Frequency domain identification of FIR models in the presence of additive input–output noise

Published Version:
Frequency domain identification of FIR models in the presence of additive input–output noise / Soverini, Umberto; Söderström, Torsten. - In: AUTOMATICA. - ISSN 0005-1098. - ELETTRONICO. - 115:5(2020), pp. 108879.1-108879.10. [10.1016/j.automatica.2020.108879]

Availability:
This version is available at: https://hdl.handle.net/11585/789539 since: 2021-01-19

Published:
DOI: http://doi.org/10.1016/j.automatica.2020.108879

Frequency domain identification of FIR models in the presence of additive input–output noise

Umberto Soverini[a], Torsten Söderström[b]

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

Abstract

This paper describes a new approach for identifying FIR models from a finite number of measurements, in the presence of additive and uncorrelated white noise. In particular, two different frequency domain algorithms are proposed. The first algorithm is based on some theoretical results concerning the dynamic Frisch scheme. The second algorithm maps the FIR identification problem into a quadratic eigenvalue problem. Both methods resemble in many aspects some other identification algorithms, originally developed in the time domain. The features of the proposed methods are compared with each other and with those of some time domain algorithms by means of Monte Carlo simulations.

Keywords: System identification; FIR models; Discrete Fourier Transform.

1. Introduction

The estimation of finite impulse response (FIR) models from noise–corrupted data is an important problem in many signal processing applications [11, 20]. Most of the solutions proposed in the literature assume that only the output measurements are affected by noise. However, as observed in [3], in many practical situations the presence of an additive noise on the input is important as well and must be taken into account.

It is a well known result that the least–squares (LS) method gives biased parameter estimates when the input is affected by noise. On the contrary, a consistent estimate can be obtained by using the total least–squares (TLS) approach [3]. In this case, however, the ratio between the input and output noise variances must be *a priori* known.

There are several methods in the literature for how to get consistent parameter estimates for noise–corrupted FIR models. One option is to use the instrumental variables (IV) approach. Such IV algorithms are computationally efficient, but suffer from poor estimation precision [24].

In order to remove the bias of the LS estimates, many bias–compensated techniques have been proposed with several, different strategies, see e.g [5, 6]. The compensated least–squares (CLS) methods are quite attractive, since the particular structure of the FIR models allows to develop iterative least–squares algorithms that are particularly suited for on–line implementations, see e.g. [7] and the reference therein.

In this work the identification of FIR systems corrupted by additive white noise is addressed by using a frequency domain approach. In particular, two different frequency domain algorithms are proposed and their features are compared with each other and with some time domain methods.

From a theoretical point of view, there is a full equivalence between time and frequency domain identification methods, as shown for example in [1]. From the practical point of view, the decision to implement a time or a frequency domain algorithm can strongly depend on the user choices and on the specific applications. In most experimental situations the observations are collected as samples of time signals, so that a Fourier transformation is required before implementing a frequency domain algorithm. However, there also exist occasions in which the data are more easily available as frequency samples since they are collected by a frequency analyzer which directly provides the Fourier transforms of the time signals [14, 18].

The approach to be described in this paper for frequency domain identification of noisy FIR models has some similarities with related errors–in–variables (EIV) problems [32, 33], where though the underlying structure is different. A common theme is that the estimation problem is formulated in the frequency domain in a way that takes all transient effects into account, as originally pointed out in [19]. As many other EIV approaches, also the proposed methods make use of the extended normal equations and require that some further equations are introduced, in order to get unique and consistent parameter estimates [24]. This being said about the similarities of the current paper and [32, 33], it is also crucial to note that the way the optimization criterions are introduced and analyzed is very highly tied to the specific model formulation/parameterization, and therefore there are also large differences in the contents.

The organization of the paper is as follows. Section 2 defines the FIR plus noise identification problem in the frequency domain, while Section 3 introduces a novel frequency domain description of the FIR processes. Section 4 discusses some contexts for the identification of EIV models. In particular, the section describes the GIVE framework, originally proposed in [23] and the dynamic Frisch scheme, originally proposed in [2]. Sections 5 proposes a possible identification criterion, that can be directly formulated in the frequency domain. In particular, this criterion takes advantage of a set of equations similar to the High Order Yule Walker (HOYW) equations. The method can be considered as the application to FIR models of the frequency domain approach proposed in [31]. In Section 6, it is shown that as an alternative approach, the FIR identification problem can be reformulated as a quadratic eigenvalue problem involving only the output noise variance. The obtained quadratic eigenvalue problem is solved by mapping it into a generalized eigenvalue problem. The method can be considered as the frequency counterpart of the time domain approach proposed in [5]. In Section 7 some general issues of the FIR identification problem are discussed. In particular, the section treats the problem of estimating the correct order of the model and analyzes some practical aspects concerning the filtering operations in the frequency domain. In Section 8 the performance of the two proposed methods is tested by means of Monte Carlo simulations. Finally some concluding remarks are reported in Section 9.

2. Problem statement

Consider the following FIR system

$$d(t) = H(z^{-1})\, x(t), \tag{1}$$

where $x(t)$ and $d(t)$ denote the input and output, and $H(z^{-1})$ is the following polynomial in the backward shift operator z^{-1}

$$H(z^{-1}) = h_0 + h_1 z^{-1} + \cdots + h_{M-1} z^{1-M}. \tag{2}$$

The observations of $x(t)$ and $d(t)$ are both affected by additive noise, so that the available signals are

$$u(t) = x(t) + n_i(t) \tag{3}$$
$$y(t) = d(t) + n_o(t). \tag{4}$$

The following assumptions are made.

A1. The length M of the FIR model is assumed as *a priori* known and $h_{M-1} \neq 0$.

A2. The true input $x(t)$ can be either a zero–mean ergodic process or a quasi–stationary bounded deterministic signal, i.e. such that the limit

$$\lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} x(t)\, x(t-\tau) \tag{5}$$

exists $\forall\tau$ [14]. Moreover, $x(t)$ is considered as persistently exciting of a sufficiently high order.

A3. The additive noises $n_i(t)$ and $n_o(t)$ are zero–mean ergodic white processes with *unknown* variances σ_i^* and σ_o^*.

A4. $n_i(t)$, $n_o(t)$ and $x(t)$ are mutually uncorrelated.

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. For $\{u(t)\}_{t=0}^{N-1}$, the corresponding Discrete Fourier Transform (DFT) is defined as

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

where $\omega_k = 2\pi k/N$ and $k = 0, \ldots, N-1$. Similarly, let $Y(\omega_k)$ be the DFT of $\{y(t)\}_{t=0}^{N-1}$. In the 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 a FIR system of type (1)–(4), under Assumptions A1–A4. Estimate the system parameters h_i $(i = 0, \ldots, M-1)$ and the noise variances σ_i^*, σ_o^*.

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

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

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 two algorithms proposed in the Sections 5 and 6 yield consistent estimates of the system parameters by using only the first $N_{\mathrm{half}} = \mathrm{ceil}\,((N+1)/2)$ samples $U(\omega_k)$, $Y(\omega_k)$, $k = 0, \ldots, \mathrm{floor}\left(\frac{N}{2}\right)$, see also Remark 5.

