A bias-corrected estimator for nonlinear systems with output-error type model structures ⋆

Dario Piga [a], Roland Tóth [b]

[a] Istituto Dalle Molle di Studi sull'Intelligenza Artificiale, Scuola Universitaria Professionale della Svizzera Italiana, Galleria 1, 6928 Manno, Switzerland.

[b] Control Systems Group, Department of Electrical Engineering, Eindhoven University of Technology, P.O. Box 513, 5600 MB, Eindhoven, The Netherlands.

Abstract

Parametric identification of linear time-invariant (LTI) systems with output-error (OE) type of noise model structures has a well-established theoretical framework. Different algorithms, like instrumental-variables approaches or prediction error methods (PEMs), have been proposed in the literature to compute a consistent parameter estimate for linear OE systems. Although the prediction error method provides a consistent parameter estimate also for nonlinear output-error (NOE) systems, it requires to compute the solution of a nonconvex optimization problem. Therefore, an accurate initialization of the numerical optimization algorithms is required, otherwise they may get stuck in a local minimum and, as a consequence, the computed estimate of the system might not be accurate. In this paper, we propose an approach to obtain, in a computationally efficient fashion, a consistent parameter estimate for output-error systems with polynomial nonlinearities. The performance of the method is demonstrated through a simulation example.

Key words: Bias-corrected least-squares estimate, nonlinear system identification, output-error models.

1 Introduction

Parametric identification of *linear time-invariant* (LTI) systems with *output-error* (OE) noise models enjoys a well-established theoretical framework. Different identification techniques have been proposed in the literature to compute a consistent estimate of the system parameters, like *instrumental variables* based approaches [16]; prediction-error methods (PEMs) [14,6] and *bias-compensated least-squares* algorithms, where the standard *least square* (LS) estimate is properly modified in order to remove the bias introduced by the noise [17,2,21,22,9]. Among the aforementioned identification algorithms, only the PEM approach is guaranteed to provide a consistent estimate of the parameters of nonlinear systems with an output-error noise model. Specifically, in the PEM, the system parameters are estimated by minimizing the 2-norm of the difference between the measured output of the data-generating system and the simulated model output. This leads, also in the linear case, to a nonconvex optimization problem. Although, under mild assumptions, the global minimum of the minimized cost function is a consistent estimate of the systems parameters, the numerical optimization algorithms (e.g., gradient methods) can get trapped in local minima, which might lead to an inaccurate estimate of the system, in particular when the initial conditions of the optimization algorithm are not "close" to the global minimum or when complex nonlinear models have to be estimated (see, e.g., [13]).

Significant efforts have been spent in recent years to develop numerical efficient algorithms for parametric identification of *nonlinear output-error* (NOE) systems. In particular, an instrumental-variable based approach providing a consistent estimate for *linear-parameter-varying* systems under zero-mean colored noise conditions, e.g., output-error or *Box-Jenkins* setting, is proposed by Laurain *et al.* in [12]. In the context of block-oriented identification, different algorithms for parametric identification of Hammerstein-like and Wiener-like structures with output-error noise models are presented

⋆ This paper was not presented at any IFAC meeting. Corresponding author Dario Piga. Tel. +41-(0)58-666-6664. Fax +41-(0)58-666-6661.

⋆⋆This work was supported by the Netherlands Organization for Scientific Research (NWO, grant. no.: 639.021.127).

Email addresses: dario.piga@supsi.ch (Dario Piga), r.toth@tue.nl (Roland Tóth).

in [5,20,11,3,4,23]. In the more general framework of nonlinear *errors-in-variables* (EIV) models (i.e., when all the regressor variables are contaminated by error or measurement noise), identification schemes for systems described by continuous nonlinear functions are presented in [7,19,1]. In these contributions, every moment of the noise is assumed to be *a-priori* known. In [18], a generalization of the Koopmans-Levin's method [8], originally developed for EIV linear system identification, is properly extended to handle identification of static systems described by polynomial functions, under the assumption that the structure of the noise covariance matrix is known. In [10], Jun and Bernstein propose a method which is able to consistently estimate the parameters of nonlinear systems described by third or lower order polynomials without assuming that the noise covariance is known.

In this paper, we present a novel approach to consistently estimate the parameters of polynomial output-error systems with Gaussian-distributed measurement noise. One of the main benefits of the algorithm proposed in this paper is its ability to compute a consistent estimate of the system parameters with a modest computational complexity and without assuming to know the variance of the noise corrupting the data. The paper is organized as follows. In Section 2, the considered estimation problem is introduced. A consistent estimate of the system parameters is derived in Section 3 under the assumption that the variance of the noise affecting the output measurements is *a-priori* known. The latter assumption is relaxed in Section 4 in order to extend the applicability of the proposed method to a more general setting. The effectiveness of the presented identification procedure is shown in Section 5 through a simulation example.

2 Problem description

Consider a discrete-time, *single-input single-output* (SISO) data-generating system $\mathcal{S}_{\mathrm{o}}$ described by the *nonlinear output-error* (NOE) structure:

$$x(t)=h^{\mathrm{o}}(x(t-1),\ldots,x(t-n_{\mathrm{a}}),u(t),\ldots,u(t-n_{\mathrm{b}})), \tag{1a}$$

$$y(t)=x(t)+e_{\mathrm{o}}(t), \tag{1b}$$

where $u(t)$ is the measured input at time instant t, $x(t)$ and $y(t)$ are the noise-free and the noise-corrupted output, respectively, and $e_{\mathrm{o}}(t)$ is a stationary white Gaussian noise, independent of $x(t)$ and $u(t)$, with zero mean and finite variance σ_{e}^2. The function $h^{\mathrm{o}}(.)$ is a real-valued multivariate polynomial, which is parameterized as follows:

$$\begin{aligned}&h^{\mathrm{o}}(x(t-1),\ldots,x(t-n_{\mathrm{a}}),u(t),\ldots,u(t-n_{\mathrm{b}}))=\\&=\sum_{i=1}^{n_\theta}\theta_i^{\mathrm{o}}\psi_i(x(t-1),\ldots,x(t-n_{\mathrm{a}}),u(t),\ldots,u(t-n_{\mathrm{b}})),\end{aligned} \tag{2}$$

where $\theta_i^{\mathrm{o}}\in\mathbb{R}$ (with $i=1,\ldots,n_\theta$) are the unknown parameters to be estimated, while $\psi_i:\mathbb{R}^{n_\theta}\to\mathbb{R}$ (with $i=1,\ldots,n_\theta$) are a-priori chosen functions belonging to the canonical polynomial basis in the variables $x(t-1),\ldots,x(t-n_{\mathrm{a}}),u(t),\ldots,u(t-n_{\mathrm{b}})$. It is worth remarking that the assumption that h^{o} is a polynomial or it can be well-approximated by polynomial functions is realistic in many applications where the nonlinearities characterizing the systems are smooth enough.

