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

The Frisch scheme for EIV system identification: time and frequency domain formulations

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: Several estimation methods have been proposed for identifying errors–in–variables systems, where both input and output measurements are corrupted by noise. One of the more interesting approaches is the Frisch scheme. The method can be applied using either time or frequency domain representations. This paper investigates the general mathematical and geometrical aspects of the Frisch scheme, illustrating the analogies and the differences between the time and frequency domain formulations.

Keywords: System identification; EIV models; Frisch scheme; Discrete Fourier Transform.

1. INTRODUCTION

Representations where errors or measurement noises are present on both inputs and outputs are usually called errors–in–variables (EIV) models. This class of models plays an important role in many engineering applications like, for instance, array signal processing, fault detection, blind channel equalization, image processing, etc. (Van Huffel and Lemmerling, 2002).

The identification of EIV models has been deeply investigated in the literature. Many time and frequency domain methods have been developed for its solution. An exhaustive presentation of this subject can be found in the recent book (Söderström, 2018), where many different approaches are deeply analyzed and compared with each other. Among these methods, one can find the bias–eliminated least squares (BELS), the Frisch scheme, the generalized instrumental variables estimates (GIVE), the covariance matching and the maximum likelihood.

In particular, the Frisch scheme is one of the more interesting approaches for the EIV identification. It has its roots in (Frisch, 1934), where the problem was treated in the static case. Subsequently, the problem has been proposed for identifying dynamic systems in (Beghelli et al., 1990) and the further elaboration of robust selection criteria has allowed its application to real data. A complete treatment of this subject can be found in (Guidorzi et al., 2008; Söderström, 2018).

The aim of this paper is not to propose a new solution, but rather it takes the EIV identification problem as a motivation/background for studying some mathematical and geometrical aspects that arise when the Frisch scheme is used as solution method. In particular, by making use of the Discrete Fourier Transform (DFT) properties, the paper gives a thorough analysis of the analogies and differences between the time and frequency domain formulations. With reference to the frequency domain, some results reported in the paper concern the general aspects of the finite–dimensional discrete–time dynamic systems. In particular, it is shown how the behavior of noise–free input–output signals of finite length can be equivalently represented by periodic sequences.

The organization of the paper is as follows. Section 2 defines the EIV identification problem in the time and in the frequency domains. In Section 3 the EIV identification problem is developed in the time domain and the related Frisch scheme problem is presented. In Section 4 the EIV identification problem is presented with a frequency domain formulation. Firstly, an in–depth analysis of the noise–free case is performed. Then, the Frisch Scheme problem is formulated in the frequency domain. Section 5 exploits the geometric properties of the Frisch scheme in order to characterize the solution sets defined in the previous Sections 3 and 4. The related main theorems are recalled. The proofs can be found in (Soverini and Söderström, 2019b), one of the proofs has not appeared before. In Section 6 the Frisch method is discussed within the GIVE method, originally proposed in (Söderström, 2011), which represents a more general framework for the identification of EIV models. Finally, some concluding remarks are reported in Section 7.

2. STATEMENT OF THE PROBLEM

Consider the linear time–invariant SISO system described by the linear difference equation

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

where $\hat{u}(t)$, $\hat{y}(t)$ are the noise–free input and output and $A(z^{-1})$, $B(z^{-1})$ are polynomials in the backward shift operator z^{-1}

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

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

In the errors–in–variables (EIV) environment the input and output measurements are both affected by additive noise, so that the available signals are

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

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

The following assumptions are made.

A1. System (1) is asymptotically stable.
A2. $A(z^{-1})$ and $B(z^{-1})$ do not share any common factor.
A3. The order n of the system is assumed as *a priori* known.
A4. The noise–free input $\hat{u}(t)$ can be either a zero–mean ergodic process or a quasi–stationary bounded deterministic signal, i.e. such that the limit

$$\lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} \hat{u}(t)\, \hat{u}(t-\tau) \tag{6}$$

exists $\forall\tau$ (Ljung, 1999). Moreover, $\hat{u}(t)$ is considered as persistently exciting of a sufficiently high order.
A5. The additive noises $\tilde{u}(t)$ and $\tilde{y}(t)$ are zero–mean ergodic white processes with unknown variances λ_u^* and λ_y^*, respectively.
A6. $\tilde{u}(t)$, $\tilde{y}(t)$ and $\hat{u}(t)$ are mutually uncorrelated.

Remark 1. The results proposed in the following can be easily extended to the case when the polynomials $A(z^{-1})$ and $B(z^{-1})$ have different degrees n_a and n_b. Also the presence of a delay in the polynomial $B(z^{-1})$ can be easily taken into account. However, for simplicity of exposition in this paper the polynomials $A(z^{-1})$ and $B(z^{-1})$ have the structure (2)–(3).

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

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

where $\omega_k = 2\pi k/N$ and $k = 0, \ldots, N-1$. Similarly, let $Y(\omega_k)$ be the DFT of $\{y(t)\}_{t=0}^{N-1}$. The classical EIV identification problem can be stated as follows.

Identification problem – time domain. Let $u(t)$, $y(t)$ be a set of noisy measurements generated by an EIV system of type (1)–(5), under Assumptions A1–A6. Estimate the system parameters α_i $(i = 1, \ldots, n)$, β_i $(i = 0, \ldots, n)$ and the noise variances λ_u^*, λ_y^*.

Identification problem – frequency domain. Let $U(\omega_k)$, $Y(\omega_k)$ be a set of noisy DFT measurements generated by an EIV system of type (1)–(5), under Assumptions A1–A6. Estimate the system parameters α_i $(i = 1, \ldots, n)$, β_i $(i = 0, \ldots, n)$ and the noise variances λ_u^*, λ_y^*.

