THE FRISCH SCHEME IN ALGEBRAIC AND DYNAMIC IDENTIFICATION PROBLEMS

ROBERTO GUIDORZI, ROBERTO DIVERSI AND UMBERTO SOVERINI

This paper considers the problem of determining linear relations from data affected by additive noise in the context of the Frisch scheme. The loci of solutions of the Frisch scheme and their properties are first described in the algebraic case. In this context two main problems are analyzed: the evaluation of the maximal number of linear relations compatible with data affected by errors and the determination of the linear relation actually linking the noiseless data. Subsequently the extension of the Frisch scheme to the identification of dynamical systems is considered for both SISO and MIMO cases and the problem of its application to real processes is investigated. For this purpose suitable identification criteria and model parametrizations are described. Finally two classical identification problems are mapped into the Frisch scheme, the blind identification of FIR channels and the identification of AR + noise models. This allows some theoretical and practical extensions.

Keywords: system identification, errors-in-variables models, Frisch scheme, linear systems

AMS Subject Classification: 93E12

1. INTRODUCTION

The search for connections between observations ("laws of nature") is at the basis of the development of scientific knowledge and can be traced back at least some thousand years as shown, for instance, by the large amount of clay tablets concerning the so-called astronomical diaries compiled by the Mesopotamian astronomers. This search for knowledge is characterized by two basic steps, the necessity of performing a quantification of observations, i.e. the transformation of observations into numerical entities and the subsequent extraction of relations between the obtained values, to be used for interpretation, prediction, control or other purposes. If we observe n different variables

$$x_1, \; x_2, \; \ldots, \; x_n \tag{1}$$

and denote with

$$x_{1i}, \; x_{2i}, \; \ldots, \; x_{ni} \tag{2}$$

the values that these variables assume at the ith observation, the search for a law describing the behavior of the process that has generated the observations is the

search for a mathematical relation

$$f(x_1, x_2, \ldots, x_n) = 0 \tag{3}$$

satisfied by every set of observations, i.e. such that, for every i,

$$f(x_{1i}, x_{2i}, \ldots, x_{ni}) = 0. \tag{4}$$

Even assuming that the considered process is actually governed by a law of the type (3), the observations will never satisfy, in all practical situations, relation (4) because of the errors that will be inevitably introduced during the quantification step (e.g. noise in analog systems, finite number of possible values and noise in digital environments etc.). The deduction of the law behind the observations is thus a problem that does not admit any solution because the observations will not satisfy, in general, any relation. All procedures leading to the extraction of abstract relations from real data rely, in fact, on modified observations. This modification process should be carried out on the basis of the exact knowledge of the nature of the errors since it can affect, in a very substantial way, the final result.

As a simple example to illustrate this point, let us assume the existence of a linear relation, described by the scalars $\alpha_1, \alpha_2, \ldots, \alpha_n$, linking the variables (1):

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n = 0. \tag{5}$$

The same relation can be described also, when $\alpha_i \neq 0$, by means of the equivalent, asymmetric relation

$$x_i = \beta_1\, x_1 + \cdots + \beta_{i-1}\, x_{i-1} + \beta_{i+1}\, x_{i+1} + \cdots + \beta_n\, x_n \tag{6}$$

where $\beta_j = -\alpha_j/\alpha_i$. Let us assume also that only one of the observations, x_{ki} is affected by zero-mean additive errors. By denoting with $\hat{x}_{ji}$ the true values, with $\tilde{x}_{ji}$ the observation errors and with x_{ji} the actual observations, these quantities will be linked by the relations

$$x_{ji} \;=\; \hat{x}_{ji} \quad \text{for } j \neq k \tag{7}$$
$$x_{ki} \;=\; \hat{x}_{ki} + \tilde{x}_{ki}. \tag{8}$$

This is the well known context of the least squares and we can easily find the optimal and asymptotically unbiased solution (Gauss–Markov theorem) by means of the least squares algorithm. Note, however, that this can be done if and only if we actually know *which* variable is affected by observation errors since, from a geometrical point of view, we perform an orthogonal projection of the vector of the N noisy observations $[x_{k1}\, x_{k2}\, \ldots\, x_{kN}]^T$ on the (hyper)plane defined by the vectors of noiseless observations $[x_{j1}\, x_{j2}\, \ldots\, x_{jN}]^T$ $(j \neq k)$ and *substitute* $[x_{k1}\, x_{k2}\, \ldots\, x_{kN}]^T$ with its orthogonal projection. If we remain in this context (a single observation is affected by errors) but with no *a priori* information about which of the variables is noisy, the problem cannot be solved. In fact it is impossible to select the correct solution among the n possible ones that can be obtained by considering as affected by errors in turn x_1, x_2, $\ldots$, x_n. Moreover, if the errors are not restricted to a single variable, none of these n solutions will be, even asymptotically, correct.

This elementary example outlines two basic points:

1. The extraction of models from data affected by errors requires an intermediate step consisting in the deduction of new data from the available ones. The model will then be deduced from these new data, not from the original ones.

2. In absence of precise information on the errors it is possible to formulate different assumptions, each leading, in general, to extract different sets of data from the observations and, consequently, to different models.

Every systematic procedure to deduce a model (or a family of models) from data affected by errors is defined as a "scheme" [29, 30]. Many different schemes have been investigated and described in the literature. None of these schemes can be considered, *per se*, as superior to any other; what changes is simply the set of assumptions. As a consequence, the results that can be obtained by applying different schemes to the same set of data depend essentially on the "distance" between the assumptions behind the scheme and the actual situation; all other claims are related more to faith than to science. A very complete analysis of the assumptions behind different schemes has been proposed by Kalman in [29, 30, 31].

The content of this paper falls inside the Errors-in-Variables (EIV) context that assumes the presence of additive noise on all variables. This is a challenging environment that has seen an increasing amount of research only during the last decades. One of the appealing features of the models that can be deduced in EIV contexts concerns their intrinsic capability of relying on a limited set of *a-priori* assumptions [41, 42]. This feature suggests the use of EIV models in applications like, for instance, diagnosis, where the interest is focused on a realistic description of a process rather than on other aspects, like prediction. For a complete overview on EIV identification see [37] and the references therein.

This paper concerns the scheme proposed by the Nobel prize Ragnar Frisch in 1934 [17] and its application to the problem of deducing linear relations from noisy observations concerning both algebraic processes (this term is used here to denote static processes that can be described by sets of relations linking measures performed at the same time) and dynamic processes (i.e. processes described by difference equations). The Frisch scheme is an interesting compromise between the great generality of the EIV environment and the possibility of performing real applications. Moreover, the Frisch scheme encompasses some other important schemes like, for instance, Least Squares and the Eigenvector Method and plays, consequently, an important role also from a conceptual point of view.

The Frisch scheme does not lead, at least in the algebraic case, to a single solution but to a whole family of solutions compatible with a given set of noisy observations. This fact has often diverted the attention towards simpler schemes leading to a single solution so that the smart environment proposed by Frisch has not received, for many decades, the attention that it deserved.

As it will be shown in the following, the analysis of the Frisch scheme leads to two separate loci of solutions, one in the parameter space and the other in the space of the noise variances; of course the points of these loci are linked by well defined relations. Some fundamental results [29, 31] describe these maps as well as the shape of the loci in the parameter space under specific conditions. Unfortunately, however, the locus of solutions in the parameter space can be easily defined only when the data

are compatible with a single linear relation; in all other cases the performed analyses have evidentiated the extremely complex structure of this locus, that prevents its practical use [3]. The investigation of the properties of the locus of solutions in the noise space has, on the contrary, offered a key for a deeper analysis that shows that this locus does never degenerate and enjoys some consistent properties [19, 20].

A problem of great importance in the econometric field consists in determining the maximal number of linear relations compatible with a given set of noisy data. The importance attributed to this problem is due to the fact that its solution is considered as linked to the extraction of the maximal information from the data [35]. The solution of this problem in the context of the Frisch scheme has been possible only by making reference to the properties of the locus of solutions in the noise space [21]; other approaches have led to the computation of an upper bound for this number [43].

When the data are generated by a linear time-invariant dynamic process and the Frisch context is used for its identification, it is necessary to consider the loci of solutions under the constraints imposed by the time shift properties of dynamic systems [5]. It can be surprising to discover that, in this respect, the dynamic case can be seen as a subcase of the algebraic one and that the shift properties of dynamic systems lead (in general) to a unique solution [2, 8, 39]. Moreover this solution is linked to the solution of the maximal corank problem in the algebraic case.

All previous statements are true when the assumptions behind the Frisch scheme are exactly fullfilled and this can be assumed, at most, only in asymptotic conditions. In all practical cases this cannot be achieved not only because real data sets are necessarily limited but also because of a whole series of violations due to non linearity, non stationarity etc. The development of Frisch identification procedures requires thus the introduction of suitable criteria [4, 10, 13]. The Frisch scheme in the identification of dynamic processes enjoys some peculiarities that make it particularly suitable for the solution of specific problems like filtering and fault detection and isolation [9, 25].

The purpose of this paper is to outline some results obtained in the analysis of the Frisch scheme in its original algebraic context and, in particular, to discuss the solution of the problem of determining the maximal number of linear relations compatible with a given set of noisy data. Other relevant topics of this analysis concern the properties of the loci of the Frisch solutions in the noise and parameter spaces as well as the possibility of obtaining single solutions. This discussion is carried out in Section 2. The second part of the paper regards the extension of the original algebraic environment to the dynamic one, where it is shown how the Frisch scheme can lead, differently from the algebraic case, to a single solution. This topic and the associated algorithms required for the application to real processes are discussed in Section 3, first in the SISO case and then in the MIMO one. Section 4 recalls how some classical problems can be mapped into a dynamical Frisch identification problem and how this can lead to extend the limits of previous approaches. The cases mentioned in this section concern the blind identification of FIR transmission channels and the identification of noisy autoregressive models. Some concluding remarks are finally reported in Section 5.

2. THE FRISCH SCHEME IN THE ALGEBRAIC CASE

2.1. Estimating linear relations from noisy data: statement of the problem

Consider the linear algebraic process (5); by denoting with X the $N \times n$ matrix whose rows contain the N observations (2)

$$X = \begin{bmatrix} x_{11} & x_{21} & \ldots & x_{n1} \\ x_{12} & x_{22} & \ldots & x_{n2} \\ \vdots & \vdots & & \vdots \\ x_{1N} & x_{2N} & \ldots & x_{nN} \end{bmatrix}, \tag{9}$$

relation (5) can be written in the form

$$X A = 0 \tag{10}$$

where

$$A = \begin{bmatrix} \alpha_1\,\alpha_2\,\ldots\,\alpha_n \end{bmatrix}^T. \tag{11}$$

In the more general case of q linear relations between the variables, A will be a $(n \times q)$ matrix with columns given by the q sets of coefficients describing the $q = n - \operatorname{rank} X$ independent linear relations linking the data. Relation (10) can be rewritten also by substituting X with

$$\Sigma = \frac{X^T X}{N} \tag{12}$$

i.e., under the assumption of null mean value of the variables, with the sample covariance matrix of the data. In absence of noise and in presence of linear relations Σ will be singular and positive semidefinite

$$\Sigma \geq 0. \tag{13}$$

Every solution, A, with maximal rank, of the equation

$$\Sigma A = 0 \tag{14}$$

is a basis of $\ker \Sigma$. When the data are corrupted by noise, $\operatorname{rank} X = n$, no linear relations are compatible with the observations and Σ is positive definite

$$\Sigma > 0. \tag{15}$$

In situations of this kind, linear relations can be extracted only by modifying X or Σ, i.e. the data.

