Communications on Stochastic Analysis
Vol. 5, No. 1 (2011) 233-250

Serials Publications
www.serialspublications.com

# UNDAMPED HARMONIC OSCILLATOR DRIVEN BY ADDITIVE GAUSSIAN WHITE NOISE: A STATISTICAL ANALYSIS

N. LIN* AND S. V. LOTOTSKY*

Abstract. While consistency of the maximum likelihood estimator of the drift matrix in a multi-dimensional continuous time Ornstein-Uhlenbeck process holds under rather general conditions, little is known about the rate of convergence and the limiting distribution of the estimator when the underlying process is not ergodic. The objective of the paper is to investigate these questions for an important example of a non-ergodic two-dimensional Ornstein-Uhlenbeck process: an undamped harmonic oscillator.

## 1. Introduction

The objective of this paper is to study parameter estimation in the equation

$$\ddot{X}(t) + a\dot{X}(t) + \theta X(t) = \dot{W}(t),\ t > 0, X(0) = \dot{X}(0) = 0, \tag{1.1}$$

or the equivalent system

$$dX = Y dt,\ dY = -(\theta X + aY)dt + dW(t) \tag{1.2}$$

with a standard Brownian motion $W = W(t)$. The solution of (1.1) is a Gaussian process

$$X(t) = \int_0^t \phi(t-s)dW(s), \tag{1.3}$$

where the function $\phi(t)$ is the unique solution of the initial value problem

$$\ddot{\phi} + a\dot{\phi}(t) + \theta\phi(t) = 0,\ \phi(0) = 0,\ \dot{\phi}(0) = 1. \tag{1.4}$$

We construct the maximum likelihood estimators $\widehat{a}_T$, $\widehat{\theta}_T$ of $a$ and $\theta$ from the continuous time observations of $(X(t), \dot{X}(t),\ t \in [0, T])$, and study their limiting distribution, as $T \to \infty$, when $\theta = c^2$ and $a = 0$. Figure 1 presents sample trajectories of the process $X$ for $c = \pi$ and three different values of $a$: $a = 0$ (pure oscillations) $a = 0.05$ (damped oscillations), and $a = -0.05$ (amplified oscillations); the realization of the Brownian motion is the same in all three simulation. The procedure used to produce Figure 1 is described at the end of Section 5. The rationale behind the choice of the parameters for simulations was as follows: $c = \pi$

Received 2010-3-3.
2000 *Mathematics Subject Classification.* Primary 62F12; Secondary 62F03, 62M07, 62M09.
*Key words and phrases.* Damping, exact discretization, hypothesis testing, maximum likelihood estimation, non-normal limiting distribution, second-order stochastic equation.
* The work of both authors was partially supported by the NSF Grant DMS-0803378.

ensures an integer-valued period of the underlying oscillations; small absolute values of $a$ ensure that the free motion does not die out or explode too quickly.

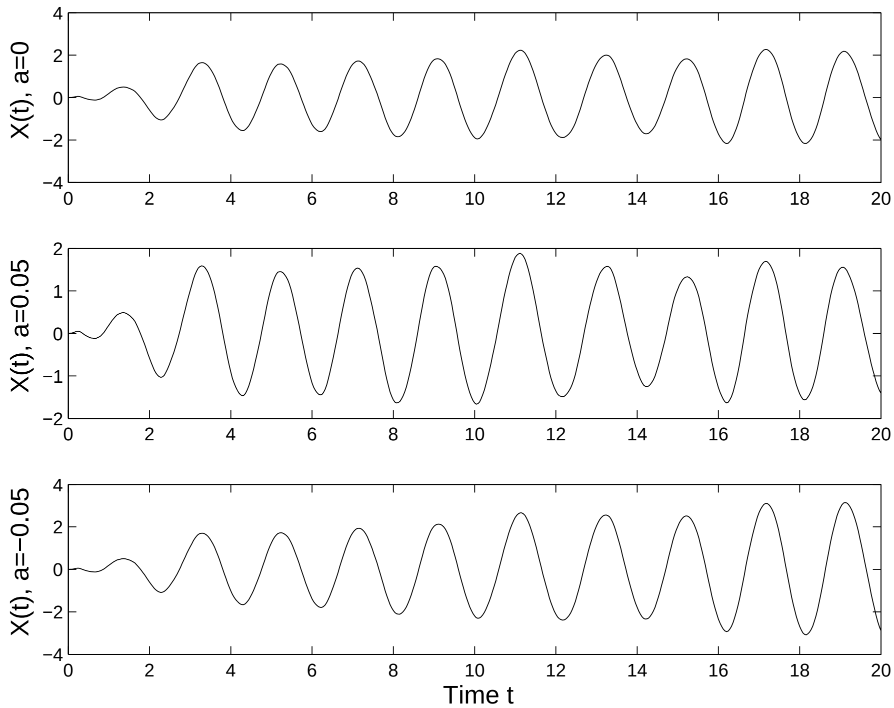


FIGURE 1. Sample trajectories of the process $X$

Statistical inference for multi-dimensional linear and bi-linear diffusion processes in continuous time has become an active area of research; the paper by Basak and Lee [2] provides a comprehensive and up-to date survey of the literature on the subject. Even though every differential equation of order two and higher can be reduced to a system, linear stochastic equations deserve a separate analysis:

(1) Similar to deterministic linear equations, the solution is easier to study than in the general matrix case;
(2) The unknown coefficients form a vector rather than matrix, which allows analysis of estimators of individual coefficients in a much more convenient way;
(3) The conditions in the matrix setting require certain non-degeneracy of the diffusion, which may or may not hold for the higher-order linear stochastic equation.

While strong consistency of the estimators has been recently established by Basak and Lee [2] for a large class of systems, which includes (1.2), the limiting distribution, together with the proper normalization, remains an open problem except when the process is ergodic and the estimator is asymptotically normal with rate $\sqrt{T}$.

Equation (1.1) defines an ergodic process if and only if $a > 0$ and $\theta > 0$, that is, when the free motion of the system is asymptotically stable. The undamped harmonic oscillator, with $a = 0$, $\theta > 0$, becomes a natural example of a non-ergodic process. If the asymptotic distribution of $\widehat{a}_T$ is known, then testing the hypothesis $H_0 : a = 0$ against the alternative $H_1 : a > 0$ can be accomplished, at least when the observation interval is large.

Here is the main result of the paper. Let $w_1$ and $w_2$ be two independent standard Brownian motions. Define random variables $\Psi$ and $\Upsilon$ by

$$\Psi = \frac{2 - w_1^2(1) - w_2^2(1)}{\int_0^1 w_1^2(t)dt + \int_0^1 w_2^2(t)dt}, \quad \Upsilon = \frac{\int_0^1 w_1(t)dw_2(t) - \int_0^1 w_2(t)dw_1(t)}{\int_0^1 w_1^2(t)dt + \int_0^1 w_2^2(t)dt}. \tag{1.5}$$

We also define the following random variables:

$$\mathbf{X}_T = \int_0^T X^2(t)dt, \ \mathbf{Y}_T = \int_0^T \dot{X}^2(t)dt, \tag{1.6}$$

and write $\overset{d}{=}$ to denote equality in distribution.

