FRISCH FILTERING OF NOISY SIGNALS

Paolo Guidorzi
DIENCA, University of Bologna, Italy
e-mail:paolo.guidorzi@mail.ing.unibo.it

Roberto Guidorzi
DEIS, University of Bologna, Italy
e-mail:rguidorzi@deis.unibo.it

ABSTRACT

The Frisch scheme, that considers additive independent noises on the measures of the input and output of a process, has recently led to the development of specific identification procedures. The obtained models, that have already been used to implement smoothing procedures congruent with the scheme, are here used to develop a filtering algorithm.

1 INTRODUCTION

The Frisch scheme was originally introduced by Ragnar Frisch [1] as a conceptual tool to describe the problem of estimating linear relations from data affected by additive noise. Differently from most common schemes, it leads, in the algebraic case, to a whole family of models compatible with the noisy data. The extension to the dynamical case has been performed in recent years [2] and it has been shown that, differently from the algebraic case, it is possible to obtain a single model of the process. This is true, however, only when the assumptions behind the scheme are exactly fulfilled and this never happens in the real world; the application to real (MISO) processes requires thus the introduction of suitable cost functions [3]. The extension to the multivariable case faces also congruence problems and has been performed even more recently [4].

Dynamic Frisch models describe also the covariance matrices of the additive noises; this information can be used to design a Kalman filter for the optimal estimation of the state or of the output of the process. Frisch identification and Kalman filtering can thus constitute an alternative to the simultaneous estimation of the parameters of a model and of its state through extended Kalman filtering [5] [6], that requires the *a priori* knowledge of the covariance matrices of the noises on the state and on the output.

The decomposition of the input/output sequences of a process into orthogonal noiseless and noise parts at the basis of the Frisch scheme has also been the basis of a smoothing procedure that allows the separation of the noiseless part of a sequence (constituted by a regular behaviour) from the additive noise [7]. This result is thus similar to that obtained by Roorda [8], and Roorda and Heij [9]. These decompositions as well as the ways leading to their determination are, however, based on different approaches; Roorda and Heij minimize a quadratic function of the deviation of the observed behaviour from its regular part while the solution considered in [7] introduces the additional requirement of the compatibility of the regular part with a Frisch model. In both cases the concept of behaviour, introduced by Willems [10], and errors–in–variables models play central roles.

This paper extends the smoothing procedure described in [7] to filtering by developing an on–line procedure that requires, at every step, only an update of preceding computations.

The content is organized as follows. Section 2 recalls the context of the Frisch scheme and defines Frisch smoothing and filtering. Section 3 describes the solution of the smoothing problem while Section 4 develops the filtering algorithm. A numerical example compares, in Section 5, Kalman filtering with Frisch smoothing and filtering. Some future developments are described in Section 6 and short concluding remarks are finally reported in Section 7.

2 STATEMENT OF THE PROBLEM

The observed process can be described, in the context of the Frisch scheme, by a state–space model of the type

$$\hat{x}(t+1) = A\,\hat{x}(t) + B\,\hat{u}(t) \tag{1}$$
$$\hat{y}(t) = C\,\hat{x}(t) \tag{2}$$
$$y(t) = \hat{y}(t) + \tilde{y}(t) \tag{3}$$
$$u(t) = \hat{u}(t) + \tilde{u}(t) \tag{4}$$

where $\hat{x} \in R^n$, $\hat{y} \in R^m$, $\hat{u} \in R^r$ and $\tilde{y}$ and $\tilde{u}$ are independent white processes with unknown covariance matrices Q and R that are assumed diagonal (the noises on the input and output components are considered as mutually independent). Model (1)–(4) can be easily rewritten in a form suitable for Kalman filtering as

$$\hat{x}(t+1) = A\,\hat{x}(t) + B\,u(t) - B\,\tilde{u}(t) \tag{5}$$
$$y(t) = C\,\hat{x}(t) + \tilde{y}(t). \tag{6}$$

As described in [11], equations (1) and (2) define, under the assumption of complete observability, a canonical MFD model

$$P(z)\,\hat{y}(t) = Q(z)\,\hat{u}(t) \tag{7}$$

where z denotes the unitary advance operator. Limiting the considerations which follow to MISO systems, model (7) can be written as

$$\begin{aligned} \hat{y}(t+n) &= \sum_{k=1}^{n} \alpha_k \, \hat{y}(t+k-1) + \\ & \sum_{j=1}^{r} \sum_{k=1}^{n} \beta_{jk} \, \hat{u}_j(t+k-1) \, . \end{aligned} \tag{8}$$