2.2. Assumptions behind estimation schemes

If no assumptions are introduced, *any* set of noisy data is compatible with *any* solution. The assumptions usually introduced on the errors (noise) to restrict the number of admissible solutions are the following:

1. The noise is additive; every observation is the sum of an unknown exact part $\hat{x}_i$, and of a noise term $\tilde{x}_i$:

$$x_i = \hat{x}_i + \tilde{x}_i \tag{16}$$

2. The mean value of $\hat{x}_i$ and $\tilde{x}_i$ is null:

$$\sum_{t=1}^{N} \hat{x}_{it} = 0, \qquad \sum_{t=1}^{N} \tilde{x}_{it} = 0. \tag{17}$$

3. The sequences of noise samples are orthogonal to the sequences of noiseless variables:

$$\sum_{t=1}^{N} \tilde{x}_{it} \, \hat{x}_{jt} = 0 \quad \forall \, i, j. \tag{18}$$

Under these assumptions:

$$X = \hat{X} + \tilde{X} \tag{19}$$
$$\hat{X}^T \tilde{X} = 0 \tag{20}$$
$$\Sigma = \hat{\Sigma} + \tilde{\Sigma} \tag{21}$$
$$\Sigma > 0 \tag{22}$$
$$\tilde{\Sigma} \geq 0 \ \text{ or } \ \tilde{\Sigma} > 0 \tag{23}$$
$$\hat{\Sigma} \geq 0 \quad \text{and} \quad \det \hat{\Sigma} = 0. \tag{24}$$

The problem of determining linear relations compatible with noisy data can be formulated as follows:

Problem 1. (Kalman [29, 30]) Given a sample covariance matrix of noisy observations, Σ, determine positive definite or semidefinite noise covariance matrices $\tilde{\Sigma}$ such that

$$\hat{\Sigma} = \Sigma - \tilde{\Sigma} \geq 0 \quad \text{and} \quad \det \hat{\Sigma} = 0. \tag{25}$$

Any basis of $\ker \hat{\Sigma}$ will describe a set of linear relations compatible with the data and with assumptions $(16) - (18)$.

2.3. The Frisch scheme

This scheme, proposed by Ragnar Frisch in 1934 [17], is based on assumptions $(16) -$ (18) and on the further assumption of mutual independence of the noise sequences

$$\sum_{t=1}^{N} \tilde{x}_{it} \, \tilde{x}_{jt} = 0 \quad \forall \, i \neq j. \tag{26}$$

As a consequence of (26), the sample covariance matrix of the noise will be diagonal. By introducing the suffix n to denote the dimension of square matrices, we will thus have

$$\tilde{\Sigma}_n = \mathrm{diag} \left[\, \tilde{\sigma}_1^2, \ldots, \tilde{\sigma}_n^2 \, \right] \geq 0 \qquad \text{or} \qquad > 0 \tag{27}$$

where $\tilde{\sigma}_1^2, \ldots, \tilde{\sigma}_n^2$ are the sample variances of the noise terms $\tilde{x}_1, \ldots, \tilde{x}_n$. Every positive definite or semidefinite diagonal matrix $\tilde{\Sigma}_n$ such that

$$\hat{\Sigma}_n = \Sigma_n - \tilde{\Sigma}_n \geq 0 \quad \text{and} \quad \det \hat{\Sigma} = 0 \tag{28}$$

is a *solution* of the Frisch scheme. The corresponding point $P = (\tilde{\sigma}_1^2, \ldots, \tilde{\sigma}_n^2) \in \mathbb{R}^n$ can be considered as an admissible solution in the noise space while the parameters α_i in (5) or β_i in (6) define the associate solution in the parameter space.

It can be observed that the set of parameters (5) or (6) refers only to the case of $\dim \ker \hat{\Sigma}_n = 1$. A solution in the noise space can, however, be associated also with noise covariance matrices $\tilde{\Sigma}_n$ such that $\dim \ker \hat{\Sigma}_n > 1$ that correspond to multiple independent linear relations between the columns (rows) of $\hat{\Sigma}_n$. The maximal dimension of $\ker \hat{\Sigma}_n$ will be denoted in the following as $\text{Maxcor}_F(\Sigma_n)$ (maximal corank of Σ_n under the assumptions of the Frisch scheme) [30].

2.3.1. Properties of the solutions in the noise space

A problem of great importance in the analysis of the properties and in the application of the Frisch scheme concerns the description of the loci of the solutions in the noise and parameter spaces. While the locus of the solutions in the parameter space has nice properties only in a well defined case (compatibility of the data with a single linear relation under the assumptions of the Frisch scheme, i.e. $\text{Maxcor}_F(\Sigma_n) = 1$), the locus of the solutions in the noise space is the same in every situation and is described by the following theorem [5].

Theorem 1. All admissible solutions in the noise space lie on a convex (hyper)surface $\mathcal{S}(\Sigma_n)$ whose concavity faces the origin and whose intersections with the coordinate axes are the points $(0, \ldots, \tilde{\sigma}_i^2, \ldots, 0)$ corresponding to the n least squares solutions (see Figure 1).
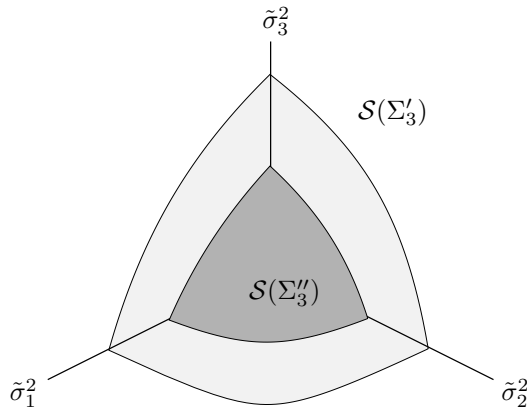

Fig. 1. Loci $\mathcal{S}(\Sigma_3)$ of admissible noise points for $n = 3$ and different amounts of noise.

Definition 1. (Guidorzi [19]) The (hyper)surface $\mathcal{S}(\Sigma_n)$ will be called *singularity (hyper)surface* of Σ_n because its points define noise covariance matrices $\tilde{\Sigma}_n$ associated with singular matrices $\hat{\Sigma}_n$.

A problem of great relevance concerns the conditions under which a covariance matrix is compatible with more linear relations i. e. the evaluation of $\mathrm{Maxcor}_F(\Sigma_n)$. A fundamental result concerning this problem is the following [30].

Theorem 2. $\mathrm{Maxcor}_F(\Sigma_n) = 1$ if and only if all entries of Σ_n^{-1} are positive or can be made positive (Frobenius-like according to the definition of Kalman [30]) by changing the sign of some variables.

Under the conditions of Theorem 2, the locus of solutions in the parameter space is described by the following theorem that shows the great relevance of the n least squares solutions (that correspond, as is well known, to the assumption that only one variable is affected by errors):

Theorem 3. When $\mathrm{Maxcor}_F(\Sigma_n) = 1$, the coefficients $\alpha_1, \ldots, \alpha_n$ of all linear relations compatible with the Frisch scheme lie (by normalizing one of the coefficients to 1) inside the simplex whose vertices are defined by the n least squares solutions (see Figure 2).

Other important properties of the loci of solutions in the noise and parameter spaces are described by the following theorems.

Theorem 4. When $\mathrm{Maxcor}_F(\Sigma_n) = 1$ the points of the simplex of solutions in the parameter space are linked by a one-to-one relation (isomorphism) to the points of $\mathcal{S}(\Sigma_n)$.

Theorem 5. (Schachermayer and Deistler [36]) When $\mathrm{Maxcor}_F(\Sigma_n) > 1$, $\mathcal{S}(\Sigma_n)$ is nonuniformly convex.

Theorem 6. (Guidorzi [21]) All points of $\mathcal{S}(\Sigma_n)$ where $\mathrm{corank}(\Sigma_n) = k$ $(k > 1)$ are accumulation points for those where $\mathrm{corank}(\Sigma_n) = k - 1$.

2.3.2. Computation of $\mathrm{Maxcor}_F(\Sigma_n)$

Despite its simple formulation, the problem of determining the maximal number of linear relations compatible with a set of noisy data in the context of the Frisch scheme remained unsolved for many years. One of the reasons is probably due to the focus of many researches on the locus of the solutions in the parameter space and to the practical impossibility of describing this locus, when $\mathrm{Maxcor}_F(\Sigma_n) > 1$, except than in elementary cases. An upper bound to $\mathrm{Maxcor}_F(\Sigma_n)$ has been given in [43]; geometric conditions to evaluate $\mathrm{Maxcor}_F(\Sigma_n)$ have, instead, been given in [21] on the basis of the analysis of the properties of the locus of noise space solutions.
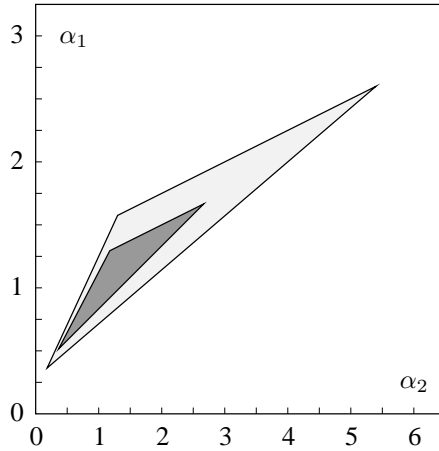
Fig. 2. Loci of admissible parameters for $n = 3$ and different amounts of noise.

Define, to this purpose, the singularity (hyper)surface $\mathcal{S}(\Sigma_{n/r})$ as the locus of the points $(\tilde{\sigma}_1^2, \ldots, \tilde{\sigma}_r^2) \in \mathbb{R}^r$ such that

$$\Sigma_n - \text{diag}\left[\tilde{\sigma}_1^2, \ldots, \tilde{\sigma}_r^2, 0, \ldots, 0\right] \geq 0 \tag{29}$$

and Σ_r as the sample covariance matrix of the first r variables. Then the following geometric relations hold:

Theorem 7. (Guidorzi and Stoian [28]) $\mathcal{S}(\Sigma_{n/r})$ lies always under or on $\mathcal{S}(\Sigma_r)$ (see Figure 3).

Theorem 8. (Guidorzi [21]) $\text{Maxcor}_F(\Sigma_n) \geq q$ if and only if $\mathcal{S}(\Sigma_{n-q+1}) \cap \mathcal{S}(\Sigma_{n/n-q+1}) \neq \{0\}$ for every subset of $n-q+1$ variables, i.e. for every permutation of the data leading to different subgroups in the first $n-q+1$ positions.

Theorem 8 allows the straightforward formulation of an algorithm for computing $\text{Maxcor}_F(\Sigma_n)$ by testing whether it is $\geq 2, 3, \ldots$ until the required conditions are no longer satisfied.

Remark 1. The existence of common points between different singularity hypersurfaces can be easily and efficiently verified by relying on the radial parametrization of these surfaces described in [26]. This parametrization will be recalled in the subsection concerning the identification of multivariable systems.
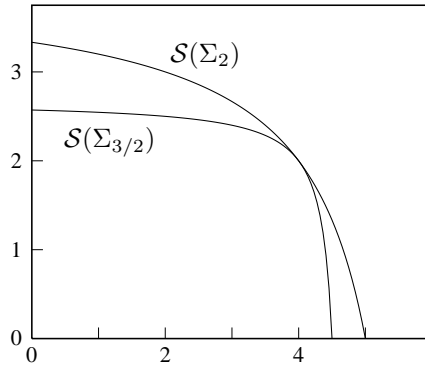
Fig. 3. Common points between $\mathcal{S}(\Sigma_2)$ and $\mathcal{S}(\Sigma_{3/2})$ in
a (3×3) covariance matrix with $\mathrm{Maxcor}_F(\Sigma_3) = 2$.