**Theorem 1.1.** *The maximum likelihood estimators $\widehat{a}_T$ and $\widehat{\theta}_T$ of $a$, $\theta$ in* (1.1) *are explicitly computable given the observations of*

$$X(t), \dot{X}(t), \ 0 \le t \le T:$$

$$\begin{aligned} \widehat{a}_T &= \frac{2X^2(T)\big(X(T)\dot{X}(T) - \mathbf{Y}_T\big) - 2\mathbf{X}_T\big(\dot{X}^2(T) - T\big)}{4\mathbf{X}_T\,\mathbf{Y}_T - X^4(T)}, \\ \widehat{\theta}_T &= \frac{X^2(T)\big(\dot{X}^2(T) - T\big) - 4\mathbf{Y}_T\big(X(T)\dot{X}(T) - \mathbf{Y}_T\big)}{4\mathbf{X}_T\,\mathbf{Y}_T - X^4(T)}, \end{aligned} \tag{1.7}$$

*and are strongly consistent in the limit $T \to \infty$. For the undamped harmonic oscillator ($\theta = c^2 > 0$, $a = 0$),*

$$\lim_{T\to\infty} T\,\widehat{a}_T \overset{d}{=} \Psi, \quad \lim_{T\to\infty} T(\widehat{\theta}_T - c^2) \overset{d}{=} 2c\Upsilon.$$

*If $T$ is sufficiently large, then the null hypothesis $H_0 : a = 0$ (no damping) can be rejected in favor of the alternative $H_1 : a > 0$ (damping is present) at the level of significance $\alpha$ if $T\widehat{a}_T > \gamma_\alpha$, where $\mathbb{P}(\Psi > \gamma_\alpha) = \alpha$.*

The advantage of the test statistic $T\widehat{a}_T$ is that no knowledge of the frequency $c$ is necessary. The disadvantage is that the resulting test is asymptotic and is only guaranteed to work in the limit $T \to \infty$.

To prove Theorem 1.1, we first consider the models with only one unknown parameter. In Section 2 we study estimation of $\theta$ when $a = 0$, and in Section 3, estimation of $a$ when $\theta = c^2$ is known. The proof of the main result is in Section 4. Estimators in Sections 2 and 3, being different from (1.7), can also be of independent interest. Section 5 presents a reduction of the continuous-time equation (1.1) to a second-order auto-regression and discusses the corresponding estimation problems in discrete-time setting. While analytically challenging, the question of discrete-time inference must be addressed to some extend as soon as one tries to implement continuous-time estimators numerically.

Through the rest of the paper, we fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with standard Brownian motion $W = W(t)$; $\mathbb{E}$ denotes the expectation with respect to $\mathbb{P}$. We use notations $\dot{g}(t)$, $\ddot{g}(t)$ for the first and second time derivatives of the function $g$ and write $\overset{d}{=}$ to indicate equality in distribution.

## 2. Estimation of Frequency With no Damping

Consider the second-order stochastic equation

$$\ddot{X}(t) + \theta\, X(t) = \dot{W}(t),\ \ t > 0, X(0) = \dot{X}(0) = 0. \tag{2.1}$$

We interpret (2.1) as a system of two equations

$$dX = Y dt;\ dY = -\theta\, X dt + dW(t);$$

that is, $Y(t) = \dot{X}(t)$. The objective is to estimate $\theta$ from the observations of the solution $(X(t), \dot{X}(t))$ for $0 \le t \le T$.

The process $Y = Y(t)$ is a diffusion-type process in the sense of Liptser and Shiryaev; see [7, Definition 4.2.7]. Therefore, by Theorem 7.6 in [7], the measure $P_T^Y$ generated by $(Y(t),\ 0 \le t \le T)$ in the space of continuous functions is absolutely continuous with respect to the corresponding measure $P_T^W$ generated by the Brownian motion $(W(t), 0 \le t \le T)$, and the likelihood ratio is

$$\frac{dP_T^Y}{dP_T^W}(Y) = \exp\left( -\int_0^T \theta X(t) dY(t) - \frac{1}{2}\int_0^T (\theta^2 X(t))^2 dt \right).$$

This results in the maximum likelihood estimator $\widehat{\theta}_T$ of $\theta$:

$$\widehat{\theta}_T = -\frac{\int_0^T X(t) dY(t)}{\int_0^T X^2(t) dt} = -\frac{\int_0^T X(t) d\dot{X}(t)}{\int_0^T X^2(t) dt}, \tag{2.2}$$

which is different from the one considered by Basak and Lee [2].

First, we establish the limiting distribution of the estimator. Here and below, expression

$$\lim_{T\to\infty} \xi(T) \overset{d}{=} \zeta$$

means that the random process $\xi = \xi(t),\ t \ge t_0$, converges in distribution to the random variable $\zeta$.

**Theorem 2.1.** *If $\theta = c^2 > 0$, then*

$$\lim_{T\to\infty} T\big(\widehat{\theta}_T - \theta\big) \overset{d}{=} 2c\Upsilon, \tag{2.3}$$

*where the random variable $\Upsilon$ is defined in* (1.5).

*Proof.* Let $B = \{b_n,\ n \ge 1\}$ be a sequence of positive numbers such that $\lim_{n\to\infty} b_n = +\infty$. Then (2.3) is equivalent to

$$\lim_{n\to\infty} b_n\big(\widehat{\theta}_{b_n} - c^2\big) \overset{d}{=} 2c\Upsilon$$

for every sequence $B$. Working with a sequence is just a technical modification that would allow us to use some standard limit theorems that are stated for sequences.

According to (1.3) and (1.4),

$$X(t) = \frac{1}{c}\int_0^t \sin(c(t-s))dW(s), \tag{2.4}$$

$$\mathbb{E}X^2(t) = \frac{t}{2c^2} - \frac{\sin(2ct)}{4c^3}. \tag{2.5}$$

For $t \in [0,1]$, define the processes

$$\begin{aligned} M^n(t) &= \frac{1}{\sqrt{cb_n}}\int_0^{b_n t} \cos(cs)\, dW(s), \\ N^n(t) &= \frac{1}{\sqrt{cb_n}}\int_0^{b_n t} \sin(cs)\, dW(s). \end{aligned} \tag{2.6}$$

It follows from (2.4) and (2.11) that

$$\frac{1}{b_n}\int_0^{b_n} X(s)dW(s) = \int_0^1 \big(M^n(t)dN^n(t) - N^n(t)dM^n(t)\big) \tag{2.7}$$

$$\frac{1}{b_n^2}\int_0^{b_n} X^2(s)ds = \frac{1}{c}\int_0^1 \big(\sin(cb_nt)N^n(t) - \cos(cb_nt)M^n(t)\big)^2 dt \tag{2.8}$$

$$b_n\big(\widehat{\theta}_{b_n} - c^2\big) = \frac{c\int_0^1 \big(M^n(t)dN^n(t) - N^n(t)dM^n(t)\big)}{\int_0^1 \big(\sin(cb_nt)N^n(t) - \cos(cb_nt)M^n(t)\big)^2 dt}. \tag{2.9}$$

