APPLICATION OF INSTRUMENTAL VARIABLE METHOD TO THE IDENTIFICATION OF HAMMERSTEIN-WIENER SYSTEMS

GRZEGORZ MZYK†

†Wrocław University of Technology, Institute of Engineering Cybernetics, ul. Janiszewskiego 11/17, 50-372 Wrocław, Poland, grmz@diuna.ict.pwr.wroc.pl

Abstract. Application of least squares and instrumental variables to recovering parameters of nonlinear complex dynamic block-oriented systems is examined. For a system with the Hammerstein-Wiener structure the instrumental variable algorithm is designed and compared with the least squares algorithm for estimating system parameters. The advantages of the proposed instrumental variable estimator are discussed and in particular its weak consistency, even in the presence of correlated noise, is shown. The problem of generating optimal values of instrumental variables is analysed, rate of convergence of the proposed estimator is evaluated and simulation examples are included. The paper provides an extension of the results introduced in [1].

Key Words. System identification, complex systems, parameter estimation, nonlinear models.

1. INTRODUCTION

Instrumental variable method is one of the most popular and universal methods used for determining parameters of single-element linear dynamic objects. The main aim of this paper is applying of the instrumental variable method to identify parameters of nonlinear dynamic systems with composite structure, and comparing it with the traditional least squares method. Applicability analysis is made on the example of the system with Hammerstein-Wiener structure described in detail in section 2. The Hammerstein-Wiener system can be viewed as a serial connection of two typical cascade structures: the Hammerstein system (Fig. 1) and the Wiener system (Fig. 2). In spite of occurring in many fields ([6],[7],[9],[11],[13]), the Hammerstein-Wiener structure has received little attention in the literature. In section 3 the least squares algorithm proposed recently in [1] for the white noise case is presented and its asymptotic bias is shown in the presence of correlated noise. To reduce the bias, the new estimator based on the instrumental variable method is proposed in section 4. The procedure is based on the ideas presented in [10] and [12] for single-element linear systems. The proposed method is next compared with the least squares algorithm given in [1]. Asymptotic behaviour is studied, and the advantages of the proposed instrumental variable estimator are discussed. In particular its weak consistency, even in the presence of correlated noise, is shown. Also the problem of generating optimal values of instrumental variables is analysed, and rate of convergence of the proposed estimator is evaluated. Comparative simulation study results are presented.

Fig. 1. Hammerstein system.

Fig. 2. Wiener system.
2. STATEMENT OF THE PROBLEM

2.1. System under consideration

We consider a single input - single output asymptotically stable discrete-time nonlinear complex system described by the equation (see [1] and Fig. 3):

\[
y_k = \sum_{j=1}^{p} a_j N_2(y_{k-j}) + \sum_{j=0}^{q} b_j N_1(u_{k-j}) + z_k
\]  

(1)

where \( N_1(u) = \sum_{j=0}^{p} c_j f_j(u) \) and \( N_2(y) = \sum_{j=0}^{q} d_j g_j(y) \).

Signals \( y_k, u_k \) and \( z_k \) are the system output, input and disturbance at time \( k \), respectively. We assume that the input \( u_k \) is bounded, white noise, disturbance \( z_k \) is random, bounded \((|z_k| < z_{\text{max}})\), zero-mean, correlated and independent of \( u_k \). Let us also assume that \( g_0(),...g_p() \) and \( f_0()...f_q() \) are \( a \text{ priori} \) known smooth nonlinear functions and orders \( p, q, n, m \) are known as well. Let

\[
a = (a_1,...,a_p)^T \quad b = (b_0,...,b_q)^T \quad c = (c_1,...,c_m)^T \quad d = (d_1,...,d_q)^T
\]

(2)

denote the unknown system parameter vectors.

System (1) is more general than classical Hammerstein system often met in applications (Fig. 1) and differs from widely discussed in the literature Wiener-Hammerstein system, where two linear dynamical objects surround one static nonlinear element ([2],[3],[4],[5]).

