BIAS ANALYSIS IN PERIODIC SIGNALS MODELING USING NONLINEAR ODE'S [1]

E. Abd-Elrady and T. Söderström

Division of Systems and Control, Dept. of Information Technology, Uppsala University, P.O. Box 337, SE-751 05 Uppsala, Sweden. Email: {ea,ts}@it.uu.se

Abstract: Second-order nonlinear ordinary differential equations (ODE's) can be used for modeling periodic signals. The right hand side function of the ODE model is parameterized in terms of polynomial basis functions. The least squares (LS) algorithm for estimating the coefficients of the polynomial basis gives biased estimates at low signal to noise ratios (SNRs). This is due to approximating the states of the ODE model using finite difference approximations from the noisy measurements. An analysis for this bias is given in this paper.

Keywords: Bias, Discretization, Identification, Least squares, Nonlinear systems.

1. INTRODUCTION

Modeling of periodic signals is a fundamental problem in many applications, see (Abd-Elrady, 2004; Abd-Elrady, 2005; Stoica and Moses, 1997). 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 and predator-prey systems, see (Khalil, 2002; Perko, 1991).

A second-order nonlinear ODE is used in this paper to model periodic signals. The periodic signal is modeled as a function of the states of the ODE. The right hand side of this ODE is parameterized using polynomial basis functions. The ODE approach is expected to obtain highly accurate models by estimating only a few parameters. Different estimators were developed in (Wigren *et al.*, 2003*a*; Wigren *et al.*, 2003*b*; Abd-Elrady *et al.*, 2004) using the same idea.

The LS estimator (Wigren *et al.*, 2003*b*) gives biased estimates especially at low SNRs. This is due to the fact that the derivatives of the modeled signal evaluated using Euler approximations are highly contaminated with noise at low SNRs.

A bias analysis for the LS estimate is given in this paper assuming that the periodic signal is contaminated with zero mean Gaussian white noise. The objective 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 2 introduces the ODE model. Section 3 reviews the LS algorithm given in (Wigren *et al.*, 2003*b*). Different estimation errors are discussed in Section 4. In Section 5, bias analysis and discretization errors for some simple systems are evaluated. Section 6 gives a comparative simulation study between three different Euler approximation techniques. Conclusions appear in Section 7.

2. THE ODE MODEL

2.1 Measurements

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

$$z(kh) = y(kh) + e(kh), \; k = 1, \cdots, N \qquad (1)$$

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

where $y(t)$ is the *periodic* continuous signal to be modeled, $y(kh)$ its sampled value, h the sampling interval and $e(kh)$ is zero mean Gaussian white noise, *i.e.*
$e(kh) \in N(0, \lambda^2)$, $\mathrm{E}[e(kh)e(kh+jh)] = \delta_{j,0}\lambda^2$.

2.2 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.*

$$\dot{\boldsymbol{x}} = f(\boldsymbol{x}, \boldsymbol{\theta}), \quad y = h(\boldsymbol{x}). \tag{2}$$

As shown in (Wigren and Söderström, 2003; Wigren *et al.*, 2003*b*), 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 $x_1 = y(t)$ and $x_2 = \dot{y}(t)$, the model given in (2) becomes

$$\begin{aligned} \begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} &= \begin{pmatrix} x_2(t) \\ f\big(x_1(t), x_2(t), \boldsymbol{\theta}\big) \end{pmatrix} \\ y(t) &= \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}. \end{aligned} \tag{4}$$

2.3 Parameterization

The right hand side of the second state equation of (4) is expanded in terms of known basis functions, modeling the right hand side as a truncated superposition of these functions. Polynomial basis functions were chosen in (Wigren *et al.*, 2003*b*) to model $f\big(x_1(t), x_2(t), \boldsymbol{\theta}\big)$, *i.e.*

$$f\big(x_1(t), x_2(t), \boldsymbol{\theta}\big) = \sum_{l=0}^{L} \sum_{m=0}^{M} \theta_{l,m} x_1^l(t) x_2^m(t) \tag{5}$$

$$\boldsymbol{\theta} = \begin{pmatrix} \theta_{0,0} \cdots \theta_{0,M} \cdots \theta_{L,0} \cdots \theta_{L,M} \end{pmatrix}^T. \tag{6}$$

