KYBERNETIKA — VOLUME *18* (1982), NUMBER 1

RECURSIVE PARAMETER ESTIMATION OF REGRESSION MODEL WHEN THE INTERVAL OF POSSIBLE VALUES IS GIVEN

MIROSLAV KÁRNÝ

The paper presents aposteriori maximum likelihood estimation of coefficients of regression model when prior bounds on their possible values are given. Its objective is twofold:
— conceptually clarify the problem in which the difference between experience accumulation and decision-making became apparent
— to describe modification of Fletcher-Jackson algorithm [1] which admits recursive parameter estimation in apriori given interval.

1. INTRODUCTION

The main feature of the Bayesian approach to the system identification [2] is its capability to feed prior information into the problem.

The simplest but important prior information is that about the range S of possible values of the unknown parameters Θ. This type of information can be, in principle, introduced in a very simple way:

The support of the prior probability density function (which exists in the case assumed) $p(\Theta)$ coincides with S i.e.

$$p(\Theta) \propto \mathscr{L}(\Theta, 0)\, \chi_S(\Theta) \tag{1}$$

where the symbol $\propto$ means proportionality, the function $\chi_S(\Theta)$ is the indicator of the set S and $\mathscr{L}(\Theta, 0)$ is a nonnegative function for which

$$0 < I_{(0)} = \int_S \mathscr{L}(\Theta, 0)\, \mathrm{d}\Theta < \infty\,. \tag{2}$$

It is known, however, that only relatively narrow range of identification problems can be exactly converted into practically feasible algorithms.

The one of important exceptions is the linear-in-parameters multivariable normal regression model (briefly fundamental model) [3] with conjugate prior distribution

function [4], [5]. In this case the Bayesian identification leads to the algorithm which formally coincides with the (recursive) least squares.

This coincidence often results into the opinion that Bayesian approach is something superfluous, at least for the set of linear-in-parameters models. But assuming the case of the fundamental model with bounded range of possible values of parameters we arrive to different solutions using different interpretations of the standard least squares (different projection see e.g. [6], [7]).

We shall try to clarify the source of this diversity.

It can be easily found that in the bounded range case also for fundamental model the full Bayesian solution is algorithmically unfeasible. There are at least two ways to overcome this trouble:

To select conjugate prior distribution which is practically equal to zero out of the range S.

This is often satisfying solution. But as an apriori unknown value of the likelihood function is decisive for what can be taken as zero probability event, this solution may be dangerous, when point estimates are needed.

To restrict oneself to the case of point estimates maximizing the aposteriori likelihood function.

The main part of this paper deals with this problem.

2. PROBLEM FORMULATION AND PRELIMINARIES

Let us assume that a controlled system can be described by multivariate linear-in-parameters normal regression model (briefly fundamental model)

$$p(y_{(t)} \mid d^{(t-1)}, u_{(t)}, \Theta) = N(\hat{y}_{(t)}, \Omega^{-1}) \tag{1}$$

where we denote

$t \in \{1, 2, \dots\}$ – discrete time

$p(\alpha \mid \beta)$ – probability density function of α conditioned on β

$y_{(t)}$ – ν-vector of outputs

$u_{(t)}$ – μ-vector of inputs

$d_{(t)} = (y_{(t)}, u_{(t)})$ – measured data

$d^{(t)} = (d_{(1)}, d_{(2)}, \dots, d_{(t)})$ – process history

$N(\hat{y}, \Omega^{-1}) = \left|\frac{\Omega}{2\pi}\right|^{1/2} \exp\{-\frac{1}{2}(y - \hat{y})^T \Omega (y - \hat{y})\}$

$\hat{y}_{(t)} = P^T z_{(t)}$ – regression function

$z_{(t)} = z(d^{(t-1)}, u_{(t)})$ – known ϱ-dimensional function

$\Theta = (P, \Omega)$ – unknown system parameters

We assume that

$$\Theta \in S = \{(P, \Omega) : \underline{P} \leqq P \leqq \overline{P}, \Omega > 0\} \tag{2}$$

where the given bounds $\underline{P}$, $\overline{P}$ satisfy

$$-\infty < \underline{P} < \overline{P} < +\infty \tag{3}$$

and vector inequalities in (2) and (3) are interpreted entrywise. The inequality $\Omega > 0$ means that Ω is a positive definite matrix.