2.2. Identification task

The aim of the identification is to recover the true parameter vectors \( a, b, c \) and \( d \) of the system (given by (2)), using only the overall system input-output measurements \( (u_k, y_k) \) \((k=1...N) \) obtained in the experiment. Note that parametrization of system (1) is not unique. For some nonzero constants \( \alpha \) and \( \beta \) systems with parameter vectors \( a, b, c, d \) and \( \beta a, \beta b, \beta c/\alpha, d/\beta \) are indistinguishable from the input-output point of view (see (1)). To obtain a unique parametrization let us take the following assumptions ([1]): (a) \( \theta_{ad} = \theta_{bc}^T \) and \( \theta_{ab} = \theta_{bc} \) are not both zero; (b) \( |a|_2 = 1 \) and \( |b|_2 = 1 \), where \( ||*||_2 \) denotes the Euclidean norm; (c) signs of the first nonzero elements of \( a \) and \( b \) are positive. Define:

\[
\theta = (b_0 c_1,...,b_q c_m,...,b_0 c_1,...,b_q c_m)^T
\]

(3)

the aggregated parameters of the system, and let \( \phi_k \) be the generalized input (regressor) vector:

\[
\phi_k = (f_1(u_k),...,f_p(u_k),...,f_1(u_{k-n}),...,f_p(u_{k-n})), g_1(y_{k-1}),...,g_q(y_{k-1}),...,g_1(y_{k-p}),...,g_q(y_{k-p}))^T
\]

Equation (1) can be rewritten in the form \( y_k = \theta^T \phi_k + z_k \), and the general measurement equation can be written in the form:

\[
Y_N = \Phi_N \theta + Z_N
\]

(4)

where

\[
Y_N = (y_1,...,y_N)^T, \quad Z_N = (z_1,...,z_N)^T, \quad \Phi_N = (\phi_1,...,\phi_N)^T.
\]

3. LEAST SQUARES IDENTIFICATION

For further comparison with the instrumental variable method we start from presenting of the two stage identification algorithm, based on the least squares method and singular value decomposition (see [1]).

3.1. The algorithm

**Step 1:** Calculate the least squares estimate:

\[
\hat{\theta}^{(LS)} = (\Phi_N^T \Phi_N)^{-1} \Phi_N^T Y_N
\]

(5)
of the aggregated vector $\theta$ and then construct therefrom evaluations $\hat{\theta}_{ad}^{(LS)}$ and $\hat{\theta}_{bc}^{(LS)}$ of $\Theta_{ad}$ and $\Theta_{bc}$ (see assumption a).

**Step 2:** Perform the singular value decomposition of evaluations $\hat{\theta}_{ad}^{(LS)}$ and $\hat{\theta}_{bc}^{(LS)}$:

$$\hat{\theta}_{ad}^{(LS)} = \sum_{i=1}^{\min \{p,n\}} \delta_i \xi_i \xi_i^T$$

and next calculate the estimates:

$$\hat{b} = s_p \hat{\mu}_1, \quad \hat{c} = s_s \sigma_1 \hat{\nu}_1, \quad \hat{a} = s_s \hat{\xi}_1, \quad \hat{d} = s_s \hat{\zeta}_1$$

where $s_p$ and $s_s$ denote the signs of the first nonzero elements in vectors $\mu_1$ and $\xi_1$, respectively.

Singualr value decomposition allows to split aggregated matrix of parameters on product of two parameter vectors. It was shown in [1] that

$$(\hat{\mu}_1, \sigma_1, \hat{\nu}_1) = \arg \min_{c \in \mathbb{R}^{m \times n}, \theta \in \mathbb{R}^n} \| \hat{\theta}_{bc}^{(LS)} - \theta \theta^T \|_F^2$$

**3.2. Limit properties**

The following theorems were proved in [1]:

**T1:** For any $N \to \infty$, if $\Phi_N$ is of full column rank and the disturbance $z_k = 0$, then:

$$\hat{b} = b, \quad \hat{c} = c, \quad \hat{a} = a, \quad \hat{d} = d$$

(7)

**T2:** Let the disturbance $z_k$ be white with zero mean, finite variance and independent of input $u_k$. Let moreover $\Phi$ be persistently exciting. Then if $N \to \infty$:

$$\hat{b} - p.l. \to b, \quad \hat{c} - p.l. \to c,$$

$$\hat{a} - p.l. \to a, \quad \hat{d} - p.l. \to d$$

(8)

