Recursive Algebraic Frisch Scheme: a Particle-Based Approach

Stefano Massaroli * **Renato Myagusuku** * **Federico Califano** **
Claudio Melchiorri ** **Atsushi Yamashita** * **Hajime Asama** *

* *Department of Precision Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan;*
email: {massaroli,miyagusuku,yamashita,asama}@robot.t.u-tokyo.ac.jp
** *Department of Electrical, Electronic and Information Engineering, University of Bologna, Viale Risorgimento 2, Bologna 40100, Italy;*
email: {federico.califano2,claudio.melchiorri}@unibo.it

Abstract: This paper presents a novel identification procedure. The proposed method consists in a recursive formulation of the algebraic Frisch scheme, which is an estimation procedure based on mild *a priori* assumptions in the context of error-in-variables (EIV) schemes. Simulations have been performed to show the validity of the new methodology.

Keywords: System identification, Recursive estimation, Linear systems, Frisch scheme.

1. INTRODUCTION

This work deals with the problem of extracting linear relations from a set of data affected by additive noise. Most of the algorithms proposed in the literature (e.g., *least mean squares*) seek a single solution for this problem. This uniqueness can be achieved only by introducing *a priori* assumptions in the estimation procedure whose validity cannot be ascertained on data. A complete analysis of different estimation schemes has been proposed by Kalman (1982), Söderström and Stoica (1989), Ljung (1998), Söderström (2007, 2012), Isermann and Münchhof (2014). Further analysis of the current state of the art of system identification can be found in (Ljung, 2010) while newer trends are presented in the survey paper (Pillonetto et al., 2014).

The most important feature of the *Algebraic Frisch Scheme* [1], originally introduced by Frisch (1934), is to employ loose assumptions leading to a whole family of solutions compatible with a given set of data rather than a unique solution. In this context, according to Kalman (1982), this estimation scheme has fewer *prejudices* with respect to most other identification procedures. The properties of this scheme and its *loci* of solutions have been deeply analyzed and developed in (Guidorzi, 1991); (Guidorzi, 1995); (Guidorzi et al., 2008). The Frisch scheme has been also extended (in the dynamic case) for the identification of multivariable EIV models (Diversi and Guidorzi, 2017). Under assumptions that will be discussed in the following sections, if only one underlying linear relation is present, the Frisch estimation scheme leads to a convex set of solutions represented by a simplex in an Euclidean space, where any point of the simplex represents a feasible solution in the context of the Frisch scheme. This peculiarity introduces a *post-identification* degree of freedom in the selection of the parameters among all the feasible solutions compatible with the hypothesis of the Frisch scheme. This aspect may represent a resource in engineering contexts since the knowledge of physical aspects of the system to be identified can help in this selection, mixing a pure black-box estimation step with a reasonable white-box selection procedure.

The main contribution of this paper is the introduction of a novel recursive estimation scheme based on the Frisch scheme, which is able to restrict the searching space of solutions with respect to the traditional batch approach without introducing additional assumptions. The proposed method is based on the idea of performing several identifications of the same process using partial data sets at different times to obtain a smaller final solution space. Using the proposed algorithm there is a consistent improvement of the estimation precision with respect to the standard Frisch scheme in only few iterations. This property makes the proposed algorithm suitable for real-time identification.

The *Particle-based Recursive Frisch* (PRF) scheme is presented and applied to a simple EIV model to prove its effectiveness. The PRF estimation scheme can be successfully used in a robust control framework as a robust controller could be designed in such a way to deal with model uncertainties ranging in the space of admissible solutions identified with the PRF scheme. This paper is organized as follows: Sec. 2 introduces some background on EIV schemes, with a particular insight on the Frisch scheme, enlightening the properties of its solutions spaces. In Sec. 3 the proposed PRF identification procedure is presented. Simulation experiments and results are shown in Sec. 4. In Sec. 5 conclusions and future work are drawn.

2. THE FRISCH ESTIMATION SCHEME: BACKGROUND

As every systematic procedure that deduces a model starting from data affected by errors, the Frisch Scheme

[1] In this paper we always deal with the *algebraic* Frisch scheme. This specification will be omitted from now on

relies on the *modification* of the observed data. This is mainly due to the fact that a deterministic mathematical relation able to fit all the observations does not exist. The main difference between different estimation schemes is the set of *a priori* assumptions that are necessary to estimate linear relations from noisy data. Different schemes use different assumptions and modify observations in different ways.

2.1 Common Assumptions for EIV Estimation Schemes

Consider the linear relation

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n = 0 \tag{1}$$

which relates n variables x_i by means of n coefficients α_i. If m measurements of the variables have been collected, the *observation matrix* can be defined as follows

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{pmatrix} \quad , \quad \mathbf{X} \in \mathbb{R}^{m \times n} \tag{2}$$

