Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Frequency domain maximum likelihood identification of noisy input–output models

Umberto Soverini * **Torsten Söderström** **

* *Department of Electrical, Electronic and Information Engineering, University of Bologna, Italy (e-mail: umberto.soverini@unibo.it)*
** *Department of Information Technology, Uppsala University, Sweden (e-mail: ts@it.uu.se)*

Abstract: This paper deals with the identification of errors–in–variables (EIV) models corrupted by additive and uncorrelated white Gaussian noises when the noise–free input is an arbitrary signal, not necessarily periodic. In particular, a frequency domain maximum likelihood (ML) estimator is proposed. As some other EIV estimators, this method assumes that the ratio of the noise variances is known.

Keywords: Linear dynamic systems; System identification; Errors–in–variables models; Discrete Fourier Transform; Maximum likelihood identification.

1. INTRODUCTION

In this paper the problem of identifying a linear dynamic system from noisy input–output measurements is addressed. System representations where both inputs and outputs are affected by additive errors are called errors–in–variables (EIV) models and play an important role in several engineering applications. The identification of EIV models has been deeply investigated in the literature and many solutions have been proposed with different approaches, see (Söderström, 2007, 2012) and the references therein.

In many EIV contexts the additive noises are assumed as white. In these cases, if the assumptions of Gaussianity are fulfilled, it is feasible to use a maximum likelihood (ML) approach.

ML estimation methods are commonly used in dynamic system identification, and several approaches have been developed in time and frequency domain. In a recent paper (Agüero et al., 2010) the relation between time and frequency domain versions of ML estimation methods has been investigated. It has been shown that the results of the estimation problem do not depend on the domain chosen to describe the available data (i.e. time or frequency). Instead, it is the choice of the likelihood function, i.e., which parameters are to be estimated and what data is assumed available, that leads to different solutions.

In this work the EIV ML problem is addressed by using frequency domain techniques, when the noise-free input is an arbitrary sequence and the noise variance ratio is known. The latter is a standard assumption, e.g. for all TLS–based estimators in EIV problems (Van Huffel and Lemmerling, 2002).

The proposed method can be considered as a frequency domain version of the ML solution described in (Diversi et al., 2007). The two methods seem to be equivalent, as it appears from the numerical examples, in accordance with the assertion given in (Agüero et al., 2010). The determination of the finite Cramér Rao lower bound is not developed here; however the result does not differ from the one already presented in (Diversi et al., 2007).

The organization of the paper is as follows. Section 2 defines the EIV identification problem in frequency domain. Section 3 describes the features of the identification procedure in case of noise–free data and Section 4 presents the ML solution of the original EIV identification problem. Section 5 describes the Koopmans-Levin solution for frequency data. In Section 6 the effectiveness of the proposed algorithm is verified by means of Monte Carlo simulations. Finally some concluding remarks are reported in Section 7.

2. STATEMENT OF THE PROBLEM

Consider the linear time–invariant SISO system described in Figure 1. The noise–free input and output $\hat{u}(t)$, $\hat{y}(t)$ are linked by the linear difference equation

$$A(z^{-1})\,\hat{y}(t) = B(z^{-1})\,\hat{u}(t), \tag{1}$$

where $A(z^{-1})$ and $B(z^{-1})$ are polynomials in the backward shift operator z^{-1}

$$A(z^{-1}) = 1 + \alpha_1\, z^{-1} + \cdots + \alpha_n\, z^{-n} \tag{2}$$

$$B(z^{-1}) = \beta_0 + \beta_1\, z^{-1} + \cdots + \beta_n\, z^{-n}. \tag{3}$$

In the EIV environment the input and output measurements are assumed as corrupted by additive noise so that the available observations are

$$u(t) = \hat{u}(t) + \tilde{u}(t) \tag{4}$$

$$y(t) = \hat{y}(t) + \tilde{y}(t). \tag{5}$$

The following assumptions are made.

- A1. The system (1) is asymptotically stable.
- A2. $A(z^{-1})$ and $B(z^{-1})$ do not share any common factor.
- A3. The order n of the system is assumed as *a priori* known.
- A4. The noise–free input $\hat{u}(t)$ is a quasi–stationary bounded deterministic signal (Ljung, 1999) and is persistently exciting of sufficiently high order.
- A5. $\tilde{u}(t)$ and $\tilde{y}(t)$ are zero–mean, mutually uncorrelated Gaussian white processes with variances λ_u and λ_y.
- A6. The noise variances λ_u and λ_y are unknown but their ratio $\rho = \lambda_y/\lambda_u$ is assumed as known, with $0 < \rho < \infty$.