3. TIME DOMAIN SETUP

In this section a time domain description for the EIV model (1)–(5) is introduced. This setup has been originally developed in (Beghelli et al., 1990) and subsequently proposed in several papers, see e.g. (Diversi et al., 2006; Guidorzi et al., 2008; Söderström, 2018).

3.1 The noise–free case

Introduce the parameter vectors

$$\theta_\alpha = [1\, \alpha_1 \ldots \alpha_n]^T \tag{8}$$

$$\theta_\beta = [\beta_0\, \beta_1 \ldots \beta_n]^T \tag{9}$$

and define the following vector

$$\theta = [\ \theta_\alpha^T\ \ -\theta_\beta^T\ ]^T, \tag{10}$$

with dimension

$$p_\theta = 2n + 2. \tag{11}$$

By defining the vectors

$$\hat{\varphi}(t) = [\, \hat{y}(t)\, \ldots\, \hat{y}(t-n)\ \hat{u}(t)\, \ldots\, \hat{u}(t-n)\,]^T \tag{12}$$

$$\varphi(t) = [\, y(t)\, \ldots\, y(t-n)\ u(t)\, \ldots\, u(t-n)\,]^T \tag{13}$$

$$\tilde{\varphi}(t) = [\, \tilde{y}(t)\, \ldots\, \tilde{y}(t-n)\ \tilde{u}(t)\, \ldots\, \tilde{u}(t-n)\,]^T, \tag{14}$$

the EIV model (1)–(5) can be rewritten in the following form

$$\hat{\varphi}^T(t)\, \theta = 0 \tag{15}$$

$$\varphi(t) = \hat{\varphi}(t) + \tilde{\varphi}(t). \tag{16}$$

In the noise–free case, from equation (15) a matrix form can be derived, by computing the following $p_\theta \times p_\theta$ sample covariance matrix

$$\hat{\Sigma}_t = \frac{1}{N-n} \sum_{t=n}^{t=N-1} \hat{\varphi}(t)\, \hat{\varphi}^T(t), \tag{17}$$

where the subscript t has been introduced to denote that the matrix $\hat{\Sigma}_t$ has been computed with the time domain data $\hat{u}(t)$, $\hat{y}(t)$, and to distinguish it from the matrix $\hat{\Sigma}_f$ that will be defined in Section 4. It then holds

$$\hat{\Sigma}_t\, \theta = 0. \tag{18}$$

It is worth recalling that the matrix $\hat{\Sigma}_t$ can also be rewritten in the following way, resembling the procedure of Section 5 for the frequency domain data. Define the Hankel matrices

$$\hat{X}_y = \begin{bmatrix} \hat{y}(n) & \ldots & \hat{y}(0) \\ \hat{y}(n+1) & \ldots & \hat{y}(1) \\ \vdots & & \vdots \\ \hat{y}(N-1) & \ldots & \hat{y}(N-n-1) \end{bmatrix} \tag{19}$$

$$\hat{X}_u = \begin{bmatrix} \hat{u}(n) & \ldots & \hat{u}(0) \\ \hat{u}(n+1) & \ldots & \hat{u}(1) \\ \vdots & & \vdots \\ \hat{u}(N-1) & \ldots & \hat{u}(N-n-1) \end{bmatrix} \tag{20}$$

and construct the matrix of the input–output samples

$$\hat{X} = [\hat{X}_y \,|\, \hat{X}_u]. \tag{21}$$

The equation (1) for $t = n, \ldots, N-1$ can be rewritten as

$$\hat{X}\, \theta = 0. \tag{22}$$

Relation (18) is then obtained, by computing the sample covariance matrix as follows

$$\hat{\Sigma}_t = \frac{1}{(N-n)} (\hat{X}^T\, \hat{X}). \tag{23}$$

Remark 2. Because of Assumptions A2 and A4, relation (15), or relation (1), cannot be satisfied by polynomials $A(z^{-1})$ and $B(z^{-1})$ with degree lower than n. Therefore, when the number of equations is such that $N - n \geq 2n - 1$, i.e $N \geq 3n - 1$, the matrix $\hat{\Sigma}_t$ in (17), or in (23), is positive semidefinite, with only one zero eigenvalue, i.e.

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

3.2 The noisy case

When $N \to \infty$, we can define the covariance matrix

$$\hat{\Sigma}_t = E\, [\hat{\varphi}(t)\, \hat{\varphi}^T(t)], \tag{25}$$

where $E[\cdot]$ denotes the mathematical expectation.

In the presence of noise, because of Assumptions A5–A6, we obtain the following $p_\theta \times p_\theta$ positive definite noisy covariance matrix

$$\Sigma_t = E\, [\varphi(t)\, \varphi^T(t)] = \hat{\Sigma}_t + \tilde{\Sigma}_t^*, \tag{26}$$

where

$$\tilde{\Sigma}_t^* = \begin{bmatrix} \lambda_y^* I_{n+1} & 0 \\ 0 & \lambda_u^* I_{n+1} \end{bmatrix}. \qquad (27)$$

Starting from an assumed knowledge of the noisy matrix Σ_t, it is then possible to write the following system of $2n+2$ algebraic non–linear equations

$$\left(\Sigma_t - \tilde{\Sigma}_t^*\right) \theta = 0, \qquad (28)$$

with $2n+3$ unknowns, i.e. the $2n+1$ free coefficients of θ and the two variances λ_u^* and λ_y^*.

