Abstract: The identification of dynamical models in the errors-in-variables (EIV) framework has seen renewed interest during the last decades. One of the main advantages of such a framework is the symmetrical treatment of all variables. One important EIV identification method is the Frisch scheme, which yields estimates for the measurement noise variances as well as the model parameters. Building on a recursive version of this scheme, which has recently been developed, this paper focuses on the need for and a method of incorporating adaptivity.

Keywords: Errors-in-variables, Exponential forgetting, Frisch scheme, Identification, Recursive estimation

1. INTRODUCTION

By considering noise disturbances on both inputs and outputs, the EIV framework encompasses the classical case in which only the output noise is taken into account (Söderström et al., 2002; Linden et al., 2006). An interesting approach for the identification of dynamical systems within this framework is the so-called Frisch scheme (Beghelli et al., 1990). It was originally used to treat a static algebraic regression problem (Frisch, 1934) without making any assumptions on the relative amount of noise on the variables. These rather loose constraints on the required a-priori knowledge yield a whole family of solutions. The extension of the Frisch scheme to dynamical multiple-input single-output (MISO) linear time-invariant (LTI) systems (Beghelli et al., 1990) leads theoretically to a single solution. In practice, however, a criterion to determine an optimal single solution has to be utilised. One approach, which is based on a comparison of the statistical properties of the residuals with those predicted from a model, was proposed in (Diversi et al., 2003). It has been further analysed in (Söderström, 2006b), where the accuracy of the estimation method with respect to the Cramer-Rao lower bound has been investigated. An extension to this method has been proposed in (Söderström, 2006a) for the case of correlated output noise.

In this paper, a recursive Frisch scheme identification approach, which has recently been developed in (Meyer et al., 2006) is extended to enhance its online applicability. It is shown that by incorporating adaptivity via the introduction of exponential forgetting, the algorithm is able to compensate for the systematic errors which arise in the original scheme (Meyer et al., 2006). Moreover, the adaptive recursive Frisch scheme is able to deal, to a certain degree, with linear time-varying systems.
2. THE EIV FRAMEWORK, ASSUMPTIONS FOR THE FRISCH SCHEME AND NOTATION

When a model of a system is created, it is often convenient to lump the uncertainties, e.g. due to model mismatch and measurement errors, and to introduce this as 'noise' at the system outputs. In practice, however, it is quite normal that both the system inputs and outputs are measured quantities and thus, it may be more realistic to assume that the measured input signal is also affected by noise. Such a setup defines the basis of the EIV approach, which is illustrated in Figure 1.

![Fig. 1. Errors-in-variables setup.](image)

In the single-input single-output (SISO) purely dynamic case, the noise-free input and output signal are denoted by $u_0(k)$ and $y_0(k)$, respectively. The input signal is assumed to be ergodic and stationary. Both signals are linked by the following difference equation

$$A(q^{-1})y_0(k) = B(q^{-1})u_0(k)$$

where

$$A(q^{-1}) = 1 + a_1 q^{-1} + ... + a_n q^{-n_a}$$  \hspace{0.5cm} (2a)

$$B(q^{-1}) = b_1 q^{-1} + ... + b_{n_b} q^{-n_b}$$  \hspace{0.5cm} (2b)

are polynomials in the backward shift operator $q^{-1}$, which is defined such that $x(k)q^{-1} = x(k-1)$ and $n_b \leq n_a$. The polynomials are assumed to describe an asymptotically stable system, which is observable and controllable. The parameter vector describing the linear relationship is given by

$$\theta_0 = [a_1 ... a_{n_a} b_1 ... b_{n_b}]^T$$

and its extended version is denoted

$$\bar{\theta}_0 = [1 \theta_0^T]^T$$

Hence, an alternative expression for (1) is given by

$$\bar{\varphi}_0^T(k)\bar{\theta}_0 = 0$$

where

$$\bar{\varphi}_0(k) = [-y_0(k) -y_0(k-1) ... -y_0(k-n_a)$$

$$u_0(k-1) ... u_0(k-n_b)]^T$$

is the extended regressor vector.