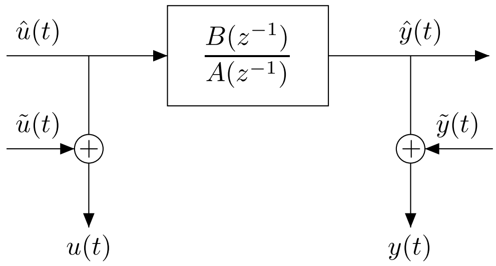

Fig. 1. Errors–in–variables model

Let $\{u(t)\}_{t=0}^{N-1}$ and $\{y(t)\}_{t=0}^{N-1}$ be a set of input and output observations at N equidistant time instants. The corresponding Discrete Fourier Transforms (DFTs) are defined as

$$U(\omega_k) = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} u(t)\, \mathrm{e}^{-j\omega_k t} \tag{6}$$

$$Y(\omega_k) = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} y(t)\, \mathrm{e}^{-j\omega_k t}\ , \tag{7}$$

where $\omega_k = 2\pi k/N$ and $k = 0, \ldots, N-1$. The system transfer function is represented as

$$G(\mathrm{e}^{-j\omega_k}) = \frac{B(\mathrm{e}^{-j\omega_k})}{A(\mathrm{e}^{-j\omega_k})}. \tag{8}$$

The DFTs defined in (6) and (7) can be expressed in matrix form by introducing the $N \times N$ Fourier matrix F_N (Agüero et al., 2010) whose entries are defined as follows

$$F_N = [f_{ik}] \tag{9}$$

$$f_{ik} = \frac{1}{\sqrt{N}} \mathrm{e}^{-j\frac{2\pi}{N}(i-1)(k-1)} \quad i, k = 1, \ldots, N\ . \tag{10}$$

In can be proved that matrix F_N is unitary, i.e. $F_N F_N^H = I$, where $(\cdot)^H$ denotes the transpose and conjugate operation.

Defining the following vectors in time and frequency domain

$$v_u = [u(0) \ldots u(N-1)]^T \tag{11}$$

$$v_y = [y(0) \ldots y(N-1)]^T \tag{12}$$

$$V_U = [U(\omega_0) \ldots U(\omega_{N-1})]^T \tag{13}$$

$$V_Y = [Y(\omega_0) \ldots Y(\omega_{N-1})]^T \tag{14}$$

the relations (6) and (7) can be represented by the linear transformations

$$V_U = F_N\, v_u \tag{15}$$

$$V_Y = F_N\, v_y\ . \tag{16}$$

In frequency domain, the problem under investigation can be stated as follows.

Problem 1. Let $U(\omega_k)$, $Y(\omega_k)$ be a set of noisy measurements generated by an EIV system of type (1)–(5), under Assumptions A1–A6, where $\omega_k = 2\pi k/N$ and $k = 0, \ldots, N-1$. Estimate the system parameters α_i $(i = 1, \ldots, n)$, β_i $(i = 0, \ldots, n)$ and the noise variances λ_u, λ_y.

In the following we focus on the case when N is much larger then n, as we are interested in consistency and asymptotic properties in general.

3. THE NOISE–FREE CASE

This section develops a model analysis in absence of noise and describes the basic ideas of a two-step iterative algorithm, where every step relies on a least squares formulation.

With reference to the noise–free signals $\hat{u}(t)$ and $\hat{y}(t)$, definitions similar to (11)–(14) and (15)–(16) hold, i.e.

$$\hat{v}_u = [\hat{u}(0) \ldots \hat{u}(N-1)]^T \tag{17}$$

$$\hat{v}_y = [\hat{y}(0) \ldots \hat{y}(N-1)]^T \tag{18}$$

$$\hat{V}_U = [\hat{U}(\omega_0) \ldots \hat{U}(\omega_{N-1})]^T \tag{19}$$

$$\hat{V}_Y = [\hat{Y}(\omega_0) \ldots \hat{Y}(\omega_{N-1})]^T\ , \tag{20}$$

where

$$\hat{V}_U = F_N\, \hat{v}_u \tag{21}$$

$$\hat{V}_Y = F_N\, \hat{v}_y\ . \tag{22}$$

For the subsequent analysis it is convenient to introduce also the auxiliary process

$$\eta(t) = \frac{1}{A(z^{-1})}\, \hat{u}(t)\ , \tag{23}$$

which allows to represent the system (1) with the following relations

$$\hat{u}(t) = A(z^{-1})\, \eta(t) \tag{24}$$

$$\hat{y}(t) = B(z^{-1})\, \eta(t)\ , \tag{25}$$

assuming that the n initial conditions $\{\eta(t)\}_{t=-n}^{-1}$ are known. Defining the vector

$$\eta = [\eta(-n), \ldots, \eta(-1), \eta(0), \ldots, \eta(N-1)]^T \tag{26}$$

the relations (24) and (25) can be written in matrix form as

$$\hat{v}_u = \mu_\alpha\, \eta \tag{27}$$

$$\hat{v}_y = \mu_\beta\, \eta\ , \tag{28}$$

where μ_α and μ_β are the $N \times (N+n)$ Toeplitz matrices

$$\mu_\alpha = \begin{bmatrix} \alpha_n & \ldots & \alpha_1 & 1 & 0 & \ldots & 0 \\ 0 & \alpha_n & \ldots & \alpha_1 & 1 & \ldots & 0 \\ \vdots & & \ddots & & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & \alpha_n & \ldots & \alpha_1 & 1 \end{bmatrix} \tag{29}$$

$$\mu_\beta = \begin{bmatrix} \beta_n & \ldots & \beta_1 & \beta_0 & 0 & \ldots & 0 \\ 0 & \beta_n & \ldots & \beta_1 & \beta_0 & \ldots & 0 \\ \vdots & & \ddots & & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & \beta_n & \ldots & \beta_1 & \beta_0 \end{bmatrix}\ . \tag{30}$$

