The 30th Annual Conference of the IEEE Industrial Electronics Society, November 2 - 6, 2004, Busan, Korea

Least-Correlation Estimates for Errors-in-Variables Nonlinear Models

Byung-Eul Jun[1] and Dennis S. Bernstein[2]

[1]Principal Researcher, Agency for Defense Development, Daejeon 305-600, Korea
[2]Professor, Aerospace Eng., University of Michigan, Ann Arbor, MI 48109, USA
email: [1]bejun@unitel.co.kr, [2]dsbaero@umich.edu

Abstract—**In this paper, we introduce a method of parameter estimation working on errors-in-variables nonlinear models whose all variables are corrupted by noise. Main idea is to augment the parameters and the regressors of the linear regressor models by even-order components of noises and by appropriate constants, respectively, and to employ the method of least correlation, which has a capability to cope with errors-in-variables models, for the extended models. Analysis shows that for the polynomial nonlinearity of up to third order, the estimate converge to the true parameters as the number of samples increases toward infinity. We discuss the expected performance of the estimates applied to fourth or higher-order polynomial nonlinear models. Monte Carlo simulations of simple numerical examples support the analytical results.**

I. INTRODUCTION

The method of least-squares works on nonlinear systems provided that they are described by the linear regression model and the measurements composing regressor vector are free from noise. A generalized approach to modelling noise is to view all variables as contaminated by noise, called errors-in-variables (EIV) models [8, 11, 15]. The noises included in regressors make the estimation problem challenging. The nonlinearities in system models make the problem more challenging. There are previous contributions for EIV nonlinear models which are described by polynomials [13], Volterra models [9, 10], Wiener-Hammerstein models [12], and general functions [2, 6, 14]. Vajk and Hetthessy's work [13] is a generalization of a classical eigenvalue-decomposition method to polynomial nonlinear systems with a priori knowledge about noise. In the works for Volterra models, they assume that the input is a amplitude-modulated cyclostationary signal [10] or they consider the polyperiodically time-varying models [9]. Tan and Godfrey [12] identify the linear subsystems of a Wiener-Hammerstein model through the measurement of the second-order Volterra kernel in frequency domain.

This paper is an extension of the least-correlation estimate for EIV linear models [4] to a type of EIV nonlinear systems described by polynomials. The least-correlation estimate has a capability to find out the best fit to the given structure of EIV linear models without a priori knowledge about noise, but direct application of the least-correlation estimate to nonlinear systems yields error-prone results affected by the noises included in regressor vector. We introduce a way of augmenting the regressor vector and extending the parameter vector of the linear regression model which is equivalent to the original nonlinear system. Applying the method of least correlation to the new models with the augmented regressors and the extended parameters gives the least-correlation estimates for nonlinear systems. Analysis shows that the estimates for systems with second-order or third-order nonlinearity converge to the mathematical expectation of the extended parameter vector as the number of samples increases to infinity. Expected performance of the method applied to fourth or higher-order systems are discussed. Monte Carlo simulations for simple examples support the analytic results.

II. SYSTEM MODELS AND ASSUMPTIONS

Consider the discrete-time Volterra models [1]

$$z(t) = z_0 + \sum_{\ell=1}^{L} \chi_M^\ell(t) + \eta_1(t), \qquad (1)$$

$$\chi_M^\ell(t) = \sum_{i_1=0}^{M} \cdots \sum_{i_\ell=0}^{M} \sum_{j=1}^{n} \alpha_\ell^j(i_1, \cdots, i_\ell) \times u_j(t-i_1)\cdots u_j(t-i_\ell), \qquad (2)$$

where $z(t) \in \mathbb{R}$ is the system response at tth sampling, z_0 is a constant, $\eta_1(t) \in \mathbb{R}$ denotes possible errors in modelling, $u_j(t)$ is the jth element of input vector $u(t) \in \mathbb{R}^n$, L denotes the nonlinear degree of model, and M is its dynamic order [1]. The equations (1)-(2) state a general Volterra model, but we will impose some restrictions on it in later sections.

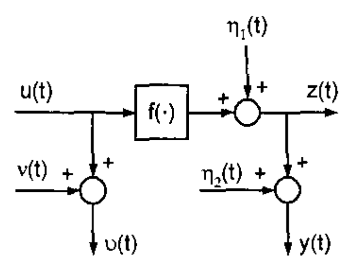

Fig. 1. Description of errors-in-variables nonlinear dynamic models

Suppose that both $z(t)$ and $u(t)$ are measured in noise as depicted in Fig.1. Let $y(t) \in \mathbb{R}$ and $v(t) \in \mathbb{R}^n$ denote available measurements of $z(t)$ and $u(t)$, respectively, *i.e.*,

$$y(t) = z(t) + \eta_2(t), \qquad (3)$$

$$v(t) = u(t) + \nu(t), \qquad (4)$$

where $\eta_2(t) \in \mathbb{R}$ and $\nu(t) \in \mathbb{R}^n$ denote measurement noises. Taking into account the noises in all variables is referred to errors-in-variables models [7, 8, 15]. Employing (3)-(4) to

(1)-(2) gives the errors-in-variables model:

$$y(t) = y_0 + \sum_{\ell=1}^{L} \chi_M^\ell(t) + \eta(t), \quad (5)$$

$$\chi_M^\ell(t) = \sum_{i_1=0}^{M} \cdots \sum_{i_\ell=0}^{M} \sum_{j=1}^{n} \alpha_\ell^j(i_1, \cdots, i_\ell) \{v_j(t-i_1) \cdots \cdots v_j(t-i_\ell) - w_j(t, i_1, \cdots, i_\ell)\}, \quad (6)$$

$$\begin{aligned} w_j(\cdot) &= v_j(t-i_1)\cdots v_j(t-i_{\ell-1})\nu_j(t-i_\ell) \\ &+ v_j(t-i_1)\cdots v_j(t-i_{\ell-2})\nu_j(t-i_{\ell-1}) \\ &\times [v_j(t-i_\ell) - \nu_j(t-i_\ell)] \\ &\vdots \\ &+ \nu_j(t-i_1)[v_j(t-i_2) - \nu_j(t-i_2)]\cdots \\ &\times [v_j(t-i_\ell) - \nu_j(t-i_\ell)], \end{aligned} \quad (7)$$

where $\eta(t) \triangleq \eta_1(t) + \eta_2(t)$. Now let us state the estimation problem.

