Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Blind identification of two–channel FIR systems: a frequency domain approach

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: torsten.soderstrom@it.uu.se)*

Abstract: This paper describes a new approach for the blind identification of a two–channel FIR system from a finite number of output measurements, in the presence of additive and uncorrelated white noise. The proposed approach is based on frequency domain data and, as a major novelty, it enables the estimation to be frequency selective. The features of the proposed method are analyzed by means of Monte Carlo simulations. The benefits of filtering the data and using only part of the frequency domain are highlighted by means of a numerical example.

Keywords: Blind identification; FIR systems; Discrete Fourier Transform.

1. INTRODUCTION

Blind system identification is a fundamental problem in signal processing aimed at estimating the channel impulse response by using the output observations only, since the input signal is either unknown or not easily accessible. Problems of this type arise in many important applications such as data transmission, speech recognition, reverberation cancellation, seismic deconvolution, biomedical data analysis, image restoration (Abed–Meraim et al., 1997b; Giannakis et al., 2001). In most of these applications, the unknown system is described by a single–input multi–output (SIMO) finite impulse response (FIR) model, whose outputs are affected by additive noise. For this problem, two different approaches are usually considered: methods based on second–order statistics (SOS) and approaches based on higher–order statistics (HOS). HOS methods are based on optimization techniques employing gradient–based algorithms and are characterized by the drawbacks of slow convergence and local minima. Since it was recognized in (Tong et al., 1991) that the problem can be solved by using SOS only, the study of blind identification with SOS has had a rapid development. Many techniques have been proposed, such as the subspace algorithm (Moulines et al., 1995), the cross–relation algorithm (Xu et al., 1995) and the two–step maximum likelihood algorithm (Hua, 1996). A review of many SOS methods can be found in (Tong and Perreau, 1998). Most of the existing approaches assume the presence of the same amount of additive white noise on all the channels or, at least, the a priori knowledge of the ratio of the noises' variances. A solution for the case of unknown unbalanced noise environments has been proposed in (Diversi et al., 2005, 2007), on the basis of Errors–in–Variables (EIV) identification techniques.

The method proposed in this paper relies on the cross–relation property and deals with the blind identification of two–channel FIR systems, whose outputs are affected by different and unknown amounts of additive white noise. As a major novelty, the algorithm makes use of frequency domain data. This feature offers an easy implementation of the filtering operation, which is reduced to the selection of an appropriate limited band of the signal spectrum. This property can be advantageously used for obtaining good identification results even when the additive noise is not white, provided that some information of its characteristic is available.

The organization of the paper is as follows. Section 2 defines the blind identification problem in the frequency domain, while Section 3 introduces a novel frequency domain description for the noisy two–channel FIR model. Section 4 discusses some contexts for the identification of EIV models. In particular, the GIVE framework, originally proposed in (Söderström, 2011), and the dynamic Frisch scheme, originally proposed in (Beghelli et al., 1990), are briefly recalled. Sections 5 describes a possible identification criterion, that can be directly formulated in the frequency domain. This criterion takes advantage of two sets of equations similar to the High Order Yule Walker (HOYW) equations. The method can be considered as the application of the approach proposed in (Soverini and Söderström, 2015) to the blind identification of FIR models. In Section 6 the effectiveness of the proposed method is verified by means of Monte Carlo simulations and the advantages of the filtering operations in the frequency domain are illustrated by means of a numerical example. Finally some concluding remarks are reported in Section 7.

2. STATEMENT OF THE PROBLEM

Consider a two–channel FIR system whose outputs $x_1(t)$, $x_2(t)$ are linked to the input $u(t)$ by the convolution model

$$x_1(t) = H_1(z^{-1})u(t) = \sum_{k=0}^{L} h_1(k)\, u(t-k) \tag{1}$$

$$x_2(t) = H_2(z^{-1})u(t) = \sum_{k=0}^{L} h_2(k)\, u(t-k), \tag{2}$$

where $H_1(z^{-1})$, $H_2(z^{-1})$ are polynomials of degree L in the unitary delay operator z^{-1} of the type

$$H_i(z^{-1}) = h_i(0) + h_i(1)\, z^{-1} + \cdots + h_i(L)\, z^{-L} \quad i = 1, 2. \tag{3}$$

This model is useful to describe the case of a single unknown source and multiple spatially or/and temporally distributed sensors (Tong and Perreau, 1998; Xu et al., 1995).

The channel outputs are affected by additive noise, so that the available signals are

$$y_1(t) = x_1(t) + n_1(t) \tag{4}$$
$$y_2(t) = x_2(t) + n_2(t). \tag{5}$$

The following assumptions are made.

A1. The order L of the FIR channels is assumed as *a priori* known.
A2. The input $u(t)$ can be either a zero–mean ergodic process or a quasi–stationary bounded deterministic signal (Ljung, 1999).
A3. The additive noises $n_1(t)$ and $n_2(t)$ are zero–mean ergodic white processes with *unknown* variances σ_1^* and σ_2^*.
A4. $n_1(t)$, $n_2(t)$ and $u(t)$ are mutually uncorrelated.

In order to guarantee the system identifiability (up to a scalar factor), the following additional assumptions must be introduced (Hua and Wax, 1996).

A5. The polynomials $H_1(z^{-1})$ and $H_2(z^{-1})$ do not share any common factor.
A6. The input $u(t)$ is persistently exciting of order $2L + 1$.
A7. The number of the available samples N satisfies the condition $N \geq 3L + 1$.

Let $\{y_1(t)\}_{t=0}^{N-1}$ and $\{y_2(t)\}_{t=0}^{N-1}$ be the sets of output observations at N equidistant time instants. For $\{y_1(t)\}_{t=0}^{N-1}$, the corresponding Discrete Fourier Transform (DFT) is defined as

$$Y_1(\omega_k) = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} y_1(t)\, \mathrm{e}^{-\mathrm{j}\,\omega_k t} \tag{6}$$

where $\omega_k = 2\pi k/N$ and $k = 0, \ldots, N-1$. Similarly, let $Y_2(\omega_k)$ be the DFT of $\{y_2(t)\}_{t=0}^{N-1}$. In the frequency domain, the problem under investigation can be stated as follows.

