Kybernetika

Miroslav Kárný

Quantification of prior knowledge about global characteristics of linear normal model

Kybernetika, Vol. 20 (1984), No. 5, 376--385

Persistent URL: http://dml.cz/dmlcz/124485

KYBERNETIKA — VOLUME 20 (1984), NUMBER 5

QUANTIFICATION OF PRIOR KNOWLEDGE ABOUT GLOBAL CHARACTERISTICS OF LINEAR NORMAL MODEL

MIROSLAV KÁRNÝ

Bayesian approach to system identification requires from the user to collect and express his prior information about the identified system. The paper presents a way in which a global prior knowledge can be quantified. Two simple cases, pulse response smoothness and knowledge of the static gain of a system modelled by mixed autoregressive-regressive model, are elaborated in detail. These cases which are of practical importance illustrate main steps in the quantification.

The significant and favourable effect of this built-in prior information on the start of self-tuning control is demonstrated on an example.

1. INTRODUCTION

The systematic Bayesian approach to the system identification has been developed by Peterka and described in [10]. Several papers have been published concerning different aspects of this theory (see e.g. [5], [6], [7], [8], [9]). The present paper belongs to this series. It tries to describe the way in which a prior knowledge about global characteristics of an identified system can be incorporated into the prior distribution. Two special cases of practical importance are elaborated in detail. The treated pulse response smoothness and knowledge of the static gain are sufficiently simple and at the same time make it possible to demonstrate the crucial steps of the procedure. The illustrative example supports practical importance of the presented solution.

2. CONJUGATE PRIOR PROBABILITY DENSITY FUNCTION FOR NORMAL REGRESSION MODELS

To make the paper self-contained and to fix the notation used some standard results are summarized here. The relation between the vector input $u_{(t)}$ and the ν-dimensional vector output $y_{(t)}$ is described by the conditional probability density function (c.p.d.f.)

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

This represents the c.p.d.f. of the output $y_{(t)}$ at point y under condition that the data $y_{(\tau)}$ and $u_{(\tau)}$, $\tau = 1, 2, \ldots, t - 1$ have been observed, the input $u_{(t)}$ has been chosen and the finite dimensional vector of unknown parameters takes the value θ.

The special case of a normal multivariate regression model having the c.p.d.f.

(2) $$p_{t|t-1}(y \mid u, \theta) = N(P^{\mathrm{T}} z_{(t)}, R) = \\ = \left|2\pi R\right|^{-1/2} \exp\left\{-\tfrac{1}{2}(y - P^{\mathrm{T}} z_{(t)})^{\mathrm{T}} R^{-1}(y - P^{\mathrm{T}} z_{(t)})\right\}$$

with unknown parameters $\theta = \{P, R\}$ is assumed. P is the (ϱ, v) real matrix of the regression coefficients and $R > 0$ is the (v, v) positive definite symmetric covariance matrix. The regressor ϱ-vector $z_{(t)}$ is a known function of data in the condition of the c.p.d.f. (1). A prior knowledge is quantified by the conjugate Gauss-Wishart p.d.f. [3]

(3) $$p(\theta) \propto (\left|R \otimes C\right|)^{-\vartheta/2} \exp\left\{-\tfrac{1}{2} \operatorname{tr}\left[R^{-1}((P - \bar{P})^{\mathrm{T}} C^{-1}(P - \bar{P}) + \Lambda)\right]\right\}$$

The quantities $\bar{P}$, C, Λ and ϑ are at disposal to express prior information about P and R. The above mentioned p.d.f. possesses computationally advantageous property to be self-reproducing. The aposterior p.d.f. of θ has the same form (3) being parametrized by the multivariate sufficient statistic $\bar{P}_{(t)}$, $C_{(t)}$, $\Lambda_{(t)}$ and $\vartheta_{(t)}$. $\bar{P}$ and $\bar{R} = \Lambda/(\vartheta - \varrho)$ $(\vartheta > \varrho)$ are the mathematical expectation of P and R respectively, the best point estimates of P and R for a rather wide class of optimality criteria. C is a positive definite symmetric matrix which determines the covariance matrix of P as the Kronecker product of $\bar{R}$ and C, $\bar{R} \otimes C$, namely

(4) $$\mathsf{E}(P_{ij} - \bar{P}_{ij})(P_{kl} - \bar{P}_{kl}) = C_{ik} \,.\, \bar{R}_{jl} \quad i, k = 1, 2, \ldots, \varrho \quad j, l = 1, 2, \ldots, v$$

The above interpretation gives a direct hint how to choose $\bar{P}$ and $\bar{R}$. The best prior guess of P and R has to be selected. We shall omit the choice of ϑ, which is important for description of R only, $\vartheta = \varrho + v$ can be used. It implies that only the matrix C has to be chosen. This choice forms the core of our paper.