Problem 1. Given the system model (1)-(2) and the measurement model (3)-(4), determine an estimate of the system parameters $\alpha_\ell^j(i_1, \cdots, i_\ell)$ based on available measurements $v(t)$ and $y(t)$.

Identification problems frequently work with signals that are described as stochastic processes with deterministic components. For a common framework for deterministic and stochastic signals, we employ the definition of quasi-stationary signals and the notation

$$\bar{E}[f(t)] \triangleq \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} E[f(t)] \quad (8)$$

which works on the deterministic components as well as the stochastic parts of quasi-stationary signal $f(t)$, where E denotes the mathematical expectation [7, p.34]. We implicitly assume that the limit in (8) exists when $\bar{E}$ is used.

We introduce the following assumptions.

A1. The structure of system model, the nonlinear degree L, the dynamic order M, and the number of inputs n are known a priori.

A2. The measured signals $v(t)$ and $y(t)$ are quasi-stationary and jointly quasi-stationary.

A3. The noises $\eta(t)$ and $\nu(t)$ are zero-mean, stationary and $E[\nu_j^L(t)] = 0, j = 1, \cdots, n$ for all odd L, and they are finitely autocorrelated with $v(t)$, that is, there exists $\tau > 0$ such that

$$\bar{E}\left[v(t)\nu^T(t-k)\right] = 0, \quad (9)$$

$$\bar{E}\left[v(t)\eta(t-k)\right] = 0 \quad (10)$$

for all $|k| \geq \tau$.

A4. For τ in A3, none of the elements of $v(t)$ is constant for all measurements, and $v(t)$ satisfies

$$\operatorname{rank}\left\{\bar{R}_{vv}(t, t-\tau, N) + \bar{R}_{vv}(t-\tau, t, N)\right\} = n, \quad (11)$$

where N denotes the number of samples and the empirical correlation $\bar{R}_{vv}(t_1, t_2)$ is defined by

$$\bar{R}_{vv}(t_1, t_2, N) \triangleq \frac{1}{N-\tau} \sum_{t=1+\tau}^{N} v(t_1)v^T(t_2) \quad (12)$$

with either $t_1 = t, t_2 = t - \tau$ or $t_1 = t - \tau, t_2 = t$ and $\tau \triangleq |t_1 - t_2|$.

Assumptions A3 and A4 express the idea that the correlations between signals are stronger than those between noises as well as those between signals and noises. Conditions (9) and (10) are equivalent to

$$\bar{E}\left[u(t)\nu^T(t-k)\right] = 0, \quad \bar{E}\left[\nu(t)\nu^T(t-k)\right] = 0, \quad (13)$$

$$\bar{E}\left[u(t)\eta(t-k)\right] = 0, \quad \bar{E}\left[\nu(t)\eta(t-k)\right] = 0, \quad (14)$$

for all $|k| \geq \tau > 0$, respectively, owing to (4) and the arbitrariness of $u(t)$, $\nu(t)$, and $\eta(t)$.

III. Main Results

We will confine the discussions in this section to static nonlinear systems with multi-input single-output. Some results can be extended with ease to multi-input, multi-output systems.

A. Second-order Nonlinear Systems

Consider a second-order nonlinear model

$$z(t) = \sum_{j=1}^{n} a_j u_j^2(t) + \eta_1(t) \quad (15)$$

which is reduced from the Volterra model (1)-(2) at

$$z_0 = 0,\ M = 0,\ L = 2,\ \alpha_1(0) = 0, \\ \alpha_2^j(0,0) = a_j, j = 1, \cdots, n. \quad (16)$$

Using (3)-(4) to (15) yields

$$y(t) = \sum_{j=1}^{n} a_j (v_j(t) - \nu_j(t))^2(t) + \eta(t). \quad (17)$$

Equation (17) can also be obtained from (5)-(7) under the condition (16). It is clear that not only the method of least squares but also the least-correlation estimate [4] generates error-prone results for the errors-in-variables nonlinear models.