Problem 1. Let $Y_1(\omega_k)$, $Y_2(\omega_k)$ be the sets of noisy measurements generated by a two–channel FIR system of type (1)–(5), under Assumptions A1–A7. Estimate the coefficients of $H_1(z^{-1})$, $H_2(z^{-1})$ (up to a scalar factor) and the variances σ_1^*, σ_2^*.

Remark 1. For real–valued signals, the following consideration holds for every N, even or odd, see also (McKelvey, 2002). Let $s(t)$ denote either $y_1(t)$ or $y_2(t)$. It can be observed that for $k = 0, \ldots, \text{floor}\left(\frac{N}{2}\right)$

$$\begin{aligned} S(\omega_{N-1-k}) &= \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} s(t)\, \mathrm{e}^{-\mathrm{j}\,\frac{N-1-k}{N} 2\pi t} \\ &= \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} s(t)\, \mathrm{e}^{-\mathrm{j}\,\frac{-(1+k)}{N} 2\pi t} = \bar{S}(\omega_{1+k}), \end{aligned} \tag{7}$$

where $\bar{S}(\cdot)$ is the complex conjugate of $S(\cdot)$. Thus, a redundant information is used when the full data set is considered. In fact, it is worth observing that the algorithm proposed in Sections 5 yields consistent estimates of the system parameters by using only the first $N_{\text{half}} = \text{ceil}\big((N+1)/2\big)$ samples $Y_1(\omega_k)$, $Y_2(\omega_k)$, $k = 0, \ldots, \text{floor}\left(\frac{N}{2}\right)$.

3. A FREQUENCY DOMAIN SETUP

In this section a new frequency domain description for the noisy two–channel FIR model (1)–(5) is introduced. This setup has been originally developed in (Soverini and Söderström, 2014a,b) with reference to the identification of errors–in–variables systems.

In absence of noise, relations (1) and (2) lead immediately to the following well–known cross-relation property (Xu et al., 1995)

$$H_2(z^{-1})x_1(t) = H_1(z^{-1})x_2(t). \tag{8}$$

Working with frequency data, the previous relation must be modified as follows. Similarly to equation (6), let $X_1(\omega_k)$ and $X_2(\omega_k)$ be the DFTs of the signals $x_1(t)$ and $x_2(t)$ appearing in equation (8). It is a well–known fact (Pintelon et al, 1997; McKelvey, 2002) that for finite N the DFTs $X_1(\omega_k)$ and $X_2(\omega_k)$ exactly satisfy an extended relation that includes also a transient term, i.e.

$$H_2(\mathrm{e}^{-j\omega_k})\, X_1(\omega_k) = H_1(\mathrm{e}^{-j\omega_k})\, X_2(\omega_k) + T(\mathrm{e}^{-j\omega_k}), \tag{9}$$

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

$$T(z^{-1}) = \tau_0 + \tau_1\, z^{-1} + \cdots + \tau_{L-1}\, z^{-L+1} \tag{10}$$

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

By considering the whole number of frequencies, eq. (9) can be rewritten in a matrix form. For this purpose, introduce the parameter vectors

$$\theta_1 = [h_1(0)\, h_1(1) \ldots h_1(L)]^T \tag{11}$$
$$\theta_2 = [h_2(0)\, h_2(1) \ldots h_2(L)]^T \tag{12}$$
$$\theta_\tau = [\,\tau_0 \ldots \tau_{L-1}\,]^T \tag{13}$$

and define the following vector Θ, with dimension (cf. Ass. A7)

$$p = 3L + 2\,, \tag{14}$$

containing the whole set of parameters

$$\Theta = [\,\theta_2^T \;\, -\theta_1^T \;\, -\theta_\tau^T\,]^T. \tag{15}$$

In absence of noise, the parameter vector (15) can be recovered by means of the following procedure. Define the row vectors

$$Z_{L+1}(\omega_k) = [1 \;\, \mathrm{e}^{-j\omega_k} \;\ldots\; \mathrm{e}^{-j(L-1)\omega_k} \;\, \mathrm{e}^{-jL\omega_k}\,] \tag{16}$$
$$Z_L(\omega_k) = [1 \;\, \mathrm{e}^{-j\omega_k} \;\ldots\; \mathrm{e}^{-j(L-1)\omega_k}\,], \tag{17}$$

whose entries are constructed with multiple frequencies of ω_k, and construct the following matrices

$$\Pi = \begin{bmatrix} Z_{L+1}(\omega_0) \\ \vdots \\ Z_{L+1}(\omega_{N-1}) \end{bmatrix} \qquad \Psi = \begin{bmatrix} Z_L(\omega_0) \\ \vdots \\ Z_L(\omega_{N-1}) \end{bmatrix} \tag{18}$$

of dimension $N \times (L+1)$ and $N \times L$, respectively.

From the DFT samples $X_1(\omega_k)$, $X_2(\omega_k)$ construct the following $N \times N$ diagonal matrices

$$V_{X_1}^{diag} = \text{diag}\,[X_1(\omega_0),\, X_1(\omega_1), \ldots, X_1(\omega_{N-1})] \tag{19}$$
$$V_{X_2}^{diag} = \text{diag}\,[X_2(\omega_0),\, X_2(\omega_1), \ldots, X_2(\omega_{N-1})]. \tag{20}$$

Compute the $N \times (L+1)$ matrices

$$\Phi_{X_1} = V_{X_1}^{diag}\, \Pi \quad \Phi_{X_2} = V_{X_2}^{diag}\, \Pi \tag{21}$$

and construct the $N \times p$ matrix

$$\hat{\Phi} = [\Phi_{X_1} \mid \Phi_{X_2} \mid \Psi]. \tag{22}$$

Thus, eq. (9) for $k = 0, \ldots, N-1$ can be rewritten as

$$\hat{\Phi}\, \Theta = 0. \tag{23}$$

It then holds

$$\hat{\Sigma}\,\Theta = 0, \tag{24}$$

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

$$\hat{\Sigma} = \frac{1}{N}(\hat{\Phi}^H \hat{\Phi}). \tag{25}$$

Remark 2. Because of assumption A5, relation (9) cannot be satisfied by polynomials $H_1(z^{-1})$ and $H_2(z^{-1})$ with order lower than L. Therefore, the matrix $\hat{\Sigma}$ in (25) is positive semidefinite, with only one zero eigenvalue, i.e.

