Combining the Frisch scheme and Yule–Walker equations for identifying multivariable errors–in–variables models

Roberto Diversi and Roberto Guidorzi

Abstract—Errors–in–Variables (EIV) models, i.e. models whose stochastic environment considers measurement errors on both inputs and outputs are intrinsically more realistic than representations assuming an exact knowledge of the input but are also more difficult to estimate. The difficulties increase in a non trivial way passing from the SISO and MISO cases to the MIMO one. This paper proposes a procedure for EIV identification of MIMO processes based on the Frisch scheme that assumes additional white noises on all inputs and outputs and shows its effectiveness by means of Monte Carlo simulations.

I. INTRODUCTION

The identification of errors–in–variables (EIV) models is a challenging problem that, despite the advantages that it can offer in many contexts, has received an increasing attention only in the last three decades [1]–[16]. The EIV environment enjoys some peculiarities not shared by other classes of models; among them, the symmetrical presence of additive noise on all variables (inputs and outputs if a model orientation is considered). EIV models are particularly suitable when the interest is mainly focused on the inner laws that describe the considered process and/or on filtering applications rather than on the prediction of its future behavior. This class of models has thus found many applications in system engineering, signal processing and econometrics [17], [18].

The identification of errors–in–variables models is more complex than that of traditional equation error models like, for instance, ARX and ARMAX where the input is considered as exactly known. An interesting EIV identification approach is based on the Frisch scheme, i.e. on the stochastic context originally introduced by Ragnar Frisch in the static case for evaluating the family of linear relations linking sets of noisy observations compatible with the data covariance matrix [2], [19]. The extension of the original algebraic context to the dynamic one leads to the use of EIV models where the measures of the inputs and outputs are affected by additive and mutually uncorrelated white noises.

The dynamic case shares many interesting algebraic and geometric properties with the static one like the existence of convex singularity hypersurfaces defining the family of solutions that can be associated with a set of observations. Moreover, the dynamic properties of the system allow the asymptotic determination of the exact model of the process that has generated the data, not just the definition of a class of compatible models [4]. This nice property is based on the existence of a common point between the singularity hypersurfaces associated with all models with orders greater than that of the process that has generated the data. Unfortunately this property cannot be observed on finite sequences so that it is necessary to introduce suitable cost functions in order to extract a single model from the data [8], [9], [10].

The extension of the Frisch scheme to the identification of multi–input multi–output (MIMO) errors–in–variables models introduces further problems not present in the SISO and MISO cases [6], [20]. In fact, the whole systems can be viewed, in this case, as a set of \( m \) (\( m \) = number of outputs) interconnected subsystems, that can be associated with \( m \) singularity hypersurfaces so that these subsystems could be separately identified by means of the same procedures used for the MISO case. An approach of this kind requires \( m \) separate searches and, even worse, leads to \( m \) different estimates for every input and output noise variance with a consequent loss of congruence for the global model. It would be thus preferable to select cost functions treating the model as a whole and not as a set of separate models. A geometric approach to the solution of this problem has been proposed in [6], [15] where multivariable EIV models have been no longer associated to single points on singularity hypersurfaces but to directions in the noise space. This approach, that relies, from a computational standpoint, on the easy computation of the intersection between a straight line from the origin and a singularity hypersurface [21], offers the additional advantage of an easy introduction of congruent cost functions [16].

The identification procedure introduced in this paper takes advantage of the properties of both the dynamic Frisch scheme and high order Yule–Walker equations and is based on the aforementioned geometric approach so that congruent estimations of the system parameters and noise variances can be obtained. The method is characterized by a good estimation accuracy and gives better results than the approach introduced in [16], as shown by the performed Monte Carlo simulations.

