IFAC Symposium Adaptive Systems in Control and Signal Processing, 1989, Glasgow

Self-tuning control with alternative sets of uncertain process models

V. Peterka
Institute of Information Theory and Automation
Pod vodárenskou věží 4, 182 08 Prague
Czechoslovakia

Abstract. Self-tuning control using an ARMA, or a theoretically equivalent Delta model to represent the controlled process is considered. If the order n of the model, the time delay d, and the parameters of the moving-average part (the $c-$parameters) are chosen a priori, then the remaining parameters of the autoregression and regression part ($a-$ and $b-$parameters), which are left for real-time estimation, determine the set of models which can be covered by the preselected model structure. In the paper more than one prior choices of $\{n, d, c\}$ are considered as uncertain hypotheses. Using the Bayesian approach the probabilities of particular hypotheses, conditioned on the observed data, are determined recursively in real time. Well feasible algorithms are developed for this purpose and demonstrated on simulated examples. It is shown that positional and incremental models can be considered as just special choices of the $c-$parameters.

Keywords: Adaptive control, Self-tuning control, Stochastic control, Direct digital control, Identification, Bayes methods, Parameter estimation, Hypotheses comparison.

1 Introduction

To assume a model structure when identifying a given process means to restrict the set of models which are considered as possible candidates to the set covered by the parameters which are left free for estimation. In self-tuning and adaptive control most often the ARMA models

$$y(t) + \sum_{i=1}^{n} a_i\, y(t-i) = \sum_{i=0}^{n} b_i\, u(t-d-i) + \sum_{i=0}^{n} c_i\, e(t-i) \quad (1)$$

or the equivalent but, for high sampling rates, numerically more robust Delta models (Goodwin, 1985; Peterka, 1986a 1986b)

$$\Delta^n y(t) + \sum_{i=1}^{n} a_i^*\, \Delta^{n-i} y(t-i) = \\ = \sum_{i=0}^{n} b_i^*\, \Delta^{n-i} u(t-d-i) + \sum_{i=0}^{n} c_i^*\, \Delta^{n-i} e(t-i) \quad (2)$$

are used as internal representation of the controlled process. In (2) Δ means the *backwards* difference operator

$$\Delta x(k) = x(k) - x(k-1) \qquad \Delta^i x(k) = \Delta^{i-1} x(k) - \Delta^{i-1} x(k-1)$$

As usually, y and u denote the output and the input of the process while e is a discrete white noise with unknown but constant variance

$$\mathsf{E}[e^2(t)] = \rho \quad (3)$$

The possible process delay is denoted by d and the time indexing of the discrete input u is chosen so that $u(k)$ *precedes* the output sample $y(k)$, ie. $y(k-1)$ is available when $u(k)$ is being generated, but $y(k)$ not. (Some authors use other convention and denote by $u(k-1)$ the input denoted here by $u(k)$.)

It is relatively easy and well feasible to estimate in real time the parameters a and b of ARMA models, or a^* and b^* of Delta models, however under the assumption that the parameters n, d and c or c^* are known or suitably chosen. Unfortunately, no generally valid rules for the prior choice of these parameters are available at present time. Note that the often used and well tried regression models are just a special case of ARMA models with the prior choice $c_i = 0$ for $i \neq 0$. On the other hand, the simulation experiments indicate that, when compared with low order regression models, introduction of apriori chosen $c-$parameters can improve the performance of the self-tuning controller (Clarke and others, 1987). Fortunately, the performance of the controller is usually not very sensitive with respect to this choice and even a relatively rough prior choice may be well sufficient. It is also clear that a certain range of the possible delay d can be covered by the estimated $b-$parameters if the model order n is chosen sufficiently high.

The above discussed experience indicates that it were desirable to consider more than one prior choices of the parameters c, d, n simultaneously and to develop algorithms which could select automatically the most suitable one on the basis of the observed real data. This motivated the research reported in this paper.

The paper is organized in the following manner. In Section 2 some preliminaries are recalled and the problem is formally stated. Its solution is given in Section 3. In Section 4 the relation between positional and incremental models is discussed and it is shown that both these types of models can be covered by two different choices of the $c-$parameters and thus considered and compared in parallel. Simulated examples in Section 5 illustrate the theory.

2 Problem statement

Before the problem can be formally stated some facts have to be recalled. They are selected from (Peterka,1986a 1986b).

