Bias Analysis in Least Squares Estimation of Periodic Signals Using Nonlinear ODE's*

E. Abd-Elrady†, T. Söderström

Systems and Control, Dept. of Information Technology, Uppsala University,

P.O. Box 337, SE-751 05 Uppsala, Sweden.

Email: emad.abdelrady@it.uu.se, torsten.soderstrom@it.uu.se

Tel.: +46-18-4713389, Fax: +46-18-511925.

Abstract

Periodic signals can be modeled by means of second-order nonlinear ordinary differential equations (ODE's). The right hand side function of the ODE is parameterized in terms of known basis functions. The least squares algorithm developed for estimating the coefficients of these basis functions gives biased estimates, especially at low signal to noise ratios. This is due to noise contributions to the periodic signal and its derivatives evaluated using finite difference approximations. In this paper an analysis for this bias is given.

Keywords: Bias analysis, Discretization errors, Identification, Least squares, Nonlinear systems, Periodic motion.

I. Introduction

Modeling of periodic signals is a fundamental problem in many applications. Examples include vibration analysis, overtone analysis in power networks and measurement of linearity in electronic power amplifiers, see [1-5]. Many systems that generate periodic signals can be described by second-order nonlinear ODE's with polynomial right hand sides. Examples include tunnel diodes, pendulums, negative-resistance oscillators and predator-prey systems, see [6-10].

In this paper the periodic signal is modeled as a function of the states of a general second-order nonlinear ODE, and by introducing a polynomial parameterization of the right hand side of this ODE. Therefore, the approach of this paper is expected to obtain highly accurate models by estimating only a few parameters. Estimators based on Kalman filter and extended Kalman filter (EKF) were developed in [11] using the same idea. A least squares (LS) estimation algorithm was derived in [12]. Also, an algorithm based on the Markov estimate was introduced in [13].

As mentioned in [12, 13], the LS estimator is expected to give biased estimates especially at low signal to noise ratios (SNRs). This is due to two reasons. First, the assumption on the additive noise in the linear regression equation to be white is not valid. In [13] this problem was solved by estimating the covariance matrix of the noise and implementing a Markov estimate based algorithm. Second, derivatives of the modeled signal evaluated using Euler approximations are highly contaminated with noise at low SNRs. Hence, the

*This work was supported by Swedish Research Council for Engineering Sciences under contract 98-654.

† Corresponding author.

regressors and regressed variable of the linear regression equation are also contaminated with noise, and the problem becomes an error-in-variables (EIV) problem.

In this paper, an analysis of the bias in the LS estimate of two nonlinear systems is given assuming that the measurement noise is zero mean, white and Gaussian. The goal is to study the effect of the sampling interval, the signal to noise ratio and the system parameters on the estimation bias.

The paper is organized as follows. Section II introduces the details on the ODE model. Section III reviews the LS estimation algorithm given in [12]. Different estimation errors are discussed in Section IV. In Section V, bias analysis and discretization errors for some simple systems are evaluated. Section VI gives a comparative numerical study between three different Euler approximation techniques. Conclusions appear in Section VII.

II. The ODE model

A. Measurements

The discrete time measured signal $z(kh)$ is given by

$$z(kh) = y(kh) + e(kh) \tag{1}$$

where $y(t)$ is the continuous time signal to be modeled, $y(kh)$ its sampled value, $e(kh)$ is the discrete time measurement noise and h the sampling interval. It is assumed here that $y(t)$ is periodic, *i.e.*

C1: $y(t+T) = y(t), \ \forall t \in R, 0 < T < \infty$.

Furthermore, $e(kh)$ is assumed to be zero mean Gaussian white noise, *i.e.*

C2: $e(kh) \in N(0, \lambda^2), \mathrm{E}[e(kh)e(kh+jh)] = \delta_{j,0}\lambda^2$.

where E is the expectation operator.

B. Model Structures

The idea here is to model the generation of the signal $y(t)$ by means of a nonlinear ODE parameterized with an unknown parameter vector $\boldsymbol{\theta}$, *i.e.*

$$\begin{aligned} \dot{\boldsymbol{x}} &= F(\boldsymbol{x}, \boldsymbol{\theta}) \\ y &= h(\boldsymbol{x}) \end{aligned} \tag{2}$$

As shown in [12], it can often be assumed that the second order ODE

$$\ddot{y}(t) = f\Big(y(t), \dot{y}(t), \boldsymbol{\theta}\Big) \tag{3}$$

generates the periodic signal that is measured. Thus choosing the state variables as follows

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} y(t) \\ \dot{y}(t) \end{pmatrix} \tag{4}$$

the model given in (2) becomes

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \begin{pmatrix} x_2(t) \\ f\Big(x_1(t), x_2(t), \boldsymbol{\theta}\Big) \end{pmatrix} \tag{5}$$

$$y(t) = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}. \tag{6}$$

This model depends only on the parameters of the second right hand side function of (5), a fact that should be advantageous from a computational and performance point of view, see [12] for more details.

It is proved in [14] that the ODE model described in Eqs. (5)-(6) is sufficient to model a large class of periodic signals provided that the following condition holds

C3: The phase plane plot that is constructed from the periodic signal $y(t)$ and its first derivative $\dot{y}(t)$ lacks any intersections or limiting cases such as corners, stops and cusps.

Remark 1. $\mathcal{C}3$ is needed to exclude signal classes that have intersected phase plane plots since these classes need higher order ODEs to be modeled accurately, see [14] for more details.

C. Parameterization

The right hand side of the second state equation of (5) is expanded in terms of known basis functions, modeling the right hand side as a truncated superposition of these functions. A polynomial basis functions were chosen in [11-14] to model $f\big(x_1(t), x_2(t), \boldsymbol{\theta}\big)$. In this case, a suitable parameterization is

$$f\Big(x_1(t), x_2(t), \boldsymbol{\theta}\Big) = \sum_{l=0}^{L} \sum_{m=0}^{M} \theta_{l,m} x_1^l(t) x_2^m(t) \tag{7}$$

$$\boldsymbol{\theta} = \begin{pmatrix} \theta_{0,0} \cdots \theta_{0,M} \cdots \theta_{L,0} \cdots \theta_{L,M} \end{pmatrix}^T \tag{8}$$

III. The least squares algorithm

In order to formulate the model in a linear regression form, note that Eqs. (5), (7) and (8) result in the model

$$\dot{x}_1(t) = x_2(t) \tag{9}$$

$$\dot{x}_2(t) = \boldsymbol{\phi}^T\Big(x_1(t), x_2(t)\Big)\ \boldsymbol{\theta} \tag{10}$$

where

$$\boldsymbol{\phi}^T\Big(x_1(t), x_2(t)\Big) = \begin{pmatrix} 1 & \cdots & x_2^M(t) & \cdots & x_1^L(t) & \cdots & x_1^L(t) x_2^M(t) \end{pmatrix} \tag{11}$$

To estimate the parameter vector $\boldsymbol{\theta}$ from (10), some approximations are needed. Since $x_1(t)$, $x_2(t)$ and $\dot{x}_2(t)$ are not known, their estimates should be used. In this case, the second state equation (10) results in (at $t = kh$)

$$\widehat{\dot{x}}_2(kh) = \boldsymbol{\phi}^T\Big(\widehat{x}_1(kh), \widehat{x}_2(kh)\Big)\ \boldsymbol{\theta} + \varepsilon(kh) \tag{12}$$

The expression (12) follows by performing a Taylor series expansion of the regression vector $\boldsymbol{\phi}^T\big(x_1(kh), x_2(kh)\big)$ around $\big(\widehat{x}_1(kh) \quad \widehat{x}_2(kh)\big)^T$. In (12) the combined regression error, $\varepsilon(kh)$, has been introduced. It can not be expected to be white since the Taylor series expansion in (12) produces a sum of noise samples that are delayed with a varying number of sampling periods.

Assuming that data are available at times $kh - Nh, \cdots, kh$ and defining the vectors and matrices

$$\boldsymbol{Z}_N = \begin{pmatrix} \widehat{\dot{x}}_2(kh) & \cdots & \widehat{\dot{x}}_2(kh - Nh) \end{pmatrix}^T \tag{13}$$

$$\boldsymbol{\Phi}_N = \begin{pmatrix} \boldsymbol{\phi}^T\Big(\widehat{x}_1(kh), \widehat{x}_2(kh)\Big) \\ \vdots \\ \boldsymbol{\phi}^T\Big(\widehat{x}_1(kh - Nh), \widehat{x}_2(kh - Nh)\Big) \end{pmatrix} \tag{14}$$

$$\boldsymbol{\varepsilon}_N = \begin{pmatrix} \varepsilon(kh) & \cdots & \varepsilon(kh - Nh) \end{pmatrix}^T \tag{15}$$

the following regression equation results

$$\boldsymbol{Z}_N = \boldsymbol{\Phi}_N \ \boldsymbol{\theta} + \boldsymbol{\varepsilon}_N \tag{16}$$

Hence the LS estimate $\widehat{\boldsymbol{\theta}}_N^{LS}$ is given by, see [15]

$$\widehat{\boldsymbol{\theta}}_N^{LS} = (\boldsymbol{\Phi}_N^T \boldsymbol{\Phi}_N)^{-1} \ \boldsymbol{\Phi}_N^T \ \boldsymbol{Z}_N. \tag{17}$$

The LS estimate (17) has been studied in [12] with $\widehat{x}_1(kh)$, $\widehat{x}_2(kh)$ and $\widehat{\dot{x}}_2(kh)$ evaluated using finite difference approximation. It is shown in [12] that the LS algorithm gives considerably accurate models at high SNRs and further research is needed to extend the operating region toward low SNRs. This is because of noise amplification during the differentiation process.

In [16] the automated spectral analysis (ASA) technique [17] is used to find more accurate estimates for the noise free periodic signal $x_1(t)$ and its time derivatives $x_2(t)$ and $\dot{x}_2(t)$. An accurate estimate $\widehat{x}_1(t)$ is found by eliminating noise corruption on the modeled signal and more accurate estimates $\widehat{x}_2(t)$ and $\widehat{\dot{x}}_2(t)$ are found by differentiating $\widehat{x}_1(t)$ in the frequency domain. Although the approach of [16] shows a significant improvement in the performance compared to using finite difference approximations, it is still interesting to study the bias of the approach in [12].

IV. Estimation Errors

As mentioned in the previous section, in order to estimate the parameter vector $\boldsymbol{\theta}$ from the linear regression equation (10), some estimates of the modeled signal $x_1(kh)$, its first derivative $x_2(kh)$ and second derivative $\dot{x}_2(kh)$ need to be found. Estimating these quantities using finite difference approximations from the measured data leads to some estimation errors. It is expected that the LS estimates will suffer from two estimation errors, namely: *random noise errors* and *discretization errors*. Random noise errors results due to differentiating additive measurement noise. On the other hand, discretization errors are caused by approximating the signal derivatives using finite approximations.

To investigate how the estimate $\widehat{\boldsymbol{\theta}}_N^{LS}$ behaves when the data length N becomes large, consider the following

$$\boldsymbol{R} = \overline{\mathrm{E}}\left[\boldsymbol{\phi}\Big(\widehat{x}_1(t),\widehat{x}_2(t)\Big)\ \boldsymbol{\phi}^T\Big(\widehat{x}_1(t),\widehat{x}_2(t)\Big)\right] \tag{18}$$

$$\boldsymbol{r} = \overline{\mathrm{E}}\left[\boldsymbol{\phi}\Big(\widehat{x}_1(t),\widehat{x}_2(t)\Big)\ \widehat{\dot{x}}_2(t)\right] \tag{19}$$

where

$$\boldsymbol{\phi}^T\Big(\widehat{x}_1(t),\widehat{x}_2(t)\Big) = \Big(\ 1\ \cdots\ \widehat{x}_2^M(t)\ \cdots\ \widehat{x}_1^L(t)\ \cdots\ \widehat{x}_1^L(t)\widehat{x}_2^M(t)\ \Big) \tag{20}$$

and

$$\overline{\mathrm{E}}f(t) = \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\mathrm{E}f(t) \tag{21}$$

Remark 2. $\overline{\mathrm{E}}$ is used instead of the ordinary expectation E to account for noise-free signals. For fully random signals $\overline{\mathrm{E}} = \mathrm{E}$.

In this case the asymptotic parameter vector estimate $\bar{\boldsymbol{\theta}}$ is given by

$$\bar{\boldsymbol{\theta}} = \boldsymbol{R}^{-1}\ \boldsymbol{r} = \boldsymbol{\theta}_0 + \tilde{\boldsymbol{\theta}}_b \tag{22}$$

where $\boldsymbol{\theta}_0$ is the true parameter vector and $\tilde{\boldsymbol{\theta}}_b$ is the bias vector. Similarly

$$\boldsymbol{R} = \boldsymbol{R}_0 + \tilde{\boldsymbol{R}}_b \tag{23}$$

$$\boldsymbol{r} = \boldsymbol{r}_0 + \tilde{\boldsymbol{r}}_b \tag{24}$$

where

$$\begin{aligned}\boldsymbol{R}_0 &= \overline{\mathrm{E}}\left[\boldsymbol{\phi}\Big(x_1(t),x_2(t)\Big)\ \boldsymbol{\phi}^T\Big(x_1(t),x_2(t)\Big)\right]\\ \boldsymbol{r}_0 &= \overline{\mathrm{E}}\left[\boldsymbol{\phi}\Big(x_1(t),x_2(t)\Big)\ \dot{x}_2(t)\right]\end{aligned} \tag{25}$$

and $\tilde{\boldsymbol{R}}_b$ and $\tilde{\boldsymbol{r}}_b$ are the bias contributions to $\boldsymbol{R}_0$ and $\boldsymbol{r}_0$, respectively, due to using estimated states $\widehat{x}_1(t)$ and $\widehat{x}_2(t)$ instead of the true states.

Now using (22)-(24) gives

$$\begin{aligned}\bar{\boldsymbol{\theta}} &= (\boldsymbol{R}_0 + \tilde{\boldsymbol{R}}_b)^{-1}\ (\boldsymbol{r}_0 + \tilde{\boldsymbol{r}}_b)\\ &= \underbrace{\boldsymbol{R}_0^{-1}\boldsymbol{r}_0}_{\boldsymbol{\theta}_0} + \underbrace{(\boldsymbol{R}_0 + \tilde{\boldsymbol{R}}_b)^{-1}(\tilde{\boldsymbol{r}}_b - \tilde{\boldsymbol{R}}_b\boldsymbol{\theta}_o)}_{\tilde{\boldsymbol{\theta}}_b}\end{aligned} \tag{26}$$

Remark 3. The bias vector $\tilde{\boldsymbol{\theta}}_b$ is a contribution of random noise errors and discretization errors. The contributions of random noise errors and discretization errors are denoted in this paper by $\tilde{\boldsymbol{\theta}}_n$ and $\tilde{\boldsymbol{\theta}}_d$ respectively. Hence

$$\tilde{\boldsymbol{\theta}}_b = \tilde{\boldsymbol{\theta}}_n + \tilde{\boldsymbol{\theta}}_d \tag{27}$$

where

$$\tilde{\boldsymbol{\theta}}_n = (\boldsymbol{R}_0 + \tilde{\boldsymbol{R}}_n)^{-1}(\tilde{\boldsymbol{r}}_n - \tilde{\boldsymbol{R}}_n\boldsymbol{\theta}_o) \tag{28}$$

$$\tilde{\boldsymbol{\theta}}_d = (\boldsymbol{R}_0 + \tilde{\boldsymbol{R}}_d)^{-1}(\tilde{\boldsymbol{r}}_d - \tilde{\boldsymbol{R}}_d\boldsymbol{\theta}_o) \tag{29}$$

Also, here $\tilde{\boldsymbol{R}}_n$ and $\tilde{\boldsymbol{r}}_n$ are the bias contributions to $\boldsymbol{R}_0$ and $\boldsymbol{r}_0$, respectively, due to random noise errors; $\tilde{\boldsymbol{R}}_d$ and $\tilde{\boldsymbol{r}}_d$ are the bias contributions to $\boldsymbol{R}_0$ and $\boldsymbol{r}_0$, respectively, due to discretization errors.