For $v_j(t)$ in the terms where the odd-order of $\nu_j(t)$ appears in the expansion of (17), substituting $v_j(t) = u_j(t) + \nu_j(t)$ gives

$$y(t) = \sum_{j=1}^{n} a_j \left[v_j^2(t) - \nu_j^2(t) - 2u_j(t)\nu_j(t)\right] + \eta(t). \quad (18)$$

Let (18) be the linear regression form

$$y(t) = \psi_a^T(t)\theta_a(t) + e(t) \quad (19)$$

with the error $e(t)$, the augmented regressor vector $\psi_a(t) \in \mathbb{R}^{n+1}$ and the extended parameter vector $\theta_a(t) \in \mathbb{R}^{n+1}$ defined by

$$e(t) = \eta(t) - 2\sum_{j=1}^{n} a_j u_j(t)\nu_j(t) \quad (20)$$

$$\psi_a(t) \triangleq \left[\ v_1^2(t) \ \cdots \ v_n^2(t) \ -1\ \right]^T, \quad (21)$$

$$\theta_a(t) \triangleq \left[\ a^T \ \sum_{j=1}^{n} a_j \nu_j^2(t)\ \right]^T, \quad (22)$$

respectively, where $a^T \triangleq [\ a_1 \ \cdots \ a_n\ ]$. Given an arbitrary estimate $\bar{\theta}_a$ and $N_{g(\tau)} \triangleq N - g(\tau)$, consider a criterion

$$J^2(\bar{\theta}_a, \tau, N) = \left(\frac{1}{N_\tau}\left(Y_0 - \Psi_0\bar{\theta}_a\right)^T\left(Y_\tau - \Psi_\tau\bar{\theta}_a\right)\right)^2, \quad (23)$$

where Y_0, Y_τ, Ψ_0 and Ψ_τ are defined by

$$Y_0 \triangleq \begin{bmatrix} y(N) \\ y(N_1) \\ \vdots \\ y(1+\tau) \end{bmatrix}, \quad Y_\tau \triangleq \begin{bmatrix} y(N_\tau) \\ y(N_{\tau+1}) \\ \vdots \\ y(1) \end{bmatrix}, \qquad (24)$$

$$\Psi_0 \triangleq \begin{bmatrix} \psi_a^T(N) \\ \psi_a^T(N_1) \\ \vdots \\ \psi_a^T(1+\tau) \end{bmatrix}, \quad \Psi_\tau \triangleq \begin{bmatrix} \psi_a^T(N_\tau) \\ \psi_a^T(N_{\tau+1}) \\ \vdots \\ \psi_a^T(1) \end{bmatrix}. \qquad (25)$$

The necessary and sufficient condition to minimize (23) with respect to $\bar{\theta}_a$ yields the least-correlation estimate [4]

$$\hat{\theta}_a(\tau, N) = \left(\Psi_{0/\tau}^T \Psi_{\tau/0}\right)^{-1} \Psi_{0/\tau}^T Y_{\tau/0}, \qquad (26)$$

where the relevant matrices and vectors are composed of

$$\Psi_{0/\tau} \triangleq \begin{bmatrix} \Psi_0 \\ \Psi_\tau \end{bmatrix}, \Psi_{\tau/0} \triangleq \begin{bmatrix} \Psi_\tau \\ \Psi_0 \end{bmatrix}, Y_{\tau/0} \triangleq \begin{bmatrix} Y_\tau \\ Y_0 \end{bmatrix}. \qquad (27)$$

The matrix $\Psi_{0/\tau}^T \Psi_{\tau/0}$ has full rank since each component of $v(t)$ is independent and is not constant owing to A4. The estimate (26) has following property.

Theorem 2. Suppose that A1-A4 are satisfied. Then as the number of samples increases to infinity, the least-correlation estimate (26) for the system (19)-(22) converges to the expectation of the extended parameter vector, that is,

$$\lim_{N\to\infty} \hat{\theta}_a(\tau, N) = \bar{E}[\theta_a(t)] = \begin{bmatrix} a^T & \sum_{j=1}^n a_j \sigma_{\nu_j}^2 \end{bmatrix}^T, \qquad (28)$$

where $\sigma_{\nu_j}^2$ is the variance of $\nu_j(t)$.

Proof. Refer to Appendix A. □

Theorem 2 addresses the consistency of the extended estimate in the sense that the first n estimates of $\hat{\theta}_a(\tau, N)$ converge to the true parameters θ as N goes to infinity. Above results for the least-correlation estimate can be extended to multi-input multi-output systems just as the extension of least-squares estimate [3, pp.97-98].

B. Third-order Nonlinear Systems

For the third-order nonlinear system

$$z(t) = \sum_{j=1}^n a_j u_j^3(t) + \eta_1(t), \qquad (29)$$

the errors-in-variables models can be written as

$$y(t) = \sum_{j=1}^n a_j \left(v_j(t) - \nu_j(t)\right)^3 + \eta(t). \qquad (30)$$

Rearranging (30) for the even-order terms of $\nu_j(t)$ to be with $v_j(t)$ and for the odd-order terms of $\nu_j(t)$ to be with $u_j(t)$ yields the linear regression model (19) with

$$e(t) = \eta(t) - \sum_{j=1}^n a_j \left(3u_j^2(t)\nu_j(t) - 2\nu_j^3(t)\right), \quad (31)$$

$$\psi_a(t) = \begin{bmatrix} v_1^3(t) & \cdots & v_n^3(t) & -3v^T(t) \end{bmatrix}^T, \quad (32)$$

$$\theta_a(t) = \begin{bmatrix} a^T & a_1\nu_1^2(t) & \cdots & a_n\nu_n^2(t) \end{bmatrix}^T. \quad (33)$$

The method of least correlation applied to this model gives an estimate with following property.

Theorem 3. Suppose that A1-A4 are satisfied. Then as the number of samples goes to infinity, the least-correlation estimate $\hat{\theta}_a(\tau, N)$ applied for the system (19) with (31)-(33) converges to the expectation of the extended parameter vector $\theta_a(t)$ in (33), that is,

$$\lim_{N\to\infty} \hat{\theta}_a(\tau, N) = \begin{bmatrix} a^T & a_1\sigma_{\nu_1}^2 & \cdots & a_n\sigma_{\nu_n}^2 \end{bmatrix}^T, \quad (34)$$

where $\sigma_{\nu_j}^2$ is the variance of $\nu_j(t)$.