The organization of the paper is as follows. Section II defines the EIV MIMO identification problem. Section III describes the properties of the dynamic Frisch scheme in the multivariable case while Section IV introduces the new identification procedure. The results obtained in some Monte Carlo simulations are reported in Section V. Short concluding remarks are finally given in Section VI.
II. PROBLEM STATEMENT

Consider the MIMO EIV model represented in Fig. 1, described by the equations

\[
Q(z) y_0(t) = P(z) u_0(t) \quad (1)
\]
\[
u(t) = u_0(t) + \tilde{u}(t) \quad (2)
\]
\[
y(t) = y_0(t) + \tilde{y}(t), \quad (3)
\]

where \(u_0(t) \in \mathbb{R}^r\), \(y_0(t) \in \mathbb{R}^m\) are the noise–free input and output

\[
u(t), y(t) \quad \text{are the available observations affected by the additive noises} \quad \tilde{u}(t), \tilde{y}(t) \quad \text{and} \quad Q(z), P(z) \quad \text{are} \quad (m \times m) \quad \text{and} \quad (m \times r) \quad \text{polynomial matrices in the unitary advance operator} \quad z (y(t) = y(t + 1)).
\]

By selecting a suitable minimal parametrization [22], [23], model (1) can be partitioned into the set of \(m\) relations

\[
y_{0i}(t + \nu_i) = \sum_{j=1}^{m} \sum_{k=1}^{\nu_{ij}} \alpha_{ijk} y_{0j}(t + k - 1) + \sum_{j=1}^{\nu_{ij}} \sum_{k=1}^{r} \beta_{ijk} u_{0j}(t + k - 1) \quad i = 1, \ldots, m
\]

where the integers \(\nu_i \quad (i = 1, \ldots, m)\) are the Kronecker observability invariants of the system while the integers \(\nu_{ij}\) are completely defined by these invariants through the relations

\[
\nu_{ij} = \nu_i \quad \text{for} \quad i = j
\]
\[
\nu_{ij} = \min(\nu_i + 1, \nu_j) \quad \text{for} \quad i > j
\]
\[
\nu_{ij} = \min(\nu_i, \nu_j) \quad \text{for} \quad i < j.
\]

The maximal observability invariant will be denoted as \(\nu_M\)

\[
\nu_M = \max_i \{\nu_i\} \quad i = 1, \ldots, m;
\]

the system order \(n\), i.e. the McMillan degree of \(Q(z)^{-1} P(z)\), is given by

\[
n = \sum_{i=1}^{m} \nu_i.
\]
for $i = 1, \ldots, m$, whose entries are noise–free, noisy and noise samples respectively. By defining also the parameter vectors

$$\theta^*_i = \left[ \alpha_{i11} \ldots \alpha_{i1\nu_i} \alpha_{i21} \ldots \alpha_{i2 \nu_i} \ldots \right. $$

$$\alpha_{i11} \ldots \alpha_{i1 \nu_i} \ 1 \ldots \alpha_{im1} \ldots \alpha_{im \nu_{im}}$$

$$\beta_{i11} \ldots \beta_{i1 (\nu_i + 1)} \ldots \beta_{ir1} \ldots \beta_{ir (\nu_i + 1)} \right]^T$$

(10)

for $i = 1, \ldots, m$, and taking into account (6), (2), (3) it is possible to write

$$\varphi_0^T(t) \theta^*_i = 0\quad (11)$$

$$\varphi_i(t) = \varphi_0 i(t) + \varphi_i(t),$$

(12)

for $i = 1, \ldots, m$.

With reference to the $i$–th subsystem, define the covariance matrices of noise–free data, noisy data and noise

$$\Sigma_{0i} = E [\varphi_0 i(t) \varphi_0^T(t)]$$

$$\Sigma_i = E [\varphi_i(t) \varphi_i^T(t)]$$

$$\Sigma_i^* = E [\tilde{\varphi}_i(t) \tilde{\varphi}_i^T(t)]$$

From (11), (12) and assumption A5 it follows that

$$\Sigma_{0i} \theta^*_i = 0$$

$$\Sigma_i = \Sigma_{0i} + \Sigma_i^*$$

where, because of assumptions A4 and A5

$$\Sigma_i^* = \text{diag} \left[ \alpha_{y_11}^2 I_{\nu_1}, \ldots, \alpha_{y_i1}^2 I_{\nu_i}, \ldots, \alpha_{y_m1}^2 I_{\nu_m} \right.$$  

$$\alpha_{y_i1}^2 I_{\nu_i + 1}, \ldots, \alpha_{y_i}^2 I_{\nu_i}, \ldots, \alpha_{y_m}^2 I_{\nu_m} + 1, \ldots, \alpha_{y_m}^2 I_{\nu_m} + 1 \right].$$