If L denotes the length of the available sequences, it is possible to define the Hankel matrices

$$X_k(y) = \begin{bmatrix} y(1) & \dots & y(k) \\ y(2) & \dots & y(k+1) \\ \vdots & \ddots & \vdots \\ y(N) & \dots & y(k+N-1) \end{bmatrix}, \tag{9}$$

$$X_k(u_i) = \begin{bmatrix} u_i(1) & \dots & u_i(k) \\ u_i(2) & \dots & u_i(k+1) \\ \vdots & \ddots & \vdots \\ u_i(N) & \dots & u_i(k+N-1) \end{bmatrix}, \tag{10}$$

the matrix of input/output samples

$$X_k = \left[X_{k+1}(y) \; X_k(u_1) \; \dots \; X_k(u_r) \right] \tag{11}$$

where $k + N \leq L$, $N > (r+1)k$, and the sample covariance matrices Σ_k given by

$$\Sigma_k = \frac{1}{N} X_k^T X_k = \begin{bmatrix} \Sigma(yy) & \Sigma(yu_1) & \dots & \Sigma(yu_r) \\ \Sigma(u_1 y) & \Sigma(u_1 u_1) & \dots & \Sigma(u_1 u_r) \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma(u_r y) & \Sigma(u_r u_1) & \dots & \Sigma(u_r u_r) \end{bmatrix}. \tag{12}$$

Denoting with $\sigma_{\tilde{u}_i}$ $(i = 1, \dots, r)$ and $\sigma_{\tilde{y}}$ the variances of the components of $\tilde{u}(t)$ and $\tilde{y}(t)$, the assumptions behind the Frisch scheme lead, when $N \to \infty$, to the following relations

$$X_k = \hat{X}_k + \tilde{X}_k, \qquad \Sigma_k = \hat{\Sigma}_k + \tilde{\Sigma}_k \tag{13}$$

where

$$\tilde{\Sigma}_k = \begin{bmatrix} R & 0 \\ 0 & Q \end{bmatrix} = \text{diag}\left[\sigma_{\tilde{y}} I_{k+1}, \, \sigma_{\tilde{u}_1} I_k, \dots, \sigma_{\tilde{u}_r} I_k\right]. \tag{14}$$

The identification problem, in the context of the Frisch scheme, consists in determining the order and the parameters of model (8), or of any equivalent state–space model, and the covariance matrix of the noises (14) on the basis of the knowledge of the noisy sequences $u(\cdot)$, $y(\cdot)$ or, equivalently, of the sequence of increasing–dimension matrices Σ_k for $k = 1, 2, \dots$ The solution of this problem can be obtained by minimizing a suitable cost function (given, for instance, by the whiteness or by the variance of the innovations of the Kalman filter associated with the model) on the convex (hyper)surface describing, in the noise space, the whole family of admissible solutions [2], [3].

The context of the Frisch scheme allows defining the following practically relevant problems.

Problem 1: Frisch smoothing. Given a sequence of noise–corrupted input–output sequences $u(\cdot)$, $y(\cdot)$ satisfying the assumptions of the Frisch scheme and the process model, extract the noiseless sequences $\hat{u}(\cdot)$, $\hat{y}(\cdot)$.

Remark 1. The noiseless sequences $\hat{u}(\cdot)$, $\hat{y}(\cdot)$ obtained in the solution of Problem 1 constitute the regular part of the data i.e. a regular behaviour compatible with the process model.

Problem 2: Frisch filtering. Given an increasing sequence of noisy samples $u(1)$, $y(1), \dots, u(t)$, $y(t)$, estimate, as any new sample becomes available, the regular part $\hat{u}(t)$, $\hat{y}(t)$.

3 FRISCH SMOOTHING

Consider the parameters of model (8) arranged into the vector

$$\theta = \left[\alpha_1, \dots, \alpha_n, -1, \, \beta_{11}, \dots, \beta_{1n}, \dots, \beta_{r1}, \dots, \beta_{rn}\right]^T ; \tag{15}$$

because of relation (8), it follows that

$$\hat{X}_n \, \theta = 0 \tag{16}$$

so that

$$X_n \, \theta = \tilde{X}_n \, \theta \, . \tag{17}$$

Equations (16) and (17) describe essentially the time shift invariance of the regular part of the data; some editing on equation (16) allows to obtain the equivalent form