The bias vector $\tilde{\boldsymbol{\theta}}_b$ will depend on the sampling interval h and the derivative approximations. Derivative approximations can always be chosen by the user. A general approximation for the derivatives of the modeled signal is [18]

$$D^n f(t) = \frac{1}{h^n}\sum_j \Gamma_{n,j} f(t+jh) \tag{30}$$

which satisfy the following conditions

$$\sum_j \Gamma_{n,j} j^v = \begin{cases} 0 & v < n \\ n! & v = n \end{cases} \tag{31}$$

The frequency function of the differentiator (30) satisfies

$$D^n(e^{j\omega}) = (i\omega)^n + O(|\omega|^{n+1}) \tag{32}$$

where $O(.)$ denotes the ordo concept. This means that the low-frequency asymptote of the differentiator frequency function is hence $(i\omega)^n$. See [18, 19] for more details.

Remark 4. As can be noticed from (31), the minimal number of terms in (30) is $n+1$. In such a case D^n will be a high-pass filter. If the number of coefficients is increased, the gained degree of freedom can be used to decrease the high frequency gain to avoid differentiating the noise as much as possible. See [20] for more details about design of differentiating filters.

In this paper the estimation of the parameter vector will be considered for the following *three* simple finite difference approximations of $x_2(kh)$ and $\dot{x}_2(kh)$ which are special cases from (30) (in all approximations, we chosed $\widehat{x}_1(kh) = z(kh)$):

$\mathcal{A}1$: Euler backward approx. (EB)

$$\begin{aligned} \widehat{x}_2(kh) &= \frac{1}{h}\sum_{j=-1}^{0} \Gamma_{1,j}\ \widehat{x}_1(kh+jh),\ \ \Gamma_{1,-1} = -1 \text{ and } \Gamma_{1,0} = 1 \\ \widehat{\dot{x}}_2(kh) &= \frac{1}{h^2}\sum_{j=-2}^{0} \Gamma_{2,j}\ \widehat{x}_1(kh+jh),\ \ \Gamma_{2,-2} = 1,\ \Gamma_{2,-1} = -2 \text{ and } \Gamma_{2,0} = 1 \end{aligned} \tag{33}$$

$\mathcal{A}2$: Euler forward approx. (EF)

$$\begin{aligned}
\widehat{x}_2(kh) &= \frac{1}{h}\sum_{j=0}^{1}\Gamma_{1,j}\ \widehat{x}_1(kh+jh),\ \Gamma_{1,0}=-1 \text{ and } \Gamma_{1,1}=1 \\
\widehat{\dot{x}}_2(kh) &= \frac{1}{h^2}\sum_{j=0}^{2}\Gamma_{2,j}\ \widehat{x}_1(kh+jh),\ \Gamma_{2,0}=1,\ \Gamma_{2,1}=-2 \text{ and } \Gamma_{2,2}=1
\end{aligned} \tag{34}$$

$\mathcal{A}3$: Euler center approx. (EC)

$$\begin{aligned}
\widehat{x}_2(kh) &= \frac{1}{h}\sum_{j=-1}^{1}\Gamma_{1,j}\ \widehat{x}_1(kh+jh),\ \Gamma_{1,-1}=-\frac{1}{2},\ \Gamma_{1,0}=0 \text{ and } \Gamma_{1,1}=\frac{1}{2} \\
\widehat{\dot{x}}_2(kh) &= \frac{1}{h^2}\sum_{j=-2}^{2}\Gamma_{2,j}\ \widehat{x}_1(kh+jh),\ \Gamma_{2,-2}=\frac{1}{4},\ \Gamma_{2,-1}=0,\ \Gamma_{2,0}=-\frac{1}{2},\ \Gamma_{2,1}=0, \text{ and } \Gamma_{2,2}=\frac{1}{4}
\end{aligned} \tag{35}$$

Remark 5. $\mathcal{A}1$-$\mathcal{A}3$ are chosen as examples for finite difference approximation. Many other different approximations can be considered. For example, one can think about evaluating $\widehat{x}_2(kh)$ and $\widehat{\dot{x}}_2(kh)$ using two different finite difference approximations.

It is well known from the numerical analysis literature that EC approximation gives lower discretization error compared to EB and EF approximations. This is because EC approximation gives a higher order (in h) remainder term. Hence the EC approximation provides smaller remainder term (for small h) than EB and EF approximations. Also, the discretization error in this case is expected to decrease as h decreases. On the other hand, random noise error gives both bias and variance contributions to the LS estimates. It is expected that the noise error increases as h decreases since the noise will be highly amplified (h appears in the denominator of the differentiator). Small values of h will only be suitable in the case the signal to noise ratio is high. The effect of h on the noise errors can be somewhat lower if more general derivative approximations such as (30) are used.

In the next section some simple systems will be considered to analyze different estimation errors. The aim of this study is to know if the bias in the LS estimates will follow the same expectations as for discretization errors and random noise errors, and if these two errors are additive so we can find an optimal sampling interval h for each system.

V. Explicit Analysis for Simple Systems

In this section the study will concentrate on the following two nonlinear systems:

$\mathcal{S}1$:

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \begin{pmatrix} x_2 \\ -\eta x_1^3 \end{pmatrix} \tag{36}$$

$\mathcal{S}2$:

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \begin{pmatrix} x_2 \\ \alpha x_1 + \beta x_2 + \gamma x_1^2 x_2 \end{pmatrix} \tag{37}$$

Remark 6. Note that $\mathcal{S}1$ does not have a *unique stable* periodic orbit as the case in $\mathcal{S}2$. Therefore, the amplitude of the periodic signal generated by $\mathcal{S}1$ is fully determined by the initial state.

A. Random noise errors

Now let $\widehat{x}_1(kh) = x_1(kh) + \tilde{x}_1(kh)$ where $\tilde{x}_1(kh)$ represents the noise contribution. It is clear from Eqs. (1) and (4) that $\tilde{x}_1(kh) = e(kh)$. Similarly, we can evaluate $\tilde{x}_2(kh)$ and $\tilde{\dot{x}}_2(kh)$ for Eqs. (33)-(35). Therefore, the noise contribution in $\widehat{x}_2(kh)$ and $\widehat{\dot{x}}_2(kh)$ are as follows:

$\mathcal{A}1$: (EB)

$$\begin{aligned}\tilde{x}_2(kh) &= \frac{e(kh) - e(kh-h)}{h}\\ \tilde{\dot{x}}_2(kh) &= \frac{e(kh) - 2e(kh-h) + e(kh-2h)}{h^2}\end{aligned} \tag{38}$$

$\mathcal{A}2$: (EF)

$$\begin{aligned}\tilde{x}_2(kh) &= \frac{e(kh+h) - e(kh)}{h}\\ \tilde{\dot{x}}_2(kh) &= \frac{e(kh+2h) - 2e(kh+h) + e(kh)}{h^2}\end{aligned} \tag{39}$$

$\mathcal{A}3$: (EC)

$$\begin{aligned}\tilde{x}_2(kh) &= \frac{e(kh+h) - e(kh-h)}{2h}\\ \tilde{\dot{x}}_2(kh) &= \frac{e(kh+2h) - 2e(kh) + e(kh-2h)}{4h^2}\end{aligned} \tag{40}$$

Remark 7. Note that by $\mathcal{C}2$, we have $\overline{\mathrm{E}}(\tilde{x}_1^i) = 0, \forall i$ odd, $\overline{\mathrm{E}}(\tilde{x}_1^2) = \lambda^2$, $\overline{\mathrm{E}}(\tilde{x}_1^4) = 3\lambda^4$, and $\overline{\mathrm{E}}(\tilde{x}_1^6) = 15\lambda^6$.

Remark 8. Note that for noise-free periodic data

$$\begin{aligned}\overline{\mathrm{E}}(x_1^i x_2) &= \frac{1}{T}\int_0^T x_1^i(t)x_2(t)dt = \frac{1}{T}\int_0^T x_1^i(t)\dot{x}_1(t)dt\\ &= \frac{1}{T(i+1)}\left[x_1^{i+1}(T) - x_1^{i+1}(0)\right] = 0,\ \forall\ i.\end{aligned} \tag{41}$$

Similarly, $\overline{\mathrm{E}}(x_2^i \dot{x}_2) = 0,\ \forall\ i.$

In the following two examples, the bias contribution due to random noise errors is analyzed for $\mathcal{S}1$ and $\mathcal{S}2$.

Example 1. *Random noise errors of* $\mathcal{S}1$.

Here $\boldsymbol{\phi} = -\widehat{x}_1^3$. Thus $\boldsymbol{R} = \overline{\mathrm{E}}(\widehat{x}_1^6)$ and $\boldsymbol{r} = \overline{\mathrm{E}}(-\widehat{x}_1^3\widehat{\dot{x}}_2)$. Straightforward calculations give

$$\begin{aligned}\boldsymbol{R} &= \overline{\mathrm{E}}(x_1 + \tilde{x}_1)^6\\ &= \underbrace{\overline{\mathrm{E}}(x_1^6)}_{\boldsymbol{R}_0} + \underbrace{15\lambda^2\overline{\mathrm{E}}(x_1^4) + 45\lambda^4\overline{\mathrm{E}}(x_1^2) + 15\lambda^6}_{\tilde{\boldsymbol{R}}_n}\end{aligned} \tag{42}$$

Similarly,

$$
\begin{aligned}
\boldsymbol{r} &= \overline{\mathrm{E}}[-(x_1+\tilde{x}_1)^3(\dot{x}_2+\dot{\tilde{x}}_2)] \\
&= -\overline{\mathrm{E}}(x_1^3\dot{x}_2) - 3\overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(\tilde{x}_1\dot{\tilde{x}}_2) - 3\lambda^2\overline{\mathrm{E}}(x_1\dot{x}_2) - \overline{\mathrm{E}}(\tilde{x}_1^3\dot{\tilde{x}}_2) \\
&= \underbrace{\eta\overline{\mathrm{E}}(x_1^6)}_{\boldsymbol{r}_0} \underbrace{-3\overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(\tilde{x}_1\dot{\tilde{x}}_2) + 3\lambda^2\eta\overline{\mathrm{E}}(x_1^4) - \overline{\mathrm{E}}(\tilde{x}_1^3\dot{\tilde{x}}_2)}_{\tilde{\boldsymbol{r}}_n}
\end{aligned}
\tag{43}
$$

Since $\boldsymbol{R}$ and $\boldsymbol{r}$ are scalars in this case, Eq. (28) gives

$$
\bar{\boldsymbol{\theta}} = \underbrace{\frac{\boldsymbol{r}_0}{\boldsymbol{R}_0}}_{\eta} + \underbrace{\frac{\tilde{\boldsymbol{r}}_n\boldsymbol{R}_0 - \boldsymbol{r}_0\tilde{\boldsymbol{R}}_n}{\boldsymbol{R}_0\boldsymbol{R}}}_{\tilde{\boldsymbol{\theta}}_n}
\tag{44}
$$

Thus we have

$$
\tilde{\boldsymbol{\theta}}_n = \frac{\tilde{\boldsymbol{r}}_n - \eta\tilde{\boldsymbol{R}}_n}{\boldsymbol{R}}
\tag{45}
$$

Equation (45) shows that unbiased estimate of η can be obtained if both $\boldsymbol{r}_0$ and $\boldsymbol{R}_0$ are perturbed proportionally to the ratio between them, *i.e.* $\boldsymbol{r}_0/\boldsymbol{R}_0 = \tilde{\boldsymbol{r}}_n/\tilde{\boldsymbol{R}}_n$.

Now, the bias $\tilde{\boldsymbol{\theta}}_n$ can be evaluated for $\mathcal{A}1$ - $\mathcal{A}3$ replacing $\overline{\mathrm{E}}(\tilde{x}_1\dot{\tilde{x}}_2)$ and $\overline{\mathrm{E}}(\tilde{x}_1^3\dot{\tilde{x}}_2)$ by their corresponding values

TABLE I

Values of some expectations.

Quantity	EB	EF	EC
$\overline{\mathrm{E}}(\tilde{x}_2^2)$	$2\lambda^2/h^2$	$2\lambda^2/h^2$	$\lambda^2/2h^2$
$\overline{\mathrm{E}}(\tilde{x}_1\tilde{x}_2)$	λ^2/h	$-\lambda^2/h$	0
$\overline{\mathrm{E}}(\tilde{x}_1^2\tilde{x}_2^2)$	$4\lambda^4/h^2$	$4\lambda^4/h^2$	$\lambda^4/2h^2$
$\overline{\mathrm{E}}(\tilde{x}_1^3\tilde{x}_2)$	$3\lambda^4/h$	$-3\lambda^4/h$	0
$\overline{\mathrm{E}}(\tilde{x}_1^4\tilde{x}_2^2)$	$18\lambda^6/h^2$	$18\lambda^6/h^2$	$3\lambda^6/2h^2$
$\overline{\mathrm{E}}(\tilde{x}_2\dot{\tilde{x}}_2)$	$3\lambda^2/h^3$	$-3\lambda^2/h^3$	0
$\overline{\mathrm{E}}(\tilde{x}_1\dot{\tilde{x}}_2)$	λ^2/h^2	λ^2/h^2	$-\lambda^2/2h^2$
$\overline{\mathrm{E}}(\tilde{x}_1^3\dot{\tilde{x}}_2)$	$3\lambda^4/h^2$	$3\lambda^4/h^2$	$-3\lambda^4/2h^2$
$\overline{\mathrm{E}}(\tilde{x}_1^2\tilde{x}_2\dot{\tilde{x}}_2)$	$5\lambda^4/h^3$	$-5\lambda^4/h^3$	0

as indicated in Table I which can be straightforward derived using Eqs. (38)-(40). Hence, we have

$$
\tilde{\boldsymbol{\theta}}_n^{EB} = \tilde{\boldsymbol{\theta}}_n^{EF} = \frac{-3\lambda^2(\frac{1}{h^2} - 15\lambda^2\eta)\overline{\mathrm{E}}(x_1^2) - 12\lambda^2\eta\overline{\mathrm{E}}(x_1^4) - 3\lambda^4(\frac{1}{h^2} + 5\lambda^2\eta)}{\overline{\mathrm{E}}(x_1^6) + 15\lambda^2\overline{\mathrm{E}}(x_1^4) + 45\lambda^4\overline{\mathrm{E}}(x_1^2) + 15\lambda^6}
\tag{46}
$$

$$
\tilde{\boldsymbol{\theta}}_n^{EC} = \frac{3\lambda^2(\frac{1}{2h^2} - 15\lambda^2\eta)\overline{\mathrm{E}}(x_1^2) - 12\lambda^2\eta\overline{\mathrm{E}}(x_1^4) + 3\lambda^4(\frac{1}{2h^2} - 5\lambda^2\eta)}{\overline{\mathrm{E}}(x_1^6) + 15\lambda^2\overline{\mathrm{E}}(x_1^4) + 45\lambda^4\overline{\mathrm{E}}(x_1^2) + 15\lambda^6}
\tag{47}
$$