3. THE LEAST SQUARES ALGORITHM

Now Eqs. (4)-(6) result in the model

$$\begin{aligned} \dot{x}_1(t) &= x_2(t) \\ \dot{x}_2(t) &= \boldsymbol{\phi}^T\big(x_1(t), x_2(t)\big)\ \boldsymbol{\theta} \end{aligned} \tag{7}$$

where

$$\begin{aligned} &\boldsymbol{\phi}^T\big(x_1(t), x_2(t)\big) = \\ &\big(1 \ \cdots \ x_2^M(t) \ \cdots \ x_1^L(t) \ \cdots \ x_1^L(t) x_2^M(t) \big). \end{aligned} \tag{8}$$

To estimate the parameter vector $\boldsymbol{\theta}$ from (7), 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 (7) results in (at $t = kh$)

$$\widehat{\dot{x}}_2(kh) = \boldsymbol{\phi}^T\big(\widehat{x}_1(kh), \widehat{x}_2(kh)\big)\ \boldsymbol{\theta} + \varepsilon(kh). \tag{9}$$

The expression (9) 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 (9) the combined regression error, $\varepsilon(kh)$, has been introduced.

The LS estimate of $\boldsymbol{\theta}$ has been studied in (Wigren *et al.*, 2003*b*) with $\widehat{x}_1(kh)$, $\widehat{x}_2(kh)$ and $\widehat{\dot{x}}_2(kh)$ evaluated using finite difference approximation. It was shown that the LS algorithm gives considerably accurate models at high SNRs and further research is needed to extend the operating region toward low SNRs.

4. ESTIMATION ERRORS

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 LS estimate behaves when the data length N becomes large, consider

$$\boldsymbol{R} = \overline{\mathrm{E}}\Big[\boldsymbol{\phi}\big(\widehat{x}_1(t), \widehat{x}_2(t)\big)\ \boldsymbol{\phi}^T\big(\widehat{x}_1(t), \widehat{x}_2(t)\big)\Big] \tag{10}$$

$$\boldsymbol{r} = \overline{\mathrm{E}}\Big[\boldsymbol{\phi}\big(\widehat{x}_1(t), \widehat{x}_2(t)\big)\ \widehat{\dot{x}}_2(t)\Big] \tag{11}$$

where

$$\overline{\mathrm{E}} f(t) = \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} \mathrm{E} f(t). \tag{12}$$

Remark 1. $\overline{\mathrm{E}}$ 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

$$\bar{\boldsymbol{\theta}} = \boldsymbol{R}^{-1}\ \boldsymbol{r} = \boldsymbol{\theta}_0 + \tilde{\boldsymbol{\theta}}_b \tag{13}$$

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

$$\boldsymbol{R} = \boldsymbol{R}_0 + \tilde{\boldsymbol{R}}_b \tag{14}$$

$$\boldsymbol{r} = \boldsymbol{r}_0 + \tilde{\boldsymbol{r}}_b \tag{15}$$

where

$$\begin{aligned} \boldsymbol{R}_0 &= \overline{\mathrm{E}}\Big[\boldsymbol{\phi}\big(x_1(t), x_2(t)\big)\ \boldsymbol{\phi}^T\big(x_1(t), x_2(t)\big)\Big] \\ \boldsymbol{r}_0 &= \overline{\mathrm{E}}\Big[\boldsymbol{\phi}\big(x_1(t), x_2(t)\big)\ \dot{x}_2(t)\Big] \end{aligned} \tag{16}$$

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 (13)-(15) gives

$$\begin{aligned} \bar{\boldsymbol{\theta}} &= (\boldsymbol{R}_0 + \tilde{\boldsymbol{R}}_b)^{-1}\ (\boldsymbol{r}_0 + \tilde{\boldsymbol{r}}_b) \\ &= \underbrace{\boldsymbol{R}_0^{-1}\boldsymbol{r}_0}_{\boldsymbol{\theta}_0} + \underbrace{(\boldsymbol{R}_0 + \tilde{\boldsymbol{R}}_b)^{-1}(\tilde{\boldsymbol{r}}_b - \tilde{\boldsymbol{R}}_b\boldsymbol{\theta}_o)}_{\tilde{\boldsymbol{\theta}}_b} \end{aligned} \tag{17}$$

Remark 2. The bias vector $\tilde{\boldsymbol{\theta}}_b$ is a contribution of random noise errors and discretization errors.