Both the ARMA model (1) and the Delta model (2) can be given the following state-space representation

$$A\begin{bmatrix} y(t) \\ s(t) \end{bmatrix} = Hs(t-1) + bu(t-d) + ce(t) \qquad (4)$$

where

$$A = \begin{bmatrix} 1 & 0 \\ a & I_n \end{bmatrix} \qquad H = \begin{bmatrix} I_n \\ 0 \end{bmatrix} + \mu \begin{bmatrix} 0 \\ I_n \end{bmatrix} \qquad (5)$$

$$a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \qquad b = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{bmatrix} \qquad c = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{bmatrix} \qquad (6)$$

For ARMA models $\mu = 0$ while for Delta models $\mu = 1$ and $a_i = a_i^*, b_i = b_i^*, c_i = c_i^*$. Since A is always invertible the model (3) can be decomposed into the more popular state equation

$$s(t) = A_s s(t-1) + b_s u(t-d) + c_s e(t) \qquad (7)$$

and the output equation

$$y(t) = s_1(t-1) + b_0 u(t-d) + c_0 e(t) \qquad (8)$$

with obvious definitions of A_s, b_s and c_s , and with $s_1(k)$ denoting the first component of the state $s(k)$. Note that in case of time delay $d > 0$ $s(t)$ is only the significant part of the state. The full state of the model actually is $\{s(t), u(t-1), \cdots u(t-d)\}$.

Suppose that the observation of the process starts at $t = 1$ and denote the set of input/output data observed up to and including the sampling interval k by

$$\mathcal{D}_k = \{u(1), y(1), u(2), y(2), \cdots, u(k), y(k)\} \qquad (9)$$

and the parameters left for real-time identification by

$$\Theta^T = [a^T, b^T] \qquad (10)$$

Let the prior uncertainty of the parameters Θ and of the initial state $s(0)$ be described by normal and mutually independent probability distributions

$$\mathrm{p}(\Theta) = \mathcal{N}(\hat{\Theta}(0), \rho R_\Theta(0)) \qquad \mathrm{p}(s(0)) = \mathcal{N}(\hat{s}(0), \rho R_s(0)) \qquad (11)$$

where the means $\hat{\Theta}(0)$ and $\hat{s}(0)$ can be understood as prior geusses the uncertainty of which is characterized by the covariance matrices $\rho R_\Theta(0)$ and $\rho R_s(0)$.

If $e(t)$ is assumed to be normal with zero mean and with the variance ρ then the conditional probability density $\mathrm{p}(y(t)|u(t), \mathcal{D}_{t-1}, \rho)$ which does not contain the unknown parameters Θ is also normal

$$\mathrm{p}(y(t)|u(t), \mathcal{D}_{t-1}, \rho) = \mathcal{N}(\hat{y}(t|t-1), \rho \tilde{d}_y(t)) = \\ = \left(2\pi\rho\tilde{d}_y(t)\right)^{-1/2} \exp\left\{-\frac{(y(t) - \hat{y}(t|t-1))^2}{2\rho\tilde{d}_y(t)}\right\} \qquad (12)$$

where the conditional mean

$$\hat{y}(t|t-1) = \mathrm{E}[y(t)|u(t), \mathcal{D}_{t-1}, \rho]$$

can be understood as the one step ahead prediction of the process output. The uncertainty of this prediction is described by the variance

$$\mathrm{E}\left[(y(t) - \hat{y}(t|t-1))^2 | u(t), \mathcal{D}_{t-1}, \rho\right] = \rho\tilde{d}_y(t) \qquad (13)$$

Since all probability distributions involved are normal their evolution can be reduced to algebraic recursions evolving the corresponding conditional means and covariances as reported in (Peterka, 1986a) and in more detail in (Peterka, 1986b). To make the paper self-contained the recursions, by which the observed data are compressed to sufficient statistics determining $\hat{y}(t|t-1)$ and $\tilde{d}_y(t)$, are summarized in Appendix. Here it is important to note that neither $\hat{y}(t|t-1)$ nor $\tilde{d}_y(t)$ depend on ρ which can be unknown and enters all covariances only as a factor.

To formulate the problem discussed informally in Introduction suppose that a finite number N of choices of the parameters $\{n, d, c\}$ has been made but it is not a priori known which one describes the given process best. Hence, N hypotheses

$$H_j : \{n, d, c\} = \left\{n_{(j)}, d_{(j)}, c_{(j)}\right\} \quad j = 1, 2, \cdots N \qquad (14)$$

are given but they are uncertain and it is up to the observed data $\mathcal{D}_t$ to reduce this uncertainty.

Following the Bayesian philosophy the uncertainty can be described by probability distribution and to solve our problem means to determine the probabilities

$$p(H_j|\mathcal{D}_t) = \Pr\left[\{n, d, c\} = \{n_{(j)}, d_{(j)}, c_{(j)}\}|\mathcal{D}_t\right] \quad j = 1, \cdots N \qquad (15)$$

As only N hypothesis are considered as possible candidates it must hold

$$\sum_{j=1}^{N} p(H_j|\mathcal{D}_t) = 1 \qquad (16)$$

Using this relation it is easily obtained

$$\mathrm{p}(H_j|\mathcal{D}_t) = \left[1 + \sum_{i \neq j} r_{ij}(t)\right]^{-1} \qquad (17)$$

where

$$r_{ij}(t) = \frac{p(H_i|\mathcal{D}_t)}{p(H_j|\mathcal{D}_t)} \qquad (18)$$

Since for the given purpose the recursive computation is of interest the problem will be formulated as follows : *Given* $r_{ij}(t-1)$ *and the data pair* $\{u(t), y(t)\}$ *determine* $r_{ij}(t)$.