Proof. Refer to Appendix B. □

C. Higher-order Nonlinear Systems

Consider the higher-order nonlinear systems

$$z(t) = \sum_{j=1}^n a_j u_j^L(t) + \eta_1(t), \quad L \geq 4. \qquad (35)$$

The errors-in-variables model of (35)

$$y(t) = \sum_{j=1}^n a_j \left(v_j(t) - \nu_j(t)\right)^L + \eta(t) \qquad (36)$$

can be rearranged for even L to

$$\begin{aligned} y(t) &= \sum_{j=1}^n a_j \left[\beta_0 v_j^L(t) + \beta_2 v_j^{L-2}(t)\, \nu_j^2(t) + \cdots \right. \\ &\quad \left. + \beta_{L-2} v_j^2(t) \nu_j^{L-2}(t) + \beta_L \nu_j^L(t)\right] + e(t), \quad (37) \\ e(t) &= \sum_{j=1}^n a_j \left[\beta_1 u_j^{L-1}(t)\nu_j(t) + \beta_3 u_j^{L-3}(t)\nu_j^3(t) + \cdots \right. \\ &\quad \left. + \beta_{L-1} u_j(t) \nu_j^{L-1}(t)\right] + \eta(t), \quad (38) \end{aligned}$$

and for odd L to

$$\begin{aligned} y(t) &= \sum_{j=1}^n a_j \left[\beta_0 v_j^L(t) + \beta_2 v_j^{L-2}(t)\nu_j^2(t) + \cdots \right. \\ &\quad \left. + \beta_{L-1} v_j(t) \nu_j^{L-1}(t)\right] + e(t), \quad (39) \\ e(t) &= \sum_{j=1}^n a_j \left[\beta_1 u_j^{L-1}(t)\nu_j(t) + \beta_3 u_j^{L-3}(t)\nu_j^3(t) + \cdots \right. \\ &\quad \left. + \beta_L \nu_j^L(t)\right] + \eta(t). \quad (40) \end{aligned}$$

Table I shows the coefficients β_ℓ in (37)-(40) for the nonlinear systems with up to 8th-order nonlinearity.

TABLE I

EXAMPLES OF THE COEFFICIENTS $\beta_0 = 1, \beta_\ell, \ell = 1, \cdots, 8$

L	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
1	-1							
2	-2	-1						
3	-3	-3	2					
4	-4	-6	8	5				
5	-5	-10	20	25	-16			
6	-6	-15	40	75	-96	-61		
7	-7	-21	70	175	-336	-427	272	
8	-8	-28	112	350	-896	-1708	2176	1385

Each of (37) and (39) is expressed as the linear regression model (19) with the error (38) and the augmented vectors

$$\begin{aligned} \psi_a^T(t) &= \begin{bmatrix} v_{(L)}'^T & v_{(L-2)}'^T & \cdots & v_{(2)}'^T & \beta_L \end{bmatrix}, \quad (41) \\ \theta_a^T(t) &= \begin{bmatrix} \nu_{(0)}'^T & \nu_{(2)}'^T & \cdots & \nu_{(L-2)}'^T & \gamma_L(t) \end{bmatrix} (42) \end{aligned}$$

for even L, and with the error (40) and the augmented vectors

$$\psi_a^T(t) = \left[\begin{array}{cccc} v_{(L)}^{'T} & v_{(L-2)}^{'T} & \cdots & v_{(1)}^{'T} \end{array} \right], \quad (43)$$

$$\theta_a^T(t) = \left[\begin{array}{cccc} \nu_{(0)}^{'T} & \nu_{(2)}^{'T} & \cdots & \nu_{(L-1)}^{'T} \end{array} \right] \quad (44)$$

for odd L, respectively, where $\gamma_L(t) \triangleq \sum_{j=1}^{n} a_j \nu_j^L(t)$. Each component of $v'_{(\ell)}(t)$, $\nu'_{(\ell)}(t)$, $\ell = 0, 1, \cdots, L$ in (41)-(44) is defined by

$$v_{(\ell)}^{'T}(t) \triangleq \beta_{L-\ell} [\, v_1^\ell(t) \;\cdots\; v_n^\ell(t) \,],$$
$$\nu_{(\ell)}^{'T}(t) \triangleq [\, a_1 \nu_1^\ell(t) \;\cdots\; a_n \nu_n^\ell(t) \,].$$

For (19) with either (38) and (41)-(42) or (40) and (43)-(44), applying the method of least correlation yields the estimate $\hat{\theta}_a(\tau, N)$. But the estimates applied to the nonlinear systems described by fourth or higher-order polynomials are not free from bias even if the number of samples are sufficiently large. We can figure it out as follows.

$$\lim_{N \to \infty} \hat{\theta}_a(\tau, N) = E[\theta_a(t)] + R_{\psi_a \psi_a}^{-1}(\tau) \mathcal{E}(t, t-\tau), \quad (45)$$

where each component of $\mathcal{E}(t, t-\tau)$ is given by (77)-(78) for even L and by (81) for odd L in Appendix C.

IV. Numerical Example

Consider the simple nonlinear model

$$z(t) = a u^L(t), \; L = 2, 3, 4, \quad (46)$$
$$u(t) = \sqrt{2} \sin 2\pi t, \quad (47)$$

and the measurements model (3)-(4). For simplicity, we assume that $\eta(t) = 0$ and $\nu(t)$ is white Gaussian with variance σ_ν^2. Each simulation chooses σ_ν^2 such that the signal-to-noise ratio, defined by

$$\text{SNR}_i = 10 \log_{10} \left(\bar{E}\left[u^2(t)\right] / E\left[\nu^2(t)\right] \right) \; [\text{dB}] \quad (48)$$

is realized.