These contributions are denoted by $\tilde{\boldsymbol{\theta}}_n$ and $\tilde{\boldsymbol{\theta}}_d$, respectively. Hence

$$\tilde{\boldsymbol{\theta}}_b = \tilde{\boldsymbol{\theta}}_n + \tilde{\boldsymbol{\theta}}_d. \tag{18}$$

The bias vector $\tilde{\boldsymbol{\theta}}_b$ will depend on the sampling interval (h) and the derivative approximations. 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 $\dot{x}_2(kh)$ (in all approximations, we chosed $\widehat{x}_1(kh) = z(kh)$ and $i = 1, 2$):

- $\mathcal{A}1$: Euler backward approx. (EB)
$$\widehat{\dot{x}}_i(kh) = \big(\widehat{x}_i(kh) - \widehat{x}_i(kh-h)\big)/h. \tag{19}$$
- $\mathcal{A}2$: Euler forward approx. (EF)
$$\widehat{\dot{x}}_i(kh) = \big(\widehat{x}_i(kh+h) - \widehat{x}_i(kh)\big)/h. \tag{20}$$
- $\mathcal{A}3$: Euler center approx. (EC)
$$\widehat{\dot{x}}_i(kh) = \big(\widehat{x}_i(kh+h) - \widehat{x}_i(kh-h)\big)/(2h). \tag{21}$$

Remark 3. $\mathcal{A}1$-$\mathcal{A}3$ are chosen as examples for finite difference approximation. Many other different approximations can be considered. See (Söderström *et al.*, 1997) for more details.

It is well known from the numerical analysis literature that EC approximation gives lower discretization error compared to EB and EF approximations. Also, the discretization error is expected to decrease as h decreases. On the other hand, random noise error is expected to increase as h decreases since the noise will be highly amplified for small h. Small values of h will only be suitable when the SNR is high.

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_{opt}) that achieves the lowest bias for each system.

5. EXPLICIT ANALYSIS FOR SIMPLE SYSTEMS

Consider the following two nonlinear systems:

$\mathcal{S}1$:

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \begin{pmatrix} x_2 \\ -\eta x_1^3 \end{pmatrix}. \tag{22}$$

$\mathcal{S}2$:

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \begin{pmatrix} x_2 \\ \alpha x_1 + \beta x_2 + \gamma x_1^2 x_2 \end{pmatrix}. \tag{23}$$

Remark 4. 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.

5.1 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 Eq. (1) that $\tilde{x}_1(kh) = e(kh)$. Similarly, we can evaluate $\tilde{x}_2(kh)$ and $\tilde{\dot{x}}_2(kh)$ for Eqs. (19)-(21). Therefore, the noise contribution in $\widehat{x}_2(kh)$ and $\widehat{\dot{x}}_2(kh)$ are as follows:

- $\mathcal{A}1$: (EB)
$$\tilde{x}_2(kh) = \frac{e(kh) - e(kh-h)}{h}$$
$$\tilde{\dot{x}}_2(kh) = \frac{e(kh) - 2e(kh-h) + e(kh-2h)}{h^2}$$
- $\mathcal{A}2$: (EF)
$$\tilde{x}_2(kh) = \frac{e(kh+h) - e(kh)}{h}$$
$$\tilde{\dot{x}}_2(kh) = \frac{e(kh+2h) - 2e(kh+h) + e(kh)}{h^2}$$
- $\mathcal{A}3$: (EC)
$$\tilde{x}_2(kh) = \frac{e(kh+h) - e(kh-h)}{2h}$$
$$\tilde{\dot{x}}_2(kh) = \frac{e(kh+2h) - 2e(kh) + e(kh-2h)}{4h^2}$$

In the following two examples, the bias contribution due to random noise errors is analyzed.