(18)

Note that $\Sigma_{0i}$ is a positive semidefinite matrix whose null space has dimension one and defines the parameter vector $\theta^*_i$. This matrix can be obtained from the noisy data covariance matrix $\Sigma_i$ if the noise variances $\sigma^2_{y_1}, \ldots, \sigma^2_{y_m}, \sigma^2_{u_1}, \ldots, \sigma^2_{u_r}$ are known.

Consider now the set of all diagonal matrices $\tilde{\Sigma}_i(P)$ of type (18) that exhibit the same properties of $\Sigma_i^*$, i.e. the set of points $P = (\tilde{\sigma}_{12}, \ldots, \tilde{\sigma}_{m2}, \tilde{\sigma}_{13}, \ldots, \tilde{\sigma}_{m3}, \ldots, \tilde{\sigma}_{m r+1}) \in \mathcal{R}^{m+r}$ such that

$$\Sigma_i - \tilde{\Sigma}_i(P) \geq 0, \quad \min \text{eig}(\Sigma_i - \tilde{\Sigma}_i(P)) = 0, \quad (19)$$

where

$$\tilde{\Sigma}_i(P) = \text{diag} \left[ \tilde{\sigma}_{12}^2 I_{\nu_1}, \ldots, \tilde{\sigma}_{i2}^2 I_{\nu_i}, \ldots, \tilde{\sigma}_{m2}^2 I_{\nu_m}, \right.$$  

$$\tilde{\sigma}_{13}^2 I_{\nu_1 + 1}, \ldots, \tilde{\sigma}_{i3}^2 I_{\nu_i + 1}, \ldots, \tilde{\sigma}_{m3}^2 I_{\nu_m + 1}, \tilde{\sigma}_{m+1}^2 I_{\nu_{m+1}}, \ldots, \tilde{\sigma}_{m+r}^2 I_{\nu_{m+r}} \right].$$

(20)

It is possible to prove the following result [4].

**Theorem 1:** The set of points $P$ satisfying (19) defines a hypersurface $S(\Sigma_i)$ belonging to the first orthant of $\mathcal{R}^{m+r}$. Every point $P$ of $S(\Sigma_i)$ is an admissible solution in the noise space and can thus be associated with a possible vector of parameters $\theta_i(P)$ given by

$$\left( \Sigma_i - \tilde{\Sigma}_i(P) \right) \theta_i(P) = 0.$$

(21)

As an example, Fig. 2 shows a typical hypersurface for a system with one input and two outputs.
(3) it is easy to show that

\[ \varphi^T(t) \Theta^* = 0 \]  

\[ \varphi(t) = \varphi_0(t) + \tilde{\varphi}(t). \]  

Introduce also the covariance matrices

\[ \Sigma_0 = E[\varphi_0(t) \varphi_0^T(t)] \]  

\[ \Sigma = E[\varphi(t) \varphi^T(t)] \]  

\[ \tilde{\Sigma}^* = E[\tilde{\varphi}(t) \tilde{\varphi}^T(t)]. \]  

From (27) and (28) it follows immediately that

\[ \Sigma_0 \Theta^* = 0 \]  

\[ \Sigma = \Sigma_0 + \tilde{\Sigma}^*. \]  

where

\[ \tilde{\Sigma}^* = \text{diag} \left[ \sigma_{y_1}^2 I_{\nu_1}, \ldots, \sigma_{y_m}^2 I_{\nu_m+1}, \sigma_{u_1}^2 I_{\nu_M+1}, \ldots, \sigma_{u_k}^2 I_{\nu_M+1} \right]. \]  

Matrix \( \Sigma_0 \) is thus positive semidefinite with a null space of dimension \( m \).

