TRACKING OF SLOWLY VARYING PARAMETERS BY DIRECTIONAL FORGETTING

Rudolf Kulhavý, Miroslav Kárný

Institute of Information Theory and Automation, Czechoslovak Academy of Sciences, Prague, Czechoslovakia

Abstract. The problem of real-time identification of a stochastic (possibly controlled) system with unknown slowly varying parameters is considered. A new technique for parameter tracking is proposed. The basic idea consists in an application of the exponential forgetting only to the marginal (data-conditioned) probability density function on the smallest subspace of parameter space influenced by the current data. Thus, only that piece of information which is substituted by a new one can be forgotten. This idea is applied to the linear normal multivariate regression model and elaborated into algorithmic details. Superiority of the new algorithm, which can be viewed as an improvement of exponentially forgotten least squares, is demonstrated by two simple examples.

Keywords. Adaptive control; Bayes methods; data handling; identification; parameter estimation; time-varying systems.

INTRODUCTION

One popular procedure how to achieve adaptivity of control and/or prediction algorithms can be decomposed in the following steps: i) An underlying model of the solved problem (control, prediction) is identified under the assumption that its unknown parameters are constant. ii) The resulting identification procedure is then modified to achieve the ability to track (relatively) slow changes of parameters.

The present paper describes a new technique for the parameter tracking, which overcomes known numerical problems appearing with identification under insufficiently informative data when commonly used technique of "exponential forgetting" is applied (phenomena as "bursting" or "covariance windup" are typical examples, see e.g. Åström (1980)). The essence of the technique consists in reducing the exponential forgetting only to the smallest subspace of the unknown parameters about which the current data carry a new piece of information. It will be shown that in case of regression-type models this way results in a simple modification of recursive least squares which can be realized by an effective and numerically safe algorithm.

BAYESIAN ESTIMATION AND SLOWLY VARYING PARAMETERS

A stochastic (controlled) system is to be identified. The finite set of data $D_{(t)}$ is measured on it at the time points indexed by t=1,2,... . The directly manipulated data, i.e. system inputs, are collected into a μ-vector $u_{(t)}$, and the rest of $D_{(t)}$ forms a υ-vector of system outputs $y_{(t)}$.

The input-output relation is described by a conditional probability density function (c.p.d.f. in the sequel) known up to a finite set of unknown time-varying parameters $\theta_{(t)}$. In accordance with the Bayesian approach, $\theta_{(t)}$ is taken as a multivariate random variable.

The short-hand notation for the c.p.d.f. of a random variable, say $x_{(t)}$, conditioned on variables $D_{(\tau)}, \tau=1,2,\ldots,k$ and some additional random variable, say $z_{(t)}$, at the point

$$(x_{(t)};D_{(\tau)},\tau=1,2,..,k;z_{(t)})=(x_t;D_\tau,\ \tau=1,2,..,k;z_t) \quad (1)$$

will be used as follows

$$p_{x_{(t)}|D_{(1)},\ldots,D_{(k)},z_{(t)}}(x|D_1,..,D_k,z_t)= =p_{t|k}(x|z) \quad (2)$$

where y,u,... is set instead of x or z.

With this notation the input-output relation is described by

$$p_{t|t-1}(y|u,\theta) \quad (3)$$

The input generator is assumed to satisfy the natural condition of control (Peterka (1981))

$$p_{t|t-1}(u|\theta) = p_{t|t-1}(u) \quad (4)$$

Under this condition Bayesian identification, i.e. the evolution of the c.p.d.f. of $\theta_{(t)}$ conditioned on observed data, is described by

$$p_{t|t}(\theta)\propto p_{t|t-1}(y|u,\theta)p_{t|t-1}(\theta), \quad t=1,2,... \quad (5)$$

The symbol $\propto$ means proportionality, i.e., equality up to the normalizing factor independent of θ.