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

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

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

Now, the bias $\tilde{\boldsymbol{\theta}}_n$ can be evaluated using Eq. (17) for $\mathcal{A}1$-$\mathcal{A}3$ by replacing $\overline{\mathrm{E}}(\tilde{x}_1\tilde{\dot{x}}_2)$ and $\overline{\mathrm{E}}(\tilde{x}_1^3\tilde{\dot{x}}_2)$ by their corresponding values. Straightforward calculations assuming high SNR, see (Abd-Elrady and Söderström, 2004), give

$$\begin{aligned} \tilde{\boldsymbol{\theta}}_n^{EB} &= \tilde{\boldsymbol{\theta}}_n^{EF} \\ &\approx \frac{-\frac{3}{h^2}[\overline{\mathrm{E}}(x_1^2)]^2 - 12\eta\overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_1^4)}{\mathrm{SNR}\ \overline{\mathrm{E}}(x_1^6)} \end{aligned} \tag{24}$$

$$\tilde{\boldsymbol{\theta}}_n^{EC} \approx \frac{\frac{3}{2h^2}[\overline{\mathrm{E}}(x_1^2)]^2 - 12\eta\overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_1^4)}{\mathrm{SNR}\ \overline{\mathrm{E}}(x_1^6)}. \tag{25}$$

Therefore, $\tilde{\boldsymbol{\theta}}_n$ satisfies the following relation:

$$\tilde{\boldsymbol{\theta}}_n \propto \frac{1}{h^2\mathrm{SNR}} \tag{26}$$

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

$$\boldsymbol{R} = \overline{\mathrm{E}}(\boldsymbol{\phi}\boldsymbol{\phi}^T) = \overline{\mathrm{E}} \begin{pmatrix} \widehat{x}_2^2 & \widehat{x}_1\widehat{x}_2 & \widehat{x}_1^2\widehat{x}_2^2 \\ \widehat{x}_1\widehat{x}_2 & \widehat{x}_1^2 & \widehat{x}_1^3\widehat{x}_2 \\ \widehat{x}_1^2\widehat{x}_2^2 & \widehat{x}_1^3\widehat{x}_2 & \widehat{x}_1^4\widehat{x}_2^2 \end{pmatrix}, \tag{27}$$

$$\boldsymbol{r} = \overline{\mathrm{E}}(\boldsymbol{\phi}\widehat{\dot{x}}_2) = \overline{\mathrm{E}} \begin{pmatrix} \widehat{x}_2\widehat{\dot{x}}_2 \\ \widehat{x}_1\widehat{\dot{x}}_2 \\ \widehat{x}_1^2\widehat{x}_2\widehat{\dot{x}}_2 \end{pmatrix}. \tag{28}$$

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 (for high SNR and small h^2) that, see (Abd-Elrady and Söderström, 2004),

$$\|\tilde{\boldsymbol{\theta}}_n\| \approx \frac{\varepsilon}{2h^2\mathrm{SNR}}\sqrt{\mathcal{L}^2 + \mathcal{Q}^2 + \frac{1}{\varepsilon^2}} \tag{29}$$

where

$$\begin{aligned}\mathcal{L} = \overline{\mathrm{E}}(x_1^2)\Big(&\overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_1^2x_2^2) + \overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_1^4x_2^2) \\ &- \overline{\mathrm{E}}(x_1^4x_2^2) - \overline{\mathrm{E}}(x_1^4)\overline{\mathrm{E}}(x_1^2x_2^2)\Big)\end{aligned} \tag{30}$$

$$\begin{aligned}\mathcal{Q} = \overline{\mathrm{E}}(x_1^2)\Big(&\overline{\mathrm{E}}(x_1^2x_2^2) + \overline{\mathrm{E}}(x_1^4)\overline{\mathrm{E}}(x_2^2) \\ &- \overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_2^2) - \overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_1^2x_2^2)\Big).\end{aligned} \tag{31}$$

Hence

$$\|\tilde{\boldsymbol{\theta}}_n\| \propto \frac{1}{h^2\mathrm{SNR}} \tag{32}$$

5.2 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 the solution to the ODE model described by Eq. (4) is *sufficiently* differentiable. The discretization errors for $\mathcal{A}1$-$\mathcal{A}3$ can be summarized as follows, see (Abd-Elrady and Söderström, 2004):