TABLE II

Effect of the number of samples, $L = 2 : \tau = 1$, $\text{SNR}_i = 5\text{dB}$

Least Correlation Estimates		
N	$\breve{a}$ [%]	$\breve{\sigma}_\nu^2$ [%]
10^3	2.0 ± 13	8.4 ± 52
10^4	0.4 ± 3.4	0.9 ± 14
10^5	-0.1 ± 1.1	-0.2 ± 4.6
10^6	-0.0 ± 0.4	-0.2 ± 1.4

TABLE III

Input-noise effects, $L = 2 : 10^5$ samples, $\tau = 1$

	Least Correlation		Least Squares
SNR_i	$\breve{a}$ [%]	$\breve{\sigma}_\nu^2$ [%]	$\breve{a}$ [%]
20 dB	0.0 ± 0.1	0.3 ± 7.4	-7.5 ± 0.1
10	0.0 ± 0.4	0.0 ± 4.3	-46 ± 0.2
5	0.1 ± 1.5	0.1 ± 5.6	-75 ± 0.4
0	0.4 ± 4.4	0.7 ± 8.9	-92 ± 0.1
($N = 10^6$)	(-0.1 ± 1.4)	(-0.3 ± 2.8)	(-92 ± 0.0)

Table II - Table V show the simulation results from 100 Monte Carlo runs for each case. The estimation error in each table is defined by

$$\breve{a} = \frac{1}{|a|} \left\{ \bar{\tilde{a}} \pm \bar{\sigma}(\tilde{a}) \right\}, \quad (49)$$

where $\bar{\tilde{a}}$ and $\bar{\sigma}(\tilde{a})$ denote the empirical mean and the empirical standard deviation of the parameter error, respectively.

TABLE IV

Input-noise effect, $L = 3 : 10^5$ samples, $\tau = 1$

	Least Correlation		Least Squares
SNR_i	$\breve{a}$ [%]	$\breve{\sigma}_\nu^2$ [%]	$\breve{a}$ [%]
25 dB	0.0 ± 0.1	-4.1 ± 17	-11 ± 0.1
20	0.0 ± 0.3	0.4 ± 13	-29 ± 0.2
15	0.0 ± 0.7	0.6 ± 10	-62 ± 0.2
10	0.2 ± 2.1	1.1 ± 12	-96 ± 0.2
5	1.3 ± 8.7	3.2 ± 22	-107 ± 0.1
($N = 10^6$)	(-0.0 ± 2.4)	(-0.1 ± 6.1)	(-107 ± 0.0)

TABLE V

Input-noise effect, $L = 4 : 10^5$ samples, $\tau = 1$

	Least Correlation		Least Squares
SNR_i	$\breve{a}$ [%]	$\breve{\sigma}_\nu^2$ [%]	$\breve{a}$ [%]
30 dB	0.0 ± 0.2	-4.2 ± 50	-12 ± 0.2
25	0.0 ± 0.4	-2.8 ± 40	-34 ± 0.3
20	0.1 ± 0.9	2.1 ± 29	-76 ± 0.4
15	0.2 ± 2.9	1.8 ± 31	-120 ± 0.3
10	-0.5 ± 12	-2.3 ± 51	-127 ± 0.1
($N = 10^6$)	(0.0 ± 4.1)	(0.0 ± 17)	(-127 ± 0.1)

The analysis in Section III says that the least-correlation estimate $\hat{\theta}_a(\tau)$ converges to the expectation of the extended parameter vector $\theta_a(t)$ without bias as the number of samples increases. Table II shows clear trend that both of the mean and the variance of the estimation error decrease as the number of data increases, which supports Theorem 2 without any exception. Observations from Table III - Table V give in general a confidence that the least-correlation estimate with the augmented regressors and the extended parameters works well for the errors-in-variables nonlinear models. Another thing observed on the tables is that the larger the noise power is the more data the estimate needs, which is not an unexpected results when we remind the effect of the number of samples in Table II. The proposed estimation method is not free from bias when it is employed for fourth or higher-order nonlinearity, but Table V gives a hint that the estimate can generate quite good results. Moreover, we can figure out the bias by (45) if the statistical property of noise are known a priori.

V. Concluding Remarks

We extend the method of least correlation [4], devised for errors-in-variables linear models, to the estimate for nonlinear models described by second or higher-order polynomials. The introduced way of augmenting regressors and extending parameters let the least-correlation estimate work on such a kind of nonlinear systems. Analysis shows that for second and third-order polynomial nonlinear systems the estimate with the augmented regressor vector and the expanded parameter vector converges to the mathematical expectation of the expanded parameter vector as the number of samples goes to infinity. That is, the extended least-correlation estimates applied to 2nd or 3rd-order nonlinear systems are consistent in the sense that the first n estimates of $\hat{\theta}_a(\tau, N)$ converge to the true parameters θ as the number of samples goes to infinity. We also evaluate the misfit for the estimate to be applied to fourth or higher-order nonlinear systems. Monte Carlo simulations for simple examples support the analytical results clearly. Applications of the extension method to Volterra models, Hammerstein systems and Wiener nonlinear models are mentioned.

The recursive least-correlation (RLC) algorithm [4], an equivalent realization of the least-correlation estimate, can be employed directly for the estimate with the augmented regressors and the extended parameters. Optimal input design problem for the best estimates will be one of the further works for applications.