$$\hat{\Sigma} \geq 0 \qquad \dim \ker \hat{\Sigma} = 1\,. \tag{26}$$

In the presence of noise, the previous procedure can be modified as follows. With the noisy DFT samples $Y_1(\omega_k)$, $Y_2(\omega_k)$ construct the $N \times N$ diagonal matrices

$$V_{Y_1}^{diag} = \mathrm{diag}\,[Y_1(\omega_0),\, Y_1(\omega_1), \ldots, Y_1(\omega_{N-1})] \tag{27}$$

$$V_{Y_2}^{diag} = \mathrm{diag}\,[Y_2(\omega_0),\, Y_2(\omega_1), \ldots, Y_2(\omega_{N-1})], \tag{28}$$

compute the matrices

$$\Phi_{Y_1} = V_{Y_1}^{diag}\,\Pi \quad \Phi_{Y_2} = V_{Y_2}^{diag}\,\Pi \tag{29}$$

and construct the $N \times p$ matrix

$$\Phi = [\Phi_{Y_1} \,|\, \Phi_{Y_2} \,|\, \Psi]. \tag{30}$$

Because of Assumptions A3–A4, when $N \to \infty$, we obtain the following $p \times p$ positive definite matrix

$$\Sigma = \lim_{N\to\infty} \frac{1}{N}(\Phi^H \Phi) = \hat{\Sigma} + \tilde{\Sigma}^*, \tag{31}$$

where

$$\tilde{\Sigma}^* = \mathrm{diag}\left[\,\sigma_1^*\, I_{L+1},\, \sigma_2^*\, I_{L+1},\, 0_L\right]. \tag{32}$$

The matrix $\hat{\Sigma}$ in (31) must be interpreted as the limit for $N \to \infty$ of (25). Of course, the property (24) still holds. Thus, the parameter vector Θ, defined in (15), can be obtained as the kernel of

$$\left(\Sigma - \tilde{\Sigma}^*\right)\Theta = 0. \tag{33}$$

By considering the particular structure of the matrices Σ and $\tilde{\Sigma}^*$, Problem 1 can be solved by considering a system of equations of type (33) with reduced dimensions. For this purpose, the matrix Σ, defined in (31), is partitioned as follows

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{bmatrix}, \tag{34}$$

where Σ_{11} and Σ_{22} are square matrices of dimension $L+1$ and Σ_{33} is a square matrix of dimension L. In a similar way, the matrix $\hat{\Sigma}$ in equation (31) is also partitioned.

Relation (33) can be expanded as follows

$$\hat{\Sigma}_{11}\,\theta_2 - \Sigma_{12}\,\theta_1 - \Sigma_{13}\,\theta_\tau = 0 \tag{35}$$

$$\Sigma_{21}\,\theta_2 - \hat{\Sigma}_{22}\,\theta_1 - \Sigma_{23}\,\theta_\tau = 0 \tag{36}$$

$$\Sigma_{31}\,\theta_2 - \Sigma_{32}\,\theta_1 - \Sigma_{33}\,\theta_\tau = 0. \tag{37}$$

Next (37) implies

$$\theta_\tau = \Sigma_{33}^{-1}\left(\Sigma_{31}\,\theta_2 - \Sigma_{32}\,\theta_1\right). \tag{38}$$

The expression (38) can be substituted in (35) and (36), and the following problem of reduced dimension can be defined

$$\left(R - \tilde{R}^*\right)\theta = 0, \tag{39}$$

where

$$R = \begin{bmatrix} \Sigma_{11} - \Sigma_{13}\Sigma_{33}^{-1}\Sigma_{31} & \Sigma_{12} - \Sigma_{13}\Sigma_{33}^{-1}\Sigma_{32} \\ \Sigma_{21} - \Sigma_{23}\Sigma_{33}^{-1}\Sigma_{31} & \Sigma_{22} - \Sigma_{23}\Sigma_{33}^{-1}\Sigma_{32} \end{bmatrix}, \tag{40}$$

$$\tilde{R}^* = \mathrm{diag}\left[\,\sigma_1^*\, I_{L+1},\, \sigma_2^*\, I_{L+1}\right] \tag{41}$$

and the parameter vector θ, with dimension (cf. Ass. A6)

$$p_\theta = 2L + 2\,, \tag{42}$$

is defined as

$$\theta = [\theta_2^T \;\; -\theta_1^T\,]^T. \tag{43}$$

4. INTERPRETATION AS AN ERRORS–IN–VARIABLES PROBLEM

4.1 The GIVE framework

Note that (39) consists of $2L+2$ algebraic non–linear equations. The number of unknowns is $2L+3$, i.e. the $2L+1$ free coefficients of θ and the two variances σ_1^* and σ_2^*. In the time domain, a similar set of equations has been widely studied in the identification of EIV dynamic systems. A general framework has been originally introduced in (Söderström, 2011), where the Generalized Instrumental Variable Estimation (GIVE) method was proposed with reference to SISO EIV systems affected by additive white noises. The GIVE method provides a unique general framework for the whole class of bias–compensating methods, including iterative solutions, like the BELS methods (Söderström et al, 2005). The GIVE framework leads to the following conclusions, that are common to the whole class of bias–compensating methods.