For time-varying parameters the evolution of the c.p.d.f. $p_{t|t}(\theta)\to p_{t+1|t}(\theta)$ must

be computed to complete the recursion. An exact evaluation of it requires a (probabilistic) model of the time evolution of parameters. When only a rough description "parameters vary slowly" is at disposal, some heuristic procedure has to be used. Since new data are not yet available, the simplest way how to respect the fact that the values of parameters may change in the time interval (t,t+1) is direct increasing the uncertainty of the old "estimate" $p_{t|t}(\theta)$ by the "flattening" operation (Peterka (1981))

$$p_{t+1|t}(\theta) \propto [p_{t|t}(\theta)]^{\phi} \qquad (6)$$

The factor $0<\phi\leq 1$ might depend on t, $D_{(1)},\ldots,D_{(t)}$ and for regression-type models coincides with the known exponential forgetting factor.

GENERAL IDEA OF DIRECTIONAL FORGETTING

There are often some directions in the parameter space in which the marginal c.p.d.f. of unknown parameters remains permanently (or at least for a long time interval) unchanged. The standard exponential forgetting (EF) interpreted in the previous section always flattens the whole c.p.d.f. $p_{t|t}(\theta)$ even under such experimental conditions. As a result, the c.p.d.f. with some "ridge" is formed and even a small change in the informational content of data may cause rapid changes in the shape of this c.p.d.f.

Directional forgetting (DF) proposed in this paper searches for a parameter subspace, say S_1, on which the probability distribution remains unchanged by the recently observed data (S_1 is spanned on the non-excited directions as will be shown below). The usual forgetting is applied only to the c.p.d.f. on the complementary subspace, say S_2. It is necessary to find stochastically independent S_1 and S_2 because only in that case a flattening of one c.p.d.f. does not influence the second one.

More formally, let a regular operator $F_{(t)}$ (depending on $D_{(1)},\ldots,D_{(t)}$) maps the parameter space in two (conditionally) independent subspaces. The generic point θ is transformed into

$$\phi=\{\phi_1,\phi_2\}=F_{(t)}(\theta) \qquad (7)$$

and the conditional independence of the appropriate c.p.d.f.'s is expressed by

$$p_{t|t-i}(\phi)=p_{t|t-i}(\phi_1)p_{t|t-i}(\phi_2), \quad i=0,1 \qquad (8)$$

The last data $D_{(t)}$ do not correct the c.p.d.f. of ϕ_1, therefore

$$p_{t|t}(\phi_1)=p_{t|t-1}(\phi_1) \qquad (9)$$

Apparently, the largest parameter subspace uninfluenced by data must be found, i.e. the cardinality of ϕ_1 has to achieve a maximum. When such a factorization is achieved, the DF method sets

$$p_{t+1|t}(\phi)=p_{t+1|t}(\phi_1)p_{t+1|t}(\phi_2) \propto p_{t|t}(\phi_1)[p_{t|t}(\phi_2)]^{\phi} \qquad (10)$$

After performing this flattening the original parametrization can be, of course, recovered

$$\theta=F^{-1}_{(t)}(\phi) \qquad (11)$$

APPLICATION TO LINEAR NORMAL MULTIVARIATE REGRESSION MODEL

The idea of the DF method will now be elaborated in detail for the linear normal multivariate regression model. A controlled system is described by the input-output relation

$$p_{t|t-1}(y|u,\theta)=N(P^T_{(t)}z_{(t)},\Omega^{-1}_{(t)}) \qquad (12)$$

where $N(\hat{y},\Omega^{-1})$ denotes the normal multivariate c.p.d.f. with a conditional expectation $\hat{y}$ and a positive definite precision matrix Ω. The expectation $\hat{y}$ is supposed to be the product of the matrix of regression coefficients $P_{(t)}$ and the ρ-dimensional regression vector $z_{(t)}$ representing a known function of $D_{(1)},\ldots,D_{(t-1)},u_{(t)}$.