The DFT of the vector η in (26) results in

$$E = [E(\omega_{-n}) \ldots E(\omega_{-1}) E(\omega_0) \ldots E(\omega_{N-1})]^T \tag{31}$$

by means of the linear transformation

$$E = F_{N+n}\, \eta \tag{32}$$

and *vice versa*

$$\eta = F_{N+n}^{-1}\, E\ , \tag{33}$$

where the $(N+n) \times (N+n)$ Fourier matrix F_{N+n} is constructed by following the rules given in (9)–(10).

Remark 1. Note that, since the dimension of vector η is different from that of $\hat{v}_u$ and $\hat{v}_y$, for the same value of k, the generic frequency $\omega_k = 2\pi k/(N+n)$ in (31) does not coincide with $\omega_k = 2\pi k/N$ in (19) and (20).

Thanks to (21) (22) and (33), the relations (27)–(28) can be written in the frequency domain as follows

$$\hat{V}_U = M_\alpha\, E \tag{34}$$

$$\hat{V}_Y = M_\beta\, E\ , \tag{35}$$

where

$$M_\alpha = F_N \mu_\alpha F_{N+n}^{-1} \tag{36}$$

$$M_\beta = F_N \mu_\beta F_{N+n}^{-1}\ . \tag{37}$$

By introducing the $2N$–dimensional vector

$$\hat{V} = [\hat{V}_Y^T \, \hat{V}_U^T \,]^T \tag{38}$$

and the $2N \times (N+n)$ matrix

$$M = \begin{bmatrix} M_\beta \\ M_\alpha \end{bmatrix} , \tag{39}$$

the relations (34)–(35) can be rewritten as

$$\hat{V} = M \, E \, . \tag{40}$$

Thanks to Assumption A2, the matrix M has full rank, so that the vector E can be univocally determined by the pseudoinverse relation

$$E = M^\dagger \, \hat{V} = (M^H \, M)^{-1} \, M^H \, \hat{V} \, . \tag{41}$$

Let us consider now the dynamic systems described by equations (24) and (25). As input of these systems, it is possible to apply the sequence $\{\eta(t)\}_{t=0}^{N-1}$ of length N, whose DFT is

$$E_0(\omega_k) = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} \eta(t) \, \mathrm{e}^{-j\omega_k t} \, , \tag{42}$$

with $\omega_k = 2\pi k/N$ and $k = 0, \ldots, N-1$.

By introducing the vector

$$\eta_0 = [\eta(0) \ldots \eta(N-1)]^T \tag{43}$$

and the corresponding DFT

$$E_0 = [E_0(\omega_0) \ldots E_0(\omega_{N-1})]^T \, , \tag{44}$$

relation (42) can be expressed also in matrix form

$$E_0 = F_N \, \eta_0 \, . \tag{45}$$

Remark 2. Note that the dimension of vector η_0 is equal to N and coincides with that of $\hat{v}_u$ and $\hat{v}_y$. Therefore the observation stated in Remark 1 holds also for vector E_0: for a fixed value of k the generic frequency $\omega_k = 2\pi k/N$ in (44) is equal to that of (19) and (20), but does not coincide with $\omega_k = 2\pi k/(N+n)$ in (31).

From knowledge of the vector E (31) it is possible to compute E_0 (44). In fact, denoting with I_N the unit matrix of dimension N, it is possible to define matrix

$$J = [0_{N \times n} \,|\, I_N] \, , \tag{46}$$

so that it results

$$\eta_0 = J \, \eta \, . \tag{47}$$

Thanks to (33) and (45), we obtain

$$E_0 = M_E \, E \, , \tag{48}$$

where

$$M_E = F_N \, J \, F_{N+n}^{-1} \, . \tag{49}$$

Remark 3. All the previous considerations can be summed up as follows. Assuming the system parameters α_i $(i = 1, \ldots, n)$, β_i $(i = 0, \ldots, n)$ as known, from the input–output frequency sequences $\hat{V}_U$ and $\hat{V}_Y$ it is possible to compute the vector E by means of equation (41). Then, vector E_0 can be obtained through (48).

It is a well–known fact (Pintelon et al., 1997) that for finite N, even in absence of noise, the ratio of the DFTs $\hat{Y}(\omega_k)$ and $\hat{U}(\omega_k)$ $(\omega_k = 2\pi k/N)$ is not equal to the true transfer function

$$G(\mathrm{e}^{-j\omega_k}) \neq \frac{\hat{Y}(\omega_k)}{\hat{U}(\omega_k)} \, . \tag{50}$$

In fact, it can be proved that the DFTs $\hat{Y}(\omega_k)$ and $\hat{U}(\omega_k)$ exactly satisfy an extended model that includes also a transient term, i.e.

$$A(\mathrm{e}^{-j\omega_k}) \, \hat{Y}(\omega_k) = B(\mathrm{e}^{-j\omega_k}) \, \hat{U}(\omega_k) + T(\mathrm{e}^{-j\omega_k}) \, , \tag{51}$$

where $T(z^{-1})$ is a polynomial of order $n-1$

$$T(z^{-1}) = \tau_0 + \tau_1 \, z^{-1} + \cdots + \tau_{n-1} \, z^{-n+1} \tag{52}$$

that takes into account the effects of the initial and final conditions of the experiment.