$$M_\theta \, \hat{v} = 0 \tag{18}$$

where

$$M_\theta = \left[\begin{matrix} \alpha_1 & \dots & \alpha_n & -1 & 0 & \dots & 0 \\ 0 & \alpha_1 & \dots & \alpha_n & -1 & \dots & 0 \\ \vdots & & \ddots & & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \alpha_1 & \dots & \alpha_n & -1 \end{matrix} \right. \tag{19}$$

$$\left. \begin{matrix} \dots & \beta_{r1} & \dots & \beta_{rn} & 0 & \dots & 0 \\ \dots & 0 & \beta_{r1} & \dots & \beta_{rn} & \dots & 0 \\ & \vdots & & \ddots & & \ddots & \vdots \\ \dots & 0 & \dots & 0 & \beta_{r1} & \dots & \beta_{rn} \end{matrix} \right],$$

$$\hat{v} = \left[\, \hat{y}(1) \dots \hat{y}(N+n) \mid \dots \mid \hat{u}_r(1) \dots \hat{u}_r(N+n-1) \,\right]^T. \tag{20}$$

Relation (17) can thus be rewritten in the equivalent form

$$M_\theta \, \tilde{v} = \gamma \tag{21}$$

where $\tilde{v}$ is a vector of noise samples with the same structure as $\hat{v}$ and $\gamma = M_\theta v$. Because of the structure of the $(N \times (r+1)(N+n) - r)$ matrix M_θ, $\tilde{v}$ cannot be univocally deduced from relation (21); the only solution of practical interest, i.e. that minimizing the Euclidean norm of $\tilde{v}$, is given by

$$\tilde{v} = M_\theta^+ \gamma \tag{22}$$

where M_θ^+ denotes the pseudoinverse of M_θ. Relation (22) allows, for a given model, to extract the compatible regular part of the data which minimizes the sum of the squares of the errors $\tilde{u}$ and $\tilde{y}$.

4 FRISCH FILTERING

The samples $\hat{u}(\tau)$, $\hat{y}(\tau)$ obtained by Frisch smoothing will depend, in general, on both previous and subsequent noisy samples $u(\cdot)$ and $y(\cdot)$; the only exception concerns the first sample $\hat{u}(1)$, $\hat{y}(1)$, that will depend only on subsequent samples and the last one $\hat{u}(t)$, $\hat{y}(t)$, that will depend only on previous samples.

Relation (22) could thus be used in filtering applications by performing at every step a smoothing and selecting in the whole sequence $\hat{u}(\cdot)$, $\hat{y}(\cdot)$ only the last sample. A procedure of this kind would however be non efficient because it does not rely on previous computations; moreover the amount of computations increases at every step with the dimension of M_θ. It is however possible to take advantage of the specific structure of M_θ to develop an on–line procedure; rewrite, to this purpose, M_θ (19) and $\hat{v}$ (20) as follows

$$M_\theta(N) = \begin{bmatrix} \alpha_1 & \beta_{11} & \dots & \beta_{1r} & \alpha_2 & \beta_{21} & \dots & \beta_{2r} & \dots & \beta_{nr} & -1 & 0 & \dots & \dots & \dots & \dots & \dots & 0 \\ 0 & \dots & \dots & 0 & \alpha_1 & \beta_{11} & \dots & \beta_{1r} & \dots & \beta_{(n-1)r} & \dots & \dots & \beta_{nr} & -1 & 0 & \dots & & 0 \\ \vdots & & & & & & & & & & & & & & & & & \vdots \\ 0 & \dots & \dots & \dots & \dots & \dots & \dots & 0 & \dots & \dots & \dots & \dots & \alpha_n & \beta_{n1} & \dots & \beta_{nr} & & -1 \end{bmatrix} \tag{23}$$

$$\hat{v}(N) = [\ \hat{y}(1)\ \hat{u}_1(1) \dots \hat{u}_r(1) \mid \hat{y}(2) \dots \tag{24}$$
$$\dots \mid \hat{y}(N+n-1)\ \dots \hat{u}_r(N+n-1) \mid \hat{y}(N+n)\ ]^T.$$

where the argument N denotes the number of relations described by (18) i.e. the number of rows of M_θ. Passing from N to $N+1$ (i.e. from $t = N+n$ to $t = N+n+1$), the update of $\hat{v}$ and M_θ is given by

$$\hat{v}(N+1) = \ [\ \hat{v}(N)^T\ \hat{u}_1(N+n) \dots \tag{25}$$
$$\dots \hat{u}_r(N+n) \mid \hat{y}(N+n+1)\ ]^T$$

$$M_\theta(N+1) = \begin{bmatrix} M_\theta(N) & 0_{(N\times(r+1))} \\ 0 \dots 0\ \alpha_1\, \beta_{11} \dots \alpha_n & \beta_{n1} \dots \beta_{nr}\ -1 \end{bmatrix}$$
$$= \begin{bmatrix} M_\theta(N) & 0 \\ m_1 & m_2 \end{bmatrix}. \tag{26}$$