The Frisch scheme (Beghelli et al., 1990; Guidorzi et al., 2008; Söderström, 2018). The relation (28) is the basis of the Frisch scheme method. As stated in Remark 2, the matrix $\hat{\Sigma}_t = \Sigma_t - \tilde{\Sigma}_t^*$ is singular (positive semidefinite) with one eigenvalue equal to zero and θ is the corresponding eigenvector. For a given Σ_t and assuming that $\tilde{\Sigma}_t^*$ has the simple structure of type (27), the problem is to find appropriate estimates of the noise variances and then determine the parameter vector θ.

For this purpose, one can proceed characterizing the solution set defined below. This study can be seen as a first step towards the solution of the EIV identification problem in the time domain. It turns out that the equations (28) alone are not enough to determine a unique solution.

Frisch scheme solution set – time domain. For a given Σ_t, determine the set of non–negative definite diagonal matrices of type

$$\tilde{\Sigma}_t = \mathrm{diag}\left[\, \lambda_y I_{n+1},\, \lambda_u I_{n+1}\right] \qquad (29)$$

such that

$$\Sigma_t - \tilde{\Sigma}_t \geq 0 \qquad \det\left(\Sigma_t - \tilde{\Sigma}_t\right) = 0. \qquad (30)$$

The theorems characterizing the solution set will be presented in Section 5.

4. FREQUENCY DOMAIN SETUP

In this section a frequency domain description for the EIV model (1)–(5) is introduced. This setup has been originally developed in (Soverini and Söderström, 2014, 2015).

4.1 The noise–free case

The transfer function of (1) is represented as

$$G(\mathrm{e}^{-j\omega}) = \frac{B(\mathrm{e}^{-j\omega})}{A(\mathrm{e}^{-j\omega})}. \qquad (31)$$

Similarly to equation (7), let $\hat{U}(\omega_k)$ and $\hat{Y}(\omega_k)$ be the DFTs of the signals $\hat{u}(t)$ and $\hat{y}(t)$ appearing in equation (1). It is a well–known fact (Pintelon et al., 1997; McKelvey, 2002) that for finite N, even in absence of noise, the ratio of the DFTs $\hat{Y}(\omega_k)$ and $\hat{U}(\omega_k)$ ($\omega_k = 2\pi k/N$) is not equal to the true transfer function

$$G(\mathrm{e}^{-j\omega_k}) \neq \frac{\hat{Y}(\omega_k)}{\hat{U}(\omega_k)}. \qquad (32)$$

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

$$A(\mathrm{e}^{-j\omega_k})\, \hat{Y}(\omega_k) = B(\mathrm{e}^{-j\omega_k})\, \hat{U}(\omega_k) + T(\mathrm{e}^{-j\omega_k}), \qquad (33)$$

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

$$T(z^{-1}) = \tau_0 + \tau_1 z^{-1} + \cdots + \tau_{n-1} z^{-n+1} \qquad (34)$$

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

By considering the whole number of frequencies, relation (33) can be rewritten in a matrix form. For this purpose, introduce the vector containing the transient coefficients

$$\theta_\tau = [\tau_0 \ldots \tau_{n-1}]^T \qquad (35)$$

and define the following vector

$$\Theta = [\, \theta_\alpha^T \;\; -\theta_\beta^T \;\; -\theta_\tau^T \,]^T, \qquad (36)$$

with dimension

$$p_\Theta = 3n + 2. \qquad (37)$$

In absence of noise, the parameter vector (36) can be recovered by means of the following procedure. Define the row vectors

$$Z_{n+1}(\omega_k) = [1 \;\; \mathrm{e}^{-j\omega_k} \;\; \ldots \;\; \mathrm{e}^{-j(n-1)\omega_k} \;\; \mathrm{e}^{-jn\omega_k}\,] \qquad (38)$$

$$Z_n(\omega_k) = [1 \;\; \mathrm{e}^{-j\omega_k} \;\; \ldots \;\; \mathrm{e}^{-j(n-1)\omega_k}\,], \qquad (39)$$

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

$$\Pi = \begin{bmatrix} Z_{n+1}(\omega_0) \\ \vdots \\ Z_{n+1}(\omega_{N-1}) \end{bmatrix} \qquad \Psi = \begin{bmatrix} Z_n(\omega_0) \\ \vdots \\ Z_n(\omega_{N-1}) \end{bmatrix}. \qquad (40)$$

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

Remark 3. It can be observed that the entries of the matrices in (40) are of type

$$f_{ik} = \mathrm{e}^{-j\frac{2\pi}{N}(i-1)(k-1)}, \quad i = 1, \ldots, N, \;\; k = 1, \ldots, n. \quad (41)$$

They also constitute the entries of the $N \times N$ Fourier matrix

$$F_N = \frac{1}{\sqrt{N}}\,[f_{ik}], \quad i, k = 1, \ldots, N. \qquad (42)$$

In particular, when $N = n$, it results in $F_N = \frac{1}{\sqrt{N}}\,\Psi$.