Multiplying equation (4) by $\Phi_N^T$ we get:

$$\Phi_N^T Y_N = \Phi_N^T \Phi_N \theta + \Phi_N^T Z_N$$

Hence the estimation error of vector $\theta$ can be written as:

$$\Delta = \hat{\theta}^{(LS)} - \theta = (\Phi_N^T \Phi_N)^{-1} \Phi_N^T Z_N =$$

$$= \left( \frac{1}{N} \sum_{i=1}^N \phi_i \phi_i^T \right)^{-1} \left( \frac{1}{N} \sum_{i=1}^N \phi_i z_i \right)$$

If $z_k$ is white noise with zero mean and finite variance then elements of matrix $Z_N$ are independent of elements in $\Phi_N$ and $\Delta \to 0$ as $N \to \infty$. Otherwise if $z_k$ is correlated, i.e. $E z_k \xi_i, \xi_i \neq 0$ for some $i, \xi_i \neq 0$, then the proposed estimator (5) of vector $\theta$ will be biased asymptotically, because of dependence between the noise $z_k$ and values $g_i(y_{ik})$ ($i=1...p, k=1...n$) appearing in the matrix $\Phi_N$. In consequence, estimators given by equations (6) will also be biased. This conclusion is confirmed by the simulation example reported below.

**3.3. Simulation examples**

The system explored in the experiment had the following parameters:

$$\begin{align*}
\mathbf{b} &= \mathbf{c} = \mathbf{d} = (0.5; 0.5; 0.5; 0.5)^T \\
\mathbf{f}(x) &= \cos tx \\
g(x) &= \sin l x \\
t, t = 1...4
\end{align*}$$

The inputs $u_k$ were generated according to the uniform distribution on the interval [-10;10]. Disturbance $z_k$ was first generated as white noise with the uniform distribution on [-0.01:0.01] (about 1% of the maximum noiseless signal) or [-0.05:0.05] (about 5% of the maximum noiseless signal). Experiment was repeated for $z_k$ transferred by colouring filter with transfer function $F(z) = 1 + z^{-1}$, where $z^{-1}$ denotes one step backward shift operator in time domain. Aggregated estimation error was evaluated from the formula:

$$\text{ERR}_{LS} = \left\| \hat{b}(N), \hat{c}(N), \hat{d}(N), \hat{a}(N) - [b, c, d, a] \right\|_F$$

Experimental results are shown in Fig. 4.

![Fig. 4. Aggregated least squares estimation error versus number of measurements. For correlated noise the least squares estimator is not consistent.](image-url)

**4. INSTRUMENTAL VARIABLE METHOD**

**4.1. Assumptions**

In the following $\lim_{N \to \infty} x_N = x$ for any random vectors $x_N, x \in \mathbb{R}^N$, means that $x_N \to x$ in probability as $N \to \infty$. Assume we have chosen instrumental variable matrix $\Psi_N$, satisfying for the correlated noise $z_k$ the following four conditions:
(C1): \( \dim \Psi_N = \dim \Phi_N \)
(C2): \( \psi_k \) and \( z_k \) are bounded and mutually independent
(C3): matrix \( \Psi_N^T \Phi_N \) is nonsingular with probability 1 for all \( N \)
(C4): \( \text{Plim}(\sqrt{N} \Psi_N^T \Phi_N) \) exists and is nonsingular.

**Lemma 1.** A necessary condition for the existence of an instrumental variable matrix \( \Psi_N \) satisfying (C4) is that the matrix \( \sqrt{N} \Phi_N \Phi_N \) be nonsingular, i.e. \( \phi_k \) must be persistently exciting (see Appendix A).

### 4.2. Instrumental variable algorithm

After multiplying the left hand side of equation (4) by matrix \( \Psi_N^T \) we get:

\[ \Psi_N^T Y_N = \Psi_N^T \Phi_N \theta + \Psi_N^T Z_N \]

Taking account of the conditions (C2) and (C3) above, we postulate to replace the least squares estimator (5) evaluated in step 1 by the following instrumental variable estimator:

\[ \hat{\theta}^{(IV)} = \left( \Psi_N^T \Phi_N \right)^{-1} \Psi_N^T Y_N \quad (9) \]

where \( \Psi_N = (\psi_1, \ldots, \psi_n)^T \). At the second stage we proceed analogically, by making singular value decomposition of evaluations \( \hat{\theta}_{ad}^{(IV)} \) and \( \hat{\theta}_{ad}^{(IV)} \) of \( \Theta_{ad} \) and \( \Theta_\phi \) based on \( \hat{\theta}^{(IV)} \).

