SYSTEM IDENTIFICATION FROM NOISY MEASUREMENTS BY USING INSTRUMENTAL VARIABLES AND SUBSPACE FITTING

Mats Cedervall

Systems and Control Group
Uppsala University, P. O. Box 27
S–751 03 Uppsala, Sweden

Petre Stoica

Department of Applied Electronics
Chalmers University of Technology
S–412 96 Gothenburg, Sweden

ABSTRACT

This paper considers the estimation of the parameters of a linear discrete-time system from noisy input and output measurements. The conditions imposed on the system are quite general. The proposed method makes use of an instrumental variable (IV)-vector whose cross-covariance with the system's regression vector is pre- and post-multiplied by some prechosen weights. The singular vectors of this matrix possess complete information on the system parameters. A weighted subspace fitting (WSF) method is then applied to these singular vectors to consistently estimate the parameters of the system. The proposed method is non-iterative, easy to implement and has a small computational burden. The asymptotic distribution of its estimation errors is derived and the result is used to motivate the choice of the weighting matrix in the WSF step and also to predict the estimation accuracy. A numerical example is included to illustrate the performance.

1. PROBLEM STATEMENT AND SOLUTION

Consider a linear discrete-time system described by the following difference equation ([3, 4, 5, 9]):

$$A(q^{-1})z(t) = B(q^{-1})x(t), \quad t = 1, 2, 3, \ldots \qquad (1)$$

where

$$A(q^{-1}) = 1 + a_1 q^{-1} + \ldots + a_n q^{-n} \qquad (2)$$
$$B(q^{-1}) = b_0 + b_1 q^{-1} + \ldots + b_n q^{-n}. \qquad (3)$$

For the sake of simplicity, the polynomials $A(q^{-1})$ and $B(q^{-1})$, in the unit delay operator q^{-1}, are assumed to have the same degree (which is a minor restriction). It is also assumed that:

This work was supported by the Swedish Research Council of Engineering Sciences under contract 93-669 and by Göran Gustafsson Foundation.

A1. $A(q^{-1})$ has all zeros strictly outside the unit circle, and $A(q^{-1})$ and $B(q^{-1})$ have no common factors.

The noise-corrupted measurements of $x(t)$ and $z(t)$ are denoted by

$$u(t) = x(t) + w(t) \qquad (4)$$
$$y(t) = z(t) + v(t) = \frac{B(q^{-1})}{A(q^{-1})}x(t) + v(t). \qquad (5)$$

All signals appearing above are assumed to be stationary. The following additional assumption is made on the noise sequences in (4) and (5).

A2. The input noise $w(t)$ is finitely auto-correlated and also finitely cross-correlated with $v(t)$, in the following sense:

$$\mathbf{E}[w(t-k)w(t)] = 0 \text{ for } |k| \geq L \qquad (6)$$
$$\mathbf{E}[w(t-k)v(t)] = 0 \text{ for } k \geq L \qquad (7)$$

where L is a given positive integer. Furthermore, both $w(t)$ and $v(t)$ are statistically independent of the noise-free input $x(t)$.

The above noise conditions are more general than those used in most previous works (See e.g. [1, 2]). Note that in **A2** only the input noise is required to be finitely correlated; the output noise $v(t)$ may be arbitrarily correlated.

This paper focuses on the problem of estimating the system parameters $\{a_i, b_i\}$ from samples of $\{u(t), y(t)\}_{t=1}^{N}$.

The following additional notation is required. Let

$$\varphi(t) = (-y(t-1)\ldots - y(t-p), \; u(t-1)\ldots u(t-p))^T \qquad (8)$$
$$\psi(t) = (u(t-p-L)\ldots u(t-p-L-m+1))^T \qquad (9)$$

where $p > n$ and $m \geq p + n$. Also let $\tilde{\varphi}(t)$ and $\bar{\varphi}(t)$ denote the signal and, respectively, the noise part of $\varphi(t)$.