From the DFT samples $\hat{U}(\omega_k)$ and $\hat{Y}(\omega_k)$ construct the following $N \times N$ diagonal matrices

$$\hat{V}_U^{diag} = \mathrm{diag}\,[\hat{U}(\omega_0),\, \hat{U}(\omega_1), \ldots, \hat{U}(\omega_{N-1})] \qquad (43)$$

$$\hat{V}_Y^{diag} = \mathrm{diag}\,[\hat{Y}(\omega_0),\, \hat{Y}(\omega_1), \ldots, \hat{Y}(\omega_{N-1})]. \qquad (44)$$

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

$$\hat{\Phi}_Y = \hat{V}_Y^{diag}\, \Pi \quad \hat{\Phi}_U = \hat{V}_U^{diag}\, \Pi \qquad (45)$$

and construct the $N \times p_\Theta$ matrix

$$\hat{\Phi} = [\hat{\Phi}_Y \mid \hat{\Phi}_U \mid \Psi]. \qquad (46)$$

Thus, eq. (33) for $k = 0, \ldots, N-1$ can be rewritten as

$$\hat{\Phi}\, \Theta = 0. \qquad (47)$$

It then holds

$$\hat{\Sigma}\, \Theta = 0, \qquad (48)$$

where $\hat{\Sigma}$ is the $p_\Theta \times p_\Theta$ matrix

$$\hat{\Sigma} = \frac{1}{N}(\hat{\Phi}^H \hat{\Phi}) \qquad (49)$$

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

Remark 4. Because of Assumptions A2 and A4, relation (33) cannot be satisfied by polynomials $A(z^{-1})$ and $B(z^{-1})$ with degree lower than n. Therefore, when the number of equations is such that $N \geq 3n-1$, the matrix $\hat{\Sigma}$ in (49) is positive semidefinite, with only one zero eigenvalue, i.e.

$$\hat{\Sigma} \geq 0 \qquad \dim\ker \hat{\Sigma} = 1. \qquad (50)$$

Remark 5. It is worth observing the analogies between the time domain relations (22), (18) and the frequency domain relations

(47), (48). Also Remark 2 is analogue to Remark 4. Finally, observe that both (24) and (50) hold if the data length satisfies the same condition $N \geq 3n-1$, where $3n-1$ is the number of the free coefficients of the vector Θ in (36). In other words, also in the time domain, the data length must take into account the transient effect, due to the initial conditions, even if they do not explicitly appear as unknowns in the system of equations (22). Of course, one could expand the model (15) involving n additional parameters describing the initial conditions and then consider the equations (22) for $t = 0, \ldots, N-1$. However, the net effect is null, since n additional equations are used, and n more unknowns are introduced.

Remark 6. If the signals $\hat{u}(t)$ and $\hat{y}(t)$ happen to be N–periodic, then the term $T(\mathrm{e}^{-\mathrm{j}\omega_h})$ in equation (33) is identically zero (Pintelon and Schoukens, 2012). In this case, the matrix in (46) can be reduced to the $N \times (2n+2)$ matrix

$$\hat{\Phi}_{per} = [\hat{\Phi}_Y \mid \hat{\Phi}_U]. \tag{51}$$

It then holds

$$\hat{\Phi}_{per}\, \theta = 0 \tag{52}$$

and

$$\hat{\Sigma}_{per}\, \theta = 0, \tag{53}$$

where $\hat{\Sigma}_{per}$ is the $(2n+2) \times (2n+2)$ positive semidefinite matrix

$$\hat{\Sigma}_{per} = \frac{1}{N}(\hat{\Phi}_{per}^H\, \hat{\Phi}_{per}) \tag{54}$$

and θ is the $2n+2$ parameter vector defined in (10).

In the following it will be shown how it is possible to reorganize the equations in order to eliminate θ_τ. Two ways are possible. One can start from the equation (47) that involves the "data" matrix $\hat{\Phi}$. Alternatively, it is possible to consider the equation (48) and work on the "covariance" matrix $\hat{\Sigma}$.

Method a Relation (47) can be expanded as follows

$$\hat{\Phi}_Y\, \theta_\alpha - \hat{\Phi}_U\, \theta_\beta - \Psi\, \theta_\tau = 0. \tag{55}$$

From (55) we obtain

$$\begin{aligned}\theta_\tau &= (\Psi^H\, \Psi)^{-1}\Psi^H\, \hat{\Phi}_Y\, \theta_\alpha - (\Psi^H\, \Psi)^{-1}\Psi^H\, \hat{\Phi}_U\, \theta_\beta \\ &= (\Psi^H\, \Psi)^{-1}\Psi^H\Big(\hat{\Phi}_Y\, \theta_\alpha - \hat{\Phi}_U\, \theta_\beta\Big).\end{aligned} \tag{56}$$

The expression (56) can then be substituted into (55). Defining the matrix

$$\bar{\Psi} = \Psi(\Psi^H\, \Psi)^{-1}\Psi^H, \tag{57}$$

it is possible to construct the new "data" matrix

$$\bar{\Phi} = [\,(I-\bar{\Psi})\,\hat{\Phi}_Y \mid (I-\bar{\Psi})\,\hat{\Phi}_U\,], \tag{58}$$

such that

$$\bar{\Phi}\,\theta = 0. \tag{59}$$

Remark 7. From (56), one can observe that only two ways are possible, in order to have $\theta_\tau = 0$, which correspond to the following conditions

$$\hat{\Phi}_Y\, \theta_\alpha - \hat{\Phi}_U\, \theta_\beta = 0 \tag{60}$$

$$(\Psi^H\, \Psi)^{-1}\Psi^H = 0. \tag{61}$$

The condition (60) means that the sequences $\hat{U}(\omega_k)$, $\hat{Y}(\omega_k)$ are N–periodic, with a finite value N. In fact, (60) coincides with the relation (52). The condition (61) holds only when $N \to \infty$. In fact, it is possible to show that

$$(\Psi^H\, \Psi)^{-1}\Psi^H \to 0 \;\text{ when }\; N \to \infty, \text{ with velocity } \frac{1}{N}. \tag{62}$$

The proof of (62) can be found in (Soverini and Söderström, 2019b).