Then, equation (1) can be rewritten as

$$\mathbf{X}\mathbf{A} = 0 \tag{3}$$

where $\mathbf{A} = (\alpha_1 \ \alpha_2 \ \cdots \ \alpha_n)^T$. In general, $\mathbf{A} \in \mathbb{R}^{n \times q}$ where q is the number of independent linear relations linking the data. By defining the *sample covariance matrix* as $\mathbf{\Sigma} = \frac{\mathbf{X}^T\mathbf{X}}{m}$, (3) is equivalent to

$$\mathbf{\Sigma}\mathbf{A} = 0 \tag{4}$$

since $\mathbf{\Sigma}$ and $\mathbf{X}$ share the same null space. Assume now that noise affects the variables in an additive way, i.e., $x_i = \hat{x}_i + \tilde{x}_i$ where $\hat{x}_i$ is the exact value of the variable and $\tilde{x}_i$ is its corresponding noise. Due to this noise, $\mathbf{\Sigma}$ is positive definite and thus it is impossible to extract from its null space any linear relation, represented by matrix $\mathbf{A}$. This can only be achieved by modifying the matrix $\mathbf{\Sigma}$. Under the common assumptions of EIV schemes which relies in zero mean value of both noise and noiseless samples and in orthogonality between them, see (Guidorzi et al., 2008), it can be derived that $\mathbf{\Sigma} = \hat{\mathbf{\Sigma}} + \tilde{\mathbf{\Sigma}}$, for $\hat{\mathbf{\Sigma}} = \frac{\hat{\mathbf{X}}^T\hat{\mathbf{X}}}{m}$ and $\tilde{\mathbf{\Sigma}} = \frac{\tilde{\mathbf{X}}^T\tilde{\mathbf{X}}}{m}$, with $\hat{\mathbf{X}}$, $\tilde{\mathbf{X}}$ both defined as in (2) by replacing x_i with $\hat{x}_i$ and $\tilde{x}_i$. The problem of finding linear relations compatible with noisy data can be then formulated as follows:

EIV Estimation Problem (Kalman, 1982) Given $\mathbf{\Sigma}$, determine $\tilde{\mathbf{\Sigma}}$ such that

$$\hat{\mathbf{\Sigma}} = \mathbf{\Sigma} - \tilde{\mathbf{\Sigma}} \geq 0 \quad , \quad \det(\hat{\mathbf{\Sigma}}) = 0 \tag{5}$$

Any base of null($\hat{\mathbf{\Sigma}}$) will span a space describing a set of linear relations compatible with the data and the assumptions, i.e. a solution for $\hat{\mathbf{\Sigma}}\mathbf{A} = 0$ can be found.
Differences among estimation schemes lie in the assumption made on the noise variables and consequently on $\tilde{\mathbf{\Sigma}}$. Two estimation schemes of interest in this work are now described.

2.2 Ordinary Least Squares (OLS)

This estimation scheme assumes only one variable affected by noise, while the remaining ones are noise-free. If noise affects the i-th variable, it follows that

$$\tilde{\mathbf{\Sigma}} = \text{diag}\left(0, \cdots, 0, \tilde{\sigma}_i^2, 0, \cdots, 0\right) \tag{6}$$

where $\tilde{\sigma}_i^2$ is the variance of $\tilde{x}_i$. The solution of the estimation problem is given by

$$\tilde{\sigma}_i^2 = \frac{\det(\mathbf{\Sigma})}{\det(\mathbf{\Sigma}_i)} \tag{7}$$

where $\mathbf{\Sigma}_i$ is the matrix built by deleting the i-th row and the i-th column of $\mathbf{\Sigma}$, see (Guidorzi, 2003). In fact $\tilde{\sigma}_i^2$, as defined in (7), represents the maximum amount of noise compatible with the condition presented in (5). Some important aspects of this estimation scheme are:

- In general, the noisy variable is the one considered as *linearly dependent*, whose coefficient is normalized to 1.
- The OLS solutions lead to the minimization of the squared estimation error:
$$\mathbf{A} = \arg\min\left(\mathbf{e}^T\mathbf{e}\right) \tag{8}$$
where
$$\mathbf{e}^T\mathbf{e} = \| [\mathbf{X}]_i - \mathbf{X}_i\mathbf{A}\|^2 \tag{9}$$
being $[\mathbf{X}]_i$ the i-th column of $\mathbf{X}$ and $\mathbf{X}_i$ obtained by deleting and the i-th column of $\mathbf{X}$. The elements of $\mathbf{A} \in \mathbb{R}^{n-1}$ define the dependence relation of the orthogonal projection of $[\mathbf{X}]_i$ on the subspace spanned by the columns of $\mathbf{X}_i$ from these vectors. When no assumptions are made about the noisy variable, the OLS scheme leads to n different solutions.
- When the noisy variable is known to be the i-th, the estimation problem can be reformulated in the most common *non-kernel* form: $[\mathbf{X}]_i = \mathbf{X}_i\mathbf{A}$. In this case the minimization of the squared error is simply obtained in the pseudo-inverse closed form solution $\mathbf{A} = (\mathbf{X}_i^T\mathbf{X}_i)^{-1}\mathbf{X}_i^T[\mathbf{X}_i]$, which is unique.

2.3 The Frisch Scheme

This scheme is based on the assumption of mutual independence of the noise samples, i.e. the error covariance matrix is in the form

$$\tilde{\mathbf{\Sigma}} = \text{diag}\left(\tilde{\sigma}_1^2, \tilde{\sigma}_2^2, \cdots, \tilde{\sigma}_n^2\right) \tag{10}$$

Given a sample covariance matrix $\mathbf{\Sigma}$, every diagonal matrix $\tilde{\mathbf{\Sigma}}$ satisfying (5) is a solution of the Frisch estimation scheme. Notice that the n OLS solutions are included. This means that the a priori assumptions made for the Frisch scheme are milder and more general than those of any estimator providing a single solution like the OLS. On the other hand, the computation of solutions in closed-form is more difficult, since there are infinite possible solutions. To overcome this problem, we use some of the results reported in (Guidorzi et al., 2008), in particular, the following definition and theorem.