These considerations can be applied to the dynamic systems (24) and (25). In frequency domain, these systems can be expressed by the relations

$$\hat{U}(\omega_k) = A(\mathrm{e}^{-j\omega_k}) \, E_0(\omega_k) + T^u(\mathrm{e}^{-j\omega_k}) \tag{53}$$

$$\hat{Y}(\omega_k) = B(\mathrm{e}^{-j\omega_k}) \, E_0(\omega_k) + T^y(\mathrm{e}^{-j\omega_k}) \tag{54}$$

where $T^u(z^{-1})$ and $T^y(z^{-1})$ are polynomials of order $n-1$

$$T^u(z^{-1}) = \tau_0^u + \tau_1^u \, z^{-1} + \cdots + \tau_{n-1}^u \, z^{-n+1} \tag{55}$$

$$T^y(z^{-1}) = \tau_0^y + \tau_1^y \, z^{-1} + \cdots + \tau_{n-1}^y \, z^{-n+1} \, . \tag{56}$$

Let us define the row vectors

$$Z_{n+1}(\omega_k) = [\mathrm{e}^{-jn\omega_k} \, \mathrm{e}^{-j(n-1)\omega_k} \, \ldots \mathrm{e}^{-j\omega_k} \, 1] \tag{57}$$

$$Z_n(\omega_k) = [\mathrm{e}^{-j(n-1)\omega_k} \, \ldots \mathrm{e}^{-j\omega_k} \, 1] \, , \tag{58}$$

whose entries are constructed with multiple frequencies of ω_k.

By introducing the parameter vectors

$$\theta_\alpha = [\alpha_n \ldots \alpha_1 1 \,]^T \tag{59}$$

$$\theta_\beta = [\beta_n \ldots \beta_1 \beta_0]^T \tag{60}$$

$$\theta_u = [\tau_{n-1}^u \cdots \tau_0^u]^T \tag{61}$$

$$\theta_y = [\tau_{n-1}^y \cdots \tau_0^y]^T \, , \tag{62}$$

equations (53)–(54) can be written, at every frequency ω_k ($k = 0, \ldots, N-1$), as follows

$$\hat{U}(\omega_k) = Z_{n+1}(\omega_k) \, \theta_\alpha \, E_0(\omega_k) + Z_n(\omega_k) \, \theta_u \tag{63}$$

$$\hat{Y}(\omega_k) = Z_{n+1}(\omega_k) \, \theta_\beta \, E_0(\omega_k) + Z_n(\omega_k) \, \theta_y \, . \tag{64}$$

Equivalently, it results in

$$\hat{U}(\omega_k) = M(\omega_k) \, \theta_U \tag{65}$$

$$\hat{Y}(\omega_k) = M(\omega_k) \, \theta_Y \, , \tag{66}$$

where

$$M(\omega_k) = [E_0(\omega_k) \, Z_{n+1}(\omega_k) \,|\, Z_n(\omega_k)] \tag{67}$$

and

$$\theta_U = [\theta_\alpha^T \, \theta_u^T \,]^T \tag{68}$$

$$\theta_Y = [\theta_\beta^T \, \theta_y^T \,]^T \, . \tag{69}$$

Finally, defining the $N \times (2n+1)$ matrix

$$\Omega = \begin{bmatrix} M(\omega_0) \\ \vdots \\ M(\omega_{N-1}) \end{bmatrix} \tag{70}$$

and recalling definitions (19)–(20), equations (53)–(54) can be written in the following matrix forms

$$\hat{V}_U = \Omega \, \theta_U \tag{71}$$

$$\hat{V}_Y = \Omega \, \theta_Y \, . \tag{72}$$

If $N > (2n+1)$ the matrix Ω has full column rank and the vectors θ_U and θ_Y can be univocally determined by the pseudoinverse relations

$$\theta_U = \Omega^\dagger \, \hat{V}_U = (\Omega^H \, \Omega)^{-1} \, \Omega^H \, \hat{V}_U \tag{73}$$

$$\theta_Y = \Omega^\dagger \, \hat{V}_Y = (\Omega^H \, \Omega)^{-1} \, \Omega^H \, \hat{V}_Y \, . \tag{74}$$

Remark 4. The observations given in Remark 3 can now be completed as follows. *Vice versa*, from knowledge of vector E_0, it is possible to construct the row vector $M(\omega_k)$ (67) and the matrix Ω (70). Finally, the system parameters θ_α and θ_β can be recovered, together with the auxiliary vectors θ_u and θ_y, by using equations (73)–(74).