A similar result can be proved also for the matrix $\bar{\Psi}$, defined in (57) and appearing in (58). In fact, it can be seen that its entries $\bar{\psi}_{ij}$ $(i,j = 1,\ldots,N)$ are bounded by $|\bar{\psi}_{ij}| \leq n/N$. Thus, the following property holds

$$\bar{\Psi} \to 0 \;\text{ when }\; N \to \infty, \text{ with velocity } \frac{n}{N}. \tag{63}$$

From (58), thanks to (63), it readily follows that the sequences $\hat{U}(\omega_k)$, $\hat{Y}(\omega_k)$ become asymptotically periodic, when $N \to \infty$.

It is worth observing that relations (58)–(59) have been derived from the relations (46)–(47). Looking at (58), one can wonder if it is possible to go further, asking if and how it is possible to map a structure of type (46)–(47), related to non periodic data, into a reduced structure of type (51)–(52), where equivalent periodic data are linked by the same system parameters θ_α, θ_β. For this purpose, the following result can be proved. A constructive proof is reported in (Soverini and Söderström, 2019b).

Existence of equivalent periodic signals. Given two non periodic DFT sequences of length N

$$\hat{V}_U = [\hat{U}(\omega_0)\ldots\hat{U}(\omega_{N-1})]^T \tag{64}$$

$$\hat{V}_Y = [\hat{Y}(\omega_0)\ldots\hat{Y}(\omega_{N-1})]^T \tag{65}$$

such that relations (46)–(47) hold, it is "generically" possible to find (up to a scalar factor) two equivalent periodic sequences with length N

$$\hat{V}_C = [\hat{C}(\omega_0)\ldots\hat{C}(\omega_{N-1})]^T \tag{66}$$

$$\hat{V}_D = [\hat{D}(\omega_0)\ldots\hat{D}(\omega_{N-1})]^T \tag{67}$$

such that

$$[\hat{\Phi}_C \mid \hat{\Phi}_D]\,\theta = 0, \tag{68}$$

where

$$\hat{\Phi}_C = \hat{V}_C^{diag}\,\Pi \quad \hat{\Phi}_D = \hat{V}_D^{diag}\,\Pi \tag{69}$$

and $\hat{V}_C^{diag}$, $\hat{V}_D^{diag}$ are defined according to (43)–(44).

Method b Partition the matrix $\hat{\Sigma}$, defined in (49), as follows

$$\hat{\Sigma} = \begin{bmatrix}\hat{\Sigma}_{11} & \hat{\Sigma}_{12} & \hat{\Sigma}_{13}\\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} & \hat{\Sigma}_{23}\\ \hat{\Sigma}_{31} & \hat{\Sigma}_{32} & \hat{\Sigma}_{33}\end{bmatrix}, \tag{70}$$

where $\hat{\Sigma}_{11}$ and $\hat{\Sigma}_{22}$ are square matrices of dimension $n+1$ and $\hat{\Sigma}_{33}$ is a square matrix of dimension n. Relation (48) can be expanded as follows

$$\hat{\Sigma}_{11}\,\theta_\alpha - \hat{\Sigma}_{12}\,\theta_\beta - \hat{\Sigma}_{13}\,\theta_\tau = 0 \tag{71}$$

$$\hat{\Sigma}_{21}\,\theta_\alpha - \hat{\Sigma}_{22}\,\theta_\beta - \hat{\Sigma}_{23}\,\theta_\tau = 0 \tag{72}$$

$$\hat{\Sigma}_{31}\,\theta_\alpha - \hat{\Sigma}_{32}\,\theta_\beta - \hat{\Sigma}_{33}\,\theta_\tau = 0. \tag{73}$$

From (73) we obtain

$$\theta_\tau = \hat{\Sigma}_{33}^{-1}\Big(\hat{\Sigma}_{31}\,\theta_\alpha - \hat{\Sigma}_{32}\,\theta_\beta\Big). \tag{74}$$

The expression (74) can then be substituted into (71) and (72). Defining the following matrices

$$\hat{\Sigma}_{red} = \begin{bmatrix}\hat{\Sigma}_{11} & \hat{\Sigma}_{12}\\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22}\end{bmatrix} \quad \hat{T} = \begin{bmatrix}\hat{\Sigma}_{13}\hat{\Sigma}_{33}^{-1}\hat{\Sigma}_{31} & \hat{\Sigma}_{13}\hat{\Sigma}_{33}^{-1}\hat{\Sigma}_{32}\\ \hat{\Sigma}_{23}\hat{\Sigma}_{33}^{-1}\hat{\Sigma}_{31} & \hat{\Sigma}_{23}\hat{\Sigma}_{33}^{-1}\hat{\Sigma}_{32}\end{bmatrix} \tag{75}$$

it is possible to define the following problem of reduced dimension

$$\hat{\Sigma}_f = \hat{\Sigma}_{red} - \hat{T} \tag{76}$$

$$\hat{\Sigma}_f\,\theta = 0. \tag{77}$$

The subscript f denotes that the matrix $\hat{\Sigma}_f$ has been computed with the DFTs $\hat{U}(\omega_k)$, $\hat{Y}(\omega_k)$, by using the frequency domain representation (33).

Remark 8. The following results have been verified numerically.

1. The matrix $\hat{\Sigma}_f$ in (76) coincides with the positive semidefinite matrix
$$\hat{\Sigma}_f = \frac{1}{N}(\bar{\Phi}^H \bar{\Phi}), \tag{78}$$
where $\bar{\Phi}$ is defined in (58).
2. The matrix $\hat{\Sigma}_f$ in (76) or (78) coincides with the matrix $\hat{\Sigma}_t$ in (23) if $\hat{\Sigma}_t$ is obtained by dividing with N, instead of $(N-n)$.