Note that $M^n$ and $N^n$ are continuous square-integrable martingales, and

$$\lim_{n\to\infty}\langle M^n\rangle(t) = \lim_{n\to\infty}\langle N^n\rangle(t) = \frac{t}{2c},$$

$$\lim_{n\to\infty}\langle M^n, N^n\rangle(t) = \lim_{n\to\infty}\frac{1}{2cb_n}\int_0^{b_n t}\sin(2cs)ds = 0.$$

By Theorem VIII.3.11 in [5], the pair $(M^n, N^n)$ converges in distribution to the two-dimensional process $(w_1, w_2)/\sqrt{2c}$, where $w_1, w_2$ are independent standard Brownian motions. Then, after expanding the square and integrating by parts,

$$\begin{aligned} \lim_{n\to\infty}&\int_0^1 \big(\sin(cb_nt)N^n(t) - \cos(cb_nt)M^n(t)\big)^2 dt \\ &\overset{d}{=} \frac{1}{4c}\int_0^1 (w_1^2(t) + w_2^2(t))dt. \end{aligned} \tag{2.10}$$

Next, let us consider the four-dimensional process

$$B^n(t) = \left(M^n(t),\ N^n(t),\ \int_0^t M^n(s)dN^n(s),\ \int_0^t N^n(s)dM_n(s)\right),$$

$0 \le t \le 1$. By Proposition VI.6.13 and Theorem VI.6.22 in [5], this process converges in distribution to

$$B(t) = \left(\frac{w_1(t)}{\sqrt{2c}},\ \frac{w_2(t)}{\sqrt{2c}},\ \frac{1}{2c}\int_0^t w_1(s)dw_2(s),\ \frac{1}{2c}\int_0^t w_1(s)dw_2(s)\right).$$

By the continuous mapping theorem (see, for example, Billingsley [3, Corollary 1.5.1]),

$$\lim_{n\to\infty} \frac{\int_0^1 \big(M^n(t)dN^n(t) - N^n(t)dM^n(t)\big)}{\int_0^1 \big(\sin(cb_nt)N^n(t) - \cos(cb_nt)M^n(t)\big)^2 dt} \overset{d}{=} 2\Upsilon,$$

which concludes the proof of Theorem 2.1. □

Theorem 2.1 is the key to establishing strong consistency of $\widehat{\theta}_T$.

**Theorem 2.2.** *If $\theta = c^2 > 0$, then the estimator $\widehat{\theta}_T$ is strongly consistent in the large sample asymptotic $T \to \infty$:*

$$\lim_{T\to\infty} \widehat{\theta}_T = \theta$$

*with probability one.*

*Proof.* It follows from (2.2) that

$$\widehat{\theta}_T - \theta = -\frac{\int_0^T X(t)dW(t)}{\int_0^T X^2(t)dt}. \tag{2.11}$$

The process $Z(t) = \int_0^t X(s)dW(s)$ is a continuous square-integrable martingale with quadratic characteristic

$$\langle Z\rangle(t) = \int_0^t X^2(s)ds,$$

so that

$$\widehat{\theta}_T - \theta = -\frac{Z(T)}{\langle Z\rangle(T)}.$$

By the strong law of large numbers for martingales (see, for example, [6, Corollary 1 to Theorem 2.6.10]), to complete the proof of the theorem it remains to show that

$$\int_0^\infty X^2(t)dt = +\infty \tag{2.12}$$

with probability one.

Define the random process

$$Q(t) = \int_0^t X^2(s)ds;$$

for notational convenience we switch from $\langle Z\rangle$ to $Q$. The random process $Q = Q(t,\omega)$ is non-decreasing and therefore the limit

$$Q_\infty(\omega) = \lim_{t\to\infty} Q(t,\omega) = \int_0^\infty X^2(s,\omega)ds,$$

finite or infinite, exists for every elementary outcome $\omega \in \Omega$. We need to show that $\mathbb{P}(Q_\infty = +\infty) = 1$, that is, for every sufficiently large $C > 0$ and every sufficiently small $\varepsilon > 0$,

$$\mathbb{P}(Q_\infty > C) > 1 - \varepsilon. \tag{2.13}$$

Fix $C$ and $\varepsilon$. Since $Q_\infty > Q(T)$ for every $T > 0$, we have

$$\mathbb{P}(Q_\infty > C) \geq \mathbb{P}(Q(T) > C) = \mathbb{P}\left(\frac{Q(T)}{T^2} > \frac{C}{T^2}\right). \tag{2.14}$$

By (2.8) and (2.10), for every $x > 0$,

$$\lim_{T\to\infty} \mathbb{P}\left(\frac{Q(T)}{T^2} > x\right) = \mathbb{P}(\xi > x), \tag{2.15}$$

where $\xi$ is an absolutely continuous non-negative random variable and $\mathbb{P}(\xi = 0) = 0$. Accordingly, there exists a $\delta > 0$ such that

$$\mathbb{P}(\xi > \delta) > 1 - \varepsilon.$$

By (2.14), for all $T$ such that $C < T^2\delta$,

$$\mathbb{P}(Q_\infty > C) > \mathbb{P}\left(\frac{Q(T)}{T^2} > \delta\right).$$

Passing to the limit $T \to \infty$ in the last inequality completes the proof of Theorem 2.2 □

## 3. Testing for Damping With Known Frequency

Consider the stochastic differential equation

$$\ddot{X}(t) + a\dot{X}(t) + c^2 X(t) = \dot{W}(t),\ t > 0, X(0) = \dot{X}(0) = 0, \tag{3.1}$$

where $c > 0$ is known. We interpret the equation as a system

$$dX = Y dt;\ dY = (-c^2 X - aY)dt + dW(t).$$

The process $Y = Y(t)$ is a diffusion-type process in the sense of Liptser and Shiryaev; see [7, Definition 4.2.7]. Therefore, by Theorem 7.6 in [7], the measure $P_T^Y$ generated by $(Y(t),\ 0 \leq t \leq T)$ in the space of continuous functions is absolutely continuous with respect to the corresponding measure $P_T^W$ generated by the Brownian motion $(W(t), 0 \leq t \leq T)$, and

$$\frac{dP_T^Y}{dP_T^W}(Y) = \exp\left(-\int_0^T (c^2 X(t) + aY(t))dY(t) - \frac{1}{2}\int_0^T (c^2 X(t) + aY(t))^2 dt\right).$$

Therefore, the maximum likelihood estimator $\widehat{a}_T$ of $a$ is

$$\widehat{a}_T = -\frac{\int_0^T Y(t)dY(t) + c^2\int_0^T X(t)Y(t)dt}{\int_0^T Y^2(t)dt}$$

or, keeping in mind that $Y = \dot{X}$,

$$\widehat{a}_T = \frac{2T - \dot{X}^2(T) - c^2 X^2(T)}{2\int_0^T Y^2(t)dt}. \tag{3.2}$$