The problem wil be solved for *natural conditions of control* (Peterka, 1981) which means that when generating $u(t)$ no more information about the hypotheses and the estimated parameters is used then contained in the observed data $\mathcal{D}_{t-1}$ and in priors. Thus it is assumed

$$\mathrm{p}(H_j|u(t), \mathcal{D}_{t-1}) = \mathrm{p}(H_j|\mathcal{D}_{t-1}) \qquad (19)$$

Note that (19) holds automatically when the input $u(t)$ is generated as a deterministic function of $\mathcal{D}_{t-1}$.

3 Solution

Under natural conditions of control the Bayes formula yields

$$\mathrm{p}(H_j|\mathcal{D}_t) = \mathrm{p}(H_j|y(t), u(t), \mathcal{D}_{t-1}) = \\ = \frac{\mathrm{p}(y(t)|u(t), \mathcal{D}_{t-1}, H_j)\ \mathrm{p}(H_j|\mathcal{D}_{t-1})}{\sum_{i=1}^{N} \mathrm{p}(y(t)|u(t), \mathcal{D}_{t-1}, H_i)\ \mathrm{p}(H_i|\mathcal{D}_{t-1})}$$

Hence, using the simplified notation

$$\mathrm{p}_k(y(t)|u(t), \mathcal{D}_{t-1}) = \mathrm{p}(y(t)|u(t), \mathcal{D}_{t-1}, H_k) \qquad (20)$$

the evolution of the probability ratio (18) is given by

$$r_{ij}(t) = \frac{\mathrm{p}_i(y(t)|u(t), \mathcal{D}_{t-1})}{\mathrm{p}_j(y(t)|u(t), \mathcal{D}_{t-1})} r_{ij}(t-1) \qquad (21)$$

This shows that the probability density $p(y(t)|u(t), \mathcal{D}_{t-1})$ has to be determined for each hypothesis. However, only $p(y(t)|u(t), \mathcal{D}_{t-1}, \rho)$ is given by (12). Since the variance ρ is not known and can be different for different hypotheses it has to be eliminated

$$p(y(t)|u(t), \mathcal{D}_{t-1}) = \int p(y(t)|u(t), \mathcal{D}_{t-1}, \rho)\, p(\rho|\mathcal{D}_{t-1})\, d\rho \tag{22}$$

However, this means to determine $p(\rho|\mathcal{D}_{t-1})$, i.e. to estimate (in Bayesian sense) the variance ρ for each of the hypotheses.

From the Bayes formula it follows

$$p(\rho|\mathcal{D}_t) \propto p(y(t)|u(t), \mathcal{D}_{t-1}, \rho)\, p(\rho|\mathcal{D}_{t-1}) \tag{23}$$

where $\propto$ means proportionality (equality up to the factor not depending on ρ). Considering the normal form of the density (12) and introducing the prediction error

$$\varepsilon(t) = y(t) - \hat{y}(t|t-1) \tag{24}$$

it is obtained

$$p(\rho|\mathcal{D}_t) \propto \rho^{-t/2} \exp\left\{-\frac{1}{2\rho}\sum_{k=1}^{t}\frac{\varepsilon^2(k)}{\tilde{d}_y(k)}\right\}\ p(\rho)$$

This shows that the self-reproducing form of $p(\rho|\mathcal{D}_t)$ is

$$p(\rho|\mathcal{D}_t) = \frac{1}{\Gamma(\tau(t)/2)}\left(\frac{\lambda(t)}{2}\right)^{\tau(t)/2} \times \\ \times \rho^{-(\tau(t)/2+1)} \exp\left\{-\frac{\lambda(t)}{2\rho}\right\} \tag{25}$$

where $\Gamma(.)$ is the gamma function and

$$\tau(t) = \tau(t-1) + 1 \tag{26}$$

$$\lambda(t) = \lambda(t-1) + \frac{\varepsilon^2(t)}{\tilde{d}_y(t)} \tag{27}$$

The prior uncertainty of ρ is then characterized by $\tau(0)$ and $\lambda(0)$ which, in case of very high prior uncertainty of ρ, have to be chosen very small.

After substituting (25) into (22) and performing the integration it is obtained that the uncertainty of the one-step ahead prediction of the process output under lack of knowledge of the variance ρ and the parameters Θ is described by the Student distribution

$$p(y(t)|u(t), \mathcal{D}_{t-1}) = \frac{\Gamma(\tau(t)/2)}{\Gamma(\tau(t-1)/2)} \times \\ \times \left(\pi\tilde{d}_y(t)\lambda(t-1)\right)^{-1/2}\left(1 + \frac{(y(t) - \hat{y}(t|t-1))^2}{\tilde{d}_y(t)\lambda(t-1)}\right)^{-\tau(t)/2} \tag{28}$$

which can be written also in the following form more suitable for the present purpose

$$p(y(t)|u(t), \mathcal{D}_{t-1}) = \frac{\Gamma(\tau(t)/2)}{\Gamma(\tau(t-1)/2)} \times \\ \times \left(\pi\tilde{d}_y(t)\right)^{-1/2} \frac{\lambda(t-1)^{\tau(t-1)/2}}{\lambda(t)^{\tau(t)/2}} \tag{29}$$