We shall start (Section 3) with a discussion of commonly used diagonal form of C. It will permit to show the weak point of this choice and to find further improvements (see Sections 4,5).

3. CONFIDENCE INTERVAL AS A GUIDE FOR THE CHOICE OF C

The commonly used choice of C is

(5) $$C = \operatorname{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_\varrho^2)$$

The form (5) corresponds to the assumption that the rows of P are uncorrelated i.e. independent in the present case. This choice with commonly used diagonal $\bar{R}$ implies independence of all entries in P. It follows that we assign the same probability (the same degree of belief) to realistic as well as unrealistic coefficient configurations. Moreover in connection with least square interpretation of evolution of the statistic $\bar{P}, C, \bar{R}, \vartheta$, see e.g. [9], rather big values of σ_i's are often recommended (near to the inverse of computational precision). Such a choice coincides with almost uniform prior distribution i.e. with the case when little is a apriori known about P. It, of course, results in drastical changes of the c.p.d.f. of θ through the transient part of recursive identification. Any careful choice of $\bar{P}$ is then meaningless. However, in self-tuning control rapid changes of point estimates can cause short-time instabilities in the closed loop. This observation and interpretation (4) led to the conclusion that in this "diagonal" case σ_i's have to be chosen so that the confidence intervals (at some high significance level) coincide with practically expected range of coefficient values. It is known that in the present case the confidence intervals take the form

$$(6) \qquad \langle -\alpha \sqrt{(\bar{R}_{jj})}\, \sigma_i + \bar{P}_{ij}\,, \quad \bar{P}_{ij} + \alpha \sqrt{(\bar{R}_{jj})}\, \sigma_i \rangle \quad i = 1, 2, \ldots, \varrho \quad j = 1, 2, \ldots, \nu$$

where

$$(7) \qquad \alpha \approx 2 \div 3$$

This choice substantially improves transient behaviour of identification. The danger caused by a wrong selection of these intervals is decreased substantially in practical applications when exponential [8] or directional [7] forgetting is used.

The improvement which results from choosing σ_i's with intervals (6) being the expected range of possible values, however, does not overcome inconsistency of the assumed independence of different entries of P.

4. SYSTEMS WITH SMOOTH PULSE RESPONSE

Let us assume the pulse-response model

$$(8) \qquad y_{(t)} = \sum_{i=0}^{n} H_i u_{(t-i)} + v_{(t)}$$

where the random term $v_{(t)}$ is taken as a normal random walk, a rather realistic and simple model of drifts. The process $v_{(t)}$ is generated by

$$(9) \qquad v_{(t)} = v_{(t-1)} + e_{(t)}$$

where

$$(10) \qquad p_{t|t-1}(e \mid u, \theta) = N(0, R)$$

i.e. $e_{(t)}$ is normal white noise. On such a model a simple selftuning controller can be based [1]. It is an easy excersise to show that the relations (8), (9), (10) determine a model of the form (2) for the increment $\Delta y_{(t)} = y_{(t)} - y_{(t-1)}$ instead of $y_{(t)}$ with

(11) $$P^{\mathrm{T}} = [H_n, H_{n-1}, \ldots, H_0]$$

(12) $$Z_{(t)}^{\mathrm{T}} = [\Delta u_{(t-n)}^{\mathrm{T}}, \Delta u_{(t-n+1)}^{\mathrm{T}}, \ldots, \Delta u_{(t)}^{\mathrm{T}}]$$

Rather often it can be expected that the parameters H_i of the model (8) have been obtained by sampling of *smooth* pulse response.