Definition 1. The *maximum corank* of a sample covariance matrix $\mathbf{\Sigma}$ is defined as the maximal number of linear relations that can be extracted from $\mathbf{\Sigma}$ under the assumptions of the Frisch scheme and it is denoted by $\text{Maxcor}_{\mathcal{F}}(\mathbf{\Sigma})$:

$$\text{Maxcor}_{\mathcal{F}}(\mathbf{\Sigma}) := \max_{\tilde{\mathbf{\Sigma}} \in \mathcal{D}} \left\{ \dim\left(\text{null}(\hat{\mathbf{\Sigma}})\right) \right\} \tag{11}$$

being $\mathcal{D}$ the set of all the diagonal matrices satisfying (5), i.e., all the solutions of the Frisch scheme.

Theorem 2. (Guidorzi et al., 2008) If $\text{Maxcor}_{\mathcal{F}}(\mathbf{\Sigma}) = 1$, the coefficient vectors $\mathbf{A} \in \mathbb{R}^{n-1}$ of all linear relations compatible with the Frisch scheme lie (by normalizing

one of the coefficients to 1) inside a simplex $\mathfrak{S}$ in the parameter space whose vertices are defined by the n OLS solutions. Furthermore, all the points of the simplex in the parameter space are linked by a one-to-one relation to *all* the solutions of the Frisch scheme.

Under the hypothesis of the latter theorem, the solution space (i.e., the simplex) can be easily derived by computing the n OLS solutions. However, in order to select one solution, additional information must be present (see e.g., (Guidorzi et al., 2008)).
The main contribution of this work is the development of an algorithm which aims at reducing the solution space. The decrease in size is intended with respect to the simplex that would merge by computing a single batch estimation by means of the classical Frisch scheme. The main advantage of this new method is the ability of shrinking the volume of the solutions set in the parameters space without introducing further assumptions. This volume can be considered indeed as a performance index of the estimation scheme. In the classical batch estimation scheme the volume of the simplex is proportional to the amount of noise affecting the system. If there is no noise, the simplex collapses to one point, i.e. the true parameters vector.

3. PARTICLE-BASED RECURSIVE METHOD

After the first m measurements, let us keep observing the linear system described by (1). At time instant t_k the state of the observed system is

$$\mathbf{x}(t_k) = (x_1(t_k)\ x_2(t_k)\ \cdots\ x_n(t_k)) \tag{12}$$

The *observation matrix* at the time instant t_k is updated as

$$\mathbf{X}(t_k) = \begin{pmatrix} \mathbf{x}(t_{k-m}) \\ \mathbf{x}(t_{k-m+1}) \\ \vdots \\ \mathbf{x}(t_k) \end{pmatrix} \in \mathbb{R}^{m\times n} \tag{13}$$

This update rule relies on the collection of m successive observations of the state that are stacked in a matrix. Note that any observation sample is used in one observation matrix only, maximizing the amount of new information contained in it. Therefore, a new estimation can be performed with a frequency m times slower than the one of state observation since m new samples are needed to build a new observation matrix. The *sample covariance matrix* at time t_k is:

$$\mathbf{\Sigma}(t_k) = \frac{\mathbf{X}(t_k)^T\mathbf{X}(t_k)}{m} \tag{14}$$

According to Theorem 2, if $\text{Maxcor}_{\mathcal{F}}(\mathbf{\Sigma}(t_k)) = 1$, the solutions of $\mathbf{\Sigma}(t_k)\mathbf{A} = 0$ lie in the simplex whose vertices $\mathbf{A}_i(t_k) \in \mathbb{R}^{n-1}$ are defined by the n OLS solutions in which one entry is normalized to 1 (or -1):

$$\mathbf{A}_i(t_k) : \text{span}(\mathbf{A}_i(t_k)) = \text{null}(\mathbf{\Sigma}(t_k) - \tilde{\mathbf{\Sigma}}_i(t_k)) \tag{15}$$

$$\text{where } \tilde{\mathbf{\Sigma}}_i(t_k) = \text{diag}(0,\cdots,0,\sigma_i^2(t_k),0,\cdots,0) \tag{16}$$

$$\text{and } \tilde{\sigma}_i^2(t_k) = \frac{\det(\mathbf{\Sigma}(t_k))}{\det(\mathbf{\Sigma}_i(t_k))} \tag{17}$$

being $\mathbf{\Sigma}_i(t_k)$ the matrix obtained by deleting the i-th row and the i-th column of $\mathbf{\Sigma}(t_k)$.
Without any loss of generality, let us suppose to normalize the last coefficient α_n to 1. Therefore, under the hypothesis of the Frisch scheme and if $\text{Maxcor}_{\mathcal{F}}(\mathbf{\Sigma}(t_k)) = 1$ the solution of the identification problem belongs the simplex $\mathfrak{S}(t_k) \subset \mathbb{R}^{n-1}$ whose vertices are defined as in (15) for any value of k.
Intuitively, if several simplices are computed at different iterations, it seems reasonable to seek the parameters value inside the intersection of all the simplices.
Although the Frisch scheme allows a simple derivation of the vertices of the simplices of solutions in the parameters space, the computation of their intersection from the knowledge of the vertices is not trivial. However, it is rather easy to check if a point of the parameters space belongs to a simplex, thanks to its convexity.
The *Particle-based* method consists in creating a set of static points in a region of the parameters space, called *particles*, each of which will be considered as a possible solution of the estimation problem. At each iteration, a new simplex is computed and the particles not belonging to the intersection of this simplex with the previous one are discarded. It turns out that this approach leads to a final solution space which is smaller with respect to the simplex that would merge when a single batch identification is performed with all the data. The working principle of the method is highlighted by the results shown in Sec. 4. In the following the procedure is described in detail.