1 Since the number of the unknowns is larger than the number of equations, some further equations need to be used in addition to the system (39) in order to find a unique estimate of θ. It can result in an over–determined system of equations. In the time domain, a natural solution is to exploit the high–order Yule–Walker equations, where the noise variances are not present. Indeed, these are the equations exploited also by the method proposed in this paper. In Section 5 we will see how these equations can be written in the frequency domain.
2 In the general case the parameter estimates are obtained as the solution of the solution to an overdetermined system of equations, leading to an optimization problem. A usual solution strategy consists in forcing some equations to hold exactly, while the others are minimized in a weighted least squares sense. This will be described as introduction of a search criterion in Section 5.
3 This second aspect does not affect the statistical properties of the estimates, since the asymptotic accuracy depends only on the set of equations used to define the problem and not on the way the equations are solved (Söderström et al, 2005). Nevertheless, in practice, different identification algorithms that are based on the same set of equations can lead to different estimation results, in terms of computational complexity and speed of convergence.

4.2 The Frisch scheme context

The purpose of this subsection is to recall the Frisch scheme (Beghelli et al., 1990; Guidorzi et al., 2008; Söderström, 2018) for developing the estimation algorithm of Section 5. This can be viewed as a possible numerical strategy to solve the ‘exact’ equations within the GIVE framework, as stated in the point 2.

Starting from an assumed knowledge of the noisy matrix R in (40), the determination of the system parameter vector θ and of the noise variances σ_1^*, σ_2^* in eq. (39) can be seen as a Frisch scheme problem. This problem can be solved both in the time and in the frequency domain. In fact, the properties of the locus of the solutions in the noise plane $\mathcal{R}^2$ are the same. In the following only the main result of the Frisch scheme is recalled. All the technical aspects, with the related proofs, are reported in (Soverini and Söderström, 2019).

Consider the set of non–negative definite diagonal matrices of type

$$\tilde{R} = \text{diag}\left[\,\sigma_1\, I_{L+1},\, \sigma_2\, I_{L+1}\right] \qquad (44)$$

such that

$$R - \tilde{R} \geq 0 \qquad \det\left(R - \tilde{R}\right) = 0. \qquad (45)$$

Main Result. The set of all matrices $\tilde{R}$ satisfying the conditions (45) defines the points $P = (\sigma_1, \sigma_2)$ of a continuous curve $\mathcal{S}(R)$ belonging to the first quadrant of the noise space $\mathcal{R}^2$. The curve $\mathcal{S}(R)$ describes a convex set in the first quadrant of $\mathcal{R}^2$, whose concavity faces the origin. When $N \to \infty$, the point $P^* = (\sigma_1^*, \sigma_2^*)$, associated with the true variances of $n_1(t)$ and $n_2(t)$, belongs to $\mathcal{S}(R)$.

Corollary. Every point $P = (\sigma_1, \sigma_2)$ of $\mathcal{S}(R)$ can be associated with a noise matrix of type $\tilde{R}(P)$ in (44) and with a coefficient vector $\theta(P)$, satisfying the relation

$$\left(R - \tilde{R}(P)\right)\, \theta(P) = 0. \qquad (46)$$

The coefficient vector $\theta(P^*)$, associated with $\tilde{R}(P^*) = \tilde{R}^*$, is characterized (after a normalization of its first entry to 1) by the true system parameter vector, i.e. $\theta(P^*) = \theta$.

5. A SEARCH CRITERION BASED ON HOYW–TYPE EQUATIONS

As asserted in Section 4, the determination of the point P^* on $\mathcal{S}(R)$ leads to the solution of Problem 1. Unfortunately, the theoretic properties of $\mathcal{S}(R)$ described so far do not allow to distinguish point P^* from the other points of the curve. Some additional conditions must be added to define a unique estimate. One possibility is to introduce a search condition, i.e. an optimization criterion, taking the equation (46) as a constraint.

In this section we will describe a possible search criterion. This criterion is analogue to that reported in (Soverini and Söderström, 2015) with reference to frequency domain identification of EIV systems.

Select the integer $q \geq 3L + 1$. Analogously to (16), consider the row vector

$$Z_{q+L+1}(\omega_k) = [1\ \mathrm{e}^{-j\omega_k}\ \ldots \mathrm{e}^{-j(q+L)\omega_k}] \qquad (47)$$

and extract from it the q–dimensional row vector

$$Z_q^h(\omega_k) = [\mathrm{e}^{-j(L+1)\omega_k}\ \ldots \mathrm{e}^{-j(q+L)\omega_k}]. \qquad (48)$$

Then, construct the following $N \times q$ matrix

$$\Pi^h = \begin{bmatrix} Z_q^h(\omega_0) \\ \vdots \\ Z_q^h(\omega_{N-1}) \end{bmatrix}. \qquad (49)$$

Due to the complete symmetry of the problem it is preferable to treat the two system outputs in the same way. Thus, with reference to the noise–free data, we can compute the following $N \times q$ matrices

$$\Phi^h_{X_1} = V^{diag}_{X_1}\, \Pi^h \qquad \Phi^h_{X_2} = V^{diag}_{X_2}\, \Pi^h \qquad (50)$$

and define the $q \times p$ matrices

$$\Sigma^h_{X_1} = \frac{1}{N}\big((\Phi^h_{X_1})^H\, \hat{\Phi}\big) \quad \Sigma^h_{X_2} = \frac{1}{N}\big((\Phi^h_{X_2})^H\, \hat{\Phi}\big). \qquad (51)$$

Because of (23) we have

$$\Sigma^h_{X_1}\, \Theta = 0 \qquad \Sigma^h_{X_2}\, \Theta = 0. \qquad (52)$$

In an analogous way, in the noisy case, we can compute the $N \times q$ matrices

$$\Phi^h_{Y_1} = V^{diag}_{Y_1}\, \Pi^h \qquad \Phi^h_{Y_2} = V^{diag}_{Y_2}\, \Pi^h \qquad (53)$$

and define the $q \times p$ matrices

$$\Sigma^h_{Y_1} = \frac{1}{N}\big((\Phi^h_{Y_1})^H\, \Phi\big) \quad \Sigma^h_{Y_2} = \frac{1}{N}\big((\Phi^h_{Y_2})^H\, \Phi\big). \qquad (54)$$

Because of Assumptions A3–A4, when $N \to \infty$, it results in

$$\Sigma^h_{Y_1} = \Sigma^h_{X_1} \qquad \Sigma^h_{Y_2} = \Sigma^h_{X_2}. \qquad (55)$$

It is thus possible to write

$$\Sigma^h_{Y_1}\, \Theta = 0 \qquad \Sigma^h_{Y_2}\, \Theta = 0. \qquad (56)$$

Remark 4. The two linear sets of q equations appearing in (56) do not involve the noise variances σ_1^*, σ_2^* and correspond to the time domain high order Yule–Walker equations. Each set of equations could be directly used to obtain an estimate of the parameter vector Θ if $q \geq 3L + 1$. The two sets of equations are analogue to two instrumental variable (IV) methods in the time domain, where delayed outputs (for example, in $\Sigma^h_{Y_1}$) and delayed inputs (for example, in $\Sigma^h_{Y_2}$) are used as instruments.