3. A frequency domain setup

In this section a new frequency domain description for the noisy FIR model (1)–(4) is introduced. This setup has been originally developed in [29, 30] with reference to the identification of errors–in–variables systems. In this respect, the maximum likelihood solution has been analyzed in depth in [27].

3.1. The noise–free case

Similarly to equation (6), let $X(\omega_k)$ and $D(\omega_k)$ be the DFTs of the signals $x(t)$ and $d(t)$ appearing in equation (1). It is a well–known fact [16, 19] that for finite N, even in absence of noise, the DFTs $X(\omega_k)$ and $D(\omega_k)$ exactly satisfy an extended model that includes also a transient term, i.e.

$$D(\omega_k) = H(\mathrm{e}^{-\mathrm{j}\,\omega_k})\, X(\omega_k) + T(\mathrm{e}^{-\mathrm{j}\,\omega_k}), \tag{8}$$

where $T(z^{-1})$ is a polynomial of order $M-2$

$$T(z^{-1}) = \tau_0 + \tau_1 z^{-1} + \cdots + \tau_{M-2} z^{-M+2} \tag{9}$$

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

By considering the whole number of frequencies, eq. (8) can be rewritten in a matrix form. For this purpose, introduce the parameter vectors

$$h = [\,h_0\ h_1 \dots h_{M-1}\,]^T \tag{10}$$

$$h_\tau = [\,\tau_0 \dots \tau_{M-2}\,]^T \tag{11}$$

and define the following vector

$$\Theta = [\,1\ \ -h^T\ \ -h_\tau^T\,]^T, \tag{12}$$

with dimension

$$p = 2M. \tag{13}$$

In absence of noise, the parameter vector (12) can be recovered by means of the following procedure. Define the row vectors

$$Z_M(\omega_k) = [1\ \ \mathrm{e}^{-\mathrm{j}\,\omega_k}\ \dots\ \mathrm{e}^{-\mathrm{j}\,(n-1)\omega_k}\ \ \mathrm{e}^{-\mathrm{j}\,(M-1)\omega_k}\,] \tag{14}$$

$$Z_{M-1}(\omega_k) = [1\ \ \mathrm{e}^{-\mathrm{j}\,\omega_k}\ \dots\ \mathrm{e}^{-\mathrm{j}\,(M-2)\omega_k}\,], \tag{15}$$

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

$$\Pi = \begin{bmatrix} Z_M(\omega_0) \\ \vdots \\ Z_M(\omega_{N-1}) \end{bmatrix} \qquad \Psi = \begin{bmatrix} Z_{M-1}(\omega_0) \\ \vdots \\ Z_{M-1}(\omega_{N-1}) \end{bmatrix} \tag{16}$$

of dimension $N \times M$ and $N \times (M-1)$, respectively.

Using the DFT samples $D(\omega_k)$ and $X(\omega_k)$ construct the following $N \times N$ diagonal matrix

$$V_X^{diag} = \mathrm{diag}\,[X(\omega_0),\ X(\omega_1), \dots, X(\omega_{N-1})] \tag{17}$$

and the N–dimensional column vector

$$V_D = [D(\omega_0),\ D(\omega_1), \dots, D(\omega_{N-1})]^T. \tag{18}$$

Then, compute the $N \times M$ matrix

$$\Phi_X = V_X^{diag}\,\Pi \tag{19}$$

and construct the $N \times p$ matrix

$$\hat{\Phi} = [V_D\,|\,\Phi_X\,|\,\Psi]. \tag{20}$$

Thus, eq. (8) for $k = 0, \dots, N-1$ can be rewritten as

$$\hat{\Phi}\,\Theta = 0. \tag{21}$$

It then holds

$$\hat{\Sigma}\,\Theta = 0, \tag{22}$$

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

$$\hat{\Sigma} = \frac{1}{N}(\hat{\Phi}^H\,\hat{\Phi}), \tag{23}$$

and $(\cdot)^H$ denotes the transpose and conjugate operation.

Remark 2. Since $d(t)$ is generated by the FIR model (1), the relation (8) cannot be satisfied by a polynomial $H(z^{-1})$ with order lower than $M-1$. Therefore, the matrix $\hat{\Sigma}$ in (23) is positive semidefinite, with only one null eigenvalue, i.e.

$$\hat{\Sigma} \geq 0 \qquad \dim\ker\hat{\Sigma} = 1. \tag{24}$$

Remark 3. If the signals $x(t)$ and $d(t)$ happen to be N–periodic, then the term $T(\mathrm{e}^{-\mathrm{j}\omega_h})$ in equation (8) is identically zero. In this case, the matrix in (20) can be reduced to the $N \times (M+1)$ matrix

$$\hat{\Phi}_{per} = [V_D\,|\,\Phi_X]. \tag{25}$$

It then holds

$$\hat{\Sigma}_{per}\,\theta = 0, \tag{26}$$

where $\hat{\Sigma}_{per}$ is the $(M+1)\times(M+1)$ positive semidefinite matrix

$$\hat{\Sigma}_{per} = \frac{1}{N}(\hat{\Phi}_{per}^H\,\hat{\Phi}_{per}) \tag{27}$$

and θ is the $M+1$ parameter vector

$$\theta = [\,1\ \ -h^T\,]^T. \tag{28}$$

In the following it will be shown how it is possible to reorganize the equations as they would have been generated by a periodic system, for every value of N. For details, see Section 4 in [35].

Partition the matrix $\hat{\Sigma}$, defined in (23), as follows

$$\hat{\Sigma} = \begin{bmatrix} \hat{\sigma}_{11} & \hat{\Sigma}_{12} & \hat{\Sigma}_{13} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} & \hat{\Sigma}_{23} \\ \hat{\Sigma}_{31} & \hat{\Sigma}_{32} & \hat{\Sigma}_{33} \end{bmatrix}, \tag{29}$$

where $\hat{\sigma}_{11}$ is a scalar, $\hat{\Sigma}_{22}$ is a square matrix of dimension M and $\hat{\Sigma}_{33}$ is a square matrix of dimension $M-1$. Relation (22) can be expanded as follows

$$\hat{\sigma}_{11} - \hat{\Sigma}_{12}\,h - \hat{\Sigma}_{13}\,h_\tau = 0 \tag{30}$$

$$\hat{\Sigma}_{21} - \hat{\Sigma}_{22}\,h - \hat{\Sigma}_{23}\,h_\tau = 0 \tag{31}$$

$$\hat{\Sigma}_{31} - \hat{\Sigma}_{32}\,h - \hat{\Sigma}_{33}\,h_\tau = 0. \tag{32}$$

From (32) we obtain

$$h_\tau = \hat{\Sigma}_{33}^{-1}\left(\hat{\Sigma}_{31} - \hat{\Sigma}_{32}\,h\right). \tag{33}$$

The expression (33) can then be substituted into (30) and (31). Defining the following matrices

$$\hat{\Sigma}_{red} = \begin{bmatrix} \hat{\sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{bmatrix}, \qquad \hat{T} = \begin{bmatrix} \hat{\Sigma}_{13}\hat{\Sigma}_{33}^{-1}\hat{\Sigma}_{31} & \hat{\Sigma}_{13}\hat{\Sigma}_{33}^{-1}\hat{\Sigma}_{32} \\ \hat{\Sigma}_{23}\hat{\Sigma}_{33}^{-1}\hat{\Sigma}_{31} & \hat{\Sigma}_{23}\hat{\Sigma}_{33}^{-1}\hat{\Sigma}_{32} \end{bmatrix}, \tag{34}$$

it is possible to reduce the non–periodic case into an equivalent periodic one, with lower dimensions (cf. (26))

$$\hat{\Sigma}_{per} = \hat{\Sigma}_{red} - \hat{T} \tag{35}$$

$$\hat{\Sigma}_{per}\,\theta = 0, \tag{36}$$

where θ has been defined in (28).