Let us rewrite the data-generating system $\mathcal{S}_{\mathrm{o}}$ as

$$y(t)=\sum_{i=1}^{n_\theta}\theta_i^{\mathrm{o}}\psi_i\big(y(t-1)-e_{\mathrm{o}}(t-1),\ldots,u(t-n_{\mathrm{b}})\big)+e_{\mathrm{o}}(t). \tag{3}$$

By introducing the matrix notation

$$\begin{aligned}\theta_{\mathrm{o}}&=\begin{bmatrix}\theta_1^{\mathrm{o}} & \ldots & \theta_{n_\theta}^{\mathrm{o}}\end{bmatrix}^\top,\\ \varphi_{\mathrm{o}}(t)&=[\ \psi_1(x(t-1)\ldots x(t-n_{\mathrm{a}})\ u(t)\ldots u(t-n_{\mathrm{b}})),\\&\ \ \vdots\\&\ \ \ \psi_{n_\theta}(x(t-1)\ldots x(t-n_{\mathrm{a}})\ u(t)\ldots u(t-n_{\mathrm{b}}))],\end{aligned}$$

the system in (3) can be rewritten in the compact form

$$y(t)=\varphi_{\mathrm{o}}^\top(t)\theta_{\mathrm{o}}+e_{\mathrm{o}}(t). \tag{4}$$

Let us introduce the following parametric model $\mathcal{M}_\theta$ to describe the system $\mathcal{S}_{\mathrm{o}}$:

$$\begin{aligned}y(t)=&\sum_{i=1}^{n_\theta}\theta_i\psi_i\big(y(t-1),\ldots,y(t-n_{\mathrm{a}}),u(t),\ldots,u(t-n_{\mathrm{b}})\big)+\\&+\varepsilon(t)=\varphi^\top(t)\theta+\varepsilon(t),\end{aligned} \tag{5}$$

with $\varepsilon(t)$ denoting the residual term. The vectors $\theta\in\mathbb{R}^{n_\theta}$ and $\varphi(t)\in\mathbb{R}^{n_\theta}$ are defined as

$$\begin{aligned}\theta&=\begin{bmatrix}\theta_1, & \ldots, & \theta_{n_\theta}\end{bmatrix}^\top,\\ \varphi(t)&=[\ \psi_1(y(t-1),\ldots,y(t-n_{\mathrm{a}}),u(t),\ldots,u(t-n_{\mathrm{b}})),\\&\ \ \vdots\\&\ \ \ \psi_{n_\theta}(y(t-1),\ldots,y(t-n_{\mathrm{a}}),u(t),\ldots,u(t-n_{\mathrm{b}}))].\end{aligned}$$

It worth mentioning that the structure of the data-generating system $\mathcal{S}_{\mathrm{o}}$ is not known in practice. Thus, in order to guarantee that the true system belongs to the chosen model class, an over-parameterized model $\mathcal{M}_\theta$ can be considered. The goal of this contribution is to compute a consistent estimate of the system parameters θ_o based on a set of observed input/output data $\mathcal{D}_N = \{u(t), y(t)\}_{t=1}^N$ generated by $\mathcal{S}_\text{o}$.

The proposed algorithm is based on a proper modification of the *least-squares* methods. Before introducing the developed approach, let us first review the asymptotical properties of the *least-square* algorithm, which aims at minimizing the ℓ_2-loss function $\mathcal{V}(\theta, \mathcal{D}_N)$ defined as

$$\begin{aligned}\mathcal{V}(\theta, \mathcal{D}_N) &= \sum_{t=1}^N \frac{\varepsilon^2(t)}{N} = \sum_{t=1}^N \frac{1}{N}\left(y(t) - \varphi^\top(t)\theta\right)^2 = \\ &= \frac{1}{N}\|Y - \Phi\theta\|_2^2, \end{aligned} \tag{6}$$

where $Y = [y(1), \dots, y(N)]^\top \in \mathbb{R}^N$ and $\Phi \in \mathbb{R}^{N, n_\theta}$ is the noise-corrupted regressor matrix defined as

$$\Phi = \begin{bmatrix} \varphi(1) & \cdots & \varphi(N) \end{bmatrix}^\top. \tag{7}$$

The LS-estimate $\hat{\theta}_\text{LS}$ is then the argument minimizing the cost function $\mathcal{V}(\theta, \mathcal{D}_N)$ over $\theta \in \mathbb{R}^{n_\theta}$, i.e.,

$$\hat{\theta}_\text{LS} = \arg\min_{\theta \in \mathbb{R}^{n_\theta}} \frac{1}{N}\|Y - \Phi\theta\|_2^2 = \left(\frac{\Phi^\top\Phi}{N}\right)^{-1} \frac{\Phi^\top Y}{N}. \tag{8}$$

In order to compute the difference between $\hat{\theta}_\text{LS}$ and the true system parameters θ_o, let us write the output signal in (4) as

$$\begin{aligned} y(t) &= \varphi_\text{o}^\top(t)\theta_\text{o} + e_\text{o}(t) = \\ &= \varphi_\text{o}^\top(t)\theta_\text{o} + \varphi^\top(t)\theta_\text{o} - \varphi^\top(t)\theta_\text{o} + e_\text{o}(t) = \\ &= \varphi^\top(t)\theta_\text{o} + \Delta\varphi(t)\theta_\text{o} + e_\text{o}(t), \end{aligned}$$

with $\Delta\varphi(t) = \left(\varphi_\text{o}^\top(t) - \varphi^\top(t)\right)$. Let us stack the vectors $\Delta\varphi(t)$ and the noise samples $e_\text{o}(t)$, with $t = 1, \dots, N$, into the matrix $\Delta\Phi \in \mathbb{R}^{N, n_\theta}$ and into the vector $E_\text{o} \in \mathbb{R}^N$, respectively:

$$\Delta\Phi = \begin{bmatrix} \Delta\varphi^\top(1) \\ \vdots \\ \Delta\varphi^\top(N) \end{bmatrix} = \underbrace{\begin{bmatrix} \varphi_\text{o}^\top(1) \\ \vdots \\ \varphi_\text{o}^\top(N) \end{bmatrix}}_{\Phi_\text{o}} - \underbrace{\begin{bmatrix} \varphi^\top(1) \\ \vdots \\ \varphi^\top(N) \end{bmatrix}}_{\Phi}, \tag{9a}$$

$$E_\text{o} = [e_\text{o}(1) \ \dots \ e_\text{o}(N)]^\top. \tag{9b}$$

Based on the above definitions, the difference between the estimate $\hat{\theta}_\text{LS}$ and the true parameter vector θ_o can be expressed as:

$$\begin{aligned} \hat{\theta}_\text{LS} - \theta_\text{o} &= \left(\frac{\Phi^\top\Phi}{N}\right)^{-1} \frac{\Phi^\top Y}{N} - \theta_\text{o} = \\ &= \left(\frac{\Phi^\top\Phi}{N}\right)^{-1} \frac{\Phi^\top\left(\Phi\theta_\text{o} + \Delta\Phi\theta_\text{o} + E_\text{o}\right)}{N} - \theta_\text{o} = \\ &= \underbrace{\left(\frac{\Phi^\top\Phi}{N}\right)^{-1} \frac{\Phi^\top\Delta\Phi}{N}\theta_\text{o}}_{B_\Delta(\theta_\text{o}, \Phi, \Delta\Phi)} + \underbrace{\left(\frac{\Phi^\top\Phi}{N}\right)^{-1} \frac{\Phi^\top E_\text{o}}{N}}_{B_\text{E}}. \end{aligned} \tag{10}$$

Eq. (10) shows that the estimate $\hat{\theta}_\text{LS}$ is not consistent, i.e., $\lim_{N\to\infty} \hat{\theta}_\text{LS} - \theta_\text{o} \neq 0$. In fact, although the term B_E is guaranteed to converge to 0 as the number of measurements N goes to infinity, $B_\Delta(\theta_\text{o}, \Phi, \Delta\Phi)$ does not converge to 0 in general. The bias term $B_\Delta(\theta_\text{o}, \Phi, \Delta\Phi)$ will be referred in the sequel as noise-induced bias.