APPENDIX

A. Proof of Theorem 2

With the empirical correlations defined by

$$\bar{R}_{\psi_a\psi_a}(t_1,t_2,N) \triangleq \frac{1}{N_\tau}\sum_{t=1+\tau}^{N}\psi_a(t_1)\psi_a^T(t_2), \quad (50)$$

$$\bar{r}_{\psi_a y}(t_1,t_2,N) \triangleq \frac{1}{N_\tau}\sum_{t=1+\tau}^{N}\psi_a(t_1)y(t_2), \quad (51)$$

where $\tau \triangleq |t_1-t_2|$ and $t_1=t, t_2=t-\tau$ or $t_1=t-\tau, t_2=t$, the estimate (26) is rewritten as

$$\begin{aligned}\hat{\theta}_a(\tau,N) &= \left\{\bar{R}_{\psi_a\psi_a}(t,t-\tau,N)+\bar{R}_{\psi_a\psi_a}(t-\tau,t,N)\right\}^{-1}\\ &\times \left\{\bar{r}_{\psi_a y}(t,t-\tau,N)+\bar{r}_{\psi_a y}(t-\tau,t,N)\right\}. \quad (52)\end{aligned}$$

Using (19) to (51) gives

$$\begin{aligned}\bar{r}_{\psi_a y}(t_1,t_2,N) &\triangleq \bar{\mathbf{t}}_{\psi_a\psi_a\theta_a}(t_1,t_2,t_2,N)\\ &+\bar{r}_{\psi_a e}(t_1,t_2,N), \quad (53)\end{aligned}$$

where the empirical bicorrelation $\bar{\mathbf{t}}_{\psi_a\psi_a\theta_a}(t_1,t_2,t_2,N)$ [5] and the empirical correlation $\bar{r}_{\psi_a e}(t_1,t_2,N)$ are defined by

$$\bar{\mathbf{t}}_{\psi_a\psi_a\theta_a}(t_1,t_2,t_2,N) \triangleq \frac{1}{N_\tau}\sum_{t=1+\tau}^{N}\psi_a(t_1)\psi_a^T(t_2)\theta_a(t_2),$$

$$\bar{r}_{\psi_a e}(t_1,t_2,N) \triangleq \frac{1}{N_\tau}\sum_{t=1+\tau}^{N}\psi_a(t_1)e(t_2),$$

respectively, and the others are given similarly.

Suppose that N increases to infinity. Then $\bar{R}_{\psi_a\psi_a}(t_1,t_2,N)$ converges to the corresponding mathematical correlation $R_{\psi_a\psi_a}(t_1-t_2)$ due to the ergodic theory [7, Theorem 2.3 in p.43], that is,

$$\lim_{N\to\infty}\frac{1}{N_\tau}\sum_{t=1+\tau}^{N}\psi_a(t_1)\psi_a^T(t_2) = \bar{E}\left[\psi_a(t_1)\psi_a^T(t_2)\right] \quad (54)$$

or equivalently

$$\lim_{N\to\infty}\bar{R}_{\psi_a\psi_a}(t_1,t_2,N) = R_{\psi_a\psi_a}(t_1-t_2). \quad (55)$$

Moreover, the correlation depends on the absolute difference of time, $\tau=|t_1-t_2|$, owing to A2, that is

$$R_{\psi_a\psi_a}(t_1-t_2) = R_{\psi_a\psi_a}(t_2-t_1) = R_{\psi_a\psi_a}(\tau). \quad (56)$$

Componentwise expression of $R_{\psi_a\psi_a}(\tau)$ is

$$R_{\psi_a\psi_a}(\tau) = \bar{E}\begin{bmatrix} v_{(2)}(t)v_{(2)}^T(t-\tau) & -v_{(2)}(t)\\ -v_{(2)}^T(t-\tau) & 1\end{bmatrix}, \quad (57)$$

where $v_{(\ell)}^T(t) \triangleq [\ v_1^\ell(t)\ \cdots\ v_n^\ell(t)\ ], \ell=2$.

Applying above approach to (53) gives

$$\begin{aligned}r_{\psi_a y}(\tau) &= \mathbf{t}_{\psi_a\psi_a\theta_a}(\tau,\tau)+r_{\psi_a e}(\tau), \quad (58)\\ r_{\psi_a y}(-\tau) &= \mathbf{t}_{\psi_a\psi_a\theta_a}(-\tau,-\tau)+r_{\psi_a e}(-\tau). \quad (59)\end{aligned}$$

Each term of (58)-(59) is evaluated as follows:

$$\begin{aligned}\mathbf{t}_{\psi_a\psi_a\theta_a}(\tau,\tau) &= R_{\psi_a\psi_a}(\tau)E[\theta_a(t)], \quad (60)\\ \mathbf{t}_{\psi_a\psi_a\theta_a}(-\tau,-\tau) &= R_{\psi_a\psi_a}(-\tau)E[\theta_a(t-\tau)]\\ &= R_{\psi_a\psi_a}(\tau)E[\theta_a(t)], \quad (61)\\ r_{\psi_a e}(\tau) &= \bar{E}[\psi_a(t)\eta(t-\tau)]\\ &- \sum_{j=1}^{n}a_j\bar{E}[\psi_a(t)u_j(t-\tau)\nu_j(t-\tau)]\\ &= 0, \quad (62)\\ r_{\psi_a e}(-\tau) &= \bar{E}[\psi_a(t-\tau)\eta(t)]\\ &- \sum_{j=1}^{n}a_j\bar{E}[\psi_a(t-\tau)u_j(t)\nu_j(t)]\\ &= 0. \quad (63)\end{aligned}$$