For noise–free data, the ratio

$$\hat{\rho} = \frac{\|\hat{T}\|_F}{\|\hat{\Sigma}_{red}\|_F} \tag{37}$$

can give a measure of the effect of the transient term, where $\|\cdot\|_F$ is the Frobenius norm of a matrix.

Remark 4. Given the input sequence $x(t)$, the ratio $\hat{\rho}$ takes into account both the parameters and the order of the FIR system. Note that $\hat{\rho}$ is a function of the data length N. In particular, $\hat{\rho} \to 0$ if $N \to \infty$, and

$$\hat{\rho}_{max} = \hat{\rho}\,(N_{min}), \tag{38}$$

where $N_{min} = 2M - 1$ is the minimum length of the input–output sequence, i.e. the minimum number of equations so that relation (24) holds.

3.2. The noisy case

In the presence of noise, the previous considerations can be modified as follows. With the noisy input–output DFT samples $U(\omega_k)$, $Y(\omega_k)$ construct the $N \times N$ diagonal matrix

$$V_U^{diag} = \text{diag}\,[U(\omega_0),\, U(\omega_1), \ldots, U(\omega_{N-1})] \tag{39}$$

and the N–dimensional column vector

$$V_Y = [Y(\omega_0),\, Y(\omega_1), \ldots, Y(\omega_{N-1})]^T. \tag{40}$$

Then, compute the matrix

$$\Phi_U = V_U^{diag}\,\Pi \tag{41}$$

and construct the $N \times p$ matrix

$$\Phi = [V_Y\,|\,\Phi_U\,|\,\Psi]. \tag{42}$$

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

$$\Sigma = \frac{1}{N}(\Phi^H\,\Phi) = \hat{\Sigma} + \tilde{\Sigma}^*, \tag{43}$$

where

$$\tilde{\Sigma}^* = \text{diag}\,[\,\sigma_o^*,\, \sigma_i^*\, I_M,\, 0_{M-1}]. \tag{44}$$

From (12), (22) and (43), the parameter vector Θ, defined in (12), can be obtained as the kernel of

$$(\Sigma - \tilde{\Sigma}^*)\,\Theta = 0, \tag{45}$$

after normalizing the first entry of Θ to 1.

Remark 5. The previous considerations hold also when only a subset of the whole frequency range is used, i.e. $\omega_k \in W = [\omega_i,\, \omega_f]$, with $i \geq 0$ and $f \leq \text{floor}\,(\frac{N}{2})$, containing $L = f - i + 1$ frequencies. The subset W must be chosen by the user on the basis of *a priori* knowledge of the frequency properties of the FIR system and of the noise–free input $X(\omega_k)$. The choice of $W = [\omega_i,\, \omega_f]$ reduces the number of the entries in (16) and (39)–(40), and all the related equations must be consequently modified. Some care must be used when the algorithms of Sections 5 and 6 make use of the whole DFT data set $U(\omega_k)$, $Y(\omega_k)$, $k = 0, \ldots, N-1$ [31]. In this case, two distinct sets of frequencies must be jointly considered, the set $W_1 = [\omega_i,\, \omega_f]$, with $i \geq 0$ and $f \leq \text{floor}\,(\frac{N}{2})$ and the set $W_2 = [\omega_{N-1-f},\, \omega_{N-1-i}]$. The total number of frequencies used in the algorithms will be $2L$. By considering a new matrix Φ with $2L$ rows, expressions (43)–(44) must be modified as follows

$$\Sigma = \frac{1}{2L}(\Phi^H\,\Phi) = \hat{\Sigma} + \tilde{\Sigma}^*\,, \tag{46}$$

where

$$\tilde{\Sigma}^* = \frac{N}{2L}\,\text{diag}\,[\,\sigma_o^*,\, \sigma_i^*\, I_M,\, 0_{M-1}]. \tag{47}$$

Remark 6. The noise Assumptions A3–A4 are necessary in order to obtain a diagonal matrix $\tilde{\Sigma}^*$, as defined in (44), when $N \to \infty$. On the other hand, one can observe that for large N the effect of the transient polynomial $T(z^{-1})$ is negligible since it vanishes at rate $O\,(1/\sqrt{N})$ [18] and the data could be treated as periodic, as done in Remark 3. This is a common procedure used in frequency domain identification, see e.g. [13, 22].

An alternative solution is to proceed as in Section 3.1. Partition the matrix Σ, defined in (43), according to the matrix $\hat{\Sigma}$ in equation (29). Expanding relation (45) as in (30)–(31), we obtain

$$\hat{\sigma}_{11} - \Sigma_{12}\,h - \Sigma_{13}\,h_\tau = 0 \tag{48}$$

$$\Sigma_{21} - \hat{\Sigma}_{22}\,h - \Sigma_{23}\,h_\tau = 0 \tag{49}$$

$$\Sigma_{31} - \Sigma_{32}\,h - \Sigma_{33}\,h_\tau = 0. \tag{50}$$

Next (50) implies

$$h_\tau = \Sigma_{33}^{-1}\left(\Sigma_{31} - \Sigma_{32}\,h\right). \tag{51}$$

Substitute now the expression (51) in (48)–(49), define the matrices

$$\Sigma_{red} = \begin{bmatrix} \sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \qquad T = \begin{bmatrix} \Sigma_{13}\Sigma_{33}^{-1}\Sigma_{31} & \Sigma_{13}\Sigma_{33}^{-1}\Sigma_{32} \\ \Sigma_{23}\Sigma_{33}^{-1}\Sigma_{31} & \Sigma_{23}\Sigma_{33}^{-1}\Sigma_{32} \end{bmatrix} \tag{52}$$

and set

$$R = \Sigma_{red} - T. \tag{53}$$

Defining the matrix

$$\tilde{R}^* = \text{diag}\,[\,\sigma_o^*,\, \sigma_i^*\, I_M], \tag{54}$$

it is then possible to derive the system of equations (cf. (36))

$$\left(R - \tilde{R}^*\right)\theta = 0. \tag{55}$$

Also in this case, the ratio

$$\rho = \frac{\|T\|_F}{\|\Sigma_{red}\|_F} \tag{56}$$

can give a measure of the effect of the transient term, starting from the available noisy data.

Remark 7. Simulation experiences have shown that, for every value of N, $\rho > \hat{\rho}$ holds, and the effect of the noise

$$\delta = \frac{\rho - \hat{\rho}}{\hat{\rho}} \tag{57}$$

decreases as the amount of noise decreases (see Figure 2). Since the ratio $\hat{\rho}$ is greater when the number of data N is shorter, the effect of the transient term is of major importance in the system parameter estimates when the data sequences are short and affected by a small amount of noise.