The joint c.p.d.f. for unknown parameters saves the Gauss-Wishart form through the objective step of identification (5). It will be apparent that also the application of the DF technique does not affect it, i.e. supposing the prior c.p.d.f. $p_{1|0}(\theta)$ of this form, we get for t=1,2,...

$$p_{t|t-i}(\theta) \propto \qquad i=0,1 \qquad (13)$$

$$\propto |\Omega|^{\frac{\theta(t|t-i)}{2}} \exp\{-\tfrac{1}{2}\mathrm{tr}(\Omega \begin{bmatrix} -I \\ P \end{bmatrix}^T V_{(t|t-i)} \begin{bmatrix} -I \\ P \end{bmatrix})\}$$

The sufficient statistic characterizing the distribution is formed by a symmetric positive definite $(\upsilon+\rho)\times(\upsilon+\rho)$-matrix

$$V_{(t|t-i)}>0, \quad i=0,1 \qquad (14)$$

and a scalar

$$\theta_{(t|t-i)}>\rho-2, \quad i=0,1 \qquad (15)$$

where the inequalities (14),(15) are valid for any t=1,2,... when a proper prior distribution, i.e. given by statistics fulfilling them, is selected. It can be found from (5) that it holds (see Peterka (1981))

$$\theta_{(t|t)}=\theta_{(t|t-1)}+1 \qquad (16)$$

$$V_{(t|t)}=V_{(t|t-1)}+d_{(t)}d^T_{(t)} \qquad (17)$$

where

$$d^T_{(t)}=[y^T_{(t)},z^T_{(t)}] \qquad (18)$$

Construction of Directional Forgetting

The limited space allows to give only a hint how to find the mapping $F_{(t)}$ with prescribed features. Firstly, the exponent of the c.p.d.f. $p_{t|t-1}(\theta)$ is transformed into a quadratic form with the unit kernel by

$$[\Omega,P^T]\rightarrow F[-I,P^T]G^{-1}_{(t|t-1)}=[\underbrace{Q_\upsilon}_{\upsilon},\underbrace{Q_\rho}_{\rho}]\}\upsilon \qquad (19)$$

where the upper triangular Choleski square roots F,G are used

$$\Omega=F^TF, \quad V^{-1}_{(t|t-1)}=G^T_{(t|t-1)}G_{(t|t-1)} \qquad (20)$$

New data imply a modification of the unit kernel by (at most) 1-rank matrix

$f_{(t|t-1)}f^T_{(t|t-1)}$, where $f_{(t|t-1)}=G_{(t|t-1)}d_{(t)}$.

Rotating of the parameter space by a suitable orthogonal matrix $\Gamma_{\rho(t|t-1)}$

$$[Q_\upsilon, Q_\rho]\rightarrow[Q_\upsilon, Q_\rho\Gamma_{\rho(t|t-1)}]=\Phi \qquad (21)$$

transforms $f_{(t|t-1)}$ into a vector having (at most) $\upsilon+1$ leading nonzero elements. Results obtained after the appropriate forgetting (10) (Φ_2 being formed by (at most) $\upsilon+1$ leading columns of Φ) and the inverse transformation may be shown to be independent of the selected special form $F_{(t)}$.

Results

The form of the c.p.d.f. (13) suits well for the selection of a transformation $F_{(t)}$. However, resulting relations may be better described if the following equivalent form is used. Partitioning $V_{(t|t-i)}$

$$V_{(t|t-i)}=\begin{bmatrix} V_y & V^T_{zy} \\ V_{zy} & V_z \end{bmatrix}\begin{matrix}\}\upsilon\\ \}\rho\end{matrix}_{(t|t-i)}, \quad i=0,1 \qquad (22)$$

and defining

$$C=V_z^{-1},\quad \hat{P}=V_z^{-1}V_{zy},\quad \Lambda=V_y-V^T_{zy}V_z^{-1}V_{zy}, \qquad (23)$$

the relation (13) can be rewritten as

$$p_{t|t-i}(\theta)\propto|\Omega|^{\frac{\theta_{(t|t-i)}}{2}}\exp\{-\tfrac{1}{2}\mathrm{tr}[\Omega(\Lambda_{(t|t-i)}+[P-\hat{P}_{(t|t-i)}]^T C^{-1}_{(t|t-i)}[P-\hat{P}_{(t|t-i)}])]\}$$