Summing up we come to the following

Result 1 *Under the assumptions of Section 2 and assuming the same priors $\tau(0)$ and $\lambda(0)$ for all hypotheses the evolution of the probability ratio $r_{ij}(t) = p(H_i|\mathcal{D}_t)/p(H_j|\mathcal{D}_t)$ for growing t is*

$$r_{ij}(t) = \kappa_{ij}(t) r_{ij}(t-1) \tag{30}$$

where the factor $\kappa_{ij}(t)$ is determined by the following relations

$$\varepsilon_{(k)}(t) = y(t) - \hat{y}_{(k)}(t|t-1) \tag{31}$$

$$\lambda_{(k)}(t) = \lambda_{(k)}(t-1) + \frac{\varepsilon^2_{(k)}(t)}{\tilde{d}_{y,(k)}(t)} \tag{32}$$

$$\tau(t) = \tau(t-1) + 1 \tag{33}$$

$$\kappa_{ij}(t) = \left(\frac{\tilde{d}_{y,(j)}(t)}{\tilde{d}_{y,(i)}(t)}\right)^{1/2}\left(\frac{\lambda_{(i)}(t-1)}{\lambda_{(j)}(t-1)}\right)^{\tau(t-1)/2}\left(\frac{\lambda_{(j)}(t)}{\lambda_{(i)}(t)}\right)^{\tau(t)/2} \tag{34}$$

where $\hat{y}_{(k)}(t|t-1)$ and $\tilde{d}_{y,(k)}(t)$ mean $\hat{y}(t|t-1)$ and $\tilde{d}_y(t)$ under the hypothesis H_k.

Remark 1. To reduce the influence of the priors associated with each of the compared hypotheses it may be suitable to start the comparing at some time instant $t_0 > 0$ and to choose $r_{ij}(t_0) = 1$, i.e. $p(H_k|\mathcal{D}_{t_0}) = 1/N$ for all k. Then it follows from (30) and (34) that the probability ratio $r_{ij}(t)$ for $t > t_0$ can be calculated according to the following formulae.

$$r_{ij}(t) = (\Phi_{ij}(t))^{1/2}\left(\frac{\lambda_{(j)}(t)}{\lambda_{(i)}(t)}\right)^{\tau(t)/2} \tag{35}$$

$$\Phi_{ij}(t) = \frac{\tilde{d}_{y,(j)}(t)}{\tilde{d}_{y,(i)}(t)}\Phi_{ij}(t-1) \tag{36}$$

$$\Phi_{ij}(t_0) = \left(\frac{\lambda_{(i)}(t_0)}{\lambda_{(j)}(t_0)}\right)^{\tau(t_0)/2} \tag{37}$$

4 Incremental vs. positional models

Both the ARMA model (1) and the Delta model (2) are, in general, positional models. This means that the zero levels for input and output signals have to be chosen in mutual correspondence. For a steady-state zero input the steady-state mean value of the output has to be zero. If this cannot be guaranteed with sufficient accuracy an absolute term has to be added and estimated as an additional parameter of the model. For instance, the often used regression model of this type reads

$$y(t) = -\sum_{i=1}^{n} a_i y(t-i) + \sum_{i=1}^{n} b_i u(t-d-i) + k_0 + e(t) \tag{38}$$

In some industrial applications, especially when the process is corrupted by nonstationary drifts or unpredictable load changes, the incremental models appear more appropriate (Peterka, 1984; Clarke and others, 1987).

$$y(t) = y(t-1) - \sum_{i=1}^{n} a_i \Delta y(t-i) + \sum_{i=1}^{n} b_i \Delta u(t-d-i) + e(t) \tag{39}$$

Taking the difference of (38)

$$\Delta y(t) = -\sum_{i=1}^{n} a_i \Delta y(t-i) + \sum_{i=1}^{n} b_i \Delta u(t-d-i) + e(t) - e(t-1) \tag{40}$$

the absolute term k_0 disappears and it is seen that the positional model (38) can be considered as an incremental ARMA model (sometimes called ARIMA or CARIMA) with special choice of the c-parameters, namely $c_0 = 1$ and $c_1 = -1$.

Generally, incremental ARMA models cover positional ARMA models if the factor $(1 - \zeta)$ is incorporated into the polynomial

$$c(\zeta) = c_0 + c_1\zeta + c_2\zeta^2 + \cdots + c_n\zeta^n \tag{41}$$

The similar holds, of course, for Delta models where to represent a positional model by an incremental one means

Appendix

Summary of real-time parameter and state estimation algorithms

Consider the state representation (4) of an ARMA or Delta model. If the incremental variant of the model is of interest replace in (4) and in the sequal $y(t)$ by $\Delta y(t)$ and $u(t)$ by $\Delta u(t)$.