In the next section, we propose an algorithm to eliminate the noise-induced bias $B_\Delta(\theta_\text{o}, \Phi, \Delta\Phi)$, thus obtaining a consistent estimate of the true system parameters θ_o.

3 A bias-corrected LS estimate

In order to correct the noise-induced bias introduced by the LS estimate, first note that $B_\Delta(\theta_\text{o}, \Phi, \Delta\Phi)$ depends on the true parameters θ_o and the noise-free regressor matrix Φ_o (see the definition of $\Delta\Phi$ in (9a)). As a consequence, such a bias cannot be computed based on the observed input/output data and thus it cannot be directly subtracted from the LS estimate $\hat{\theta}_\text{LS}$.

Inspired by (10), the following corrected LS estimate is introduced:

$$\tilde{\theta}_\text{CLS} = \hat{\theta}_\text{LS} - B_\Delta(\tilde{\theta}_\text{CLS}, \Phi, \Delta\Phi), \tag{11}$$

with $B_\Delta(\tilde{\theta}_\text{CLS}, \Phi, \Delta\Phi)$ being

$$B_\Delta(\tilde{\theta}_\text{CLS}, \Phi, \Delta\Phi) = \left(\frac{\Phi^\top\Phi}{N}\right)^{-1} \frac{\Phi^\top\Delta\Phi}{N}\tilde{\theta}_\text{CLS}. \tag{12}$$

Algebraic manipulations of (11) lead to the following expression of $\tilde{\theta}_\text{CLS}$:

$$\tilde{\theta}_\text{CLS} = \left(\frac{\Phi^\top\Phi + \Phi^\top\Delta\Phi}{N}\right)^{-1} \frac{\Phi^\top Y}{N}. \tag{13}$$

Property 1 Let us assume that the following limit exists:

$$\lim_{N\to\infty} \left(\frac{\Phi^\top\Phi + \Phi^\top\Delta\Phi}{N}\right)^{-1}. \tag{14}$$

Then, $\tilde{\theta}_{\rm CLS}$ is a consistent estimate of the true system parameters $\theta_{\rm o}$, i.e.,

$$\lim_{N\to\infty} \tilde{\theta}_{\rm CLS} = \theta_{\rm o} \ \text{ w.p. 1.} \tag{15}$$

Proof: See Appendix 7.1. ■

Note that the estimate $\tilde{\theta}_{\rm CLS}$ does not explicitly depend on the true system parameters $\theta_{\rm o}$. However, it cannot be computed since it depends on the matrix $\Phi^\top \Delta\Phi$, which is unknown. In fact, $\Delta\Phi$ depends, by definition, on the noise-free regressor matrix $\Phi_{\rm o}$ (see (9a)). In order to overcome such a problem, the estimate $\tilde{\theta}_{\rm CLS}$ is modified by replacing $\Phi^\top \Delta\Phi$ in (11) with the matrix Ψ (constructed through Algorithm 1, see later) which depends on the measured output $y(t)$ and satisfies the following condition:

C1 $\displaystyle\lim_{N\to\infty} \frac{1}{N}\Phi^\top\Delta\Phi = \lim_{N\to\infty}\frac{1}{N}\Psi$ w.p. 1.

From (11), the new corrected LS estimate is then given by:

$$\hat{\theta}_{\rm CLS} = \hat{\theta}_{\rm LS} - B_\Delta(\hat{\theta}_{\rm CLS}, \Psi), \tag{16}$$

that is:

$$\hat{\theta}_{\rm CLS} = \left(\frac{\Phi^\top\Phi + \Psi}{N}\right)^{-1}\frac{\Phi^\top Y}{N}. \tag{17}$$

Property 2 Let us assume that the limit in (14) exists. Then, $\hat{\theta}_{\rm CLS}$ is a consistent estimate of the true system parameters $\theta_{\rm o}$, i.e.,

$$\lim_{N\to\infty} \hat{\theta}_{\rm CLS} = \theta_{\rm o} \ \text{ w.p. 1.} \tag{18}$$

Proof: See Appendix 7.2. ■

The matrix Ψ satisfying condition **C1** can be constructed through the following Algorithm.

Algorithm 1 [Construction of Ψ]

A1.1 Compute the analytical expression of the matrix $\mathbb{E}\left\{\Phi^\top\Delta\Phi\right\}$. Note that, since polynomial nonlinearities are considered, the entries of $\mathbb{E}\left\{\Phi^\top\Delta\Phi\right\}$ are described by an affine combination of the monomials $x(t), x^2(t), x^3(t), \ldots, x(t)x(t-1), \ldots, x(t)x(t-1)x(t-n_{\rm a}), \ldots$.

A1.2 For each monomial $x(t), x^2(t), x^3(t), \ldots, x(t)x(t-1), \ldots, x(t)x(t-1)x(t-n_{\rm a}), \ldots$, compute the coefficients $\alpha_1(t),\ \alpha_2(t),\ \alpha_3(t), \ldots, \alpha_n(t), \ldots, \alpha_m(t), \ldots$, such that:

- $x(t) = \mathbb{E}\left\{y(t) + \alpha_1(t)\right\}$
- $x^2(t) = \mathbb{E}\left\{y^2(t) + \alpha_2(t)\right\}$
- $x^3(t) = \mathbb{E}\left\{y^3(t) + \alpha_3(t)\right\}$
- $\vdots$
- $x(t)x(t-1) = \mathbb{E}\left\{y(t)y(t-1) + \alpha_n(t)\right\}$
- $\vdots$
- $x(t)x(t-1)x(t-n_{\rm a}) = \mathbb{E}\{y(t)y(t-1)y(t-n_{\rm a}) + \alpha_m(t)\}$

$\vdots$

To illustrate this construction, we develop the computation of the α coefficients through a recursive procedure. Consider first $\alpha_1(t)$. Then,

$$\begin{aligned} x(t) &= \mathbb{E}\left\{y(t)+\alpha_1(t)\right\} = \mathbb{E}\left\{x(t)+e_{\rm o}(t)+\alpha_1(t)\right\} = \\ &= x(t) + \mathbb{E}\left\{\alpha_1(t)\right\}. \end{aligned} \tag{19}$$

Equation (19) implies that $\alpha_1(t) = 0$ and $x(t) = \mathbb{E}\left\{y(t)\right\}$ for all $t = 1, \ldots, N$.