2.4. Computation of a single solution in the context of the Frisch scheme

The properties of the loci of solutions in the noise and parameter spaces show that, given a (sample) data covariance matrix Σ_n it is impossible to discriminate any solution (i.e. any decomposition of Σ) against any other, unless additional information (e.g. noise variance ratios) is available. It is however possible to take advantage of the differences between data belonging to two finite sequences and estimate, under some conditions, the linear relation actually linking the noiseless data. For this purpose we will introduce some abstract definitions and conditions that will eventually lead to algorithms appliable in real cases.

2.4.1. Complete sets of data for the Frisch scheme

The definitions and properties that follow concern the asymptotic case (infinite sequence of data).

Definition 2. (Guidorzi [19]) Two noise-free data covariance matrices of the same linear algebraic process, $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ are defined as *independent* if

$$\dim \ker \hat{\Sigma}_1 = \dim \ker \hat{\Sigma}_2 = \dim \ker \left(\hat{\Sigma}_1 - \hat{\Sigma}_2 \right) = 1. \tag{30}$$

Property 1. If $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ are independent there exists a unique (modulo scaling) vector A satisfying the conditions

$$\hat{\Sigma}_1 \, A = \hat{\Sigma}_2 \, A = \left(\hat{\Sigma}_1 - \hat{\Sigma}_2 \right) A = 0. \tag{31}$$

Definition 3. (Guidorzi [19]) Two noisy data covariance matrices of the same linear algebraic process, $\Sigma_1 > 0$ and $\Sigma_2 > 0$ are defined as *independent* if

$$\dim \ker \left(\Sigma_1 - \Sigma_2 \right) = 1. \tag{32}$$

Theorem 9. (Guidorzi [19]) Two independent noisy covariance matrices, Σ_1 and Σ_2, satisfy the following conditions under the Frisch scheme

$$\dim \ker \left(\Sigma_1 - \tilde{\Sigma} \right) = \dim \ker \left(\Sigma_2 - \tilde{\Sigma} \right) = 1 \tag{33}$$

$$\left(\Sigma_1 - \Sigma_2 \right) A = \left(\Sigma_1 - \tilde{\Sigma} \right) A = \left(\Sigma_2 - \tilde{\Sigma} \right) A = 0, \tag{34}$$

where $A = [\alpha_1\, \alpha_2\, \ldots\, \alpha_n]^T$ defines the process model (5) and $\tilde{\Sigma} \geq 0$ is a diagonal matrix satisfying the conditions

$$\Sigma_1 - \tilde{\Sigma} \geq 0, \quad \det \left(\Sigma_1 - \tilde{\Sigma} \right) = 0, \tag{35}$$

$$\Sigma_2 - \tilde{\Sigma} \geq 0, \quad \det \left(\Sigma_2 - \tilde{\Sigma} \right) = 0. \tag{36}$$

Theorem 10. (Guidorzi [19]) Among all points common to the hypersurfaces of admissible noise points associated with the independent noisy covariance matrices Σ_1 and Σ_2, one and only one point is mapped, according to Σ_1 and Σ_2, into the same point of the parameter space (see Figures 4 and 5).

Corollary 1. The Frisch scheme leads to a unique solution determined by every pair of independent noisy data covariance matrices of the process.

Corollary 2. Two independent noisy data covariance matrices of a process constitute a *complete* set of data for the Frisch scheme.
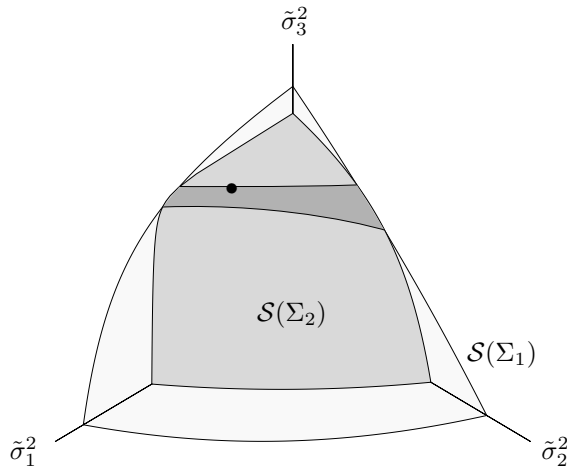

Fig. 4. Admissible noise points ($n = 3$).

2.4.2. Determination of the Frisch solution from real data

In all practical cases, even when two data sets are available, it is worthless testing whether they meet the independence conditions. Theorem 10, however, allows defining a consistent criterion to search for solutions even when the intersection between
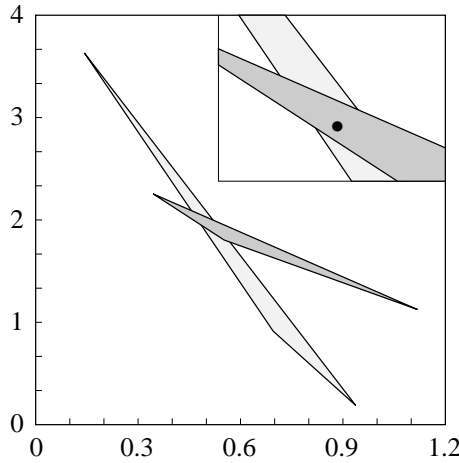
Fig. 5. Admissible model parameters $(n = 3)$.

$\mathcal{S}(\Sigma_1)$ and $\mathcal{S}(\Sigma_2)$ does not contain any point mapped, by Σ_1 and Σ_2 into the same point of the parameter space.

Criterion 1. (Guidorzi and Diversi [23]) Consider a pair of covariance matrices Σ_1 and Σ_2 and their loci of solutions, $\mathcal{S}(\Sigma_1)$, $\mathcal{S}(\Sigma_2)$ in the noise space. The best approximation of the actual noise variances will be given by the point $P \in \mathcal{S}(\Sigma_1) \cap \mathcal{S}(\Sigma_2)$ that minimizes the Euclidean norm of the distance between the parameter vectors A' and A'' associated to P by Σ_1 and Σ_2.

Remark 2. Criterion 1 is consistent since the cost function $f(P) = \|A' - A''\|_2$ annihilates when Σ_1 and Σ_2 are independent.

Remark 3. Once that the minimum of $f(P)$ has been found, two solutions, A' and A'' will be available and their distance is a measure of the reliability of the procedure. Their mean value can be taken as problem solution.

Remark 4. It can be observed that the outlined procedure can be applied even when the simplexes associated with Σ_1 and Σ_2 do not share common points.

Example 1. Two independent sets of noise-free data, concerning $N = 100$ observations of 3 variables are characterized by the sample covariance matrices

$$\hat{\Sigma}_1 \quad = \quad \frac{\hat{X}_1^T \hat{X}_1}{N} = \begin{bmatrix} 3 & 12 & -12 \\ 12 & 56 & -52 \\ -12 & -52 & 50 \end{bmatrix}$$

$$\hat{\Sigma}_2 \quad = \quad \frac{\hat{X}_2^T \hat{X}_2}{N} = \begin{bmatrix} 14 & 18 & -37 \\ 18 & 36 & -54 \\ -37 & -54 & 101 \end{bmatrix}.$$

$\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ have rank 2 and are associated with the same linear relation described by $\alpha_1 = 2$, $\alpha_2 = 0.5$ and $\alpha_3 = 1$. A Monte Carlo simulation of 100 runs has been performed by generating, in every run, two independent sets of three Gaussian white sequences and by adding these sequences to the noise-free data in order to obtain the noisy ones. The results are reported in Table 1.

Table 1. True and estimated values of the coefficients α_1 and α_2.

	α_1	α_2
true	2	0.5
estim. 1	2.0320 ± 0.1437	0.4945 ± 0.0610
estim. 2	2.0305 ± 0.1424	0.4802 ± 0.0724

3. THE FRISCH SCHEME IN THE DYNAMIC CASE

3.1. The SISO case

The extension of the Frisch scheme to the identification of dynamical processes can rely on some properties that, differently from the algebraic case, lead to a single solution also when a single sequence of data is available. To allow a simpler formulation of the problem, the SISO case will be firstly considered while the identification of MIMO systems will be treated only in a second time.

Consider a dynamic SISO system of order n described by the input-output model

$$\hat{y}(t+n) = \sum_{k=1}^{n} \alpha_k \, \hat{y}(t+k-1) + \sum_{k=1}^{n+1} \beta_k \, \hat{u}(t+k-1) \tag{37}$$

where $\hat{u}(t)$ denotes the input at time t and $\hat{y}(t)$ the output. Consider also noisy input/output observations, $u(t)$ and $y(t)$ given by

$$u(t) = \hat{u}(t) + \tilde{u}(t) \tag{38}$$
$$y(t) = \hat{y}(t) + \tilde{y}(t) \tag{39}$$

where $\tilde{u}(t)$ and $\tilde{y}(t)$ are white processes with zero mean, mutually uncorrelated and uncorrelated with $\hat{u}(t)$ (see Figure 6).

Define now the Hankel matrices

$$X_k(y) = \begin{bmatrix} y(1) & \dots & y(k) \\ y(2) & \dots & y(k+1) \\ \vdots & \ddots & \vdots \\ y(N) & \dots & y(k+N-1) \end{bmatrix}, \tag{40}$$

$$X_k(u) = \begin{bmatrix} u(1) & \dots & u(k) \\ u(2) & \dots & u(k+1) \\ \vdots & \ddots & \vdots \\ u(N) & \dots & u(k+N-1) \end{bmatrix}, \tag{41}$$
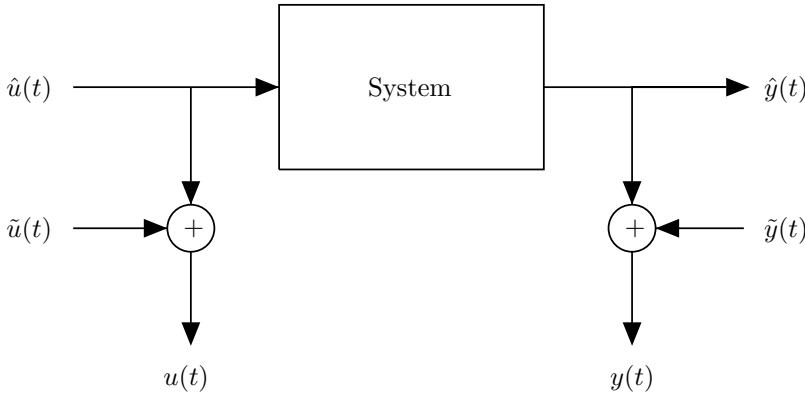
Fig. 6. The dynamic Frisch scheme context.

the matrix of input/output samples

$$X_k = \begin{bmatrix} X_{k+1}(y) \ X_{k+1}(u) \end{bmatrix} \tag{42}$$

and the sample covariance matrices Σ_k given by

$$\Sigma_k = \frac{X_k^T X_k}{N} = \begin{bmatrix} \Sigma(yy) & \Sigma(yu) \\ \Sigma(uy) & \Sigma(uu) \end{bmatrix}. \tag{43}$$

Denoting with $\tilde{\sigma}_u^{2*}$ and $\tilde{\sigma}_y^{2*}$ the variances of $\tilde{u}(t)$ and $\tilde{y}(t)$ and with P^* the point

$$P^* = \left(\tilde{\sigma}_y^{2*}, \tilde{\sigma}_u^{2*} \right), \tag{44}$$

the previous assumptions establish that, when $N \to \infty$

$$\Sigma_k = \hat{\Sigma}_k + \tilde{\Sigma}_k^* \tag{45}$$

where

$$\tilde{\Sigma}_k^* = \operatorname{diag}\left[\tilde{\sigma}_y^{2*} I_{k+1}, \tilde{\sigma}_u^{2*} I_{k+1} \right]. \tag{46}$$

The identification problem, in the context of the Frisch scheme, consists in determining the order and the parameters of model (37), or of any equivalent state-space model, and the additive noise variances $\tilde{\sigma}_y^{2*}$, $\tilde{\sigma}_u^{2*}$ on the basis of the knowledge of the noisy sequences $u(\cdot)$, $y(\cdot)$ or, equivalently, of the sequence of increasing-dimension matrices Σ_k for $k = 1, 2, \ldots$.