Substituting (60)-(63) into (58)-(59) yields

$$r_{\psi_a y}(\tau) = R_{\psi_a\psi_a}(\tau)E[\theta_a(t)] = r_{\psi_a y}(-\tau). \quad (64)$$

Finally as the number of samples goes to infinity, the least-correlation estimate $\hat{\theta}_a(\tau,N)$ can be expressed as

$$\lim_{N\to\infty}\hat{\theta}_a(\tau,N) = R_{\psi_a\psi_a}^{-1}(\tau)r_{\psi_a y}(\tau), \quad (65)$$

and applying (56) and (64) to (65) yields (28). □

B. Proof of Theorem 3

Equations (50)-(56) in Appendix A work on the third-order model (19) with (31)-(33) as they are. For this third-order model, $R_{\psi_a\psi_a}(\tau)$ is expressed as

$$R_{\psi_a\psi_a}(\tau) = \begin{bmatrix} R_{v^3v^3}(\tau) & -3R_{v^3v^1}(\tau)\\ -3R_{v^1v^3}(\tau) & 9R_{v^1v^1}(\tau)\end{bmatrix}, \quad (66)$$

where the sub-matrices are given by

$$\begin{aligned}R_{v^\ell v^k}(\tau) &\triangleq \bar{E}[v_{(\ell)}(t)v_{(k)}^T(t-\tau)], \quad (67)\\ v_{(\ell)}^T(t) &\triangleq [\ v_1^\ell(t)\ \cdots\ v_n^\ell(t)\ ] \quad (68)\end{aligned}$$

for $\ell,k=1,3$. Similarly the bicorrelation $\mathbf{t}_{\psi_a\psi_a\theta_a}(\tau,\tau)$ is written as

$$\mathbf{t}_{\psi_a\psi_a\theta_a}(\tau,\tau) = R_{\psi_a\psi_a}(\tau)\bar{E}[\theta_a(t)] \quad (69)$$

and $r_{\psi_a e}(\tau)$ is evaluated as

$$\begin{aligned}r_{\psi_a e}(\tau) &= \bar{E}[\psi_a(t)\eta(t-\tau)]\\ &- 3\bar{E}\left[\psi_a(t)\sum_{j=1}^{n}a_ju_j^2(t-\tau)\nu_j(t-\tau)\right]\\ &+ 2\bar{E}\left[\psi_a(t)\sum_{j=1}^{n}a_j\nu_j^3(t-\tau)\right]\\ &= 0. \quad (70)\end{aligned}$$

From above results, $r_{\psi_a y}(\tau)$ is evaluated as

$$r_{\psi_a y}(\tau) = R_{\psi_a\psi_a}(\tau)E[\theta_a(t)] \quad (71)$$

by using (69)-(70). Finally applying (71) to

$$\lim_{N\to\infty} \hat{\theta}_a(\tau, N) = R^{-1}_{\psi_a\psi_a}(\tau) r_{\psi_a y}(\tau) \quad (72)$$

yields (28). □

C. Proof of Equation (45)

Refer to (50)-(56) in Appendix A. First suppose that L is even. Then $R_{\psi_a\psi_a}(\tau)$ is expressed as

$$R_{\psi_a\psi_a}(\tau) = \begin{bmatrix} R_{v'^L v'^L}(\tau) \cdots R_{v'^L v'^2}(\tau) & r_{v'^L} \\ \vdots \quad \cdots \quad \vdots & \vdots \\ R_{v'^2 v'^L}(\tau) \cdots R_{v'^2 v'^2}(\tau) & r_{v'^2} \\ r_{v'^L} \quad \cdots \quad r_{v'^2} & \beta_L^2 \end{bmatrix}, \quad (73)$$

where each component matrix or vector is given by

$$R_{v'^\ell v'^k}(\tau) \triangleq \bar{E}[v'_{(\ell)}(t) v'^T_{(k)}(t-\tau)], \quad (74)$$

$$r_{v'^\ell} \triangleq \beta_L \beta_{L-\ell} \bar{E}[v_{(\ell)}(t)] = \beta_L \beta_{L-\ell} \bar{E}[v_{(\ell)}(t-\tau)], \quad (75)$$

for $\ell, k \in \{L, L-2, \cdots, 2\}$. The bicorrelation $\mathbf{t}_{\psi_a\psi_a\theta_a}(\tau,\tau)$ is expressed as

$$\mathbf{t}_{\psi_a\psi_a\theta_a}(\tau,\tau) = R_{\psi_a\psi_a}(\tau)\bar{E}[\theta_a(t)] + \mathcal{E}(t, t-\tau) \quad (76)$$

where the first and kth component of

$$\mathcal{E}(t, t-\tau) = [\ \mathcal{E}_0 \quad \mathcal{E}_2 \quad \cdots \quad \mathcal{E}_{L-2} \quad \mathcal{E}_L\ ]$$

is evaluated as

$$\mathcal{E}_0 = \sum_{\ell=\mathcal{L}_e} \beta_L \Big(\bar{E}[v'^T_{(\ell)}(t-\tau) \nu'_{(L-\ell)}(t-\tau)] - r_{v'^\ell} E[\nu'_{(L-\ell)}(t)] \Big), \quad (77)$$

$$\mathcal{E}_k = \sum_{\ell=\mathcal{L}_e} \Big(\bar{E}[v'_{(k)}(t) v'^T_{(\ell)}(t-\tau) \nu'_{(L-\ell)}(t-\tau)] - R_{v'^L v'^\ell}(\tau) E[\nu'_{(L-\ell)}(t)] \Big), \quad (78)$$

respectively, for $\mathcal{L}_e = \{L-2, \cdots, 2\}$. The cross-correlation $r_{\psi_a e}(\tau)$ is evaluated as

$$\begin{aligned} r_{\psi_a e}(\tau) &= \bar{E}[\psi_a(t)\eta(t-\tau)] \\ &+ \beta_1 \bar{E}\left[\psi_a(t) \sum_{j=1}^{n} a_j u_j^{L-1}(t-\tau) \nu_j(t-\tau)\right] \\ &\vdots \\ &+ \beta_{L-1} \bar{E}\left[\psi_a(t) \sum_{j=1}^{n} a_j u_j(t-\tau) \nu_j^{L-1}(t-\tau)\right] \\ &= 0. \end{aligned} \quad (79)$$