This means that the second differences

(13) $$\delta_i = H_{i+2} - 2H_{i+1} + H_i \quad i = 0, 1, 2, \ldots, n - 2$$

are apriori expected to be rather near to zero. If smoothness is the only prior information at disposal, the range of possible values of δ_i is substantially narrower than that of $H - \bar{H}$.

It is of course reasonable to select $\bar{P}^{\mathrm{T}} = [\bar{H}_n, \ldots, \bar{H}_0]$ consistently so that

(14) $$0 = \bar{H}_{i+2} - 2\bar{H}_{i+1} + \bar{H}_i \quad i = 0, 1, \ldots, n - 2$$

From equations (13), (14) it can be seen that

(15) $$\delta_i = (H_{i+2} - \bar{H}_{i+2}) - 2(H_{i+1} - \bar{H}_{i+1}) + (H_i - \bar{H}_i) \quad i = 0, 1, \ldots, n - 2$$

To replace H_i fully by δ_i's it remains to define also δ_{n-1} and δ_n. Assuming the finite length of the pulse response it is taken for sure that $H_i = \bar{H}_i = 0$ for $i > n$. Then it is natural to introduce

(16) $$\delta_{n-1} = -2(H_n - \bar{H}_n) + (H_{n-1} - \bar{H}_{n-1})$$

(17) $$\delta_n = H_n - \bar{H}_n$$

Relations (15), (16) and (17) can be rewritten in the matrix form

(18) $$L^{-1}(P - \bar{P}) = \delta$$

where

(19) $$L^{-1} = \begin{bmatrix} I_v & & & & \\ -2I_v & \ddots & & & \\ I_v & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & I_v & -2I_v & I_v \end{bmatrix}$$

I_v being the (v, v) unit matrix and

(20) $$\delta^{\mathrm{T}} = [\delta_n^{\mathrm{T}}, \delta_{n-1}^{\mathrm{T}}, \ldots, \delta_0^{\mathrm{T}}]$$

Definitions (15), (16), (17), (20) imply that

(21) $$\mathsf{E}[\delta] = 0$$

Let us assume that smoothness is the only prior information available i.e. δ_t,

$i = 0, 1, \ldots, n$ are taken to be independent. To maintain the simple self-reproducing form of the c.p.d.f. (3) we shall assume that

(22) $$\mathsf{E}\delta_i\delta_i^{\mathrm{T}} = \gamma^2\bar{R} \quad i = 0, 1, 2, \ldots, n$$

where the factor γ^2 has to be specified. It can be chosen on the basis of the confidence intervals similar to (6).

The above considerations determine

$$\mathsf{E}\delta\delta^{\mathrm{T}} = \bar{R} \otimes \begin{bmatrix} \gamma^2 I_v & & & & & \\ & \gamma^2 I_v & & & & \\ & & \gamma^2 I_v & & & \\ & & & \cdot & & \\ & & & & \cdot & \\ & & & & & \gamma^2 I_v \end{bmatrix} = \bar{R} \otimes \bar{I}^{\gamma}$$

Using (18) and (23) we have

(24) $$L^{-1}C(L^{-1})^{\mathrm{T}} = \bar{I}^{\gamma}$$

We shall write down the solution of (24) in terms of the LD-factorisation of C, suitable for initialization of numerically stable LD-factorization based filtering (in the style of UD-factorisation of Bierman [2])

(25) $$C = LDL^{\mathrm{T}}$$

L is a lower triangular matrix with unit diagonal and the matrix D is a diagonal one. It is apparent that

(26) $$D = \bar{I}^{\gamma}$$

and

(27) $$L = \begin{bmatrix} I_v & & & \\ -2I_v & \ddots & & \\ I_v & \ddots & \ddots & \\ & \ddots & \ddots & \ddots \\ & I_v & -2I_v & I_v \end{bmatrix}^{-1} = \begin{bmatrix} I_n & & & & \\ 2I_v & \ddots & & & \\ 3I_v & \ddots & \ddots & & \\ \vdots & \ddots & \ddots & \ddots & \\ (n+1)I_v & \cdots & 3I_v & 2I_v & I_v \end{bmatrix}$$

Quantification of the prior information about the smoothness of the pulse response $\{H_i\}$ can be summarized as follows:

- choose the best prior gues $\{\bar{H}_i\}$ of $\{H_i\}$ having the second order differences equal to zero (14)
- start up recursive estimation with L given by (27).

