Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Gradient-based approaches for recursive Frisch scheme identification

J. G. Linden * **B. Vinsonneau** * **K. J. Burnham** *

* *Control Theory and Applications Centre, Coventry University, UK, (email: j.linden@coventry.ac.uk)*

Abstract: An algorithm for recursive Frisch scheme system identification of linear single-input single-output errors-in-variables systems is developed. For the update of the estimated model parameters, a recursive bias-compensating least squares algorithm, which is based on the well-known recursive least squares technique, is considered. The estimate of the output measurement noise variance is determined using a conjugate gradient method, which tracks the smallest eigenvalue of a slowly varying matrix. For the update of the input measurement noise estimate, a steepest gradient search is applied. It tracks the minimum of a model selection cost function, which is based on a set of high order Yule-Walker equations.

Keywords: Errors in variables identification; Recursive identification; Estimation and filtering.

1. INTRODUCTION

Linear time-invariant (LTI) errors-in-variables (EIV) models are characterised by an exact linear relationship between inputs and outputs where both quantities are assumed to be corrupted by measurement noise (Söderström et al. [2002], Söderström [2007b]). One interesting approach for the identification of dynamical systems within this framework is the so-called Frisch scheme (Beghelli et al. [1990]), which yields estimates of the model parameters as well as the measurement noise variances. Whilst the Frisch scheme leads theoretically to a single solution, a model selection criterion is required to be utilised in practice to determine an 'optimal' single solution from a set of admissible Frisch solutions. For this purpose, a covariance matching criterion, which is based on a comparison of the statistical properties of the residuals with those predicted from a model, has been proposed (Diversi et al. [2003]). An attempt to develop a recursive algorithm based on this criterion, which simply applies the offline Frisch equations to approximately updated covariance matrices, has been considered in Linden et al. [2006]. The major drawback of this procedure is the computational burden as well as a systematic error, which is introduced due to the approximation of the residuals. Recently, an alternative model selection criterion based on a set of high order Yule-Walker (YW) equations has been presented (Diversi et al. [2006]). This criterion seems more appropriate for the development of a recursive Frisch scheme algorithm, since no approximations are necessary for the update of the covariance matrices.

This paper develops a recursive Frisch scheme identification algorithm based on the YW criterion. Fast algorithms within this framework are discussed in Linden et al. [2007], whereas the coloured output noise case is considered in Linden and Burnham [2008].

The paper is organised as follows. Section 2 presents the EIV identification problem and introduces some notational conventions. Section 3 reviews the offline Frisch scheme. The new recursive approach is presented in Section 4, followed by a numerical example in Section 5. Conclusions are given in Section 6.

2. PROBLEM STATEMENT AND NOTATION

In this paper, a discrete-time, LTI single-input single-output (SISO) EIV system is considered, which is described by

$$A(q^{-1})y_{0_i} = B(q^{-1})u_{0_i}, \tag{1}$$

where i is an integer valued time index and

$$A(q^{-1}) \triangleq 1 + a_1 q^{-1} + ... + a_{n_a} q^{-n_a}, \tag{2a}$$

$$B(q^{-1}) \triangleq b_1 q^{-1} + ... + b_{n_b} q^{-n_b} \tag{2b}$$

are polynomials in the backward shift operator q^{-1}, which is defined such that $x_i q^{-1} = x_{i-1}$. The noise-free input u_{0_i} and output y_{0_i} are unknown and only the measurements

$$u_i = u_{0_i} + \tilde{u}_i, \tag{3a}$$

$$y_i = y_{0_i} + \tilde{y}_i \tag{3b}$$

are available, where $\tilde{u}_i$ and $\tilde{y}_i$ denote input and output measurement noise, respectively. Such a setup is depicted in Figure 1. The following assumptions are introduced:

A1 The dynamic system (1) is asymptotically stable, i.e. $A(q^{-1})$ has all zeros inside the unit circle.
A2 All system modes are observable and controllable, i.e. $A(q^{-1})$ and $B(q^{-1})$ have no common factors.
A3 The polynomial degrees n_a and n_b are known *a priori* with $n_b \leq n_a$.

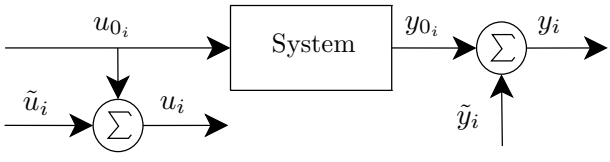

Fig. 1. Errors-in-variables setup.

A4 The true input u_{0_i} is a zero-mean ergodic process and is persistently exciting of sufficiently high order.
A5 The sequences $\tilde{u}_i$ and $\tilde{y}_i$ are zero-mean, ergodic, white noises with unknown variances $\sigma_{\tilde{u}}$ and $\sigma_{\tilde{y}}$, respectively.
A6 The sequences $\tilde{u}_i$ and $\tilde{y}_i$ are mutually uncorrelated and uncorrelated with u_{0_i} and y_{0_i}, respectively.