It is assumed that $u_0(k)$ and $y_0(k)$ are affected by measurement errors. The noisy signals can be decomposed into

$$u(k) = u_0(k) + \tilde{u}(k)$$  \hspace{0.5cm} (7a)

$$y(k) = y_0(k) + \tilde{y}(k)$$  \hspace{0.5cm} (7b)

where $\tilde{\cdot}$ denotes additive zero-mean white noise sequences with variances

$$E[\tilde{y}^2(k)] = \sigma_y, \quad E[\tilde{u}^2(k)] = \sigma_u$$

and $E[\cdot]$ is the expected value operator. The noise sequences $\tilde{u}(k)$ and $\tilde{y}(k)$ are assumed to be mutually uncorrelated and also uncorrelated with the noise free signal $u_0(k)$, hence $y_0(k)$.

The identification problem in the EIV framework can be defined as follows:

**Problem 1.** Given noisy input and output measurements of a dynamical system as well as $n_a$ and $n_b$, determine an estimate for $\sigma_u$, $\sigma_y$ and $\theta_0$, which minimises appropriate user defined cost criteria.

3. THE FRISCH SCHEME

The Frisch scheme provides estimates for $\sigma_u$, $\sigma_y$ and $\theta_0$ for a LTI dynamical system. Moreover, it can also be utilised to determine the polynomial orders $n_a$ and $n_b$ (cf. Beghelli et al., 1990). However, in this work the polynomial orders are assumed to be known in advance.

3.1 Problem statement

Instead of the expression in (5), one can consider

$$\Sigma_{\bar{\varphi}_0} \theta_0 = 0$$

where

$$\Sigma_{\bar{\varphi}_0} = E[\bar{\varphi}_0(k)\bar{\varphi}_0^T(k)] \in \mathbb{R}^{(n_a+n_b+1)\times(n_a+n_b+1)}$$

is the noise-free covariance matrix, which is singular positive semidefinite, with rank($\Sigma_{\bar{\varphi}_0}$) = $n_a+n_b$ (i.e. rank-one deficient).

For the noisy measurement case, one obtains the approximate relationship

$$\Sigma_{\bar{\varphi}} \theta_0 \approx 0$$

with

$$\Sigma_{\bar{\varphi}} = E[\bar{\varphi}(k)\bar{\varphi}^T(k)]$$

where $\bar{\varphi}(k)$ is the extended noisy regressor vector defined similarly to (6), with $u_0(k)$ and $y_0(k)$ replaced by $u(k)$ and $y(k)$, respectively. Following the stated assumptions, the covariance matrix $\Sigma_{\bar{\varphi}}$ can be decomposed into

$$\Sigma_{\bar{\varphi}} = \Sigma_{\bar{\varphi}_0} + \Sigma_{\bar{\varphi}}$$

where
\[ \Sigma_{\hat{\phi}} = \begin{bmatrix} \sigma_y I_{n_a+1} & 0 \\ 0 & \sigma_u I_{n_b} \end{bmatrix} \]  

is the noise covariance matrix. Note that in the noisy case, the covariance matrix \( \Sigma_{\hat{\phi}} \) is of full rank, i.e., positive definite. Moreover, it can be approximated by the sample covariance matrix

\[ \Sigma_{\hat{\phi}} \approx \hat{\Sigma}_{\phi}(N) = \frac{1}{N} \sum_{k=1}^{N} \varphi(k) \varphi^T(k) \]  

where \( N \) is the number of observations. The Frisch scheme identification problem can thus be stated as follows (Kalman, 1982):

**Problem 2.** Given a sample covariance matrix \( \hat{\Sigma}_{\phi} \) of noisy observations, determine the positive definite or semidefinite covariance matrices \( \hat{\Sigma}_{\phi} \) such that

\[ \hat{\Sigma}_{\phi_n} = \hat{\Sigma}_{\phi} - \hat{\Sigma}_{\phi} \geq 0 \]  

Thus, one aims to find suitable noise variances \( \hat{\sigma}_u \) and \( \hat{\sigma}_y \), such that \( \hat{\Sigma}_{\phi_n} \) is singular positive semidefinite, i.e., \( \hat{\Sigma}_{\phi_n} \) is rank-one deficient and the corresponding overdetermined system of equations

\[ \hat{\Sigma}_{\phi_n} \hat{\theta} = \hat{0} \]  

is solvable. Here, \( \hat{\theta} \) denotes the estimate of \( \theta_0 \).