For α_2, we have:

$$\begin{aligned} x^2(t) &= \mathbb{E}\left\{y^2(t) + \alpha_2(t)\right\} = \\ &= \mathbb{E}\left\{(x(t) + e_{\rm o}(t))^2 + \alpha_2(t)\right\} = \\ &= x^2(t) + \sigma_{\rm e}^2 + \mathbb{E}\left\{\alpha_2(t)\right\}. \end{aligned} \tag{20}$$

Therefore, a possible choice is $\alpha_2(t) = -\sigma_{\rm e}^2$, which provides $x^2(t) = \mathbb{E}\left\{y^2(t)\right\} - \sigma_{\rm e}^2$ for all $t = 1, \ldots, N$.

In case $d > 2$, the values of $\alpha_d(t)$ can be recursively computed on the basis of the (previously computed) unbiased estimates of $x(t), x^2(t), \ldots, x^{d-1}(t)$. For instance, a possible choice of α_3 is given by:

$$\begin{aligned} x^3(t) &= \mathbb{E}\left\{y^3(t) + \alpha_3(t)\right\} = \\ &= \mathbb{E}\left\{(x(t) + e_{\rm o}(t))^3 + \alpha_3(t)\right\} = \\ &= x^3(t) + 3x(t)\sigma_{\rm e}^2 + \mathbb{E}\left\{\alpha_3(t)\right\}. \end{aligned} \tag{21}$$

Eq. (21) implies that $\alpha_3(t)$ should be such that:

$$\mathbb{E}\left\{\alpha_3(t)\right\} = -3x(t)\sigma_{\rm e}^2. \tag{22}$$

Since, based on the previous computation, $x(t) = \mathbb{E}\left\{y(t)\right\}$, from Eq. (22) we get:

$$\mathbb{E}\left\{\alpha_3(t)\right\} = -3x(t)\sigma_{\rm e}^2 = \mathbb{E}\left\{-3y(t)\sigma_{\rm e}^2\right\}. \tag{23}$$

This means that a possible choice for $\alpha_3(t)$ is $\alpha_3(t) = -3y(t)\sigma_{\rm e}^2$. Thus, $x^3(t) = \mathbb{E}\left\{y^3(t) - 3y(t)\sigma_{\rm e}^2\right\}$ for all $t = 1, \ldots, N$.

As far as the computation of the coefficient $\alpha_m(t)$ satisfying the condition $x(t)x(t-1)x(t-n_{\rm a}) = \mathbb{E}\left\{y(t)y(t-1)y(t-n_{\rm a}) + \alpha_m(t)\right\}$ is concerned, we have

$$\begin{aligned}
&x(t)x(t-1)x(t-n_{\mathrm{a}}) = \\
=&\mathbb{E}\left\{y(t)y(t-1)y(t-n_{\mathrm{a}})+\alpha_m(t)\right\} = \\
=&\mathbb{E}\left\{x(t)x(t-1)x(t-n_{\mathrm{a}})+\alpha_m(t)\right\},
\end{aligned} \tag{24}$$

which implies that a possible choice of $\alpha_m(t)$ is $\alpha_m(t) = 0$ for all $t = 1, \ldots, N$. Note that Eq. (24) follows from the assumption that the noise process e_{o} is white.

A1.3 The matrix Ψ is constructed by replacing, in the analytical expression of $\mathbb{E}\left\{\Phi^\top \Delta\Phi\right\}$, the monomials $x(t), x^2(t), x^3(t), \ldots, x(t)x(t-1), \ldots, x(t)x(t-1)x(t-n_{\mathrm{a}}), \ldots$ with $y(t)+\alpha_1(t), y^2(t)+\alpha_2(t), y^3(t)+\alpha_3(t), \ldots, y(t)y(t-1)+\alpha_n(t), \ldots, y(t)y(t-1)y(t-n_{\mathrm{a}})+\alpha_m(t), \ldots$. ■

An illustrative example on the construction of the matrix Ψ is reported in Appendix 7.4.

Property 3 The matrix Ψ, computed through Algorithm 1, satisfies condition **C1** under the assumption that the noise-free output sequence $\{x(t)\}_{t=1}^{\infty}$ is bounded, i.e., there exists a constant $M_x > 0$ such that

$$|x(t)| \leq M_x \quad \text{for all } t = 1, 2, \ldots. \tag{25}$$

Proof: See Appendix 7.3. ■

The application of the proposed identification scheme is limited, in principle, to the case when the value of the noise variance σ_{e}^2 is available, either because σ_{e}^2 is *a-priori* known or because it can be estimated through a set of dedicated experiments. In the next section, we present an algorithm to extend the applicability of the developed identification procedure to the case when the estimate of σ_{e}^2 is not *a-priori* available.

4 Estimation with unknown noise variance

In order to compute a relation between the noise variance σ_{e}^2 and the system parameters θ_{o}, let us rewrite the minimal value of the loss function $\mathcal{V}(\theta, \mathcal{D}_N)$ as

$$\begin{aligned}
&\mathcal{V}(\hat{\theta}_{\mathrm{LS}}, \mathcal{D}_N) = \frac{1}{N}\|Y-\Phi\hat{\theta}_{\mathrm{LS}}\|_2^2 = \\
&= \frac{1}{N}\|\Phi_{\mathrm{o}}\theta_{\mathrm{o}} + E_{\mathrm{o}} - \Phi\hat{\theta}_{\mathrm{LS}}\|_2^2 = \\
&= \frac{1}{N}\left(\|E_{\mathrm{o}}\|_2^2 + \|\Phi_{\mathrm{o}}\theta_{\mathrm{o}} - \Phi\hat{\theta}_{\mathrm{LS}}\|_2^2 - 2E_{\mathrm{o}}^\top\left(\Phi_{\mathrm{o}}\theta_{\mathrm{o}} - \Phi\hat{\theta}_{\mathrm{LS}}\right)\right).
\end{aligned} \tag{26}$$

As the number of measurements N goes to infinity, the term $\frac{1}{N}\|E_{\mathrm{o}}\|_2^2$ converges (w.p. 1) to σ_{e}^2, while $\frac{1}{N}E_{\mathrm{o}}^\top\left(\Phi_{\mathrm{o}}\theta_{\mathrm{o}} - \Phi\hat{\theta}_{\mathrm{LS}}\right)$ converges (w.p. 1) to 0 because of the independence of $e_{\mathrm{o}}(t)$ from the noise-free and noise-corrupted regressors $\varphi_{\mathrm{o}}(t)$ and $\varphi(t)$. Based on such considerations, from Eq. (26), it follows that:

$$\begin{aligned}
&\lim_{N\to\infty}\mathcal{V}(\hat{\theta}_{\mathrm{LS}}, \mathcal{D}_N) = \lim_{N\to\infty}\frac{1}{N}\|Y-\Phi\hat{\theta}_{\mathrm{LS}}\|_2^2 = \\
&= \sigma_{\mathrm{e}}^2 + \lim_{N\to\infty}\frac{1}{N}\left(\theta_{\mathrm{o}}^\top\Phi_{\mathrm{o}}^\top\Phi_{\mathrm{o}}\theta_{\mathrm{o}} + \hat{\theta}_{\mathrm{LS}}^\top\Phi^\top\Phi\hat{\theta}_{\mathrm{LS}} - 2\theta_{\mathrm{o}}^\top\Phi_{\mathrm{o}}^\top\Phi\hat{\theta}_{\mathrm{LS}}\right).
\end{aligned} \tag{27}$$