Let the prior uncertainty of the estimated parameters Θ and of the initial state $s(0)$ be described by normal and mutually independent probability densities (11). Then the joint probability distribution for the uncertain parameters Θ, ordered according to (10), and the state $s(t)$, conditioned on the data $\mathcal{D}_t$ (9) and for given (preselected as a hypothesis) n, d and c, is also normal

$$p(\Theta, s(t)|\mathcal{D}_t) = \mathcal{N}\left(\begin{bmatrix} \hat{\Theta}(t) \\ \hat{s}(t|t) \end{bmatrix}, \rho R_{\Theta,s}\right) \qquad (42)$$

If all of the parameters a and b are estimated, then the number of estimated parameters is $m = 2n+1$. Introduce the following statistics :

- A rectangular $(n \times (m+1))$-matrix $X(t)$ partitioned in the following two ways

$$X(t) = \underbrace{[-Y(t)}_{n+1}, \underbrace{U(t)]}_{n+1} = \underbrace{[-Y^*(t)}_{1}, \underbrace{Z(t)]}_{m} \qquad (43)$$

  where $Y^*(t)$ is the first column of $Y(t)$.

- A lower triangular $(n \times n)$-matrix $L_s(t)$ with unit diagonal elements and a diagonal $(n \times n)$-matrix $D_s(t)$ (actually a vector of n diagonal elements).

- A lower triangular $(m \times m)$-matrix $L_\Theta(t)$ with unit diagonal elements and a diagonal $(m \times m)$-matrix $D_\Theta(t)$.

Then the positive definite matrix $R_{\Theta,s}(t)$, proportional to the joint covariance matrix of (42), can be written in the following factorized LDL^T form

$$R_{\Theta,s}(t) = \begin{bmatrix} L_\Theta(t) & 0 \\ Z(t)L_\Theta(t) & L_s(t) \end{bmatrix} \times \begin{bmatrix} D_\Theta(t) & 0 \\ 0 & D_s(t) \end{bmatrix} \begin{bmatrix} L_\Theta^T(t) & L_\Theta^T(t)Z^T(t) \\ 0 & L_s^T(t) \end{bmatrix} \qquad (44)$$

$$\hat{s}(t|t) = X(t) \begin{bmatrix} 1 \\ \hat{\Theta}(t) \end{bmatrix} \qquad (45)$$

Suppose that the statistics $L_s(t-1), D_s(t-1), X(t-1), L_\Theta(t-1), D_\Theta(t-1)$ are given from the previous step of the recursion or, for $t = 1$ as priors with $X(0) = 0$. Then the statistics can be updated with respect to the newly observed input-output pair $\{u(t), y(t)\}$ according to the following procedure.

[1] Using the given c and the relation

$$[c\ ,\ HL_s(t-1)] \begin{bmatrix} 1 & 0 \\ 0 & D_s(t-1) \end{bmatrix} \begin{bmatrix} c^T \\ L_s^T(t-1)H^T \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \tilde{c}(t) & L_s(t) \end{bmatrix} \begin{bmatrix} d_y(t) & 0 \\ 0 & D_s(t) \end{bmatrix} \begin{bmatrix} 1 & \tilde{c}^T(t) \\ 0 & L_s^T(t) \end{bmatrix} \qquad (46)$$

determine $L_s(t), D_s(t), \tilde{c}(t)$ and $d_y(t)$.

[2] Construct the row vector $z(t)$ as the first row $Z_1(t-1)$ of the matrix $Z(t-1)$ modified so that $u(t-d)$ is added to the element with the column index $(n+1)$, i.e. to the element which meets $\hat{b}_0$ in (45).

$$z_j(t) = Z_{1,j}(t-1) \quad j \neq n+1 \qquad (47)$$

$$z_{n+1}(t) = Z_{1,n+1}(t-1) + u(t-d) \qquad (48)$$

and determine the prediction error

$$\varepsilon(t) = y(t) - \hat{y}(t|t-1) = y(t) + Y_1^*(t-1) - z(t)\hat{\Theta}(t-1) \qquad (49)$$

[3] Given $d_y(t)$ from step [1] and $z(t)$ from step [2] update $L_\Theta(t-1)$ and $D_\Theta(t-1)$ according to the relations

$$f(t) = z(t)L_\Theta(t-1) \qquad (50)$$

$$\begin{bmatrix} 1 & f(t) \\ 0 & L_\Theta(t-1) \end{bmatrix} \begin{bmatrix} d_y(t) & 0 \\ 0 & D_\Theta(t-1) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ f^T(t) & L_\Theta^T(t-1) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ g_\Theta(t) & L_\Theta(t) \end{bmatrix} \begin{bmatrix} \tilde{d}_y(t) & 0 \\ 0 & D_\Theta(t) \end{bmatrix} \begin{bmatrix} 1 & g_\Theta^T(t) \\ 0 & L_\Theta^T(t) \end{bmatrix} \qquad (51)$$

and using the gain vector $g_\Theta(t)$ produced by (51) update the parameter estimates

$$\hat{\Theta}(t) = \hat{\Theta}(t-1) + g_\Theta(t)\varepsilon(t) \qquad (52)$$

The relation (51) also produces $\tilde{d}_y(t)$ which can be used, together with the prediction error (49) for comparing the hypotheses according to the Result 1 of Section 3.