$$i=0,1 \qquad (24)$$

The final algorithm describing the recursion of the statistics (23) and θ has two branches according to the value of the quantity

$$\zeta_{(t|t-1)} = z^T_{(t)}C_{(t|t-1)}z_{(t)} \qquad (25)$$

Moreover, the prediction error

$$\hat{e}_{(t|t-1)} = y_{(t)}-\hat{P}^T_{(t|t-1)}z_{(t)} \qquad (26)$$

is used in the evaluation of sufficient statistic under DF.

The regular case: $\zeta_{(t|t-1)}\neq 0$.

$$\theta_{(t+1|t)} = \phi(\theta_{(t|t-1)}-\rho+2)+\rho-1 \qquad (27)$$

$$\Lambda_{(t+1|t)} = \phi(\Lambda_{(t|t-1)}+\frac{\hat{e}_{(t|t-1)}\hat{e}^T_{(t|t-1)}}{1+\zeta_{(t|t-1)}}) \qquad (28)$$

$$\hat{P}_{(t+1|t)} = \hat{P}_{(t|t-1)}+\frac{C_{(t|t-1)}z_{(t)}}{1+\zeta_{(t|t-1)}}\hat{e}^T_{(t|t-1)} \qquad (29)$$

$$C_{(t+1|t)}=C_{(t|t-1)}-\frac{C_{(t|t-1)}z_{(t)}z^T_{(t)}C_{(t|t-1)}}{\varepsilon^{-1}_{(t|t-1)}+\zeta_{(t|t-1)}} \qquad (30)$$

where the weighting factor

$$\varepsilon_{(t|t-1)} = \phi-\frac{1-\phi}{\zeta_{(t|t-1)}} \qquad (31)$$

The singular case: $\zeta_{(t|t-1)}=0$.

$$\theta_{(t+1|t)}=\phi(\theta_{(t|t-1)}-\rho+1)+\rho \qquad (32)$$

$$\Lambda_{(t+1|t)}=\phi(\Lambda_{(t|t-1)}+\hat{e}_{(t|t-1)}\hat{e}^T_{(t|t-1)}) \qquad (33)$$

$$\hat{P}_{(t+1|t)}=\hat{P}_{(t|t-1)},\quad C_{(t+1|t)}=C_{(t|t-1)} \qquad (34)$$

The major and substantial innovation lies in (30). In case of EF the statistic C evolves according to

$$C_{(t+1|t)} = \frac{1}{\phi}(C_{(t|t-1)}-\frac{C_{(t|t-1)}z_{(t)}z^T_{(t)}C_{(t|t-1)}}{1+\zeta_{(t|t-1)}}) \qquad (35)$$

and it is known that at least some elements of C may quickly grow if the available data are noninformative. Note that recently Hägglund (1983) derived a relation similar to Eq. (30), however, under less general conditions and from a different point of view.

GEOMETRICAL INTERPRETATION OF DIRECTIONAL FORGETTING

It is well-known that the algorithm evaluating the sufficient statistic $\theta,\Lambda,\hat{P},C$ formally coincides with recursive least squares (using the EF technique). This coincidence permits to understand the proposed algorithm as a modification of least-squares algorithm.

As mentioned above, the substantial difference between the EF and DF methods consists in a different evaluation of C. To understand it better, let us define a basis of ρ-dimensional Euclidean$_\rho$ space formed by the conjugate vectors $\{a_i\}_{i=1}$

$$a_i^T C_{(t|t-1)}a_j=\delta_{ij};\quad i,j=1,2,\ldots,\rho,$$

$$\delta_{ij}=\begin{cases}1, & i=j\\ 0, & i\neq j\end{cases},\qquad a_1=\frac{z_{(t)}}{\sqrt{\zeta_{(t|t-1)}}} \qquad (36)$$