For high signal to noise ratio (SNR), it can be assumed that $\overline{\mathrm{E}}(x_1^6) \gg 15\lambda^2\overline{\mathrm{E}}(x_1^4)$ and that the effect of the terms which contain λ^4 and λ^6 can be neglected. Hence,

$$\tilde{\boldsymbol{\theta}}_n^{EB} = \tilde{\boldsymbol{\theta}}_n^{EF} \approx \frac{-\frac{3\lambda^2}{h^2}\overline{\mathrm{E}}(x_1^2) - 12\lambda^2\eta\overline{\mathrm{E}}(x_1^4)}{\overline{\mathrm{E}}(x_1^6)} \tag{48}$$

$$\tilde{\boldsymbol{\theta}}_n^{EC} \approx \frac{\frac{3\lambda^2}{2h^2}\overline{\mathrm{E}}(x_1^2) - 12\lambda^2\eta\overline{\mathrm{E}}(x_1^4)}{\overline{\mathrm{E}}(x_1^6)} \tag{49}$$

Also, multiplying the numerators of Eqs. (48)-(49) by $\frac{\overline{\mathrm{E}}(x_1^2)}{\overline{\mathrm{E}}(x_1^2)}$, and noticing that SNR$=\overline{\mathrm{E}}(x_1^2)/\lambda^2$ give

$$\tilde{\boldsymbol{\theta}}_n^{EB} = \tilde{\boldsymbol{\theta}}_n^{EF} \approx \frac{-\frac{3}{h^2}[\overline{\mathrm{E}}(x_1^2)]^2 - 12\eta\overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_1^4)}{\mathrm{SNR}\ \overline{\mathrm{E}}(x_1^6)} \tag{50}$$

$$\tilde{\boldsymbol{\theta}}_n^{EC} \approx \frac{\frac{3}{2h^2}[\overline{\mathrm{E}}(x_1^2)]^2 - 12\eta\overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_1^4)}{\mathrm{SNR}\ \overline{\mathrm{E}}(x_1^6)} \tag{51}$$

Therefore, $\tilde{\boldsymbol{\theta}}_n$ satisfies the following relations for high SNR and for small h;

$$\tilde{\boldsymbol{\theta}}_n \propto \frac{1}{\mathrm{SNR}} \tag{52}$$

$$\tilde{\boldsymbol{\theta}}_n \propto \frac{1}{h^2} \tag{53}$$

Example 2. *Random noise errors for $\mathcal{S}2$.*

In this case, we have $\boldsymbol{\phi}^T = (\widehat{x}_2 \quad \widehat{x}_1 \quad \widehat{x}_1^2\widehat{x}_2)$ and $\boldsymbol{\theta} = (\beta \quad \alpha \quad \gamma)^T$. Therefore

$$\boldsymbol{R} = \overline{\mathrm{E}}(\boldsymbol{\phi}\boldsymbol{\phi}^T) = \overline{\mathrm{E}}\begin{pmatrix} \widehat{x}_2^2 & \widehat{x}_1\widehat{x}_2 & \widehat{x}_1^2\widehat{x}_2^2 \\ \widehat{x}_1\widehat{x}_2 & \widehat{x}_1^2 & \widehat{x}_1^3\widehat{x}_2 \\ \widehat{x}_1^2\widehat{x}_2^2 & \widehat{x}_1^3\widehat{x}_2 & \widehat{x}_1^4\widehat{x}_2^2 \end{pmatrix} \tag{54}$$

and

$$\boldsymbol{r} = \overline{\mathrm{E}}(\boldsymbol{\phi}\widehat{\dot{x}}_2) = \overline{\mathrm{E}}\begin{pmatrix} \widehat{x}_2\widehat{\dot{x}}_2 \\ \widehat{x}_1\widehat{\dot{x}}_2 \\ \widehat{x}_1^2\widehat{x}_2\widehat{\dot{x}}_2 \end{pmatrix} \tag{55}$$

Similarly as done in Example 1, replacing $\widehat{x}_1$, $\widehat{x}_2$ and $\widehat{\dot{x}}_2$ by $(x_1 + \tilde{x}_1)$, $(x_2 + \tilde{x}_2)$ and $(\dot{x}_2 + \tilde{\dot{x}}_2)$, it follows that

$$\boldsymbol{R}(1,1) = \overline{\mathrm{E}}(x_2^2) + \overline{\mathrm{E}}(\tilde{x}_2^2) \tag{56}$$

$$\boldsymbol{R}(1,2) = \overline{\mathrm{E}}(\tilde{x}_1\tilde{x}_2) \tag{57}$$

$$\boldsymbol{R}(1,3) = \overline{\mathrm{E}}(x_1^2x_2^2) + \lambda^2\overline{\mathrm{E}}(x_2^2) + \overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(\tilde{x}_2^2) + \overline{\mathrm{E}}(\tilde{x}_1^2\tilde{x}_2^2) \tag{58}$$

$$\boldsymbol{R}(2,2) = \overline{\mathrm{E}}(x_1^2) + \lambda^2 \tag{59}$$

$$\boldsymbol{R}(2,3) = 3\overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(\tilde{x}_1\tilde{x}_2) + \overline{\mathrm{E}}(\tilde{x}_1^3\tilde{x}_2) \tag{60}$$

$$\boldsymbol{R}(3,3) = \overline{\mathrm{E}}(x_1^4x_2^2) + 6\lambda^2\overline{\mathrm{E}}(x_1^2x_2^2) + 3\lambda^4\overline{\mathrm{E}}(x_2^2) + \overline{\mathrm{E}}(x_1^4)\overline{\mathrm{E}}(\tilde{x}_2^2) + 6\overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(\tilde{x}_1^2\tilde{x}_2^2) + \overline{\mathrm{E}}(\tilde{x}_1^4\tilde{x}_2^2) \tag{61}$$

$$\boldsymbol{r}(1) = \overline{\mathrm{E}}(\tilde{x}_2\tilde{\dot{x}}_2) \tag{62}$$

$$\boldsymbol{r}(2) = \alpha\overline{\mathrm{E}}(x_1^2) + \overline{\mathrm{E}}(\tilde{x}_1\tilde{\dot{x}}_2) \tag{63}$$

$$\boldsymbol{r}(3) = \beta\overline{\mathrm{E}}(x_1^2x_2^2) + \gamma\overline{\mathrm{E}}(x_1^4x_2^2) + 2\alpha\overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(\tilde{x}_1\tilde{x}_2) + \overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(\tilde{x}_2\tilde{\dot{x}}_2) + \overline{\mathrm{E}}(\tilde{x}_1^2\tilde{x}_2\tilde{\dot{x}}_2) \tag{64}$$

Hence, using (28) and Table I, $\tilde{\boldsymbol{\theta}}_n$ can be evaluated numerically for $\mathcal{A}1$ - $\mathcal{A}3$.

Asymptotic expressions for the random noise bias on the parameters β, α and γ (denoted as $\tilde{\beta}_n$, $\tilde{\alpha}_n$ and $\tilde{\gamma}_n$, respectively) can be evaluated using (28), Table I, and Eqs. (56)-(64). Straightforward calculations assuming high SNR and small sampling interval h for $\mathcal{A}3$, see Appendix I, give

$$\tilde{\beta}_n \approx \frac{1}{2h^2\mathrm{SNR}} \frac{\beta\overline{\mathrm{E}}(x_1^2)\Big(\overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_1^2x_2^2) - \overline{\mathrm{E}}(x_1^4x_2^2)\Big) + \gamma\overline{\mathrm{E}}(x_1^2)\Big(\overline{\mathrm{E}}(x_1^4)\overline{\mathrm{E}}(x_1^2x_2^2) - \overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_1^4x_2^2)\Big)}{\Big(\overline{\mathrm{E}}(x_2^2)\overline{\mathrm{E}}(x_1^4x_2^2) - [\overline{\mathrm{E}}(x_1^2x_2^2)]^2\Big)} \tag{65}$$

$$\tilde{\alpha}_n \approx \frac{-1}{2h^2\mathrm{SNR}} \tag{66}$$

$$\tilde{\gamma}_n \approx \frac{1}{2h^2\mathrm{SNR}} \frac{\beta\overline{\mathrm{E}}(x_1^2)\Big(\overline{\mathrm{E}}(x_1^2x_2^2) - \overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_2^2)\Big) + \gamma\overline{\mathrm{E}}(x_1^2)\Big(\overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_1^2x_2^2) - \overline{\mathrm{E}}(x_1^4)\overline{\mathrm{E}}(x_2^2)\Big)}{\Big(\overline{\mathrm{E}}(x_2^2)\overline{\mathrm{E}}(x_1^4x_2^2) - [\overline{\mathrm{E}}(x_1^2x_2^2)]^2\Big)} \tag{67}$$

Also, for $\beta = -\gamma = \varepsilon$, $\|\tilde{\boldsymbol{\theta}}_n\|$ can be approximated as

$$\|\tilde{\boldsymbol{\theta}}_n\| \approx \frac{\varepsilon}{2h^2\mathrm{SNR}}\sqrt{L^2 + Q^2 + \frac{1}{\varepsilon^2}} \tag{68}$$

where

$$L = \overline{\mathrm{E}}(x_1^2)\Big(\overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_1^2x_2^2) + \overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_1^4x_2^2) - \overline{\mathrm{E}}(x_1^4x_2^2) - \overline{\mathrm{E}}(x_1^4)\overline{\mathrm{E}}(x_1^2x_2^2)\Big) \tag{69}$$

$$Q = \overline{\mathrm{E}}(x_1^2)\Big(\overline{\mathrm{E}}(x_1^2x_2^2) + \overline{\mathrm{E}}(x_1^4)\overline{\mathrm{E}}(x_2^2) - \overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_2^2) - \overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_1^2x_2^2)\Big) \tag{70}$$

Hence

$$\|\tilde{\boldsymbol{\theta}}_n\| \propto \frac{1}{\mathrm{SNR}} \tag{71}$$

$$\|\tilde{\boldsymbol{\theta}}_n\| \propto \frac{1}{h^2} \tag{72}$$

B. Discretization errors

In this section the evaluation of discretization errors is considered. The data are assumed to be noise-free (*i.e.* $\widehat{x}_1(kh) = x_1(kh)$) and the estimates $\widehat{x}_2(kh)$ and $\widehat{\dot{x}}_2(kh)$ are chosen as one of $\mathcal{A}1$ - $\mathcal{A}3$.

The discretization error contributions to $\widehat{x}_2(kh)$ and $\widehat{\dot{x}}_2(kh)$ can be evaluated using Taylor series expansions assuming

$\mathcal{C}4$: The solution to the ODE model described by Eqs. (5)-(6) is sufficiently differentiable.

The discretization errors for $\mathcal{A}1$-$\mathcal{A}3$ can be summarized as follows, see Appendix II:

A1: (EB)

$$\begin{aligned}\tilde{x}_2(kh) &= -\frac{h}{2}D^2x_1(kh) + \frac{h^2}{6}D^3x_1(kh) - \frac{h^3}{24}D^4x_1(kh) + O(h^4)\\ \tilde{\dot{x}}_2(kh) &= -hD^2x_2(kh) + \frac{7h^2}{12}D^3x_2(kh) + O(h^3)\end{aligned} \tag{73}$$

A2: (EF)

$$\begin{aligned}\tilde{x}_2(kh) &= \frac{h}{2}D^2x_1(kh) + \frac{h^2}{6}D^3x_1(kh) + \frac{h^3}{24}D^4x_1(kh) + O(h^4)\\ \tilde{\dot{x}}_2(kh) &= hD^2x_2(kh) + \frac{7h^2}{12}D^3x_2(kh) + O(h^3)\end{aligned} \tag{74}$$

A3: (EC)

$$\begin{aligned}\tilde{x}_2(kh) &= \frac{h^2}{6}D^3x_1(kh) + \frac{h^4}{120}D^5x_1(kh) + O(h^6)\\ \tilde{\dot{x}}_2(kh) &= \frac{h^2}{3}D^3x_2(kh) + O(h^4)\end{aligned} \tag{75}$$

In the next two examples, discretization errors of $\mathcal{S}1$ and $\mathcal{S}2$ are evaluated.

Example 3. *Discretization errors of* $\mathcal{S}1$.

Here $\boldsymbol{\phi} = -x_1^3$. Thus $\boldsymbol{R} = \overline{\mathrm{E}}(x_1^6)$ and $\boldsymbol{r} = \overline{\mathrm{E}}(-x_1^3\widehat{\dot{x}}_2)$. Eq. (22) gives

$$\bar{\boldsymbol{\theta}} = \frac{\overline{\mathrm{E}}(-x_1^3\widehat{\dot{x}}_2)}{\overline{\mathrm{E}}(x_1^6)} = \underbrace{\frac{\overline{\mathrm{E}}(-x_1^3\dot{x}_2)}{\overline{\mathrm{E}}(x_1^6)}}_{\eta} + \underbrace{\frac{\overline{\mathrm{E}}(-x_1^3\tilde{\dot{x}}_2)}{\overline{\mathrm{E}}(x_1^6)}}_{\tilde{\boldsymbol{\theta}}_d} \tag{76}$$

Noticing that (*cf.* Remark 8)

$$\overline{\mathrm{E}}(x_1^3D^2x_2) = \overline{\mathrm{E}}(x_1^3D\dot{x}_2) = \overline{\mathrm{E}}(x_1^3D\left[-\eta x_1^3\right]) = -3\eta\overline{\mathrm{E}}(x_1^5\dot{x}_1) = 0 \tag{77}$$

and using Eqs. (73)-(75), it follows that

$$\tilde{\boldsymbol{\theta}}_d^{EB} = \tilde{\boldsymbol{\theta}}_d^{EF} = -\frac{7h^2}{12}\frac{\overline{\mathrm{E}}(x_1^3D^3x_2)}{\overline{\mathrm{E}}(x_1^6)} + O(h^3) \tag{78}$$

$$\tilde{\boldsymbol{\theta}}_d^{EC} = -\frac{h^2}{3}\frac{\overline{\mathrm{E}}(x_1^3D^3x_2)}{\overline{\mathrm{E}}(x_1^6)} + O(h^4) \tag{79}$$

We have also $x_1^3D^3x_2 = -6\eta x_1^4x_2^2 + 3\eta^2x_1^8$. Therefore Eqs. (78)-(79) can alternatively be written as

$$\tilde{\boldsymbol{\theta}}_d^{EB} = \tilde{\boldsymbol{\theta}}_d^{EF} = \frac{7h^2\eta}{4}\frac{\overline{\mathrm{E}}(2x_1^4x_2^2 - \eta x_1^8)}{\overline{\mathrm{E}}(x_1^6)} + O(h^3) \tag{80}$$

$$\tilde{\boldsymbol{\theta}}_d^{EC} = h^2\eta\frac{\overline{\mathrm{E}}(2x_1^4x_2^2 - \eta x_1^8)}{\overline{\mathrm{E}}(x_1^6)} + O(h^4) \tag{81}$$

The discretization errors of $\mathcal{S}1$ can be reduced to be of $O(h^3)$ by considering the general differentiator form given in (30), see Appendix III. Also, it is concluded from Eqs. (80)-(81) that

$$\tilde{\boldsymbol{\theta}}_d \propto h^2 \tag{82}$$

Now the total estimation bias $\tilde{\boldsymbol{\theta}}$ can be found for $\mathcal{S}1$ by the addition of Eqs. (80)-(81) and Eqs. (50)-(51) respectively. Thus we have

$$\tilde{\boldsymbol{\theta}}^{EB} = \tilde{\boldsymbol{\theta}}^{EF} \approx \frac{7h^2\eta}{4} \frac{\overline{\mathrm{E}}(2x_1^4x_2^2 - \eta x_1^8)}{\overline{\mathrm{E}}(x_1^6)} + \frac{-\frac{3}{h^2}[\overline{\mathrm{E}}(x_1^2)]^2 - 12\eta\overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_1^4)}{\mathrm{SNR}\ \overline{\mathrm{E}}(x_1^6)} \tag{83}$$

$$\tilde{\boldsymbol{\theta}}^{EC} \approx h^2\eta \frac{\overline{\mathrm{E}}(2x_1^4x_2^2 - \eta x_1^8)}{\overline{\mathrm{E}}(x_1^6)} + \frac{\frac{3}{2h^2}[\overline{\mathrm{E}}(x_1^2)]^2 - 12\eta\overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_1^4)}{\mathrm{SNR}\ \overline{\mathrm{E}}(x_1^6)} \tag{84}$$