For noise–free data, the ratio
$$\hat{\rho} = \frac{||\hat{T}||_F}{||\hat{\Sigma}_{red}||_F} \tag{79}$$
may be taken as a measure of the effect of the transient term, where $||\cdot||_F$ is the Frobenius norm of a matrix.

Remark 9. Given the input–output DFTs $\hat{U}(\omega_k)$, $\hat{Y}(\omega_k)$, the ratio $\hat{\rho}$ depends on both the parameters and the order of the system. Note that $\hat{\rho}$ is a function of the data length N. In particular, $\hat{\rho} \to 0$ if $N \to \infty$, and
$$\hat{\rho}_{max} = \hat{\rho}\,(N_{min}), \tag{80}$$
where $N_{min} = 3n - 1$ is the minimum length of the input–output sequence, i.e. the minimum number of equations so that relation (50) holds.

4.2 The noisy case

In the presence of noise, the previous procedure can be modified as follows. With the noisy input–output DFT samples $U(\omega_k)$ and $Y(\omega_k)$ construct the $N \times N$ diagonal matrices
$$V_U^{diag} = \text{diag}\,[U(\omega_0),\, U(\omega_1), \ldots, U(\omega_{N-1})] \tag{81}$$
$$V_Y^{diag} = \text{diag}\,[Y(\omega_0),\, Y(\omega_1), \ldots, Y(\omega_{N-1})]. \tag{82}$$
Then, compute the matrices
$$\Phi_Y = V_Y^{diag}\,\Pi \quad \Phi_U = V_U^{diag}\,\Pi \tag{83}$$
and construct the $N \times p_\Theta$ matrix
$$\Phi = [\Phi_Y \,|\, \Phi_U \,|\, \Psi]\,. \tag{84}$$
Because of Assumptions A5–A6, when $N \to \infty$, we obtain the following $p_\Theta \times p_\Theta$ positive definite matrix
$$\Sigma = \lim_{N\to\infty} \frac{1}{N}(\Phi^H \Phi) = \hat{\Sigma} + \tilde{\Sigma}^*\,, \tag{85}$$
where
$$\tilde{\Sigma}^* = \text{diag}\left[\,\lambda_y^*\, I_{n+1},\, \lambda_u^*\, I_{n+1},\, 0\, I_n\right]. \tag{86}$$
From (48) and (85), the parameter vector Θ, defined in (36), can be obtained as the kernel of
$$\left(\Sigma - \tilde{\Sigma}^*\right)\Theta = 0\,, \tag{87}$$
where the first entry is normalized to 1.

Remark 10. The noise Assumptions A5–A6 are necessary in order to obtain a diagonal matrix $\tilde{\Sigma}^*$, as defined in (86), 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 with order $O\,(1/\sqrt{N})$ (Pintelon and Schoukens, 2012) and the data could be treated as periodic, as done in Remark 6. This is a common procedure used in many frequency domain identification approaches, see e.g. (Ljung, 1993; Smith, 2014).

At this point, one can proceed as in Section 4.1, in order to eliminate θ_τ. Partition the matrix Σ, defined in (85), according to the matrix $\hat{\Sigma}$ in equation (70). Expanding relation (87) as in (71)–(73), we obtain
$$\hat{\Sigma}_{11}\,\theta_\alpha - \Sigma_{12}\,\theta_\beta - \Sigma_{13}\,\theta_\tau = 0 \tag{88}$$
$$\Sigma_{21}\,\theta_\alpha - \hat{\Sigma}_{22}\,\theta_\beta - \Sigma_{23}\,\theta_\tau = 0 \tag{89}$$
$$\Sigma_{31}\,\theta_\alpha - \Sigma_{32}\,\theta_\beta - \Sigma_{33}\,\theta_\tau = 0. \tag{90}$$
Next (90) implies
$$\theta_\tau = \Sigma_{33}^{-1}\left(\Sigma_{31}\theta_\alpha - \Sigma_{32}\,\theta_\beta\right). \tag{91}$$
Substitute now the expression (91) in (88)–(89), define the matrices
$$\Sigma_{red} = \begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix},\quad T = \begin{bmatrix}\Sigma_{13}\Sigma_{33}^{-1}\Sigma_{31} & \Sigma_{13}\Sigma_{33}^{-1}\Sigma_{32}\\ \Sigma_{23}\Sigma_{33}^{-1}\Sigma_{31} & \Sigma_{23}\Sigma_{33}^{-1}\Sigma_{32}\end{bmatrix} \tag{92}$$
and set
$$R = \Sigma_{red} - T. \tag{93}$$
Defining the matrix
$$\tilde{R}^* = \text{diag}\left[\,\lambda_y^*\, I_{n+1},\, \lambda_u^*\, I_{n+1}\right] \tag{94}$$
it is then possible to derive the system of equations (cf. (77))
$$\left(R - \tilde{R}^*\right)\theta = 0. \tag{95}$$
Equation (95) constitutes a system of $2n+2$ algebraic non–linear equations with $2n+3$ unknowns, i.e. the $2n+1$ free coefficients of θ and the two variances λ_u^* and λ_y^*.

Remark 11. With reference to relation (95) we can repeat considerations similar to those given after equation (28) and we can define an analogue Frisch scheme problem in the frequency domain. Similarly to the time domain, we can characterize the properties of the solution set defined below, and this study can be seen as a first step towards the solution of the EIV identification problem in the frequency domain. It is worth observing the analogy of the time domain relations (29), (30) with the following relations (96), (97).