Introducing the parameter vectors

$$\theta \triangleq \begin{bmatrix} a^T & b^T \end{bmatrix}^T = \begin{bmatrix} a_1 & \dots & a_{n_a} & b_1 & \dots & b_{n_b} \end{bmatrix}^T, \tag{4a}$$

$$\bar{\theta} \triangleq \begin{bmatrix} \bar{a}^T & b^T \end{bmatrix}^T = \begin{bmatrix} 1 & \theta^T \end{bmatrix}^T, \tag{4b}$$

an alternative description of (1)-(3) is given by

$$\bar{\varphi}_{0_i}^T \bar{\theta} = 0, \tag{5a}$$

$$\bar{\varphi}_i = \bar{\varphi}_{0_i} + \tilde{\bar{\varphi}}_i, \tag{5b}$$

where

$$\varphi_{0_i} \triangleq [-y_{0_{i-1}} \ \dots \ -y_{0_{i-n_a}} \ u_{0_{i-1}} \ \dots \ u_{0_{i-n_b}}]^T, \tag{6a}$$

$$\bar{\varphi}_{0_i} \triangleq [-y_{0_i} \ \varphi_{0_i}^T]^T, \tag{6b}$$

$$\varphi_i \triangleq [-y_{i-1} \ \dots \ -y_{i-n_a} \ u_{i-1} \ \dots \ u_{i-n_b}]^T, \tag{6c}$$

$$\bar{\varphi}_i \triangleq [-y_i \ \varphi_i^T]^T, \tag{6d}$$

$$\tilde{\varphi}_i \triangleq [-\tilde{y}_{i-1} \ \dots \ -\tilde{y}_{i-n_a} \ \tilde{u}_{i-1} \ \dots \ \tilde{u}_{i-n_b}]^T, \tag{6e}$$

$$\tilde{\bar{\varphi}}_i \triangleq [-\tilde{y}_i \ \tilde{\varphi}_i^T]^T. \tag{6f}$$

The identification problem is now given by:

Problem 1. Given an incrementally increasing number k of measured input-output samples

$$Z^k \triangleq \{u_1, y_1, ..., u_i, y_i, ..., u_k, y_k\}, \tag{7}$$

determine an estimate of the augmented parameter vector

$$\vartheta \triangleq [a_1 \ \dots \ a_{n_a} \ b_1 \ \dots \ b_{n_b} \ \sigma_{\tilde{y}} \ \sigma_{\tilde{u}}]^T. \tag{8}$$

Throughout this paper, the convention is made that estimated quantities are marked by a $\hat{}$ while time dependent quantities have a sub- or superscript k, e.g. $\hat{\vartheta}_k$.

3. THE FRISCH SCHEME

As an alternative formulation of (5a), the following expression can be considered

$$\Sigma_{\bar{\varphi}_0} \bar{\theta} = 0, \tag{9}$$

where

$$\Sigma_{\bar{\varphi}_0} \triangleq E[\bar{\varphi}_{0_i} \bar{\varphi}_{0_i}^T] \in \mathbb{R}^{(n_a+n_b+1)\times(n_a+n_b+1)} \tag{10}$$

is the noise-free covariance matrix, which is singular positive semidefinite, with $\text{rank}(\Sigma_{\bar{\varphi}_0}) = n_a + n_b$ (i.e. rank-one deficient). Here, $E[\cdot]$ denotes the expected value operator. Due to the stated assumptions, the noise-free covariance matrix can be decomposed into

$$\Sigma_{\bar{\varphi}_0} = \Sigma_{\bar{\varphi}} - \Sigma_{\tilde{\bar{\varphi}}}, \tag{11}$$

where

$$\Sigma_{\bar{\varphi}} \triangleq E[\bar{\varphi}_i \bar{\varphi}_i^T] \tag{12}$$

is the covariance matrix of the noisy data and

$$\Sigma_{\tilde{\bar{\varphi}}} \triangleq E[\tilde{\bar{\varphi}}_i \tilde{\bar{\varphi}}_i^T] = \begin{bmatrix} \sigma_{\tilde{y}} I_{n_a+1} & 0 \\ 0 & \sigma_{\tilde{u}} I_{n_b} \end{bmatrix}. \tag{13}$$

Note that, in the noisy case, the covariance matrix $\Sigma_{\bar{\varphi}}$ is generally of full rank. Moreover, it can be approximated by the sample covariance matrix

$$\hat{\Sigma}_{\bar{\varphi}}^k \triangleq \frac{1}{k} \sum_{i=1}^{k} \bar{\varphi}_i \bar{\varphi}_i^T \approx \Sigma_{\bar{\varphi}}. \tag{14}$$

The Frisch scheme identification problem can thus be re-expressed as follows:

Problem 2. Given a sample covariance matrix $\hat{\Sigma}_{\bar{\varphi}}^k$ of noisy observations, determine $\hat{\Sigma}_{\tilde{\bar{\varphi}}}^k$ such that

$$\hat{\Sigma}_{\bar{\varphi}_0}^k = \hat{\Sigma}_{\bar{\varphi}}^k - \hat{\Sigma}_{\tilde{\bar{\varphi}}}^k \geq 0 \qquad \text{and} \qquad \det(\hat{\Sigma}_{\bar{\varphi}_0}^k) = 0. \tag{15}$$

Problem 2 does not yield a unique solution, but the set $(\sigma_{\tilde{u}}, \sigma_{\tilde{y}})$ which satisfies (15) defines a convex curve in the first quadrant of the noise space $\mathbb{R}^2$, where each point can be uniquely mapped into the parameter space $\mathbb{R}^{n_a+n_b}$. This means that a particular solution can be characterised by the input measurement noise variance, since a given $\sigma_{\tilde{u}}$ uniquely defines $\sigma_{\tilde{y}}$, hence θ (Beghelli et al. [1990]). These functional relationships in the Frisch scheme may be formalised as

$$\sigma_{\tilde{y}}(Z^k, \sigma_{\tilde{u}}) \triangleq \lambda_{\min}\left[A_k\left(Z^k, \sigma_{\tilde{u}}\right)\right], \tag{16a}$$

$$\theta(Z^k, \sigma_{\tilde{y}}, \sigma_{\tilde{u}}) \triangleq -\left(\hat{\Sigma}_{\varphi}^k - \begin{bmatrix} \sigma_{\tilde{y}} I_{n_a} & 0 \\ 0 & \sigma_{\tilde{u}} I_{n_b} \end{bmatrix}\right)^{-1} \hat{\xi}_k, \tag{16b}$$

where $\lambda_{\min}$ is the minimum eigenvalue operator and

$$A_k\left(Z^k, \sigma_{\tilde{u}}\right) \triangleq \hat{\Sigma}_y^k - \hat{\Sigma}_{yu}^k \left[\hat{\Sigma}_u^k - \sigma_{\tilde{u}} I_{n_b}\right]^{-1} \hat{\Sigma}_{uy}^k. \tag{17}$$