4. MAXIMUM LIKELIHOOD IDENTIFICATION

Defining the noise vectors

$$\tilde{v}_u = [\tilde{u}(0)\dots\tilde{u}(N-1)]^T \tag{75}$$

$$\tilde{v}_y = [\tilde{y}(0)\dots\tilde{y}(N-1)]^T \tag{76}$$

$$\tilde{v} = [\tilde{v}_y^T\,\tilde{v}_u^T]^T \tag{77}$$

and the corresponding DFTs

$$\tilde{V}_U = [\tilde{U}(\omega_0)\dots\tilde{U}(\omega_{N-1})]^T \tag{78}$$

$$\tilde{V}_Y = [\tilde{Y}(\omega_0)\dots\tilde{Y}(\omega_{N-1})]^T \tag{79}$$

$$\tilde{V} = [\tilde{V}_Y^T\,\tilde{V}_U^T]^T\,, \tag{80}$$

from equations (40) and (71)–(72), the EIV model (1)–(5) can be written in the following compact forms

$$V = [V_Y^T\,V_U^T]^T = \hat{V} + \tilde{V} = M\,E + \tilde{V} \tag{81}$$

$$V = [V_Y^T\,V_U^T]^T = \hat{V} + \tilde{V} = \Gamma\,\Theta + \tilde{V}\,, \tag{82}$$

where

$$\Gamma = \begin{bmatrix}\Omega & 0\\ 0 & \Omega\end{bmatrix} \qquad \Theta = \begin{bmatrix}\theta_U\\ \theta_Y\end{bmatrix}. \tag{83}$$

Thanks to Assumption A5 we have

$$\tilde{\Sigma} = E\,[\tilde{v}\,\tilde{v}^T\,] = E\,[\tilde{V}\,\tilde{V}^H\,] = \begin{bmatrix}\lambda_y\,I_N & 0\\ 0 & \lambda_u\,I_N\end{bmatrix}, \tag{84}$$

where $E[\cdot]$ is the mathematical expectation. Because of A6 it results

$$\tilde{\Sigma} = \lambda_u \begin{bmatrix}\rho I_N & 0\\ 0 & I_N\end{bmatrix} = \lambda_u\,W\,, \tag{85}$$

i.e. $\tilde{\Sigma}$ is known up to a single multiplicative constant. Since the distribution of the measurement noise is Gaussian, the conditional probability density function of the complex–valued vector V in (81) is

$$p(V|\theta_Y, \theta_U, E, \lambda_u) = \tag{86}$$
$$\frac{1}{\pi^{2N}\det\tilde{\Sigma}}\exp\left\{-(V - M\,E)^H\,\tilde{\Sigma}^{-1}(V - M\,E)\right\}.$$

The log-likelihood function is thus given by

$$L(\theta_Y, \theta_U, E, \lambda_u) = \text{const} - 2N\log\lambda_u$$
$$-\frac{1}{\lambda_u}\,(V - M\,E)^H\,W^{-1}(V - M\,E), \tag{87}$$

or, equivalently, because of (82)

$$L(\theta_Y, \theta_U, E, \lambda_u) = \text{const} - 2N\log\lambda_u$$
$$-\frac{1}{\lambda_u}\,(V - \Gamma\,\Theta)^H\,W^{-1}(V - \Gamma\,\Theta)\,. \tag{88}$$

Since Γ and W are block–diagonal matrices, the loss function (88) can be rewritten as follows

$$L(\theta_Y, \theta_U, E, \lambda_u) = \text{const} - 2N\log\lambda_u$$
$$-\frac{1}{\lambda_u}\,(V_U - \Omega\,\theta_U)^H\,(V_U - \Omega\,\theta_U)$$
$$-\frac{1}{\lambda_u\,\rho}\,(V_Y - \Omega\,\theta_Y)^H\,(V_Y - \Omega\,\theta_Y)\,. \tag{89}$$

Therefore $L(\theta_Y, \theta_U, E, \lambda_u)$ is maximized when

$$E = \left(M^H W^{-1} M\right)^{-1} M^H W^{-1} V \tag{90}$$

$$\theta_Y = (\Omega^H\Omega)^{-1}\Omega^H V_Y \tag{91}$$

$$\theta_U = (\Omega^H\Omega)^{-1}\Omega^H V_U\,, \tag{92}$$

with

$$\lambda_u = \frac{1}{2N}\,(V - M\,E)^H\,W^{-1}(V - M\,E)\,. \tag{93}$$

Remark 5. The ML estimation problem as formulated in (89) turns out to have the same algebraic form as the EIV problem for a static linear system where all the latent variables (corresponding to E here) are estimated. For such cases it is known that the estimated parameters are consistent, but the estimated noise variance (corresponding to λ_u) is asymptotically off by 50 percent (Fuller, 1987). For this reason in the next algorithm we modify the estimate of λ_u by doubling the result of (93).

We can now formulate an iterative algorithm for the maximum likelihood estimation of θ_Y, θ_U, E and λ_u structured as follows.

Algorithm 1.

(1) Let θ_α^0 and θ_β^0 an initial estimate of the system parameters (59) (60). Define

$$\Theta^0 = [\theta_\alpha^{0^T}\;\theta_\beta^{0^T}]^T \tag{94}$$

and set $\Theta^k = \Theta^0$.