Frisch scheme solution set – frequency domain. Assigned R, determine the set of non–negative definite diagonal matrices of type
$$\tilde{R} = \text{diag}\left[\,\lambda_y\, I_{n+1},\, \lambda_u\, I_{n+1}\right] \tag{96}$$
such that
$$R - \tilde{R} \geq 0 \qquad \det\left(R - \tilde{R}\right) = 0. \tag{97}$$
The theorems characterizing the solution set will be presented in Section 5.

Remark 12. As a final consideration of this section, we can state that to take into account the transient effect, by including the vector θ_τ in (36), has a twofold meaning, theoretical and practical. From the theoretical point of view, this is the only way that allows a one to one correspondence between the time and the frequency domain approaches, as proved in the paper. From the practical point of view, if the vector θ_τ is not considered, it means that the data are treated as periodic even when they are not. When the data length N is large, this effect can be considered as negligible for many practical applications, see Remark 10. However, it is never null. With reference to the identification of FIR models, this aspect is well illustrated in (Soverini and Söderström, 2019a) by the numerical example 2.

5. THE FRISCH SCHEME CONTEXT

The purpose of this subsection is to exploit some geometric properties of the Frisch scheme (Beghelli et al., 1990; Guidorzi et al., 2008; Söderström, 2018) in order to characterize the solution sets defined in Sections 3.2 and 4.2.

It must be observed that, interchanging the notation Σ_t with R, the formulation of the Frisch scheme in the time domain, involving the equations (27)–(30), is formally equal to the formulation in the frequency domain, involving the equations (94)–(97). In fact, both Σ_t and R are real–valued, positive definite matrices, with dimension $p_\theta \times p_\theta$. Moreover, the matrix $\tilde{\Sigma}_t^*$ in (27) coincides with $\tilde{R}^*$ in (94), as well the structure of the matrices $\tilde{\Sigma}_t$ in (29) coincides with that of the matrices $\tilde{R}$ in (96).

On the basis of the previous observation, it can be concluded that the Frisch scheme can be studied in the time and in the frequency domain. The properties of the locus of the solutions in the noise plane $\mathcal{R}^2$ will be the same. In the following we formulate the theorems by making reference to the frequency domain problem, equal theorems hold for the time domain.

Remark 13. The theorems concern some geometric properties of the equations (96)–(97). It is worth recalling that these equations arise from the study of the EIV dynamic system defined by the relations (1)–(5) under the Assumptions A1–A6. This setup is known as the "dynamic" Frisch scheme (Beghelli et al., 1990), since it can be considered as a subcase of the classical scheme proposed by the Nobel prize laureate Ragnar Frisch in 1934 (Frisch, 1934), with reference to the algebraic processes. The Frisch scheme is an interesting compromise between the generality of the EIV environment and the possibility of performing real applications. It must be observed that the Frisch scheme encompasses some other important methods, like the Least Squares method and the Eigenvector method. In the algebraic case, the Frisch scheme does not lead to the determination of a single optimal solution. Rather, it leads to the determination of a whole family of solutions which are compatible with a given set of noisy observations. It will be shown that this fact is true also for the dynamic case, if we limit its study to the set of equations (96)–(97). Indeed, in the dynamic case, the shift–invariant property of the process allows to add further equations and this fact makes it is possible to determine a unique solution. This aspect will be only briefly discussed in Section 6, since it goes beyond the aim of the paper.

Partition the positive definite matrix R as follows

$$R = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}, \tag{98}$$

where R_{11} and R_{22} are square matrices of dimension $n + 1$. Then, the following theorems hold. The proofs can be found in (Soverini and Söderström, 2019b). In particular, the proof of Theorem 3 has not appeared anywhere before.

Theorem 1. The maximal admissible value of the input noise variance λ_u^{max} compatible with the conditions (97) is obtained when $\lambda_y = 0$ and is given by

$$\lambda_u^{max} = \min \operatorname{eig}\left(R_{22} - R_{21}\, R_{11}^{-1}\, R_{12}\right). \tag{99}$$

Similarly, the maximal admissible value of the output noise variance λ_y^{max} compatible with the conditions (97) is obtained when $\lambda_u = 0$ and is given by

$$\lambda_y^{max} = \min \operatorname{eig}\left(R_{11} - R_{12}\, R_{22}^{-1}\, R_{21}\right). \tag{100}$$

Theorem 2. The set of all matrices $\tilde{R}$ satisfying the conditions (97) defines the points $P = (\lambda_u, \lambda_y)$ of a continuous curve $\mathcal{S}(R)$ belonging to the first quadrant of the noise space $\mathcal{R}^2$. This curve defines λ_y uniquely from λ_u, and vice versa.

Theorem 3. The curve $\mathcal{S}(R)$ in *Theorem 2* describes a convex set in the first quadrant of $\mathcal{R}^2$, whose concavity faces the origin.

Corollary 1. Every point $P = (\lambda_u, \lambda_y)$ of $\mathcal{S}(R)$ can be associated with a noise matrix of type $\tilde{R}(P)$ (96) and with a coefficient vector $\theta(P)$, satisfying the relation

$$\left(R - \tilde{R}(P)\right)\, \theta(P) = 0. \tag{101}$$

Corollary 2. When $N \to \infty$, the point $P^* = (\lambda_u^*, \lambda_y^*)$ belongs to $\mathcal{S}(R)$ and the corresponding coefficient vector $\theta(P^*)$ is characterized (after a normalization of its first entry to 1) by the true system parameter vector, i.e. $\theta(P^*) = \theta$.