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

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

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

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

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

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

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. (13) gives

$$\bar{\boldsymbol{\theta}} = \frac{\overline{\mathrm{E}}(-x_1^3\widehat{\dot{x}}_2)}{\overline{\mathrm{E}}(x_1^6)} = \underbrace{\frac{\overline{\mathrm{E}}(-x_1^3\dot{x}_2)}{\overline{\mathrm{E}}(x_1^6)}}_{\eta} + \underbrace{\frac{\overline{\mathrm{E}}(-x_1^3\tilde{\dot{x}}_2)}{\overline{\mathrm{E}}(x_1^6)}}_{\tilde{\boldsymbol{\theta}}_d} \tag{33}$$

Taking into account that $x_1^3D^3x_2 = -6\eta x_1^4x_2^2 + 3\eta^2x_1^8$, straightforward calculations give

$$\tilde{\boldsymbol{\theta}}_d^{EB} = \tilde{\boldsymbol{\theta}}_d^{EF} \approx \frac{7h^2\eta}{4}\frac{\overline{\mathrm{E}}(2x_1^4x_2^2 - \eta x_1^8)}{\overline{\mathrm{E}}(x_1^6)} \tag{34}$$

$$\tilde{\boldsymbol{\theta}}_d^{EC} \approx h^2\eta\frac{\overline{\mathrm{E}}(2x_1^4x_2^2 - \eta x_1^8)}{\overline{\mathrm{E}}(x_1^6)}. \tag{35}$$

It can be concluded from Eqs. (34)-(35) that

$$\tilde{\boldsymbol{\theta}}_d \propto h^2 \tag{36}$$

Now the total estimation bias $\tilde{\boldsymbol{\theta}}$ can be found for $\mathcal{S}1$ by the addition of Eqs. (34)-(35) and Eqs. (24)-(25), respectively. Differentiating $|\tilde{\boldsymbol{\theta}}|$ w.r.t. h to find the optimal sampling interval value that minimize the total bias $\tilde{\boldsymbol{\theta}}$ gives, see (Abd-Elrady and Söderström, 2004) for details,

$$h_{opt}^{EB} = h_{opt}^{EF} \approx \left|\frac{6[\overline{\mathrm{E}}(x_1^2)]^2}{7\eta\ \mathrm{SNR}\ \overline{\mathrm{E}}(2x_1^4x_2^2 - \eta x_1^8)}\right|^{\frac{1}{4}} \tag{37}$$

$$\begin{aligned}h_{opt}^{EC} \approx \Bigg(&\frac{6\overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_1^4)}{\mathrm{SNR}\ \overline{\mathrm{E}}(2x_1^4x_2^2 - \eta x_1^8)} \times \\ &\Bigg(1 - \sqrt{1 - \frac{\mathrm{SNR}\ \overline{\mathrm{E}}(2x_1^4x_2^2 - \eta x_1^8)}{24\eta[\overline{\mathrm{E}}(x_1^4)]^2}}\Bigg)\Bigg)^{\frac{1}{2}}\end{aligned} \tag{38}$$

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

$$\boldsymbol{R} = \overline{\mathrm{E}}(\boldsymbol{\phi}\boldsymbol{\phi}^T) = \overline{\mathrm{E}} \begin{pmatrix} \widehat{x}_2^2 & x_1\widehat{x}_2 & x_1^2\widehat{x}_2^2 \\ x_1\widehat{x}_2 & x_1^2 & x_1^3\widehat{x}_2 \\ x_1^2\widehat{x}_2^2 & x_1^3\widehat{x}_2 & x_1^4\widehat{x}_2^2 \end{pmatrix}, \tag{39}$$

$$\boldsymbol{r} = \overline{\mathrm{E}}(\boldsymbol{\phi}\widehat{\dot{x}}_2) = \overline{\mathrm{E}} \begin{pmatrix} \widehat{x}_2\widehat{\dot{x}}_2 \\ x_1\widehat{\dot{x}}_2 \\ x_1^2\widehat{x}_2\widehat{\dot{x}}_2 \end{pmatrix}. \tag{40}$$

Straightforward calculations in (Abd-Elrady and Söderström, 2004) for $\mathcal{A}3$ at high SNR show that

$$\|\tilde{\boldsymbol{\theta}}_d\| \propto h^2 \tag{41}$$

and the analytic expression of h_{opt}^{EC} is given by Eq. (42).