Consider now the family of all diagonal matrices \( \tilde{\Sigma}(P) \) of type (34) which exhibit the same properties of \( \tilde{\Sigma}^* \), i.e. the set of points \( P \in \mathbb{R}^{m+r} \) such that

\[ \Sigma - \tilde{\Sigma}(P) \geq 0, \quad \min\text{eig}(\Sigma - \tilde{\Sigma}(P)) = 0, \]  

where \( \tilde{\Sigma}(P) \) has the structure (34). This set is described by the following theorem that constitutes a simple extension of the result reported in [4].

**Theorem 3:**  The set of points satisfying (35) defines an hypersurface \( S(\Sigma) \) belonging to the first orthant of \( \mathbb{R}^{m+r} \) and lying under \( S(\Sigma_1), \ldots, S(\Sigma_m) \). Because of (32) and (33), the point \( P^* \), characterized by the true noise variances, belongs to \( S(\Sigma) \) and \( \Sigma - \tilde{\Sigma}(P^*) = \Sigma_0 \).

The properties of \( S(\Sigma) \) play an important role in the solution of Problem 1, as it will be shown in the next section.

IV. IDENTIFICATION OF MIMO EIV MODELS

Although of great theoretical interest, the properties described by Theorems 2 and 3 cannot be directly used in practice. In fact, in presence of a finite number \( N \) of noisy data, the covariance matrices \( \Sigma_i, i = 1, \ldots, m \) and \( \Sigma \) must be replaced by their sample estimates so that the associated hypersurfaces do no longer share any common point.

It is of course possible to identify the \( m \) subsystems separately by using one of the selection criteria proposed for the SISO or MISO cases [8, 9, 10]. The generic subsystem \( i \) would thus be treated as a MISO model where all outputs except the \( i \)-th one play the role of inputs. Nevertheless, an approach of this kind is characterized by two main drawbacks. Firstly, it is necessary to solve \( m \) optimization problems of complexity \( m + r \). This increases the computational burden but it is much more important to observe that this procedure leads to \( m \) different estimates for every input and output noise variance with a consequent loss of congruence. To overcome these problems, the search procedure and the selection criterion should treat the \( m \) subsystems as a whole MIMO system.

For this reason, the solution of Problem 1 will be based on the geometric approach proposed in [6, 15], where multivariable EIV models have been no longer associated to single points on singularity hypersurfaces but to directions in the noise space. This approach relies on the computation of the intersection between a straight line from the origin and a singularity hypersurface described by the following theorem [21].

**Theorem 4:**  Let \( \xi = (\xi_1, \xi_2, \ldots, \xi_{m+r}) \) be a generic point of the first orthant of \( \mathbb{R}^{m+r} \) and \( \rho \) the straight line from the origin trough \( \xi \). Its intersection with \( S(\Sigma) \) is the point \( P_i \) given by

\[ P_i = \frac{\xi}{\lambda_{M_i}}, \quad \lambda_{M_i} = \max \text{eig} \left( \Sigma_0^{-1} \tilde{\Sigma}_i \right) \]  

where

\[ \tilde{\Sigma}_i = \text{diag} \left[ \xi_1 I_{\nu_1}, \ldots, \xi_m I_{\nu_m}, \xi_{m+1} I_{\nu_{M+1}}, \ldots, \xi_{m+r} I_{\nu_{M+1}} \right]. \]

**Remark 3:**  To compute the intersection \( \hat{P} \) between \( S(\Sigma) \) and \( \rho \), (36) will be replaced with

\[ \hat{P} = \frac{\xi}{\lambda_{M}}, \quad \lambda_{M} = \max \text{eig} \left( \Sigma^{-1} \tilde{\Sigma} \right), \]

where \( \tilde{\Sigma} \) has the structure (34).

It is thus possible to consider the \( m \) intersections \( P_1, P_2, \ldots, P_m \) of a line \( \rho \) from the origin with the \( m \) singularity surfaces \( S(\Sigma_1), \ldots, S(\Sigma_m) \) that define univocally the parameters \( \theta_1(P_1), \theta_2(P_2), \ldots, \theta_m(P_m) \) i.e. a whole MIMO model. The use of this approach, which is valid also for finite numbers of data, still leads to \( m \) different values of each noise variance. However, since \( P_1, P_2, \ldots, P_m \) belong to the same straight line from the origin, the ratios between the noise variances are univocally defined.