3.1 Initialization of the Procedure

Let us define $\mathbf{X}(t_0)$ as

$$\mathbf{X}(t_0) = \begin{pmatrix} \mathbf{x}(t_0) \\ \mathbf{x}(t_1) \\ \vdots \\ \mathbf{x}(t_m) \end{pmatrix} \tag{18}$$

If we calculate the corresponding sample covariance matrix $\mathbf{\Sigma}(t_0) = \frac{\mathbf{X}(t_0)^T\mathbf{X}(t_0)}{m}$, the *initial simplex* $\mathfrak{S}(t_0)$ is then uniquely defined by its vertices which are computed as in (15). Being $l(t_0)$ the initial number of particles, the region of the parameter space bounded by the initial simplex is filled uniformly with the $l(t_0)$ particles $\mathbf{p}_j \in \mathbb{R}^{n-1}$, with the procedure described in Appendix A.
The initial *particle set* is defined as

$$\mathcal{P}(t_0) := \{\mathbf{p}_j : j = 1,2,\ldots,l(t_0)\} \tag{19}$$

3.2 Updating Rule of the Solutions Set

Being $\mathfrak{S}(t_k)$ the simplex corresponding to $\mathbf{\Sigma}(t_k)$ and $\mathcal{P}(t_k)$, $l(t_k)$ the particles set and the number of particles at the k-th time instant, in each iteration the algorithm is developed according to the *pseudo-code* reported in Algorithm 1.
$\mathfrak{S}(t_k)$ is initially derived from $\mathbf{\Sigma}(t_k)$ by means of (15). Then the updated particle set $\mathcal{P}(t_k)$ is assembled with the elements of $\mathcal{P}(t_{k-1})$ lying inside $\mathfrak{S}(t_k)$. Thus, $\mathcal{P}(t_k)$ has the following properties:

$$\mathcal{P}(t_k) = \{\mathbf{p}_j : \mathbf{p}_j \in \mathfrak{S}(t_h) \quad \forall h = 0,1,\ldots,k\} \tag{20}$$

$$= \left\{\mathbf{p}_j : \mathbf{p}_j \in \bigcap_{h=0}^{k} \mathfrak{S}(t_h)\right\} \tag{21}$$

From this simple idea it is possible to derive interesting consequences. Indeed, it represents a different way to use

Algorithm 1: PRF psudo-code

Data: $\mathcal{P}(t_{k-1})$, $l(t_k)$, $\mathbf{\Sigma}(t_k)$
Result: $\mathcal{P}(t_k)$
Compute $\mathfrak{S}(t_k)$;
$j = 1$;
while $j \leq l(t_k)$ **do**
  if $\mathbf{p}_j \in \mathfrak{S}(t_k)$ **then**
    $\mathbf{p}_j \in \mathcal{P}(t_k)$;
  else
    discard $\mathbf{p}_j$
  end
  $j = j + 1$
end

a set of data, in the contest of the Frisch scheme, as an alternative to performing a unique batch identification. The advantages of this approach, with respect to the standard case, are analyzed in the next section, also on the basis of the choice m, which represents the length of the partial data set used for a single identification step. Furthermore, thanks to its fast convergence time, this new scheme turns out to be suitable for online identification.

4. SIMULATION RESULTS AND DISCUSSION

4.1 Case of Study and Simulation Setup

To study the effectiveness of the proposed procedure the following linear model has been analyzed.

$$\alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 = 0 \tag{22}$$

where

$$\alpha_1 = -1.5 \quad \alpha_2 = 1.2 \quad \alpha_3 = 2 \quad \alpha_4 = 1 \tag{23}$$

A Monte Carlo Simulation of 100 runs has been performed in which x_1, x_2 and x_3 have been generated as three independent sequences of zero mean random numbers with unitary variance, while x_4 has been obtained as

$$x_4 = -(\alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3) \tag{24}$$

Due to the value of the variances of x_1, x_2, x_3 and the coefficients α_i $(i = 1, \ldots, 4)$ x_4 presents a standard deviation of about 2.77. Independent zero mean additive Gaussian noises has been successively added to all the variables, with variances $\tilde{\sigma}_1^2$, $\tilde{\sigma}_2^2$, $\tilde{\sigma}_3^2$ and $\tilde{\sigma}_4^2$. Therefore, the state of the system at any instant t_k can be described as follows:

$$\mathbf{x}(t_k) = (x_1(t_k)\ x_2(t_k)\ x_3(t_k)\ x_4(t_k)) \tag{25}$$

and the *observation matrix* has been built according to (13). Let m_{tot} be the number of samples in the dataset. The total number of PRF iterations is $N = (m_{tot}/m) - 1$ (excluding the initialization). The parameter m_{tot} has been set to be 10^7 and the initial number of particles 10^6. The initialization of $\mathbf{X}$ has been done using the first m samples. At each iteration the vertices of the simplex of solutions have been computed as described in Section 3 normalizing to 1 the last entry of each OLS solution. For comparison, the *total simplex* $\mathfrak{S}_{tot}$ has been computed with the whole dataset, i.e the simplex obtained with the classical Frisch estimation scheme using all the m_{tot} observations of the state. In order to evaluate the efficiency of the proposed scheme the following two performance indexes are defined:

- The *relative mean error* of the particles $e_{rm}(t_k)$, which is computed at each iteration as:

$$e_{rm}(t_k) = \frac{1}{l(t_k)\|\hat{\mathbf{A}}\|} \sum_{j=1}^{l(t_k)} \|\mathbf{p}_j(t_k) - \hat{\mathbf{A}}\| \tag{26}$$

  where $\hat{\mathbf{A}}$ is the vector containing true values of the parameters corresponding to the first three variables

$$\hat{\mathbf{A}} = (\alpha_1\ \alpha_2\ \alpha_3)^T = (-1.5\ 1.2\ 2)^T \tag{27}$$

- The *volume* V_{sol} of the *solution space*, namely, the volume of the convex polytope of the parameters space delimiting the particles [2].