Let $R_{\psi\varphi}$ denote the following covariance matrix

$$R_{\psi\varphi} = \mathbf{E}\{\psi(t)\varphi^T(t)\} \quad (m \times 2p) \qquad (10)$$

0-7803-2431-5/95 $4.00 © 1995 IEEE

and similarly for $R_{\psi\psi}$ and $R_{\varphi\varphi}$. The vector $\psi(t)$ is usually called an IV vector because it is uncorrelated with the noise part of $\varphi(t)$, and hence it can be used to extract the signal part of $\varphi(t)$. Let

$$R = R_{\psi\psi}^{-1/2} R_{\psi\varphi} R_{\varphi\varphi}^{-1/2} \tag{11}$$

It can be proved [6] that under a mild "persistence-of-excitation" assumption on the noise-free input $x(t)$,

$$\text{rank}[R] = n + p. \tag{12}$$

In view of (12) we can write the singular value decomposition (SVD) of R as

$$R = \overset{n+p}{\begin{pmatrix} U & \Xi \end{pmatrix}} \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} S^T \\ G^T \end{pmatrix} \begin{matrix} n+p \\ p-n \end{matrix} = U \Sigma S^T. \tag{13}$$

Finally, let

$$\widetilde{G} = \begin{pmatrix} 1 & a_1 & \dots & a_n & 0 & b_0 & \dots & b_n & & 0 \\ & \ddots & \ddots & & \ddots & & \ddots & & \ddots & \\ 0 & & 1 & a_1 & \dots & a_n & 0 & b_0 & \dots & b_n \end{pmatrix}^T \tag{14}$$

A key result, proved in [6], is that

$$\widetilde{G}^T R_{\varphi\varphi}^{1/2} S = 0. \tag{15}$$

$R_{\varphi\varphi}^{1/2} S$ in equation (15) is consistently estimable from the available samples, and $\widetilde{G}$ is linear in the unknown parameters. Hence (15) can be used for parameter estimation in a rather obvious manner as explained in the following. Let

$$\widehat{R} = \widehat{R}_{\psi\psi}^{-1/2} \widehat{R}_{\psi\varphi} \widehat{R}_{\varphi\varphi}^{-1/2} \tag{16}$$

where $\widehat{R}_{\psi\psi} = \frac{1}{N} \sum_{t=1}^{N} \psi(t)\psi^T(t)$ and $\widehat{R}_{\psi\varphi}$, $\widehat{R}_{\varphi\varphi}$ are similarly defined. Furthermore, let

$$\widehat{S} \;=\; \text{the matrix made from the } (n+p) \text{ principal right singular vectors of } \widehat{R}. \tag{17}$$

Once $\widehat{S}$ and $\widehat{R}_{\varphi\varphi}^{1/2}$ have been computed, the next step should be to obtain parameter estimates by minimizing a suitable scalar function of the matrix $\widetilde{G}^T \widehat{R}_{\varphi\varphi}^{1/2} \widehat{S}$ (cf (15)). In finite samples the accuracy with which the columns of $\widehat{S}$ are determined can be expected to be proportional to the corresponding singular values. In addition, better performance may be expected if not only $\widehat{S}$ but also the matrix $\widetilde{G}^T \widehat{R}_{\varphi\varphi}^{1/2}$ in the sample counterpart of (15) would be orthogonal [7, 8]. These two arguments suggest the following Frobenius-norm loss function whose minimization with respect to $\theta = (1\ a_1 \dots a_n\ b_1 \dots b_n)$ yields the parameter estimates:

$$f(\theta) = \| \left(\widetilde{G}^T \widehat{R}_{\varphi\varphi} \widetilde{G} \right)^{-1/2} \widetilde{G}^T \widehat{R}_{\varphi\varphi}^{1/2} \widehat{S} \widehat{\Sigma} \|_F^2 \tag{18}$$

where $\widehat{\Sigma}$ denotes the sample analogue of Σ defined in (13). Minimization of (18) is a highly nonlinear problem. Fortunately, owing to the special structure of (18) it can be shown that parameter estimates having the same asymptotic properties as the minimizer of (18) can be obtained by the following multistep but non-iterative procedure.

Step 1 Obtain initial parameter estimates by minimizing (18) with $\widetilde{G}^T \widehat{R}_{\varphi\varphi} \widetilde{G}$ replaced by

$$\begin{bmatrix} I & 0 \end{bmatrix} \widehat{R}_{\varphi\varphi} \begin{bmatrix} I \\ 0 \end{bmatrix}. \tag{19}$$

Step 2 Use the initial estimate of Step 1 to determine a consistent estimate of $(\widetilde{G}^T \widehat{R}_{\varphi\varphi} \widetilde{G})^{-1}$. Minimize (18) using this consistent estimate of the "weighting matrix".

Step 3 (Optional). Reiterate Step 2 using the most recent estimate of θ.