The evolution of the covariance ellipsoids defined by

$$\zeta_{(t|k)}(x)\triangleq x^T C_{(t|k)}x = \text{const.},\quad t\geq k \qquad (37)$$

for an arbitrary ρ-vector x expressed in the a-basis may be easily evaluated for $(t|t-1)\rightarrow(t|t)\rightarrow(t+1|t)$ using (30) and (35). It follows from the analysis of these ellipsoids that only the diameter of the covariance ellipsoid $\zeta_{(t|t-1)}(x)=\text{const.}$ which is changed by objective data, i.e. the diameter in the direction $z_{(t)}$, is modified (decreased) by DF. On the other hand, using the EF technique all the conjugate diameters are decreased. When a sequence of noninformative data appears, the length of some diameters may converge to zero, in other words, the marginal variances in appropriate directions approach to infinity. The above analysis is illustrated for the two-dimensional case ($\rho=2$) in Fig. 1.

NUMERICAL IMPLEMENTATION

The resulting relations (27)-(31), and (32)-(34) may be algorithmized in many ways. It is, however, well-known that especially recursion (30), as a counter-part of (35)

for ordinary least squares, is numerically sensitive: positive definite matrix $C_{(t+1|t)}$ has to be obtained as a result of subtraction. Square-root version has been found as a successful tool how to deal with this numerical problem. It led us to select LD--factorizations for the implementation.

Let us assume that the LD-factorization of the statistic C is given

$$C_{(t|t-1)}=L_{(t|t-1)}D_{(t|t-1)}L^T_{(t|t-1)} \qquad (38)$$

where $L_{(t|t-1)}$ is a lower triangular matrix with the unit diagonal and $D_{(t|t-1)}$ is a positive diagonal matrix. Using (38) the relation (30) may be rewritten as

$$\tilde{L}\tilde{D}\tilde{L}^T = L\left[D - \frac{Dff^TD}{\varepsilon^{-1}+f^TDf}\right]L^T \qquad (39)$$

where instead of time indices the tilde is used to denote updated matrices and the ρ-vector f is defined by

$$f = L^Tz \qquad (40)$$

LD-factorization of the kernel of the right--hand side of Eq. (39), say $H\tilde{D}H^T$, can be found explicitely because the diagonal matrix is modified only by the 1-rank matrix. Then only the matrix multiplication LH has to be performed to complete the covariance updating (for details see e.g. Bierman (1977)). The final algorithm describing the updating of $C_{(t|t-1)}$ by new data $z_{(t)}$ with the usage of DF is realized by the FORTRAN subroutine LDDIF given in Appendix.

Notice that two tests are employed in this algorithm. If $\zeta<\delta_1$, where δ_1 is quite small, switching off the identification of P is applied, i.e. the computation passes to the singular branch. If $|\varepsilon\zeta|<\delta_2$, where $\delta_2<<1$, the evolution of the statistic C is suppressed because in such a case only unimportant differences of successive values of C appear (this fact may be easily verified from (30)).

The singularity in the computation of $\frac{1}{\varepsilon}$ was consciously left in the algorithm to reach the minimum execution time. It is worth mentioning that the condition $|\varepsilon\zeta|>\delta_2$ ensures $|\varepsilon|>\delta_3$, where δ_3 is the "zero--divide" value, just for

$$\phi > \frac{1+\delta_2}{1+\frac{\delta_2}{\delta_3}} \triangleq \phi_{crit.} \qquad (41)$$

In other case, the explicit test on $|\varepsilon|$ must be added. In practice this is unnecessary because a reasonable choice of δ_2 in the interval $(10^{-6}, 10^{-2})$ for $\delta_3 \approx$ machine zero implies as a rule inapplicably small value of $\phi_{crit.}$.