(2) Construct, with the entries of Θ^k, the matrix M^k as in (39) and compute (41)

$$E^k = \left(M^{k^H} W^{-1} M^k\right)^{-1} M^{k^H} W^{-1} V\,; \tag{95}$$

then compute the vector E_0^k by means of (48) and construct the matrix Ω^k (70).

(3) Compute

$$\theta_U^{k+1} = (\Omega^{k^H}\Omega^k)^{-1}\Omega^{k^H}\,V_U \tag{96}$$

$$\theta_Y^{k+1} = (\Omega^{k^H}\Omega^k)^{-1}\Omega^{k^H}\,V_Y \tag{97}$$

and extract vectors θ_α^{k+1}, θ_β^{k+1} from (68) and (69). Set

$$\Theta^{k+1} = [\theta_\alpha^{(k+1)^T}\;\theta_\beta^{(k+1)^T}]^T\,. \tag{98}$$

(4) Test if

$$\frac{\|\Theta^{k+1} - \Theta^k\|}{\|\Theta^k\|} < \varepsilon \tag{99}$$

where ε is an assigned convergence threshold.
Set $\Theta^k = \Theta^{k+1}$ and repeat steps 2 and 3 until the condition (99) is fulfilled.

(5) Let E^*, Θ^* be the final values of E^k, Θ^k. Compute the estimate of the noise variances

$$\lambda_u^* = \frac{1}{N}\,(V - M^*\,E^*)^H\,W^{-1}(V - M^*\,E^*)$$
$$\lambda_y^* = \rho\,\lambda_u^*\,, \tag{100}$$

where M^* (39) can be constructed with the entries of Θ^* according to (29)–(30) and (36)–(37).

By inserting (93) into (87), it can be shown that Algorithm 1 yields an iterative solution of the following weighted least squares problem

$$\min_{\Theta,E} J(\Theta, E)\,, \tag{101}$$

where

$$J(\Theta, E) = \|V - M\,E\|^2_{W^{-1}} = \|V - \Gamma\,\Theta\|^2_{W^{-1}}\,. \tag{102}$$

Note that steps (2) and (3) solve univocally the quadratic minimization problems

$$\min_{E} J(\Theta^k, E) \tag{103}$$

and

$$\min_{\Theta} J(\Theta, E^k) \ , \tag{104}$$

respectively, so that

$$J(\Theta^k, E^k) \leq J(\Theta^k, E^{k-1}) \leq J(\Theta^{k-1}, E^{k-1}) \ . \tag{105}$$

Since $J(\Theta, E) \geq 0$ and every step decreases the value of $J(\Theta, E)$, the convergence of Algorithm 1 to a (local) minimum is guaranteed.

Remark 6. Algorithm 1 is based on an alternating projection procedure, see for example (Grigoriadis and Skelton, 1996). Its structure is completely similar to the one proposed in (Diversi et al., 2007). Thus, for a deeper discussion about an efficient implementation of the algorithm the reader can consult that paper and the references therein reported.

5. THE KOOPMANS–LEVIN SOLUTION

The initial estimate Θ^0 in step 1 can significantly affect the convergence of the algorithm to the global minimum point, with a modest number of iterations. A suitable starting point can be obtained by means of the following frequency domain solution, analogue to the time domain Koopmans–Levin solution (Fernando and Nicholson, 1985).

With reference to system (51)

$$A(\mathrm{e}^{-j\omega_k})\, \hat{Y}(\omega_k) = B(\mathrm{e}^{-j\omega_k})\, \hat{U}(\omega_k) + T(\mathrm{e}^{-j\omega_k}) \ , \tag{106}$$

where

$$A(z^{-1}) = 1 + \alpha_1 z^{-1} + \cdots + \alpha_n z^{-n} \tag{107}$$

$$B(z^{-1}) = \beta_0 + \beta_1 z^{-1} + \cdots + \beta_n z^{-n} \tag{108}$$

$$T(z^{-1}) = \tau_0 + \tau_1 z^{-1} + \cdots + \tau_{n-1} z^{-n+1} \ , \tag{109}$$

construct with the row vectors (57) and (58) the following matrices of dimension $N \times (n+1)$ and $N \times n$, respectively

$$\Pi = \begin{bmatrix} Z_{n+1}(\omega_0) \\ \vdots \\ Z_{n+1}(\omega_{N-1}) \end{bmatrix} \qquad \Psi = \begin{bmatrix} Z_n(\omega_0) \\ \vdots \\ Z_n(\omega_{N-1}) \end{bmatrix} . \tag{110}$$

With the noise–free input–output DFTs (19) and (20) construct the following $N \times N$ diagonal matrices

$$\hat{V}_U^{diag} = \begin{bmatrix} \hat{U}(\omega_0) & 0 & \dots & 0 \\ 0 & \hat{U}(\omega_1) & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & \hat{U}(\omega_{N-1}) \end{bmatrix} \tag{111}$$

$$\hat{V}_Y^{diag} = \begin{bmatrix} \hat{Y}(\omega_0) & 0 & \dots & 0 \\ 0 & \hat{Y}(\omega_1) & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & \hat{Y}(\omega_{N-1}) \end{bmatrix} . \tag{112}$$