### 3.2 Solution for the off-line problem

Once \( \hat{\sigma}_u \) and \( \hat{\sigma}_y \) are determined, the model parametrisation \( \hat{\theta} \) can be obtained by solving the last \( n_a + n_b \) equations of (17), which can be rewritten as

\[ \left( \hat{\Sigma}_{\phi} - \begin{bmatrix} \hat{\sigma}_u I_{n_a} & 0 \\ 0 & \hat{\sigma}_y I_{n_b} \end{bmatrix} \right) \hat{\theta} = -\xi \]  

where \( \hat{\Sigma}_{\phi} \) is obtained by deleting the first row and column of \( \hat{\Sigma}_{\phi} \), \( \xi \) is the first column of \( \hat{\Sigma}_{\phi_0} \) from row 2 to \( n_a + n_b + 1 \) and \( \hat{\theta} \) denotes the estimate of \( \theta_0 \). The relationship between \( \hat{\sigma}_u \) and \( \hat{\sigma}_y \) is given by applying the Schur complement (Beghelli et al., 1990)

\[ \hat{\sigma}_y(\hat{\sigma}_u) = \lambda_{\min} \left( \hat{\Sigma}_{\phi_n} - \hat{\Sigma}_{\phi_n}^{-1} \hat{\Sigma}_{\phi_n} \hat{\sigma}_u I_{n_b} \right) \]  

where \( \lambda_{\min}(\cdot) \) is the least eigenvalue operator and

\[ \hat{\Sigma}_{\phi} = \begin{bmatrix} \hat{\Sigma}_{\phi_n} & \hat{\Sigma}_{\phi_n} \hat{\sigma}_u \\ \hat{\Sigma}_{\phi_n} \hat{\sigma}_u & \hat{\Sigma}_{\phi_n} \hat{\sigma}_u \hat{\sigma}_y \end{bmatrix} \]  

with \( \hat{\Sigma}_{\phi_n} \in \mathbb{R}^{(n_a+1) \times (n_a+1)} \) and \( \hat{\Sigma}_{\phi_u} \in \mathbb{R}^{n_b \times n_b} \). Note that the estimated model parameters \( \hat{\theta} \) are completely characterised by \( \hat{\sigma}_u \) and \( \hat{\sigma}_y \) and since the latter is obtained via (19), the solution of Problems 1 and 2 reduces to that of determining a suitable value for \( \hat{\sigma}_u \). Several approaches have been provided in literature to obtain \( \hat{\sigma}_u \) (cf. Beghelli et al., 1990). The approach adopted here was proposed in (Diversi et al., 2003), with further analysis being made in (Söderström, 2006b). Essentially, it compares the statistical properties of the computed residuals with those predicted from a certain model.

Firstly, the residuals are given by

\[ \varepsilon(k, \hat{\theta}) = A(q^{-1}) y(k) - B(q^{-1}) u(k) \]  

and their sample auto-covariance can be determined by

\[ \hat{r}_\varepsilon(\kappa, N) = \frac{1}{N} \sum_{i=1}^{N} \varepsilon(l, \hat{\theta}) \varepsilon(l + \kappa, \hat{\theta}) \]  

where \( \kappa \) denotes a time shift. These are then compared with the theoretical auto-covariance elements

\[ \hat{r}_{\varepsilon_0}(\kappa) = E \left[ \varepsilon_0(l, \hat{\theta}) \varepsilon_0(l + \kappa, \hat{\theta}) \right] \]  

in which the theoretical residuals based on the unobserved noise sequences are given by

\[ \varepsilon_0(l, \hat{\theta}) = \hat{A}(q^{-1}) \hat{y}(l) - \hat{B}(q^{-1}) \hat{u}(l) \]  