It also follows that

$$\widehat{a}_T - a = -\frac{\int_0^T \dot{X}(t)dW(t)}{\int_0^T \dot{X}^2(t)dt}. \tag{3.3}$$

Note that (3.2) is not the same as the estimator considered by Basak and Lee [2]. If $a = 0$, then (2.4) implies

$$\dot{X}(t) = \int_0^t \cos(c(t-s))dW(s).$$

**Theorem 3.1.** *If $a = 0$ and $c > 0$, then*

$$\mathbb{P}(\lim_{T\to\infty} \widehat{a}_T = 0) = 1 \tag{3.4}$$

*and*

$$\lim_{T\to\infty} T\,\widehat{a}_T \stackrel{d}{=} \Psi, \tag{3.5}$$

*with $\Psi$ defined in* (1.5). *Thus, for sufficiently large $T$, the null hypothesis $H_0 : a =$ 0 (no damping) can be rejected in favor of the alternative $H_1 : a > 0$ (damping is present) at the level of significance $\alpha$ if $\widehat{a}_T > \gamma_\alpha/T$, where $\mathbb{P}(\Psi > \gamma_\alpha) = \alpha$.*

*Proof.* The arguments are similar to the proof of Theorem 2.1. We start by establishing (3.5). Let $B = (b_n,\, n \geq 1)$ be a sequence such that $b_n > 0$ and $\lim_{n\to\infty} b_n = +\infty$. With the martingales $M^n$ and $N^n$ defined in (2.6),

$$\frac{1}{b_n}\int_0^{b_n} \dot{X}(s)dW(s) = c\int_0^1 \big(M^n(t)dM^n(t) + M^n(t)dM^n(t)\big)$$

$$\frac{1}{b_n^2}\int_0^{b_n} \dot{X}^2(s)ds = c\int_0^1 \Big(\cos(cb_nt)M^n(t) + \sin(cb_nt)N^n(t)\Big)^2 dt$$

$$b_n\left(\widehat{a}_{b_n} - a\right) = -\frac{\int_0^1 \big(M^n(t)dM^n(t) + N^n(t)dN^n(t)\big)}{\int_0^1 \Big(\cos(cb_nt)M^n(t) + \sin(cb_nt)N^n(t)\Big)^2 dt}.$$

The proof of Theorem 2.1 shows that

$$\begin{aligned}\lim_{n\to\infty}\int_0^1 \big(M^n(t)dM^n(t) + M^n(t)dM^n(t)\big)&\\ &\stackrel{d}{=} \frac{1}{2c}\int_0^1 (w_1(t)dw_1(t) + w_2(t)dw_2(t))\\ &= \frac{w_1^2(1) + w_2^2(1) - 2}{4c},\end{aligned}$$

$$\begin{aligned}\lim_{n\to\infty}\int_0^1 \Big(\cos(cb_nt)M^n(t) + \sin(cb_nt)N^n(t)\Big)^2 dt&\\ &\stackrel{d}{=} \frac{1}{4c}\int_0^1 (w_1^2(t) + w_2^2(t))dt,\end{aligned}$$

and, by the continuous mapping theorem,

$$\lim_{n\to\infty} b_n\left(\widehat{a}_{b_n} - a\right) \stackrel{d}{=} \Psi.$$

To prove (3.4), note that

$$\widehat{a}_T - a = -\frac{Z(T)}{\langle Z\rangle(T)},\ \text{where } Z(t) = \int_0^t \dot{X}(s)dW(s).$$

The proof that $\lim_{T\to\infty}\langle Z\rangle(T) = +\infty$, that is,

$$\int_0^\infty \dot{X}^2(t)dt = +\infty$$

with probability one, is identical to the proof of (2.12). □

To conclude the section, we comment briefly about connections with the discrete time case. In [4, Corollary 3.3.8], Chan and Wei consider the parameter estimation problem for the second-order auto-regression

$$x_n = 2\cos\theta\, x_{n-1} + \alpha\, x_{n-2} + \varepsilon_n \tag{3.6}$$

and show that if $\alpha = -1$, then, for all $\theta \in (0, \pi)$, the least squares estimator $\hat{\alpha}_n$ of $\alpha$ satisfies

$$\lim_{n\to\infty} n(\hat{\alpha}_n + 1) \stackrel{d}{=} \Psi. \tag{3.7}$$

Let us discretize (3.1) using a uniform time step $h$ as follows:

$$\frac{X_n - 2X_{n-1} + X_{n-2}}{h^2} + a\frac{X_{n-1} - X_{n-2}}{h} + c^2 X_{n-1} = \frac{\xi_n}{\sqrt{h}}.$$

Then

$$X_n = (2 - ah - c^2h^2)X_{n-1} + (ah - 1)X_{n-2} + h^{3/2}\xi_n, \tag{3.8}$$

which is of the same form as (3.6), with $2\cos\theta = 2 - ah - c^2h^2$ and $\alpha = ah - 1$. In particular, if $a = 0$, then $\alpha = -1$, and then estimation of $\alpha = -1$ in (3.6) is equivalent to estimation of $a = 0$ in (3.8). Since the limiting distribution of $n(\hat{\alpha}_n + 1)$ in (3.6) does not depend on $\theta$, the limiting distribution of $n\hat{a}_n$ in (3.8) does not depend on $h$, suggesting that the result should continue to hold in the limit $h \to 0$. Theorem 3.1 shows that this is indeed the case, which is rather remarkable, because in general there is little connection between the estimators in continuous time and the corresponding estimators for discretized models. In Section 5 we will see how (3.1) can be discretized exactly, leading to a finite-difference equation that is very different from (3.6) and (3.8).

## 4. Testing for Damping With Unknown Frequency

Consider the stochastic differential equation

$$\ddot{X}(t) + a\dot{X}(t) + \theta\, X(t) = \dot{W}(t),\ t > 0, X(0) = \dot{X}(0) = 0,$$

which we interpret as a system of two equations

$$dX = Y dt;\ dY = (-\theta\, X - aY)dt + dW(t).$$

If $\theta$ is unknown, then testing $a = 0$ vs $a > 0$ requires a joint estimation of $a$ and $\theta$.