4. Overview of related EIV approaches

4.1. The GIVE framework

Note that (55) consists of $M+1$ algebraic non–linear equations. The number of unknowns is $M+2$, i.e. the elements of h in (10) and the two variances σ_o^* and σ_i^*. In the time domain, a similar set of equations has been largely studied in the identification of EIV dynamic systems. A general framework has been originally introduced in [23], where the Generalized Instrumental Variable Estimation (GIVE) method was proposed with reference to SISO EIV systems affected by additive white noises. The method has been generalized to the case of correlated noises in [25]. The GIVE method provides a unique general framework for the whole class of bias–compensating methods, including iterative solutions, like the BELS methods [26]. The GIVE frameworks 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 (55) 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 two methods 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 an optimization problem. An usual solution strategy consists in forcing some of the over–determined system of equations to hold exactly, while the others are minimized in a weighted least squares sense.

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 [26]. 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. In particular, it will be shown that the over–determined systems of equations exploited by the two proposed methods differ only for one equation, but the way they are solved is completely different.

4.2. The Frisch scheme context

The purpose of this subsection is to recall the Frisch scheme [2, 9, 24] for developing the estimation algorithm of Section 5. This can be thought 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 (53), the determination of the system parameter vector θ and of the noise variances σ_i^*, σ_o^* in eq. (55) can be seen as a Frisch scheme problem. This problem can be solved both in the time and in the frequency domain. Tn fact, the properties of the locus of the solutions in the noise plane $\mathcal{R}^2$ are the same. An in–depth description of these properties can be found in [2, 9, 24]. In the following only the main result of the Frisch scheme is recalled. All the technical aspects and proofs are reported in [34].

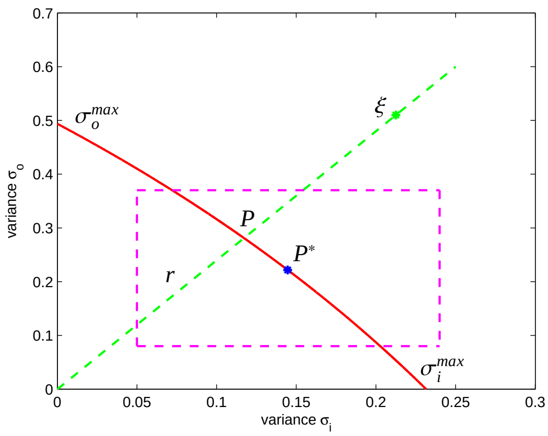

Figure 1: Typical shape of $\mathcal{S}(R)$

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

$$\tilde{R} = \text{diag}\,[\,\sigma_o,\ \sigma_i\, I_M] \tag{58}$$

such that

$$R - \tilde{R} \geq 0 \qquad \det(R - \tilde{R}) = 0. \tag{59}$$

Main Result. The set of all matrices $\tilde{R}$ satisfying the conditions (59) defines the points $P = (\sigma_i, \sigma_o)$ 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. Asymptotically, when $N \to \infty$, the point $P^* = (\sigma_i^*, \sigma_o^*)$ belongs to $\mathcal{S}(R)$.

As an example, Figure 1 reports the curve $\mathcal{S}(R)$ of the numerical Example 1 of Section 8.

Remark 8. In many practical situations it is possible to have some information about the lower and upper bounds of the noise covariances σ_i^* and σ_o^*, for example when other measurements taken in different experimental conditions are available. In these cases, the search of the point P^* on the curve $\mathcal{S}(R)$ can be bounded within a limited area in the noise plane $\mathcal{R}^2$, e.g. see the box depicted in Figure 1.

5. A criterion based on HOYW–type equations

As stated in the Main Result, the determination of the point P^* on $\mathcal{S}(R)$ leads to the solution of Problem 1, thanks to (55). 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 condition must be added to define a *unique* estimate.

In this section we will describe a possible search criterion. This criterion is analogue to that reported in [31] with reference to frequency domain identification of EIV systems.

Select the integer $q \geq 2M - 1$. The value of the parameter q is a user choice. In general, this value can affect the quality of the estimates. A good choice, from simulation experience, is $q = 2M - 1$. Analogously to (14), consider the row vector

$$Z_{q+M}(\omega_k) = [1 \; \mathrm{e}^{-\mathrm{j}\,\omega_k} \; \ldots \, \mathrm{e}^{-\mathrm{j}\,(M-1+q)\omega_k}] \tag{60}$$

and extract from it the q–dimensional row vector

$$Z_q^h(\omega_k) = [\mathrm{e}^{-\mathrm{j}\,M\omega_k} \; \ldots \, \mathrm{e}^{-\mathrm{j}\,(M-1+q)\omega_k}]. \tag{61}$$

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

$$\Pi^h = \left[Z_q^h(\omega_0) \; Z_q^h(\omega_1) \; \ldots \; Z_q^h(\omega_{N-1}) \right]^T. \tag{62}$$

and compute the $N \times q$ matrix

$$\hat{\Phi}^h = V_X^{diag} \, \Pi^h. \tag{63}$$

Define now the $q \times p$ matrix

$$\hat{\Sigma}^h = \frac{1}{N}((\hat{\Phi}^h)^H \, \hat{\Phi}). \tag{64}$$

Because of (21) we have

$$\hat{\Sigma}^h \, \Theta = 0. \tag{65}$$

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

$$\Phi^h = V_U^{diag} \, \Pi^h \tag{66}$$

and define the $q \times p$ matrix

$$\Sigma^h = \frac{1}{N}((\Phi^h)^H \, \Phi). \tag{67}$$

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

$$\Sigma^h = \hat{\Sigma}^h. \tag{68}$$

It is thus possible to write

$$\Sigma^h \, \Theta = 0. \tag{69}$$

It is not difficult to show that the dimensions of equation (69) can be reduced. For this purpose, partition matrix Σ^h as follows

$$\Sigma^h = \begin{bmatrix} \Sigma_1^h & \Sigma_2^h & \Sigma_3^h \end{bmatrix}, \tag{70}$$

where the matrices Σ_1^h, Σ_2^h, Σ_3^h have dimensions $q \times 1$, $q \times M$, $q \times (M-1)$, respectively. Thanks to (51), equation (69) can be reduced to

$$R^h \, \theta = 0, \tag{71}$$

where θ has been defined in (28) and

$$R^h = \begin{bmatrix} \Sigma_1^h - \Sigma_3^h \Sigma_{33}^{-1} \Sigma_{31} & \Sigma_2^h - \Sigma_3^h \Sigma_{33}^{-1} \Sigma_{32} \end{bmatrix}. \tag{72}$$

Remark 9. Equation (69) constitutes a set of q equations, analogue to the time domain high order Yule–Walker equations, that does not involve the noise variances σ_i^*, σ_o^*. These equations could be directly used to obtain an estimate of the parameter vector Θ if $q \geq 2M - 1$.

Remark 10. Equation (69) are analogue to an instrumental variable (IV) method in the time domain, where delayed outputs are used as instruments. These equations can be solved by using a total least squares approach. The main advantage of the method is the computational efficiency. On the other hand, the obtained estimation precision is often poor [24].

Remark 11. The set of $(M + 1)$ non–linear equations (55) can be joined to the set of q linear equations (71), with $q \geq 2M - 1$ and can be settled within the GIVE framework, as described in Subsection 4.1. Thus, the results in [23] and [24] 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 (55) are treated as a constraint that must be exactly satisfied, while the equations (71) must hold approximately. In other words, the search for P^* along $\mathcal{S}(R)$ can be performed by minimizing a quadratic cost function.