[4] The vector $\tilde{c}(t)$ produced in step [1] is used to update the submatrices $Y(t-1)$ and $U(t-1)$ of the statistic $X(t-1)$.

$$Y(t) = (\mu I_n + C(t))Y(t-1) + C^*(t)y(t) \qquad (53)$$

$$U(t) = (\mu I_n + C(t))U(T-1) + C^*(t)u(t-d) \qquad (54)$$

where

$$C(t) = \left[-\tilde{c}(t) \,\middle|\, \begin{matrix} I_{n-1} \\ 0 \end{matrix}\right] \qquad C^*(t) = \left[-\tilde{c}(t) \,\middle|\, I_n\right] \qquad (55)$$

where $\mu = 0$ for ARMA case and $\mu = 1$ for Delta case. Note that in (53) and (54) $u(t)$ and $y(t)$ are scalars by which the matrix $C^*(t)$ is multiplied.

[5] To obtain the state estimate $\hat{s}(t|t)$ use (45) .

Remark a. The equality (46) expresses two possible ways how a given nonnegative definite matrix can be factorized. Note that the right hand side is an LDL^T factorization. This modification of factorization can be performed in a numerically save and efficient way using the algorithm of dyadic reduction (Peterka 1986a, 1986b). The same holds also for the equality (51).

Remark b. The time varying vector $\tilde{c}(t)$ produced in the first step of the recursion converges for growing t to $\tilde{c}(\infty)$ which defines the polynomial

$$\tilde{c}(\varsigma) = 1 + \tilde{c}_1\varsigma + \tilde{c}_2\varsigma^2 + \cdots + \tilde{c}_n\varsigma^n \qquad (56)$$

If the given polynomial (41) has all roots outside or on the unit circle then

$$\tilde{c}(\varsigma) = \frac{1}{c_0}c(\varsigma) \qquad (57)$$

When some of the roots lie on the unit circle and $D_s(0) \gg 0$ then the approaching to (57) goes from outside the unit circle. It should be emphasized that to pass this transient part of the filtering (53) and (54) is essential in this case which is of special interest as explained in Section 4. If the given polynomial $c(\varsigma)$ had some of its roots inside the unit circle than the obtained $\tilde{c}(\varsigma)$ is the reflection of $c(\varsigma)$ which guarantees the stability of the filters (53) and (54).

Remark c. The relations (50), (51), and (52) are nothing else than a numerically reliable realization of recursive least squares, similar to Bierman's UD filter. However, it is important to note how the regressor $z(t)$ and the regressant $(y(t) + Y_1^*(t-1))$ have to be prepared.

to set $c_n^* = 0$. However, it should be emphasized that the asymptotic polynomial theory cannot handle these cases and for process identification, namely for the necessary prefiltering of processed data, more general algorithms listed in Appendix have to be employed. It should be also noted that, especially in case of fast sampling, the c-polynomials (41) with root lying close to the unit circle, typically with factors close to $(1+\varsigma)$, are rather a rule than an exception.

The Result of Section 3 makes it possible to consider both incremental and positional models in parallel (for the price of doubled computational load, of course) or can serve as an objective tool for making the choice on the basis of real data.

5 Simulated examples

The purpose of the first example is to demonstrate that the case of regulation, when the setpoint of a self-tuning controller is fixed and the goal is to compensate unobservable random disturbances, is more sensitive with respect to the prior choice of the $c-$parameters than the case of a servo, when the closed control loop is well excited by the external command signal. The controlled process was simulated using the model

$$(1 - 0.85\varsigma)y(t) = (0.06 + 0.09\varsigma)u(t) + 0.15v(t)$$

where ς is the backwards shift operator (often denoted by z^{-1} or q^{-1}) and $\{v(t)\}$ is a strongly correlated noise generated as the output of a filter driven by the discrete, approximately normal white noise $\{e(t)\}$.

$$(1 - 0.92\varsigma)v(t) = 0.2(1 + \varsigma)e(t)$$

Hence, the overall second order ARMA model is

$$a(\varsigma)y(t) = b(\varsigma)u(t) + c(\varsigma)e(t)$$

$$a(\varsigma) = (1 - 0.85\varsigma)(1 - 0.92\varsigma) = 1 - 1.77\varsigma + 0.782\varsigma^2$$

$$b(\varsigma) = (0.06+0.09\varsigma)(1-0.92\varsigma) = 0.06+0.0348\varsigma+0.0828\varsigma^2$$

$$c(\varsigma) = 0.03(1 + \varsigma)$$

Simulating the situation when it is a priori known that the neighboring noise samples are strongly correlated the following two hypotheses were considered as a priori equally likely.

$$H_1: \quad n = 2,\ d = 0,\ c(\varsigma) = 1 + \varsigma$$

$$H_2: \quad n = 2,\ d = 0,\ c(\varsigma) = 1 + 0.9\varsigma$$

Note that H_1 was actually true. The selftuning controller started with priors $\hat{s}(0) = 0$, $L_s(0) = I_2$, $D_s(0) = 10^{10}I_2$, $\hat{\Theta}(0) = [0,0,0,0,1]$, $L_\Theta = I_5$, $D_\Theta(0) = 10^4 I_5$ for the both hypotheses.