$$h_{opt}^{EC} \approx \left| \frac{9}{(\mathrm{SNR})^2} \times \frac{\frac{|\boldsymbol{R}_0|^2}{[\overline{\mathrm{E}}(x_1^2)]^2} + \varepsilon^2 [\overline{\mathrm{E}}(x_1^2)]^2 (T_1^2 + T_5^2)}{\frac{4|\boldsymbol{R}_0|^2}{[\overline{\mathrm{E}}(x_1^2)]^4} T_4^2 + 4(T_2^2 + T_6^2) + \varepsilon^2 (T_3^2 + T_7^2) + 4\varepsilon (T_2 T_3 + T_6 T_7)} \right|^{\frac{1}{8}} \tag{42}$$

where

$$\begin{aligned}
T_1 &= \overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_1^2 x_2^2) - \overline{\mathrm{E}}(x_1^4 x_2^2) - \overline{\mathrm{E}}(x_1^4)\overline{\mathrm{E}}(x_1^2 x_2^2) + \overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_1^4 x_2^2)\\
T_2 &= \overline{\mathrm{E}}(x_1^4 x_2^2)\overline{\mathrm{E}}(x_2 D^3 x_2) - \overline{\mathrm{E}}(x_1^2 x_2^2)\overline{\mathrm{E}}(x_1^2 x_2 D^3 x_2)\\
T_3 &= \overline{\mathrm{E}}(x_1^2 x_2^2)\overline{\mathrm{E}}(x_1^2 x_2 D^2 x_2) - \overline{\mathrm{E}}(x_1^4 x_2^2)\overline{\mathrm{E}}(x_2 D^2 x_2) - \overline{\mathrm{E}}(x_1^2 x_2^2)\overline{\mathrm{E}}(x_1^4 x_2 D^2 x_2) + \overline{\mathrm{E}}(x_1^4 x_2^2)\overline{\mathrm{E}}(x_1^2 x_2 D^2 x_2)\\
T_4 &= \overline{\mathrm{E}}(x_1 D^3 x_2) - \frac{\varepsilon}{2}\overline{\mathrm{E}}(x_1 D^2 x_2) + \frac{\varepsilon}{2}\overline{\mathrm{E}}(x_1^3 D^2 x_2)\\
T_5 &= \overline{\mathrm{E}}(x_1^2 x_2^2) - \overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_2^2) - \overline{\mathrm{E}}(x_1^2)\overline{\mathrm{E}}(x_1^2 x_2^2) + \overline{\mathrm{E}}(x_1^4)\overline{\mathrm{E}}(x_2^2)\\
T_6 &= \overline{\mathrm{E}}(x_2^2)\overline{\mathrm{E}}(x_1^2 x_2 D^3 x_2) - \overline{\mathrm{E}}(x_1^2 x_2^2)\overline{\mathrm{E}}(x_2 D^3 x_2)\\
T_7 &= \overline{\mathrm{E}}(x_1^2 x_2^2)\overline{\mathrm{E}}(x_2 D^2 x_2) - \overline{\mathrm{E}}(x_2^2)\overline{\mathrm{E}}(x_1^2 x_2 D^2 x_2) - \overline{\mathrm{E}}(x_1^2 x_2^2)\overline{\mathrm{E}}(x_1^2 x_2 D^2 x_2) + \overline{\mathrm{E}}(x_2^2)\overline{\mathrm{E}}(x_1^4 x_2 D^2 x_2)
\end{aligned}$$

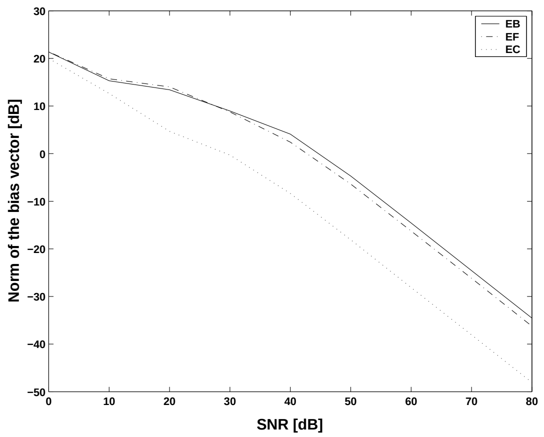

Fig. 1. $\|\tilde{\boldsymbol{\theta}}_n\|$ vs SNR.