Differentiating $|\tilde{\boldsymbol{\theta}}|^2$ using Eqs. (83)-(84) w.r.t. the sampling interval h to find the optimal sampling interval value that minimize the norm of the total bias $\tilde{\boldsymbol{\theta}}$ gives

$$h_{opt}^{EB} = h_{opt}^{EF} \approx \left| \frac{6[\overline{\mathrm{E}}(x_1^2)]^2}{7\eta\ \mathrm{SNR}\ \overline{\mathrm{E}}(2x_1^4x_2^2 - \eta x_1^8)} \right|^{\frac{1}{4}} \tag{85}$$

$$h_{opt}^{EC} \approx \left(\frac{6\overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_1^4)}{\mathrm{SNR}\ \overline{\mathrm{E}}(2x_1^4x_2^2 - \eta x_1^8)} \left(1 - \sqrt{1 - \frac{\mathrm{SNR}\ \overline{\mathrm{E}}(2x_1^4x_2^2 - \eta x_1^8)}{24\eta[\overline{\mathrm{E}}(x_1^4)]^2}} \right) \right)^{\frac{1}{2}} \tag{86}$$

Remark 9. Differentiating $|\tilde{\boldsymbol{\theta}}|^2$ for $\mathcal{S}1$ means solving $2\tilde{\boldsymbol{\theta}}\frac{d\tilde{\boldsymbol{\theta}}}{dh} = 0$. Since $\tilde{\boldsymbol{\theta}}^{EB} = 0$ does not have any real roots, expression (85) was found by solving $\frac{d\tilde{\boldsymbol{\theta}}^{EB}}{dh} = 0$. Expression (86) was found by solving $\tilde{\boldsymbol{\theta}}^{EC} = 0$ for positive real roots.

Example 4. *Discretization errors of* $\mathcal{S}2$.
In this case, we have $\boldsymbol{\phi}^T = (\widehat{x}_2 \quad x_1 \quad x_1^2\widehat{x}_2)$. Therefore

$$\boldsymbol{R} = \overline{\mathrm{E}}(\boldsymbol{\phi}\boldsymbol{\phi}^T) = \overline{\mathrm{E}} \begin{pmatrix} \widehat{x}_2^2 & x_1\widehat{x}_2 & x_1^2\widehat{x}_2^2 \\ x_1\widehat{x}_2 & x_1^2 & x_1^3\widehat{x}_2 \\ x_1^2\widehat{x}_2^2 & x_1^3\widehat{x}_2 & x_1^4\widehat{x}_2^2 \end{pmatrix} \tag{87}$$

and

$$\boldsymbol{r} = \overline{\mathrm{E}}(\boldsymbol{\phi}\widehat{\dot{x}}_2) = \overline{\mathrm{E}} \begin{pmatrix} \widehat{x}_2\widehat{\dot{x}}_2 \\ x_1\widehat{\dot{x}}_2 \\ x_1^2\widehat{x}_2\widehat{\dot{x}}_2 \end{pmatrix} \tag{88}$$

Now using Eqs. (73)-(75), discretization errors on each component of $\boldsymbol{R}$ and $\boldsymbol{r}$ can be computed. The results are given in Table II, see Appendix IV. Also, discretization errors can be evaluated as a function of α, β and γ by using $Dx_2 = \alpha x_1 + \beta x_2 + \gamma x_1^2 x_2$. The results in this case are given in Table III, see Appendix IV. Finally $\tilde{\boldsymbol{\theta}}_d$ can be evaluated numerically for different Euler approximations using (29).

In Appendix V, asymptotic expressions for the discretization bias on the parameters β, α and γ (denoted as $\tilde{\beta}_d$, $\tilde{\alpha}_d$ and $\tilde{\gamma}_d$, respectively) using $\mathcal{A}3$ are derived at high SNR and it is proved that

$$\|\tilde{\boldsymbol{\theta}}_d\| \propto h^2 \tag{89}$$

Now using Eqs. (65)-(67) and Eqs. (142)-(144) the norm of the total estimation bias is given by

$$\|\tilde{\boldsymbol{\theta}}\| \approx \sqrt{(\tilde{\beta}_n + \tilde{\beta}_d)^2 + (\tilde{\alpha}_n + \tilde{\alpha}_d)^2 + (\tilde{\gamma}_n + \tilde{\gamma}_d)^2} \tag{90}$$

TABLE II

Discretization errors of $\mathcal{S}2$.

Quantity	EB	EF	EC
$\tilde{\boldsymbol{R}}_d(1,1)$	$\frac{h^2}{3}\overline{\mathrm{E}}(x_2D^2x_2)+\frac{h^2}{4}\overline{\mathrm{E}}(Dx_2)^2+O(h^3)$	$\frac{h^2}{3}\overline{\mathrm{E}}(x_2D^2x_2)+\frac{h^2}{4}\overline{\mathrm{E}}(Dx_2)^2+O(h^3)$	$\frac{h^2}{3}\overline{\mathrm{E}}(x_2D^2x_2)+O(h^4)$
$\tilde{\boldsymbol{R}}_d(1,2)$	$-\frac{h}{2}\overline{\mathrm{E}}(x_1Dx_2)+O(h^2)$	$\frac{h}{2}\overline{\mathrm{E}}(x_1Dx_2)+O(h^2)$	$\frac{h^2}{6}\overline{\mathrm{E}}(x_1D^2x_2)+O(h^4)$
$\tilde{\boldsymbol{R}}_d(1,3)$	$-h\overline{\mathrm{E}}(x_1^2x_2Dx_2)+O(h^2)$	$h\overline{\mathrm{E}}(x_1^2x_2Dx_2)+O(h^2)$	$\frac{h^2}{3}\overline{\mathrm{E}}(x_1^2x_2D^2x_2)+O(h^4)$
$\tilde{\boldsymbol{R}}_d(2,2)$	0	0	0
$\tilde{\boldsymbol{R}}_d(2,3)$	$-\frac{h}{2}\overline{\mathrm{E}}(x_1^3Dx_2)+O(h^2)$	$\frac{h}{2}\overline{\mathrm{E}}(x_1^3Dx_2)+O(h^2)$	$\frac{h^2}{6}\overline{\mathrm{E}}(x_1^3D^2x_2)+O(h^4)$
$\tilde{\boldsymbol{R}}_d(3,3)$	$-h\overline{\mathrm{E}}(x_1^4x_2Dx_2)+O(h^2)$	$h\overline{\mathrm{E}}(x_1^4x_2Dx_2)+O(h^2)$	$\frac{h^2}{3}\overline{\mathrm{E}}(x_1^4x_2D^2x_2)+O(h^4)$
$\tilde{\boldsymbol{r}}_d(1)$	$-h\overline{\mathrm{E}}[x_2D^2x_2+\frac{1}{2}(Dx_2)^2]+O(h^2)$	$h\overline{\mathrm{E}}[x_2D^2x_2+\frac{1}{2}(Dx_2)^2]+O(h^2)$	$h^2\overline{\mathrm{E}}[\frac{1}{6}Dx_2D^2x_2+\frac{1}{3}x_2D^3x_2]+O(h^4)$
$\tilde{\boldsymbol{r}}_d(2)$	$-h\overline{\mathrm{E}}(x_1D^2x_2)+O(h^2)$	$h\overline{\mathrm{E}}(x_1D^2x_2)+O(h^2)$	$\frac{h^2}{3}\overline{\mathrm{E}}(x_1D^3x_2)+O(h^4)$
$\tilde{\boldsymbol{r}}_d(3)$	$-h\overline{\mathrm{E}}[x_1^2x_2D^2x_2+\frac{1}{2}x_1^2(Dx_2)^2]+O(h^2)$	$h\overline{\mathrm{E}}[x_1^2x_2D^2x_2+\frac{1}{2}x_1^2(Dx_2)^2]+O(h^2)$	$h^2\overline{\mathrm{E}}[\frac{1}{6}x_1^2Dx_2D^2x_2+\frac{1}{3}x_1^2x_2D^3x_2]+O(h^4)$

Differentiating $\|\tilde{\boldsymbol{\theta}}\|^2$ w.r.t. the sampling interval h gives (for $\beta=-\gamma=\varepsilon$)

$$h_{opt}^{EC}\approx\left|\frac{9}{(\mathrm{SNR})^2}\times\frac{\frac{|\boldsymbol{R}_0|^2}{[\overline{\mathrm{E}}(x_1^2)]^2}+\varepsilon^2[\overline{\mathrm{E}}(x_1^2)]^2(T_1^2+T_5^2)}{\frac{4|\boldsymbol{R}_0|^2}{[\overline{\mathrm{E}}(x_1^2)]^4}T_4^2+4(T_2^2+T_6^2)+\varepsilon^2(T_3^2+T_7^2)+4\varepsilon(T_2T_3+T_6T_7)}\right|^{\frac{1}{8}} \tag{91}$$

where

$$T_1=\overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_1^2x_2^2)-\overline{\mathrm{E}}(x_1^4x_2^2)-\overline{\mathrm{E}}(x_1^4)\overline{\mathrm{E}}(x_1^2x_2^2)+\overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_1^4x_2^2) \tag{92}$$

$$T_2=\overline{\mathrm{E}}(x_1^4x_2^2)\overline{\mathrm{E}}(x_2D^3x_2)-\overline{\mathrm{E}}(x_1^2x_2^2)\overline{\mathrm{E}}(x_1^2x_2D^3x_2) \tag{93}$$

$$T_3=\overline{\mathrm{E}}(x_1^2x_2^2)\overline{\mathrm{E}}(x_1^2x_2D^2x_2)-\overline{\mathrm{E}}(x_1^4x_2^2)\overline{\mathrm{E}}(x_2D^2x_2)-\overline{\mathrm{E}}(x_1^2x_2^2)\overline{\mathrm{E}}(x_1^4x_2D^2x_2)+\overline{\mathrm{E}}(x_1^4x_2^2)\overline{\mathrm{E}}(x_1^2x_2D^2x_2) \tag{94}$$

$$T_4=\overline{\mathrm{E}}(x_1D^3x_2)-\frac{\varepsilon}{2}\overline{\mathrm{E}}(x_1D^2x_2)+\frac{\varepsilon}{2}\overline{\mathrm{E}}(x_1^3D^2x_2) \tag{95}$$

$$T_5=\overline{\mathrm{E}}(x_1^2x_2^2)-\overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_2^2)-\overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_1^2x_2^2)+\overline{\mathrm{E}}(x_1^4)\overline{\mathrm{E}}(x_2^2) \tag{96}$$

$$T_6=\overline{\mathrm{E}}(x_2^2)\overline{\mathrm{E}}(x_1^2x_2D^3x_2)-\overline{\mathrm{E}}(x_1^2x_2^2)\overline{\mathrm{E}}(x_2D^3x_2) \tag{97}$$

$$T_7=\overline{\mathrm{E}}(x_1^2x_2^2)\overline{\mathrm{E}}(x_2D^2x_2)-\overline{\mathrm{E}}(x_2^2)\overline{\mathrm{E}}(x_1^2x_2D^2x_2)-\overline{\mathrm{E}}(x_1^2x_2^2)\overline{\mathrm{E}}(x_1^2x_2D^2x_2)+\overline{\mathrm{E}}(x_2^2)\overline{\mathrm{E}}(x_1^4x_2D^2x_2) \tag{98}$$

Similarly, analytical expressions for h_{opt}^{EB} and h_{opt}^{EF} can be derived by repeating the analysis of Appendix I and Appendix V for $\mathcal{A}1$ and $\mathcal{A}2$, respectively.

VI. Numerical Study

In this section a numerical study of the estimation errors of $\mathcal{S}1$ and $\mathcal{S}2$ using $\mathcal{A}1$ - $\mathcal{A}3$ is given. The study is based on numerical evaluation of the truncated analytical expressions derived in Section V. In the numerical calculation of the derived expressions, $\overline{\mathrm{E}}(f(t))$ was approximated by $\frac{1}{N}\sum_{t=1}^{N}f(t)$.

TABLE III

Discretization errors of $\mathcal{S}2$ as a function of the system parameters.