The procedure was initially tested with $m = 2000$, $\tilde{\sigma}_1^2 = \tilde{\sigma}_2^2 = \tilde{\sigma}_3^2 = 0.16$ and $\tilde{\sigma}_4^2 = 1.23$. In Fig. 1 the convex hull

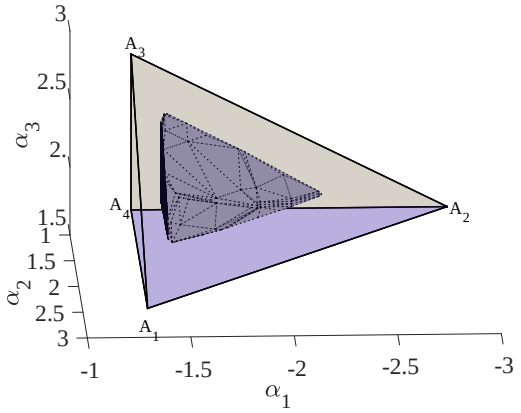

Fig. 1. Comparison between the total simplex defined by the vertices A_i computed as in (15) with the whole data set and the convex hull enclosing the particles calculated for $m = 2000$, $\tilde{\sigma}_1^2 = \tilde{\sigma}_2^2 = \tilde{\sigma}_3^2 = 0.16$ and $\tilde{\sigma}_4^2 = 1.23$

enclosing the particles remained at the end of the PRF identification is compared with the total simplex $\mathfrak{S}_{tot}$. The volume difference of the two solution spaces shows the improvement in the estimation precision given by the PRF. The convergence of the PRF can be observed in Fig. 2 where the time evolution of the relative mean error of the particles is presented. It can be noticed that e_{rm} rapidly decays with respect to the number of iterations reaching a nonzero constant value.

The volume V_{sol} of the solution space also decreases rapidly as it can be ascertained in Fig. 3. In this figure the time evolution of V_{sol} is compared with the one of V_k, the volume of the simplices obtained at different iterations and with the volume of the total simplex V_{tot} Results show that V_{sol} is halved after only ten iterations and in the worst case it becomes smaller than V_{tot} after three iterations. From the outcome of the simulations can be noticed that, among all the Monte Carlo iterations, V_{tot} changed of about 0.01% of its value [3], remarking the robustness of the standard Frisch scheme identification. It also turns out that V_{tot} corresponds always to the average of the volumes of the simplices obtained at different iterations.

[2] The geometrical properties of convex polytopes and useful methods to compute their volume can be found in (Henk et al., 2004)

[3] For this reason, in Fig. 3, only the average among the Monte Carlo runs is represented

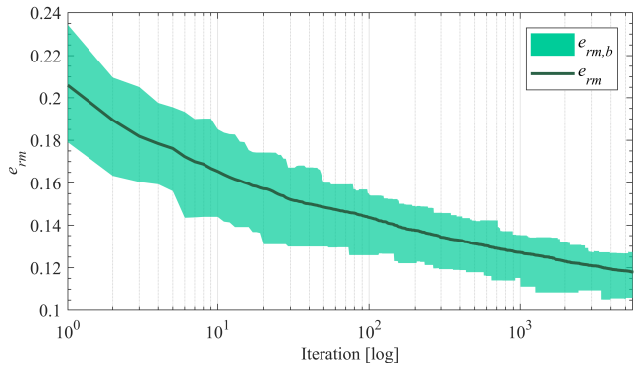

Fig. 2. Time evolution of relative mean error e_{rm} for $m = 2000$, $\tilde{\sigma}_1^2 = \tilde{\sigma}_2^2 = \tilde{\sigma}_3^2 = 0.16$ and $\tilde{\sigma}_4^2 = 1.23$. The solid line is the average of the values obtained in different Monte Carlo runs while the colored area ($e_{rm,b}$) is enclosed by the absolute minimum and maximum values of e_{rm} in the Monte Carlo runs

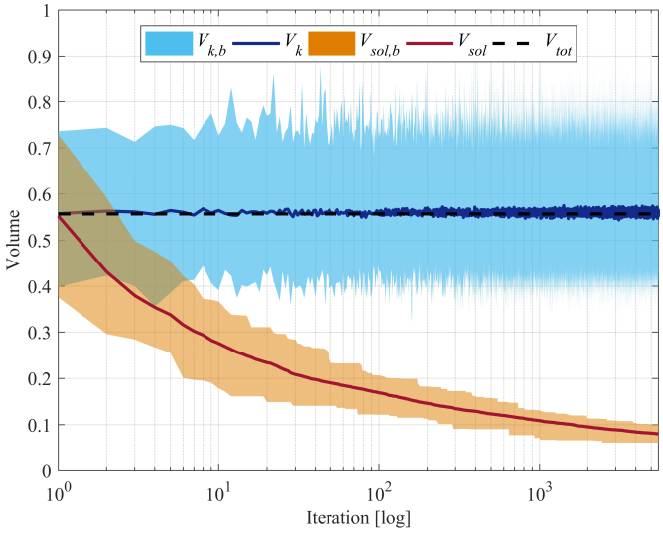

Fig. 3. Time evolution of the volume of the convex hull enclosing the particles V_{sol} compared with the volume of the simplices obtained at each iteration V_k and the volume of the total simplex V_{tot} for $m = 2000$, $\tilde{\sigma}_1^2 = \tilde{\sigma}_2^2 = \tilde{\sigma}_3^2 = 0.16$ and $\tilde{\sigma}_4^2 = 1.23$. The two solid lines are the average of V_{sol} and V_k obtained in different Monte Carlo runs, the dashed black line the average of V_{tot} while the colored areas ($V_{k,b}$ and $V_{sol,b}$) are enclosed by the absolute minimum and maximum values of V_{sol} and V_k in the Monte Carlo runs