It is now necessary to introduce a selection criterion to identify the “true” direction in the noise space, i.e. that associated with the true model. The criterion introduced in the next subsection takes advantage of the properties of both the high–order Yule–Walker equations and the hypersurface \( S(\Sigma) \).

A. The selection criterion

Let us define the \( r q \times 1 \) (being \( q \) a user–chosen parameter) vector of delayed inputs

\[ \varphi^d_u(t) = [u_1(t-q) \ldots u_1(t-1) \quad u_2(t-q) \ldots u_2(t-1) \ldots u_r(t-q) \ldots u_r(t-1)]^T, \]

that, because of (2), satisfies the condition

\[ \varphi^d_u(t) = \varphi^d_{u0}(t) + \tilde{\varphi}^d_u(t), \]
where
\[ \varphi(t) = [u(t-q), \ldots, u(t-1)]^T, \]
\[ u(t) = \varphi(t) \theta, \]
and has been constructed with the entries of the covariance matrices \( \Sigma_i \) and \( \Sigma \).

Relation (46) represents a set of high order Yule–Walker equations that can be used to define a selection criterion.

For this purpose, consider a direction \( \rho \) in the noise space, its intersections \( P_1, \ldots, P_m \) with \( S(\Sigma_1), \ldots, S(\Sigma_m) \) and the associated MIMO model \( \theta_1(P_1), \theta_2(P_2), \ldots, \theta_m(P_m) \). Relation (46) allows the introduction of the cost function
\[ J(P_1, \ldots, P_m) = J(\rho) \]
\[ = \text{trace} \left( \Theta^T(P_1, \ldots, P_m) (\Sigma^d)^T \Sigma^d \Theta(P_1, \ldots, P_m) \right), \]
where \( \Theta(P_1, \ldots, P_m) = \Theta(\rho) \) has the same structure of \( \Theta^* \) and has been constructed with the entries of \( \theta_1(P_1), \ldots, \theta_m(P_m) \). This function satisfies the conditions
1) \( J(\rho) \geq 0 \)
2) \( J(\rho) = 0 \) if \( P_1 = \cdots = P_m = P^* \),
where the second condition follows directly from (46). Problem 1 can thus be solved by minimizing \( J(\rho) \) in the noise space. Of course, when the covariance matrices are replaced by their sample estimates relation (46) does no longer hold exactly and the minimum of the cost function (47) will be greater than zero.

**B. The identification algorithm**

The whole identification procedure is summarized by the following algorithm.

**Algorithm 1.**

1) Compute, on the basis of the noisy observations \( u(1), \ldots, u(N), y(1), \ldots, y(N) \), the sample estimates of the covariance matrices \( \Sigma_1, \ldots, \Sigma_m \) and \( \Sigma \):
\[ \hat{\Sigma}_i = \frac{1}{N - \nu M} \sum_{t=1}^{N-\nu M} \varphi_i(t) \varphi_i^T(t), \]
\[ \hat{\Sigma} = \frac{1}{N - \nu M} \sum_{t=1}^{N-\nu M} \varphi(t) \varphi^T(t). \]

2) Choose a value of the delay \( q \) in (39) and compute the sample estimate of the covariance matrix \( \Sigma^d \):
\[ \hat{\Sigma}^d = \frac{1}{N - \nu M - q} \sum_{t=q+1}^{N} \varphi(t) \varphi^T(t). \]