Model (37) implies, for every input sequence persistently exciting of order $n + 1$, the nonsingularity of $\hat{\Sigma}_1, \ldots, \hat{\Sigma}_{n-1}$ and the singularity of $\hat{\Sigma}_k$ for $k \geq n$. For any value of k (lower, equal or larger than n), a point $P = (\tilde{\sigma}_y^2, \tilde{\sigma}_u^2)$ belonging to the first orthant of the noise space, defines an admissible solution if and only if

$$\Sigma_k - \tilde{\Sigma}_k \geq 0 \tag{47}$$

$$\dim \ker\left(\Sigma_k - \tilde{\Sigma}_k\right) = 1, \tag{48}$$

where $\tilde{\Sigma}_k$ is the noise covariance matrix defined by P

$$\tilde{\Sigma}_k = \tilde{\Sigma}_k(P) = \mathrm{diag}\,[\tilde{\sigma}_y^2\, I_{k+1},\, \tilde{\sigma}_u^2\, I_{k+1}]. \tag{49}$$

The corresponding solution in the parameter space, $\theta(P) = [\alpha_1(P), \ldots, \alpha_n(P), -1,$ $\beta_1(P), \ldots, \beta_{n+1}(P)]^T$, is univocally defined by $\ker(\Sigma_k - \tilde{\Sigma}_k)$, i.e. by the relation

$$\hat{\Sigma}_k(P)\,\theta(P) = \big(\Sigma_k - \tilde{\Sigma}_k(P)\big)\,\theta(P) = 0. \tag{50}$$

Theorem 11. (Beghelli et al. [5]) For every $k > 0$ all admissible points define a convex curve $\mathcal{S}(\Sigma_k)$ in the first quadrant of the noise plane $\mathbb{R}^2$ with a concavity facing the origin. The point $P^* = (\tilde{\sigma}_y^{2*},\, \tilde{\sigma}_u^{2*})$ associated with the actual noise variances belongs to all curves $\mathcal{S}(\Sigma_k)$ when $k \geq n$ and $\theta(P^*)$ is the true parameter vector, θ^*.

Theorem 12. (Beghelli et al. [5]) If i and j are integers with $j > i$, then $\mathcal{S}(\Sigma_j)$ lies under or on $\mathcal{S}(\Sigma_i)$.

Remark 5. $\mathcal{S}(\Sigma_k)$ partitions the noise space $\mathbb{R}^2$ into the regions of the points σ^+ associated with positive definite matrices $\hat{\Sigma}_k = \Sigma_k - \tilde{\Sigma}^+$ and of the points σ^n associated with non definite and negative definite matrices $\hat{\Sigma}_k = \Sigma_k - \tilde{\Sigma}^-$. These regions lie under and over $\mathcal{S}(\Sigma_k)$ respectively.

Remark 6. Note that the dimension of the noise space is always equal to the total number of inputs and outputs (two for the SISO case) i.e. to the number of variables, like in the algebraic case. The dimension of the parameter space depends also on the order of the process.

Remark 7. Theorem 11 can be considered as a corollary of Theorem 8 since, because of the well-known shift property of dynamical systems, $\mathrm{Maxcor}_F\,\Sigma_k = k - n + 1$ when $k \geq n$.

Remark 8. (Diversi et al. [10]) The well-known Koopmans–Levin method, described in [34, 16], assumes that the ratio of the variances of the input and output noises is *a priori* known. It can be easily shown that this approach leads to a solution belonging to the set described by Theorem 11. Denote with η the ratio $\tilde{\sigma}_y^{2*}/\tilde{\sigma}_u^{2*}$ and define

$$\tilde{\Sigma}_n^* = \tilde{\sigma}_u^{2*}\,\tilde{\Sigma}_n^\eta, \quad \text{where} \quad \tilde{\Sigma}_n^\eta = \begin{bmatrix} \eta\,I_{n+1} & 0 \\ 0 & I_{n+1} \end{bmatrix}. \tag{51}$$

The knowledge of η implies that $\tilde{\Sigma}_n^*$ is known up to the scalar $\tilde{\sigma}_u^{2*}$. In the Koopmans–Levin method, the parameter vector θ is obtained by computing the minimal value of λ that satisfies the relation

$$(\Sigma_n - \lambda\,\tilde{\Sigma}_n^\eta)\,\theta = 0. \tag{52}$$

The solution is

$$\lambda^* = \min\,\mathrm{eig}\left(\Sigma_n\,(\tilde{\Sigma}_n^\eta)^{-1}\right), \tag{53}$$
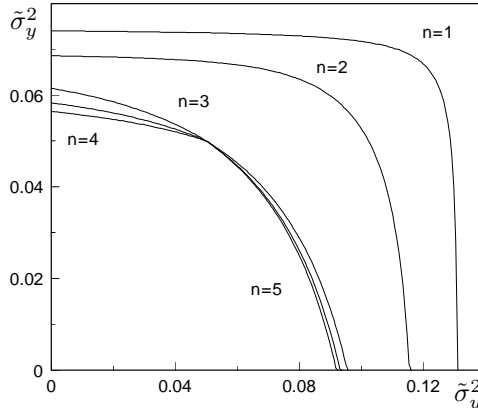
Fig. 7. Singularity curves for increasing-order models.

or, equivalently

$$\frac{1}{\lambda^*} = \max \text{ eig} \left(\Sigma_n^{-1} \, \tilde{\Sigma}_n^{\eta} \right). \tag{54}$$

Relation (54) is indeed preferable since it yields the solution also when $\tilde{\Sigma}_n^*$ is singular. Note that solutions (53) and (54) satisfy the conditions

$$\Sigma_n - \lambda^* \, \tilde{\Sigma}_n^{\eta} \geq 0, \qquad \dim \ker \left(\Sigma_n - \lambda^* \, \tilde{\Sigma}_n^{\eta} \right) = 1 \tag{55}$$

which correspond to the Frisch conditions (47) and (48), so that $\lambda^* = \tilde{\sigma}_u^{2*}$. It can also be noted that this solution coincides with the Total Least Squares one.

Example 2. Figure 7 shows the curves $\mathcal{S}(\Sigma_1), \ldots \ldots, \mathcal{S}(\Sigma_5)$ for data generated by the third order system

$$\hat{y}(t+3) = 0.4 \, \hat{y}(t+2) - 0.3 \, \hat{y}(t+1) - 0.1 \, \hat{y}(t) + 0.2 \, \hat{u}(t+2) - 0.38 \, \hat{u}(t+1) + 0.58 \, \hat{u}(t)$$

for input and output measures corrupted by white noises with variances $\sigma_u^{2*} = 0.05$ and $\sigma_y^{2*} = 0.05$. The geometric properties described by Theorems 11 and 12 can be easily observed.

3.2. Frisch identification of real processes and model selection criteria

The key property described by Theorem 11 holds only when the (asymptotic) properties assumed for the additive noise sequences (mutual orthogonality and orthogonality with the input/output sequences) hold, i.e. when $\tilde{u}(\cdot)$ and $\tilde{y}(\cdot)$ are uncorrelated
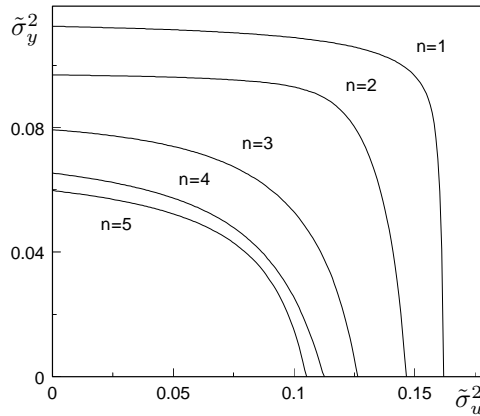
Fig. 8. Singularity curves for increasing-order models and real data.

white sequences with infinite length. In all other cases no common point between different curves can be observed. Similar consequences follow from violations on the linearity and time-invariance assumptions. Moreover, the algorithms that can be developed to estimate a single solution from real data can exhibit robustness and reliability problems that require the development of suitable criteria.

Example 3. The process considered is a natural gas reservoir converted to storage operations. The model orientation considers as input the total amount of injected/extracted gas and as output the mean reservoir pressure. The process exhibits a non stationary behavior because of the volume variations due to water encroaching. The actual amount of noise on the measures is modest, the sampling interval (1 month) is excessive with respect to the dynamics of the gas inside the porous rock but acceptable with respect to water dynamics. Figure 8 clearly shows the absence of common points between the different curves.

The number of criteria that can be developed is relatively large. Many of them, however, are not endowed with sufficient robustness degrees for real applications. As an example, it is possible to cite the selection criterion based on the minimal radial distance between two adjacent curves or, more generally, hypersurfaces. It is easy to show that such a criterion can select *any* point by properly scaling the data and insensitivity to data scaling is just *one* of the requirements for possible criteria.

The criteria described in the following exhibit a good robustness and are based on different properties.

3.2.1. The shifted relation criterion [4, 10]

This criterion is based on the following rank deficiency property of the matrices $\hat{\Sigma}_k(P^*)$ for $k \geq n$:

1. If $k \geq n$ the dimension of the null space of $\hat{\Sigma}_k(P^*)$ and, consequently, the multiplicity of its least eigenvalue, is equal to $(k - n + 1)$;

2. For $k > n$ all linear dependence relations between the columns of the matrices $\hat{\Sigma}_k(P^*)$ are described by the same set of coefficients θ^*.

When $k = n$, $\ker \hat{\Sigma}_n(P^*) = \operatorname{im} \theta^*$ while when $k = n + 1$

$$\ker \hat{\Sigma}_{n+1}(P^*) = \operatorname{im} [\,\overline{\theta}' \ \overline{\theta}''\,], \qquad (56)$$

where
$$\overline{\theta}' = [\,0 \ \alpha_1 \ \ldots \ \alpha_n - 1 \ 0 \ \beta_1 \ \ldots \ \beta_{n+1}\,]^T \qquad (57)$$

$$\overline{\theta}'' = [\,\alpha_1 \ \ldots \ \alpha_n - 1 \ 0 \ \beta_1 \ \ldots \ \beta_{n+1} \ 0\,]^T. \qquad (58)$$

Consider now the intersections $P' = (\tilde{\sigma}_y^{2'}, \tilde{\sigma}_u^{2'})$, $P'' = (\tilde{\sigma}_y^{2''}, \tilde{\sigma}_u^{2''})$ of a line from the origin with $\mathcal{S}(\Sigma_n)$ and $\mathcal{S}(\Sigma_{n+1})$, so that

$$\frac{\tilde{\sigma}_y^{2'}}{\tilde{\sigma}_u^{2'}} = \frac{\tilde{\sigma}_y^{2''}}{\tilde{\sigma}_u^{2''}}, \qquad (59)$$

and define the cost function

$$J(P', P'') = \operatorname{trace} \left([\,\overline{\theta}'(P') \ \overline{\theta}''(P')]^T \hat{\Sigma}_{n+1}(P'') \ [\overline{\theta}'(P') \ \overline{\theta}''(P')]\right), \qquad (60)$$

where $\overline{\theta}'(P'), \overline{\theta}''(P')$ have been constructed with the entries of $\theta(P')$. This function exhibits the following properties:

$$J(P', P'') \geq 0 \qquad (61)$$

$$J(P', P'') = 0 \Leftrightarrow P' = P'' = P^*. \qquad (62)$$

It is thus possible to perform the identification by searching, on $\mathcal{S}(\Sigma_n)$, for the solution that minimizes (60).