The process $Y = Y(t)$ is a diffusion-type process in the sense of Liptser and Shiryaev; see [7, Definition 4.2.7]. Therefore, by Theorem 7.6 in [7], the measure

$P_T^Y$ generated by $(Y(t),\ 0 \le t \le T)$ in the space of continuous functions is absolutely continuous with respect to the corresponding measure $P_T^W$ generated by the Brownian motion $(W(t), 0 \le t \le T)$, and

$$\frac{dP_T^Y}{dP_T^W}(Y) = \exp\left(-\int_0^T (\theta X(t) + aY(t))dY(t) - \frac{1}{2}\int_0^T (\theta\, X(t) + aY(t))^2 dt\right).$$

Keeping in mind that $Y(t) = \dot{X}(t)$, the maximum likelihood estimators $\widehat{a}_T$ and $\widehat{\theta}_T$ of $a$ and $\theta$ given the observations $(X(t), \dot{X}(t),\ 0 \le t \le T)$ are

$$\begin{aligned}\widehat{a}_T &= \frac{\left(\int_0^T X(t)\dot{X}(t)dt\right)\left(\int_0^T X(t)d\dot{X}(t)\right) - \left(\int_0^T X^2(t)dt\right)\left(\int_0^T \dot{X}(t)d\dot{X}(t)\right)}{\left(\int_0^T \dot{X}^2(t)dt\right)\left(\int_0^T X^2(t)dt\right) - \left(\int_0^T X(t)\dot{X}(t)dt\right)^2},\\ \widehat{\theta}_T &= \frac{\left(\int_0^T X(t)\dot{X}(t)dt\right)\left(\int_0^T \dot{X}(t)d\dot{X}(t)\right) - \left(\int_0^T \dot{X}^2(t)dt\right)\left(\int_0^T X(t)d\dot{X}(t)\right)}{\left(\int_0^T \dot{X}^2(t)dt\right)\left(\int_0^T X^2(t)dt\right) - \left(\int_0^T X(t)\dot{X}(t)dt\right)^2}.\end{aligned} \tag{4.1}$$

The estimators are well-defined: by the Cauchy-Schwartz inequality,

$$\left(\int_0^T \dot{X}^2(t)dt\right)\left(\int_0^T X^2(t)dt\right) > \left(\int_0^T X(t)\dot{X}(t)dt\right)^2$$

with probability one.

The amount of numerical integration required to evaluate the above expressions can be reduced using the rules of the usual and stochastic calculus and keeping in mind that the processes $X$ is continuously differentiable, the process $\dot{X}$ is a continuous semi-martingale with quadratic variation equal to $t$, and that $X(0) = \dot{X}(0) = 0$:

$$\begin{aligned}&\int_0^T X(t)\dot{X}(t)dt = \int_0^T X(t)dX(t) = \frac{X^2(T)}{2},\\ &\int_0^T \dot{X}(t)d\dot{X}(t) = \frac{\dot{X}^2(T) - T}{2},\\ &\int_0^T X(t)d\dot{X}(t) = X(T)\dot{X}(T) - \int_0^T \dot{X}^2(t)dt.\end{aligned} \tag{4.2}$$

This leads to equivalent formulas (1.7) in Introduction.

**Theorem 4.1.** *If $\theta = c^2 > 0$ and $a = 0$, then the estimators $\widehat{a}_T$ and $\widehat{\theta}_T$ are strongly consistent in the large sample asymptotic: with probability one,*

$$\lim_{T\to\infty} \widehat{a}_T = 0,\ \lim_{T\to\infty} \widehat{\theta} = c^2, \tag{4.3}$$

*and*

$$\lim_{T\to\infty} T\left(\widehat{a}_T - a\right) \overset{d}{=} \Psi,\ \ \lim_{T\to\infty} T\big(\widehat{\theta}_T - \theta\big) \overset{d}{=} 2c\Upsilon, \tag{4.4}$$

*with random variables* $\Psi$, $\Upsilon$ *defined in* (1.5). *For sufficiently large* $T$, *the null hypothesis* $H_0 : a = 0$ *(no damping) can be rejected in favor of the alternative* $H_1 : a > 0$ *(damping is present) at the level of significance* $\alpha$ *if*

$$\widehat{a}_T > \frac{\gamma_\alpha}{T},$$

*where* $\mathbb{P}(\Psi > \gamma_\alpha) = \alpha$.

*Proof.* Estimators (4.1) are a part of the estimator considered by Basak and Lee [2] for the drift matrix

$$F = \begin{pmatrix} 0 & 1 \\ -\theta & -a \end{pmatrix},$$

and then (4.3) follows from [2, Theorem 2.1].

To establish (4.4), define

$$\begin{aligned} D_T &= \frac{\left(\int_0^T X(t)\dot{X}(t)dt\right)^2}{\left(\int_0^T \dot{X}^2(t)dt\right)\left(\int_0^T X^2(t)dt\right)} \\ &= \frac{\left(\frac{X(T)}{T}\right)^4}{4\left(\frac{1}{T^2}\int_0^T \dot{X}^2(t)dt\right)\left(\frac{1}{T^2}\int_0^T X^2(t)dt\right)}. \end{aligned} \tag{4.5}$$

and rewrite (4.1) as

$$\begin{aligned} T(\widehat{a}_T - a) = \frac{1}{1-D_T}\Bigg( &-\frac{\frac{1}{T}\int_0^T \dot{X}(t)dW(t)}{\frac{1}{T^2}\int_0^T \dot{X}^2(t)dt} \\ &+ \frac{\left(\frac{1}{T}\int_0^T X(t)dW(t)\right)\left(\frac{X^2(T)}{4T^2}\right)}{\left(\frac{1}{T^2}\int_0^T \dot{X}^2(t)dt\right)\left(\frac{1}{T^2}\int_0^T X^2(t)dt\right)}\Bigg), \end{aligned} \tag{4.6}$$

$$\begin{aligned} T(\widehat{\theta}_T - \theta) = \frac{1}{1-D_T}\Bigg( &-\frac{\frac{1}{T}\int_0^T X(t)dW(t)}{\frac{1}{T^2}\int_0^T X^2(t)dt} \\ &+ \frac{\left(\frac{1}{T}\int_0^T \dot{X}(t)dW(t)\right)\left(\frac{X^2(T)}{4T^2}\right)}{\left(\frac{1}{T^2}\int_0^T \dot{X}^2(t)dt\right)\left(\frac{1}{T^2}\int_0^T X^2(t)dt\right)}\Bigg). \end{aligned} \tag{4.7}$$

If $\theta = c^2 > 0$ and $a = 0$, then

$$X(T) = \frac{1}{c}\int_0^T \sin(c(T-s))\,dW(s),\;\; \dot{X}(T) = \int_0^T \cos(c(T-s))\,dW(s).$$

In particular, $\mathbb{E}X^2(T) \le T/c^2$ and therefore.

$$\lim_{T\to\infty} \frac{X^2(T)}{T^2} = 0 \tag{4.8}$$

in probability. Next, (2.3) implies

$$\lim_{T\to\infty} \frac{\frac{1}{T}\int_0^T X(t)dW(t)}{\frac{1}{T^2}\int_0^T X^2(t)dt} \stackrel{d}{=} -2c\Upsilon,$$

and (3.5) implies

$$\frac{\frac{1}{T}\int_0^T \dot{X}(t)dW(t)}{\frac{1}{T^2}\int_0^T \dot{X}^2(t)dt} \stackrel{d}{=} -\Psi.$$

Finally, from (4.8) and the second equality in (4.5),

$$\lim_{T\to\infty} D(T) = 0$$

in probability. This establishes (4.4) and completes the proof of Theorem 4.1. □

## 5. Inference From Discrete Time Observations

The objective of this section is to provide some insight into a more realistic setting when the continuous time process $X = X(t)$ is observed at discrete moments. The corresponding estimation problem admits several formulation and a number of different solutions, and is different from the well-studied second-order auto-regression models. The emphasis below is on the ideas rather than detailed proofs, and the main goal is to suggest ways to *implement* the continuous time estimators.