Note that due to

\[ E \left[ \hat{y}^2(l) \right] = \hat{\sigma}_y, \quad E \left[ \hat{u}^2(l) \right] = \hat{\sigma}_u \]  

and the stated assumptions one obtains

\[ \hat{r}_{\varepsilon_0}(\kappa) = \hat{\sigma}_y \sum_{i=1}^{n_a-\kappa+1} \hat{a}_i - \hat{a}_{i+\kappa-1} + \hat{\sigma}_u \sum_{i=2}^{n_b-\kappa+2} \hat{b}_{i-1} - \hat{b}_{i+\kappa-1} \]  

with \( \hat{a}_0 = 1 \) and \( \kappa \leq n_a \). Recall that \( \hat{\theta} \) is completely determined by the input noise variance and thus Equations (22) and (23) only depend on \( \hat{\sigma}_u \) and the measured data. This allows the specification of a cost function

\[ V(\hat{\sigma}_u) = \delta^T W \delta \]  

with \( W \) being a positive definite diagonal weighting matrix and

\[ \delta = \begin{bmatrix} \hat{r}_\varepsilon(1) - \hat{r}_{\varepsilon_0}(1) \\ \vdots \\ \hat{r}_\varepsilon(n_d) - \hat{r}_{\varepsilon_0}(n_d) \end{bmatrix} \]  

where the value for \( n_d \leq n_u \), as well as \( W \), are chosen by the user. Thus, one can utilise an optimisation procedure to determine a value for \( \hat{\sigma}_u \) which minimises (27)
\[ \hat{\sigma}_u = \arg \min_{\sigma_u} V(\sigma_u) \quad (29) \]

This, in turn, uniquely defines \( \hat{\sigma}_y \) and \( \hat{\theta} \) yielding a solution for Problem 1.

The Frisch scheme can be summarised as follows:

Algorithm 1. (Batch procedure)
1. Choose \( n_\delta \) and \( W \) for a suitable cost function (27)
2. Apply an optimisation procedure to solve (29) by means of (18), (19), (22) and (26)
3. Determine \( \hat{\sigma}_y \) using (19)
4. Solve (18) by means of (15) to obtain \( \hat{\theta} \)

4. RECURSIVE FRISCH SCHEME

In many applications it is essential to obtain online estimates of the model parameters, while the process generating the data is running. One particular application is adaptive control, where the control action at time step \( k \) relies on a current estimate of the plant model, which is estimated using data up until \( k \). Typically, such a recursive estimation scheme must obey the following principles (Ljung, 1999): Firstly, the processing must with certainty be completed during one sample interval using a fixed and \( a\)-\( p\)-riori known amount of calculations. Secondly, the data, which is passed from one recursion step to the next, must be stored in a finite-dimensional information vector.

4.1 Fundamental on-line modifications

This subsection describes a recursive Frisch scheme approach as developed in (Meyer et al., 2006).

In order to modify Algorithm 1, the summations in Equations (15) and (22) are required to be updated recursively. Whilst the update of the covariance matrix is straightforward

\[ \hat{\Sigma}_\varphi(k) = \frac{k-1}{k} \hat{\Sigma}_\varphi(k-1) + \frac{1}{k} \hat{\varphi}(k)\hat{\varphi}^T(k) \quad (30) \]

the auto-covariance \( \hat{\epsilon}_\varphi(k, \kappa) \) cannot be computed recursively. This is because the computation of the sequence of residuals \( \{ \epsilon(l, \hat{\theta}) \}_{l=1}^k \), characterised by (21), requires the knowledge of the whole data sequences \( \{ u(l) \}, \{ y(l) \} \) for \( 1 \leq l \leq k \). However, it seems natural to approximate the residuals by

\[ \epsilon(l, \hat{\theta}(k)) \approx \epsilon(l, \hat{\theta}(l)) \quad \text{for} \ l < k \quad (31) \]