The individual block matrices in (16b) and (17) are defined by

$$\hat{\Sigma}_{\bar{\varphi}}^k \triangleq \begin{bmatrix} \hat{\Sigma}_y^k & \hat{\Sigma}_{yu}^k \\ \hat{\Sigma}_{uy}^k & \hat{\Sigma}_u^k \end{bmatrix} \triangleq \begin{bmatrix} \times & \times \\ \hat{\xi}_k & \hat{\Sigma}_{\varphi}^k \end{bmatrix}, \tag{18}$$

where $\hat{\Sigma}_u^k \in \mathbb{R}^{n_b \times n_b}$, $\hat{\xi}_k \in \mathbb{R}^{(n_a+n_b)\times 1}$ and $\times$ denotes an arbitrary entry.

In order to obtain a single solution within the Frisch scheme, a model selection criterion is required to be introduced and the approach adopted here makes use of a set of high order YW equations (Diversi et al. [2006]). In this case, $\sigma_{\tilde{u}}$, which completely characterises the solution of the Frisch scheme, is obtained by minimising the cost function

$$J_k \triangleq J(Z^k, \theta) = \|\hat{\Sigma}_{\zeta\bar{\varphi}}^k \bar{\theta}\|_2^2 \tag{19}$$

with

$$\hat{\Sigma}_{\zeta\bar{\varphi}}^k \triangleq \frac{1}{k} \sum_{i=1}^{k} \zeta_i \bar{\varphi}_i^T \approx E\left[\zeta_i \bar{\varphi}_i^T\right] = \Sigma_{\zeta\bar{\varphi}} \tag{20}$$

and

$$\zeta_i \triangleq \left[u_{i-n_b-n_\zeta} \ \dots \ u_{i-n_b-1}\right]^T \in \mathbb{R}^{n_\zeta \times 1}, \tag{21}$$

where $n_\zeta \geq n_a + n_b + 1$ is user specified. This approach can be interpreted as combining the Frisch scheme with the instrumental variable estimator (cf. e.g. Söderström and Mahata [2002], Söderström et al. [2002]), which utilises delayed inputs as instruments, exploiting the fact that

$$\Sigma_{\zeta\bar{\varphi}} \bar{\theta} = 0 \tag{22}$$

holds in the asymptotic case.

The Frisch scheme can hence be summarised as

$$\hat{\sigma}_{\tilde{u}}^k = \arg \min_{\sigma_{\tilde{u}}} J(Z^k, \theta), \tag{23a}$$

$$\hat{\sigma}_{\tilde{y}}^k = \sigma_{\tilde{y}}(Z^k, \hat{\sigma}_{\tilde{u}}^k), \tag{23b}$$

$$\hat{\theta}_k = \theta(Z^k, \hat{\sigma}_{\tilde{y}}^k, \hat{\sigma}_{\tilde{u}}^k). \tag{23c}$$

Note that since J_k depends on θ, the computation of $\hat{\sigma}_{\tilde{u}}^k$ also requires (16b), which, in turn, requires (16a) and (17).

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. Typically, such a recursive estimation scheme must obey the following principles (Ljung [1999]):

P1 The processing must with certainty be completed during one sample interval using a fixed and *a priori* known amount of calculations.
P2 The data, which is passed from one recursion step to the next, must be stored in a finite-dimensional information vector.

The following subsections develop the recursive equations for the Frisch scheme, where the common idea is to use iterative procedures which carry out a single iteration as new data arrives. Such approaches are frequently utilised within recursive prediction error-methods (see e.g. Ljung and Söderström [1983]).

4.1 Update of covariance matrices

In order to satisfy requirement P2, the covariance matrices in (14) and (20) are to be updated recursively. By weighting the ith data at time k with

$$\beta_i^k = \lambda_k \beta_i^{k-1} \quad \text{for} \quad 0 \leq i \leq k-1 \quad \text{and} \quad \beta_k^k \triangleq 1, \tag{24}$$

the update equations are given by (cf. Chapter 11.2 in Ljung [1999])

$$\hat{\Sigma}_{\bar{\varphi}}^k = \hat{\Sigma}_{\bar{\varphi}}^{k-1} + \gamma_k \left(\bar{\varphi}_k \bar{\varphi}_k^T - \hat{\Sigma}_{\bar{\varphi}}^{k-1} \right), \tag{25a}$$

$$\hat{\Sigma}_{\zeta\bar{\varphi}}^k = \hat{\Sigma}_{\zeta\bar{\varphi}}^{k-1} + \gamma_k \left(\zeta_k \bar{\varphi}_k^T - \hat{\Sigma}_{\zeta\bar{\varphi}}^{k-1} \right). \tag{25b}$$

The normalising gain γ_k is given by

$$\gamma_k \triangleq \left(\sum_{i=1}^{k} \beta_{k,i} \right)^{-1} = \frac{\gamma_{k-1}}{\lambda_k + \gamma_{k-1}}, \tag{26}$$

which reduces to $1/k$ in the case of no adaptivity, i.e. λ_k equal to 1, or to $1-\lambda$ in the case of exponential forgetting, i.e. $\lambda_k = \lambda$ where $0 < \lambda < 1$.