The entries of D have to reflect uncertainties of the second order differences $H_{i+2} - 2H_{i+1} + H_i$, the choice $D_i = \gamma^2$ for all i is usually well satisfactory.

Notice that the simplicity of the present case follows from the fact the smoothness of the pulse-response model can be expressed in terms of linear relations the number of which is sufficient to determine the structure of the matrix C uniquely.

The next section can be taken as an example how to proceed when relations between the entries of P are nonlinear and the number of relations is smaller than the number of the P-entries.

5. INFORMATION ABOUT STATIC GAIN

The value of static gain is a piece of information which is often at disposal. We shall deal with this problem restricting ourselves to single-input single-output systems in order to simplify the exposition. The system is assumed to be described by the mixed autoregressive-regressive model

$$y_{(t)} = \sum_{i=1}^{n} a_i y_{(t-i)} + \sum_{i=0}^{n} b_i u_{(t-i)} + e_{(t)} \tag{28}$$

where $e_{(t)}$ is assumed to be normal and independent of the past input-output history, i.e.

$$p_{t|t-1}(e \mid u, \theta) = N(0, R) \tag{29}$$

The assumption (29) together with (28) gives $p_{t|t-1}(y \mid u, \theta)$ in the form (2) with

$$P^{\mathrm{T}} = [b_0, b_1, \ldots, b_n, a_1, a_2, \ldots, a_n], \quad \varrho = 2n + 1 \tag{30}$$

and

$$z_{(t)}^{\mathrm{T}} = [u_{(t)}, u_{(t-1)}, \ldots, u_{(t-n)}, y_{(t-1)}, y_{(t-2)}, \ldots, y_{(t-n)}] \tag{31}$$

The static gain g of the model (28) is given by

$$g = \frac{\sum_{i=0}^{n} b_i}{1 - \sum_{i=1}^{n} a_i} \tag{32}$$

We shall linearize this nonlinear function of parameters at the value given by $\bar{P}$. The linear term of the Taylor expansion gives

$$\delta = g - \bar{g} = \frac{1}{1 - \sum_{i=1}^{n} \bar{a}_i} l_0^{\mathrm{T}}[P - \bar{P}] \tag{33}$$

where $\bar{a}_i$ and $\bar{b}_i$ are the entries of $\bar{P}$ corresponding to (30)

$$\bar{g} = \frac{\sum_{i=0}^{n} \bar{b}_i}{1 - \sum_{i=1}^{n} \bar{a}_i} \tag{34}$$

and

(35) $$l_0^{\mathrm{T}} = [\underbrace{1, 1, \ldots, 1}_{n+1}, \underbrace{\bar{g}, \bar{g}, \ldots, \bar{g}}_{n}]$$

Apparently δ defined by (33) has zero prior expectation. If we denote

(36) $$\sigma_\delta^2 = \frac{\mathrm{E}\delta^2}{\bar{R}} \frac{1}{(1 - \sum_{i=1}^{n} \bar{a}_i)^2}$$

where $\bar{R}$ is the prior expectation of R, then the relations (33) and (35) imply that

(37) $$l_0^{\mathrm{T}} C l_0 = \sigma_\delta^2$$

Moreover the value σ_δ^2 determines the range in which the guess of the gain given by $l_0^{\mathrm{T}} P$ can be found with sufficiently high probability.

Of course, the number of degrese of freedom in determining C only in accordance with (37) is too high. We shall complete the definition of C through the following reasoning: Let us take, for a fixed C, a C-orthogonal basis $l_0, l_1, \ldots, l_{\varrho-1}$ of the ϱ dimensional Euclidian space i.e.