Remark 5. The value of the parameter q is a user choice. In general, this value can affect the quality of the estimates. However its influence is not straightforward to investigate. In all cases of Section 6, q has been chosen as $q = 3L + 1$.

The two sets of equations in (56) can be collected in one set, by defining the matrix

$$\Sigma^h = \begin{bmatrix} \Sigma^h_{Y_1} \\ \Sigma^h_{Y_2} \end{bmatrix}, \qquad (57)$$

so that we obtain

$$\Sigma^h\, \Theta = 0. \qquad (58)$$

By using again (38), the system of equations (58) can be reduced to a system of equations with θ as unknown, thus reducing the number of unknowns. For this purpose, partition matrix Σ^h as follows

$$\Sigma^h = \left[\Sigma^h_1\ \Sigma^h_2\ \Sigma^h_3\right], \qquad (59)$$

where the matrices Σ^h_1 and Σ^h_2 have dimensions $2q \times (L + 1)$ and Σ^h_3 has dimension $2q \times L$. Thanks to (38), equation (58) can be reduced to

$$R^h\, \theta = 0, \qquad (60)$$

where θ has been defined in (43) and

$$R^h = \left[\Sigma^h_1 - \Sigma^h_3 \Sigma^{-1}_{33} \Sigma_{31}\quad \Sigma^h_2 - \Sigma^h_3 \Sigma^{-1}_{33} \Sigma_{32}\right]. \qquad (61)$$

Remark 6. The set of $(2L + 2)$ non–linear equations (39) can be joined to the set of $2q$ linear equations (60), with $q \geq 3L + 1$ and can be settled within the GIVE framework, as described in Subsection 4.1. Thus, the results in (Söderström, 2011) and (Söderström, 2018) can be applied, to express the statistical accuracy in terms of the theoretical asymptotic covariance matrix of the parameter estimates. More precisely, applying the

Frisch scheme described in Subsection 4.2, the equations (39) are treated as a constraint that must be exactly satisfied, while the equations (60) must hold approximately. In other words, the search for P^* along $\mathcal{S}(R)$ can be performed by minimizing a quadratic cost function.

For every point P on $\mathcal{S}(R)$, one can compute the parameter vector $\theta(P)$ defined in (46) and evaluate the cost function

$$J(P) = \|R^h\,\theta(P)\|_2^2 = \theta^T(P)(R^h)^H\,R^h\,\theta(P) \qquad (62)$$

which exhibits the following properties

$$\text{i) } J(P) \geq 0 \qquad \text{ii) } J(P) = 0 \Leftrightarrow P = P^*.$$

In general, for a finite number of data, the minimum of $J(P)$ will not be exactly obtained for $P = P^*$. Note that the search algorithm moves radially on $\mathcal{S}(R)$, thanks to the equations (64)–(66).

On the basis of the previous considerations, the following algorithm can be derived.

Algorithm 1.

(1) Compute, on the basis of the available time domain data, the DFTs $Y_1(\omega_k)$, $Y_2(\omega_k)$ with $\omega_k = 2\pi k/N$ ($k = 0, \ldots, N-1$).
(2) Compute the matrices Φ_{Y_1}, Φ_{Y_2} as in (29) and construct the matrix Φ as in (30).
(3) Compute, as in (31), the sample estimate of matrix
$$\Sigma = \frac{1}{N}(\Phi^H\Phi) \qquad (63)$$
and compute the matrix R by means of (40).
(4) Select $q \geq 3L+1$ and construct the matrix Π^h as in (49), then compute the matrices $\Phi^h_{Y_1}$, $\Phi^h_{Y_2}$ as in (53).
(5) Compute, as in (54), the sample estimate of the matrices $\Sigma^h_{Y_1}$ and $\Sigma^h_{Y_2}$. Collect them in the matrix Σ^h as in (57) and compute the matrix R^h by means of (61).
(6) Start from a generic point $\xi = (\xi_1, \xi_2)$ (a generic direction) in the first quadrant of $\mathcal{R}^2$ and compute the corresponding point $P = (\sigma_1, \sigma_2)$ on $\mathcal{S}(R)$, by means of
$$\tilde{R}_\xi = \operatorname{diag}\left[\,\xi_1\, I_{L+1},\, \xi_2\, I_{L+1}\right] \qquad (64)$$
$$\lambda_M = \max \operatorname{eig}\left(R^{-1}\tilde{R}_\xi\right) \qquad (65)$$
$$\sigma_1 = \frac{\xi_1}{\lambda_M} \qquad \sigma_2 = \frac{\xi_2}{\lambda_M}. \qquad (66)$$
(7) Compute the estimates of $\hat{R}(P)$ and $\theta(P)$ by means of the relations
$$\hat{R}(P) = R - \operatorname{diag}\left[\,\sigma_1 I_{L+1},\, \sigma_2\, I_{L+1}\right], \qquad (67)$$
$$\hat{R}(P)\,\theta(P) = 0. \qquad (68)$$
(8) Compute the value of the cost function $J(P)$ (62).
(9) Search on the curve $\mathcal{S}(R)$ for the point associated with the minimum of $J(P)$.