3) Start from a generic direction \( \rho \) in the first orhant of \( \mathcal{R}^{m+r} \).
4) Compute the intersections \( P_1, \ldots, P_m \) between \( \rho \) and \( S(\Sigma_1), \ldots, S(\Sigma_m) \).
5) Construct the diagonal matrices \( \hat{\Sigma}(P_1), \ldots, \hat{\Sigma}(P_m) \) and compute the parameter vectors \( \theta_1(P_1), \theta_2(P_2), \ldots, \theta_m(P_m) \) by means of the relations
\[ \left( \hat{\Sigma}_i - \hat{\Sigma}(P_i) \right) \theta_i(P_i) = 0, \quad i = 1, \ldots, m. \]
6) Construct, with the entries of \( \theta_1(P_1), \ldots, \theta_m(P_m) \) the matrix \( \Theta(\rho) \) of type (25).
7) Compute the cost function \( J(\rho) \) (47).
8) Move to a new direction \( \rho \pm \Delta \rho \) corresponding to a decrease of \( J(\rho) \).
9) Repeat steps 3–8 until the direction \( \rho^* \) corresponding to the minimum of \( J(\rho) \) is found.

Algorithm 1 involves the computation of the points \( P_1, \ldots, P_m \) and \( \rho \) in the noise space so that it gives \( m + 1 \) different estimations for each noise variance \( \sigma_{i1}^2, \ldots, \sigma_{im}^2, \sigma_{u1}^2, \ldots, \sigma_{um}^2 \). Even if other choices are possible we will consider as noise variance the values obtained by the point \( \rho \) on \( S(\Sigma) \). From Theorem 3 it is easy to show that these values lead to matrices \( \Sigma_i - \hat{\Sigma}(P_i) \), \( i = 1, \ldots, m \), that are all positive definite. The algorithm is thus concluded as follows

10) Compute the intersection \( \hat{P} \) between \( \rho^* \) and \( S(\Sigma) \) and consider the entries of \( \hat{P} \) as estimates of the true noise variances.

**Remark 4:** It is important to observe that the whole procedure requires only the computation of a very limited number of points of the considered singularity hypersurfaces, i.e. only their intersections with the directions tested by the adopted optimization algorithm. The whole procedure is thus very efficient from a numerical point of view.

**V. Numerical results**

The effectiveness of the proposed identification method has been tested by means of Monte Carlo simulations and compared to that of the identification procedure described in [16]. The numerical example concerns the two–input two–output model considered in [16], characterized by the polynomial matrices
\[ Q(z) = \begin{bmatrix} z^2 - 0.4 z + 0.3 & z + 0.2 \\ -0.2 z + 0.1 & z^2 - 0.4 \end{bmatrix} \]

\[ P(z) = \begin{bmatrix} 0.6 z^2 - 0.1 & 0.2 z^2 + 0.8 z + 1 \\ 0.2 z^2 + 0.5 z + 0.5 & 0.4 z^2 - z + 0.5 \end{bmatrix}. \]

The true inputs \( u_{01}(t), u_{02}(t) \) are pseudo random binary sequences of unit variance and length \( N = 2000 \). The true
The values of the noise variances are

\[ \tilde{\sigma}_{u_1}^2 = 0.1 \quad \tilde{\sigma}_{u_2}^2 = 0.1 \quad \tilde{\sigma}_{y_i}^2 = 0.75 \quad \tilde{\sigma}_{u_1}^2 = 0.22, \]

corresponding to signal-to-noise ratios of about 10 dB on both inputs and outputs.

Two experiments have been performed. In the first experiment the tested algorithms are initialized by choosing a direction \( \rho \) near the true one while in the second experiment the initial direction is far from the true one. For both cases a Monte Carlo simulation of 100 independent runs has been carried out.

The value of \( q \) in step 2 has been set to 5 since it has been observed that greater values do not lead to significant improvements. The obtained results are shown in Tables 1–3 where the true value of parameters and noise variances, the means of their estimates and the corresponding standard deviations are reported. The proposed Yule–Walker based algorithm is denoted as YW while the approach [16], based on the null space of \( \Sigma_0 \), is denoted as NS. It can be observed that the YW method is characterized by better accuracy and higher robustness with respect to the initial estimate of the noise space direction.
VI. CONCLUSIONS

This paper has proposed a MIMO EIV identification procedure based on the use of a general class of minimally parameterized models particularly suitable for diagnosis and filtering applications. This procedure is based on a geometric approach that associates EIV models with directions in the noise space and leads to the computation of both model parameters and noise variances. The performed Monte Carlo simulations show the effectiveness of this approach.

REFERENCES