Quantity	EB	EF	EC
$\tilde{\boldsymbol{R}}_d(1,1)$	$h^2[(\frac{\alpha}{3}+\frac{7\beta^2}{12})\overline{\mathrm{E}}(x_2^2)+\frac{\alpha^2}{4}\overline{\mathrm{E}}(x_1^2)+\frac{7\beta\gamma}{6}\overline{\mathrm{E}}(x_1^2x_2^2)$ $+\frac{2\gamma}{3}\overline{\mathrm{E}}(x_1x_2^3)+\frac{7\gamma^2}{12}\overline{\mathrm{E}}(x_1^4x_2^2)]+O(h^3)$	$h^2[(\frac{\alpha}{3}+\frac{7\beta^2}{12})\overline{\mathrm{E}}(x_2^2)+\frac{\alpha^2}{4}\overline{\mathrm{E}}(x_1^2)+\frac{7\beta\gamma}{6}\overline{\mathrm{E}}(x_1^2x_2^2)$ $+\frac{2\gamma}{3}\overline{\mathrm{E}}(x_1x_2^3)+\frac{7\gamma^2}{12}\overline{\mathrm{E}}(x_1^4x_2^2)]+O(h^3)$	$\frac{h^2}{3}[(\alpha+\beta^2)\overline{\mathrm{E}}(x_2^2)+2\beta\gamma\overline{\mathrm{E}}(x_1^2x_2^2)$ $+2\gamma\overline{\mathrm{E}}(x_1x_2^3)+\gamma^2\overline{\mathrm{E}}(x_1^4x_2^2)]+O(h^4)$
$\tilde{\boldsymbol{R}}_d(1,2)$	$-\frac{h}{2}\alpha\overline{\mathrm{E}}(x_1^2)+O(h^2)$	$\frac{h}{2}\alpha\overline{\mathrm{E}}(x_1^2)+O(h^2)$	$\frac{h^2}{6}[\alpha\beta\overline{\mathrm{E}}(x_1^2)+2\gamma\overline{\mathrm{E}}(x_1^2x_2^2)+\alpha\gamma\overline{\mathrm{E}}(x_1^4)]+O(h^4)$
$\tilde{\boldsymbol{R}}_d(1,3)$	$-h[\beta\overline{\mathrm{E}}(x_1^2x_2^2)+\gamma\overline{\mathrm{E}}(x_1^4x_2^2)]+O(h^2)$	$h[\beta\overline{\mathrm{E}}(x_1^2x_2^2)+\gamma\overline{\mathrm{E}}(x_1^4x_2^2)]+O(h^2)$	$\frac{h^2}{3}[(\alpha+\beta^2)\overline{\mathrm{E}}(x_1^2x_2^2)+2\beta\gamma\overline{\mathrm{E}}(x_1^4x_2^2)$ $+2\gamma\overline{\mathrm{E}}(x_1^3x_2^3)+\gamma^2\overline{\mathrm{E}}(x_1^6x_2^2)]+O(h^4)$
$\tilde{\boldsymbol{R}}_d(2,2)$	0	0	0
$\tilde{\boldsymbol{R}}_d(2,3)$	$-\frac{h}{2}\alpha\overline{\mathrm{E}}(x_1^4)+O(h^2)$	$\frac{h}{2}\alpha\overline{\mathrm{E}}(x_1^4)+O(h^2)$	$\frac{h^2}{6}[\alpha\beta\overline{\mathrm{E}}(x_1^4)+2\gamma\overline{\mathrm{E}}(x_1^4x_2^2)+\alpha\gamma\overline{\mathrm{E}}(x_1^6)]+O(h^4)$
$\tilde{\boldsymbol{R}}_d(3,3)$	$-h[\beta\overline{\mathrm{E}}(x_1^4x_2^2)+\gamma\overline{\mathrm{E}}(x_1^6x_2^2)]+O(h^2)$	$h[\beta\overline{\mathrm{E}}(x_1^4x_2^2)+\gamma\overline{\mathrm{E}}(x_1^6x_2^2)]+O(h^2)$	$\frac{h^2}{3}[(\alpha+\beta^2)\overline{\mathrm{E}}(x_1^4x_2^2)+2\beta\gamma\overline{\mathrm{E}}(x_1^6x_2^2)$ $+2\gamma\overline{\mathrm{E}}(x_1^5x_2^3)+\gamma^2\overline{\mathrm{E}}(x_1^8x_2^2)]+O(h^4)$
$\tilde{\boldsymbol{r}}_d(1)$	$-h[(\alpha+\frac{3\beta^2}{2})\overline{\mathrm{E}}(x_2^2)+\frac{\alpha^2}{2}\overline{\mathrm{E}}(x_1^2)$ $+3\beta\gamma\overline{\mathrm{E}}(x_1^2x_2^2)+2\gamma\overline{\mathrm{E}}(x_1x_2^3)+\frac{3\gamma^2}{2}\overline{\mathrm{E}}(x_1^4x_2^2)]$ $+O(h^2)$	$h[(\alpha+\frac{3\beta^2}{2})\overline{\mathrm{E}}(x_2^2)+\frac{\alpha^2}{2}\overline{\mathrm{E}}(x_1^2)$ $+3\beta\gamma\overline{\mathrm{E}}(x_1^2x_2^2)+2\gamma\overline{\mathrm{E}}(x_1x_2^3)+\frac{3\gamma^2}{2}\overline{\mathrm{E}}(x_1^4x_2^2)]$ $+O(h^2)$	$h^2[\frac{\alpha^2\beta}{6}\overline{\mathrm{E}}(x_1^2)+\gamma(\frac{19\alpha}{6}+\frac{3\beta^2}{2})\overline{\mathrm{E}}(x_1^2x_2^2)+\frac{\alpha^2\gamma}{6}\overline{\mathrm{E}}(x_1^4)$ $+\beta(\frac{5\alpha}{6}+\frac{\beta^2}{2})\overline{\mathrm{E}}(x_2^2)+3\beta\gamma\overline{\mathrm{E}}(x_1x_2^3)+\frac{3\beta\gamma^2}{2}\overline{\mathrm{E}}(x_1^4x_2^2)$ $+3\gamma^2\overline{\mathrm{E}}(x_1^3x_2^3)+\frac{\gamma^3}{2}\overline{\mathrm{E}}(x_1^6x_2^2)+\frac{2\gamma}{3}\overline{\mathrm{E}}(x_2^4)]+O(h^4)$
$\tilde{\boldsymbol{r}}_d(2)$	$-h[\alpha\beta\overline{\mathrm{E}}(x_1^2)+2\gamma\overline{\mathrm{E}}(x_1^2x_2^2)+\alpha\gamma\overline{\mathrm{E}}(x_1^4)]$ $+O(h^2)$	$h[\alpha\beta\overline{\mathrm{E}}(x_1^2)+2\gamma\overline{\mathrm{E}}(x_1^2x_2^2)+\alpha\gamma\overline{\mathrm{E}}(x_1^4)]$ $+O(h^2)$	$\frac{h^2}{3}[\alpha(\alpha+\beta^2)\overline{\mathrm{E}}(x_1^2)+2\alpha\beta\gamma\overline{\mathrm{E}}(x_1^4)+8\beta\gamma\overline{\mathrm{E}}(x_1^2x_2^2)$ $+2\gamma\overline{\mathrm{E}}(x_1x_2^3)+8\gamma^2\overline{\mathrm{E}}(x_1^4x_2^2)+\alpha\gamma^2\overline{\mathrm{E}}(x_1^6)]+O(h^4)$
$\tilde{\boldsymbol{r}}_d(3)$	$-h[(\alpha+\frac{3\beta^2}{2})\overline{\mathrm{E}}(x_1^2x_2^2)+\frac{\alpha^2}{2}\overline{\mathrm{E}}(x_1^4)$ $+3\beta\gamma\overline{\mathrm{E}}(x_1^4x_2^2)+2\gamma\overline{\mathrm{E}}(x_1^3x_2^3)+\frac{3\gamma^2}{2}\overline{\mathrm{E}}(x_1^6x_2^2)]$ $+O(h^2)$	$h[(\alpha+\frac{3\beta^2}{2})\overline{\mathrm{E}}(x_1^2x_2^2)+\frac{\alpha^2}{2}\overline{\mathrm{E}}(x_1^4)$ $+3\beta\gamma\overline{\mathrm{E}}(x_1^4x_2^2)+2\gamma\overline{\mathrm{E}}(x_1^3x_2^3)+\frac{3\gamma^2}{2}\overline{\mathrm{E}}(x_1^6x_2^2)]$ $+O(h^2)$	$h^2[\frac{\alpha^2\beta}{6}\overline{\mathrm{E}}(x_1^4)+\gamma(\frac{19\alpha}{6}+\frac{3\beta^2}{2})\overline{\mathrm{E}}(x_1^4x_2^2)+\frac{\alpha^2\gamma}{6}\overline{\mathrm{E}}(x_1^6)$ $+\beta(\frac{5\alpha}{6}+\frac{\beta^2}{2})\overline{\mathrm{E}}(x_1^2x_2^2)+3\beta\gamma\overline{\mathrm{E}}(x_1^3x_2^3)+\frac{3\beta\gamma^2}{2}\overline{\mathrm{E}}(x_1^6x_2^2)$ $+3\gamma^2\overline{\mathrm{E}}(x_1^5x_2^3)+\frac{\gamma^3}{2}\overline{\mathrm{E}}(x_1^8x_2^2)+\frac{2\gamma}{3}\overline{\mathrm{E}}(x_1^2x_2^4)]+O(h^4)$

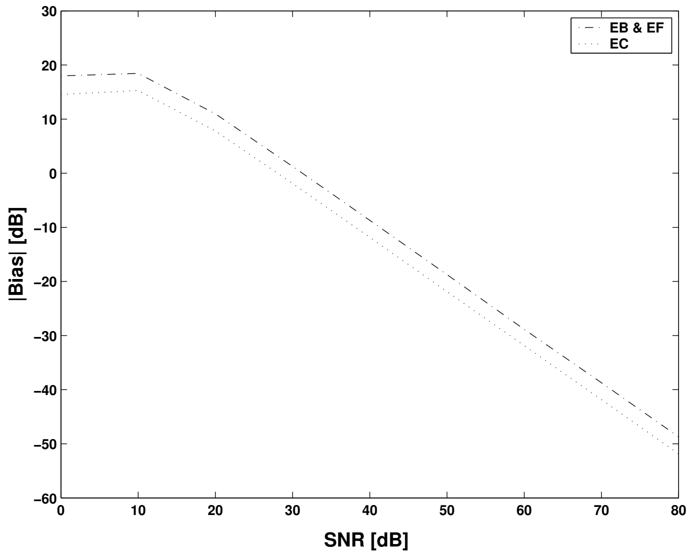

Fig. 1. $|\tilde{\boldsymbol{\theta}}_n|$ vs SNR for $\mathcal{S}1$ [$\eta = 2$, $h = 0.05$ s].

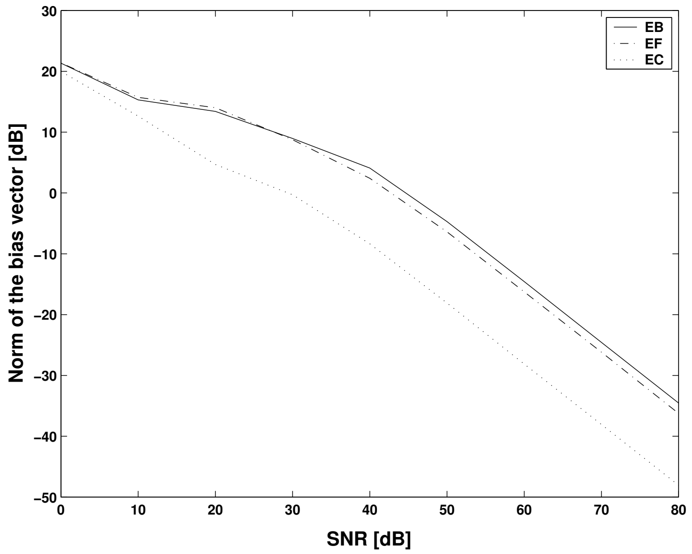

Fig. 2. $\|\tilde{\boldsymbol{\theta}}_n\|$ vs SNR for $\mathcal{S}2$ [$\alpha = -1$, $\beta = -\gamma = 2$, $h = 0.05$ s].

Example 5. *Random noise error study.*

In this example 10^4 data samples were generated from $\mathcal{S}1$ and $\mathcal{S}2$ using the Matlab routine `ode45` for $\eta = 2$, $\alpha = -1$, $\beta = 2$ and $\gamma = -2$. The initial states of $\mathcal{S}1$ and $\mathcal{S}2$ were selected as $\begin{pmatrix} x_1(0) & x_2(0) \end{pmatrix}^T = \begin{pmatrix} 0 & 1 \end{pmatrix}^T$ and $\begin{pmatrix} x_1(0) & x_2(0) \end{pmatrix}^T = \begin{pmatrix} 2 & 0 \end{pmatrix}^T$, respectively. This initial state of $\mathcal{S}2$ was chosen so that transient effects can be neglected. First, the bias error for these systems was studied at different signal to noise ratios (SNRs) with a sampling interval $h = 0.05$ s. Second, the bias was studied at different sampling intervals. In this case, the SNR was 50 dB for $\mathcal{S}1$ and 60 dB for $\mathcal{S}2$. Finally, the bias was studied for different values of the system parameters. The results are shown in Figures 1-6.

As it is shown in the log-log scale Figures 1-2, $\log \|\tilde{\boldsymbol{\theta}}_n\|$ using $\mathcal{A}1$-$\mathcal{A}3$ is proportional to $\big(-\log(\text{SNR}) + \text{constant}\big)$

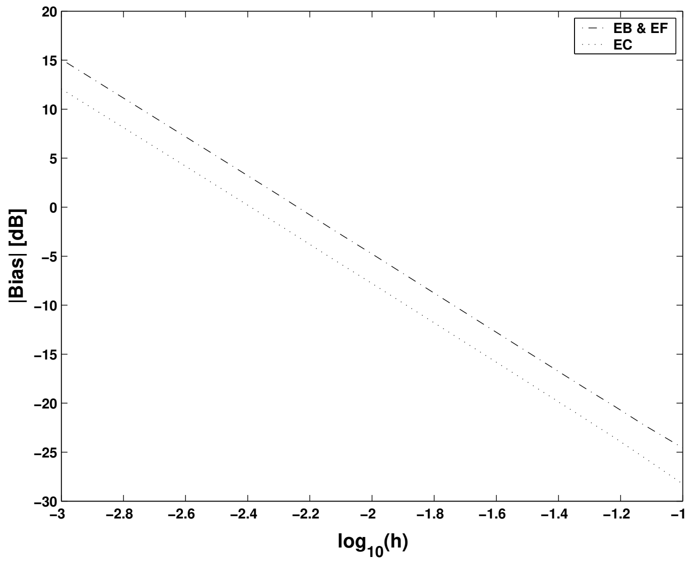

Fig. 3. $|\tilde{\boldsymbol{\theta}}_n|$ vs h for $\mathcal{S}1$ [$\eta = 2$, SNR $= 50$ dB].

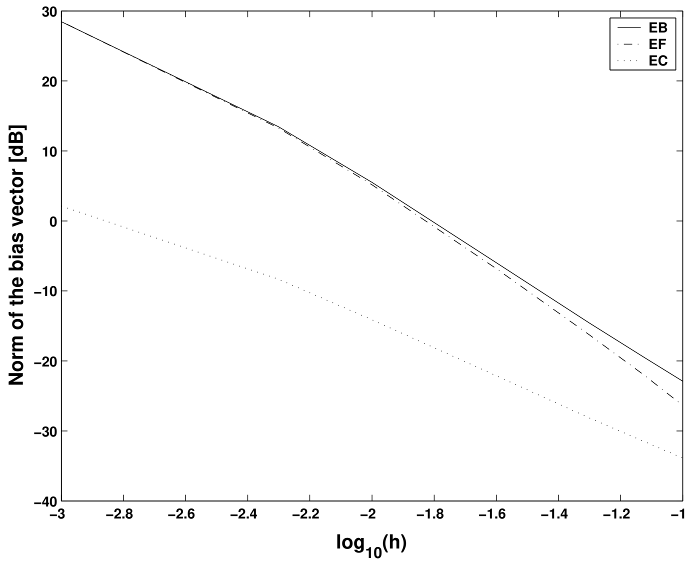

Fig. 4. $\|\tilde{\boldsymbol{\theta}}_n\|$ vs h for $\mathcal{S}2$ [$\alpha = -1$, $\beta = -\gamma = 2$, SNR $= 60$ dB].

(for moderate and high SNR). Also, we can conclude from the log-log scale Figures 3-4 that $\log \|\tilde{\boldsymbol{\theta}}_n\|$ is proportional to $\big(-\log(h) + \text{constant}\big)$. These results match the derived asymptotic results in Eqs. (52)-(53) and Eqs. (71)-(72), and our earlier expectation that random noise errors and its corresponding bias $\tilde{\boldsymbol{\theta}}_n$ can be reduced by increasing the sampling time h or using smaller h whenever the SNR is high. Also, it can be noticed that $\mathcal{A}3$ gives the lowest bias as expected.

Figures 5-6 show that the random noise bias increases as the value of η in $\mathcal{S}1$ and ε in $\mathcal{S}2$ ($\beta = \varepsilon$, $\gamma = -\varepsilon$) is increased. This is so because the nonlinear dynamics of the systems $\mathcal{S}1$ and $\mathcal{S}2$ become more effective as η and ε respectively increases. Hence, the signals generated by $\mathcal{S}1$ and $\mathcal{S}2$ become more nonlinearly distorted and the accuracy of the finite difference approximations decreases.

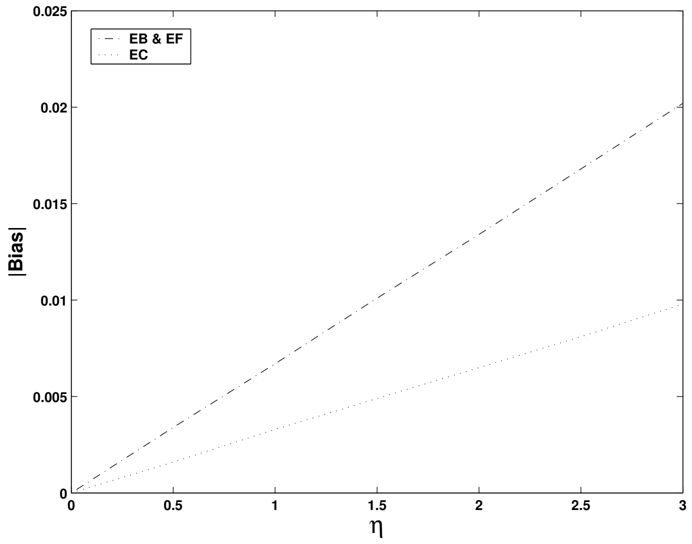

Fig. 5. $|\tilde{\boldsymbol{\theta}}_n|$ vs η for $\mathcal{S}1$ [$h = 0.05$ s, SNR = 50 dB].