For odd L, the correlation is written as

$$R_{\psi_a\psi_a}(\tau) = \begin{bmatrix} R_{v'^L v'^L}(\tau) \cdots R_{v'^L v'^1}(\tau) \\ \vdots \quad \cdots \quad \vdots \\ R_{v'^1 v'^L}(\tau) \cdots R_{v'^1 v'^1}(\tau) \end{bmatrix} \quad (80)$$

and the kth component of

$$\mathcal{E}(t, t-\tau) = [\ \mathcal{E}_1 \quad \mathcal{E}_3 \quad \cdots \quad \mathcal{E}_{L-2} \quad \mathcal{E}_L\ ]$$

is evaluated as

$$\mathcal{E}_k = \sum_{\ell=\mathcal{L}_o} \Big(\bar{E}[v'_{(k)}(t) v'^T_{(\ell)}(t-\tau) \nu'_{(L-\ell)}(t-\tau)] - R_{v'^L v'^\ell}(\tau) E[\nu'_{(L-\ell)}(t)] \Big), \quad (81)$$

for $\mathcal{L}_o = \{L-2, \cdots, 3\}$, and

$$\begin{aligned} r_{\psi_a e}(\tau) &= \bar{E}[\psi_a(t)\eta(t-\tau)] \\ &+ \beta_1 \bar{E}\left[\psi_a(t) \sum_{j=1}^{n} a_j u_j^{L-1}(t-\tau) \nu_j(t-\tau)\right] \\ &\vdots \\ &+ \beta_L \bar{E}\left[\psi_a(t) \sum_{j=1}^{n} a_j \nu_j^{L}(t-\tau)\right] \\ &= 0. \end{aligned} \quad (82)$$

Employing (76) with either (77)-(79) or (81)-(82) to (58), and applying the results to (72) yields (45). □

References

[1] F. J. Doyle III, R. K. Pearson, and B. A. Qgunnike, *Identification and Control Using Volterra Models*. Springer, 2002.

[2] D. Hermey and G. A. Watson, "Fitting Data with Errors in Variables Using the Huber M-estimator," *SIAM J. Sci. Comput.*, Vol. 20, pp. 1276-1298, 1999.

[3] R. Johansson, *System Modeling and Identification*, Prentice-Hall, New Jersey, 1993.

[4] B. E. Jun and D. S. Bernstein, "Least-Correlations Estimates for Errors-in-Variables Models," IECON 2004.

[5] T. Koh and E. J. Powers, "Second-Order Volterra Filtering and Its Application to Nonlinear System Identification," *IEEE Trans. Acoustics, Speech, and Signal Processing*, Vol. ASSP-33, No. 6, pp. 1445-1455, Dec. 1985.

[6] T. Li, "Robust and Consistent Estimation of Nonlinear Errors-in-Variables Models," *Journal of Econometrices*, Vol. 110, pp. 1-26, 2002.

[7] L. Ljung, *System Identification - Theory for the User, 2nd Edition*, Prentice-Hall, New Jersey, 1999.

[8] W. Scherrer and M. Deistler, "A Structure Theory for Linear Dynamic Errors-in-Variables Models," *SIAM J. Control Optimization*, Vol. 36, No. 6, pp. 2148-2175, Nov. 1998.

[9] D. Mattera, "Identification of Polyperiodic Volterra Systems by means of Input-output Noisy Measurements," *Signal Processing*, Vol. 75, pp. 41-50, 1999.

[10] D. Mattera, "Higher-order Cyclostationary-based Methods for Identifying Volterra Systems by Input-output Noisy Measurements," *Signal Processing*, Vol. 67, pp. 77-98, 1998.

[11] T. Soderstrom, U. Soverini, and K. Mahata, "Perspectives on Errors-in-variables Estimation for Dynamic Systems," *Signal Processing*, Vol. 82, pp. 1139-1154, 2002.

[12] A. H. Tan and K. Godfrey, "Identification of Wiener-Hammerstein Models Using Linear Interpolation in the Frequency Domain (LIFED)," *IEEE Trans. Instrumentation and Measurement*, Vol. 51, No. 3, pp. 509-521, June 2002.

[13] I. Vajk and J. Hetthessy, "Identification of Nonlinear Errors-in-Variables Models," *Automatica*, Vol. 39, pp. 2099-2107, 2003.

[14] G. Vandersteen, Y. Rolain, J. Schoukens, and P. Pintelon, "On the Use of System Identification for Accurate Parametric Modeling of Nonlinear Systems Using Noisy Measurements," *IEEE Trans. Instrumentation and Measurement*, Vol. 45, pp. 605-609, 1996.

[15] S. van Huffel and P. Lemmerling (Ed.), *Total Least Squares Techniquea and Errors-in-Variables Modeling: Analysis, Algorithms and Applications*, Kluwer Academic Publisher, 2002.