Remark 7. Algorithm 1 makes reference to the case of N data. As observed in Remark 1, if only $N_{\text{half}} = \text{floor}\,(N/2)$ data are considered, the algorithm must be modified in a straightforward fashion, by substituting N with N_{half} starting from step (2).

Remark 8. As stated in *Problem 1*, the coefficients of the FIR systems $H_1(z^{-1})$ and $H_2(z^{-1})$, i.e. the entries of the vector θ in (43), can be estimated up to a scalar factor, only. Thus, a natural choice for comparing the identification results is to normalize θ and its estimate $\hat{\theta}$ to unit norm vectors. However, at a generic i–th iteration of the algorithm, the normalization to unity can lead to some numerical problems as far as the evaluation of the cost function $J(P)$ in (62) is concerned. Better results can be obtained by taking into account the symmetry of the problem, avoiding any preference between the two vectors θ_1 and θ_2. For this reason, at the i-th iteration of the algorithm, before evaluating the cost function $J(P)$, the estimate $\hat{\theta}_i$ obtained in (68) has been normalized as follows

$$\theta_a = \frac{\hat{\theta}_i}{\hat{\theta}_i(1)}, \quad \theta_b = \frac{\hat{\theta}_i}{\hat{\theta}_i(2L+2)} \quad \hat{\theta}_i = \frac{\theta_a + \theta_b}{2}. \qquad (69)$$

6. NUMERICAL EXAMPLES

In this section, the performance of the proposed blind identification algorithm is evaluated and compared with those of other methods by means of numerical simulations.

Example 1. The simulations have been performed on the following two–channel FIR system of order $L = 5$, already considered in (Abed–Meraim et al., 1997a; Diversi et al., 2005)

$$H_1(z^{-1}) = -\,1.1836 + 0.4906\,z^{-1} - 0.3093\,z^{-2} + 0.4011\,z^{-3} + 0.1269\,z^{-4} - 1.8522\,z^{-5} \quad (70)$$
$$H_2(z^{-1}) = 1.2965 + 0.0525\,z^{-1} + 0.3410\,z^{-2} - 0.0260\,z^{-3} + 0.3991\,z^{-4} + 0.8817\,z^{-5}. \quad (71)$$

The input signal is an i.i.d. sequence with length $N = 200$ of binary variables, assuming the values $\{-1, +1\}$ with equal probability. The noiseless outputs $x_1(t)$ and $x_2(t)$ have been corrupted by additive Gaussian white noise sequences corresponding to different values of the Signal–to–Noise Ratio (SNR). The SNR on the i–th channel is defined, in dB, as

$$\text{SNR} = 10\log_{10}\frac{E[x_i^2]}{E[n_i^2]} = 10\log_{10}\frac{E[x_i^2]}{\sigma_i^*} \qquad i = 1, 2 \qquad (72)$$

where $E[\cdot]$ denotes the mathematical expectation. Since $x_1(t)$ and $x_2(t)$ have different variances, equal SNRs on every channel correspond to different values of σ_1^* and σ_2^*. For every SNR condition, a Monte Carlo simulation of $N_r = 100$ runs has been carried out. As performance index, the following Normalized Root Mean Square projection Error has been considered

$$\text{NRMSE} = \sqrt{\frac{1}{N_r}\sum_{i=1}^{N_r}\xi\big(\theta, \hat{\theta}^i\big)}\,, \qquad (73)$$

where $\hat{\theta}^i$ is the estimate of θ obtained at the i-th trial and

$$\xi\big(\theta, \hat{\theta}^i\big) = \sin^2(\phi) = 1 - \Big(\frac{\theta^T\,\hat{\theta}^i}{||\theta||\;||\hat{\theta}^i||}\Big)^2 \qquad (74)$$

measures the angle $\phi \in [0, \pi/2]$ between the two directions given by the vectors θ and $\hat{\theta}^i$ (Morgan et al., 1998). The algorithm is considered as convergent if it results $\sqrt{\xi\big(\theta, \hat{\theta}^i\big)} < 0.5$ (-6 dB). The proposed Algorithm 1 has been compared with the time domain Frisch Shifted Relation algorithm described in (Diversi et al., 2005), denoted with Frisch-SR. Figure 1 reports the NRMSE versus the SNR for the Algorithm 1 (solid line) and the Frisch-SR method (dashed line). Table 1 reports the empirical means of the estimates of the parameter θ (scaled to unit norm) and of the noise variances, together with the corresponding standard deviations, obtained with the Algorithm 1 and with the Frisch-SR algorithm, for SNR=10dB. Figure 1 and Table 1 show that the two identification methods yield good and comparable results.

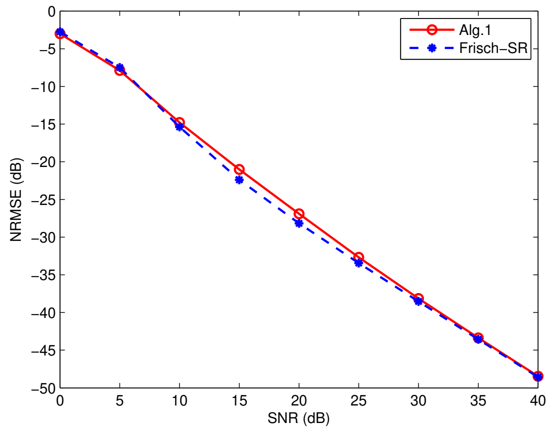

Fig. 1. NMRSE versus SNR: Alg.1 (solid), Frisch-SR (dashed).

Table 1. True and estimated parameters - $N = 200$, SNR$= 10$dB.