In Fig.1 the result of simulation is recorded for the case of regulation and for $\mathsf{E}[e^2(t)] = 0.2$. The first 45 samples were used as the passive learning phase of the controller, however, the comparison of the hypotheses was switched on right from the start. In Fig.2 the same controller with the same priors was applied in servo mode for $\mathsf{E}[e^2] = 0.02$. Compare the evolution of the conditional probability of H_1 in these two operating modes. The evolutions of the parameter estimates, given in the attached tables, clearly show that in regulation mode all free parameters were properly estimated from the beginning of identification while in servo mode the parameters of the deterministic component of the simulated process were identified first and the less important parameters of the stochastic component were incorporated only relatively slowly.

The second example has been chosen to demonstrate the behaviour of the algorithms in case when the assumptions, on which their derivation is based, are not fulfilled. Only the more difficult and more interesting case of regulation is reported here. The following continuous stochastic process was simulated

$$y = \frac{4(s + 0.75)}{3(s + 2)(s + 1)(s + 0.5)}\, u + \frac{s + 3}{3(s + 2)(s + 1)(s + 0.5)}\, v$$

$$v = \frac{1}{s + 0.2}\, e$$

The self-tuning regulator sampled the process output with the sampling period of 0.5 sec. The drifting character and the intensity of the noise, which is to be compensated, is seen in Fig.3 where, besides the controlled output, also the uncontrolled output is plotted. Note that the order of the corresponding overall ARMA model is 4. A second order incremental model was used to represent the controlled process and the following three hypotheses were considered as potential candidates.

$H_1: \ n = 2, d = 0, c(\varsigma) = 1$ (purely incremental model)

$H_2: \ n = 2, d = 0, c(\varsigma) = 1+\varsigma$ (strongly correlated noise)

$H_3: \ n = 2, d = 0, c(\varsigma) = 1 - \varsigma$ (positional model)

Three self-tuning controllers, based on the three above hypotheses, ran in parallel (using instead of ARMA the equivalent Delta representation of the controlled process) but only the process input generated by the controller with the instantly highest probability of the associated hypothesis was actually applied. Both the self-tuning regulation and the hypotheses comparison was active from the start of the process and all three controllers used the same priors as in the first example. The result of the experiment is documented in Fig.3. Note that the positional model (H_3) was eliminated very fast. The hypothesis H_2 was winning in the most part of the self-tuning process but finally it appeared to be practically equivalent to H_1 in spite of the seemingly drastic difference in the assumed $c-$parameters.

References

Clarke D.W.,C.Mohtadi and P.S.Tuffs (1987) . Generalized predictive control.*Automatica 23*,39–50

Goodwin G.G. (1985) . Some observations on robust estimation and control. *7th IFAC/IFORS Symposium on Identification and System Parameter Estimation*, York, UK, Preprints Vol.1, 851–859

Peterka, V. (1981) . Bayesian Approch to System Identification. In: (P. Eykhoff, ed.) *Trends and Progress in System Identification*. Chapter 8, Pergamon Press

Peterka, V. (1986a) . Algorithms for LQG self-tuning control based on input-output Delta models. *IFAC Workshop on Adaptive Systems in Control and Signal Processing*, Lund, Sweden

Peterka, V. (1986b) . *Control of Uncertain Processes : Applied Theory and Algorithms*. Supplement to the journal Kybernetika, *22*, No 3–6, 102 pp.