3.2.2. The covariance-matching criterion [9]

Consider the residual $\gamma(t)$ of the EIV process

$$\gamma(t) = \alpha_1 \, y(t) + \cdots + \alpha_n \, y(t + n - 1) - y(t + n) + \beta_1 \, u(t) + \cdots + \beta_{n+1} \, u(t + n) \quad (63)$$

that can also be written as

$$\gamma(t) = \alpha_1 \, \tilde{y}(t) + \cdots + \alpha_n \, \tilde{y}(t + n - 1) - \tilde{y}(t + n) + \beta_1 \, \tilde{u}(t) + \cdots + \beta_{n+1} \, \tilde{u}(t + n), \quad (64)$$

i.e., as the sum of two MA processes driven by the white noises $\tilde{y}(t)$ and $\tilde{u}(t)$. Because of the assumptions on $\tilde{y}(t)$ and $\tilde{u}(t)$, the autocorrelations of $\gamma(t)$, $r_\gamma(k) = \mathrm{E}\,[\gamma(t)\,\gamma(t - k)]$, are given by

$$r_\gamma(0) = \tilde{\sigma}_y^{2*} \sum_{i=1}^{n+1} \alpha_i^2 + \tilde{\sigma}_u^{2*} \sum_{i=1}^{n+1} \beta_i^2 \qquad (65)$$

$$r_\gamma(k) = \tilde{\sigma}_y^{2*} \sum_{i=1}^{n-k+1} \alpha_i \, \alpha_{i+k} + \tilde{\sigma}_u^{2*} \sum_{i=1}^{n-k+1} \beta_i \, \beta_{i+k} \quad \text{for } k = 1, \ldots, n \qquad (66)$$

$$r_\gamma(k) = 0 \quad \text{for } k > n \qquad (67)$$

where $\alpha_{n+1} = -1$. Define now, for every point $P = (\tilde{\sigma}_y^2, \tilde{\sigma}_u^2)$ of $\mathcal{S}(\Sigma_n)$ the vector

$$r_k(P) = [\, r_\gamma(0, P) \, r_\gamma(1, P) \, \dots \, r_\gamma(k, P)]^T, \tag{68}$$

with entries computed by means of $(65)-(67)$ using the variances $(\tilde{\sigma}_y^2, \tilde{\sigma}_u^2)$ and the parameters $\theta(P)$. Compute also, by means of the available data and $\theta(P)$, the sample vector

$$\bar{r}_k(P) = [\, \bar{r}_\gamma(0, P) \, \bar{r}_\gamma(1, P) \, \dots \, \bar{r}_\gamma(k, P)]^T, \tag{69}$$

where

$$\bar{r}_\gamma(k, P) = \frac{1}{N - k + 1} \sum_{t=n+k}^{N+n} \gamma(t)\, \gamma(t + k). \tag{70}$$

Since, for $N \to \infty$

$$r_k(P^*) = \bar{r}_k(P^*) = [\, r_\gamma(0) \, r_\gamma(1) \, \dots \, r_\gamma(k)]^T, \tag{71}$$

the following covariance-matching cost function can be considered

$$J(P) = \|r_k(P) - \bar{r}_k(P)\|_2\,, \tag{72}$$

where the theoretical statistical properties of $\gamma(t)$ are compared with those computed from the data.

The identification problem can thus be solved by minimizing $J(P)$ along $\mathcal{S}(\Sigma_n)$. The estimate accuracy of the covariance matching criterion has been analyzed in [38].

3.2.3. A criterion based on high-order Yule–Walker equations [13]

Define the regressor vector

$$\varphi(t) = [\, y(t - n) \, \dots \, y(t - 1) - y(t)\, u(t - n) \, \dots \, u(t)\,]^T \tag{73}$$

and the $q \times 1$ vector of delayed inputs

$$\varphi_u^h(t) = [\, u(t - n - q) \, \dots \, u(t - n - 1)]^T. \tag{74}$$

Consider then the $q \times (2n + 2)$ matrix

$$\Sigma^h = \mathrm{E}\,[\,\varphi_u^h(t)\, \varphi^T(t)]. \tag{75}$$

It is easy to show that

$$\Sigma^h = \mathrm{E}\,[\,\hat{\varphi}_u^h(t)\, \hat{\varphi}^T(t)], \tag{76}$$

where

$$\hat{\varphi}(t) = [\, \hat{y}(t - n) \, \dots \, \hat{y}(t - 1) - \hat{y}(t)\, \hat{u}(t - n) \, \dots \, \hat{u}(t)\,]^T \tag{77}$$

$$\hat{\varphi}_u^h(t) = [\, \hat{u}(t - n - q) \, \dots \, \hat{u}(t - n - 1)]^T. \tag{78}$$

Since $\hat{\varphi}^T(t)\, \theta^* = 0$ it follows that

$$\Sigma^h\, \theta^* = 0. \tag{79}$$

Relation (79) represents a set of high-order Yule–Walker equations that could be directly used to estimate the parameter vector θ^*. This approach can also be seen as an instrumental variable method that uses delayed inputs as instruments. The search for the point P^* on $\mathcal{S}(\Sigma_n)$ can be performed by means of the cost function

$$J(P) = \|\Sigma^h\,\theta(P)\|_2^2 = \theta^T(P)(\Sigma^h)^T\,\Sigma^h\,\theta(P) \tag{80}$$

that exhibits the following properties

$$J(P) \geq 0 \tag{81}$$

$$J(P^*) = 0. \tag{82}$$

The application of the described or of other possible criteria can be performed by minimizing their value on the singularity curve associated with the selected model order. This can be performed by using standard search algorithms and by selecting a suitable stop threshold. The efficiency of practical implementations can take great advantage from parametrizations of the curves that allow to perform the search by computing only a very limited number of points, like the radial parametrization described in the multivariable identification subsection.

3.3. The MIMO case

The extension of Frisch identification techniques to the MISO case is straightforward; this is not the case for MIMO processes that face conceptual and practical congruence problems not present in the single-output case.

3.3.1. The multivariable identification problem

The MIMO (purely) dynamic systems considered in this section are described by the input-output model

$$P(z)\,\hat{y}(t) = Q(z)\,\hat{u}(t), \tag{83}$$

where $\hat{u}(t) \in \mathbb{R}^r$, $\hat{y}(t) \in \mathbb{R}^m$ and $P(z)$, $Q(z)$ are $(m \times m)$ and $(m \times r)$ left coprime polynomial matrices in the unitary advance operator z. By selecting a minimal parametrization [18], model (83) can be partitioned into the set of m relations

$$\hat{y}_i(t+\nu_i) = \sum_{j=1}^{m}\sum_{k=1}^{\nu_{ij}} \alpha_{ijk}\,\hat{y}_j(t+k-1) + \sum_{j=1}^{r}\sum_{k=1}^{\nu_i} \beta_{ijk}\,\hat{u}_j(t+k-1) \tag{84}$$

where the integers ν_i $(i=1,\ldots,m)$ that appear in (84) and describe the structure of the model are the observability invariants of the system. The integers ν_{ij} are completely defined by these invariants through the relations

$$\begin{aligned}
\nu_{ij} &= \nu_i & \text{for } i = j \tag{85}\\
\nu_{ij} &= \min\,(\nu_i+1, \nu_j) & \text{for } i > j \tag{86}\\
\nu_{ij} &= \min\,(\nu_i, \nu_j) & \text{for } i < j\ . \tag{87}
\end{aligned}$$

For a complete description of the properties of the scalars $\{\nu_i, \alpha_{ijk}, \beta_{ijk}\}$ see [18]. The order of system (84) is given by

$$n = \sum_{i=1}^{m} \nu_i \; ; \tag{88}$$

ν_M will denote, in the following, the maximal observability index, i.e.

$$\nu_M = \max_i \{\nu_i, \; i = 1, \ldots, m\} \; . \tag{89}$$

In an errors-in-variables context, the noise-free signals $\hat{u}(t)$ and $\hat{y}(t)$ linked by model (84) are not directly accessible and only the noisy observations

$$u(t) = \hat{u}(t) + \tilde{u}(t) \tag{90}$$
$$y(t) = \hat{y}(t) + \tilde{y}(t) \; , \tag{91}$$

are available. In this paper the additive noises $\tilde{u}(t)$ and $\tilde{y}(t)$ satisfy the following assumptions.

1. The processes $\tilde{u}(t)$ and $\tilde{y}(t)$ are zero-mean, mutually uncorrelated white noise sequences, with unknown covariance matrices $\tilde{\Sigma}_u^* = \text{diag}\,[\tilde{\sigma}_{u_1}^{2*}, \ldots \tilde{\sigma}_{u_r}^{2*}]$ and $\tilde{\Sigma}_y^* = \text{diag}\,[\tilde{\sigma}_{y_1}^{2*}, \ldots \tilde{\sigma}_{y_m}^{2*}]$;

2. The processes $\tilde{u}(t)$ and $\tilde{y}(t)$ are uncorrelated with the the noise-free signal $\hat{u}(t)$.

The EIV MIMO identification problem can be stated as follows: given N noisy input-output observations $u(\cdot)$, $y(\cdot)$, estimate the noise covariance matrices $\tilde{\Sigma}_u^*$, $\tilde{\Sigma}_y^*$ and the coefficients α_{ijk}, β_{ijk} of model (84).