	True	Alg.1	Frisch − SR
$\theta(1)$	0.4560	0.4492 ± 0.0255	0.4496 ± 0.0306
$\theta(2)$	0.0185	0.0240 ± 0.0631	0.0173 ± 0.0475
$\theta(3)$	0.1199	0.1251 ± 0.0422	0.1221 ± 0.0500
$\theta(4)$	−0.0091	-0.0026 ± 0.0536	-0.0091 ± 0.0463
$\theta(5)$	0.1404	0.1427 ± 0.0444	0.1401 ± 0.0387
$\theta(6)$	0.3101	0.3060 ± 0.0420	0.3023 ± 0.0372
$\theta(7)$	0.4163	0.4067 ± 0.0374	0.4087 ± 0.0386
$\theta(8)$	−0.1726	-0.1604 ± 0.0810	-0.1676 ± 0.0630
$\theta(9)$	0.1088	0.1142 ± 0.0565	0.1110 ± 0.0618
$\theta(10)$	−0.1411	-0.1355 ± 0.0672	-0.1423 ± 0.0603
$\theta(11)$	−0.0446	-0.0410 ± 0.0559	-0.0424 ± 0.0579
$\theta(12)$	0.6515	0.6409 ± 0.0405	0.6425 ± 0.0492
σ_1^*	0.5022	0.4448 ± 0.1835	0.4530 ± 0.2553
σ_2^*	0.2224	0.1969 ± 0.0923	0.2057 ± 0.1239

Example 2. This numerical example illustrates the frequency domain features of the new identification method. The same two–channel FIR system of the Example 1 has been considered, driven by the same input sequence. However, in this example the output signals are affected by pink noises. Pink noise is characterized by a power spectrum that falls in frequency like $1/f$. The pink noise has been generated by using the following third–order ARMA model, suggested in (Orfanidis, 2010) at pag. 736

$$n_i(t) = g_0 \, \frac{B(z^{-1})}{A(z^{-1})} \, e_i(t) \qquad i = 1, 2 \tag{75}$$

where $e_i(t)$ is a white noise with variance σ_{e_i}, $g_0 = 0.57534$ and

$$\begin{aligned} B(z^{-1}) =& (1 - 0.98444\, z^{-1})(1 - 0.83392\, z^{-1}) \\ & \times (1 - 0.07568\, z^{-1}) \end{aligned} \tag{76}$$

$$\begin{aligned} A(z^{-1}) =& (1 - 0.99574\, z^{-1})(1 - 0.94791\, z^{-1}) \\ & \times (1 - 0.53568\, z^{-1}). \end{aligned} \tag{77}$$

The resulting power spectra of the noises $n_i(t)$ are

$$S_{n_i}(\omega_k) = g_0^2 \, \frac{|B(\mathrm{e}^{-\mathrm{j}\,\omega_k})|^2}{|A(\mathrm{e}^{-\mathrm{j}\,\omega_k})|^2} \, \sigma_{e_i} \qquad i = 1, 2 \tag{78}$$

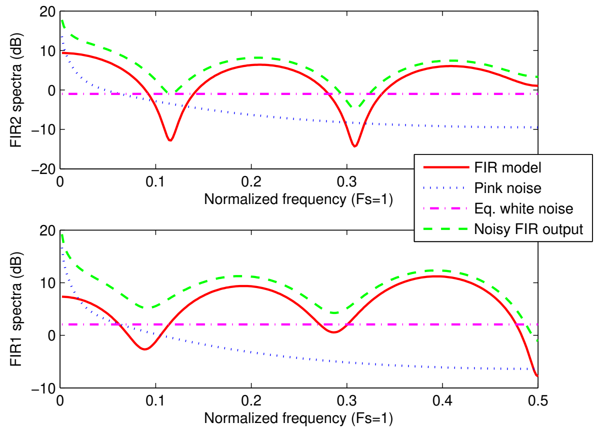

Fig. 2. FIR models (solid), pink noises (dotted), equivalent white noises (dash–dotted) and noisy signals (dashed)

By construction, asymptotically when $N \to \infty$, the variance σ_{e_i} coincides with the variance of the output noise $n_i(t)$, i.e. $\sigma_{e_i} = \sigma_i^*$. The data length in this example has been fixed to $N = 500$.

A Monte Carlo study of 100 independent runs has been performed by considering noisy output sequences, affected by additive pink noises, with variances $\sigma_1^* = 1.6086$, $\sigma_2^* = 0.7902$, corresponding to a ratio

$$10 \, \log_{10} \frac{E[x_i^2]}{E[n_i^2]} = 10 \, \log_{10} \frac{E[x_i^2]}{\sigma_i^*} \qquad i = 1, 2 \tag{79}$$

of about 5 dB on both channels.

Figure 2 shows, for the two FIR channels, the spectrum of the FIR system (solid line) and the spectrum of the additive pink noise (dotted line) together with the resulting noisy output spectrum (dashed line). The dash–dotted line reports the spectrum of the “equivalent” white noise with variance σ_i^*.

In many real situations some additional information about the system is available. In this case, for example, one could be aware that the additive noise is of pink type. Taking account of this information, the two–channel FIR system has been identified by using the Algorithm 1 within the frequency window $F = [f_i, \, f_f]$, with $f_i = 0.05$ and $f_f = 0.5$. In this way, the effect of the pink noise, acting at low frequencies, has been filtered out.

The results of the simulation are reported in the first column of Table 2. For comparison, the second column of Table 2 reports the estimates obtained by the Frisch-SR algorithm. When the whole frequency window $F = [0, \, 0.5]$ is used, also the Algorithm 1 yields bad results, similar to those of Frisch-SR. The corresponding reconstructions of the frequency responses (properly scaled) are shown in Figure 3. The advantageous effects of filtering are evident.

7. CONCLUSIONS

In this paper a novel frequency domain approach has been proposed for the blind identification of two–channel FIR systems affected by additive white noises. The estimation properties of the new algorithm have been tested and compared by means of Monte Carlo simulations. The numerical results have confirmed the good performances of the proposed method. The benefits of

Table 2. True and estimates parameters in the presence of output pink noise - $N = 500$.