We begin with the setting of Section 2 and assume that $X = X(t)$ satisfies

$$\ddot{X}(t) + \theta\, X(t) = \dot{W}(t),\ X(0) = \dot{X}(0) = 0, \tag{5.1}$$

and that available observations are $X_k = X(hk)$, $k = 0, 1, \ldots, N$, with fixed and non-random $h > 0$. In other words, the assumption is that the position (and only the position) of the randomly perturbed harmonic oscillator is observed at equally spaced time moments. Under some conditions, it could be natural to assume that both the position $X_k = X(hk)$ and the velocity $\dot{X}_k = \dot{X}(hk)$ are observed, which leads to a different estimation problem. Further modifications could be non-equally spaced observation times, either deterministic or random.

To begin, let us use (5.1) to derive the relation between the samples $X_k$ when $\theta = c^2 > 0$. Using the properties of second-order ordinary differential equations with constant coefficients,

$$\begin{aligned} X(t) &= X_k \cos\big(c(t-kh)\big) + \frac{\dot{X}_k}{c}\,\sin\big(c(t-kh)\big) \\ &\quad + \frac{1}{c}\int_{kh}^{t} \sin\big(c(t-s)\big)\, dW(s). \end{aligned} \tag{5.2}$$

Adding equations (5.2) for $t = (k+1)h$ and $t = (k-1)h$, we find

$$X_{k+1} = 2\cos(ch)X_k - X_{k-1} + \xi_{k+1},\ k \geq 1, \tag{5.3}$$

where

$$\begin{aligned} X_0 = 0,\ X_1 = \xi_1 &= \frac{1}{c}\int_0^h \sin\big(c(h-s)\big)\, dW(s), \\ \xi_{k+1} &= \frac{1}{c}\int_{kh}^{(k+1)h} \sin\big(c((k+1)h-s)\big)\, dW(s) \\ &\quad - \frac{1}{c}\int_{(k-1)h}^{kh} \sin\big(c((k-1)h-s)\big)\, dW(s). \end{aligned} \tag{5.4}$$

By direct computation,

$$\mathbb{E}\big(X_k\xi_{k+1}\big) = \mathbb{E}\big(\xi_k\xi_{k+1}\big) = \frac{ch\cos(ch) - \sin(ch)}{2c} < 0 \text{ for all } ch \in (0, \pi).$$

Let us emphasize that (5.3) is not an approximation but an exact discretizaion of (5.1): $X_k = X(kh)$ for all $k \geq 0$.

While (5.3) looks like a standard second-order auto-regression, there are several major differences from the similar models considered in the literature (for example, by Chan and Wei [4]):

(1) the random variables $\xi_k,\ k \geq 1$, are not independent, which complicates the analysis;
(2) only one parameter is unknown, which means that the "off-the-shelf" estimators, designed for all the coefficients at once, are not the best;
(3) the unknown parameter $c$ is in $\cos(ch)$, which creates identifiability problems if $ch > \pi$.

On the one hand, the sequence $X_k,\ k \geq 1$, is Gaussian (although not Markov). Therefore, the joint density $p_N^\theta$ of the vector $(X_0, \ldots, X_N)$ is known, and leads to the corresponding maximum likelihood estimator of $\theta$. On the other hand, the expression for $p_N^\theta$ is complicated and becomes increasingly complex for larger $N$. Therefore, the closed-form expression for the maximum likelihood estimator is, for all practical purposes, unavailable. A possible approach is to study the maximum likelihood estimator indirectly and then construct an approximate estimator using a suitable approximation of the function $p_N^\theta$. The work of Aït-Sahalia [1] for scalar diffusions suggests that this approach requires a serious investigation well beyond the scope of this paper.

A more straightforward approach is to consider a discrete-time approximation of (2.2) using $d\dot{X} \approx \dot{X}_{k+1} - \dot{X}_k,\ \dot{X}_k \approx (X_k - X_{k-1})/h$ and approximating the integral in the denominator by the left-point rule:

$$\widehat{\theta}_{N,h} = -\frac{\displaystyle\sum_{k=1}^{N-1} \frac{X_{k+1} - 2X_k + X_{k-1}}{h} X_k}{\displaystyle h\sum_{k=1}^{N-1} X_k^2}; \tag{5.5}$$

when $h$ is small, (5.5) is preferable to any alternative expression involving an $h^2$. When $\widehat{\theta}_{N,h} \geq 0$, we call

$$\hat{c}_{N,h} = \sqrt{\widehat{\theta}_{N,h}} \tag{5.6}$$

the D-MLE (discretized maximum likelihood estimator) of $c$. Similar to (2.2), (5.5) makes sense for all real values of $\theta$, and it follows from (5.3) that, when $\theta = c^2 > 0$,

$$\widehat{\theta}_{N,h} - c^2 = -\frac{2\cos(ch) - 2 + c^2h^2}{h^2} - \frac{\sum_{k=0}^{N-1} X_k\xi_{k+1}}{h^2\sum_{k=1}^{N-1} X_k^2}. \tag{5.7}$$

Define the random variable

$$R_N = \frac{\sum_{k=0}^{N-1} X_k\xi_{k+1}}{\sum_{k=1}^{N-1} X_k^2}.$$

Since the sequence $\xi_k,\ k \geq 1$, is stationary and weakly dependent ($\xi_k$ and $\xi_m$ are independent for all $k > m+1$), by analogy with the results of Chan and Wei [4], one would expect that $\lim_{N\to\infty} R_N = 0$ with probability one and $\lim_{N\to\infty} NR_N$ exists in distribution and is a non-degenerate random variable. Numerical experiments seem to be consistent with this conjecture. Still, dependence between $X_k$ and $\xi_{k+1}$ leads to technical complications in the theoretical analysis of $R_N$, and we will not pursue it in this paper.

Equality (5.7) suggests, and numerical experiments confirm, that $\widehat{\theta}_{N,h}$ is not a consistent estimator of $c^2$. In the limit $n \to \infty$, the bias is $2h^{-2}(1-\cos(ch)) - c^2$, which, for small $c^2h^2$, is approximately $-c^4h^2/12$. The asymptotic bias $\lim_{N\to\infty} \hat{c}_{N,h} - c$ of the D-MLE is

$$\delta = \sqrt{2h^{-2}(1-\cos(ch))} - c. \tag{5.8}$$

The relative asymptotic bias $\delta_r = \delta/c$ is a function of $ch$:

$$\delta_r = \frac{\sqrt{2\big(1-\cos(ch)\big)}}{ch} - 1; \tag{5.9}$$

for small $ch$, $\delta_r \approx -c^2h^2/24$. Numerical experiments seem to confirm these conclusions.