3.3.2. Properties of EIV MIMO systems

Consider the Hankel matrix

$$H_k(\hat{y}_i) = \begin{bmatrix} \hat{y}_i(1) & \ldots & \hat{y}_i(k) \\ \hat{y}_i(2) & \ldots & \hat{y}_i(k+1) \\ \vdots & & \vdots \\ \hat{y}_i(N) & \ldots & \hat{y}_i(k+N-1) \end{bmatrix} \; , \tag{92}$$

and the analogous matrices $H_k(y_i)$, $H_k(\tilde{y}_i)$, $H_k(\hat{u}_i)$, $H_k(u_i)$ and $H_k(\tilde{u}_i)$. Define also the multi-index $k^M = (k_1, \ldots, k_{m+r})$ and the matrix

$$\hat{H}(k^M) = \left[\, H_{k_1}(\hat{y}_1) \ldots H_{k_m}(\hat{y}_m)\, H_{k_{m+1}}(\hat{u}_1) \ldots H_{k_{m+r}}(\hat{u}_r) \,\right] .$$

Relations (84) can be used to write an overdetermined set of linear equations in the unknowns α_{ijk} and β_{ijk}. In fact, by considering the multi-index $\nu^M = (\nu_1 + 1, \ldots, \nu_m + 1, \nu_M, \ldots, \nu_M)$, relations (84) imply that

$$\hat{H}(\nu^M)\, \Theta = 0 \; , \tag{93}$$

where
$$\Theta = \begin{bmatrix} \theta_1 \ \theta_2 \ \cdots \ \theta_m \end{bmatrix} , \tag{94}$$

and
$$\theta_i = \begin{bmatrix} \alpha_{i11} \cdots \alpha_{i1\nu_{i1}} \ \underbrace{0 \cdots 0}_{(\nu_1 + 1 - \nu_{i1})} \ | \cdots | \end{bmatrix} \tag{95}$$

$$| \ \alpha_{ii1} \cdots \alpha_{ii\nu_i} \ -1 \ | \cdots | \ \alpha_{im1} \cdots \alpha_{im\nu_{im}} \ \underbrace{0 \cdots 0}_{(\nu_m + 1 - \nu_{im})} \ |$$

$$| \ \beta_{i11} \cdots \beta_{i1\nu_i} \ \underbrace{0 \cdots 0}_{(\nu_M - \nu_i)} \ | \cdots | \ \beta_{ir1} \cdots \beta_{ir\nu_i} \ \underbrace{0 \cdots 0}_{(\nu_M - \nu_i)} \ ]^T .$$

By defining the covariance matrix $\hat{\Sigma}(\nu^M)$ as

$$\hat{\Sigma}(\nu^M) = \frac{1}{N} \, \hat{H}(\nu^M)^T \hat{H}(\nu^M) = \begin{bmatrix} \hat{\Sigma}(\hat{y}_1 \hat{y}_1) & \hat{\Sigma}(\hat{y}_1 \hat{y}_2) & \ldots & \hat{\Sigma}(\hat{y}_1 \hat{u}_r) \\ \hat{\Sigma}(\hat{y}_2 \hat{y}_1) & \hat{\Sigma}(\hat{y}_2 \hat{y}_2) & \ldots & \hat{\Sigma}(\hat{y}_2 \hat{u}_r) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\Sigma}(\hat{u}_r \hat{y}_1) & \hat{\Sigma}(\hat{u}_r \hat{y}_2) & \ldots & \hat{\Sigma}(\hat{u}_r \hat{u}_r) \end{bmatrix}, \tag{96}$$

equation (93) implies that
$$\hat{\Sigma}(\nu^M) \, \Theta = 0 \ . \tag{97}$$

Define now the point P^* as

$$P^* = \left(\tilde{\sigma}_{y_1}^{2*}, \ldots \tilde{\sigma}_{y_m}^{2*}, \tilde{\sigma}_{u_1}^{2*}, \ldots \tilde{\sigma}_{u_r}^{2*} \right); \tag{98}$$

the assumptions of noise additivity and independence at the basis of the Frisch scheme lead, for $N \to \infty$, to the decomposition

$$\Sigma(\nu^M) = \hat{\Sigma}(\nu^M) + \tilde{\Sigma}^*(\nu^M), \tag{99}$$

where
$$\tilde{\Sigma}^*(\nu^M) = \text{diag} \left[\tilde{\sigma}_{y_1}^{2*} I_{\nu_1 + 1}, \ldots, \tilde{\sigma}_{y_m}^{2*} I_{\nu_m + 1}, \tilde{\sigma}_{u_1}^{2*} I_{\nu_M}, \ldots, \tilde{\sigma}_{u_r}^{2*} I_{\nu_M} \right]. \tag{100}$$

Consider now the generic subsystem i described by relation (84), the multi-index

$$\nu_i^M = (\nu_{i1}, \ldots, \nu_i + 1, \ldots, \nu_{im}, \nu_i, \ldots, \nu_i) \tag{101}$$

and the ith set of parameters

$$\eta_i = \begin{bmatrix} \alpha_{i11} \ldots \alpha_{ii1} \ldots \alpha_{ii\nu_i} \ -1 \ldots \alpha_{im\nu_{im}} \beta_{i11} \ldots \beta_{i1\nu_i} \ldots \beta_{ir1} \ldots \beta_{ir\nu_i} \end{bmatrix}^T ; \tag{102}$$

then
$$\hat{\Sigma}(\nu_i^M) \, \eta_i = \left[\Sigma(\nu_i^M) - \tilde{\Sigma}^*(\nu_i^M) \right] \eta_i = 0. \tag{103}$$

By defining the relation between multi-indices

$$k^M < h^M \quad \text{if} \quad k_i < h_i \quad \text{for} \quad i = 1, \ldots, m + r \, , \tag{104}$$

it is possible to state the following theorems whose proofs can be carried out along the lines considered in [5, 21] for the MISO case.

Theorem 13. For every structure $\xi = (\nu_1, \ldots, \nu_m)$, the admissible noise-space solutions associated with the ith subsystem, i.e. the locus of points $(\tilde{\sigma}_1^2, \ldots \tilde{\sigma}_{m+r}^2)$ such that

$$\hat{\Sigma}(\nu_i^M) = \Sigma(\nu_i^M) - \tilde{\Sigma}(\nu_i^M) \geq 0 , \tag{105}$$

is a convex hypersurface $\mathcal{S}(\Sigma(\nu_i^M))$ belonging to the first orthant of $\mathbb{R}^{m+r}$ whose concavity faces the origin (singularity hypersurface).

Theorem 14. *(the onion theorem)* If k_i^M and h_i^M are multi-indices with $h_i^M > k_i^M$, then $\mathcal{S}(\Sigma(h_i^M))$ lies under $\mathrm{S}(\Sigma(k_i^M))$.

Theorem 15. All hypersurfaces $\mathcal{S}(\Sigma(k_i^M))$, $(i = 1, \ldots, m)$ with $k_i^M > \nu_i^M$ have the single common point P^* corresponding to the actual variances of the noise on the data.

Theorems 13, 14 and 15 give a picture of the multivariable case similar to the pictures of the SISO and MISO cases. The existence of a single point (exact noise variances) common to the singularity hypersurfaces associated with the different subsystems, allows to solve the MIMO identification problem in a way similar to the SISO and MISO cases, by computing, in a congruent way, the parameters of every subsystem, defined by the kernels of the matrices $\hat{\Sigma}(\nu_i^M)$.

This, unfortunately, is no longer true in all real cases concerning limited sequences of data and/or real data that do not fulfill exactly the assumptions of the Frisch scheme. The solution of this problem [22, 27] has required the introduction of new parametrizations of the Frisch singularity surfaces that associate models to *all directions* in the noise space.

3.3.3. Radial parametrization for Frisch singularity hypersurfaces [26]

Radial parametrizations can be used effectively in both algebraic and dynamic cases for computing the points of $\mathcal{S}(\Sigma_n)$ and also to perform fast searches on $\mathcal{S}(\Sigma_n)$ to minimize a given cost function. It is important to note that such a minimization can be performed by computing only the points requested by the adopted search procedure. The same procedure can also be used for the direct computation of the distance between two singularity hypersurfaces along a given direction.

Radial parametrizations will be first described for the algebraic case and then extended to the dynamic one.

Let $\xi = (\xi_1, \ldots, \xi_n)$ be a generic point in the first orthant of $\mathbb{R}^n$; the intersection, $P = (\tilde{\sigma}_1^2, \ldots, \tilde{\sigma}_n^2)$, between the straight line through the origin and ξ with $\mathcal{S}(\Sigma_n)$ satisfies the conditions

$$\Sigma_n - \tilde{\Sigma}_n \geq 0, \qquad \dim \ker(\Sigma_n - \tilde{\Sigma}_n) = 1 \tag{106}$$

and

$$\lambda P = \xi \quad \text{with} \quad \lambda > 0. \tag{107}$$

It follows that

$$\det\left(\Sigma_n - \frac{1}{\lambda}\tilde{\Sigma}_n^{\xi}\right) = 0 \tag{108}$$

where

$$\tilde{\Sigma}_n^{\xi} = \mathrm{diag}\left[\,\xi_1,\ldots,\xi_n\,\right]. \tag{109}$$

Relation (108) is equivalent $(\Sigma_n > 0)$ to

$$\det\left(\lambda\,I - \Sigma_n^{-1}\tilde{\Sigma}_n^{\xi}\right) = 0 \tag{110}$$

so that the solution compatible with condition (106) is given by

$$P = \frac{\xi}{\lambda_M} \tag{111}$$

with

$$\lambda_M = \max\,\mathrm{eig}\left(\Sigma_n^{-1}\tilde{\Sigma}_n^{\xi}\right). \tag{112}$$

The points of $\mathcal{S}(\Sigma_n)$ associated with straight lines from the origin can thus be obtained by computing Σ_n^{-1} and the intersection between any line and $\mathcal{S}(\Sigma_n)$ by means of (111) and (112).

These results define a parametrization of singularity hypersurfaces that associates their points with the sheaf of lines from the origin in the first orthant. The parametrization of the singularity hypersurfaces associated with the dynamic case can be performed in the same way by considering the block-structure of $\tilde{\Sigma}_n$ associated with dynamical models.

Consider, for this purpose, the point ξ belonging to the first orthant of $\mathbb{R}^{m+r}$; the intersection between the straight line from the origin to ξ and $\mathcal{S}(\Sigma(\nu_i^M))$ is given by

$$P = \frac{\xi}{\lambda_M} \tag{113}$$

with

$$\lambda_M = \max\,\mathrm{eig}\left(\Sigma(\nu_i^M)^{-1}\,\tilde{\Sigma}^{\xi}(\nu_i^M)\right) \tag{114}$$

where

$$\tilde{\Sigma}^{\xi}(\nu_i^M) = \mathrm{diag}\left[\,\xi_1 I_{\nu_{i1}},\ldots,\xi_i I_{\nu_i+1},\ldots,\xi_m I_{\nu_{im}},\xi_{m+1}I_{\nu_i},\ldots,\xi_{m+r}I_{\nu_i}\,\right]. \tag{115}$$

3.3.4. EIV MIMO identification

The identification of real processes requires, as in the SISO or MISO cases, the definition of suitable selection criteria since the common point described by Theorem 15 will no longer exist. The criteria described for the SISO (or MISO) case, as well as others, could be applied to every subsystem and this would lead to a complete parametrization of the multivariable model.

It must however be noted that a procedure of this kind would lead to an incongruent solution because the identification of the different subsystems would lead to the estimation of (slightly) different points in the noise space and the corresponding

variances would be different and characterized by different ratios. As a consequence, the obtained model would be no longer associated with a single point in the noise space but with a set of points and this constitutes a serious limit to the congruence of the model and to some of its possible applications like, for instance, filtering.

The solution of this problem can be obtained by using radial parametrizations that allow to establish a one to one relation between *directions* in the noise space and model parameters and by suitable extensions of the selection criteria. To simplify the exposition we will consider, in the following, only the extension of the Shifted Relation Criterion.

The extension of this criterion to the multivariable case can be performed by considering the rank deficiency properties of the matrices $\hat{\Sigma}(\nu^M + \ell)$ ($\ell \geq 0$), where the compact notation $\nu^M + \ell$ stands for $(\nu_1 + 1 + \ell, \ldots, \nu_m + 1 + \ell, \nu_M + \ell, \ldots, \nu_M + \ell)$, that can be summarized as follows:

1. If $\ell \geq 0$ the dimension of the null space of $\hat{\Sigma}(\nu^M + \ell)$ and, consequently, the multiplicity of its (null) least eigenvalue, is equal to $(\ell + 1)m$;

2. For $\ell > 0$ all linear dependence relations between the vectors of the matrices $\hat{\Sigma}(\nu^M + \ell)$ can be described by the same sets of coefficients $\eta_1, \ldots, \eta_m$.