On the basis of the previous considerations, it is possible to develop an algorithm for the identification of the FIR plus noise models. A detailed description of the procedure can be found in [34]. In the following this algorithm is denoted as FIR1-FD.

6. A subspace approach

The approach proposed in this section is analogue to that described in [4] for the identification of autoregressive models affected by additive noise. This alternative method is not directly based on the Frisch scheme, but exploits the set of equations (69) together with the equations (45). It will be shown that the FIR plus noise identification problem can be mapped into a quadratic eigenvalue problem that, in turn, can be solved as a generalized eigenvalue problem. The principle has been applied to some other identification problems as well [5, 32].

Consider again equation (45), expanded as in (48)–(50). The last $2M - 1$ equations (49)–(50) can be written as

$$\begin{bmatrix} \Sigma_{21} & \Sigma_{22} - \sigma_i^* \, I_M & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{bmatrix} \Theta = 0. \tag{73}$$

Equation (73) contains $2M+1$ unknowns, i.e. σ_i^* and the entries of Θ. The equations (69) and (73) can be combined together in order to obtain the following nonlinear system of $2M - 1 + q$ equations

$$\begin{bmatrix} \Sigma_{21} & \Sigma_{22} - \sigma_i^* \, I_M & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \\ & \Sigma^h & \end{bmatrix} \Theta = 0. \tag{74}$$

It can be observed that the dimensions of the equation (74) can be reduced. For this purpose, consider the matrix R defined in (53) and partitioned as follows

$$R = \begin{bmatrix} r_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}, \tag{75}$$

where r_{11} is a scalar and R_{22} is a square matrix of dimension M.

Moreover, consider the matrix R^h defined in (72). Thanks to (51), equation (74) can be reduced to

$$\begin{bmatrix} R_{21} & R_{22} - \sigma_i^* \, I_M \\ & R^h \end{bmatrix} \theta = 0. \tag{76}$$

Remark 12. Observe that the set of equations (76) differ from the set of equations (55) and (71) by only one equation, in fact the first row in (55) is missing. Also the set of equations (76) can be settled within the GIVE framework of Subsection 4.1. As shown in the following, all the equations (76) are minimized in a total least squares sense, by solving a minimal eigenvalue problem. Of course, the asymptotic analysis proposed in [23] and [24] can be applied also in this case.

The equations (76) can be compactly rewritten as

$$(S - \sigma_i^* J)\,\theta = 0, \tag{77}$$

where

$$S = \begin{bmatrix} R_{21} & R_{22} \\ & R^h \end{bmatrix} \qquad J = \begin{bmatrix} 0_{M\times 1} & I_M \\ & 0_{q\times(M+1)} \end{bmatrix}. \tag{78}$$

Equations (78) represent a non–square generalized eigenvalue problem. In order to solve it, in the following we pursue the approach proposed in [4].

Multiplying both sides of (77) by $\left(S - \sigma_i^* J\right)^T$ leads to the equation

$$\left(A_2\,\sigma_i^{2*} + A_1\,\sigma_i^* + A_0\right)\,\theta = 0, \tag{79}$$

where

$$A_0 = S^T S \qquad A_1 = -\left(S^T J + J^T S\right) \qquad A_2 = J^T J. \tag{80}$$

The $M+1$ coefficients of θ can thus be estimated by solving the following quadratic eigenvalue problem (QEP)

$$\left(A_2\,\lambda^2 + A_1\,\lambda + A_0\right)\,v = 0. \tag{81}$$

Equations (81) can be solved by rewriting them as

$$A_2\,v'\,\lambda + A_1\,v\,\lambda + A_0\,v = 0, \tag{82}$$

where $v' = \lambda v$. Thus, the following $(2M + 2)$–dimensional generalized eigenvalue problem can be derived

$$(P - \lambda\,Q)\,\eta = 0, \tag{83}$$

where

$$P = \begin{bmatrix} A_0 & 0 \\ 0 & I_{M+1} \end{bmatrix} \qquad Q = \begin{bmatrix} -A_1 & -A_2 \\ I_{M+1} & 0 \end{bmatrix} \qquad \eta = \begin{bmatrix} v \\ v' \end{bmatrix}. \tag{84}$$

Following the discussion of [4], it can be stated that, asymptotically when the the number of data $N \to \infty$, the only real–valued eigenvalue solving (83) is σ_i^* and the first $M + 1$ entries of the corresponding eigenvector η^* are, after a normalization of the first entry to 1, the entries of θ, i.e.

$$\eta_0 = \frac{\eta^*}{\eta^*(1)} = [\,\theta^T \;\; \sigma_i^* \theta^T\,]^T. \tag{85}$$

Remark 13. Forming the matrix product $S^T S$ in (80) is the critical point of the proposed subspace approach, due to the potential loss of numerical precision. With a finite number of data, all the eigenvalues solving (83) will exhibit, in general, a small imaginary part. A criterion leading to good results consists in choosing the eigenvalue having the smallest modulus [4].

On the basis of the previous considerations, it is possible to develop a second algorithm for the FIR identification problem. A detailed description of the procedure can be found in [34]. In the following this algorithm is denoted as FIR2-FD.

7. General considerations

7.1. Estimation of the model order

A possible method for estimating the order $M - 1$ of the FIR model (1) can be developed on the basis of the following observation. In the time domain, the process

$$e(t) = y(t) - H(z^{-1})\,u(t) \tag{86}$$

describes the equation error of the FIR model (1)–(4). By substituting the relations (3) and (4) in (86), we obtain the alternative representation

$$e(t) = n_o(t) - H(z^{-1})\,n_i(t). \tag{87}$$

Because of Assumptions A3-A4, the equation (87) proves that the equation error $e(t)$ is an MA process of order $M - 1$, whose autocovariance function $r_e(k) = E\,[e(t)\,e(t-k)]$ is given by

$$r_e(0) = \sigma_o^* + \sigma_i^* \sum_{i=0}^{M-1} h_i^2 \tag{88}$$

$$r_e(k) = \sigma_i^* \sum_{i=0}^{M-k-1} h_i\,h_{i+k} \qquad \text{for } k = 1, \ldots, M-1 \tag{89}$$

$$r_e(k) = 0 \qquad \text{for } k > M - 1. \tag{90}$$

Thus, a possible way for estimating the order $M - 1$ of (1) consists in applying one of the proposed algorithms for an increasing sequence of orders n. Once the parameters $h_0, \ldots, h_n$ of $H(z^{-1})$ and the variances σ_o^*, σ_i^* have been estimated by means of the algorithm FIR1-FD or FIR2-FD, it is possible to compute the estimates of $r_e(k)$ by using (88)–(90). The model order $M-1$ can be estimated by observing that the cost function

$$J_n = \sum_{k=-n}^{n} r_e(k) \qquad n = 1, 2, \ldots \tag{91}$$

stabilizes at a constant value for $n \geq M - 1$. Note that J_n represents the value of the spectrum of the MA process $e(t)$ at zero frequency.

It is worth observing that the algorithm FIR1-FD is to be preferred in the computation of the quantities (88)–(90). In fact, in many cases the algorithm FIR2-FD yields worse estimates of σ_i^*, σ_o^*.