Using the recursively updated covariance matrices (25), it is now possible to evaluate the Frisch equations (23) at each time step. Although such an algorithm may satisfy P1 and P2, it cannot be considered to be a recursive scheme since only the trivial covariance matrix update operations are performed. However, this repeatedly applied Frisch scheme (RAFS) is used for comparison purposes in the subsequent development, since it exhibits all the characteristic properties of the Frisch scheme.

A fully recursive Frisch scheme (RFS) is proposed in the following subsections, by developing update equations for $\sigma_{\tilde{u}}$, $\sigma_{\tilde{y}}$ and θ.

4.2 Update of θ

In order to obtain a recursive expression for $\hat{\theta}_k$, an approach is adopted here, similar to that in Sagara and Wada [1977], Zheng and Feng [1989], where the bias of the recursive least squares (RLS) estimate is compensated at each time step k. It is assumed that estimates of $\hat{\sigma}_{\tilde{u}}^k$ and $\hat{\sigma}_{\tilde{y}}^k$ have already been obtained. The update equations for the latter two quantities are given in the remainder of this Section.

Equation (1) can be rewritten as

$$A(q^{-1})y_i - B(q^{-1})u_i = A(q^{-1})\tilde{y}_i - B(q^{-1})\tilde{u}_i \triangleq e_i, \tag{27}$$

where the residual e_i is the difference of two moving average processes. This allows the formulation of a linear regression problem

$$y_i = \varphi_i^T \theta + e_i \tag{28}$$

and the application of the least squares (LS) estimator. It is well known that the LS estimate

$$\hat{\theta}_k^{\mathrm{LS}} = \left[\sum_{i=1}^{k} \varphi_i \varphi_i^T \right]^{-1} \sum_{i=1}^{k} \varphi_i y_i \tag{29}$$

is asymptotically biased in the presence of measurement noise. An explicit expression for the bias can be obtained by substituting (28) in (29) which yields

$$\hat{\theta}_k^{\mathrm{LS}} = \theta + \left[\sum_{i=1}^{k} \varphi_i \varphi_i^T \right]^{-1} \sum_{i=1}^{k} \varphi_i e_i. \tag{30}$$

By using $e_i = -\tilde{\varphi}_i^T \theta + \tilde{y}_i$, and dividing by k it follows that

$$\frac{1}{k} \sum_{i=1}^{k} \varphi_i \varphi_i^T \left(\hat{\theta}_k^{\mathrm{LS}} - \theta \right) = -\frac{1}{k} \sum_{i=1}^{k} \varphi_i \tilde{\varphi}_i \theta + \frac{1}{k} \sum_{i=1}^{k} \varphi_i \tilde{y}_i, \tag{31}$$

which becomes, in the asymptotic case, i.e. for $k \to \infty$

$$\Sigma_\varphi \left(\theta^{\mathrm{LS}} - \theta \right) = - \begin{bmatrix} \sigma_{\tilde{y}} I_{n_a} & 0 \\ 0 & \sigma_{\tilde{u}} I_{n_b} \end{bmatrix} \theta \tag{32}$$

or, equivalently,

$$\theta = \theta^{\mathrm{LS}} + \Sigma_\varphi^{-1} \begin{bmatrix} \sigma_{\tilde{y}} I_{n_a} & 0 \\ 0 & \sigma_{\tilde{u}} I_{n_b} \end{bmatrix} \theta, \tag{33}$$

where Σ_φ is obtained by deleting the first row and column of $\Sigma_{\bar{\varphi}}$.

Equation (33) gives rise to a recursive form and if the noise variances are known (or estimated), it is possible to apply the RLS estimator and compensate for the bias at each time step k. This gives the update equation

$$\hat{\theta}_k = \hat{\theta}_k^{\mathrm{LS}} + P_k \hat{D}_k \hat{\theta}_{k-1} \tag{34}$$

with

$$\hat{D}_k \triangleq \begin{bmatrix} \hat{\sigma}_{\tilde{y}}^k I_{n_a} & 0 \\ 0 & \hat{\sigma}_{\tilde{u}}^k I_{n_b} \end{bmatrix}, \tag{35}$$

whilst the corresponding RLS equations are given by

$$\hat{\theta}_k^{\mathrm{LS}} = \hat{\theta}_{k-1}^{\mathrm{LS}} + L_k \left(y_k - \varphi_k^T \hat{\theta}_{k-1}^{\mathrm{LS}} \right), \tag{36a}$$

$$L_k = \frac{P_{k-1}\varphi_k}{\varphi_k^T P_{k-1} \varphi_k + \frac{1-\gamma_k}{\gamma_k}}, \tag{36b}$$

$$P_k = \frac{1}{1-\gamma_k} \left(P_{k-1} - \frac{P_{k-1}\varphi_k \varphi_k^T P_{k-1}}{\varphi_k^T P_{k-1} \varphi_k + \frac{1-\gamma_k}{\gamma_k}} \right). \tag{36c}$$

Note that by utilising the above normalised gain version of RLS and applying the matrix inversion lemma (see Ch. 11.2 in Ljung [1999]), P_k can be utilised in (34) to compute $[\hat{\Sigma}_\varphi^k]^{-1}$ recursively, which avoids the matrix inversion.