For example, when $\ell = 1$ it can be easily verified that

$$\ker\left[\hat{\Sigma}(\nu^M + 1)\right] = \operatorname{im} \bar{\Theta} = \operatorname{im}\left[\begin{array}{cccc} \bar{\theta}'_1 & \bar{\theta}''_1 & \cdots & \bar{\theta}'_m & \bar{\theta}''_m \end{array}\right], \tag{116}$$

where

$$\bar{\theta}'_i = \left[\begin{array}{cccccc} 0 & \alpha_{i11} & \cdots & \alpha_{i1\nu_{i1}} & \underbrace{0 \cdots 0}_{\nu_1 + 1 - \nu_{i1}} & | \cdots | \end{array}\right. \tag{117}$$

$$0\,\alpha_{ii1}\cdots\alpha_{ii\nu_i}-1 \,|\, \cdots \,|\, 0\,\alpha_{im1}\cdots\alpha_{im\nu_{im}}\underbrace{0\cdots0}_{\nu_m+1-\nu_{im}}\,|$$

$$\left.|\,0\,\beta_{i11}\cdots\beta_{i1\nu_i}\underbrace{0\cdots0}_{\nu_M-\nu_i}\,|\,\cdots\,|\,0\,\beta_{ir1}\cdots\beta_{ir\nu_i}\underbrace{0\cdots0}_{\nu_M-\nu_i}\right]^T$$

$$\bar{\theta}''_i = \left[\begin{array}{ccccc} \alpha_{i11} & \cdots & \alpha_{i1\nu_{i1}} & \underbrace{0\cdots0}_{\nu_1+1-\nu_{i1}} & 0 & | \cdots | \end{array}\right. \tag{118}$$

$$\alpha_{ii1}\cdots\alpha_{ii\nu_i}-1\,0\,|\,\cdots\,|\,\alpha_{im1}\cdots\alpha_{im\nu_{im}}\underbrace{0\cdots0}_{\nu_m+1-\nu_{im}}\,0\,|$$

$$\left.|\,\beta_{i11}\cdots\beta_{i1\nu_i}\underbrace{0\cdots0}_{\nu_M-\nu_i}\,0\,|\,\cdots\,|\,\beta_{ir1}\cdots\beta_{ir\nu_i}\underbrace{0\cdots0}_{\nu_M-\nu_i}\,0\right]^T.$$

It is thus possible to define, as an extension of (60), the cost function

$$f(P) = \left\| \bar{\Theta}^T \hat{\Sigma}(\nu^M + 1)\,\bar{\Theta} \right\|, \tag{119}$$

associated with the generic point $P = (\tilde{\sigma}_1^2, \ldots \tilde{\sigma}_{m+r}^2) \in \mathcal{S}(\Sigma(\nu^M + 1))$. By denoting with δ a generic direction, the cost function (119) can be thus defined as $f(\delta)$ and the intersections of the straight line from the origin with direction δ with the hypersurfaces $\mathcal{S}(\Sigma(\nu_i^M))$, $(i = 1, \ldots, m)$ allow estimating all subsystem parameters.

A selection criterion based on the minimization of this function is consistent because $f(\delta)$ annihilates, in the asymptotic case, only in the point associated with the actual noise variances.

In all other cases the identification procedure can be performed by using radial parametrizations that allow to establish a one to one relation between *directions* in the noise space and model parameters and by performing a search in order to minimize expression (119). The ratio of the noise variances is defined by the direction associated with the minimum of $f(\delta)$ while the values of the variances to be associated with the model are defined by the intersection of the selected line with $\mathcal{S}(\Sigma(\nu^M + 1))$. This assures the fulfillment of the condition $\hat{\Sigma}(\nu^M + 1) \geq 0$.

The robustness of the proposed procedure has been tested on sequences obtained by simulation from the following single-input two-output model

$$P(z) = \begin{bmatrix} z^2 - 0.4z + 0.3 & 0.1975 \\ -0.2026z + 0.1013 & z - 0.4 \end{bmatrix}$$

$$Q(z) = \begin{bmatrix} 0.3426z + 0.7194 \\ 0.7979 \end{bmatrix}.$$

The input sequence $\hat{u}(\cdot)$ is a PRBS normalized between -1 and 1 with unit variance and length $N = 300$. The variances of the noiseless output sequences $\hat{y}_1(\cdot)$, $\hat{y}_2(\cdot)$ are equal to 1. A Monte Carlo simulation of 100 independent runs has been performed by adding to the noise-free sequences different Gaussian white noise realizations with variances

$$\tilde{\sigma}_u^{2*} = 0.04 \qquad \tilde{\sigma}_{y_1}^{2*} = 0.16 \qquad \tilde{\sigma}_{y_2}^{2*} = 0.36 ,$$

which correspond to amounts of 20%, 40% and 60% in standard deviation. The identification results are summarized in Tables $2-4$ where the true value of parameters and noise variances, the means of their estimates and the corresponding standard deviations are reported.

These results show a good performance of the proposed approach also in presence of poor signal to noise ratios (SNR).

4. APPLICATIONS OF FRISCH IDENTIFICATION TECHNIQUES

4.1. Blind identification of SIMO FIR systems

The blind identification of dynamic systems is of great relevance in many fields like telecommunications, sismology, radioastronomy, etc. The purpose is the reconstruction of the transfer function of a transmission channel starting from noisy measurements performed only on its output [1, 40].

Blind identification relies on linear models describing a set of parallel channels driven by an unknown sequence and characterized by a finite impulse response (FIR). These models can describe a single unknown source in presence of multiple spatially and/or temporally distributed sensors. In the two-channel case the process is described (see Figure 9) by the model

Table 2. True and estimated parameters of $P(z)$.

	α_{111}	α_{112}	α_{121}	α_{211}	α_{212}	α_{221}
true	0.3000	-0.4000	0.19750	0.1013	-0.2026	-0.4000
ident.	0.2781 ± 0.08	-0.4246 ± 0.09	0.2363 ± 0.14	0.090 ± 0.10	-0.2165 ± 0.12	-0.3748 ± 0.16

Table 3. True and estimated parameters of $Q(z)$.

	β_{111}	β_{112}	β_{121}
true	0.7194	0.3426	0.7979
ident.	0.7167 ± 0.04	0.3423 ± 0.03	0.7997 ± 0.06

Table 4. True and estimated variances of $\tilde{u}(t), \tilde{y}_1(t), \tilde{y}_2(t)$.

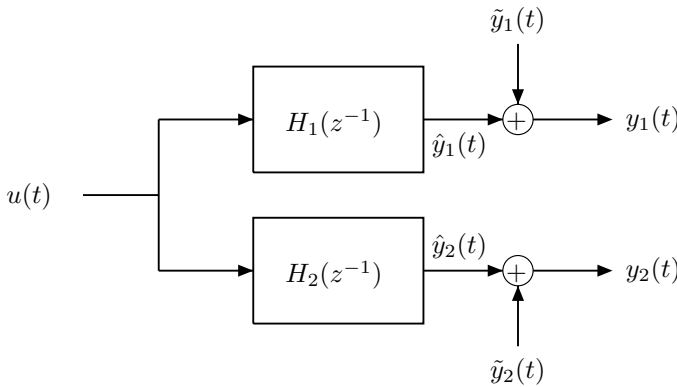|	$\tilde{\sigma}_u^*$	$\tilde{\sigma}_{y_1}^*$	$\tilde{\sigma}_{y_2}^*$
true	0.0400	0.1600	0.3600
ident.	0.0400 ± 0.02	0.1537 ± 0.02	0.3746 ± 0.04

Fig. 9. Two-channel FIR system.

$$\hat{y}_i(t) = H_i(z^{-1}) \, u(t) = \sum_{k=0}^{n} h_i(k) \, u(t-k), \quad i = 1, 2 \qquad (120)$$

$$H_i(z^{-1}) = h_i(0) + h_i(1) \, z^{-1} + \cdots + h_i(n) \, z^{-n}, \quad i = 1, 2 \qquad (121)$$

$$y_i(t) = \hat{y}_i(t) + \tilde{y}_i(t), \quad i = 1, 2 \qquad (122)$$

where $\tilde{y}_1(t)$ and $\tilde{y}_2(t)$ are mutually uncorrelated white noises, uncorrelated with $u(t)$ and with unknown variances $\tilde{\sigma}_{y_1}^{2*}, \tilde{\sigma}_{y_2}^{2*}$. Relations (120) lead immediately to the well known cross-relation property

$$H_2(z^{-1}) \, \hat{y}_1(t) = H_1(z^{-1}) \, \hat{y}_2(t). \qquad (123)$$

It is thus possible to write the relation

$$\left[X_{n+1}(\hat{y}_1) \ X_{n+1}(\hat{y}_2) \right] h = 0, \tag{124}$$

where

$$h = \left[h_2(n) \ \cdots \ h_2(0) \ -h_1(n) \ \cdots \ -h_1(0) \right]^T \tag{125}$$

and

$$X_{n+1}(\hat{y}_i) = \begin{bmatrix} \hat{y}_i(1) & \dots & \hat{y}_i(n+1) \\ \vdots & & \vdots \\ \hat{y}_i(N) & \dots & \hat{y}_i(N+n) \end{bmatrix}, \quad i = 1, 2. \tag{126}$$

Define now the covariance matrix

$$\hat{\Sigma}_n = \lim_{N \to \infty} \frac{1}{N} \left[X_{n+1}(\hat{y}_1) \ X_{n+1}(\hat{y}_2) \right]^T \left[X_{n+1}(\hat{y}_1) \ X_{n+1}(\hat{y}_2) \right]. \tag{127}$$

It follows that

$$\hat{\Sigma}_n \, h = 0 \tag{128}$$

$$\Sigma_n = \hat{\Sigma}_n + \tilde{\Sigma}_n^* \tag{129}$$

$$\tilde{\Sigma}_n^* = \text{diag} \left[\tilde{\sigma}_{y_1}^{2*} \, I_{n+1}, \ \tilde{\sigma}_{y_2}^{2*} \, I_{n+1} \right], \tag{130}$$

where Σ_n and $\tilde{\Sigma}_n^*$ can be obtained by inserting $X_{n+1}(y_i)$ and $X_{n+1}(\tilde{y}_i)$ in (127).

The blind identification problem has thus been mapped into an errors-in-variables identification problem that can be solved (in the case of two channels) by using the identification procedures described for the SISO case or with more specific procedures [11]. The multichannel case is more complex and cannot be reconducted to the MISO or MIMO cases; a procedure solving the blind multichannel identification problem is described in [24].

In both cases the proposed approaches extend the existing blind channel identification procedures to the case of unbalanced channel noises.

4.2. Identification of noisy autoregressive models

Autoregressive (AR) models are commonly used in a wide range of engineering applications, like spectral estimation, speech and image processing, noise cancellation etc.

A considerable attention has been dedicated, in the literature, to the problem of estimating AR models from signals corrupted by white noise. In this case the estimates obtained with classical AR identification methods (least-squares, Yule–Walker equations) are poor, particularly for low signal-to-noise ratio conditions [32, 33].

Consider the noisy AR model

$$x(t) = \alpha_1 \, x(t-1) + \cdots + \alpha_n \, x(t-n) + e(t), \tag{131}$$

$$y(t) = x(t) + w(t), \tag{132}$$

where $x(t)$ is the noise-free AR signal, $e(t)$ is the driving noise and $y(t)$ is the available observation affected by the additive noise $w(t)$; $e(t)$ and $w(t)$ are zero-mean white processes, mutually uncorrelated, with unknown variances σ_e^{2*} and σ_w^{2*}.