Clearly all steps above are covered by the quadratic problem:

$$\min_{\theta} \text{tr} \left[\widetilde{G} W \widetilde{G}^T \widehat{\Delta} \right] \tag{20}$$

where $\widehat{\Delta} = \widehat{R}_{\varphi\varphi}^{1/2} \widehat{S} \widehat{\Sigma}^2 \widehat{S}^T \widehat{R}_{\varphi\varphi}^{1/2}$ and W takes on different values depending on the step under discussion. It can be shown that the general problem (20) can be solved in the following efficient way. We let $\bar{\Omega}$ denote the Kronecker product $W \otimes \widehat{\Delta}$ from which the rows and columns corresponding to the zeros in $\text{vec}(\widetilde{G})$ are eliminated, and also denote by Ω the following matrix

$$\Omega = \begin{bmatrix} I & \cdots & I \end{bmatrix} \bar{\Omega} \begin{bmatrix} I & \cdots & I \end{bmatrix}^T. \tag{21}$$

Then the TLS (Total Least Squares) solution to (20) is given by

$$\hat{\theta} \;=\; \text{the minimum eigenvector of } \Omega, \text{ with the first component normalized to one} \tag{22}$$

2. ASYMPTOTIC DISTRIBUTION

This section establishes the asymptotic distribution of the parameter estimates introduced previously, under the assumption that the data are Gaussian distributed. First, we note that the consistency of the method can be proved using the same analysis technique as in [6], to which we refer for details. Since $\hat{\theta}$ is a consistent estimate of θ we can make use of a standard asymptotic Taylor series expansion to write [3, 5]

$$\hat{\theta} - \theta \simeq -\left(f''(\theta)\right)^{-1} f'(\theta) \tag{23}$$

where the symbol $\simeq$ denotes a first-order approximate equality in which all terms that do not affect the asymptotics (w.r.t. N) have been neglected, and where $f''(\theta)$ and $f'(\theta)$ are the Hessian matrix and, respectively, the gradient vector of $f(\theta)$, evaluated at the *true* parameter values. The derivation of appropriate asymptotic approximations for $f'(\theta)$ and $f''(\theta)$ can easily be done as follows. Let

$$\tilde{g}'_k = \frac{\partial \tilde{g}}{\partial \theta_k} \tag{24}$$

where $\tilde{g} = \mathrm{vec}(\tilde{G})$. Then

$$f'_k(\theta) = 2\tilde{g}'^T_k \left(W \otimes \hat{\tilde{\Delta}}\right) \tilde{g} \tag{25}$$

and, noting that $\tilde{g}''_{k,j} = 0$,

$$f''_{k,j}(\theta) \simeq 2\tilde{g}'^T_k \left(W \otimes R^{1/2}_{\varphi\varphi} S \Sigma^2 S^T R^{1/2}_{\varphi\varphi}\right) \tilde{g}'_j \tag{26}$$

From the above calculations and the rule $\mathrm{vec}(ABC) = (C^T \otimes A)\,\mathrm{vec}(B)$ it readily follows that

$$\hat{\theta} - \theta \simeq -H^{-1} D^T \,\mathrm{vec}(\hat{\Delta}\tilde{G}W) \tag{27}$$

where

$$D = \begin{pmatrix} \tilde{g}'_1 & \cdots & \tilde{g}'_{2n+1} \end{pmatrix} \tag{28}$$

$$H = D^T \left(W \otimes R^{1/2}_{\varphi\varphi} S \Sigma^2 S^T R^{1/2}_{\varphi\varphi}\right) D. \tag{29}$$

Hence, in order to compute the large-sample covariance of $\hat{\theta}$ it is necessary to derive the asymptotic covariance matrix of $\mathrm{vec}(\hat{\Delta}\tilde{G}W)$. Towards that end, note from the sample analogue of (13) that

$$\begin{aligned} \hat{\Delta}\tilde{G} &= \hat{R}^{1/2}_{\varphi\varphi} \hat{R} \hat{U} \hat{\Sigma}^{-1} \hat{\Sigma}^2 \hat{\Sigma}^{-1} \hat{U}^T \hat{R} \hat{R}^{1/2}_{\varphi\varphi} \tilde{G} \\ &= \hat{R}^T_{\psi\varphi} \hat{R}^{-1/2}_{\psi\psi} \hat{U} \hat{U}^T \hat{R}^{-1/2}_{\psi\psi} \hat{R}_{\psi\varphi} \tilde{G} \end{aligned} \tag{30}$$