4.3 Update of $\sigma_{\tilde{y}}$

In order to compute $\hat{\sigma}_{\tilde{y}}^k$, the least eigenvalue of $\hat{A}_k \triangleq A_k\left(Z^k, \hat{\sigma}_{\tilde{u}}\right)$ in (17) is to be determined. This could be

achieved via the singular value decomposition (SVD), which generally requires $O(n^3)$ flops for a $n \times n$ matrix (Golub and Van Loan [1996]). When only a few singular triplets or eigenpairs are required, more efficient algorithms exist which only track the subspace corresponding to one or more singular values (Comon and Golub [1990]). Such an approach is feasible when the corresponding matrix (hence the singular triplets) 'varies slowly' with time, which is assumed to be the case here. Since the matrix update $\Delta A = \hat{A}_k - \hat{A}_{k-1}$ is generally of full rank $n_a + 1$, a gradient based algorithm, which requires $O(n^2)$ flops [1], is applied in the subsequent development, in order to determine a recursive expression for $\hat{\sigma}^k_{\tilde{y}}$. More specifically, an iterative conjugate gradient method similar to that proposed in Chen et al. [1986] is used, where one iteration per recursion is applied.

Suppose the eigensystem is given by

$$\hat{A}_k x_k = \lambda_k x_k, \tag{37}$$

where $x_k \in \mathbb{R}^{n_a+1}$ and λ_k is a scalar parameter. Then the minimum eigenvalue can be obtained by minimising the Rayleigh quotient (Golub and Van Loan [1996])

$$\lambda_k \triangleq R(x_k) = \frac{x_k^T \hat{A}_k x_k}{x_k^T x_k}. \tag{38}$$

Utilising a gradient-based approach, the update equations for the minimum eigenvalue are given by (cf. Chen et al. [1986])

$$x_k = \bar{x}_{k-1} + \mu^x_{k-1}\psi^x_{k-1}, \tag{39a}$$
$$\bar{x}_k = x_k / \left(x_k^T x_k\right)^{1/2}, \tag{39b}$$
$$\hat{\sigma}^k_{\tilde{y}} = \bar{x}_k^T \hat{A}_k \bar{x}_k, \tag{39c}$$

where the scalar μ^x_k denotes the stepsize and $\psi^x_k \in \mathbb{R}^{n_a+1}$ the update direction. The conjugate gradient update direction is given by

$$r_k = \hat{\sigma}^k_{\tilde{y}}\bar{x}_k - \hat{A}_k\bar{x}_k, \tag{40a}$$
$$q_{k-1} = -\left(r_k^T \hat{A}_k \psi^x_{k-1}\right) / \left({\psi^x_{k-1}}^T \hat{A}_k \psi^x_{k-1}\right), \tag{40b}$$
$$\psi^x_k = r_k + q_{k-1}\psi^x_{k-1}, \tag{40c}$$

whilst the step size is chosen as

$$\mu^x_k = \left(r_{k-1}^T r_{k-1}\right) / \left(r_{k-1}^T \hat{A}_k r_{k-1}\right). \tag{41}$$

The algorithm is initialised with a guess of x_0 and $\hat{A}_0$, such that

$$\bar{x}_0 = x_0 / \left(x_0^T x_0\right)^{1/2}, \tag{42a}$$
$$\psi^x_0 = r_0 = R(x_0)\bar{x}_0 - \hat{A}_0\bar{x}_0. \tag{42b}$$

Remark 3. (Frisch-like character). One of the significant characteristics of the Frisch scheme is that the estimated model belongs to a class characterised by a convex curve in the noise space, where the functional relationship between $\sigma_{\tilde{u}}$ and $\sigma_{\tilde{y}}$ defining a locus of solutions is given by (16a). However, computing the output noise variance recursively via (39)-(41) will inevitably introduce an error, which means that the set $(\hat{\sigma}^k_{\tilde{u}}, \hat{\sigma}^k_{\tilde{y}})$ will not exactly lie on this convex curve. This would, strictly speaking, only imply a 'Frisch-like' character for the solution. However, as illustrated in Section 5, the set $(\hat{\sigma}^k_{\tilde{u}}, \hat{\sigma}^k_{\tilde{y}})$ can converge to the convex curve, after the initialisation transients have decayed.

Remark 4. (Error detection and re-initialisation). An accurate computation of $\hat{\sigma}^k_{\tilde{y}}$ is important within the recursive Frisch scheme, since $\hat{\sigma}^k_{\tilde{y}}$ effects $\hat{\theta}_k$ in (34), which in turn will effect $\hat{\sigma}^{k+1}_{\tilde{u}}$ (see below). It is crucial that the initialisation of x_0 is sufficiently close to the true eigenvector corresponding to the smallest eigenvalue of A_0. Otherwise the estimate of $\sigma_{\tilde{y}}$ may exhibit an oscillation between different solutions, a behaviour which is then propagated to $\hat{\sigma}^k_{\tilde{u}}$ and $\hat{\theta}_k$ due to the previously stated dependencies. Note that the residual r_k in (40) can be utilised to compute an accuracy measure, say $\rho_k = ||r_k||_2^2$, for the eigenpair computation, with which an oscillating condition may be detected. An appropriate re-initialisation of (42) can be invoked in order to overcome this potential problem.

4.4 Update of $\sigma_{\tilde{u}}$

For the recursive computation of $\hat{\sigma}^k_{\tilde{u}}$, a gradient-based iterative search algorithm can be utilised, where one iteration per recursion is applied.