The formal parameters of the subroutine LDDIF are related to the symbols used as follows:
CALL state: LD = $L_{(t|t-1)}$ together with $D_{(t|t-1)}$ placed on the diagonal, Z=$z_{(t)}$, NZ=$\rho(\geq 2)$, DZ=ϕ, G and H = arbitrary ρ-vectors, AC1= δ_1, AC2= δ_2;
RETURN state: LD=$L_{(t+1|t)}$ together with $D_{(t+1|t)}$ placed on the diagonal, DZ=$\zeta_{(t|t-1)}$, G=$C_{(t|t-1)}z_{(t)}$, H=auxiliary vector, the rest is unchanged.

Numerical demands of the algorithm LDDIF are comparable (or even less) with hitherto used algorithms with EF. Demands on the computer memory increased by ρ memory cells.

Returning from this algorithm the statistics θ and $\hat{P}$ may be evaluated directly according to (27) and (29), or (32) and (34) depending on the value of ζ and with the usage of the vector Cz. The evolution of Λ may be computed by some algorithm based on Choleski or LD-factorization of Λ, or Λ^{-1} (see e.g. Bierman (1977)).

ILLUSTRATIVE EXAMPLES

The following examples demonstrate the effect of the DF technique for two different sources of noninformativity.

Regression with Stopped Changes of Regressor

One-output (υ=1) two-input (μ=2) system was simulated with

$$p_{t|t-1}(y|u,\theta)=N(\beta_1u_{1(t)}+\beta_2u_{2(t)}+k_{(t)},\Omega^{-1})$$

where the unknown parameters took the nominal values

$$P^T_{(t)}=[\beta_1,\beta_2,k_{(t)}]=[1,-1,\sin(\frac{2\pi t}{200})],$$

$$\Omega^{-1}=0.0025$$

and the regressor was $z^T_{(t)}=[u_{1(t)},u_{2(t)},1]$. The inputs $u_{i(t)}$ were mutually independent discrete white noises having N(0,1) distribution up to the time t=100, then $u_{i(t)}=u_{i(100)}$ for t>100. Note that the above model is usually written as $y_{(t)}=P^T_{(t)}z_{(t)}+e_{(t)}$ where $e_{(t)}$ is discrete white noise $N(0,\Omega^{-1})$.

The typical evolution of the parameter estimates and the prediction error for the DF technique with the rather small value ϕ=0.25 and for $C_{(1|0)}=I$, $\hat{P}_{(1|0)}=0$ are drawn in Fig. 2. The evolution of the statistic C stopped in a few steps after the time t=100 (the test $|\varepsilon\zeta|\leq 10^{-6}$ for suppressing the evolution of C was used). As a result of this fact, the vector $\frac{Cz}{1+\zeta}$ was constant in next steps and the point estimates $\hat{P}_i$ copied the prediction error (26) with the appropriate gains given by the initial excitation of system.

Under the same circumstances, even with ϕ much more closer to the unit (e.g. ϕ =0.95), the EF method quickly resulted in a numerical "breakdown" of identification consisting in rapid increasing some elements of the matrix C corresponding to ill-excited directions in the P parameter space and in "going off" of the point estimates $\hat{P}$.

Closed-loop Self-tuning Control with Rare Changes of External Disturbance

Single input (μ=1) single output (υ=1) system with an external disturbance was simulated described by

$$p_{t|t-1}(y|u,\theta)=N(b_0u_{(t)}+b_1u_{(t-1)}+d_1v_{(t-1)}+$$
$$+d_2v_{(t-2)}+a_1y_{(t-1)},\Omega^{-1})$$

where the nominal values of the involved parameters were

$$P^T=[b_0,b_1,d_1,d_2,a_1]=$$
$$=[1.01,\ -1.00,\ 0.10,\ -0.08,\ -0.98],$$
$$\Omega^{-1}=1.0$$

and the regressor was

$$z^T_{(t)}=[u_{(t)},u_{(t-1)},v_{(t-1)},v_{(t-2)},y_{(t-1)}].$$

The external disturbance $v_{(t)}$ was modelled as the random walk with the $N(0,10^{-4})$ - distributed increments. Moreover, two steps occurred in $v_{(t)}$ (from 0 to -1 at t=1 and to +1 at t=10). The simulated system was controlled to the zero setpoint by the self--tuning controller with the predictor P-1A2B2D and the penalty on input increments given by the weight ω=0.01 (see Böhm and others (1984) for details). Initial conditions were $u_{(1)}=1$, $\hat{P}_{(1|0)}=P$ and $C_{(1|0)}=I$, the forgetting factor was ϕ=0.95. Runs for both forgetting methods were performed to compare their behaviour.