Compute the matrices

$$\hat{\Phi}_B = \hat{V}_U^{diag}\, \Pi \quad \hat{\Phi}_A = \hat{V}_Y^{diag}\, \Pi \quad \hat{\Phi}_T = \Psi \ . \tag{113}$$

Defined with $p = 3n + 2$ the whole number of the parameters, construct the $N \times p$ matrix

$$\hat{\Phi} = [\hat{\Phi}_A \,|\, \hat{\Phi}_B \,|\, \hat{\Phi}_T] \ . \tag{114}$$

It then holds

$$\hat{\Sigma}_p\, \Theta_{KL} = 0 \ , \tag{115}$$

where $\hat{\Sigma}_p$ is the $p \times p$ matrix

$$\hat{\Sigma}_p = \frac{1}{N}(\hat{\Phi}^H \hat{\Phi}) \tag{116}$$

and Θ_{KL} is the p–dimensional parameter vector

$$\Theta_{KL} = [\theta_\alpha^T \ -\theta_\beta^T \ -\theta_\tau^T\,]^T \ , \tag{117}$$

with θ_α, θ_β defined in (59), (60) and

$$\theta_\tau = [\tau_{n-1} \dots \tau_0]^T \ . \tag{118}$$

Similar considerations hold for the noisy input–output DFTs (13) and (14). Construct the $N \times N$ diagonal matrices

$$V_U^{diag} = \begin{bmatrix} U(\omega_0) & 0 & \dots & 0 \\ 0 & U(\omega_1) & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & U(\omega_{N-1}) \end{bmatrix} \tag{119}$$

$$V_Y^{diag} = \begin{bmatrix} Y(\omega_0) & 0 & \dots & 0 \\ 0 & Y(\omega_1) & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & Y(\omega_{N-1}) \end{bmatrix} , \tag{120}$$

compute the matrices

$$\Phi_B = V_U^{diag}\, \Pi \quad \Phi_A = V_Y^{diag}\, \Pi \quad \Phi_T = \Psi \tag{121}$$

and construct the $N \times p$ matrix

$$\Phi = [\Phi_A \,|\, \Phi_B \,|\, \Phi_T] \ . \tag{122}$$

Because of Assumptions A5 and A6 we obtain the following $p \times p$ positive definite matrix

$$\Sigma_p = \frac{1}{N}(\Phi^H \Phi) = \hat{\Sigma}_p + \tilde{\Sigma}_p \ , \tag{123}$$

where

$$\tilde{\Sigma}_p = \begin{bmatrix} \lambda_y I_{n+1} & 0 & 0 \\ 0 & \lambda_u I_{n+1} & 0 \\ 0 & 0 & 0_n \end{bmatrix} \tag{124}$$

$$= \lambda_u \begin{bmatrix} \rho I_{n+1} & 0 & 0 \\ 0 & I_{n+1} & 0 \\ 0 & 0 & 0_n \end{bmatrix} = \lambda_u\, \Lambda \ . \tag{125}$$

It is a well-known result (Guidorzi et al., 2008) that the parameter vector Θ_{KL}, defined in (117), can be obtained as the kernel of

$$\left(\Sigma_p - \lambda_u\, \Lambda\right) \Theta_{KL} = 0 \ , \tag{126}$$

with

$$\frac{1}{\lambda_u} = \max \text{eig}\left(\Sigma_p^{-1}\, \Lambda\right) . \tag{127}$$

6. NUMERICAL RESULTS

As a first case, the proposed algorithm has been tested on sequences generated by a second–order model of type (1), already proposed in (Diversi et al., 2007)

$$A(z^{-1}) = 1 - 0.5\, z^{-1} + 0.3\, z^{-2} \tag{128}$$

$$B(z^{-1}) = 2 - 1.2\, z^{-1} - 0.6\, z^{-2} \ . \tag{129}$$

The input is a pseudo random binary sequence with unit variance and length $N = 250$. A Monte Carlo simulation of 100 independent runs has been performed by adding to the noise–free sequences $\hat{u}(\cdot)$, $\hat{y}(\cdot)$ different Gaussian white noise realizations with variances $\lambda_u = 0.1$, $\lambda_y = 0.6$, corresponding to a signal to noise ratio (SNR) of about 10 dB on both input and output.