Next we make use of the decomposition $\varphi(t) = \tilde{\varphi}(t) + \bar{\varphi}(t)$ and the fact that $\tilde{\varphi}^T(t)\tilde{G} = 0$ to see that $\hat{R}_{\psi\varphi}\tilde{G}$ tends to zero as N tends to infinity. It follows that $\hat{R}^T_{\psi\varphi}\hat{R}^{-1/2}_{\psi\psi}\hat{U}\hat{U}^T\hat{R}^{-1/2}_{\psi\psi}$ in (30) can be replaced by its true value without changing the asymptotic properties of $\hat{\Delta}\tilde{G}$. Hence we obtain the following asymptotically valid expression for $\mathrm{vec}(\hat{\Delta}\tilde{G}W)$

$$\mathrm{vec}(\hat{\Delta}\tilde{G}W) = (W\tilde{G}^T \otimes K) \left[\frac{1}{N}\sum_{t=1}^{N} \bar{\varphi}(t) \otimes \psi(t)\right]. \tag{31}$$

where $K \triangleq R^T_{\psi\varphi} R^{-1/2}_{\psi\psi} U U^T R^{-1/2}_{\psi\psi}$. Now we make the assumption that the data is Gaussian distributed. (Note that, up to this point, no assumption on the distributional properties of the data has been utilized.) Under the Gaussian assumption the following covariance matrix can readily be derived [6]

$$\begin{aligned} &\mathrm{cov}\left[\frac{1}{N}\sum_{t=1}^{N} \bar{\varphi}(t) \otimes \psi(t)\right] \\ &\quad = \frac{1}{N^2}\sum_{l=-N}^{N} (N - |l|)\left[\Gamma_{\bar{\varphi}}(l) \otimes \Gamma_{\psi}(l)\right] \end{aligned} \tag{32}$$

$$\Gamma_{\bar{\varphi}}(l) = \mathbf{E}[\bar{\varphi}(t)\bar{\varphi}^T(t-l)] \;\; (2p \times 2p) \tag{33}$$

$$\Gamma_{\psi}(l) = \mathbf{E}[\psi(t)\psi^T(t-l)] \;\; (m \times m). \tag{34}$$

By combining (27), (31) and (32) we obtain the following large-sample expression for the covariance matrix of the (Gaussian distributed) estimation error $\hat{\theta} - \theta$:

$$\mathrm{cov}(\hat{\theta}) = H^{-1} D^T \Gamma D H^{-1} \tag{35}$$

where

$$\begin{aligned} \Gamma &= (W\tilde{G}^T \otimes K)\left[\frac{1}{N^2}\sum_{l=-N}^{N} (N - |l|)\left[\Gamma_{\bar{\varphi}}(l) \otimes \Gamma_{\psi}(l)\right]\right] \\ &\quad (\tilde{G}W \otimes K^T) \\ &= \frac{1}{N^2}\sum_{l=-N}^{N} (N - |l|)\left[W\tilde{G}^T \Gamma_{\varphi}(l) \tilde{G}W \otimes K \Gamma_{\psi}(l) K^T\right] \end{aligned} \tag{36}$$

In the last equality of equation (36) we used the fact that $\tilde{\varphi}^T(t)\tilde{G} = 0$.

If $\|\Gamma_{\psi}(l) \otimes \Gamma_{\bar{\varphi}}(l)\|$ decreases rapidly as $|l|$ increases, then a suboptimal weight can be obtained by neglecting all terms in (36) except the zeroth-order term, and then minimizing the so-simplified covariance matrix (35) with respect to W. Doing so we obtain the weight used in (18):

$$W_o = (\tilde{G}^T R_{\varphi\varphi} \tilde{G})^{-1} \tag{37}$$

3. NUMERICAL EXAMPLE

Consider the system described by the equations (4), (5) with

$$\begin{aligned} A(q^{-1}) &= 1 - 1.8778q^{-1} + 0.9025q^{-2} \\ B(q^{-1}) &= 1 - q^{-1} + 0.5q^{-2}. \end{aligned}$$

The noise-free input signal is generated as the following moving average:

$$\begin{aligned} x(t) &= (0.7 + 0.8q^{-1} + 0.9q^{-2} + q^{-3} \\ &\quad + 0.9q^{-4} + 0.8q^{-5} + 0.7q^{-6}) e_x(t) \end{aligned} \tag{38}$$

where $e_x(t)$ is white noise of unit variance. The input and output noises are generated as first-order moving averages:

$$w(t) = (1 + 0.5q^{-1}) e_w(t), \;\; v(t) = (1 + 0.5q^{-1}) e_v(t) \tag{39}$$

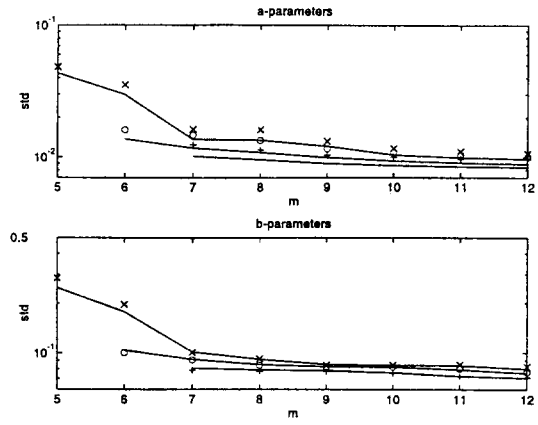

Figure 1: RMSE values for the proposed method. Solid lines = Theoretical RMSE; "×"=Empirical RMSE for $p = 3$; "o"=Empirical RMSE for $p = 4$; "+"=Empirical RMSE for $p = 5$.

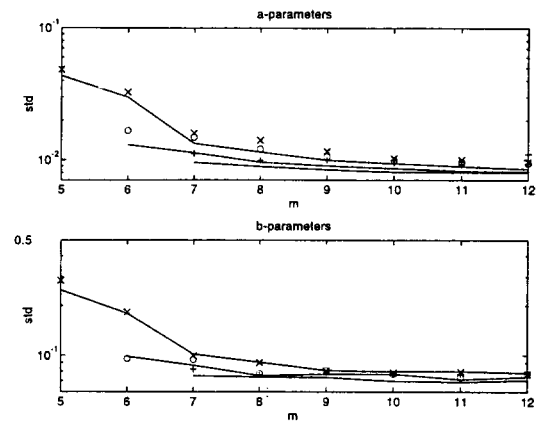

Figure 2: RMSE values for the method of [6]. Solid lines = Theoretical RMSE; "×"=Empirical RMSE for $p = 3$; "o"=Empirical RMSE for $p = 4$; "+"=Empirical RMSE for $p = 5$.

where $e_w(t)$ and $e_v(t)$ are white noise sequences that are uncorrelated with each other. The variances of $e_w(t)$ and $e_v(t)$ are selected so that the signal-to-noise ratio (SNR) equals 11dB. The SNR is defined as [6]

$$\text{SNR} = E[x^2(t)]/E[A(q^{-1})v(t) - B(q^{-1})w(t)]^2. \quad (40)$$

Both the method described in this paper (Step 3 is iterated four times) and the optimal IV-WSF procedure introduced in [6] are used for estimating the system parameters. We estimate the parameters of the system above from $N = 1000$ data samples, for $p = 3, 4, 5$ and $m = p+2, \ldots, 12$. One hundred Monte-Carlo simulations are used to determine the root-mean-squared errors (RMSE) of the various parameter estimates.

Figures 1 and 2 show the RMSE values for the methods under study. The Figures also show the theoretical standard deviations computed by using the corresponding asymptotic covariance expressions. To reduce the number of graphs we only plot: $\text{std}(\hat{a}) = [\text{std}(\hat{a}_1) + \text{std}(\hat{a}_2)]/2$ and similarly for the b-parameters.

Comparing Figures 1 and 2 we see that the proposed method and the optimal IV-WSF procedure of [6] possess similar performance. On the other hand, the method of [6] is 30 to 40 times more computationally intensive than the proposed method for this example.

4. REFERENCES

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

[2] J. A. Cadzow and O. M. Solomon. Algebraic approach to system identification. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP–34(3):462–469, June 1986.

[3] L. Ljung. *System Identification: Theory for the User*. Prentice Hall, Englewood Cliffs, N.J., 1987.

[4] T. Söderström and P. Stoica. *Instrumental Variable Methods for System Identification*. Springer Verlag, Berlin, Germany, 1983.

[5] T. Söderström and P. Stoica. *System Identification*. Prentice Hall International, London, U.K., 1989.

[6] P. Stoica, M. Cedervall, and A. Eriksson. Combined instrumental variable and subspace fitting approach to parameter estimation of noisy input-output systems. Technical Report UPTEC 94043R, Department of Technology, Uppsala University, Sweden, May 1994.

[7] P. Stoica and A. Nehorai. Performance study of conditional and unconditional direction-of-arrival estimation. *IEEE Transactions on Signal Processing*, SP-38(10):1783–1795, Oct. 1990.

[8] P. Stoica and K. C. Sharman. Maximum likelihood methods for direction-of-arrival estimation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-38(7):1132–1143, July 1990.

[9] P. Young. *Recursive Estimation and Time Series Analysis*. Springer Verlag, Berlin, 1984.