### 4.3. Limit properties

Estimation error for the algorithm (9) has the form:

\[ \Delta^{(IV)} = \hat{\theta}^{(IV)} - \theta = \left( \Psi_N^T \Phi_N \right)^{-1} \Psi_N^T Z_N \quad (10) \]

**Theorem 1.** Under conditions (C1)-(C4), estimator (9) is weakly consistent, independently of the noise correlation structure, i.e.

\[ P \lim (\Delta^{(IV)}) = 0 \]

(see Appendix B).

**Theorem 2.** Estimation error \( \Delta^{(IV)} \) has the rate of convergence of order \( \frac{1}{\sqrt{N}} \) in probability, for all choices of instruments (see Appendix C).

### 4.4. Optimal instruments

Denote

\[ \tilde{\phi}_k = (f_1(u_k), \ldots, f_m(u_k), \ldots, f_1(u_{k-m}), \ldots, f_m(u_{k-m})) \nu \]

\[ R(u_{k-1}), R(u_{k-1}), \ldots, R(u_{k-p}), \ldots, R(u_{k-p})^T \]

where

\[ R_j(u) = E \{ g_j(y_k) | u_k = \theta \} \quad j = 1, \ldots, q \]

are the regression functions. Denote

\[ Z_N^* = \sqrt{\frac{1}{N}} Z_N \]

and let

\[ Q(\Psi_N) = \max \left\| \Delta_N(\Psi_N) \right\|^2 \]

be the instruments quality index, where \( \Delta_N(\Psi_N) \) is defined in (10), and \( \| \| \) is euclidean vector norm.

**Theorem 3.** Under conditions (C1)-(C4), the value of \( Q(\Psi_N) \) (see (12)) is minimal for

\[ \Psi_N = \hat{\Phi}_N = (\hat{\phi}_1, \ldots, \hat{\phi}_N)^T \]

(see Appendix D).

### 5. Simulation Examples

For comparison, algorithms (5) and (9) were examined practically on the example system described in subsection 3.3. Instrumental variables were generated by the standard method:

\[ \psi_k = (f_1(u_k), \ldots, f_m(u_k), \ldots, f_1(u_{k-m}), \ldots, f_m(u_{k-m})) \]

\[ g_1(\hat{y}_{k-1}) \nu, \ldots, g_q(\hat{y}_{k-1}) \nu, \ldots, g_1(\hat{y}_{k-p}) \nu \]

where

\[ \hat{y}_k = \sum_{i=1}^{p} \tilde{a}_i \left( \sum_{i=1}^{p} \tilde{d}_i (\hat{y}_{k-i}) \right) + \sum_{i=1}^{q} \tilde{c}_i \left( \sum_{i=1}^{p} \tilde{c}_i (u_{k-i}) \right) \]

and \( \tilde{a}_i, \tilde{b}_j, \tilde{c}_i, \tilde{d}_i \) denote the elements of vectors \( \tilde{a}, \tilde{b}, \tilde{c}, \tilde{d} \) computed by using least squares method described in section 3. Aggregated estimation errors for the least squares method (ERR_LS) and instrumental variable method (ERR_IV), evaluated as in subsection 3.3, were computed for \( 10^5 \leq N \leq 10^5 \) measurement data and results are shown in Fig. 5. As we see, for \( N \) large - asymptotic behaviour of instrumental variable estimator is better than of least squares estimator. Convergence is rather slow, so the fundamental problem is to select the instruments in an optimal manner.
6. CONCLUSIONS

Computational complexity of the proposed instrumental variable algorithm is comparable with the least squares method. The only difficulty is generation of instruments. The second step is the computation of the singular value decomposition of two matrices with fixed dimensions. Instrumental variable method can be well applied to identification of nonlinear dynamic systems with composite (Hammerstein-Wiener) structure in the presence of correlated noise. This is the main advantage of the method. Notice that, correlation of the noise process can be connected with the presence of structural feedback in the complex system. We refer to [8] where successful application of instrumental variable method to identification of linear static complex systems with feedback is considered.