4.2 Overall Performance Analysis

Several estimation tests have been performed assigning different values to the noise variances and to the length of the observation matrix. This allows to underline how the mutual choice of these variables affects the PRF estimation, showing up some important properties of this scheme. Therefore, different simulation have been performed in which the noise variances have been chosen as $\tilde{\sigma}_i^2 = (\delta \cdot \hat{\sigma}_i)^2$, for $i = 1, 2, 3, 4$ being $\hat{\sigma}_i$ the standard deviation of the variable x_i. The coefficient δ has been valued between 0.1 and 0.75 with steps of 0.05. In addition, the estimation procedure has been tested with multiple values of m chosen between 50 and 10000.

The results of the simulations tests show that the residual [4] relative mean error of the particles, $e_{rrm} = e_{rm}(t_N)$ and the residual relative volume of the solution space $V_{rr} = V_{sol}(t_N)/V_{tot}$ strongly depend on the mutual values of δ and m, i.e., $e_{rrm} = e_{rrm}(m, \delta)$, $V_{rr} = V_{rr}(m, \delta)$. In Fig. 4 e_{rrm} and V_{rr} are represented as function of δ and m while in Fig. 5 they are shown as function of the observation matrix length m for different values of δ.

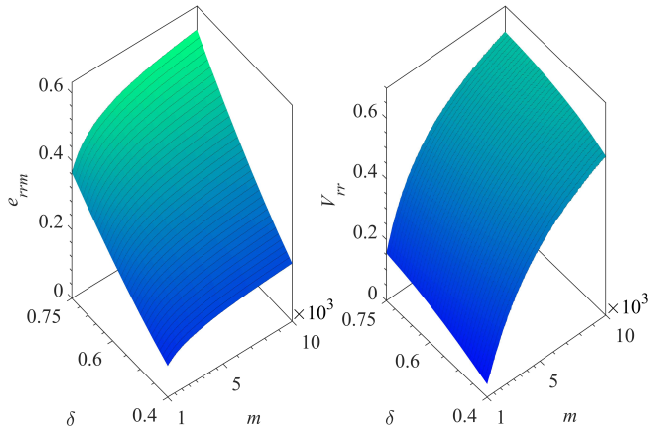

Fig. 4. As function of δ and m, respectively the ratio between noise and signal standard deviations and the length of the observation matrix, two quantities are represented. On the left, the residual relative mean error of the particles e_{rrm}. On the right, the ratio V_{rr} between the volume enclosing the particles set at the end of the PRF identification (V_{sol}) and the volume of the total simplex (V_{tot})

Residual Relative Mean Error

- Fig. 4 shows how e_{rrm} increases linearly with δ, independently from m.
- Fig. 5 shows how e_{rrm} reaches exponentially with m a constant $\xi(\delta)$ which is the relative mean error of the initial particles, having initialized the PRF with all the m_{tot} samples.
- If $l(t_0)$ is sufficiently large this value is very close to the relative mean error of the total simplex $e_{rm,tot}$, hence:

$$\xi(\delta) \approx e_{rm,tot} \tag{28}$$

where

$$e_{rm,tot} = \frac{1}{V_{tot}\|\hat{\mathbf{A}}\|} \int_{\mathfrak{S}_{tot}} \|\mathbf{p} - \hat{\mathbf{A}}\| \mathrm{d}\mathbf{p} \tag{29}$$

- If we suppose m_{tot} arbitrarily big, namely $m_{tot} \to \infty$, it yields

$$e_{rrm}(m, \delta) \approx \xi(\delta)(1 - e^{-m/\lambda}) \approx e_{rm,tot}(1 - e^{-m/\lambda})$$

- In our experimental case, the value of λ turns out to be approximatively 2000, independently from δ.

Residual Relative Volume

- Results shown in Fig. 5 exhibits that the PRF leads always to a solutions set whose volume V_{sol} is lower than the one obtained through a single identification with the whole data V_{tot} since V_{rr} is always smaller than one. Like e_{rrm}, V_{rr} shows an exponential trend