The AR+noise identification problem consists in estimating $\alpha_1, \ldots, \alpha_n$ and σ_e^{2*}, σ_w^{2*} starting from the available measurements $y(1), y(2), \ldots, y(N)$.

By defining the vectors

$$
\begin{aligned}
\varphi_x(t) &= \big[\, x(t-n) \ldots x(t-1)\ x(t) \,\big]^T, \\
\varphi_y(t) &= \big[\, y(t-n) \ldots y(t-1)\ y(t) \,\big]^T, \\
\varphi_w(t) &= \big[\, w(t-n) \ldots w(t-1)\ w(t) \,\big]^T,
\end{aligned}
$$

and the parameter vector

$$
\theta^* = \big[\, \alpha_n \cdots \alpha_1\ -1 \,\big]^T, \tag{133}
$$

it is possible to write model $(131)-(132)$ in the form

$$
\big(\varphi_x^T(t) - [0 \ldots 0\ e(t)]\big)\, \theta^* = 0 \tag{134}
$$

$$
\varphi_y(t) = \varphi_x(t) + \varphi_w(t). \tag{135}
$$

Define now the $(n+1) \times (n+1)$ covariance matrix

$$
\begin{aligned}
\hat{\Sigma}_n \;=\;& \mathrm{E}\,[\,\varphi_x(t)\,\varphi_x^T(t)\,] - \mathrm{diag}\,[\,\underbrace{0 \ldots 0}_{n}\ \sigma_e^{2*}\,] \\[2mm]
=\;& \begin{bmatrix}
r_x(0) & r_x(1) & \cdots & r_x(n) \\
r_x(1) & r_x(0) & \cdots & r_x(n-1) \\
\vdots & \vdots & \ddots & \vdots \\
r_x(n) & r_x(n-1) & \cdots & r_x(0) - \sigma_e^{2*}
\end{bmatrix},
\end{aligned}
$$

where $r_x(k) = r_x(-k) = \mathrm{E}\,[x(t)\,x(t-k)]$. From relation (134) it follows that

$$
\hat{\Sigma}_n\, \theta^* = 0. \tag{136}
$$

Because of (135), it follows that the covariance matrix of the noisy observations is given by

$$
\Sigma_n = \mathrm{E}\,[\,\varphi_y(t)\,\varphi_y^T(t)\,] = \hat{\Sigma}_n + \tilde{\Sigma}_n^* \tag{137}
$$

where

$$
\tilde{\Sigma}_n^* = \begin{bmatrix}
\sigma_w^{2*} & 0 & \cdots & \cdots & & 0 \\
0 & \sigma_w^{2*} & 0 & \cdots & & 0 \\
\vdots & & \ddots & & & \vdots \\
\vdots & & & \sigma_w^{2*} & & 0 \\
0 & \cdots & \cdots & 0 & & (\sigma_w^{2*} + \sigma_e^{2*})
\end{bmatrix} = \mathrm{diag}\,\big[\, \sigma_w^{2*}\, I_n, \sigma_s^{2*} \,\big] \tag{138}
$$

with $\sigma_s^{2*} = \sigma_w^{2*} + \sigma_e^{2*}$. Relation (137) shows that the identification of AR+noise models can be performed by means of Frisch procedures [12, 15] by introducing a

constraint on the last noise variance that must be $\geq \sigma_w^2$; this corresponds to the exclusion of the points of $\mathcal{S}(\Sigma_n)$ that do not satisfy this condition. It has also been shown that the particular structure of the covariance matrix of a noisy AR model allows to consider, as locus of compatible solutions, an interval instead of a curve [14]. This leads to a more efficient implementation of the identification algorithm.

Other applications of blind FIR identification and AR+noise procedures concern speech enhancement [6, 7].

5. CONCLUSIONS

This paper has presented an overview of several results concerning the properties of the Frisch scheme and its application to the estimation of linear relations from data affected by unknown amounts of additive noise and to the identification of dynamic processes in an Errors-in-Variables context. While it does not present new results, it integrates in an unitary view many results previously scattered in different works and underlines the links, not previously described, between the algebraic and dynamic contexts where the Frisch scheme can be applied.

REFERENCES

[1] K. Abed-Meraim, W. Qiu, and Y. Hua: Blind system identification. Proc. IEEE *85* (1997), 1310–1322.

[2] B. D. O. Anderson and M. Deistler: Identifiability of dynamic errors-in-variables models. J. Time Ser. Anal. *5* (1984), 1–13.

[3] B. D. O. Anderson, M. Deistler, and W. Scherrer: Solution set properties for static errors-in-variables problems. Automatica *32* (1996), 1031–1035.

[4] S. Beghelli, P. Castaldi, R. Guidorzi, and U. Soverini: A robust criterion for model selection in identification from noisy data. In: Proc. 9th International Conference on Systems Engineering, Las Vegas 1993, pp. 480–484.

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

[6] W. Bobillet, E. Grivel, R. Guidorzi, and M. Najim: Noisy speech de-reverberation as a SIMO system identification issue. In: Proc. IEEE Workshop on Statistical Signal Processing, Bordeaux 2005.

[7] W. Bobillet, R. Diversi, E. Grivel, R. Guidorzi, M. Najim, and U. Soverini: Speech enhancement combining optimal smoothing and errors-in-variables identification of noisy AR processes. IEEE Trans. Signal Process. *55* (2007), 5564–5578.

[8] M. Deistler: Linear errors-in-variables models. In: Time Series and Linear Systems (Lecture Notes in Control and Information Sciences; S. Bittanti, ed.). Springer–Verlag, Berlin 1986, pp. 37–67.

[9] R. Diversi, R. Guidorzi, and U. Soverini: A new criterion in EIV identification and filtering applications. In: Preprints 13th IFAC Symposium on System Identification, Rotterdam 2003, pp. 1993–1998.

[10] R. Diversi, R. Guidorzi, and U. Soverini: Frisch scheme-based algorithms for EIV identification. In: Proc. 12th IEEE Mediterranean Conference on Control and Automation, Kusadasi 2004.

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

[12] R. Diversi, R. Guidorzi, and U. Soverini: A noise-compensated estimation scheme for AR processes. In: Proc. 44th IEEE Conference on Decision and Control and 8th European Control Conference, Seville 2005, pp. 4146–4151.

[13] R. Diversi, R. Guidorzi, and U. Soverini: Yule–Walker equations in the Frisch scheme solution of errors-in-variables identification problems. In: Proc. 17th International Symposium on Mathematical Theory of Networks and Systems, Kyoto 2006, pp. 391–395.

[14] R. Diversi, R. Guidorzi, and U. Soverini: Identification of autoregressive models in the presence of additive noise. International J. Adaptive Control and Signal Process. *22* (2008), 465–481.

[15] R. Diversi, U. Soverini, and R. Guidorzi: A new estimation approach for AR models in presence of noise. In: Preprints 16th IFAC World Congress, Prague 2005.

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

[17] R. Frisch: Statistical Confluence Analysis by Means of Complete Regression Systems. Economic Institute, Pub. No. 5, Oslo University 1934.

[18] R. Guidorzi: Equivalence, invariance and dynamical system canonical modelling. Part I, Kybernetika *25* (1989), 233–257, Part II, Kybernetika *25* (1989), 386–407.

[19] R. Guidorzi: Certain models from uncertain data: the algebraic case. Systems Control Lett. *17* (1991), 415–424.

[20] R. Guidorzi: Errors-in-variables identification and model uniqueness. In: Statistical Modelling and Latent Variables (K. Haagen, D. J. Bartholomew, and M. Deistler, eds.), North Holland, Amsterdam 1993, pp. 127–150.

[21] R. Guidorzi: Identification of the maximal number of linear relations from noisy data. Systems Control Lett. *24* (1995), 159–166.

[22] R. Guidorzi: Identification of multivariable processes in the Frisch scheme context. MTNS'96, St. Louis 1996.

[23] R. Guidorzi and R. Diversi: Determination of linear relations from real data in the Frisch scheme context. In: Proc. 17th International Symposium on Mathematical Theory of Networks and Systems, Kyoto 2006, pp. 530–535.

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

[25] R. Guidorzi, R. Diversi, U. Soverini, and A. Valentini: A noise signature approach to fault detection and isolation. In: Proc. 16th International Symposium on Mathematical Theory of Networks and Systems, Leuven 2004.

[26] R. Guidorzi and M. Pierantoni: A new parametrization of Frisch scheme solutions. In: Proc. XII International Conference on Systems Science, Wroclaw 1995, pp. 114–120.

[27] R. Guidorzi, U. Soverini, and R. Diversi: Multivariable EIV identification. In: Proc. 10th IEEE Mediterranean Conference on Control and Automation, Lisboa 2002.

[28] R. Guidorzi and A. Stoian: On the computation of the maximal corank of a covariance matrix under the Frisch scheme. In: Proc. 10th IFAC Symposium on System Identification, Copenhagen 1994, pp. 171–173.

[29] R. E. Kalman: Identification from real data. In: Current Developments in the Interface: Economics, Econometrics, Mathematics (M. Hazewinkel, H. G. Rinnooy Kan, and D. Reidel, eds.), Dordrecht 1982, pp. 161–196.

[30] R. E. Kalman: Nine Lectures on Identification. (Lecture Notes on Economics and Mathematical Systems.) Springer–Verlag, Berlin (to appear).

[31] R. E. Kalman: System identification from noisy data. In: Dynamical Systems II (A. R. Bednarek and L. Cesari, eds.), Academic Press 1982, pp. 135–164.

[32] S. M. Kay: The effects of noise on the autoregressive spectral estimator. IEEE Trans. Acoustics, Speech and Signal Process. *27* (1979), 478–485.

[33] S. M. Kay: Noise compensation for autoregressive spectral estimates. IEEE Trans. Acoustics, Speech and Signal Process. *28* (1980), 292–303.

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

[35] E. Malinvaud: Méthodes statistiques de l'économétrie. Third edition. Dunod, Paris 1980.

[36] W. Schachermayer and M. Deistler: The set of observationally equivalent errors-in-variables models. Systems Control Lett. *34* (1998), 101–104.

[37] T. Söderström: Errors-in-variables methods in system identification. Automatica *43* (2007), 939–958.

[38] T. Söderström: Accuracy analysis of the Frisch estimates for identifying errors-in-variables systems. IEEE Trans. Automat. Control *52* (2007), 985–997.

[39] P. Stoica and A. Nehorai: On the uniqueness of prediction error models for systems with noisy input-output data. Automatica *23* (1987), 541–543.

[40] L. Tong and S. Perreau: Multichannel blind identification: from subspace to maximum likelihood methods. Proc. IEEE *86* (1998), 1951–1968.

[41] S. Van Huffel (ed.): Recent Advances in Total Least Squares Techniques and Errors-in-Variables Modelling. SIAM, Philadelphia 1997.

[42] S. Van Huffel and P. Lemmerling (eds.): Total Least Squares and Errors-in-Variables Modelling: Analysis, Algorithms and Applications. Kluwer Academic Publishers, Dordrecht 2002.

[43] K. G. Woodgate: An upper bound on the number of linear relations identified from noisy data by the Frisch scheme. Systems Control Lett. *24* (1995), 153–158.

Roberto Guidorzi, Roberto Diversi and Umberto Soverini, Dipartimento di Elettronica, Informatica e Sistemistica, University of Bologna, Viale del Risorgimento 2, 40136 Bologna. Italy.

e-mails: roberto.guidorzi@unibo.it, roberto.diversi@unibo.it, umberto.soverini@unibo.it