The aim is to find an expression of the form

$$\hat{\sigma}^k_{\tilde{u}} = \hat{\sigma}^{k-1}_{\tilde{u}} + \mu^u_k \psi^u_k, \tag{43}$$

where the potentially time-varying step-size μ^u_k is a user chosen positive scalar and $\psi^u_k \in \mathbb{R}$ is the update direction. Several choices for ψ^u_k are possible, a simple one being the negative steepest gradient of J_k (cf. (19)) with respect to $\sigma_{\tilde{u}}$. Consequently, $dJ_k/d\sigma_{\tilde{u}}$ is to be determined at $\hat{\vartheta}_{k-1}$ and by applying the chain rule for vector differentiation, the first order derivative is given by

$$\left.\frac{dJ_k}{d\sigma_{\tilde{u}}}\right|_{\vartheta=\hat{\vartheta}_{k-1}} = \left(\frac{d\sigma_{\tilde{y}}}{d\sigma_{\tilde{u}}}\frac{\partial\theta}{\partial\sigma_{\tilde{y}}} + \frac{\partial\theta}{\partial\sigma_{\tilde{u}}}\right)\left.\frac{dJ_k}{d\theta}\right|_{\vartheta=\hat{\vartheta}_{k-1}}. \tag{44}$$

Defining $\hat{\Sigma}^k_{\zeta\bar{\varphi}} \triangleq \left[\Sigma^k_1 \ \Sigma^k_2\right]$ and rewriting the cost function (19) as

$$\begin{aligned} J_k &= \bar{\theta}^T \left(\hat{\Sigma}^k_{\zeta\bar{\varphi}}\right)^T \hat{\Sigma}^k_{\zeta\bar{\varphi}}\bar{\theta} \\ &= \begin{bmatrix}1 & \theta^T\end{bmatrix}\begin{bmatrix}{\Sigma^k_1}^T \\ {\Sigma^k_2}^T\end{bmatrix}\begin{bmatrix}\Sigma^k_1 & \Sigma^k_2\end{bmatrix}\begin{bmatrix}1 \\ \theta\end{bmatrix} \\ &= {\Sigma^k_1}^T\Sigma^k_1 + {\Sigma^k_1}^T\Sigma^k_2\theta + \theta^T{\Sigma^k_2}^T\Sigma^k_1 + \theta^T{\Sigma^k_2}^T\Sigma^k_2\theta, \end{aligned} \tag{45}$$

one obtains, for the last term in (44),

$$\frac{dJ_k}{d\theta} = 2{\Sigma^k_2}^T\left(\Sigma^k_1 + \Sigma^k_2\theta\right) \in \mathbb{R}^{(n_a+n_b)\times 1}. \tag{46}$$

In addition, expressions for the sensitivity derivatives $\partial\theta/\partial\sigma_{\tilde{u}}$, $\partial\theta/\partial\sigma_{\tilde{y}}$ and $d\sigma_{\tilde{y}}/d\sigma_{\tilde{u}}$ are given by [2]:

Lemma 5. The sensitivity derivatives at $\hat{\vartheta}_{k-1}$ are given by

[1] In the case of a rank-one update, it is possible to track d singular triplets using $O(nd^2)$ flops only (cf. Section V in Comon and Golub [1990] or Davila [1994]).

[2] Note, that the sensitivity derivatives given here are based on the non-recursive Frisch equations (16a) and (16b). The determination of recursive expressions based on (39c) and (34) may also be possible.

$$\left.\frac{\partial\theta}{\partial\sigma_{\tilde{u}}}\right|_{\vartheta=\hat{\vartheta}_{k-1}} = \begin{bmatrix}0 & \hat{b}_{k-1}^T\end{bmatrix}\left(\hat{\Sigma}_{\varphi}^k - \hat{D}_{k-1}\right)^{-1} \in \mathbb{R}^{1\times(n_a+n_b)}, \tag{47a}$$

$$\left.\frac{\partial\theta}{\partial\sigma_{\tilde{y}}}\right|_{\vartheta=\hat{\vartheta}_{k-1}} = \begin{bmatrix}\hat{a}_{k-1}^T & 0\end{bmatrix}\left(\hat{\Sigma}_{\varphi}^k - \hat{D}_{k-1}\right)^{-1} \in \mathbb{R}^{1\times(n_a+n_b)}, \tag{47b}$$

$$\left.\frac{d\sigma_{\tilde{y}}}{d\sigma_{\tilde{u}}}\right|_{\vartheta=\hat{\vartheta}_{k-1}} = -\frac{\hat{b}_{k-1}^T\hat{b}_{k-1}}{\hat{a}_{k-1}^T\hat{a}_{k-1}} \in \mathbb{R}. \tag{47c}$$

Proof 4.1. Assuming that $\hat{\vartheta}_{k-1}$ is close to $\hat{\vartheta}_k$, the derivations are conceptually identical to those given in Appendix II.A and II.B in Söderström [2007a] when a linearisation is carried out around $\hat{\vartheta}_{k-1}$. ■

Consequently, the gradient can be computed by (44) using (46) and (47), which defines the steepest gradient update direction

$$\psi_k^u = -\left.\frac{dJ_k}{d\sigma_{\tilde{u}}}\right|_{\vartheta=\hat{\vartheta}_{k-1}}. \tag{48}$$

For the variable step size μ_k^u in (43), a constant value $\mu_k^u = \mu^u$ has been found to give excellent results in practice for the cases considered.

Remark 6. (Projection facility). In order to stabilise the algorithm during the initial phase, it might be advantageous to project the noise variances into the intervals

$$0 \leq \sigma_{\tilde{u}} \leq \sigma_{\tilde{u}}^{\max}, \tag{49a}$$

$$0 \leq \sigma_{\tilde{y}} \leq \sigma_{\tilde{y}}^{\max}, \tag{49b}$$

where $\sigma_{\tilde{u}}^{\max}$ and $\sigma_{\tilde{y}}^{\max}$ are the maximal admissible solutions for $\sigma_{\tilde{u}}$ and $\sigma_{\tilde{y}}$, respectively. These values can be computed from the data as in Beghelli et al. [1990]

$$\sigma_{\tilde{u}}^{\max} = \lambda_{\min}\left[\hat{\Sigma}_u^k - \hat{\Sigma}_{uy}^k[\Sigma_y^k]^{-1}\hat{\Sigma}_{yu}^k\right], \tag{50a}$$

$$\sigma_{\tilde{y}}^{\max} = \lambda_{\min}\left[\hat{\Sigma}_y^k - \hat{\Sigma}_{yu}^k[\Sigma_u^k]^{-1}\hat{\Sigma}_{uy}^k\right]. \tag{50b}$$

Alternatively, a positive constant can be chosen for the maximum admissible values, if such *a priori* knowledge is available.