Table 1. True and estimated parameters of $q(z^{-1})$ obtained by means of FD, TD Algorithms and FD, TD Koopmans–Levin methods

	α_1	α_2
true	-0.5	0.3
FD Alg.	-0.5006 ± 0.0601	0.3029 ± 0.0530
TD Alg.	-0.5006 ± 0.0595	0.3029 ± 0.0527
FD − KL	-0.5055 ± 0.0688	0.3109 ± 0.0706
TD − KL	-0.5055 ± 0.0688	0.3109 ± 0.0707

Table 2. True and estimated parameters of $p(z^{-1})$ obtained by means of FD, TD Algorithms and FD, TD Koopmans–Levin methods

	β_0	β_1	β_2
true	2	-1.2	-0.6
FD Alg.	2.0004 ± 0.0232	-1.2070 ± 0.1421	-0.5897 ± 0.1310
TD Alg.	2.0005 ± 0.0221	-1.2069 ± 0.1406	-0.5899 ± 0.1296
FD − KL	2.0064 ± 0.0413	-1.2179 ± 0.1632	-0.5711 ± 0.1763
TD − KL	2.0065 ± 0.0416	-1.2178 ± 0.1633	-0.5710 ± 0.1765

Table 3. True and estimated values of λ_u, λ_y obtained by means of FD, TD Algorithms and FD, TD Koopmans–Levin methods

	λ_u	λ_y
true	0.1	0.6
FD Alg.	0.0975 ± 0.0040	0.5853 ± 0.0242
TD Alg.	0.0975 ± 0.0040	0.5852 ± 0.0242
FD − KL	0.0967 ± 0.0060	0.5803 ± 0.0359
TD − KL	0.0979 ± 0.0061	0.5873 ± 0.0363

The ML estimates of the parameters and variances have been obtained by using $\varepsilon = 10^{-3}$ in step 6 of the Frequency Domain (FD) Algorithm proposed in Section 4.

Tables 1 and 2 report the empirical means of the parameter estimates together with the corresponding standard deviations, obtained with the FD methods and with the Time Domain (TD) Algorithm described in (Diversi et al., 2007). The estimates obtained by the corresponding Koopmans–Levin (KL) solutions are also reported.

Table 3 reports the empirical means of the noise variance estimates and the corresponding standard deviations, obtained with the FD and the TD Algorithms, together with the corresponding Koopmans–Levin solutions.

These tables seem to well illustrate the fact that the TD and FD methods coincide. In fact both algorithms lead to equal results, not only for the estimates but also for the standard deviations.

In order to verify the improvement of the accuracy in the estimates for increasing values of data, a Monte Carlo simulation of 100 independent runs has been also performed with $N = 125$, $N = 250$ and $N = 500$. The results confirm the expectations, as shown in Tables 4–5, with reference to the system parameters. Similar results hold for the noise variances.

Table 4. True and estimated parameters of $q(z^{-1})$ obtained by means of FD Algorithm for different data length N

	α_1	α_2
true	-0.5	0.3
$N = 125$	-0.5009 ± 0.0824	0.2997 ± 0.0736
$N = 250$	-0.5006 ± 0.0601	0.3029 ± 0.0530
$N = 500$	-0.4991 ± 0.0445	0.3018 ± 0.0343

Table 5. True and estimated parameters of $p(z^{-1})$ obtained by means of FD Algorithm for different data length N

	β_0	β_1	β_2
true	2	-1.2	-0.6
$N = 125$	1.9967 ± 0.0353	-1.1950 ± 0.1900	-0.6063 ± 0.1955
$N = 250$	2.0004 ± 0.0232	-1.2070 ± 0.1421	-0.5897 ± 0.1310
$N = 500$	2.0000 ± 0.0165	-1.2002 ± 0.1059	-0.5903 ± 0.0985

7. CONCLUSIONS

In this paper a new frequency domain ML method has been proposed for the identification of EIV models with additive white noises. The method applies for general inputs, but requires the *a priori* knowledge of the noise variance ratio. The effectiveness of the proposed algorithm has been verified by means of Monte Carlo simulations.

REFERENCES

Agüero, J.C., Yuz, J.I., Goodwin, G.C. and Delgado, R.A. (2010). On the equivalence of time and frequency domain maximum likelihood estimation. *Automatica*, 46, 260–270.

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

Fernando, K. V., Nicholson, H. (1985). Identification of linear systems with input and output noise: the Koopmans–Levin method. *IEE Proceedings*, Part D, 132, 30–36.

Fuller, W. A.(1987). *Measurement Error Models*. New York, NY, USA: Wiley.

Grigoriadis, K.M., Skelton, R. (1996). Low–order control design for LMI problems using Alternating Projection Methods. *Automatica*, 32(8), 1117–1125.

Guidorzi, R., Diversi, R. and Soverini, U. (2008). The Frisch Scheme in algebraic and dynamic identification problems. *Kybernetika*, 44(5), 585–616.

Ljung, L. (1999). *System identification–Theory for the user*. (2nd ed.), Upper Saddle River, NJ, USA: Prentice–Hall.

Pintelon, R., Schoukens, J. and Vandersteen G. (1997). Frequency domain system identification using arbitrary signals. *IEEE Transactions on Automatic Control*, 42, 1717–1720.

Söderström, T. (2007). Errors–in–Variables methods in system identification. *Automatica*, 43(6), 939–958.

Söderström, T. (2012). System identification for the errors–in–variables problem. *Transaction of the Institute of Measurement and Control*, 34(7), 780–792.

Van Huffel, S., and Lemmerling, P. (Eds.) (2002). *Total Least Squares Techniques and Errors–in–Variables Modelling: Analysis, Algorithms and Applications*. Dordrecht, The Netherlands: Kluwer Academic Publishers.