The pseudoinverse of $M_\theta(N+1)$ is thus given by

$$M_\theta^+(N+1) = M_\theta^T(N+1)\big(M_\theta(N+1)\, M_\theta^T(N+1)\big)^{-1}$$
$$= \begin{bmatrix} M_\theta(N) & 0 \\ m_1 & m_2 \end{bmatrix}^T \begin{bmatrix} M_\theta(N)\, M_\theta^T(N) & M_\theta(N)\, m_1^T \\ m_1\, M_\theta^T(N) & m_1 m_1^T + m_2 m_2^T \end{bmatrix}^{-1} \tag{27}$$

Applying a well–known matrix inversion lemma to expression (27) we obtain

$$M_\theta^+(N+1) = \begin{bmatrix} M_\theta^+(N) & 0 \\ 0 & 0 \end{bmatrix} \tag{28}$$
$$+ \begin{bmatrix} \big(I - M_\theta^+(N)\, M_\theta(N)\big)\, m_1^T \\ m_2^T \end{bmatrix} \begin{bmatrix} -m_1\, M_\theta^+(N) & 1 \end{bmatrix} \alpha$$

where the scalar α is given by

$$\alpha = 1/\big[\, m_1 m_1^T + m_2 m_2^T - m_1\, M_\theta^+(N)\, M_\theta(N)\, m_1^T\,\big]. \tag{29}$$

Expressions (28) and (29) can be used to perform an efficient on–line update of M_θ^+ and thus to obtain updated values of $\tilde{v}$, i.e. of $\tilde{u}(t)$, $\tilde{y}(t)$, $\hat{u}(t)$ and $\hat{y}(t)$.

Remark 2. The actual on–line implementation of (28)–(29) and (22) can take advantage of the peculiar structure of M_θ and of the necessity to compute, at every step, only the last $r+1$ entries of $\tilde{v}$.

Remark 3. The filtered values $\hat{u}(t)$, $\hat{y}(t)$ obtained in the sequence of filtering steps does not constitute a smoothing of the available data and a regular behaviour since they derive from data sets of different length.

5 A NUMERICAL EXAMPLE

The example which follows refers to a dynamical system with order 2, poles given by $p_{1,2} = -0.3 \pm 0.6j$ and to the same data set used in [7]. The input sequence is a PRBS with standard deviation equal to 1 and also the standard deviation of the corresponding output sequence is unitary. Sequences of Gaussian noises with standard deviations equal to 1.2 and 0.8 have been added to the input/output sequences; the amount of additive noise on the input measures is thus equal to 120% (in amplitude) while the noise on the output reaches an 80% level. A model of the process has then been estimated, in the context of the Frisch scheme, by means of the identification procedure described in [3]. Figure 1 compares the noiseless output sequence (continuous line) with the output of a Kalman filter designed using the true model and the true noise variances; the standard deviation of the error is 0.665.

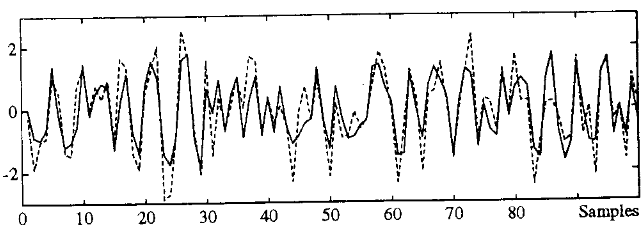

Figure 1 – Noiseless output versus Kalman filtering

Figure 2 compares the noiseless output sequence with the results of Frisch smoothing using the identified model; the standard deviation of the error is 0.638, (slightly) better than that obtained with Kalman filtering.

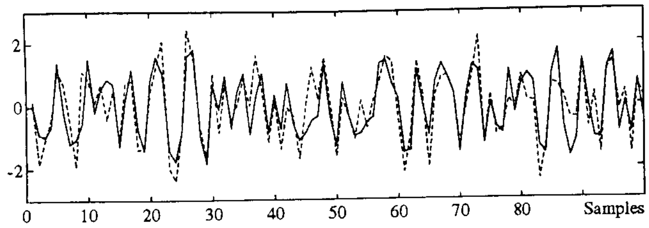

Figure 2 – Noiseless output versus Frisch smoothing

Figure 3 reports the noiseless output (continuous line) and the results of Frisch filtering performed using relations (28) and (29); the standard deviation of the error is 0.639, only slightly worse than the value obtained performing a smoothing on the whole set of data and still better than the value obtained with Kalman filtering.

