comput Math Appl 43:707–719

Basak G, Lee P (2008) Asymptotic properties of an estimator of the drift coefficients of multidimensional Ornstein-Uhlenbeck processes that are not necessarily stable. Electron J Stat 2:1309–1344
Brockwell P, Davis R, Yang Y (2007) Continuous-time Gaussian autoregression. Stat Sinica 17:63–80
Clough RW, Penzien J (1982) Dynamics of structures. McGraw-Hill, Singapore
Ibragimov I, Khasminskii R (1981) Statistics of random processes. Springer, New York
Kleptsyna M, Le Breton A, Viot M (2001) About Laplace transforms of quadratic functionals of multidimensional Gauss-Markov processes and matrix-valued Riccati differential equations. In: Menaldi J, Rofman E, Sulem A (eds) Optimal Control and PDE: Innovations and Applications, Conference in honor of Prof. A. Bensoussan on the occasion of his 60th birthday, IOS Press, Amsterdam, 248-257
Kleptsyna M, Le Breton A, Viot M (2008) On the linear-exponential filtering problem for general Gaussian processes. Siam J Control Optim 47(6):2886–2911
Koncz K (1987) On the parameter estimation of diffusional type processes with constant coefficients (elementary Gaussian processes). J Anal Math 13(1):75–91
Kucera V (1973) A review of the matrix Riccati equation. Kybernetika 9(1):42–61
Kutoyants YA (2004) Statistical inference for Ergodic diffusion processes. Springer series in statistics. Springer, London
Le Breton A (1977) Parameter estimation in a vector linear stochastic differential equation. In: Transactions of the 7th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes. Academia, Publ H Cz Acad Sc, Prague A, pp. 353–366
Lin N, Lototsky S (2011) Undamped harmonic oscillator driven by additive Gaussian white noise: a statistical analysis. Commun Stoch Anal 5(1):233–250
Lin N, Lototsky S (2014) Second-order continuous-time non-stationary Gaussian autoregression. Stat Inference Stoch Process 17:19–49
Lutes L, Sarkani S (1997) Stochastic analysis of structural and mechanical vibrations. Prentice-Hall, Upper Saddle River
Magalhães F (2010) Operational modal analysis for testing and monitoring of bridges and special structures. PhD thesis, Faculty of Engineering of the University of Porto
Peeters B, Roeck G (2001) Stochastic system identification for operational modal analysis: a review. J Dyn Syst Trans ASME 123(4):659–667
Perninge M, Knazkins V, Amelin M, Soder L (2011) Modeling the electric power consumption in a multi-area system. Eur Trans Electr Power 21(1):413–423
Rao B (1999) Statistical inference for diffusion type process. Kendall's Library of Statistics 8, Edward Arnold, London. Oxford University Press, New York
Samson A, Thieullen M (2012) A contrast estimator for completely or partially observed hypoelliptic diffusion. Stoch Proc Appl 122(7):2521–2552