(38) $$l_i^{\mathrm{T}} C l_j = 0 \quad \text{for} \quad i \neq j \quad i, j = 0, 1, 2, \ldots, \varrho - 1$$

(39) $$l_i^{\mathrm{T}} C l_i = \bar{\sigma}^2 \quad \text{where} \quad \bar{\sigma}^2 \gg \sigma_\delta^2 \quad i = 1, 2, \ldots, \varrho - 1$$

Expressing any ϱ dimensional vector $\tilde{l}$ in this basis

(40) $$\tilde{l} = \alpha_0 l_0 + \sum_{i=1}^{\varrho-1} \alpha_i l_i$$

we shall find that

(41) $$\mathrm{E}[\tilde{l}^{\mathrm{T}}(P - \bar{P})] = 0$$

and

(42) $$\mathrm{E}[\tilde{l}^{\mathrm{T}}(P - \bar{P})]^2 = [\sigma_\delta^2 \alpha_0^2 + \bar{\sigma}^2 \sum_{i=1}^{\varrho-1} \alpha_i^2] \bar{R}$$

The expression $\tilde{g} = \tilde{l}^{\mathrm{T}} P$ can be interpreted as a guess that the system has another static gain $\tilde{g}$. The error of this guess $\tilde{l}^{\mathrm{T}}(P - \bar{P})$ has the greater dispersion the greater is Euclidian norm of the projection of $\tilde{l}$ onto the subspace spanned by the directions C-orthogonal to l_0. We shall require these C-orthogonal directions to be also "geometrically" dissimilar. It is reasonable to take this dissimilarity as the I-orthogonality. Then the condition (37) can be taken as a definition of a direction in which the covariance elipsoid of P has significantly greater diameter. Relations (37), (38), (39) for fixed l_i $i = 0, 1, 2, \ldots, \varrho - 1$ represent $\varrho(\varrho + 1)/2$ linear equations for the same number of independent entries of a symmetric matrix C. By the direct inspection it can be verified that C, uniquely determined by (37), (38), (39), takes the form

$$C = \sigma_\delta^2 l_0 [l_0^T l_0]^{-2} l_0^T + \bar{\sigma}^2 (I - l_0 (l_0^T l_0)^{-1} l_0^T) \tag{43}$$

which is positive definite and independent of the special form of l_i $i = 1, \ldots, \varrho - 1$. The simple form of C admits to determine explicitly the LD-factorisation of it. This C can be rewritten as

$$C = \bar{\sigma}^2 \left[I - \frac{l_0 l_0^T}{\omega^2} \right] \tag{44}$$

where (using (35) and (43))

$$\omega^2 = \frac{(l_0^T l_0)^2}{l_0^T l_0 - \left(\frac{\sigma_\delta}{\bar{\sigma}}\right)^2} = \frac{(n + 1 + \bar{g}^2 n)^2}{n + 1 + \bar{g}^2 n - \left(\frac{\sigma_\delta}{\bar{\sigma}}\right)^2} \tag{45}$$

Using the definition of LD-factorisation and denoting

$$s_i = \omega^2 - \sum_{j=1}^{i} l_{0j}^2 = s_{i+1} + l_{0i+1}^2 \quad i = \varrho, \varrho - 1, \ldots, 1 \tag{46}$$

with

$$s_{\varrho+1} = \left(\frac{\sigma_\delta}{\bar{\sigma}}\right)^2 \Bigg| \left(1 - \frac{\sigma_\delta^2}{\bar{\sigma}^2 (n + 1 + n\bar{g}^2)}\right) \tag{47}$$

it can be found that

$$D_j = \bar{\sigma}^2 \frac{s_{j+1}}{s_j} \quad j = \varrho, \varrho - 1, \ldots, 1 \tag{48}$$

$$L_{ij} = -\frac{l_{0i} l_{0j}}{s_{j+1}} \quad j = \varrho, \varrho - 1, \ldots, 1; i < j \tag{49}$$

where l_{0i} is the i-th entry of l_0 (35)

6. ILLUSTRATIVE EXAMPLE

Simple self-tuning regulator I-nB [1] using the model (8), (9) with $v = 1$, $n = 6$ was simulated on a hybrid computer. The output of a continuous system having three identical time constants was regulated under the influence of unmeasurable disturbances.

Two typical runs are documented. The dotted lines denote the uncontrolled system output which is the same in both cases. Figure 1 shows the closed loop behaviour when a "diagonal" initialisation is used ($C = I\gamma^2$, $\gamma^2 = 1$, $\bar{P} = 0$). Figure 2 present the result obtained by using information about the smoothness of the pulse response (LD-decomposition according to (26), (27) is used with $\gamma^2 = 1$, $\bar{P} = 0$).