Possible controllers are restricted to the class which fulfils natural conditions of control [2] i.e. explicit values of unknown parameters cannot be used

$$p(u_{(t)} \mid d^{(t-1)}, \Theta) = p(u_{(t)} \mid d^{(t-1)}) .$$

Assuming conjugate prior distribution restricted to S, we obtain the following solution of the Bayesian system identification: the aposteriori probability density function of unknown parameters is given by

$$p(\Theta \mid d^{(t)}) = \frac{\mathscr{L}(\Theta, t)\, \chi_S(\Theta)}{I_{(t)}} \tag{4}$$

where ("conditional") likelihood function evolves according to

$$\mathscr{L}(\Theta, t) = \mathscr{L}(\Theta, t-1)\, p(y_{(t)} \mid d^{(t-1)}, u_{(t)}, \Theta) \tag{5}$$

$\mathscr{L}(\Theta, 0)$ is given in (1.1), (1.2) and

$$I_{(t)} = \int \mathscr{L}(\Theta, t)\, \chi_S(\Theta)\, \mathrm{d}\Theta = \int_S \mathscr{L}(\Theta, t)\, \mathrm{d}\Theta \tag{6}$$

Bayesian prediction is given by

$$p(y_{(t)} \mid d^{(t-1)}, u_{(t)}) = \frac{I_{(t)}}{I_{(t-1)}} \tag{7}$$

Specializing relations (4) to (7) for the system model (1) we have

$$\begin{aligned} \mathscr{L}(\Theta, t) &= \left|\frac{\Omega}{2\pi}\right|^{\vartheta(t)/2} \exp\left\{-\frac{1}{2}\operatorname{tr}\left(\Omega \begin{bmatrix} -I \\ P \end{bmatrix}^T V_{(t)} \begin{bmatrix} -I \\ P \end{bmatrix}\right)\right\} = \\ &= \left|\frac{\Omega}{2\pi}\right|^{\vartheta(t)/2} \exp\left\{-\tfrac{1}{2}\operatorname{tr}\left[\Omega(P^T V_{z(t)} P - P^T V_{zy(t)} - V_{zy(t)}^T P + V_{y(t)})\right]\right\} = \\ &= \left|\frac{\Omega}{2\pi}\right|^{\vartheta(t)/2} \exp\left\{-\tfrac{1}{2}\operatorname{tr}\left[\Omega((P - \hat{P}_{(t)})^T V_{z(t)} (P - \hat{P}_{(t)}) + A_{(t)})\right]\right\} \end{aligned} \tag{8}$$

where

$$\vartheta_{(t)} = \vartheta_{(t-1)} + 1 \tag{9}$$

$$V_{(t)} = \underbrace{\begin{bmatrix} V_{y(t)} & V_{zy(t)}^T \\ V_{zy(t)} & V_{z(t)} \end{bmatrix}}_{\begin{matrix} \nu & \varrho \end{matrix}} \begin{matrix} \} \nu \\ \} \varrho \end{matrix} = V_{(t-1)} + \begin{bmatrix} y_{(t)} \\ z_{(t)} \end{bmatrix} \begin{bmatrix} y_{(t)} \\ z_{(t)} \end{bmatrix}^T \tag{10}$$

$$\hat{P}_{(t)} = V_{z(t)}^{-1} V_{zy(t)} = \hat{P}_{(t-1)} + \frac{C_{(t-1)} z_{(t)}}{1 + z_{(t)}^T C_{(t-1)} z_{(t)}} \hat{e}_{(t)}^T \tag{11}$$

$$\hat{e}_{(t)} = y_{(t)} - \hat{P}_{(t-1)}^T z_{(t)} \tag{12}$$

$$\Lambda_{(t)} = V_{y(t)} - V_{zy(t)}^T V_{z(t)}^{-1} V_{zy(t)} = \Lambda_{(t-1)} + \frac{\hat{e}_{(t)} \hat{e}_{(t)}^T}{1 + z_{(t)}^T C_{(t-1)} z_{(t)}} \tag{13}$$

$$C_{(t)} = V_{z(t)}^{-1} = C_{(t-1)} - \frac{C_{(t-1)} z_{(t)} z_{(t)}^T C_{(t-1)}}{1 + z_{(t)}^T C_{(t-1)} z_{(t)}} \tag{14}$$