If $ch \leq \pi$, relation (5.3) leads to an alternative procedure for estimating $c$. We start with the least-squares estimator $\hat{z}$ of $2\cos(ch)$ by minimizing with respect to $z$ the expression

$$\sum_{k=1}^{N-1} \Big( X_{k+1} - zX_k + X_{k-1} \Big)^2;$$

the result is

$$\hat{z}_{N,h} = \frac{\displaystyle\sum_{k=1}^{N-1} (X_{k+1} + X_{k-1})X_k}{\displaystyle\sum_{k=1}^{N-1} X_k^2}. \tag{5.10}$$

It follows that

$$\hat{z}_{N,h} - 2\cos(ch) = \frac{\sum_{k=0}^{N-1} X_k \xi_{k+1}}{\sum_{k=1}^{N-1} X_k^2},$$

suggesting that $\hat{z}_{N,h}$ is a consistent estimator of $2\cos(ch)$.

Estimator $\hat{z}_{N,h}$ can be defined using discrete samples $X_k = X(kh)$ of the solution of (5.1) for all values of $\theta$. In particular if $\theta = -b^2 < 0$, then $\hat{z}_{N,h}$ is an estimator of $2\cosh(bh)$.

If $\theta = c^2$ and $|\hat{z}_{N,h}| \leq 1$, then

$$\tilde{c}_{N,h} = h^{-1} \arccos(\hat{z}_{N,h}/2) \tag{5.11}$$

is an estimator of $c$. We refer to it as the LSE (least squares estimator).

Figures 2 and 3 present sample realizations of the D-MLE $\hat{c}_{N,h}$ and the LSE $\tilde{c}_{N,h}$ for different values of $c$ and $h$. In Figure 2, $c = \pi$ and $h = 0.1$; in Figure 3, $c = 5\pi$ and $h = 0.02$. In both cases, $N = 2000$ (which explains different time scales), $ch \approx 0.3$ and the relative bias $\delta_r$ of the D-MLE is approximately $-0.004$.

The graphs do not show the highly irregular behavior of the estimators for small values of $N$.

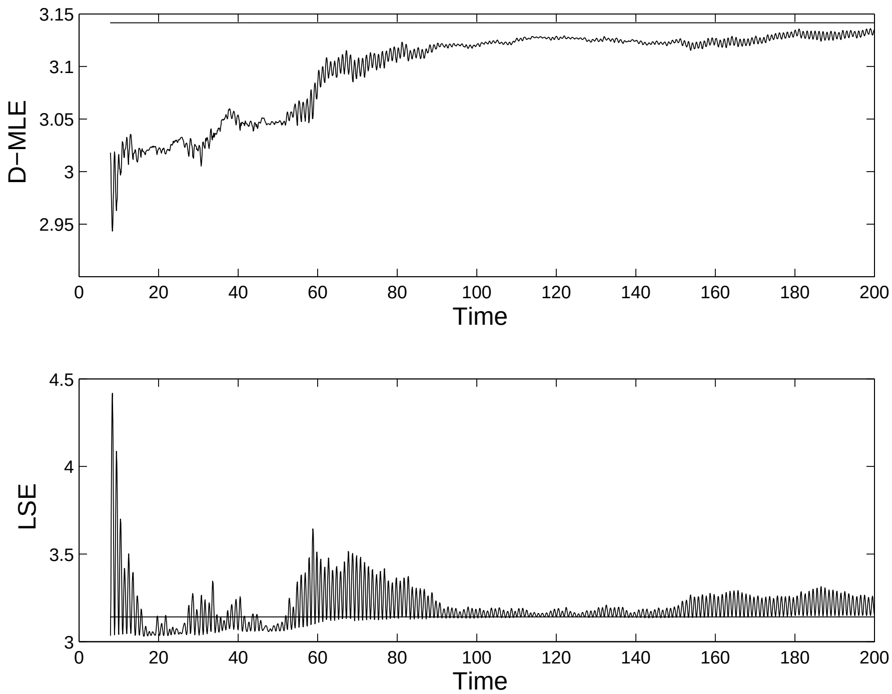


FIGURE 2. Estimators of $c$ when $c = \pi$.

To conclude this section, we note that every second-order equation

$$\ddot{X} + a\dot{X} + \theta X = \dot{W} \tag{5.12}$$

admits a time-series representation of the type (5.3). Indeed, let $u = u(t)$ and $v = v(t)$ be the fundamental family of solutions for (5.12), that is,

$$\begin{aligned} &\ddot{u} + a\dot{u} + \theta\, u = 0,\ u(0) = 1,\ \dot{u}(0) = 0,\\ &\ddot{v} + a\dot{v} + \theta\, v = 0,\ v(0) = 0,\ \dot{v}(0) = 1. \end{aligned} \tag{5.13}$$

By direct computations, and using notation $\nu = \sqrt{|\theta - (a^2/4)|}$,

$$\begin{aligned} u(t) &= \begin{cases} \big(\cos(\nu t) + \frac{a}{2\nu}\sin(\nu t)\big)\, e^{-at/2}, & \text{if } \theta > a^2/4,\\ (1 + (at/2))e^{-at/2}, & \text{if } \theta = a^2/4,\\ \big(\cosh(\nu t) + \frac{a}{2\nu}\sinh(\nu t)\big)\, e^{-at/2}, & \text{if } \theta < a^2/4, \end{cases}\\ v(t) &= \begin{cases} \frac{\sin(\nu t)}{\nu}\, e^{-at/2}, & \text{if } \theta > a^2/4,\\ t\, e^{-at/2}, & \text{if } \theta = a^2/4,\\ \frac{\sinh(\nu t)}{\nu}\, e^{-at/2}, & \text{if } \theta < a^2/4, \end{cases} \end{aligned} \tag{5.14}$$

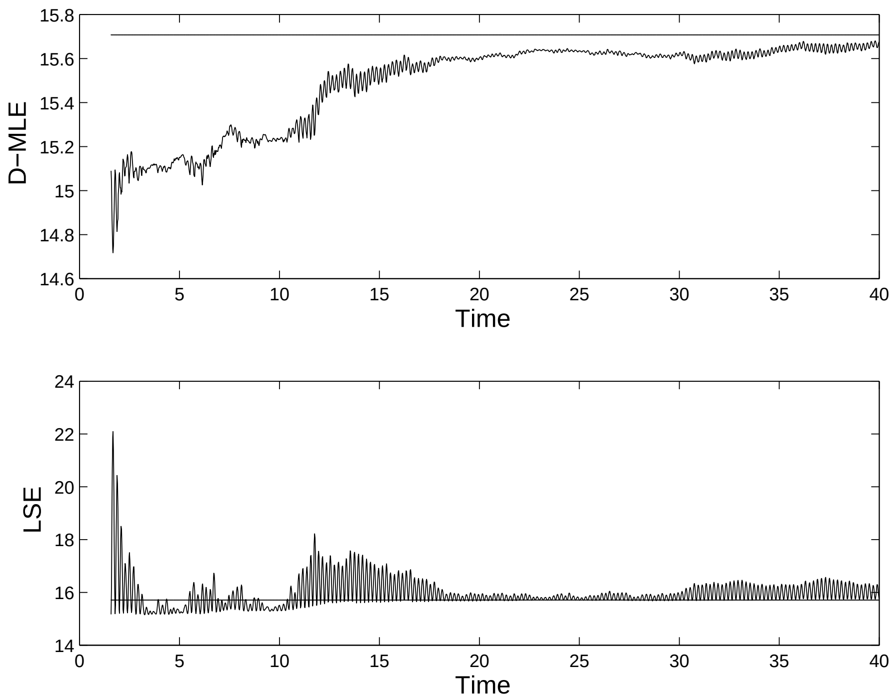


FIGURE 3. Estimators of $c$ when $c = 5\pi$.