Let us now construct from the noise-corrupted output observations $y(t)$ two matrices $\Omega' \in \mathbb{R}^{n_\theta, n_\theta}$ and $\Omega'' \in \mathbb{R}^{n_\theta, n_\theta}$ satisfying the following condition:

X1 The matrices Ω' and Ω'' are such that:

$$\lim_{N\to\infty}\frac{1}{N}\Omega' = \lim_{N\to\infty}\frac{1}{N}\Phi_{\mathrm{o}}^\top\Phi_{\mathrm{o}}, \qquad \text{w.p. } 1, \tag{28a}$$

$$\lim_{N\to\infty}\frac{1}{N}\Omega'' = \lim_{N\to\infty}\frac{1}{N}\Phi_{\mathrm{o}}^\top\Phi, \qquad \text{w.p. } 1. \tag{28b}$$

The matrices $\Omega' \in \mathbb{R}^{n_\theta, n_\theta}$ and $\Omega'' \in \mathbb{R}^{n_\theta, n_\theta}$ satisfying condition **X1** can be constructed through a procedure similar to the one described in Algorithm 1 to construct the matrix Ψ. Note that, like for the matrix Ψ, the noise variance σ_{e}^2 is needed to construct the matrices Ω' and Ω''.

A (nonlinear) relation between σ_{e}^2 and θ_{o} is then obtained by substituting (28a) and (28b) into Eq. (27), i.e.,

$$\begin{aligned}
&\lim_{N\to\infty}\mathcal{V}(\hat{\theta}_{\mathrm{LS}}, \mathcal{D}_N) = \lim_{N\to\infty}\frac{1}{N}\|Y-\Phi\hat{\theta}_{\mathrm{LS}}\|_2^2 = \\
&= \sigma_{\mathrm{e}}^2 + \lim_{N\to\infty}\frac{1}{N}\left(\theta_{\mathrm{o}}^\top\Omega'\theta_{\mathrm{o}} + \hat{\theta}_{\mathrm{LS}}^\top\Phi^\top\Phi\hat{\theta}_{\mathrm{LS}} - 2\theta_{\mathrm{o}}^\top\Omega''\hat{\theta}_{\mathrm{LS}}\right).
\end{aligned} \tag{29}$$

Note that also the matrices Ω' and Ω' depend on the (unknown) noise variance σ_{e}^2. An estimate of σ_{e}^2 and the system parameters θ_{o} can be then computed by combing Eq. (17) and Eq. (29) (for finite N). This leads to the set of nonlinear equations in the variables (θ, σ^2):

$$\theta = \left(\Phi^\top\Phi + \Psi(\sigma^2)\right)^{-1}\Phi^\top Y, \tag{30a}$$

$$\frac{1}{N}\|Y-\Phi\hat{\theta}_{\mathrm{LS}}\|_2^2 = \sigma^2 + \frac{1}{N}\left(\theta^\top\Omega'\theta + \hat{\theta}_{\mathrm{LS}}^\top\Phi^\top\Phi\hat{\theta}_{\mathrm{LS}} - 2\theta^\top\Omega''\hat{\theta}_{\mathrm{LS}}\right). \tag{30b}$$

Indeed, as $N \to \infty$, the pair $(\theta_{\mathrm{o}}, \sigma_{\mathrm{e}}^2)$ becomes a solution of the set of equations in (30). A simple numerical algorithm to compute a solution of the nonlinear system of equations in (30) is described in the following.

Algorithm 2 [Combined bias-corrected LS and noise variance estimate]

Initialization: Set an upper bound $\sigma^2_{\max}$ that can be assumed for the noise variance $\sigma^2_{\rm e}$.

A2.1 Generate a set $\{\sigma^2_i\}_{i=1}^M$ of M equally-spaced points in the interval $[0,\ \sigma^2_{\max}]$.
A2.2 **for** i from 1 to M.
A2.3 Compute $\hat{\theta}^{(i)}$ through Eq. (30a) with σ^2_i as the noise variance, i.e., $\hat{\theta}^{(i)} = \left(\Phi^\top\Phi + \Psi(\sigma^2_i)\right)^{-1}\Phi^\top Y$.
A2.4 Compute the error term $\epsilon\big(\hat{\theta}^{(i)}\big)$ (see Eq. (30b)) as

$$\epsilon\big(\hat{\theta}^{(i)}\big) = \frac{1}{N}\|Y - \Phi\hat{\theta}_{\rm LS}\|_2^2 - \sigma^2_i + \\ -\frac{1}{N}\left[\hat{\theta}^{(i)\top}\Omega'\big(\sigma^2_i\big)\,\hat{\theta}^{(i)} + \hat{\theta}^\top_{\rm LS}\Phi^\top\Phi\hat{\theta}_{\rm LS} - 2\hat{\theta}^{(i)\top}\Omega''\big(\sigma^2_i\big)\,\hat{\theta}_{\rm LS}\right]$$

A2.5 **end for**
A2.6 **return** $\hat{\theta}_{\rm CLS} = \arg\min_{i=1,\ldots,M}\left|\epsilon\big(\hat{\theta}^{(i)}\big)\right|$. ■

Note that the accuracy of Algorithm 2 can be arbitrarily increased by refining the gridding at stage A2.1, at the cost of increasing the computational load.

It is worth pointing out that the idea of computing a relation between the noise variance $\sigma^2_{\rm e}$ and the system parameters $\theta_{\rm o}$ from the minimal value of the LS criterion $\mathcal{V}(\theta,\mathcal{D}_N)$ has been inspired by the papers [21,22,9], where bias-eliminated least-squares algorithms for identification of LTI systems in the EIV framework are discussed.

Remark 1 The bias-corrected estimate $\hat{\theta}_{\rm CLS}$ in (17) is guaranteed to be consistent also in case the measurement noise $e_{\rm o}(t)$ is not Gaussian. However, if $e_{\rm o}(t)$ is not Gaussian, the matrix Ψ depends not only on the noise variance $\sigma^2_{\rm e}$ (i.e., second-order moment of the noise $e_{\rm o}(t)$), but also on higher order moments. Therefore, in order to compute the bias-corrected estimate $\hat{\theta}_{\rm CLS}$ in case of unknown moments of $e_{\rm o}(t)$, a set of nonlinear equations depending on higher order moments of $e_{\rm o}(t)$ need to be considered together with (30b). However, solving such a set of nonlinear equations might be quite demanding in terms of computational resources. ■

5 Numerical example

The capabilities of the estimation scheme proposed in the paper are now shown through a simulation example.