Recursive parts of the relations (11), (13), (14) are implied by the relation (10). The entities (9) and (10) form sufficient statistic of the problem assumed, their recursion results from Eq. (5) and initial conditions from Eq. (1.1). The condition (1.2) is satisfied for $\vartheta_{(0)} \geqq 0$, $V_{(0)} > 0$.

The form of the recursive least squares (11), (12), (13), (14) is nothing more than the alternative recursion for equivalent sufficient statistic $(\vartheta, \hat{P}, \Lambda, C)$. It implies immediately the following conclusions:

- Information contained in measured data is compressed in statistic $(\vartheta, \hat{P}, \Lambda, C)$ and any prior information cannot destroy (i.e. singularily transform) this statistic.
- When the statistic $\hat{P}_{(t)}$ is the admissible estimate of P i.e.

$$\underline{P} \leqq \hat{P}_{(t)} \leqq \overline{P} \tag{15}$$

then maximum likelihood estimate $\hat{P}_{ML(t)}$ coincides with this "least square" estimate.

- The prediction (7) is always influenced by bounds $\underline{P}$, $\overline{P}$, even when (15) is fulfilled. Apparently computation of the predictive probability density (7) is a difficult problem which can be practically solved only for small ϱ [8] and will be omitted here.
- Maximum likelihood estimates of the parameters Θ are given:

$$\hat{\Omega}_{ML} = \vartheta \left[(\hat{P}_{ML} - \hat{P})^T V_z (\hat{P}_{ML} - \hat{P}) + \Lambda \right]^{-1} \tag{16}$$

$$\hat{P}_{ML} = \underset{\underline{P} \leqq P \leqq \overline{P}}{\arg\min} \left\{ \operatorname{tr} \left[\hat{\Omega}_{ML} (P^T V_z P - P^T V_{zy} - V_{zy}^T P) \right] \right\} \tag{17}$$

It follows that for multi-output systems ($\nu > 1$) the estimation of Ω and P cannot be separated as in unbounded case. Probably some sort of bootstrap construction can be successfully used, but we restrict ourselves to the single-output case ($\nu = 1$) in which such splitting appears, i.e.

$$\hat{P}_{ML} = \underset{\underline{P} \leqq P \leqq \overline{P}}{\arg\min} \left[\tfrac{1}{2} (P^T V_z P - 2 P^T V_{zy}) \right]. \tag{18}$$

Algorithmic solution of the problem (18) is contained in the next part.

3. FLETCHER-JACKSON ALGORITHM

Searching for maximum likelihood estimates ($v = 1$) we solve the sequence of the following problems

$$\min_{\underline{P} \leqq P \leqq \overline{P}} \left[\tfrac{1}{2}(P^T V_{z(t)} P - 2V_{zy(t)}^T P)\right] \quad t = 1, 2, \ldots \tag{1}$$

We assume that the regular prior distribution is selected and sufficient condition for validity of (1.2) is fulfilled

$$V_{z(0)} > 0, \quad \vartheta_{(0)} \geqq 0. \tag{2}$$

It follows that

$$V_{z(t)} > 0 \quad \text{for all} \quad t. \tag{3}$$

Under the assumption

$$-\infty < \underline{P} < \overline{P} < +\infty \tag{4}$$

the next theorem can be easily proved.

Theorem 1. The solution of the problem (1), (3), (4) exists and it is unique.

Let us describe briefly the algorithm, which Fletcher and Jackson [1] propose for solution of problem (1), (3), (4) in the fixed time t.

The theoretical behavior of this algorithm [1] comprimes

Theorem 2. The algorithm described below needs finite number of iterations.

Algorithm.