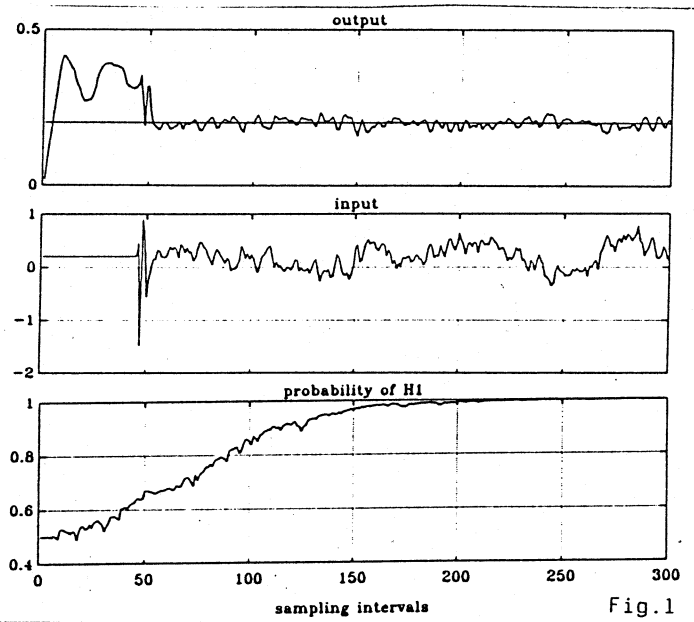

Fig.1 Example 1 - regulation mode

Evolution of parameter estimates — regulation						
time		$\hat{a}_1$	$\hat{a}_2$	$\hat{b}_0$	$\hat{b}_1$	$\hat{b}_2$
100	H_1	-1.70	0.726	0.064	0.042	-0.077
	H_2	-1.71	0.731	0.064	0.041	-0.078
200	H_1	-1.76	0.781	0.064	0.039	-0.081
	H_2	-1.77	0.785	0.063	0.039	-0.082
400	H_1	-1.76	0.777	0.062	0.036	-0.084
	H_2	-1.76	0.778	0.061	0.037	-0.084
true		-1.77	0.782	0.060	0.0348	-0.0828

Table 1: Example 1 — regulation mode

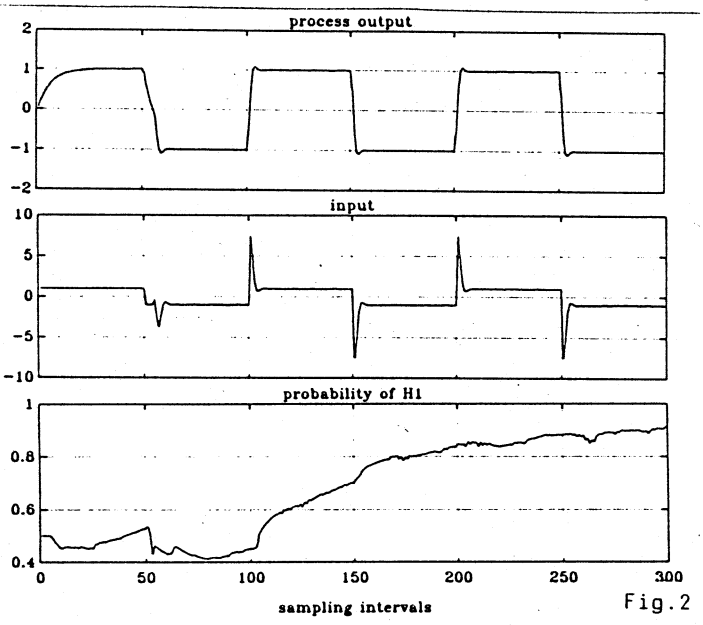

Fig.2 Example 1 - servo mode

Evolution of parameter estimates — servo						
time		$\hat{a}_1$	$\hat{a}_2$	$\hat{b}_0$	$\hat{b}_1$	$\hat{b}_2$
100	H_1	-0.82	-0.02	0.060	0.092	0.003
	H_2	-0.86	0.006	0.060	0.090	-0.001
200	H_1	-1.07	0.189	0.060	0.077	-0.020
	H_2	-1.00	0.220	0.060	0.075	-0.023
400	H_1	-1.55	0.596	0.060	0.048	-0.063
	H_2	-1.57	0.613	0.060	0.047	-0.065
1600	H_1	-1.68	0.707	0.060	0.040	-0.074
	H_2	-1.69	0.717	0.060	0.039	-0.076
true		-1.77	0.782	0.060	0.0348	-0.0828

Table 2: Example 1 — servo mode

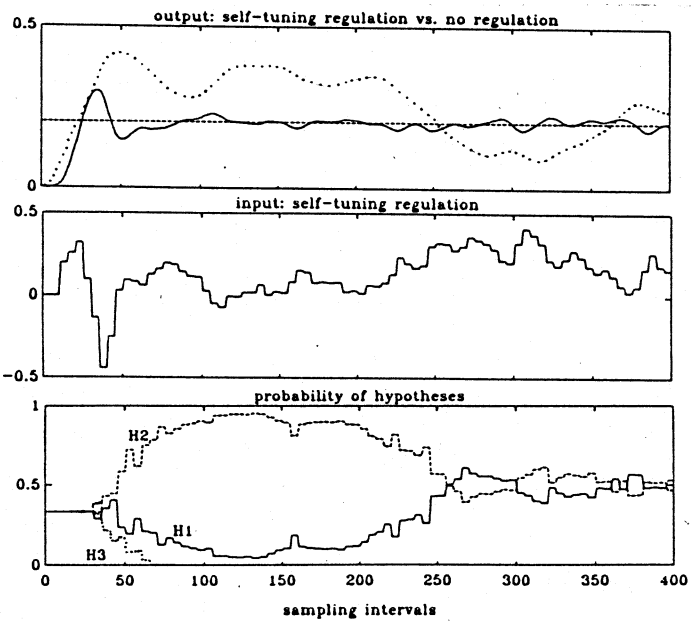

Fig.3 Example 2