5.1 Simulation setup

Consider the data-generating system $\mathcal{S}_{\rm o}$ described by the nonlinear output-error structure:

$$\begin{aligned} x(t) =& a^{\rm o}_1 x(t-1) + a^{\rm o}_2 x(t-2) + a^{\rm o}_4 x^2(t-2) + \\ &+ a^{\rm o}_6 x^3(t-1) + a^{\rm o}_{10} u(t), \\ y(t) =& x(t) + e_{\rm o}(t), \end{aligned}$$

with

$$\theta_{\rm o} = [a^{\rm o}_1\ a^{\rm o}_2\ a^{\rm o}_4\ a^{\rm o}_6\ a^{\rm o}_{10}]^\top = [-0.2\ \ 0.1\ \ 0.2\ \ -0.15\ \ 1.4]^\top.$$

Since the structure of the true system is not known in practice, the following over-parameterized model class is chosen:

$$\begin{aligned} y(t) =& a_1 y(t-1) + a_2 y(t-2) + a_3 y^2(t-1) + a_4 y^2(t-2) + \\ &+ a_5 y(t-1)y(t-2) + a_6 y^3(t-1) + a_7 y^2(t-1)y(t-2) \\ &+ a_8 y(t-1)y^2(t-2) + a_9 y^3(t-2) + a_{10} u(t) + \varepsilon(t). \end{aligned}$$

The noise measurement $e_{\rm o}(t)$ is taken as a zero-mean stationary white-noise process with Gaussian distribution $\mathcal{N}(0,\sigma^2_{\rm e})$, while the input signal $u(t)$ is a white-noise sequence with uniform distribution $\mathcal{U}(-0.5, 0.5)$. The model parameters θ are estimated from an input/output data set $\mathcal{D}_N$ of length $N = 4000$ generated by the system. In order to empirically study the statistical properties of the developed bias-correction scheme, a Monte Carlo study with $N_{\rm MC} = 1000$ runs with new noise and input realizations in each run, is carried out. In this study, $\sigma^2_{\rm e}$ is chosen to be 0.03, corresponding to an average *signal-to-noise ratio* (SNR) of 8 dB. The SNR is defined as

$$\text{SNR} = 10\log_{10}\left(\sum_{t=1}^N x^2(t) \Big/ \sum_{t=1}^N e^2_{\rm o}(t)\right).$$

5.2 Obtained results

The following three estimates of the model parameters θ are computed:

- LS estimate $\hat{\theta}_{\rm LS}$, computed by minimizing the sample variance of the residual $\varepsilon(t)$.
- PEM estimate with output error noise model. The LS estimate $\hat{\theta}_{\rm LS}$ has been used as an initial estimate for the PEM.
- Bias-corrected LS estimate $\hat{\theta}_{\rm CLS}$, computed through Eq. (17) and under the assumption that the value of the noise variance $\sigma^2_{\rm e}$ is known.
- Bias-corrected LS estimate $\hat{\theta}_{\rm CLS}$, computed through Algorithm 2, i.e., the noise variance $\sigma^2_{\rm e}$ is assumed to be *a-priori* unknown. 101 equally-spaced points in the interval $[0,\ \sigma^2_{\max}] = [0,\ 1]$ are considered in Algorithm 2.

The obtained results are reported in Table 1, which shows the average of the estimated parameters and their standard deviation over the 1000 Monte Carlo runs. Table 1 also shows that, in line with the theory, the LS method provides a biased estimate of the system parameters, while the approach proposed in the paper provides a consistent estimate of the true system parameters $\theta_{\rm o}$ also in the case when the noise variance is unknown. It is worth noting that, due to the uncertainty introduced in estimating the matrix Ψ, the bias-corrected approach provides an estimate of the model parameters with a larger variance than the LS method. Results in Table 1 also emphasize that, even if the PEM is guaranteed to provide, theoretically, a consistent estimate for NOE models, due of the usual absence of a good initialization and heavy complexity of the associated nonlinear optimization problem, it fails to provide reliable estimates. This concludes that the proposed approach offers a good tradeoff in terms consistency, variance, and required computational effort.

6 Conclusion

In this paper, we have proposed a method for computing a consistent parameter estimate for output-error systems with polynomial nonlinearities. The noise corrupting the output measurements has been assumed to be white and Gaussian with unknown variance. The underlying idea of the proposed approach is to estimate, from the measured data, the bias introduced by the LS approach. The estimated bias is guaranteed to asymptotically converge to the true one as the number of measurements increases, and it is used to correct the LS estimate. Possible extensions of the developed approach include:

- identification of nonlinear systems with different types of nonlinear parameterizations, noise-models and noise distributions;
- data-driven selection of the model structure.

7 Appendix

7.1 *Proof of Property 1*

Property 1 is proved on the basis of the following algebraic manipulations:

$$\begin{aligned}\tilde{\theta}_{\mathrm{CLS}} &= \left(\frac{\Phi^\top\Phi+\Phi^\top\Delta\Phi}{N}\right)^{-1}\frac{1}{N}\Phi^\top\underbrace{(\Phi\theta_{\mathrm{o}}+\Delta\Phi\theta_{\mathrm{o}}+E_{\mathrm{o}})}_{Y}\\ &= \left(\frac{\Phi^\top\Phi+\Phi^\top\Delta\Phi}{N}\right)^{-1}\left[\left(\frac{\Phi^\top\Phi+\Phi^\top\Delta\Phi}{N}\right)\theta_{\mathrm{o}}+\frac{\Phi^\top E_{\mathrm{o}}}{N}\right]\\ &= \theta_{\mathrm{o}}+\left(\frac{\Phi^\top\Phi+\Phi^\top\Delta\Phi}{N}\right)^{-1}\frac{\Phi^\top E_{\mathrm{o}}}{N}.\end{aligned}$$

Because of the independence between the measurement noise $e_{\mathrm{o}}(t)$ and the regressor $\varphi(t)$, the term

$$\left(\frac{\Phi^\top\Phi+\Phi^\top\Delta\Phi}{N}\right)^{-1}\frac{\Phi^\top E_{\mathrm{o}}}{N} \tag{32}$$

converges to zero with probability 1. This implies (15).

7.2 *Proof of Property 2*

Let us rewrite the bias-corrected estimate $\hat{\theta}_{\mathrm{CLS}}$ in (17) as follows

$$\begin{aligned}\hat{\theta}_{\mathrm{CLS}} &= \left(\frac{\Phi^\top\Phi+\Psi}{N}\right)^{-1}\frac{1}{N}\Phi^\top\underbrace{(\Phi\theta_{\mathrm{o}}+\Delta\Phi\theta_{\mathrm{o}}+E_{\mathrm{o}})}_{Y}\\ &= \left(\frac{\Phi^\top\Phi}{N}+\frac{\Psi}{N}\right)^{-1}\left(\frac{\Phi^\top\Phi}{N}+\frac{\Phi^\top\Delta\Phi}{N}\right)\theta_{\mathrm{o}}+\\ &\quad+\left(\frac{\Phi^\top\Phi}{N}+\frac{\Psi}{N}\right)^{-1}\frac{\Phi^\top E_{\mathrm{o}}}{N}.\end{aligned} \tag{33}$$

Because of the independence between the measurement noise $e_{\mathrm{o}}(t)$ and the regressor $\varphi(t)$, the term $\left(\frac{\Phi^\top\Phi}{N}+\frac{\Psi}{N}\right)^{-1}\frac{\Phi^\top E_{\mathrm{o}}}{N}$ converges to zero with probability 1, while, because of condition **C1**, the matrix

$$\left(\frac{\Phi^\top\Phi}{N}+\frac{\Psi}{N}\right)^{-1}\left(\frac{\Phi^\top\Phi}{N}+\frac{\Phi^\top\Delta\Phi}{N}\right) \tag{34}$$

converges to the identity matrix with probability 1. Based on the above considerations and from Eq. (33), Property 2 follows.