6. NUMERICAL STUDY

In this section a numerical study of the estimation errors of $\mathcal{S}2$ using $\mathcal{A}1$-$\mathcal{A}3$ is given. The study is based on numerical evaluation of the truncated analytical expressions derived in Section 5. A numerical study of $\mathcal{S}1$ can be found in (Abd-Elrady and Söderström, 2004). In the numerical calculation of the derived expressions, $\overline{\mathrm{E}}(f(t))$ was approximated by $\frac{1}{N}\sum_{t=1}^{N} f(t)$.

Example 5. *Random noise error study.*

In this example 10^4 samples were generated from $\mathcal{S}2$ using the Matlab routine `ode45` for $\alpha = -1$, $\beta = 2$ and $\gamma = -2$. The initial state of $\mathcal{S}2$ was selected as $\begin{pmatrix} x_1(0) & x_2(0) \end{pmatrix}^T = \begin{pmatrix} 2 & 0 \end{pmatrix}^T$. First, the bias error was studied at different SNRs with $h = 0.05$ s. Second, the bias was studied at different sampling intervals with SNR of 60 dB. The results are shown in the log-log scale Figs. 1-2.

As it is shown in Fig. 1, $\log\|\tilde{\boldsymbol{\theta}}_n\|$ using $\mathcal{A}1$-$\mathcal{A}3$ is proportional to $\big(-\log(\mathrm{SNR}) + \mathrm{constant}\big)$ (for moderate and high SNR). Also, we can conclude from Fig. 2 that $\log\|\tilde{\boldsymbol{\theta}}_n\|$ is proportional to $\big(-\log(h) + \mathrm{constant}\big)$. These results match the derived asymptotic results in Eq. (32), and our earlier expectation that random noise errors 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.

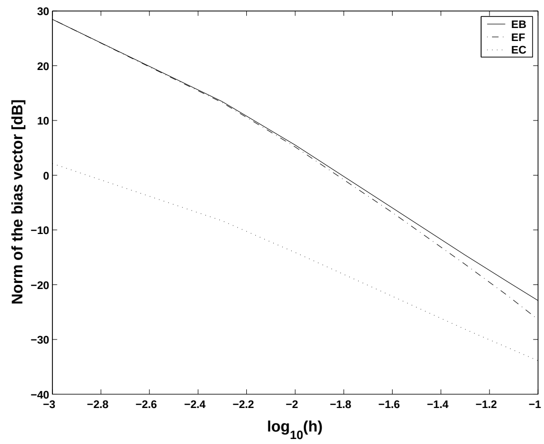

Fig. 2. $\|\tilde{\boldsymbol{\theta}}_n\|$ vs h.

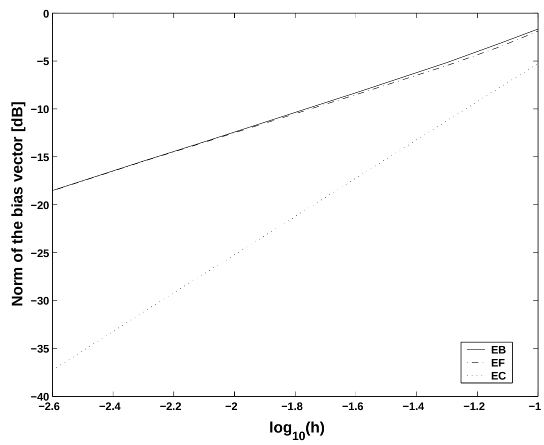

Fig. 3. $\|\tilde{\boldsymbol{\theta}}_d\|$ vs h.

Example 6. *Discretization error study.*

In this example, a noise-free data length of 10^4 samples was generated from system $\mathcal{S}2$ as done in Example 5. The bias corresponding to discretization errors ($\tilde{\boldsymbol{\theta}}_d$) was evaluated. The results are plotted in the log-log scale Fig. 3.