i.e. at time step \( k \) only the residual \( \epsilon(k, \hat{\theta}(k)) \) is calculated using the data stored in \( \hat{\varphi}(k) \) together with the current parameter estimate. This, in turn, allows an approximate recursive calculation of the auto-covariance

\[ \hat{\epsilon}_\varphi(k, \kappa) = \frac{k-1}{k} \hat{\epsilon}_\varphi(k, \kappa - 1) + \frac{1}{k} \epsilon(k, \hat{\theta}(k)) \epsilon(k + \kappa, \hat{\theta}(k)) \]

where only \( \epsilon(k + \kappa, \hat{\theta}(k)) \) for \( 1 \leq \kappa \leq n_\delta \) needs to be stored during each recursion step.

The procedure presented in (Meyer et al., 2006) can be summarised as followed:

Algorithm 2. (Recursive procedure)
1. Choose \( n_\delta \) and \( W \) for a suitable cost function (27)
2. Initialise \( \hat{\theta}(0), \hat{\Sigma}_\varphi(0) \) and \( \hat{\epsilon}_\varphi(0, \kappa) \) for \( \kappa = 1, \ldots, n_\delta \)
3. At each recursion step, apply (2)-(4) of Algorithm 1 replacing (15), (21) and (22) by (30), (31) and (32), respectively

In order to guarantee a fixed and \( a\)-\( p\)-riori known amount of calculations, the number of iterations during the optimisation at each recursion step is limited. The values for \( \hat{\theta}(0), \hat{\Sigma}_\varphi(0) \) and \( \hat{\epsilon}_\varphi(0, \kappa) \) can be initialised using Algorithm 1.

5. ADAPTIVITY

Recursive estimation algorithms facilitate the tracking of variation of system properties, i.e. to cope with slowly time-varying systems. Assuming some continuous property and slowly varying characteristics of the system, this can be achieved by reducing the importance of older observations while placing more emphasis on the more recent data and may be realised using a forgetting factor (Young, 1984; Ljung, 1999). This section describes how Algorithm 2 is modified in order to implement exponential forgetting.

The recursive algorithm requires to store the counter value \( k \) for the scaling in Equations (30) and (32) and this can lead to numerical difficulties for large values of \( k \). One benefit of using adaptivity within the recursive Frisch scheme is that the storage of \( k \) can be avoided. It may also reduce the error which arises due to the approximation of the residuals in Equation (31).

5.1 Exponential forgetting

Let \( \hat{H}_\varphi(k) \) and \( \hat{h}_\varphi(k, \kappa) \) denote the unscaled sample covariance matrix and auto-covariance, respectively, i.e.
In the case of no adaptivity, the update equations

\[ w(k) = \frac{1}{\lambda} \tilde{H}(k) \]  

(33)

\[ \tilde{r}(k, k) = \frac{1}{w(\lambda)} \tilde{h}(k, k) \]  

(34)

where \( w(\lambda) \) is a scaling factor which is equal to \( k \) in the case of no adaptivity. The update equations are then given by

\[ \tilde{H}(k) = \lambda \tilde{H}(k - 1) + \tilde{\varphi}(k) \tilde{\varphi}^T(k) \]  

(35)

\[ \tilde{h}(k, k) = \lambda \tilde{h}(k, k - 1) + \epsilon \left( k, \tilde{\theta}(k) \right) \epsilon \left( k + \kappa, \tilde{\theta}(k) \right) \]  

(36)

where \( 0 < \lambda < 1 \) is the forgetting factor. The scaling factor can hence be approximated by

\[ w(\lambda) \approx \lim_{k \to \infty} \sum_{i=1}^{k} \lambda^i = \frac{\lambda}{1 - \lambda} \]  

(37)

**Algorithm 3. (Adaptive procedure)**

The adaptive recursive Frisch scheme is essentially Algorithm 2, with (30) and (32) replaced by (33) and (34), respectively.