7.3 *Proof of Property 3*

In order to prove that Ψ satisfies condition **C1**, the following necessary lemma coming from a direct application of the Ninness's strong law of large numbers [15] is first presented.

Lemma 1 *Let $\{\nu(t)\}$ be a sequence of random variables with arbitrary correlation structure (not necessarily stationary) that is characterized by the existence of a finite value C such that*

$$\sum_{t=1}^{N}\sum_{s=1}^{N}\mathbb{E}\left\{\nu(t)\nu(s)\right\}<CN. \tag{35}$$

Then,

$$\frac{1}{N}\sum_{t=1}^{N}\nu(t)\xrightarrow{a.s.}0 \quad as \quad N\to\infty \tag{36}$$

■

Let $[\cdot]_{i,j}$ be the (i,j)-th entry of a matrix. Let us consider the term:

$$\left[\frac{1}{N}\Psi-\frac{1}{N}\Phi^\top\Delta\Phi\right]_{i,j}=\frac{1}{N}\sum_{t=1}^{N}\nu_{i,j}(t), \tag{37}$$

Table 1
Mean and standard deviation of the estimates of the parameters θ over the 1000 Monte Carlo runs.

		True value	least-squares estimate $\hat{\theta}_{\mathrm{LS}}$	PEM estimate	bias-corrected estimate $\hat{\theta}_{\mathrm{CLS}}$ (σ_{e}^2 known)	bias-corrected estimate $\hat{\theta}_{\mathrm{CLS}}$ (σ_{e}^2 unknown)
a_1	mean	-0.2	-0.2233	-0.2205	-0.2006	-0.2008
	std	–	0.0094	0.0365	0.0184	0.0186
a_2	mean	0.1	0.1009	0.0988	0.1005	0.1004
	std	–	0.0091	0.0293	0.0136	0.0137
a_3	mean	0	0.0043	-0.0045	-0.0004	-0.0006
	std	–	0.0074	0.0304	0.0143	0.0147
a_4	mean	0.2	0.1350	0.1444	0.2003	0.2006
	std	–	0.0073	0.0297	0.0128	0.0129
a_5	mean	0	-0.0065	-0.0105	-0.0002	-0.0005
	std	–	0.0111	0.0322	0.0158	0.0160
a_6	mean	-0.15	0.0234	-0.0127	-0.1489	-0.1493
	std	–	0.0168	0.1105	0.0576	0.0589
a_7	mean	0	0.0774	0.0586	0.0013	0.0013
	std	–	0.0236	0.1187	0.0506	0.0514
a_8	mean	0	0.0023	0.0066	0.0016	0.0012
	std	–	0.0217	0.0948	0.0365	0.0377
a_9	mean	0	-0.0221	-0.0051	-0.0010	-0.0012
	std	–	0.0148	0.0594	0.0324	0.0341
a_{10}	mean	1.4	1.4000	1.4001	1.4000	1.4000
	std	–	0.0070	0.0206	0.0070	0.0070

From the construction of the matrix Ψ (Algorithm 1), the random variable $\nu_{i,j}(t)$ is guaranteed to be zero-mean and it only depends on the deterministic noise-free output samples $x(t-1), \ldots, x(t-n_{\mathrm{a}})$ and on the white noise samples $e_{\mathrm{o}}(t-1), \ldots, e_{\mathrm{o}}(t-n_{\mathrm{a}})$. As a consequence, the variables $\nu_{i,j}(t)$ and $\nu_{i,j}(s)$ are stochastically independent for all t, s such that $s \geq t + n_{\mathrm{a}}$. Therefore,

$$\mathbb{E}\left\{\nu_{i,j}(t)\nu_{i,j}(s)\right\} = 0 \quad \text{for all } s \geq t + n_{\mathrm{a}}. \tag{38}$$

Note also that, since $x(t)$ is assumed to be bounded for all $t \geq 0$ and the variance σ_{e}^2 is finite, then the term $\mathbb{E}\left\{\nu_{i,j}(t)\nu_{i,j}(s)\right\}$ is bounded for any index-pair $t, s > 0$, i.e., there exists a positive constant $M_{i,j}$ such that

$$\mathbb{E}\left\{\nu_{i,j}(t)\nu_{i,j}(s)\right\} < M_{i,j} \quad \text{for all } s, t > 0. \tag{39}$$

Based on the above considerations, we have:

$$\begin{aligned}\sum_{t=1}^{N}\sum_{s=1}^{N} \mathbb{E}\{\nu_{i,j}(t)\nu_{i,j}(s)\} &= \sum_{t=1}^{N} \sum_{s=t}^{\min\{t+n_{\mathrm{a}}-1, N\}} \mathbb{E}\left\{\nu_{i,j}(t)\nu_{i,j}(s)\right\} \\ &< \sum_{t=1}^{N} n_{\mathrm{a}} M_{i,j} = n_{\mathrm{a}} M_{i,j} N.\end{aligned}$$

Therefore, from Lemma 1, it follows

$$\frac{1}{N}\sum_{t=1}^{N} \nu_{i,j}(t) \xrightarrow{a.s.} 0 \quad \text{as } N \to \infty, \tag{40}$$

or equivalently, (see Eq. (37))

$$\lim_{N\to\infty} \frac{1}{N}\left[\Psi\right]_{i,j} = \lim_{N\to\infty} \frac{1}{N}\left[\Phi^\top \Delta \Phi\right]_{i,j} \quad \text{w.p. 1.} \tag{41}$$

This proves that Ψ satisfies condition **C1**.