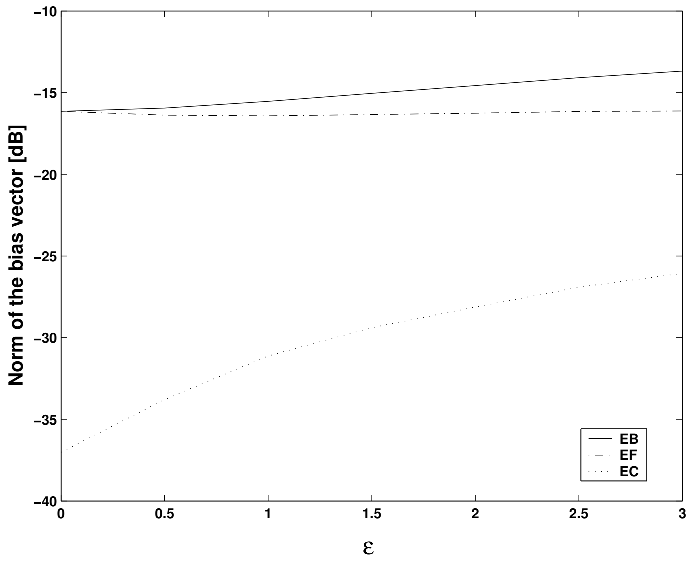

Fig. 6. $\|\tilde{\boldsymbol{\theta}}_n\|$ vs ε for $\mathcal{S}2$ [$\alpha = -1$, $\beta = -\gamma = \varepsilon$, $h = 0.05$, SNR = 60 dB].

Example 6. *Discretization error study.*

In this example, a noise-free data length of 10^4 samples was generated from the systems $\mathcal{S}1$ and $\mathcal{S}2$ as done in Example 5. The bias corresponding to discretization errors ($\tilde{\boldsymbol{\theta}}_d$) was evaluated using Eqs. (80)-(81) for $\mathcal{S}1$, and Table III and (29) for $\mathcal{S}2$. The results are plotted in the log-log scale Figures 7-8.

The results of Figures 7-8 show that $\log \|\tilde{\boldsymbol{\theta}}_d\|$ is proportional to $\big(\log(h) + \text{constant}\big)$, see Eqs. (82) and (89). This result also matches our earlier expectation that the discretization error increases as the sampling time is increased. It can be noticed also from Figures 7-8 that the slope of $\|\tilde{\boldsymbol{\theta}}_d^{EC}\|$ is lower than the slopes of $\|\tilde{\boldsymbol{\theta}}_d^{EB}\|$ and $\|\tilde{\boldsymbol{\theta}}_d^{EF}\|$, *i.e* the EC approximation ($\mathcal{A}3$) gives the lowest discretization bias.

Also, $\tilde{\boldsymbol{\theta}}_d$ can be evaluated for different system parameters. The results are shown in Figures 9-10. These

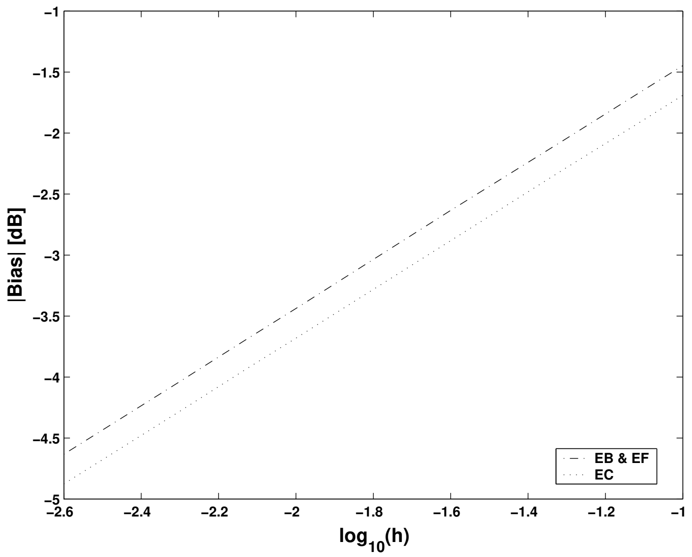

Fig. 7. $\tilde{\boldsymbol{\theta}}_d$ vs h for $\mathcal{S}1$ [$\eta = 2$].

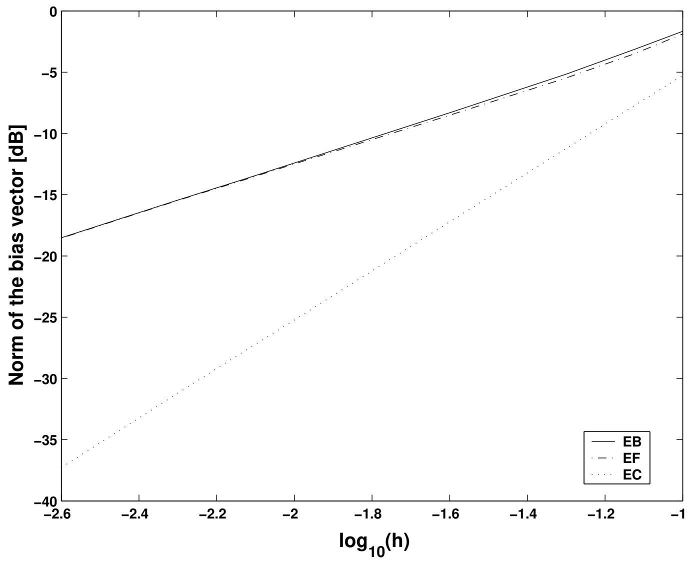

Fig. 8. $\|\tilde{\boldsymbol{\theta}}_d\|$ vs h for $\mathcal{S}2$ [$\alpha = -1$, $\beta = -\gamma = 2$].

figures show that $\|\tilde{\boldsymbol{\theta}}_d\|$ increases as the value of η in $\mathcal{S}1$ and ε in $\mathcal{S}2$ is increased. This behaviour can be explained here, as done in Example 5, by the accuracy reduction of the derivative approximations due to increasing the nonlinear dynamics of the two systems.

It can be concluded from Example 5 and Example 6 that there are two contradicting requirements to obtain an accurate estimate for the parameter vector $\boldsymbol{\theta}$ using the least squares estimation algorithm. A small sampling interval h is needed to reduce the bias contribution due to discretization errors $\tilde{\boldsymbol{\theta}}_d$, and a large h is required to reduce the bias contribution due to random noise errors $\tilde{\boldsymbol{\theta}}_n$. Therefore, it is expected that there is an optimal sampling interval h_{opt} that achieves minimum total estimation bias. This h_{opt} can be easily determined by adding the two bias contributions and plotting the results versus h as shown in Figures 11-12. Fig. 11 shows

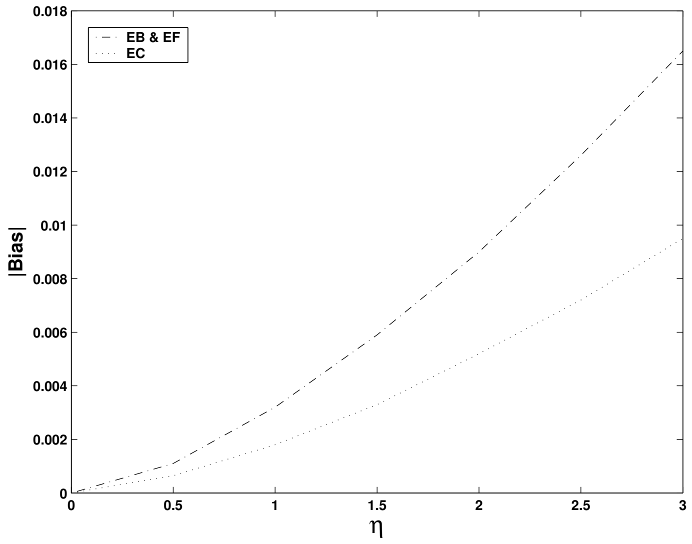

Fig. 9. $|\tilde{\boldsymbol{\theta}}_d|$ vs η for $\mathcal{S}1$ [$h = 0.05$ s].

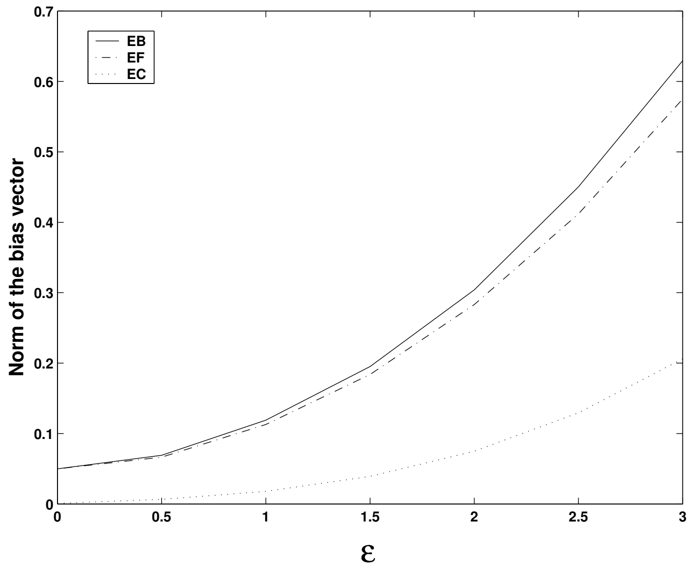

Fig. 10. $\|\tilde{\boldsymbol{\theta}}_d\|$ vs ε for $\mathcal{S}2$ [$\alpha = -1$, $\beta = -\gamma = \varepsilon$, $h = 0.05$ s].

that $h_{opt} = 0.052$ s gives minimum estimation bias for $\mathcal{S}1$. The evaluation of the approximate analytical values of h_{opt}^{EB} ($= h_{opt}^{EF}$) and h_{opt}^{EC}, see Eqs. (85)-(86), gives 0.0425 s and 0.0490 s, respectively. Fig. 12 shows that $h_{opt} = 0.03$ s for $\mathcal{S}2$ in case EB or EF approximations are used. On the other hand, if EC approximation is used, the optimal sampling interval is $h_{opt} = 0.02$ s. The expression of h_{opt}^{EC} in Eq. (91) gives 0.0190 s.

The difference between the value of h_{opt} concluded from Fig. 11 and the derived analytical expressions for $\mathcal{S}1$ is expected to be reduced as the SNR increases. This is due to the fact that the analytical expressions are derived for high SNR values. In order to verify this expectation, the total estimation bias was re-evaluted for SNR=70 dB. The results in this case are given in Fig. 13. Fig. 13 shows that $h_{opt} = 0.016$ s. The evaluation of the approximate analytical values of h_{opt}^{EB} and h_{opt}^{EC} in this case gives 0.0133 s and 0.0153 s, respectively.

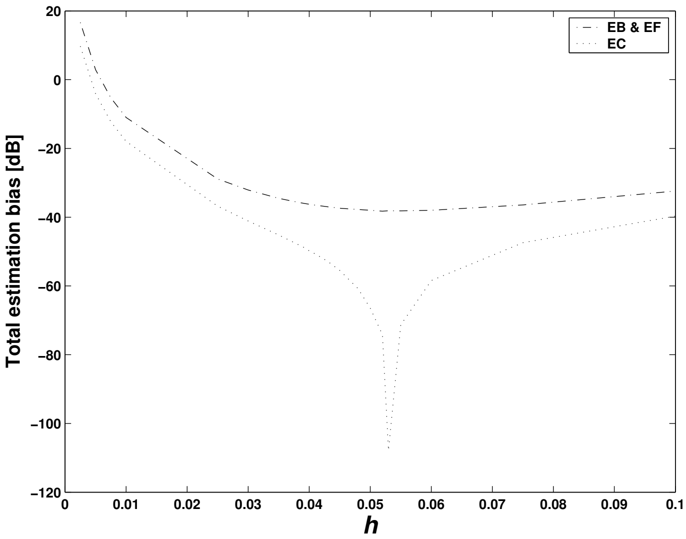

Fig. 11. $|\tilde{\boldsymbol{\theta}}|$ vs h for $\mathcal{S}1$ [$\eta = 2$, SNR=50 dB].

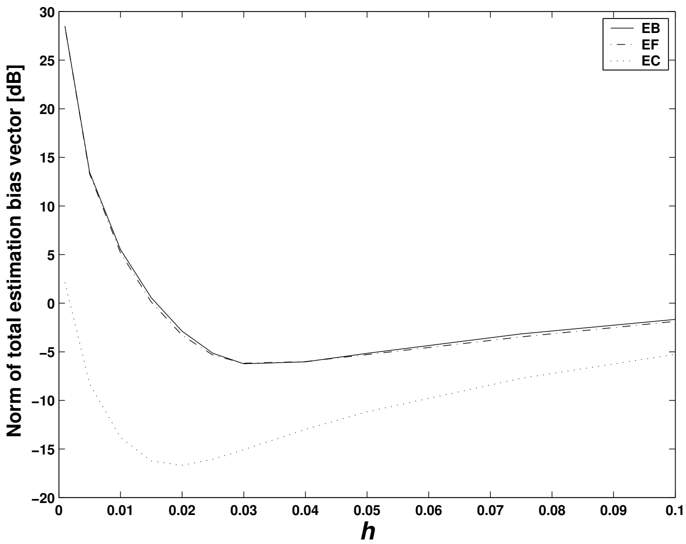

Fig. 12. $\|\tilde{\boldsymbol{\theta}}\|$ vs h for $\mathcal{S}2$ [$\alpha = -1$, $\beta = -\gamma = 2$, SNR=60 dB].

VII. Conclusions

Estimation errors in the least squares modeling of periodic signals using second-order nonlinear ODE have been studied in this paper. The estimation of the parameters of two nonlinear systems were considered using different Euler approximation techniques. The analysis of this paper shows that the sampling interval (h) plays a critical role in the performance of the LS estimation algorithm. Asymptotic analytical expressions show that

$$\|\tilde{\boldsymbol{\theta}}_n\| \propto \frac{1}{h^2\mathrm{SNR}} \quad \text{and} \quad \|\tilde{\boldsymbol{\theta}}_d\| \propto h^2$$

and that the estimation bias increases whenever the nonlinear dynamics become more effective. Hence, the existence of an optimal sampling interval h_{opt} that achieves the lowest estimation bias is expected. It is shown

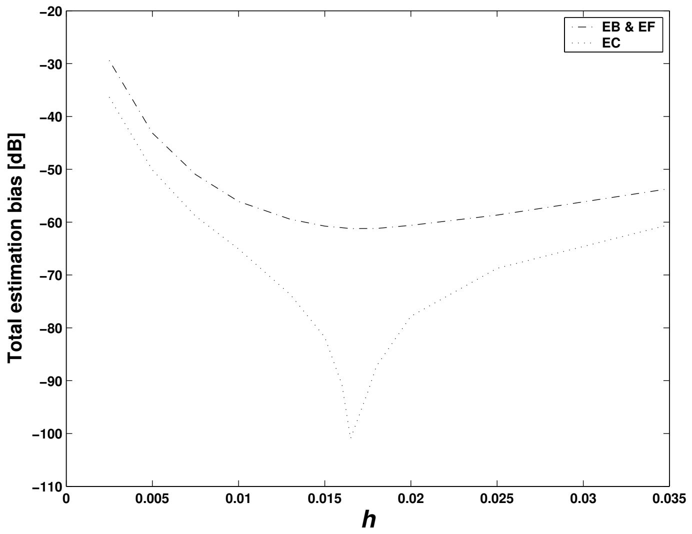

Fig. 13. $|\tilde{\boldsymbol{\theta}}|$ vs h for $\mathcal{S}1$ [$\eta = 2$, SNR=70 dB].

in the paper how an analytical expression for h_{opt} can be derived in a systematic way for different periodic nonlinear systems.