The substantially narrower (as was shown above, always bounded) range of changes of the point estimates $\hat{P}$ in case of the DF technique was striking. The appearance of output "bursts" resulted in increasing the sample variance of output to 32.93 for EF (being 1.07 for DF).

CONCLUDING REMARKS

In contrast to most of known methods, the suggested technique covers the whole range of possible data noninformativity sources (as overparametrization, linear feedback, input saturation, rare changes of external variables etc.). This fact has a considerable importance for applications. E.g., the ability to track periodic trends under heavy experimental conditions, demonstrated through the example 1, may be attractive for time series analysis.

The abstract formulation of the idea of directional forgetting in terms of Bayesian approach to system identification admits some straightforward extensions:

- an application to another class of models, e.g. to Markov chain with unknown transition probabilities;
- a combination of the idea "do not suppress old information when new one has not come" with another type of flattening operation than exponential forgetting, e.g. an increase of the conditional covariance matrix of θ by adding some suitably chosen matrix (then forgetting is induced by a random walk on "excited" parameters);
- any extension applicable to ordinary exponential forgetting. as a time and data dependence of the forgetting factor and/or an identification switching similar to Kárný (1982).

Moreover, it can be shown that the technique of directional forgetting may be derived directly as a solution of a rather generally defined optimization problem, which enables to show the close connection with exponential forgetting and to solve the question of adaptive adjustment of the forgetting factor at the same time (Kulhavý (1984)).

REFERENCES

Åström, K.J. (1980). Design principles for self-tuning regulators. In H. Unbehauen (Ed.), Methods and Applications in Adaptive Control. Springer, Berlin. pp. 1-20.

Bierman, G.J. (1977). Factorization Methods for Discrete Sequential Estimation. Academic Press, New York.

Böhm, J., A. Halousková, M. Kárný, and V. Peterka (1984). Simple LQ adaptive controllers. Proc. 9th IFAC Congress, Budapest.

Hägglund, T. (1983). The problem of forgetting old data in recursive estimation. Proc. IFAC Workshop on Adaptive System in Control and Signal Processing, San Francisco.

Kárný, M. (1982). Identification with interruptions as a anti-bursting device. Kybernetika, 18, 320-329.

Kulhavý, R. (1984). Bayesian approach to estimation of slowly varying parameters. Submitted to 6th COMPSTAT Symposium, Prague.

Peterka, V. (1981). Bayesian approach to system identification. In P. Eykhoff (Ed.), Trends and Progress in System Identification. Pergamon Press, Oxford. Chap. 8, pp. 239-304.

APPENDIX

```
      SUBROUTINE LDDIF(LD,Z,NZ,DZ,G,H,AC1,AC2)
      REAL LD(NZ,NZ),Z(NZ),G(NZ),H(NZ)
      FI=DZ
      J=NZ
      G(J)=LD(J,J)*Z(J)
      DZ=G(J)*Z(J)
      H(J)=DZ
      J1=J-1
    1 F=Z(J1)
      DO 2 I=J,NZ
    2 F=F+LD(I,J1)*Z(I)
      GJ=F*LD(J1,J1)
      G(J1)=GJ
      DZ=DZ+F*GJ
      H(J1)=DZ
      J=J1
      J1=J-1
      IF(J1)3,3,1
    3 IF(DZ.LE.AC1) GO TO 9
      HJ=FI*DZ-1.+FI
      IF(ABS(HJ).LE.AC2) GO TO 6
      HJ=DZ/HJ
      J=NZ
      A=H(J)+HJ
      LD(J,J)=LD(J,J)*HJ/A
      J1=J-1
    4 GJ=G(J1)
      C=LD(J1,J1)*A
      B=GJ/C
      A=H(J1)+HJ
      LD(J1,J1)=C/A
      DO 5 I=J,NZ
      C=LD(I,J1)
      LD(I,J1)=C-B*G(I)
    5 G(I)=G(I)+C*GJ
      J=J1
      J1=J-1
      IF(J1)9,9,4
    6 J=NZ
      J1=J-1
    7 GJ=G(J1)
      DO 8 I=J,NZ
    8 G(I)=G(I)+LD(I,J1)*GJ
      J=J1
      J1=J-1
      IF(J1)9,9,7
    9 CONTINUE
      RETURN
      END
```