	True	Alg.1 $[0.05 - 0.5]$	Frisch − SR
$\theta(1)$	0.4560	0.4448 ± 0.0143	0.3353 ± 0.0254
$\theta(2)$	0.0185	-0.0335 ± 0.0304	-0.2397 ± 0.0157
$\theta(3)$	0.1199	0.1461 ± 0.0271	0.0976 ± 0.0362
$\theta(4)$	−0.0091	-0.0246 ± 0.0293	-0.1941 ± 0.0191
$\theta(5)$	0.1404	0.1123 ± 0.0302	0.0027 ± 0.0264
$\theta(6)$	0.3101	0.2760 ± 0.0217	0.1249 ± 0.0303
$\theta(7)$	0.4163	0.4400 ± 0.0215	0.4054 ± 0.0220
$\theta(8)$	−0.1726	-0.2226 ± 0.0413	-0.4752 ± 0.0341
$\theta(9)$	0.1088	0.1852 ± 0.0350	0.2176 ± 0.0617
$\theta(10)$	−0.1411	-0.1276 ± 0.0387	-0.3125 ± 0.0387
$\theta(11)$	−0.0446	-0.0633 ± 0.0412	-0.1418 ± 0.0482
$\theta(12)$	0.6515	0.6176 ± 0.0412	0.4431 ± 0.0487
NRMSE		−15.6670 dB	−4.7524 dB

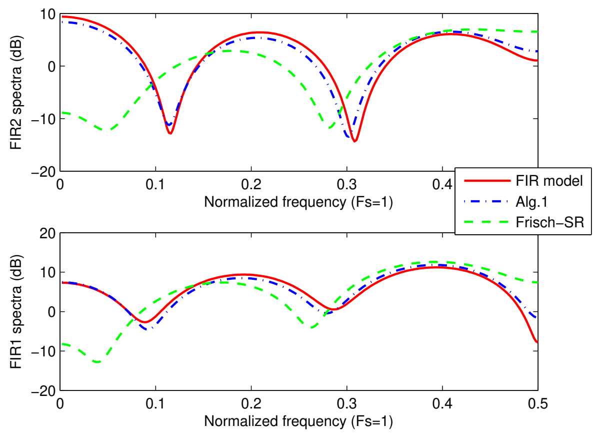

Fig. 3. FIRs spectra: true models (solid), estimated models with Alg.1 (dash-dotted) and with Frisch-SR (dashed).

filtering the data in the frequency domain have been illustrated by means of a numerical example.

REFERENCES

Abed–Meraim, K., Moulines, E. and Loubaton, P. (1997a). Prediction error method for second–order blind identification. *IEEE Transactions on Signal Processing*, 45, 694–705.

Abed–Meraim, K., Qiu, W. and Hua, Y. (1997b). Blind system identification. *Proceedings of the IEEE*, 85, 1310–1322.

Beghelli, S., Guidorzi, R. and Soverini, U. (1990). The Frisch scheme in dynamic system identification. *Automatica*, 26, 171–176.

Diversi, R., Guidorzi, R. and Soverini, U. (2005). Blind identification and equalization of two–channel FIR systems in unbalanced noise environments. *Signal Processing*, 85, 215–225.

Diversi, R., Guidorzi, R. and Soverini, U. (2007). Blind identification and equalization of multichannel FIR systems in unbalanced noise environments. *Signal Processing*, 87, 654–664.

Giannakis, G.B., Hua, Y., Stoica P. and Tong, L., Eds. (2001). *Signal Processing Advances in Wireless and Mobile Communications. Vol I: Trends in Channel Estimation and Equalization*. Englewood Cliffs, NJ, Prentice–Hall.

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

Hua, Y. (1996). Fast maximum likelihood for blind identification of multiple FIR channels. *IEEE Trans. on Signal Processing*, 44, 661–672.

Hua, Y. and Wax, M. (1996). Strict identifiability of multiple FIR channels driven by unknown arbitrary sequence. *IEEE Trans. on Signal Processing*, 44, 756–758.

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

McKelvey, T. (2002). Frequency domain identification methods. *Circuits, Systems and Signal Processing*, 21(1), 39–55.

Morgan, D.R., Benesty, J. and Sondhi, M.M. (1998). On the evaluation of estimated impulse responses. *IEEE Signal Processing Letters*, 5, 174–176.

Moulines, E., Duhamel, P. Cardoso, J.F. and Mayrargue, S. (1995). Subspace methods for the blind identification of multichannel FIR filters. *IEEE Transactions on Signal Processing*, 43, 516–525.

Orfanidis, S.J. (2010). *Introduction to Signal Processing*, Upper Saddle River, NJ, Prentice–Hall. Web page: www.ece.rutgers.edu/ orfanidi/i2sp

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. (2011). A generalized instrumental variable estimation method for errors–in–variables identification problems, *Automatica*, 47, 1656–1666.

Söderström, T. (2018). *Errors–in–Variables Methods in System Identification*, Springer, London.

Söderström, T., Hong, M. and Zheng, W.X. (2005). Convergence properties of bias–eliminating algorithms for errors–in–variables identification. *Inter. Journal of Adaptive Control and Signal Processing*, 19, 703–722.

Soverini, U. and Söderström, T. (2014a). Frequency domain maximum likelihood identification of noisy input–output models. *Proc. of the 19–th IFAC World Conference*, Cape Town, South Africa, 4625–4630.

Soverini, U. and Söderström, T. (2014b). Frequency domain EIV identification: a Frisch scheme approach. *Proc. of the 19–th IFAC World Conference*, Cape Town, South Africa, 4631–4636.

Soverini, U. and Söderström, T. (2015). Frequency domain EIV identification combining the Frisch scheme and Yule–Walker equations. *Proc. of the 14–th European Control Conference*, Linz, Austria, 2038–2043.

Soverini U. and Söderström, T. (2019). Frequency domain identification of FIR models from noisy input–output data. *Technical Report 2019-006.* Department of Information Technology, Uppsala University, Sweden.

Tong, L., Xu, G. and Kailath, T. (1991). A new approach to blind identification and equalization of multipath channels. *Proc. of the 25–th Asilomar Conf. Signals, Syst., Comput.*, 2, 856–860.

Tong, L. and Perreau, P. (1998). Multichannel blind identification: from subspace to maximum likelihood methods. *Proceedings of the IEEE*, 86, 1951–1968.

Xu, G., Liu, H., Tong, L. and Kailath, T. (1995). A least–squares approach to blind channel identification. *Signal Processing*, 62, 187–205.