7.4 *Example: structure of the matrices $\Phi^\top \Delta \Phi$ and Ψ*

Let us consider a data-generating system $\mathcal{S}_{\mathrm{o}}$ described by

$$\begin{aligned} x(t) &= a_{11}x(t-1) + a_{12}x^2(t-1) + a_{21}x(t-2), &\text{(45a)} \\ y(t) &= x(t) + e_{\mathrm{o}}(t), &\text{(45b)}\end{aligned}$$

where a_{11}, a_{12} and a_{21} are real-valued constants. Then, the noise-free and the noise-corrupted regressors $\varphi_{\mathrm{o}}(t)$ and $\varphi(t)$ are:

$$\varphi_{\mathrm{o}}(t) = \begin{bmatrix} x(t-1) \\ x^2(t-1) \\ x(t-2) \end{bmatrix}, \quad \varphi(t) = \begin{bmatrix} y(t-1) \\ y^2(t-1) \\ y(t-2) \end{bmatrix}, \tag{46}$$

and the matrix $\Phi^\top \Delta \Phi$ and its expected value are given by (42) and (43), respectively. The matrix Ψ (see (44)) is then obtained by substituting the terms $x(t-1)$ and $x^2(t-1)$ in (43) with $y(t-1)$ and $y^2(t-1) - \sigma_{\mathrm{e}}^2$, respectively.

$$\Phi^\top \Delta\Phi = \begin{bmatrix} \sum_{t=1}^{N} y(t-1)(x(t-1)-y(t-1)) & \sum_{t=1}^{N} y(t-1)\big(x^2(t-1)-y^2(t-1)\big) & \sum_{t=1}^{N} y(t-1)(x(t-2)-y(t-2)) \\ \sum_{t=1}^{N} y^2(t-1)(x(t-1)-y(t-1)) & \sum_{t=1}^{N} y^2(t-1)\big(x^2(t-1)-y^2(t-1)\big) & \sum_{t=1}^{N} y^2(t-1)(x(t-2)-y(t-2)) \\ \sum_{t=1}^{N} y(t-2)(x(t-1)-y(t-1)) & \sum_{t=1}^{N} y(t-2)\big(x^2(t-1)-y^2(t-1)\big) & \sum_{t=1}^{N} y(t-2)(x(t-2)-y(t-2)) \end{bmatrix} \tag{42}$$

$$\mathbb{E}\left\{\Phi^\top \Delta\Phi\right\} = -\begin{bmatrix} N\sigma_{\mathrm{e}}^2 & \sigma_{\mathrm{e}}^2 \sum_{t=1}^{N} 3x(t-1) & 0 \\ \sigma_{\mathrm{e}}^2 \sum_{t=1}^{N} 2x(t-1) & \sigma_{\mathrm{e}}^2 \sum_{t=1}^{N} 5x^2(t-1) + 3N\sigma_e^4 & 0 \\ 0 & \sigma_{\mathrm{e}}^2 \sum_{t=1}^{N} x(t-2) & N\sigma_{\mathrm{e}}^2 \end{bmatrix} \tag{43}$$

$$\Psi = -\begin{bmatrix} N\sigma_{\mathrm{e}}^2 & \sigma_{\mathrm{e}}^2 \sum_{t=1}^{N} 3y(t-1) & 0 \\ \sigma_{\mathrm{e}}^2 \sum_{t=1}^{N} 2y(t-1) & \sigma_{\mathrm{e}}^2 \sum_{t=1}^{N} 5\left(y^2(t-1)-\sigma_{\mathrm{e}}^2\right) + 3N\sigma_e^4 & 0 \\ 0 & \sigma_{\mathrm{e}}^2 \sum_{t=1}^{N} y(t-2) & N\sigma_{\mathrm{e}}^2 \end{bmatrix} \tag{44}$$

References

[1] S. Baran. A consistent estimator in general functional errors-in-variables models. *Metrika*, 51(2):117–132, 2000.

[2] S. Beghelli, R. Guidorzi, and U. Soverini. The frisch scheme in dynamic system identification. *Automatica*, 26(1):171–176, 1990.

[3] V. Cerone, D. Piga, and D. Regruto. Bounded error identification of Hammerstein systems through sparse polynomial optimization. *Automatica*, 48(10):2693–2698, 2012.

[4] V. Cerone, D. Piga, and D. Regruto. Computational load reduction in bounded error identification of Hammerstein systems. *IEEE Transactions on Automatic Control*, 58(5):1317–1322, 2013.

[5] F. Ding, Y. Shi, and T. Chen. Auxiliary model-based least-squares identification methods for Hammerstein output-error systems. *Systems & Control Letters*, 56(5):373–380, 2007.

[6] R. Diversi, R. Guidorzi, and U. Soverini. Maximum likelihood identification of noisy input–output models. *Automatica*, 43(3):464–472, 2007.

[7] I. Fazekas and A. G. Kukush. Asymptotic properties of an estimator in nonlinear functional errors-in-variables models with dependent error terms. *Computers & Mathematics with Applications*, 34(10):23–39, 1997.

[8] K.V. Fernando and H. Nicholson. Identification of linear systems with input and output noise: the Koopmans-Levin method. In *Control Theory and Applications, IEE Proceedings D*, volume 132, pages 30–36, 1985.

[9] M. Hong, T. Söderström, and W. X. Zheng. Accuracy analysis of bias-eliminating least squares estimates for errors-in-variables systems. *Automatica*, 43(9):1590–1596, 2007.

[10] B.E. Jun and D. S. Bernstein. Extended least-correlation estimates for errors-in-variables non-linear models. *International Journal of Control*, 80(2):256–267, 2007.

[11] V. Laurain, M. Gilson, and H. Garnier. Refined instrumental variable methods for Hammerstein Box-Jenkins models. *System Identification, Environmental Modelling, and Control System Design*, pages 27–47, 2012.

[12] V. Laurain, M. Gilson, R. Tóth, and H. Garnier. Refined instrumental variable methods for identification of LPV Box–Jenkins models. *Automatica*, 46(6):959–967, 2010.

[13] L. Lennart. Perspectives on system identification. *Annual Reviews in Control*, 34(1):1–12, 2010.

[14] L. Ljung. *System Identification, Theory for the User.* Prentince Hall, Upper Saddle River, 1999.

[15] B. Ninness. Strong laws of large numbers under weak assumptions with application. *IEEE Transactions on Automatic Control*, 45(11):2117–2122, 2000.

[16] T. Söderström and P. Stoica. *Instrumental variable methods for system identification*, volume 161. Springer-Verlag Berlin, 1983.

[17] P. Stoica and T. Söderström. Bias correction in least-squares identification. *International Journal of Control*, 35(3):449–457, 1982.

[18] I. Vajk and J. Hetthéssy. Identification of nonlinear errors-in-variables models. *Automatica*, 39(12):2099–2107, 2003.

[19] G. Vandersteen, Y. Rolain, J. Schoukens, and R. Pintelon. On the use of system identification for accurate parametric

modeling of nonlinear systems using noisy measurements. *IEEE Transactions on Instrumentation and Measurement*, 45(2):605–609, 1996.

[20] J. Wang, Q. Zhang, and L. Ljung. Revisiting Hammerstein system identification through the two-stage algorithm for bilinear parameter estimation. *Automatica*, 45(11):2627–2633, 2009.

[21] W. X. Zheng. Transfer function estimation from noisy input and output data. *International Journal of Adaptive Control and Signal Processing*, 12(4):365–380, 1998.

[22] W. X. Zheng. A bias correction method for identification of linear dynamic errors-in-variables models. *IEEE Transactions on Automatic Control*, 47(7):1142–1147, 2002.

[23] Y. Zhu. Estimation of an N–L–N Hammerstein–Wiener model. *Automatica*, 38(9):1607–1614, 2002.