7.2. Practical aspects of frequency filtering

One of the main advantages of the proposed techniques lies in the fact that they allow to perform the identification by using only a reduced number of frequencies, as described in Remark 5. However, this property must be used with some caution, for two main reasons.

First of all, it must be observed that the condition $q \geq 2M - 1$ of Remark 9, together with Assumption A2, do not assure the consistency of the IV estimate (69). They assure only the "generic consistency", i.e. the consistency with probability one, see [28] pag. 266. In other words, there may exist cases where the rank of Σ^h in (69) is lower than $2M - 1$ and consequently the proposed algorithms do not produce consistent estimates. This consideration is particularly important when only a subset of the whole frequency range is used and Assumption A2 may fail.

Moreover, in case of FIR system identification, there is another important aspect that limits the usage of a reduced number of frequencies. The frequency shape of a FIR model is strictly linked with its order, e.g. see [10]. This property has also direct consequences in FIR system identification, as far as the choice of the order is concerned. In fact, from simulation experiences, it is possible to state that, in general, in order to obtain a good identification of the FIR system the model order must be correctly chosen and equal to its nominal value. A choice of a wrong (reduced) model order leads to biased estimates. The previous property is valid also when the system is identified by using a reduced number of frequencies, within a limited window of the spectrum. Also in this case, the order of the identified FIR model $H(z^{-1})$ must be chosen equal to its nominal value. In the presence of additive white noises on the input and output the previous consideration has a direct consequence. The signal–to–noise ratios of the input and output power spectra are

$$\frac{S_x(\omega_k)}{S_{n_i}(\omega_k)} = \frac{\lim_{N\to\infty} E\left[|X(\omega_k)|^2/N\right]}{\sigma_i^*} \tag{92}$$

$$\begin{aligned}\frac{S_d(\omega_k)}{S_{n_o}(\omega_k)} &= \frac{\lim_{N\to\infty} E\left[|D(\omega_k)|^2/N\right]}{\sigma_o^*}\\ &= |H(e^{-j\omega_k})|^2\,\frac{\lim_{N\to\infty} E\left[|X(\omega_k)|^2/N\right]}{\sigma_o^*}\end{aligned} \tag{93}$$

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

It can be observed that equation (92) is characterized by a constant denominator σ_i^* for all possible frequency windows. Also equation (93) is characterized by a constant denominator σ_o^* for all possible frequency windows since in this case $H(z^{-1})$ is a transfer function with only zeros, see eq. (2). Thus, restricting the identification within a limited window of the frequency spectrum $\omega_k \in [\omega_i, \omega_f]$ leads, in general, to worse results with respect to performing the same identification procedure by using the whole set of frequencies, for the simple reason that in the former case the available information of the system $H(e^{-j\omega_k})$ are only partially exploited. Of course, this consideration is no more valid when the input and output noises are characterized by a spectrum that is not constant for all the frequencies, i.e. when the denominators in the expressions (92)–(93) are not constant for all the frequency range. In this case the identification of the FIR model by using only a limited number of frequencies can lead to great advantages, if the chosen frequency window is characterized by lower values of the noise spectra.

8. Numerical examples

The effectiveness of the considered identification algorithms has been tested by means of numerical simulations on several FIR models taken from the literature. As a general consideration, it can be said that quite often the FIR1 algorithm gives better estimates than the FIR2 method, in particular as far as

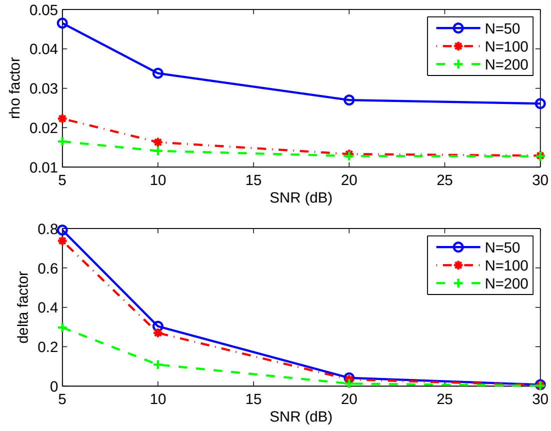

Figure 2: ρ and δ factors versus SNR.

the estimate of the noise variances is concerned. The performances have been always similar to those reported below, even for longer FIR models, like the 25 taps FIR filter proposed in [21]. In the following only two illustrative cases are presented, additional numerical examples can be found in [34].

Example 1. The proposed algorithms have been tested on sequences generated by the following FIR model, originally proposed in [3] and reconsidered several times in the literature, see e.g. [5, 6, 7] and the references therein. The FIR model has length $M = 5$ and is characterized by the coefficients

$$h = [-0.3 \quad -0.9 \quad 0.8 \quad -0.7 \quad 0.6]^T. \tag{94}$$

The input signal is the autoregressive process

$$x(t) = -0.2\,x(t-1) - 0.6\,x(t-2) + w(t), \tag{95}$$

where $w(t)$ is a Gaussian white noise with unit variance.

In order to illustrate the effect of the transient term with respect to the amount of noise, Figure 2 reports, for different data lengths N, the values of ρ, see eq. (56), and δ, see eq. (57) for different values of the signal–to–noise ratio (SNR).

To complete this analysis, a Monte Carlo simulation of 100 independent runs has been performed in the case of very short data, $N = 25$, with very small amount of noise, SNR = 60 dB. Table 1 reports the empirical means of the system parameter estimates obtained by means of the algorithm FIR1-FD, when the transient term is taken into account as described in Section 5, and when the transient term is not considered, as described in the Remarks 3 and 6 (algorithm FIR1-FD-NT). The number of the HOYW equations has been fixed to the minimal values $q = 2M - 1 = 9$ and $q = M = 5$, respectively.

A second Monte Carlo of 100 independent runs has been performed by considering noisy input and output sequences of length $N = 100$, affected by additive white noises corresponding to a SNR of 20 dB on both input and output sides.

Table 2 reports the empirical means of the system parameter estimates and of the noise variance estimates, together with

Table 1: True and estimated parameters with and without the transient term - $N = 25$, SNR = 60 dB.

	h_0	h_1	h_2	h_3	h_4
true	−0.3	−0.9	0.8	−0.7	0.6
FIR1 – FD	$-0.2998 \pm 0.6635 * 10^{-3}$	$-0.8999 \pm 0.5814 * 10^{-3}$	$0.8001 \pm 0.6782 * 10^{-3}$	$-0.7000 \pm 0.5469 * 10^{-3}$	$0.5999 \pm 0.7329 * 10^{-3}$
FIR1 – FD – NT	−0.4594 ± 0.0009	−1.1020 ± 0.0012	0.5219 ± 0.0016	−0.7924 ± 0.0011	0.4451 ± 0.0010

Table 2: True and estimated parameters obtained with FIR1-FD, FIR2-FD, FIR1-TD, FIR2-TD and TLS-TD - $N = 100$, SNR = 20 dB.