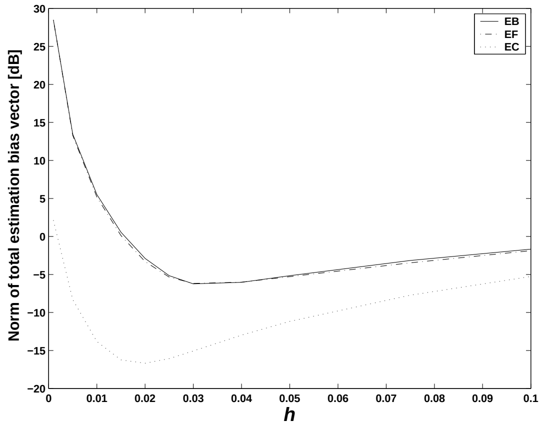

Fig. 4. $\|\tilde{\boldsymbol{\theta}}\|$ vs h.

The results of Fig. 3 show that $\log\|\tilde{\boldsymbol{\theta}}_d\|$ is proportional to $\big(\log(h)+\text{constant}\big)$, see Eq. (41). This result also matches our earlier expectation that the discretization error increases as the sampling time is increased. It can be noticed also from Fig. 3 that $\mathcal{A}3$ gives the lowest discretization bias.

It can be concluded from Examples 5 and 6 that there are two contradicting requirements to obtain an accurate estimate for the parameter vector $\boldsymbol{\theta}$ using the LS algorithm. A small h is needed to reduce $\tilde{\boldsymbol{\theta}}_d$, and a large h is required to reduce $\tilde{\boldsymbol{\theta}}_n$. Therefore, it is expected that there is an optimal sampling interval h_{opt} that achieves minimal 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 Fig. 4. Figure 4 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, is $h_{opt} = 0.02$ s. The expression of h^{EC}_{opt} in Eq. (42) gives 0.0190 s.

Also in (Abd-Elrady and Söderström, 2004), $\tilde{\boldsymbol{\theta}}_n$ and $\tilde{\boldsymbol{\theta}}_d$ were studied with the nonlinear system parameters η, α, β and γ. The results show that $\|\tilde{\boldsymbol{\theta}}_n\|$ and $\|\tilde{\boldsymbol{\theta}}_d\|$ increases as the value of these parameters is increased. This is so because the nonlinear dynamics of the systems $\mathcal{S}1$ and $\mathcal{S}2$ become more effective as these parameters increase. Then the signals generated by $\mathcal{S}1$ and $\mathcal{S}2$ become more nonlinearly distorted and the accuracy of the finite difference approximations decreases.

7. CONCLUSIONS

In this paper, estimation errors in least squares estimation of periodic signals using second-order nonlinear ODE model have been studied. The study has considered the estimation of two nonlinear systems using Euler approximations for the derivatives of the modeled signal. The bias analysis shows that

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

It is shown in the paper how an analytical expression for an optimal sampling interval h_{opt} that achieves the lowest estimation bias can be derived in a systematic way for different periodic nonlinear systems.

REFERENCES

Abd-Elrady, E. (2004). A nonlinear approach to harmonic signal modeling. *Signal Processing* **84**(1), 163–195.

Abd-Elrady, E. (2005). Nonlinear Approaches to Periodic Signal Modeling. Ph.D. thesis. Serie Acta Universitatis Uppsaliensis. Uppsala, Sweden.

Abd-Elrady, E. and T. Söderström (2004). Bias analysis in least squares estimation of periodic signals using nonlinear ODE's. Technical Report 2004-035. Dept. of Information Technology, Uppsala University. Uppsala, Sweden.

Abd-Elrady, E., T. Söderström and T. Wigren (2004). 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.

Khalil, H. K. (2002). *Nonlinear Systems*. third ed.. Prentice-Hall. Upper Saddle River, NJ, USA.

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

Söderström, T., H. Fan, B. Carlsson and S. Bigi (1997). Least squares parameter estimation of continuous time ARX models from discrete-time data. *IEEE Trans. on Automatic Control* **42**(5), 659–673.

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

Wigren, T. and T. Söderström (2003). Second order ODEs are sufficent for modeling of many periodic signals. Technical Report 2003-025. Dept. of Information Technology, Uppsala University. Uppsala, Sweden.

Wigren, T., E. Abd-Elrady and T. Söderström (2003*a*). 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*. Vol. VI. Hong Kong. pp. 669–672.

Wigren, T., E. Abd-Elrady and T. Söderström (2003*b*). Least squares harmonic signal analysis using periodic orbits of ODEs. In: *Proc. of 13th IFAC Symposium on System Identification*. Rotterdam, The Netherlands. pp. 1584–1589.