1) Set $r = 0$, select initial admissible approximation

$$\underline{P} \leqq P^r \leqq \overline{P} \tag{5}$$

and compute gradient of the minimized function

$$g(P^r) = V_z P^r - V_{zy}. \tag{6}$$

Sort P^r in two parts

$$P^r = \begin{bmatrix} {}^1P \\ {}^2P \end{bmatrix} \begin{matrix} \} k \\ \} \varrho \end{matrix} \tag{7}$$

where

$$\underline{P}_i < P_i < \overline{P}_i \quad \text{and} \quad g_i(P^r) = 0 \quad \text{for} \quad i = 1, 2, \ldots, k \quad \text{(free variables)} \tag{8}$$

$$P_i = \underline{P}_i \quad \text{or} \quad P_i = \overline{P}_i \quad \text{or} \quad \underline{P}_i < P_i < \overline{P}_i \quad \text{and} \quad g_i(p^r) \neq 0$$
$$\text{for } i = k + 1, k + 2, \ldots, \varrho \quad \text{(fixed variables)} \tag{9}$$

2) Select some fixed variable P_l, $l > k$ which will be changed by a step α. Admissible signs of changes are apparently restricted

$$\alpha > 0 \quad \text{if} \quad P_l = \bar{P}_l \tag{10}$$
$$\alpha < 0 \quad \text{if} \quad P_l = \underline{P}_l$$

We take such step size that free variables (8) can be changed in admissible set in such a way that

$$g_i(P^{r+1}) = 0 \quad \text{for} \quad i = 1, 2, \ldots, k \tag{11}$$

It can be easily found that we can achieve a smaller value of the minimized function iff for admissible α (7,) (8) such l exists that

$$g_l(P^r)\,\alpha < 0 \tag{12}$$

i.e. if such l and admissible α does not exist optimal solution is achieved.

The optimality conditions described above are apparently specialized form of Kuhn-Tucker conditions.

3) If we find some l which does not fulfil Kuhn-Tucker conditions we change P^r to P^{r+1} according to the rules

$$P_l^{r+1} = P_l^r + \alpha \tag{13}$$

$$P_i^{r+1} = P_i^r \qquad i > k\,, \quad i \neq l \tag{14}$$

$$P_i^{r+1} = P_i^r + \alpha\gamma_i \quad i = 1, 2, \ldots, k \tag{15}$$

where k-vector γ is determined from validity of the Eq. (11):

The step size α must result in an admissible P^{r+1}. There are the following possible branches:

i) The step α_0, resulting from requirement $g_l(P^{r+1}) = 0$ is admissible. Its selection moves the entry P_l^{r+1} into the part ${}^1P$, the dimension k is increased.

ii) The α_0 is inadmissible, but there is admissible step which transfer l-th entry to the opposite boundary and ${}^1P^{r+1}$ does not cross admissible bound. This step size does not change dimension of ${}^1P$ but the entry P_l^{r+1} now apparently fulfils Kuhn-Tucker conditions.

iii) Also the step size from ii) is too long. Some free variable, say P_m^{r+1}, $m \leqq k$, achieves its bound.

Using this step size we start "minor" iterations. We do not change the selected index l and try to improve the approximation changing this entry. Only the variable P_m is moved from the part ${}^1P$ to the ${}^2P$, the dimension k decreases.

These minor iterations must finish at most in $k \leqq \varrho$ steps because always the entries ${}^1P$ are moved to the opposite part ${}^2P$.

4) Re-computing P^r to P^{r+1} we re-compute also $g(P^r)$ to $g(P^{r+1})$ and return to the point 2); only in minor iterations (the case iii) of the 3)) we return to the point 3).

For effective and numerically stable computation of the vector γ (15) Fletcher and Jackson propose to use partial LDL^T factorisation. They assume V_z decomposed

$$(16) \qquad V_z = LDL^T = \begin{bmatrix} {}^1V & {}^{12}V^T \\ {}^{12}V & {}^2V \end{bmatrix} \begin{matrix} \} k \\ \\ \end{matrix} = \begin{bmatrix} {}^1L & 0 \\ {}^{12}L & {}^2L \end{bmatrix} \begin{bmatrix} {}^1D & 0 \\ 0 & {}^2D \end{bmatrix} \begin{bmatrix} {}^1L^T & {}^{12}L^T \\ 0 & {}^2L^T \end{bmatrix} \begin{matrix} \} k \\ \\ \end{matrix}$$

where L is lower triangular matrix with units on the main diagonal, D is diagonal matrix. The partition in submatrices corresponds to the partition (7).