	true	FIR1 – FD	FIR2 – FD	FIR1 – TD	FIR2 – TD	TLS – TD
h_0	−0.3	−0.2983 ± 0.0243	−0.3032 ± 0.0254	−0.2990 ± 0.0261	−0.3024 ± 0.0271	−0.2990 ± 0.0263
h_1	−0.9	−0.9039 ± 0.0261	−0.9097 ± 0.0695	−0.9022 ± 0.0271	−0.9094 ± 0.0917	−0.9050 ± 0.0238
h_2	0.8	0.7970 ± 0.0306	0.7906 ± 0.0446	0.7970 ± 0.0316	0.7953 ± 0.0534	0.7985 ± 0.0321
h_3	−0.7	−0.7040 ± 0.0238	−0.7107 ± 0.0472	−0.7031 ± 0.0243	−0.7091 ± 0.0537	−0.7046 ± 0.0233
h_4	0.6	0.5975 ± 0.0253	0.5945 ± 0.0576	0.5969 ± 0.0254	0.5978 ± 0.0735	0.5993 ± 0.0260
σ_o^*	0.0207	0.0206 ± 0.0208	0.0155 ± 0.0975	0.0234 ± 0.0225	0.0138 ± 0.1310	0.0194 ± 0.0031
σ_i^*	0.0113	0.0093 ± 0.0088	0.0858 ± 0.0278	0.0089 ± 0.0090	0.0835 ± 0.0328	0.0106 ± 0.0017

the corresponding standard deviations, obtained with the algorithms proposed in Sections 5 and 6, denoted with FIR1-FD and FIR2-FD, respectively. The number of the HOYW equations has been fixed to $q = 2M - 1 = 9$ for both algorithms. The table shows also the results obtained with the corresponding time domain algorithms, denoted with FIR1-TD and FIR2-TD. The FIR2-TD algorithm was originally described in [5]. For FIR1-TD and FIR2-TD the number of the HOYW equations has been fixed to $q = M = 5$. For comparison, the last column of the table reports also the estimates obtained with the classical (time–domain) total least–squares method [8, 12]. Of course, to obtain the TLS-TD solution, the noise variance ratio σ_o^*/σ_i^* has been considered as *a priori* known. The TLS estimate should therefore be seen as a lower (and not accessible) limit of the performance, as TLS is heavily exploiting the information of known σ_o^*/σ_i^* [24]. It can be observed that the proposed identification methods yield similar, good results. Very similar results are also obtained with the corresponding time domain versions.

For a deeper comparison of the asymptotic performances of the algorithms, a Monte Carlo simulation of 100 independent runs has been carried out with $N = 500$, by considering different values of the SNR. In every run, the SNRs on the input and output sides are equal. The normalized root mean square error

$$\text{NRMSE} = \frac{1}{\|h\|}\sqrt{\frac{1}{100}\sum_{i=1}^{100}\|\hat{h}^i - h\|^2} \tag{96}$$

has been used as performance index, where $\hat{h}^i$ denotes the estimates of h obtained in the i–th trial of the Monte Carlo simulation. The results are shown in Figure 3.

Note that for high values of the SNR, the curves tend all to be straight lines with slope −1. This means that for high SNR

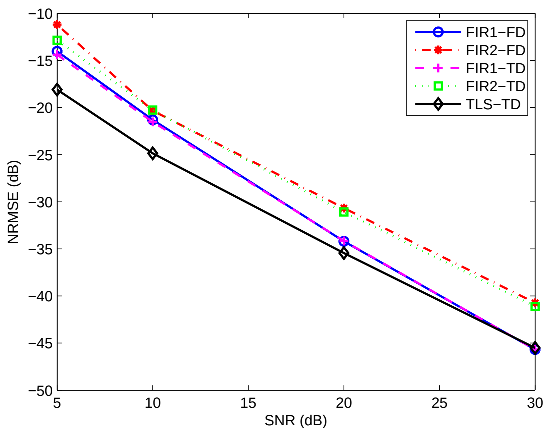

Figure 3: NMRSE versus SNR: FIR1-FD (circle, solid), FIR2-FD (dashed–dotted), FIR1-TD (dashed), FIR2-TD (dotted), TLS-TD (diamond, solid).

Table 3: True and estimated parameters in the presence of output pink noise

	true	FIR1 – FD [0.15 – 0.5]	FIR1 – FD [0 – 0.5]
h_0	1.0000	1.0061 ± 0.0106	1.0912 ± 0.1172
h_1	−0.3903	−0.3856 ± 0.0133	−0.4333 ± 0.0607
h_2	0.6240	0.6315 ± 0.0157	0.6840 ± 0.0779
h_3	−0.1912	−0.1860 ± 0.0130	−0.2132 ± 0.0345
h_4	0.2401	0.2427 ± 0.0095	0.2601 ± 0.0286
σ_o^*	0.7833	0.1096 ± 0.0038	0.6138 ± 0.0885
σ_i^*	0.0000	0.0018 ± 0.0022	0.0726 ± 0.1378

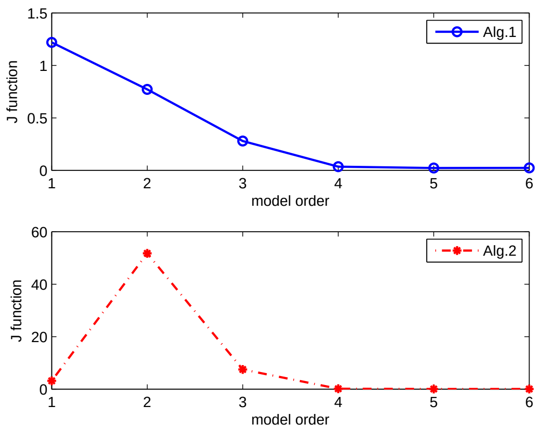

Figure 4: Function J_n versus the model order n. N=100, SNR=20 dB.

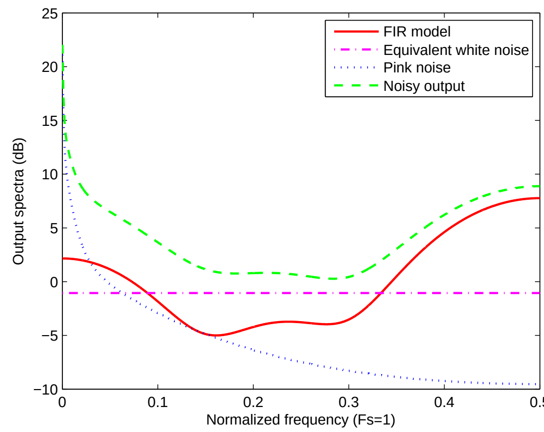

Figure 5: FIR model (solid), pink noise (dotted), equivalent white noise (dash–dotted) and noisy signal (dashed)

values, NRMSE is inversely proportional to SNR, and it holds

$$\text{NRMSE} \sim \frac{1}{\text{SNR}}. \tag{97}$$

It can be observed that the performances of the FIR1 algorithms are slightly better than those of the corresponding FIR2 algorithms.

Concluding, with reference to the case of $N = 100$ and SNR = 20 dB, the Figure 4 reports the results of the model order estimate, obtained with the cost function J_n in (91), evaluated by means of the two proposed algorithms. It can be observed that the Algorithm 1 yields a nicer, monotonic, behavior of J_n.

The next numerical example illustrates the frequency domain features of the new identification methods, described in the Subsection 7.2.