Theorem 4. Let $\xi = (\xi_1, \xi_2)$ be an arbitrary point of the first quadrant of $\mathcal{R}^2$ and r the straight line from the origin through ξ. Its intersection with $\mathcal{S}(R)$ is the point $P = (\lambda_u, \lambda_y)$, with

$$\lambda_u = \xi_1 \,/\, \lambda_M \qquad \lambda_y = \xi_2 \,/\, \lambda_M \tag{102}$$

$$\lambda_M = \max \operatorname{eig}\left(R^{-1} \tilde{R}_\xi\right) \tag{103}$$

$$\tilde{R}_\xi = \operatorname{diag}\left[\, \xi_2\, I_{n+1},\, \xi_1\, I_{n+1}\right]. \tag{104}$$

6. THE FRISCH SCHEME AND THE GIVE FRAMEWORK

The determination of the point P^* on $\mathcal{S}(R)$ leads to the solution of the identification problem. 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. However, on the contrary to the static case, the shift–invariant structure of the underlying dynamic system allows to add further equations, which lead to the determination of a unique solution.

Note that (95) consists of $2n+2$ algebraic non–linear equations. The number of unknowns is $2n+3$, i.e. the free coefficients of θ in (10) and the two variances λ_u^* and λ_y^*. 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).

To explain how the Frisch scheme can be formalized and solved within the GIVE framework, we can write the equation (95) as

$$f_1 = 0. \tag{105}$$

This is an undetermined system of equations, with $2n+2$ equations and $2n+3$ unknowns. Its general solution can therefore be described with a parametrized form, using a single parameter. There are several possibilities for this parameterization. For example, one can choose the slope ξ_2/ξ_1 of the line r defined in Theorem 4. As an alternative, one can choose one of the noise variances, λ_u or λ_y. In the GIVE context, the Frisch scheme equations (105) must be complemented with one or more equations. For example, high order Yule–Walker equations can be exploited for this purpose. For details, see (Diversi et al., 2006) for the time domain approach and (Soverini and Söderström, 2015) for the frequency domain case. These additional equations can be symbolically written as

$$f_2 \approx 0. \tag{106}$$

Several situations can occur.

1 A first case is when there is a single additional equation, so that $\dim(f_2) = 1$. Then, the total system of equations

$$f = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = 0 \tag{107}$$

has $2n+3$ equations and $2n+3$ unknowns. By solving $f_1 = 0$, one can get $f_2 = 0$ as a single equation in the remaining unknown parameter.

2 A second case is when more than one equation is added, so that $\dim(f_2) > 1$. Then, the total system of equations

$$f = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \approx 0 \tag{108}$$

is overdetermined and cannot be solved exactly. In this second case, two options are possible.

2.1 As a first option, the system of equations (108) is solved approximately in a weighted least squares sense

$$\min ||f||^2_{\mathbf{W}}, \tag{109}$$

where $\mathbf{W}$ is a possible nonnegative definite weighting matrix, chosen by the user.

2.2 Another option is to require the equations f_1 hold exactly, and to minimize the equations f_2 in a weighted least squares sense, i.e. to find

$$\min ||f_2||^2_{\mathbf{W_2}} \tag{110}$$

with $f_1 = 0$, and $\mathbf{W_2}$ is a possible weighting matrix.

Indeed, the selection criteria proposed for identifying the EIV models with the Frisch scheme (Guidorzi et al., 2008; Söderström, 2018) fall in the case 2.2. In this case it is appropriate to use one of the parameterizations defined above. Instead, the GIVE method performs directly the minimization of the condition (109), containing the unknowns λ_u, λ_y and θ_1, where θ_1 collects the $2n+1$ free parameters of θ. It is worth observing that the criterion (110) falls within (109) with a suitable choice of $\mathbf{W}$. For solving (109) it is not possible to make use of the one–dimensional parameterization of the Frisch scheme. However, one can observe that the criterion (109) is quadratic on θ_1 and it is possible to write a concentrated loss function of the two variables λ_u and λ_y that allows a robust solution of the optimization problem. For the details, see (Söderström, 2018).

Remark 14. It is worth observing that the algorithmical aspects described at the previous points do 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.

7. CONCLUSIONS

This paper has proposed an unifying framework for the time and frequency domain definitions of the Frisch scheme. By making use of the DFT properties, a thorough analysis of the analogies and differences between these two different formulations has been carried out and some new results have been reported with reference to the frequency domain. Then, the paper has also reported the main results concerning the mathematical and geometrical aspects of the Frisch scheme, that hold both in time and frequency. Finally, the paper has briefly recalled the links between the Frisch scheme and the GIVE method.

REFERENCES

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. (2006). Yule–Walker equations in the Frisch scheme solution of errors–in–variables identification problems, *17th International Symposium on Mathematical Theory of Networks and Systems*, Kyoto, Japan, 391–395.

Frisch, R. (1934). Statistical Confluence Analysis by Mean of Complete Regression Systems. *Economic Institute Pub*, Oslo University.

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

Ljung, L. (1993). Some results on identifying linear systems using frequency domain data. *32nd IEEE Conference on Decision and Control*, San Antonio, Texas, 567–569.

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

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

Pintelon R. and Schoukens, J. (2012). *System Identification: a Frequency Domain Approach*, IEEE Press, New York.

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

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

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.

Söderström T. and Stoica, P. (1989). *System Identification*, Prentice–Hall International, Hemel Hempstead, UK.

Soverini U. and Söderström, T. (2014). Frequency domain EIV identification: a Frisch scheme approach. *19th 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. *14th European Control Conference*, Linz, Austria, 2038–2043.

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

Soverini U. and Söderström, T. (2019b). The Frisch scheme: time and frequency domain aspects. *Tech. Report 2019-010.* Dept. Information Technology, Uppsala University, Sweden.

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