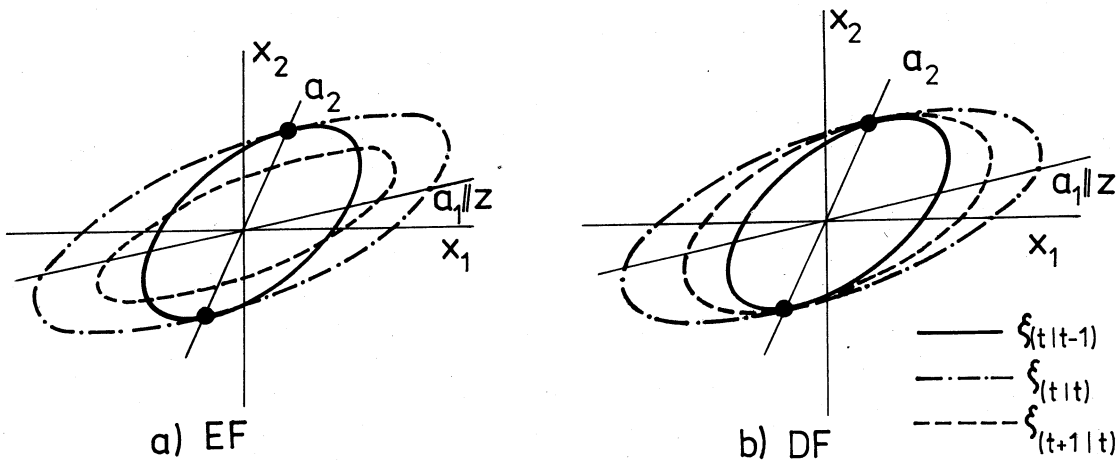

Fig. 1. Comparison of the shape of a covariance ellipse using EF and DF respectively.

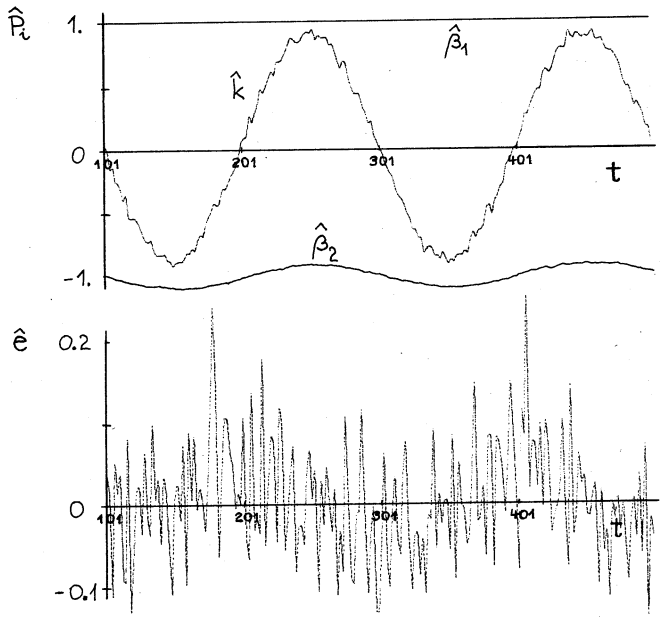

Fig. 2. Tracking of periodic trends using DF.