7. APPENDICES

7.1. Appendix A. Sketch of the proof of Lemma 1

Let $A, B \in R^{\alpha \times \beta}$ be the matrices with finite elements, and let $\det(A^{T}A) = 0$. Then there exist a nonzero vector $\xi \in R^\beta$, such that $A\xi = 0$, thus $B^{T}A\xi = 0$ and $\det(B^{T}A) = 0$.

We get Lemma 1, substituting

$A := \frac{1}{\sqrt{N}} \Phi_N$, $B := \frac{1}{\sqrt{N}} \Psi_N$.

7.2. Appendix B. Sketch of the proof of Theorem 1

Making use of Slutsky theorem [10], one can write

$$P \lim(D^{(IV)}) = P \lim\left(\left(\Psi_{N}^{T} \Phi_{N} \right)^{-1}\right) \cdot P \lim\left(\Psi_{N}^{T} Z_{N}\right)$$

Denote $G_{N} = \Psi_{N}^{T} Z_{N}$, $G_{N}^{i} = \sum_{k=1}^{N} \psi_{i,k} z_{k}$. It is easy to show that

$$E\left[\frac{1}{N} G_{N}^{i}\right] = 0$$

$$\text{var}\left[\frac{1}{N} G_{N}^{i}\right] \leq \frac{2\psi_{\text{max}}^2}{N} \sum_{k=1}^{N} |\omega(k)|$$

where $\omega(k) = E z_{i, z_{i+k}}$ is the noise correlation function.

So if $k \omega(k) \to 0$ we get

$$\lim \text{var}\left[\frac{1}{N} G_{N}^{i}\right] = 0.$$ From Tchebychev inequality:

$$P \lim\left(\frac{1}{N} G_{N}^{i}\right) = P \lim\left(\frac{1}{N} \Psi_{N}^{T} Z_{N}\right) = 0$$

and $P \lim(D^{(IV)}) = 0$.

7.3. Appendix C. Sketch of the proof of Theorem 2

Denote

$$A_{N} = \frac{1}{N} \Psi_{N}^{T} \Phi_{N} = \frac{1}{N} \sum_{k=1}^{N} \psi_{i,k} \phi_{k}^{T}$$

$$B_{N} = \frac{1}{\sqrt{N}} \Psi_{N}^{T} Z_{N} = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \psi_{i,k} z_{k}$$

$$C_{N} = \Delta^{(IV)}$$

$$D_{N} = A_{N}^{-1} B_{N}.$$ Then $C_{N} = \frac{1}{\sqrt{N}} D_{s}$. It can be shown that

$$|A_{N}^{i,j}| < \infty, \ E B_{N}^{i,j} = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} E \psi_{i,k} z_{k} = 0$$

$$\text{var} B_{N}^{i,j} \leq 2 \psi_{\text{max}}^2 \sum_{k=1}^{N} |\omega(k)| < \infty$$

$$E D_{N}^{i,j} < \infty, \ \var(D_{N}^{i,j}) < \infty$$

Thus for all number sequences $\chi_{N}$ such that

$$\lim_{N \to \infty} \chi_{N} = 0$$

it holds

$$\chi_{N} D_{N}^{i,j} = C_{N}^{i,j} \frac{N}{\sqrt{N}} \to 0$$

in probability.
Lemma 2. (see [12]) Let $M_1$ and $M_2$ be matrices with the same dimensions, if exist $(M_1^T M_1)^{-1}$, $(M_2^T M_2)^{-1}$ and $(M_1^T M_1)^{-1}$ then the matrix $D_N = (M_2^T M_1)^{-1} M_2^T M_2 (M_2^T M_2)^{-1} - (M_1^T M_1)^{-1}$ is positive semidefinite, i.e. for all vector $\eta$: $\eta^T D_N \eta \geq 0$.

Making use of Lemma 2 for $M_1 = \frac{1}{\sqrt{N}} \Phi_N$, $M_2 = \frac{1}{\sqrt{N}} \Psi_N$ and $\eta = Z_N^T$ we get (with probability 1):

$Q(\Psi_N) = \max \frac{\left\| \Delta_N (\Psi_N) \Delta_N (\Psi_N) \right\|^2}{\left\| Z_N \right\|^2}$

so $Q(\Psi_N) = \max \frac{\left\| \Delta_N (\Psi_N) \right\|^2}{\left\| Z_N \right\|^2}$ attains its lower bound for the choice $\Psi_N = \hat{\Phi}_N$.

8. REFERENCES