### 6. NUMERICAL SIMULATION

A second order linear time-varying SISO system described by

\[ G(q^{-1}, k) = \frac{b_1(k)q^{-1} + 0.5q^{-2}}{1.0 - 1.5q^{-1} + 0.7q^{-2}} \]

is simulated, where \( b_1(k) \) is a time varying parameter. The system is simulated for \( N = 10000 \) samples, the noise variances are given by \( \sigma_y = 0.5 \) and \( \sigma_a = 0.2 \) and the input signal is chosen to be a pseudo random binary sequence (PRBS) between \(-1\) and \(1\). The parameter \( b_1(k) \) is constant and equal to unity for the first thousand samples, while data for the initialisation of the recursive algorithm is collected. An initial estimate of \( \tilde{\theta}(1000) \) is obtained using the off-line version of the Frisch scheme (Algorithm 1). From \( k = 1001 \), \( b_1(k) \) is varied in a sinusoidal-like manner

\[ b_1(k) = \sin \left( \frac{2k\pi}{4000} \right) + 0.05 \sin \left( \frac{2k\pi}{1745} \right) + 1 \]  

(38)

and the recursive scheme is activated. The values for \( n_q \) and \( W \) are chosen to be \( 1 \). The simulation is run twice: first, Algorithm 2 with no adaptivity is used, then the adaptive scheme of Algorithm 3 proposed in Section 5, using a forgetting factor of \( \lambda = 0.995 \), is applied. The mean and variance of the estimated time-invariant parameters are given in Table 1. The mean square error (MSE) between the time varying parameter \( b_1(k) \) and its estimates for both cases is given in Table 2. The trajectories for the estimated noise variances and model parameters are shown in Figure 2.

It is observed, that the recursive Frisch scheme incorporating adaptivity is able to track the changes of the time-varying parameter \( b_1(k) \) (cf. Table 2 and Figure 2) while the estimate obtained by the recursive scheme with no forgetting tends to a constant value with an increasing number of samples. However, the variance of the estimates increases if adaptivity is used as seen in Table 1, which seems reasonable since less observations are effectively used in the adaptive scheme.

### 7. CONCLUSIONS

The recursive Frisch scheme has been extended to the adaptive case making use of exponential forgetting. This offers the possibility to apply the Frisch scheme online to estimate the parameters of linear time-varying systems. It also overcomes potential numerical difficulties with the existing recursive scheme. An illustrative example, in which the adaptive recursive scheme is compared with its non-adaptive counterpart, has been presented. The ability of the adaptive scheme to track changes in the system parameters has been demonstrated.

Further work concerns the reduction of computational load, which is rather high due to the need to invoke an optimisation procedure at each recursion. Regularisation may also be considered to reduce the variance of the estimates.

---

Table 1. Mean and variance for the estimates of the time-invariant parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>true value</th>
<th>no adaptivity</th>
<th>adaptivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma_y )</td>
<td>( 5 \cdot 10^{-1} )</td>
<td>5.31 ( \cdot 10^{-1} )</td>
<td>5.14 ( \cdot 10^{-1} )</td>
</tr>
<tr>
<td>( \sigma_a )</td>
<td>( 2 \cdot 10^{-1} )</td>
<td>2.03 ( \cdot 10^{-1} )</td>
<td>1.90 ( \cdot 10^{-1} )</td>
</tr>
<tr>
<td>( a_1 )</td>
<td>-1.5</td>
<td>-1.51</td>
<td>-1.51</td>
</tr>
<tr>
<td>( a_2 )</td>
<td>( 7 \cdot 10^{-1} )</td>
<td>7.09 ( \cdot 10^{-1} )</td>
<td>7.05 ( \cdot 10^{-1} )</td>
</tr>
<tr>
<td>( b_2 )</td>
<td>( 5 \cdot 10^{-1} )</td>
<td>4.87 ( \cdot 10^{-1} )</td>
<td>5.07 ( \cdot 10^{-1} )</td>
</tr>
</tbody>
</table>

Table 2. Mean square error between \( b_1(k) \) and its estimate.
Fig. 2. Estimated noise variances and model parameters. Solid: true values, feint: estimates of adaptive recursive scheme, dashed: estimates using no forgetting

REFERENCES