Remark 7. (Computational complexity). The computation time per single recursion can be approximately halved by using the RFS approach compared to the RAFS, although both algorithms are of cubic complexity w.r.t. the number of system parameters to be identified. However, fast algorithms of quadratic order are possible by accounting for the fact that the eigenvector corresponding to the smallest eigenvalue of A_k is also part of the parameter vector to be estimated (cf. Linden et al. [2007]).

Remark 8. (Classification of the RFS). The RFS can be considered to belong to the family of iterative bias-compensating LS algorithms (see e.g. Sagara and Wada [1977], Zheng and Feng [1989], Söderström [2007b] and the references within). The essential difference is that the measurement noise variances are computed in a different way.

A detailed description of the RFS algorithm developed in this Section is given in Appendix A.

5. NUMERICAL EXAMPLE

A LTI SISO system with $n_a = n_b = 2$ and given by

$$\vartheta = [-1.5\ 0.7\ 1\ 0.5\ 4\ 1]^T \tag{51}$$

(cf. (8)) is simulated for 500 samples using a zero mean, white and Gaussian distributed input signal of unity variance. The RAFS and the RFS are applied to estimate ϑ; both using $n_\zeta = n_a + n_b + 1$, whilst $\mu^u = 0.05$ is chosen for the gradient-based approach. The results are shown in Figure 2.

As expected, the RAFS yields identical estimates as its offline counterpart after $k = 500$ recursions, while the estimates of the RFS are slightly different. However, this difference is mainly due to different solutions of $\hat{\sigma}_{\tilde{u}}^k$ and, w.r.t Remark 3, it is interesting to investigate how accurate the computation of $\hat{\sigma}_{\tilde{y}}^k$ is, i.e. how exact the least eigenvalue of A_k can be approximated using the conjugate gradient method. Therefore, the experiment is repeated where the RFS utilises the same input measurement noise variance estimate as the exact algorithm, i.e. $\hat{\sigma}_{\tilde{u}}^k$ is identical in both cases. The differences in $\hat{\sigma}_{\tilde{y}}^k$ is then a measure for the accuracy of the subspace tracking algorithm in Section 4.3. It turns out that the difference between both estimates is marginal: after 500 recursion steps, for instance, it is approximately $3\cdot10^{-4}$ while it decreases to $3\cdot10^{-7}$ after 5000 iterations. This means that (at least in the example considered here) the RFS yields estimates of $\sigma_{\tilde{u}}$ and $\sigma_{\tilde{y}}$ which seem to converge to the set of admissible Frisch solutions, once the initialisation transients have decayed. Moreover, it is observed in Figure 2 that $\hat{\sigma}_{\tilde{u}}^k$ based on the RFS is 'smoother' than its exact counterpart; a property which is propagated to all elements of $\hat{\vartheta}_k$ and which might be desirable in practical applications. It is clear that the smoothness depends on the choice of μ^u.

6. CONCLUSIONS

The Frisch scheme for the identification of dynamical systems has been modified to recursively estimate the parameters and measurement noise variances of LTI SISO EIV systems. The recursive scheme belongs to the family of iterative bias-compensating LS algorithms, which compensates for the RLS bias at each recursion whilst the recursive update of measurement noise variance estimates is achieved using gradient-based techniques. A numerical example has been provided to illustrate the characteristics of the developed algorithm. Although approximations are necessary to track the smallest eigenvalue of a slowly varying matrix, the solution seems to retain the Frisch character after the initialisation transients have decayed.

Further work could concern the adaptive behaviour of the method, i.e its capability to track time-varying system parameters and to investigate convergence and consistency properties.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the inspiring discussions with Dr. R. Diversi and Prof. R. Guidorzi, as well as the anonymous reviewers for their valuable feedback.

REFERENCES

S. Beghelli, R. P. Guidorzi, and U. Soverini. The Frisch scheme in dynamic system identification. *Automatica*, 26(1):171–176, 1990.