[4] The term *residual* refers to the value at the last PRF iteration

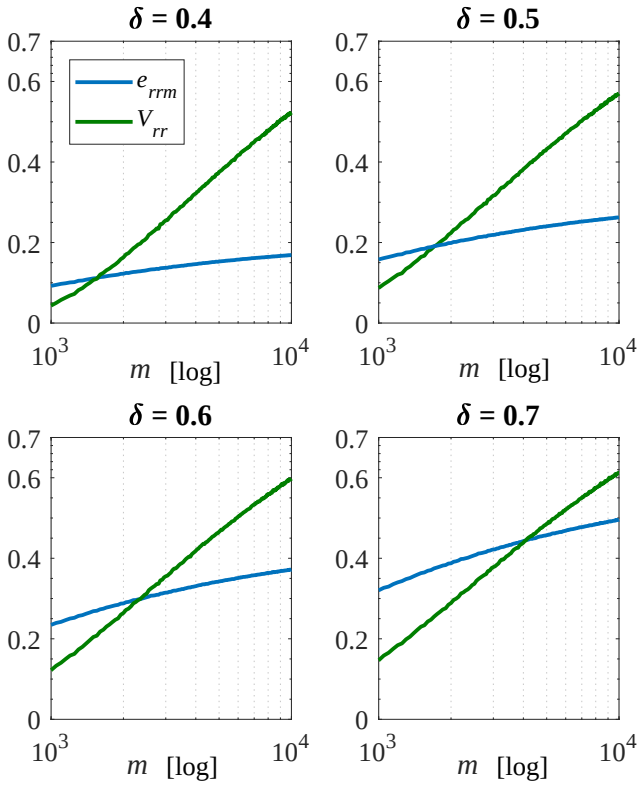

Fig. 5. For different values of δ (the ratio between noise and signal standard deviations), two significant data are here reported as function of the observation matrix length m (in log scale). With solid blue lines, the residual relative mean error of the particles e_{rrm}. With green lines the ratio V_{rr} between the volume enclosing the particles set at the end of the PRF identification (V_{sol}) and the volume of the total simplex (V_{tot})

with m and reaches a value close to 1 when m approaches m_{tot}.

- On the other hand, the growth rate of V_{rr} depends on δ. The dependency on δ is also exponential as it can be appreciated in Fig. 4.
- If we suppose m_{tot} arbitrarily big, we would have:

$$V_{rr}(m, \delta) \approx 1 - e^{-(\lambda_1 m + \lambda_2 \delta)}$$

After this analysis, it comes out that the performances of the PRF scheme strongly depend on two variables: the noise affecting the system and the number of subsets in which a given dataset is divided [5]. However, as system designers, we only have control on this latter variable since we cannot change the system noise rate.
Experimental results also shown, independently on the system noise, the optimal design of the identification algorithm is achieved by dividing the dataset in small partial subsets. In fact, the performance of the PRF degrades exponentially with m.
The results show the improvement of the performance of the PRF compared to a single batch approach for any value of noise variance and length of the observation matrix. Therefore, whenever a sufficiently large set of data is available it is always more convenient to perform the PRF scheme rather than use the whole data set for a single estimation.
However, in certain conditions, the PRF scheme may become unstable due to both numerical and statistical issues. In fact, although small values of m and δ should improve the overall performance of the PRF scheme, too small values may lead to failures of the procedure.
A robustness analysis follows in the next part.

[5] i.e., the number of iterations performed

4.3 Robustness Analysis of the PRF identification

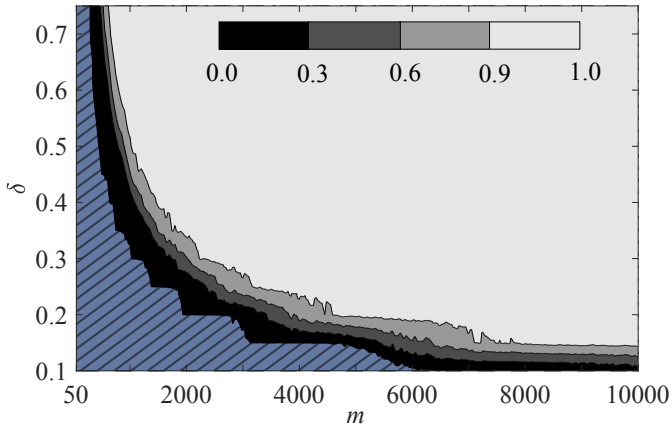

Fig. 6. Robustness map of the PRF scheme as function of m and δ. The blue area represents the unfeasible points where the algorithm failed in at least one Monte Carlo run. The gray-scale plot represents the probability of having the true parameters vector lying inside the convex hull enclosing the "survived" particles for different Monte Carlo runs

Throughout the simulations the feasibility of the PRF estimates has been constantly checked. Firstly, at each iteration the number of survived particles $l(t_k)$ has been monitored: if it dropped below one thousandth of the initial number the estimation has been considered unreliable. In fact excessive drop of particles might be symptom of ill conditioned estimates because a too small solution space should not be achieved considering the amount of noise affecting the system. Consequently, the simulation was aborted and the corresponding couple $\delta - m$ was marked as *failure* (the blue area). Then, at the end of each estimation it has been checked whether the true parameters was lying inside the PRF solutions space or not. If the true parameters vector was inside the hull the value of 1 has been assigned to the corresponding couple $\delta - m$ while, if it was outside, the value has been set to 0. Finally, the results of different Monte Carlo runs has been averaged. This data has been then used to build a map of to the robustness of the PRF procedure as function of δ and m, as in Fig. 6. When m is very small (i.e. less than 300) the PRF scheme always fails for any amount of noise. In fact, if m is too small, the observation matrix (i.e. the regressor) is not representative of the underlying system, leading to unreliable estimates. Furthermore, when also δ is small, this effect is accentuated by numerical issues (e.g. null space extraction from bad conditioned matrices, check of particles inclusion inside very narrow simplices). Therefore, the estimates result to be unreliable even for higher values of m. Note that this latter issue might occur also in the standard version of the Frisch scheme.
In order to find the optimal trade-off between performance

and reliability, m has to be chosen based on some rough estimates of δ. In addition, if the system noise in small, the above described numerical issues can be overcome by adding more independent additive noise in the system. This would stabilize the estimations without violating the assumptions of the Frisch scheme.