Using the notations $t_k = kh$, $X_k = X(t_k)$, $\dot{X}_k = \dot{X}(t_k)$, the solution of (5.12) can now be written as

$$X(t) = X_k u(t - t_k) + \dot{X}_k v(t - t_k) + \int_{t_k}^{t} v(t - s) dW(s). \tag{5.15}$$

The term containing $\dot{X}_k$ is then eliminated by considering (5.15) for $t = t_{k+1}$ and $t = t_{k-1}$. The result is

$$X_{k+1} = AX_k + BX_{k-1} + \xi_{k+1},\ k \geq 1, \tag{5.16}$$

where

$$\begin{aligned} A &= u(h) - \frac{v(h)}{v(-h)} u(-h),\ B = \frac{v(h)}{v(-h)}, \\ \xi_{k+1} &= \int_{t_k}^{t_{k+1}} v(t_{k+1} - s) dW(s) - \frac{v(h)}{v(-h)} \int_{t_{k-1}}^{t_k} v(t_{k-1} - s) dW(s); \end{aligned} \tag{5.17}$$

the values of $X_1$ and $\xi_1$ are obtained from (5.15) when $k = 0$ and $t = t_1$. Note that, by (5.17), $v(t) \neq 0$ in some neighborhood of $t = 0$. We again emphasize that (5.16) is not an approximation but an exact discretizaion of (5.12): $X_k = X(kh)$ for all $k \geq 0$.

An oscillator, damped or amplified, corresponds to $\theta = c^2$, $a \neq 0$, and $\nu^2 = c^2 - (a/2)^2 > 0$, so that

$$u(t) = e^{-at/2}\cos(\nu t) + \frac{a}{2\nu}e^{-at/2}\sin(\nu t),\ v(t) = \frac{1}{\nu}e^{-at/2}\sin(\nu t),$$

and

$$X_{k+1} = 2e^{-ah/2}\,\cos(\nu h)\,X_k - e^{-ah}\,X_{k-1} + \xi_{k+1},\ k \geq 1, \tag{5.18}$$

with

$$\begin{aligned}\xi_{k+1} = \frac{1}{\nu}\int_{t_k}^{t_{k+1}} & e^{-a(t_k-s)/2}\sin(\nu(t_k-s))dW(s)\\ - \frac{e^{-ah}}{\nu}\int_{t_{k-1}}^{t_k} & e^{-a(t_{k-1}-s)/2}\sin(\nu(t_{k-1}-s))dW(s).\end{aligned} \tag{5.19}$$

In particular, sample trajectories in Figure 1 for $a \neq 0$ were produced using (5.18) and (5.19) with $h = 0.04$; for $a = 0$, equations (5.3) and (5.4) were used, also with $h = 0.04$.

Statistical analysis of (5.18) (and more generally, (5.16)) cannot be carried out using the existing results for time series and requires a separate investigation because the random variables $\xi_k,\ k \geq 1$, are dependent, and the unknown parameters enter the regression coefficients in a rather complicated way.

## 6. Conclusions

(1) In this paper, we consider the parameter estimation problem for the stochastic differential equation

$$\ddot{X}(t) + a\dot{X}(t) + \theta\,X(t) = \dot{W},\ X(0) = \dot{X}(0) = 0.$$

We establish consistency and the rate of convergence of the maximum likelihood estimators for $a$ and $\theta$ when $a = 0$ and $\theta > 0$ (undamped harmonic oscillator), and use the result to propose a statistical procedure for testing $a = 0$ vs. $a > 0$.

(2) When $a = 0$ and $\theta > 0$, the limiting distribution of $T(\widehat{a}_T - a)$, as $T \to \infty$, does not depend on $\theta$.

(3) While maximum likelihood estimators (4.1) have the same form for all $a, \theta \in \mathbb{R}$, the rate of convergence and the limiting distribution depend on the specific values of the parameters.

(4) Time discretization of the stochastic differential equation with a constant time step $h$ leads to an exact relation between the samples $X(k) = X(kh)$:

$$X_{k+1} = AX_k + BX_{k-1} + \xi_{k+1},\ k \geq 1,$$

with *dependent* Gaussian noise sequence $\xi_k$ and explicit but complicated connection between $a$, $\theta$ and the coefficients $A$, $B$.

## References

1. Aït-Sahalia, Y.: Likelihood inference for diffusions: A survey, in: *Frontiers in Statistics: in Honor of Peter J. Bickel's 65th Birthday* 369–407, Imperial College Press, 2006.
2. Basak, G. K. and Lee, P.: Asymptotic properties of an estimator of the drift coefficients of multidimensional Ornstein-Uhlenbeck processes that are not necessarily stable, *Electron. J. Stat.* **2** (2008) 1309–1344.
3. Billingsley, P.: *Convergence of Probability Measures*, John Wiley & Sons Inc., New York, 1968.
4. Chan, N. H. and Wei, C. Z.: Limiting distributions of least squares estimates of unstable autoregressive processes, *Ann. Statist.* **16** (1988) 367–401.
5. Jacod, J. and Shiryaev, A. N.: *Limit Theorems for Stochastic Processes, 2nd ed.*, Grundlehren der Mathematischen Wissenschaften, vol. 288, Springer, 2003.
6. Liptser, R. Sh. and Shiryaev, A. N.: *Theory of Martingales*, Mathematics and its Applications (Soviet Series), vol. 49, Kluwer Academic Publishers, Dordrecht, 1989.
7. Liptser, R. Sh. and Shiryaev, A. N.: *Statistics of Random Processes, I: General theory, 2nd ed.*, Applications of Mathematics, vol. 5, Springer, 2001.

N. Lin: Department of Mathematics, USC, Los Angeles, CA 90089 USA, tel. (+1) 213 821 1480; fax: (+1) 213 740 2424
*E-mail address*: nlin@usc.edu

S. V. Lototsky: Department of Mathematics, USC, Los Angeles, CA 90089 USA, tel. (+1) 213 740 2389; fax: (+1) 213 740 2424
*E-mail address*: lototsky@usc.edu
*URL*: http://www-rcf.usc.edu/~lototsky