The similar influence has been observed for different values γ^2. Essentially the same improvement has been achieved when the information about the static gain was exploited for the P-$2A$ $3B$ self-tuning regulator [1].

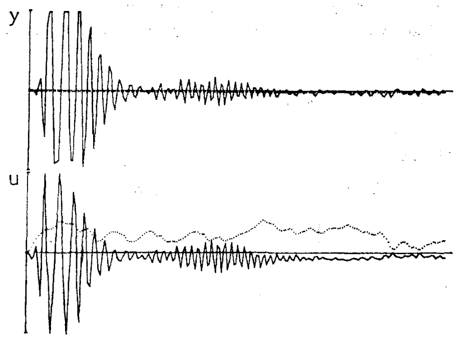

Fig. 1. Closed loop behaviour with $C = I$, $\bar{P} = O$.

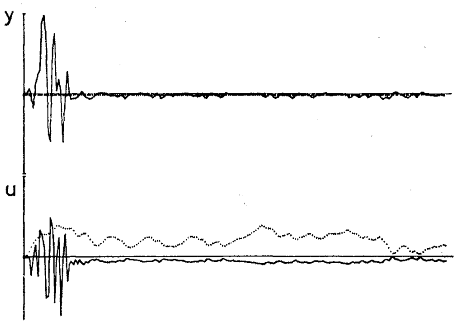

Fig. 2. Closed loop behaviour with L according to (27) and $D = I$, $\bar{P} = O$.

7. CONCLUDING REMARKS

Two practically important special cases of quantifying prior information have been elaborated in detail. It is known that the importance of a careful handling of the prior information is rather high when either limited amount of informative data is at disposal or the results of recursive identification are used in the subsequent decision problem. Typical examples of this kind are the one-shot identification with costly data or the start of self-tuning controllers.

Information about smooth character of the impulse response has been used in an algorithmically similar way in [4] where a surprisingly strong influence of it has been demonstrated on identification of the step-response.

Naturally, the cases discussed in the paper do not cover the whole range of possibilities. However, it has been outlined how to deal with superfluous degrees of freedom and with a nonlinear form of relations between the parameters.

(Received October 24, 1983.)

REFERENCES

[1] J. Böhm, A. Halousková, M. Kárný and V. Peterka: Simple LQ selftuners. Preprints 9th IFAC World Congress, Budapest 1984.

[2] G. J. Biermann: Factorization Methods for Discrete Sequential Estimation. Academic Press, New York 1977.

[3] M. H. De Groot: Optimal Statistical Decision. McGraw-Hill, New York, 1970.

[4] B. Hanuš: Ein einfaches Programm zur Ermittlung von Übergangscharakteristiken. In: "XVIII. Interactionales Ilmenau", Ilmenau 15.—19. 10. 1973. Ilmenau, TH 1973, No. 3, 25—27.

[5] M. Kárný: Bayesian estimation of model order. Problems Control Inform. Theory *9* (1980), 1, 33—46.

[6] M. Kárný: Algorithms for determining the model structure of a controlled system. Kybernetika *19* (1983), 2, 164—178.

[7] R. Kulhavý and M. Kárný: Tracking of slowly varying parameters by directional forgetting. Preprints 9th IFAC World Congress, Budapest 1984.

[8] V. Peterka: Subjective probability approach to real-time identification. Preprints 4th IFAC Symposium on Identification and System Parameter Estimation, Tbilisi, USSR, *3* (1976), vol. 3, 83—96.

[9] V. Peterka: Experience accumulation for decision making in multivariate time series. Problems Control Inform. Theory *7* (1978), 3, 143—159.

[10] V. Peterka: Bayesian approach to system identification. In: Trends and Progress in System Identification (P. Eykhoff, ed.), IFAC Series for Graduates, Research Workers and Practicing Engineers *1* (1981).

Ing. Miroslav Kárný, CSc., Ústav teorie informace a automatizace ČSAV (Institute of Information — Czechoslovak Academy of Sciences), Pod vodárenskou věží 4, 182 08 Praha 8. Czechoslovakia.