References

[1] P. Stoica and R. Moses, *Introduction to Spectral Analysis*, Prentice-Hall, Upper Saddle River, NJ, USA, 1997.

[2] E. Abd-Elrady, *Harmonic Signal Modeling Based on the Wiener Model Structure*, Licentiate thesis, Department of Systems and Control, Uppsala University, Uppsala, Sweden, 2002.

[3] E. Abd-Elrady, "A nonlinear approach to harmonic signal modeling," *Signal Processing*, vol. 84, no. 1, pp. 163–195, 2004.

[4] A. Nehorai and B. Porat, "Adaptive comb filtering for harmonic signal enhancement," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-34, pp. 1124–1138, 1986.

[5] T. Wigren and P. Händel, "Harmonic signal modeling using adaptive nonlinear function estimation," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, Atlanta, GA, May, 7–10 1996.

[6] D. W. Jordan and P. Smith, *Nonlinear Ordinary Differential Equations: An Introduction to Dynamical Systems*, Oxford University Press, Oxford, UK, third edition, 1999.

[7] H. K. Khalil, *Nonlinear Systems*, Prentice-Hall, Upper Saddle River, NJ, USA, third edition, 2002.

[8] L. Perko, *Differential Equations and Dynamical Systems*, Texts in Applied Mathematics. Springer-Verlag, NY, USA, 1991.

[9] S. H. Strogatz, *Nonlinear Dynamics and Chaos With Applications to Physics, Biology, Chemistry, and Engineering*, Addison-Wesley, 1994.

[10] S. Sastry, *Nonlinear Systems - Analysis, Stability, and Control*, Interdisciplinary Applied Mathematics. Springer-Verlag, NY, USA, 1999.

[11] T. Wigren, E. Abd-Elrady, and T. Söderström, "Harmonic signal analysis with Kalman filters using periodic orbits of nonlinear ODEs," in *Proc. of The IEEE International Conference on Acoustics, Speech, and Signal Processing*, Hong Kong, April, 6–10 2003, vol. VI, pp. 669–672.

[12] T. Wigren, E. Abd-Elrady, and T. Söderström, "Least squares harmonic signal analysis using periodic orbits of ODEs," in *Proc. of 13th IFAC Symposium on System Identification*, Rotterdam, The Netherlands, August, 27–29 2003, pp. 1584–1589.

[13] E. Abd-Elrady, T. Söderström, and T. Wigren, "Periodic signal analysis using orbits of nonlinear ODEs based on the Markov estimate," in *Proc. of 2nd IFAC Workshop on Periodic Control Systems (PSYCO 04)*, Yokohama, Japan, 2004.

[14] T. Wigren and T. Söderström, "Second order ODEs are sufficent for modeling of many periodic signals," Tech. Rep. 2003-025, Department of Information Technology, Uppsala University, Uppsala, Sweden, 2003.

[15] T. Söderström and P. Stoica, *System Identification*, Prentice-Hall International, Hemel Hempstead, United Kingdom, 1989.

[16] E. Abd-Elrady and J. Schoukens, "Least squares periodic signal modeling using orbits of nonlinear ODEs and fully automated spectral analysis," in *Proc. of 6th IFAC symposium on Nonlinear Control Systems (NOLCOS 04)*, Stuttgart, Germany, 2004.

[17] J. Schoukens, Y. Rolain, G. Simon, and R. Pintelon, "Fully automated spectral analysis of periodic signals," *IEEE Trans. on Instrumentation and Measurement*, vol. 52, no. 4, pp. 1021–1024, 2003.

[18] T. Söderström, H. Fan, B. Carlsson, and S. Bigi, "Least squares parameter estimation of continuous time ARX models from discrete-time data," *IEEE Trans. on Automatic Control*, vol. 42, no. 5, pp. 659–673, May 1997.

[19] T. Söderström and M. Mossberg, "Performance evaluation of methods for identifying continuous-time autoregressive processes," *Automatica*, vol. 36, no. 1, pp. 53–59, January 2000.

[20] B. Carlsson, *Digital Differentiating Filters and Model Based Fault Detection*, Ph.D. thesis, Department of Technology, Uppsala University, Uppsala, Sweden, 1989.

Appendix

I. Asymptotic random noise bias for $\mathcal{S}2$ using $\mathcal{A}3$

Using Table I and Eqs. (56)-(64), we have (neglecting λ^4 and λ^6 terms)

$$\boldsymbol{R}(1,1) = \overline{\mathrm{E}}(x_2^2) + \frac{\lambda^2}{2h^2} \tag{99}$$

$$\boldsymbol{R}(1,2) = 0 \tag{100}$$

$$\boldsymbol{R}(1,3) = \overline{\mathrm{E}}(x_1^2x_2^2) + \lambda^2\overline{\mathrm{E}}(x_2^2) + \frac{\lambda^2}{2h^2}\overline{\mathrm{E}}(x_1^2) \tag{101}$$

$$\boldsymbol{R}(2,2) = \overline{\mathrm{E}}(x_1^2) + \lambda^2 \tag{102}$$

$$\boldsymbol{R}(2,3) = 0 \tag{103}$$

$$\boldsymbol{R}(3,3) = \overline{\mathrm{E}}(x_1^4x_2^2) + 6\lambda^2\overline{\mathrm{E}}(x_1^2x_2^2) + \frac{\lambda^2}{2h^2}\overline{\mathrm{E}}(x_1^4) \tag{104}$$

$$\boldsymbol{r}(1) = 0 \tag{105}$$

$$\boldsymbol{r}(2) = \alpha\overline{\mathrm{E}}(x_1^2) - \frac{\lambda^2}{2h^2} \tag{106}$$

$$\boldsymbol{r}(3) = \beta\overline{\mathrm{E}}(x_1^2x_2^2) + \gamma\overline{\mathrm{E}}(x_1^4x_2^2) \tag{107}$$

Hence

$$\tilde{\boldsymbol{r}}_n = \lambda^2 \begin{pmatrix} 0 & -\frac{1}{2h^2} & 0 \end{pmatrix}^T + O(\lambda^4) \tag{108}$$

and

$$\tilde{\boldsymbol{R}}_n = \lambda^2 \begin{pmatrix} \frac{1}{2h^2} & 0 & \overline{\mathrm{E}}(x_2^2) + \frac{1}{2h^2}\overline{\mathrm{E}}(x_1^2) \\ 0 & 1 & 0 \\ \overline{\mathrm{E}}(x_2^2) + \frac{1}{2h^2}\overline{\mathrm{E}}(x_1^2) & 0 & 6\overline{\mathrm{E}}(x_1^2x_2^2) + \frac{1}{2h^2}\overline{\mathrm{E}}(x_1^4) \end{pmatrix} + O(\lambda^4) \tag{109}$$

Substituting by Eqs. (99)-(109) in Eq. (28) after evaluating $\boldsymbol{R}^{-1}$ gives

$$\tilde{\beta}_n \approx \frac{\beta\lambda^2\Big(\overline{\mathrm{E}}(x_2^2)\overline{\mathrm{E}}(x_1^2x_2^2) + \frac{1}{2h^2}\overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_1^2x_2^2) - \frac{1}{2h^2}\overline{\mathrm{E}}(x_1^4x_2^2)\Big)}{\mathcal{D}} + \frac{\gamma\lambda^2\Big(\frac{1}{2h^2}\overline{\mathrm{E}}(x_1^4)\overline{\mathrm{E}}(x_1^2x_2^2) + 6[\overline{\mathrm{E}}(x_1^2x_2^2)]^2 - \frac{1}{2h^2}\overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_1^4x_2^2) - \overline{\mathrm{E}}(x_2^2)\overline{\mathrm{E}}(x_1^4x_2^2)\Big)}{\mathcal{D}} + O(\lambda^4) \tag{110}$$

$$\tilde{\alpha}_n \approx \frac{-\lambda^2\left(\frac{1}{2h^2}+\alpha\right)}{\overline{\mathrm{E}}(x_1^2)} + O(\lambda^4) \tag{111}$$

$$\tilde{\gamma}_n \approx \frac{\beta\lambda^2\Big(\frac{1}{2h^2}\overline{\mathrm{E}}(x_1^2x_2^2) - [\overline{\mathrm{E}}(x_2^2)]^2 - \frac{1}{2h^2}\overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_2^2)\Big)}{\mathcal{D}} + \frac{\gamma\lambda^2\Big(\frac{1}{2h^2}\overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_1^2x_2^2) - 5\overline{\mathrm{E}}(x_2^2)\overline{\mathrm{E}}(x_1^2x_2^2) - \frac{1}{2h^2}\overline{\mathrm{E}}(x_1^4)\overline{\mathrm{E}}(x_2^2)\Big)}{\mathcal{D}} + O(\lambda^4) \tag{112}$$

where

$$\mathcal{D} = \overline{\mathrm{E}}(x_2^2)\overline{\mathrm{E}}(x_1^4x_2^2) - [\overline{\mathrm{E}}(x_1^2x_2^2)]^2 + O(\lambda^2) \tag{113}$$

Multiplying the numerators and denominators of Eqs. (110)-(112) by $2h^2\frac{\overline{\mathrm{E}}(x_1^2)}{\lambda^2}$ gives

$$\tilde{\beta}_n \approx \frac{\beta\overline{\mathrm{E}}(x_1^2)\Big(2h^2\overline{\mathrm{E}}(x_2^2)\overline{\mathrm{E}}(x_1^2x_2^2) + \overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_1^2x_2^2) - \overline{\mathrm{E}}(x_1^4x_2^2)\Big)}{\bar{\mathcal{D}}} + \frac{\gamma\overline{\mathrm{E}}(x_1^2)\Big(2h^2\left(6[\overline{\mathrm{E}}(x_1^2x_2^2)]^2 - \overline{\mathrm{E}}(x_2^2)\overline{\mathrm{E}}(x_1^4x_2^2)\right) + \overline{\mathrm{E}}(x_1^4)\overline{\mathrm{E}}(x_1^2x_2^2) - \overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_1^4x_2^2)\Big)}{\bar{\mathcal{D}}} \tag{114}$$

$$\tilde{\alpha}_n \approx \frac{-(1+2h^2\alpha)}{\bar{\mathcal{D}}}\times\Big(\overline{\mathrm{E}}(x_2^2)\overline{\mathrm{E}}(x_1^4x_2^2) - [\overline{\mathrm{E}}(x_1^2x_2^2)]^2\Big) \tag{115}$$

$$\tilde{\gamma}_n \approx \frac{\beta\overline{\mathrm{E}}(x_1^2)\Big(-2h^2[\overline{\mathrm{E}}(x_2^2)]^2 + \overline{\mathrm{E}}(x_1^2x_2^2) - \overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_2^2)\Big)}{\bar{\mathcal{D}}} + \frac{\gamma\overline{\mathrm{E}}(x_1^2)\Big(-10h^2\overline{\mathrm{E}}(x_2^2)\overline{\mathrm{E}}(x_1^2x_2^2) + \overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_1^2x_2^2) - \overline{\mathrm{E}}(x_1^4)\overline{\mathrm{E}}(x_2^2)\Big)}{\bar{\mathcal{D}}} \tag{116}$$

where

$$\bar{\mathcal{D}} \approx 2h^2\mathrm{SNR}\left(\overline{\mathrm{E}}(x_2^2)\overline{\mathrm{E}}(x_1^4x_2^2) - [\overline{\mathrm{E}}(x_1^2x_2^2)]^2\right) \tag{117}$$

Substituting by (117) in Eqs. (114)-(116) gives Eqs. (65)-(67).

II. Discretization error contributions to $\widehat{x}_2(kh)$ and $\widehat{\dot{x}}_2(kh)$

As an example, consider using the Euler backward approximation ($\mathcal{A}1$). We have

$$\widehat{x}_2(kh) = \frac{x_1(kh) - x_1(kh-h)}{h} \tag{118}$$

Performing a Taylor series expansion of $x_1(kh-h)$ around $x_1(kh)$, we obtain

$$x_1(kh-h) = x_1(kh) - hDx_1(kh) + \frac{h^2}{2}D^2x_1(kh) - \frac{h^3}{6}D^3x_1(kh) + \frac{h^4}{24}D^4x_1(kh) + O(h^5) \tag{119}$$

Substituting by (119) in (118), we have

$$\widehat{x}_2(kh) = \underbrace{Dx_1(kh)}_{x_2(kh)} \underbrace{-\frac{h}{2}D^2x_1(kh) + \frac{h^2}{6}D^3x_1(kh) - \frac{h^3}{24}D^4x_1(kh) + O(h^4)}_{\tilde{x}_2(kh)} \tag{120}$$

Similarly,

$$\begin{aligned} \widehat{\dot{x}}_2(kh) &= \frac{\widehat{x}_2(kh) - \widehat{x}_2(kh-h)}{h} \\ &= \frac{1}{h}[x_2(kh) - x_2(kh-h) - \frac{h}{2}D^2x_1(kh) + \frac{h}{2}D^2x_1(kh-h) + \frac{h^2}{6}D^3x_1(kh) \\ &\quad - \frac{h^2}{6}D^3x_1(kh-h) - \frac{h^3}{24}D^4x_1(kh) + \frac{h^3}{24}D^4x_1(kh-h) + O(h^4)] \end{aligned} \tag{121}$$

Thus performing the following Taylor series expansions

$$\begin{aligned} x_2(kh-h) &= x_2(kh) - hDx_2(kh) + \frac{h^2}{2}D^2x_2(kh) - \frac{h^3}{6}D^3x_2(kh) + O(h^4) \\ D^2x_1(kh-h) &= D^2x_1(kh) - hD^3x_1(kh) + \frac{h^2}{2}D^4x_1(kh) + O(h^3) \\ D^3x_1(kh-h) &= D^3x_1(kh) - hD^4x_1(kh) + O(h^2) \\ D^4x_1(kh-h) &= D^4x_1(kh) + O(h) \end{aligned}$$

we finally obtain

$$\widehat{\dot{x}}_2(kh) = \underbrace{Dx_2(kh)}_{\dot{x}_2(kh)} \underbrace{-hD^2x_2(kh) + \frac{7h^2}{12}D^3x_2(kh) + O(h^3)}_{\tilde{\dot{x}}_2(kh)} \tag{122}$$

This completes the derivation of Eq. (73). Similarly, $\widehat{x}_2(kh)$ and $\widehat{\dot{x}}_2(kh)$ can be evaluated for $\mathcal{A}2$ and $\mathcal{A}3$.