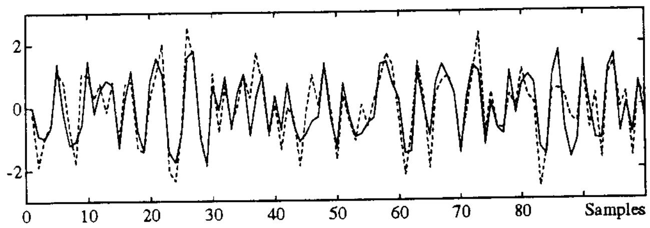

Figure 3 – Noiseless output versus Frisch filtering

6 EXTENSIONS OF FRISCH FILTERING

The filtering technique proposed in this paper is an extension of the smoothing procedure described in [7] and relies on an increasing horizon which is not strictly necessary neither in smoothing nor in filtering. This aspect of the problem can be easily realized observing Figure 4 which shows the entries of a part of M_θ; it is evident how, for every practical purpose, only a limited number of samples before and after a considered instant of time affects smoothed values.

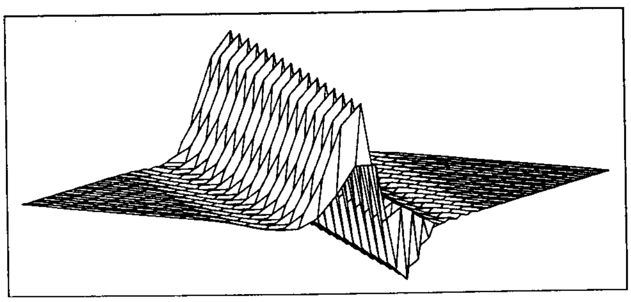

Figure 4 – Entries of (a part of) $M_\theta^+(100)$

These considerations are at the basis of the current development of FFF (Fast Frisch Filtering) algorithms that could conjugate high computational efficiencies with the same performance of extended horizon algorithms.

7 CONCLUDING REMARKS

This paper has described the extension to filtering of a previous smoothing procedure based on the Frisch scheme. The results compare well with both Kalman filtering and Frisch smoothing.

Future developments concern on one side the extension to the multivariable case and, on the other, the development of a new Fast Frisch Filtering algorithm that should exhibit the same properties of the algorithm described here and a very efficient numerical implementation.

References

[1] R. Frisch, "Statistical confluence analysis by means of complete regression systems," Pub. no. 5, Economic Institute, Oslo University, 1934.

[2] S. Beghelli, R. P. Guidorzi and U. Soverini, "The Frisch scheme in dynamic system identification", *Automatica*, vol. 26, no. 1, pp. 171–176, Jan. 1990.

[3] S. Beghelli, P. Castaldi, R. P. Guidorzi and U. Soverini, "On the application of the Frisch scheme identification", *Proc. of 3rd European Control Conference*, pp. 851–855, Rome, September 1995.

[4] R. P. Guidorzi, "Identification of multivariable processes in the Frisch scheme context", MTNS 96, St. Louis, June 1996.

[5] G. C. Goodwin and K. S. Sin, *Adaptive Filtering, Prediction and Control*. Englewood Cliffs, NJ: Prentice–Hall, 1984.

[6] N. Kalouptsidis, *Signal Processing Systems*. John Wiley & Sons, New York, 1997.

[7] R.P. Guidorzi and A. Stoian, "Frisch interpolation of noisy behaviours", 11th International Conference on Control Systems and Computer Science, Bucharest, May 1997.

[8] B. Roorda, "Algorithms for global total least squares modelling of finite multivariable time series", *Automatica*, vol. 31, pp. 391–404, 1995.

[9] B. Roorda and C. Heij, "Global total least squares modeling of multivariable time series", *IEEE Trans. on Automatic Control*, vol. 40, pp. 50–63, 1995.

[10] J.C. Willems, "From time series to linear system, part I: Finite dimensional linear time invariant systems; part II: Exact modelling; part III: Approximate modelling", *Automatica*, vol. 22/23, pp. 561–580, 675–694, 87–115, 1986, 1987.

[11] R. P. Guidorzi, "Invariants and canonical forms for systems structural and parametric identification" *Automatica*, vol. 17, pp. 117–133, 1981.