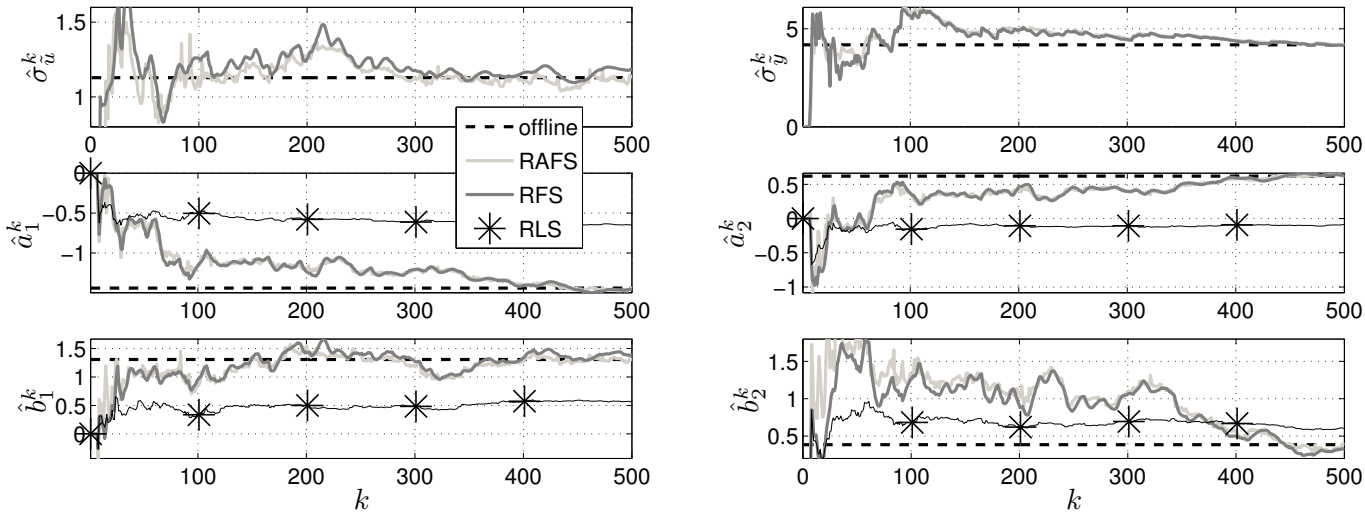

Fig. 2. Estimates of ϑ_k for system (51) using the offline Frisch scheme, the RAFS, the RFS and RLS.

H. Chen, T. K. Sarkar, S. A. Dianat, and J. D. Brulé. Adaptive spectral estimation by the conjugate gradient method. *IEEE Trans. on Acoustics, Speech, and Signal Proc.*, 34(2):272–284, 1986.

P. Comon and G. H. Golub. Tracking a few extreme singular values and vectors in signal processing. *Proc. of IEEE*, 78(8):1327–1343, 1990.

C. E. Davila. An efficient recursive total least squares algorithm for FIR adaptive filtering. *IEEE Trans. on Signal Proc.*, 42(2):268–280, 1994.

R. Diversi, R. Guidorzi, and U. Soverini. A new criterion in EIV identification and filtering applications. In *Preprints 13th IFAC Symposium on System Identification*, pages 1993–1998, Rotterdam, The Netherlands, August 2003.

R. Diversi, R. Guidorzi, and U. Soverini. Yule-Walker equations in the Frisch scheme solution of errors-in-variables identification problems. In *Proc. 17th Int. Symp. on Math. Theory of Networks and Systems*, pages 391–395, 2006.

G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore and London, 3rd edition, 1996.

J. G. Linden and K. J. Burnham. A recursive Frisch scheme algorithm for coloured output noise. To be published in Proceedings of ICINCO 2008, Funchal, Madeira, Portugal, 2008.

J. G. Linden, B. Vinsonneau, and K. J. Burnham. Recursive Frisch scheme identification incorporating adaptivity. In *Proc. DVD-ROM 21st IAR & ACD Workshop*, Nancy, France, November 2006.

J. G. Linden, B. Vinsonneau, and K. J. Burnham. Fast algorithms for recursive Frisch scheme system identification. In *Proc. CD-ROM 22nd IAR & ACD Workshop*, Grenoble, France, November 2007.

L. Ljung. *System Identification - Theory for the user*. PTR Prentice Hall Infromation and System Sciences Series. Prentice Hall, New Jersey, 2nd edition, 1999.

L. Ljung and T. Söderström. *Theory and Practice of Recursive Identification*. M.I.T. Press, Cambridge, MA, 1983.

S. Sagara and K. Wada. On-line modified least-squares parameter estimation of linear discrete dynamic systems. *Int. J. Control*, 25(3):329–343, 1977.

T. Söderström. Accuracy analysis of the Frisch scheme for identifying errors-in-variables systems. *IEEE Trans. Autom. Contr.*, 52(6):985–997, 2007a.

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

T. Söderström and K. Mahata. On instrumental variable and total least squares approaches for identification of noisy systems. *Int. J. Control*, 75(6):381–389, 2002.

T. Söderström, U. Soverini, and K. Mahata. Perspectives on errors-in-variables estimation for dynamic systems. *Signal Proc.*, 82(8):1139–1154, 2002.

W. X. Zheng and C. B. Feng. Unbiased parameter estimation of linear systems in the presence of input and output noise. *Int. J. of Adaptive Control and Signal Proc.*, 3:231–251, 1989.

Appendix A. RFS ALGORITHM

1) Choose μ^u, $n_\zeta \geq n_a+n_b+1$, $0 < \lambda_k < 1$ and $j = n_\zeta+n_b$

2) Initialise RLS with $\hat{\theta}^{\mathrm{LS}}_{n_a} = 0$, $P_{n_a} = 0.1I$ and for $k = n_a+1, ..., j$

- **a)** Compute P_k, L_k and $\hat{\theta}^{\mathrm{LS}}_k$ using (36)

3) Initialise conjugate gradient method:

- **a)** Obtain 'guess' for x_j, e.g. via SVD of A_j
- **b)** Compute $\bar{x}_j$ and ψ^x_j via (42)

4) Initialise $\hat{\Sigma}^j_{\zeta\bar{\varphi}} = 0$, $\hat{\Sigma}^j_{\bar{\varphi}} = \frac{1}{n_\zeta-1}\sum_{i=n_b+1}^{j} \bar{\varphi}_i\bar{\varphi}_i^T$ and $\gamma_j = \frac{1}{n_\zeta-1}$

5) For $k = j+1, ...$

- **a)** Update γ_k via (26)
- **b)** Update $\hat{\Sigma}^k_{\bar{\varphi}}$ and $\hat{\Sigma}^k_{\zeta\bar{\varphi}}$ via (25)
- **c)** Compute $\hat{\theta}^{\mathrm{LS}}_k$ using (36)
- **d)** Update $\hat{\sigma}^k_{\tilde{u}}$
  - **i)** Determine ψ^u_k in (48) utilising (44) and (46)-(47)
  - **ii)** Update $\hat{\sigma}^k_{\tilde{u}}$ via (43)
- **e)** Update $\hat{\sigma}^k_{\tilde{y}}$ in (39c) using (39)-(41)
- **f)** Update $\hat{\theta}_k$ in (34)