III. Discretization error of order $O(h^3)$ for $\mathcal{S}1$

For $\mathcal{S}1$, we have

$$\begin{aligned} \overline{\mathrm{E}}(x_1^6)\widehat{\eta} &= -\overline{\mathrm{E}}(\widehat{\ddot{x}}_1 x_1^3) \\ &= -\overline{\mathrm{E}}\left[\frac{1}{h^2}\sum_j \Gamma_{2,j} x_1(kh+jh) \times x_1^3(kh)\right] \\ &= -\frac{1}{h^2}\sum_j \Gamma_{2,j}\overline{\mathrm{E}}\left[\sum_{v=0}^{\infty}\frac{1}{v!}(jh)^v D^v x_1 \times x_1^3\right] \\ &= -\frac{1}{h^2}\sum_{v=0}^{\infty}\frac{1}{v!}h^v\left(\sum_j \Gamma_{2,j} j^v\right)\overline{\mathrm{E}}\left(D^v x_1 \times x_1^3\right) \end{aligned} \tag{123}$$

Now using (31) and noticing that $\overline{\mathrm{E}}(x_1^3 D^3 x_1) = \overline{\mathrm{E}}(x_1^3 D^2 x_2) = 0$ from (77), we get

$$\begin{aligned} \widehat{\eta} &= \eta - \frac{1}{\overline{\mathrm{E}}(x_1^6)} \sum_{v=4}^{\infty} \frac{1}{v!} h^{v-2} \left(\sum_j \Gamma_{2,j} j^v \right) \overline{\mathrm{E}} \left(D^v x_1 \times x_1^3 \right) \\ &= \eta + \underbrace{\tilde{\eta}_2 h^2 + \tilde{\eta}_3 h^3 + \cdots}_{\tilde{\eta}} \end{aligned} \tag{124}$$

Thus to have discretization errors of $O(h^3)$, $\tilde{\eta}_2$ should equal to 0. Since

$$\tilde{\eta}_2 = -\frac{1}{4!\overline{\mathrm{E}}(x_1^6)} \left(\sum_j \Gamma_{2,j} j^4 \right) \overline{\mathrm{E}} \left(D^4 x_1 \times x_1^3 \right) \tag{125}$$

$\tilde{\eta}_2 = 0$ if

$$\sum_j \Gamma_{2,j} j^4 = 0$$

One can design the differentiation filter so that $\sum_j \Gamma_{2,j} j^4 = 0$. For example, consider using a filter with 5 coefficients based on EF and EB approximations and including low pass filter effects with unity static gain, *i.e.*

$$H(q) = \frac{(q-\rho)}{(1-\rho)}(q-1)(1-q^{-1})\frac{(\rho - q^{-1})}{(\rho - 1)} \tag{126}$$

where ρ is a free parameter to be determined to achieve that $\sum_{j=-2}^{2} \Gamma_{2,j} j^4 = 0$. In this case we have

$$\begin{aligned} \Gamma_{2,0} &= -\frac{2(1+\rho+\rho^2)}{(1-\rho)^2} \\ \Gamma_{2,1} = \Gamma_{2,-1} &= \frac{(1+\rho)^2}{(1-\rho)^2} \\ \Gamma_{2,2} = \Gamma_{2,-2} &= -\frac{\rho}{(1-\rho)^2} \end{aligned} \tag{127}$$

Thus

$$\begin{aligned} \sum_{j=-2}^{2} \Gamma_{2,j} j^4 &= \frac{1}{(1-\rho)^2} \left[-2\rho(2)^4 + 2(1+\rho)^2(1)^4 - 2(1+\rho+\rho^2)(0)^4 \right] \\ &= \frac{2}{(1-\rho)^2} \left(\rho^2 - 14\rho + 1 \right) \end{aligned} \tag{128}$$

Thus $\sum_{j=-2}^{2} \Gamma_{2,j} j^4 = 0$ for $\rho = 7 \pm 4\sqrt{3}$. Therefore, choosing $\rho = 7 - 4\sqrt{3}$ or $\rho = 7 + 4\sqrt{3}$ gives

$$\begin{aligned} \tilde{\eta} &= \tilde{\eta}_3 h^3 + O(h^4) \\ &= -\frac{h^3 \overline{\mathrm{E}} \left(x_1^3 D^4 x_2 \right)}{5!(1-\rho)^2} \left[-2\rho(2)^5 + 2(1+\rho)^2 \right] + O(h^4) \\ &= -\frac{h^3}{15} \overline{\mathrm{E}} \left(x_1^3 D^4 x_2 \right) + O(h^4) \end{aligned} \tag{129}$$

IV. Results of Table II and Table III

Let us consider the first element in Table II, *i.e.* $\tilde{\boldsymbol{R}}_d(1,1)$ in the EB approximation. Using Eqs. (87) and (120), it follows that

$$\begin{aligned}\boldsymbol{R}(1,1) &= \overline{\mathrm{E}}\{\widehat{x}_2^2\} \\ &= \overline{\mathrm{E}}\{x_2 - \frac{h}{2}Dx_2 + \frac{h^2}{6}D^2x_2 + O(h^3)\}^2 \\ &= \underbrace{\overline{\mathrm{E}}(x_2^2)}_{\boldsymbol{R}_0(1,1)} \underbrace{-h\overline{\mathrm{E}}(x_2Dx_2) + \frac{h^2}{3}\overline{\mathrm{E}}(x_2D^2x_2) + \frac{h^2}{4}\overline{\mathrm{E}}(Dx_2)^2 + O(h^3)}_{\tilde{\boldsymbol{R}}_d(1,1)}\end{aligned} \tag{130}$$

Taking into account that $\overline{\mathrm{E}}(x_2Dx_2) = 0$, *cf.* Remark 8, it follows that

$$\tilde{\boldsymbol{R}}_d(1,1) = \frac{h^2}{3}\overline{\mathrm{E}}(x_2D^2x_2) + \frac{h^2}{4}\overline{\mathrm{E}}(Dx_2)^2 + O(h^3) \tag{131}$$

Similarly, the remaining elements of Table II can be systematically evaluated.

In Table III, $\tilde{\boldsymbol{R}}_d(1,1)$ can be also expressed as a function of the system parameters exploiting the fact that

$$Dx_2 = \alpha x_1 + \beta x_2 + \gamma x_1^2 x_2 \tag{132}$$

Hence

$$\begin{aligned}D^2x_2 &= D(\alpha x_1 + \beta x_2 + \gamma x_1^2 x_2) \\ &= \alpha x_2 + \beta Dx_2 + \gamma(2x_1x_2^2 + x_1^2Dx_2)\end{aligned} \tag{133}$$

Now reusing $Dx_2 = \alpha x_1 + \beta x_2 + \gamma x_1^2 x_2$ in (133) and substituting in (131), it follows that

$$\tilde{\boldsymbol{R}}_d(1,1) = h^2\left[\left(\frac{\alpha}{3} + \frac{7\beta^2}{12}\right)\overline{\mathrm{E}}(x_2^2) + \frac{\alpha^2}{4}\overline{\mathrm{E}}(x_1^2) + \frac{7\beta\gamma}{6}\overline{\mathrm{E}}(x_1^2x_2^2) + \frac{2\gamma}{3}\overline{\mathrm{E}}(x_1x_2^3) + \frac{7\gamma^2}{12}\overline{\mathrm{E}}(x_1^4x_2^2)\right] \tag{134}$$

Following the same procedure, the remaining terms of Table III can be evaluated.

V. Asymptotic discretization bias for $\mathcal{S}2$ using $\mathcal{A}3$

From Table II, we have

$$\tilde{\boldsymbol{r}}_d - \tilde{\boldsymbol{R}}_d\boldsymbol{\theta}_0 = h^2\begin{pmatrix}\frac{1}{3}\overline{\mathrm{E}}(x_2D^3x_2) - \frac{\beta}{6}\overline{\mathrm{E}}(x_2D^2x_2) - \frac{\gamma}{6}\overline{\mathrm{E}}(x_1^2x_2D^2x_2) \\ \frac{1}{3}\overline{\mathrm{E}}(x_1D^3x_2) - \frac{\beta}{6}\overline{\mathrm{E}}(x_1D^2x_2) - \frac{\gamma}{6}\overline{\mathrm{E}}(x_1^3D^2x_2) \\ \frac{1}{3}\overline{\mathrm{E}}(x_1^2x_2D^3x_2) - \frac{\beta}{6}\overline{\mathrm{E}}(x_1^2x_2D^2x_2) - \frac{\gamma}{6}\overline{\mathrm{E}}(x_1^4x_2D^2x_2)\end{pmatrix} + O(h^4) \tag{135}$$

$$\boldsymbol{R} = \begin{pmatrix}\overline{\mathrm{E}}(x_2^2) + \frac{h^2}{3}\overline{\mathrm{E}}(x_2D^2x_2) & \frac{h^2}{6}\overline{\mathrm{E}}(x_1D^2x_2) & \overline{\mathrm{E}}(x_1^2x_2^2) + \frac{h^2}{3}\overline{\mathrm{E}}(x_1^2x_2D^2x_2) \\ \frac{h^2}{6}\overline{\mathrm{E}}(x_1D^2x_2) & \overline{\mathrm{E}}(x_1^2) & \frac{h^2}{6}\overline{\mathrm{E}}(x_1^3D^2x_2) \\ \overline{\mathrm{E}}(x_1^2x_2^2) + \frac{h^2}{3}\overline{\mathrm{E}}(x_1^2x_2D^2x_2) & \frac{h^2}{6}\overline{\mathrm{E}}(x_1^3D^2x_2) & \overline{\mathrm{E}}(x_1^4x_2^2) + \frac{h^2}{3}\overline{\mathrm{E}}(x_1^4x_2D^2x_2)\end{pmatrix} + O(h^4) \tag{136}$$

Inverting $\boldsymbol{R}$ and substituting in (29) gives (neglecting h^4 terms)

$$\begin{aligned}\tilde{\beta}_d \approx \frac{h^2\overline{\mathrm{E}}(x_1^2)}{3|\boldsymbol{R}|}\Bigg[&\Big(\overline{\mathrm{E}}(x_1^4x_2^2)\overline{\mathrm{E}}(x_2D^3x_2) - \overline{\mathrm{E}}(x_1^2x_2^2)\overline{\mathrm{E}}(x_1^2x_2D^3x_2)\Big) + \frac{\beta}{2}\Big(\overline{\mathrm{E}}(x_1^2x_2^2)\overline{\mathrm{E}}(x_1^2x_2D^2x_2) - \overline{\mathrm{E}}(x_1^4x_2^2)\overline{\mathrm{E}}(x_2D^2x_2)\Big)\\ &+ \frac{\gamma}{2}\Big(\overline{\mathrm{E}}(x_1^2x_2^2)\overline{\mathrm{E}}(x_1^4x_2D^2x_2) - \overline{\mathrm{E}}(x_1^4x_2^2)\overline{\mathrm{E}}(x_1^2x_2D^2x_2)\Big)\Bigg] \end{aligned} \tag{137}$$

$$\tilde{\alpha}_d \approx \frac{h^2}{3|\boldsymbol{R}|}\Bigg[\Big(\overline{\mathrm{E}}(x_2^2)\overline{\mathrm{E}}(x_1^4x_2^2) - [\overline{\mathrm{E}}(x_1^2x_2^2)]^2\Big)\Big(\overline{\mathrm{E}}(x_1D^3x_2) - \frac{\beta}{2}\overline{\mathrm{E}}(x_1D^2x_2) - \frac{\gamma}{2}\overline{\mathrm{E}}(x_1^3D^2x_2)\Big)\Bigg] \tag{138}$$

$$\begin{aligned}\tilde{\gamma}_d \approx \frac{h^2\overline{\mathrm{E}}(x_1^2)}{3|\boldsymbol{R}|}\Bigg[&\Big(\overline{\mathrm{E}}(x_2^2)\overline{\mathrm{E}}(x_1^2x_2D^3x_2) - \overline{\mathrm{E}}(x_1^2x_2^2)\overline{\mathrm{E}}(x_2D^3x_2)\Big) + \frac{\beta}{2}\Big(\overline{\mathrm{E}}(x_1^2x_2^2)\overline{\mathrm{E}}(x_2D^2x_2) - \overline{\mathrm{E}}(x_2^2)\overline{\mathrm{E}}(x_1^2x_2D^2x_2)\Big)\\ &+ \frac{\gamma}{2}\Big(\overline{\mathrm{E}}(x_1^2x_2^2)\overline{\mathrm{E}}(x_1^2x_2D^2x_2) - \overline{\mathrm{E}}(x_2^2)\overline{\mathrm{E}}(x_1^4x_2D^2x_2)\Big)\Bigg] \end{aligned} \tag{139}$$

where

$$|\boldsymbol{R}| \approx \overline{\mathrm{E}}(x_1^2)\Big(\overline{\mathrm{E}}(x_2^2)\overline{\mathrm{E}}(x_1^4x_2^2) - [\overline{\mathrm{E}}(x_1^2x_2^2)]^2\Big) + O(h^2) \tag{140}$$

For small sampling interval, $|\boldsymbol{R}|$ can be approximated again to be

$$|\boldsymbol{R}| \approx \overline{\mathrm{E}}(x_1^2)\Big(\overline{\mathrm{E}}(x_2^2)\overline{\mathrm{E}}(x_1^4x_2^2) - [\overline{\mathrm{E}}(x_1^2x_2^2)]^2\Big) = |\boldsymbol{R}_0| \tag{141}$$

Substituting by (141) in Eqs. (137)-(139), we obtain

$$\begin{aligned}\tilde{\beta}_d \approx \frac{h^2\overline{\mathrm{E}}(x_1^2)}{3|\boldsymbol{R}_0|}\Bigg[&\Big(\overline{\mathrm{E}}(x_1^4x_2^2)\overline{\mathrm{E}}(x_2D^3x_2) - \overline{\mathrm{E}}(x_1^2x_2^2)\overline{\mathrm{E}}(x_1^2x_2D^3x_2)\Big)\\ &+ \frac{\beta}{2}\Big(\overline{\mathrm{E}}(x_1^2x_2^2)\overline{\mathrm{E}}(x_1^2x_2D^2x_2) - \overline{\mathrm{E}}(x_1^4x_2^2)\overline{\mathrm{E}}(x_2D^2x_2)\Big) + \frac{\gamma}{2}\Big(\overline{\mathrm{E}}(x_1^2x_2^2)\overline{\mathrm{E}}(x_1^4x_2D^2x_2) - \overline{\mathrm{E}}(x_1^4x_2^2)\overline{\mathrm{E}}(x_1^2x_2D^2x_2)\Big)\Bigg] \end{aligned} \tag{142}$$

$$\tilde{\alpha}_d \approx \frac{h^2}{3\overline{\mathrm{E}}(x_1^2)}\Big(\overline{\mathrm{E}}(x_1D^3x_2) - \frac{\beta}{2}\overline{\mathrm{E}}(x_1D^2x_2) - \frac{\gamma}{2}\overline{\mathrm{E}}(x_1^3D^2x_2)\Big) \tag{143}$$

$$\begin{aligned}\tilde{\gamma}_d \approx \frac{h^2\overline{\mathrm{E}}(x_1^2)}{3|\boldsymbol{R}_0|}\Bigg[&\Big(\overline{\mathrm{E}}(x_2^2)\overline{\mathrm{E}}(x_1^2x_2D^3x_2) - \overline{\mathrm{E}}(x_1^2x_2^2)\overline{\mathrm{E}}(x_2D^3x_2)\Big) + \frac{\beta}{2}\Big(\overline{\mathrm{E}}(x_1^2x_2^2)\overline{\mathrm{E}}(x_2D^2x_2) - \overline{\mathrm{E}}(x_2^2)\overline{\mathrm{E}}(x_1^2x_2D^2x_2)\Big)\\ &+ \frac{\gamma}{2}\Big(\overline{\mathrm{E}}(x_1^2x_2^2)\overline{\mathrm{E}}(x_1^2x_2D^2x_2) - \overline{\mathrm{E}}(x_2^2)\overline{\mathrm{E}}(x_1^4x_2D^2x_2)\Big)\Bigg] \end{aligned} \tag{144}$$

Hence, Eq. (89) is concluded from Eqs. (142)-(144).