5. CONCLUSION

In this work a novel recursive identification procedure (PRF) based on the Algebraic Frisch Scheme has been presented.
The main difference with respect to the standard approach consists in using different subsets of data to perform several identifications of the same process instead of performing a batch identification step.
Simulations have been conducted to prove the effectiveness of the proposed method. Results shown the increased reliability of the developed methodology with respect to a single identification. Furthermore, this can be achieved without introducing additional assumptions on data.
A robustness and sensitivity analysis of the algorithm, with respect to the parameters used to implement it, has been performed through Monte Carlo simulations.
Future works will extend the proposed method to time varying systems assigning dynamic features to the particles, and will apply the developed algorithm in a robust control framework.

REFERENCES

Boyd, S., Vandenberghe, L., 2004. Convex optimization. Cambridge university press.

Diversi, R., Guidorzi, R., 2017. The frisch scheme in multivariable errors-in-variables identification. European Journal of Control 37, 43–53.

Frisch, R., 1934. Statistical confluence analysis by means of complete regression systems. Vol. 5. Universitetets Økonomiske Instituut.

Guidorzi, R., 1991. Certain models from uncertain data: the algebraic case. Systems & control letters 17 (6), 415–424.

Guidorzi, R., 1995. Identification of the maximal number of linear relations from noisy data. Systems & Control Letters 24 (3), 159–165.

Guidorzi, R., 2003. Multivariable system identification: from observations to models. Bononia University Press.

Guidorzi, R., Diversi, R., Soverini, U., 2008. The frisch scheme in algebraic and dynamic identification problems. Kybernetika 44 (5), 585–616.

Henk, M., Richter-Gebert, J., Ziegler, G. M., 2004. Basic properties of convex polytopes. Handbook of discrete and computational geometry, 255–382.

Isermann, R., Münchhof, M., 2014. Identification of Dynamic Systems: An Introduction with Applications. Springer Publishing Company, Incorporated.

Kalman, R. E., 1982. Identification from real data. In: Current developments in the interface: Economics, Econometrics, Mathematics. Springer, pp. 161–196.

Ljung, L., 1998. System identification. Signal analysis and prediction, 163–173.

Ljung, L., 2010. Perspectives on system identification. Annual Reviews in Control 34 (1), 1–12.

Pillonetto, G., Dinuzzo, F., Chen, T., Nicolao, G. D., Ljung, L., 2014. Kernel methods in system identification, machine learning and function estimation: A survey. Automatica 50 (3), 657–682.

Söderström, T., 2007. Errors-in-variables methods in system identification. Automatica 43 (6), 939–958.

Söderström, T., 2012. System identification for the errors-in-variables problem. Transactions of the Institute of Measurement and Control 34 (7), 780–792.

Söderström, T., Stoica, P., 1989. System identification.

Appendix A. INITIALIZATION OF THE PARTICLES

Given a matrix $\mathbf{M}$, $\mathbf{M}_{i,j}$ denotes the element of $\mathbf{M}$ corresponding to the i-th row and the j-th column Let $\mathfrak{S}(t_0)$ be the initial simplex. It is possible to define a matrix $\mathbf{S} \in \mathbb{R}^{(n-1)\times n}$ whose columns are the vertices $\mathbf{A}_j(t_0)$ of $\mathfrak{S}(t_0)$. Then, let $\mathbf{B}$ be computed as

$$\mathbf{B} = \begin{pmatrix} \min\limits_j\{\mathbf{S}_{1,j}\} & \max\limits_j\{\mathbf{S}_{1,j}\} \\ \min\limits_j\{\mathbf{S}_{2,j}\} & \max\limits_j\{\mathbf{S}_{2,j}\} \\ \vdots & \vdots \\ \min\limits_j\{\mathbf{S}_{(n-1),j}\} & \max\limits_j\{\mathbf{S}_{(n-1),j}\} \end{pmatrix} \tag{A.1}$$

A maximum number l_i of initial particles is chosen such that $l_i = a^{n-1}$, $a \in \mathbb{N}$. It is then possible to define a matrix $\mathbf{R} \in \mathbb{R}^{(n-1)\times a}$ whose elements $\mathbf{R}_{i,j}$ are as follow:

$$\mathbf{R}_{i,j} = \frac{a-j}{a-1}\mathbf{B}_{i,1} + \frac{j-1}{a-1}\mathbf{B}_{i,2} \tag{A.2}$$

for $i = 1, \ldots, n-1$, $j = 1, \ldots, a$.
The generation of the particles can be achieved by employing the algorithm shown in the pseudo-code Algorithm 2. To carry out the particles set initialization it is needed an algorithm able to determine if a point belongs to a convex hull (i.e., a simplex), see e.g., (Boyd and Vandenberghe, 2004).

Algorithm 2: Particles set initialization pseudo-code

```
Data: R, 𝔖(t_0)
Result: 𝒫(t_0)
𝒫(t_0) = {};
for c_1 ← 1 to a do
    for c_2 ← 1 to a do
        ⋮
        for c_{n-1} ← 1 to a do
            p = (R_{1,c_1}, R_{2,c_2}, ..., R_{n-1,c_{n-1}})
            if p ∈ 𝔖(t_0) then
                add p to 𝒫(t_0)
            end
        end
        ⋮
    end
end
```

This initialization procedure leads to a number of particles $l(t_0)$ that constitute $\mathcal{P}(t_0)$ which is less than l_i, $l(t_0) < l_i$. The diminution of $l(t_0)$ with respect to l_i is only dependent on the shape of the initial simplex $\mathfrak{S}(t_0)$, e.g., it may greatly decrease if some edges of the simplex are narrowed.