Example 2. The following FIR model, with length $M = 5$, has been considered

$$h = [1.0000 \quad -0.3903 \quad 0.6240 \quad -0.1912 \quad 0.2401]^T. \tag{98}$$

The input signal $x(t)$ is a Gaussian white noise with variance $\sigma_x = 1$. In this example the input measurement noise is not present, i.e. $n_i(t) = 0$, and the output signal is affected by a pink noise. Pink noise is frequently used in music signal processing and is characterized by a power spectrum that falls in frequency like $1/f$. The pink noise has been generated by using the third–order ARMA model, suggested in [17] at pag. 736

$$n_o(t) = g_0 \, \frac{B(z^{-1})}{A(z^{-1})} \, e(t), \tag{99}$$

where $e(t)$ is a white noise with variance σ_e, $g_0 = 0.57534$ and

$$B(z^{-1}) = (1 - 0.98444\, z^{-1})(1 - 0.83392\, z^{-1})(1 - 0.07568\, z^{-1})$$
$$A(z^{-1}) = (1 - 0.99574\, z^{-1})(1 - 0.94791\, z^{-1})(1 - 0.53568\, z^{-1}).$$

The resulting power spectra of the signals $d(t)$ and $n_o(t)$ are

$$S_d(\omega_k) = |H(\mathrm{e}^{-\mathrm{j}\,\omega_k})|^2\, \sigma_x = |H(\mathrm{e}^{-\mathrm{j}\,\omega_k})|^2 \tag{100}$$

$$S_{n_o}(\omega_k) = g_0^2\, \frac{|B(\mathrm{e}^{-\mathrm{j}\omega_k})|^2}{|A(\mathrm{e}^{-\mathrm{j}\omega_k})|^2}\, \sigma_e. \tag{101}$$

By construction, asymptotically, when $N \to \infty$, the variance σ_e coincides with the variance of the output noise $n_o(t)$, i.e. $\sigma_e = \sigma_o^*$. For this reason, the data length in this example has been fixed to $N = 5000$.

A Monte Carlo of 100 independent runs has been performed by considering noisy output sequences, affected by additive pink noise, with variance $\sigma_o^* = 0.7833$, corresponding to a ratio

$$10 \log_{10} \frac{E[d(t)^2]}{E[n_o(t)^2]} = 10 \log_{10} \frac{\int_{-\pi}^{\pi} S_d(\omega_k)\, \frac{d\omega}{2\pi}}{\int_{-\pi}^{\pi} S_{n_o}(\omega_k)\, \frac{d\omega}{2\pi}}$$
$$= 10 \log_{10} \frac{\int_{-\pi}^{\pi} |H(\mathrm{e}^{-\mathrm{j}\,\omega_k})|^2 \, \frac{d\omega}{2\pi}}{\sigma_o^*} \approx 3\,\text{dB}.$$

Figure 5 shows 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_o^*.

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 FIR model has been identified by using the FIR1-FD algorithm within the frequency window $F = [f_i,\ f_f]$, with $f_i = 0.15$ 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 3. For comparison, the second column of Table 3 reports the estimates obtained by the same FIR1-FD algorithm when the whole frequency window $F = [0, 0.5]$ is used. The advantageous effects of filtering are evident. In fact the method yields accurate estimates of the system parameters. It can be observed that the results using all frequencies are definitely worse.

9. Conclusions

In this paper two novel frequency domain approaches have been proposed for the identification of FIR models affected by additive white noises. Their estimation properties have been tested and compared by means of Monte Carlo simulations. The numerical results have confirmed the good performances of the methods. The benefits of filtering the data in the frequency domain have been illustrated by means of a numerical example.

References

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

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

[3] C. E. Davila, An efficient recursive total least squares algorithm for FIR adaptive filtering. *IEEE Trans. on Signal Processing*, 42 (1994), 268–280.

[4] C. E. Davila, A subspace approach to estimation of autoregressive parameters from noisy measurements. *IEEE Trans. on Signal Processing*, 46 (1998) 531–534.

[5] R. Diversi, Noisy FIR identification as a Quadratic Eigenvalue Problem. *IEEE Trans. on Signal Processing*, 57 (2009), 4563–4568.

[6] R. Diversi, R. Guidorzi and U. Soverini, A new approach for identifying noisy input–output FIR models. *Proc. of the 3rd IEEE ISCCSP*, St. Julians, Malta, 2008, 1548–1552.

[7] D.-Z. Feng and W.X. Zheng, Bilinear equation method for unbiased identification of linear FIR systems in the presence of input and output noises. *Signal Processing*, 87 (2007), 1147–1155.

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

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

[10] F. Harris, Fixed length FIR filters with continuously variable bandwidth *Proc. of the 1st International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic System Technology* Aalborg, Denmark, 2009, 931–935.

[11] S. Haykin, *Adaptive Filter Theory*, Prentice–Hall, New Jersey, 1996.

[12] M.J. Levin, Estimation of a system pulse transfer function in the presence of noise. *IEEE Trans. on Automatic Control*, 9 (1964) 229–235.

[13] L. Ljung, Some results on identifying linear systems using frequency domain data. *Proc. of the 32nd IEEE Conference on Decision and Control*, San Antonio, Texas, 1993, 567–569.

[14] L. Ljung, *System Identification–Theory for the User* (2nd ed.), Prentice–Hall, Upper Saddle River, New Jersey, 1999.

[15] C.D. Meyer, *Matrix Analysis and Applied Linear Algebra* SIAM, Society for Industrial and Applied Mathematics, Philadelphia, 2000.

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

[17] S.J. Orfanidis, *Introduction to Signal Processing*, Prentice–Hall, Upper Saddle River, New Jersey, 2010. www.ece.rutgers.edu/ orfanidi/i2sp

[18] R. Pintelon and J. Schoukens, *System Identification: a Frequency Domain Approach* (2nd ed.), IEEE Press, New York, 2012.

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

[20] J.G. Proakis, *Digital Communications*, McGraw–Hill, New York, 1995.

[21] L.R. Rabiner, R.E. Crochiere and J.B. Allen, FIR system modeling and identification in the presence of noise and with band-limited inputs. *IEEE Trans. on Acoustic, Speech and Signal Processing*, 26 (1978), 319–333.

[22] R.S. Smith, Frequency domain subspace identification using nuclear norm minimization and Hankel matrix realizations. *IEEE Trans. on Automatic Control*, 59 (2014) 2886–2896.

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

[24] T. Söderström, *Errors–in–Variables Methods in System Identification*, Springer, London, 2018.

[25] T. Söderström, R. Diversi and U. Soverini, A unified framework for EIV identification methods when the measurement noise are mutually correlated, *Automatica*, 50 (2014), 3216–3223.

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

[27] T. Söderström and U. Soverini, Errors–in–variables identification using maximum likelihood estimation in the frequency domain. *Automatica*, 79 (2017) 131–143.

[28] T. Söderström and P. Stoica, *System Identification*, Prentice–Hall, Cambridge, MA, 1989.

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

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

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

[32] U. Soverini and T. Söderström, Frequency domain identification of autoregressive models in the presence of additive noise. *WSEAS Trans. on Signal Processing*, 12 (2016) 208–217.

[33] U. Soverini and T. Söderström, Frequency domain identification of ARX models in the presence of additive input–output noise. *Proc. of the 20th IFAC World Conference*, Toulouse, France, 2017, 6400–6405.

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

[35] U. Soverini and T. Söderström, The Frisch scheme: time and frequency domain aspects. *Technical Report 2019-010.* Department of Information Technology, Uppsala University, Sweden, 2019.