Partial decomposition means that the submatrices ${}^1L$, ${}^{12}L$ and ${}^1D$ can be computed without computation of the ${}^2L$, ${}^2D$. Moreover the following identity is used

$$(17) \qquad {}^2H = {}^2L\,{}^2D\,{}^2L^T = {}^2V - {}^{12}V\,{}^1D\,{}^{12}V^T$$

With notations (16) and (17) we can find that the vector (15) solves triangular system

$$(18) \qquad {}^1L^T\gamma = -\operatorname{col}_{l-k}({}^{12}L^T)$$

and the following recursion is valid for nonzero part of the gradient

$$(19) \qquad g_i(P^{r+1}) = g_i(P^r) + \alpha\,{}^2H_{i-k,i-k} \quad i = k+1, \quad k+2, \ldots, \varrho$$

Partial decomposition can save computational effort, but in the problem discussed it has two disadvantages

- the partial decomposition of $V_{z(t)}$ cannot be efficiently used in the sequence of problems i.e. in transformation of minimized function by new data.
- the computation in Eq. (17) may be numerically unstable: positive definite matrix ${}^2H$ must be obtained as a result of substraction.

These two points forced us to use full LDL^T decomposition. In the sequence of problems it does not increase computation effort substantially and it removes possible numerical troubles.

Now we describe in which way this modified Fletcher-Jackson algorithm can be implemented into data accumulation.

It seems to be reasonable to take as an initial approximation at time t the optimal estimate obtained in the time $t-1$ i.e.

$$(20) \qquad \hat{P}^0_{ML(t)} = \hat{P}_{ML(t-1)}$$

The relations (2.10), (2.12) and the choice (20) result in the following recursion for the initial gradient:

$$(21) \qquad \begin{aligned} g_{(t)}(\hat{P}^0_{ML(t)}) &= g_{(t)}(\hat{P}_{ML(t-1)}) = V_{z(t)}\hat{P}_{ML(t-1)} - V_{zy(t)} = \\ &= V_{z(t-1)}\hat{P}_{ML(t-1)} - V_{zy(t-1)} - z_{(t)}(y_{(t)} - \hat{P}^T_{ML(t-1)}z_{(t)})^T = \\ &= g_{(t-1)}(\hat{P}_{ML(t-1)}) - z_{(t)}\hat{e}^T_{(t)} \end{aligned}$$

Having LDL^T decomposition of the $V_{z(t-1)}$ the recursion (10) for the V_z can be written

$$L_{(t)}D_{(t)}L_{(t)}^T = L_{(t-1)}(D_{(t-1)} + f_{(t-1)}f_{(t-1)}^T)\,L_{(t-1)}^T \tag{22}$$

where the vector f (superfluous index $t-1$ will be suppressed) solve the triangular system

$$Lf = z_{(t)} \tag{23}$$

Using definition of the LDL^T decomposition we decompose

$$D + ff^T = QD_{(t)}Q^T \tag{24}$$

where the Q is the matrix of the same type as the L. Then

$$L_{(t)} = L_{(t-1)}Q \tag{25}$$

Algorithm which solve (23), (24), (25) can be written in the following compact form (vector $z_{(t)}$ is destroyed)

$$\begin{aligned}
&s = 1\\
&j = 1, 2, \ldots, \varrho\\
&c = z_j\\
&a = G_{jj}\\
&b = 1/s\\
&s = s + c * * 2/a\\
&a = a * s\\
&G_{jj} = a * b\\
&a = c/a\\
&i = j+1, j+2, \ldots, \varrho\\
&b = G_{ij}\\
&z_i = z_i - b * c\\
&G_{ij} = b + a * z_i\\
&\text{end of } i, j
\end{aligned} \tag{26}$$

It means that the use of the full LDL^T decomposition requires approximatelly only two times more multiplications than the simple re-computation of V_z according to (2.10). It requires aproximatelly ϱ^2 multiplications.

The last problem which has to be solved is reconstruction of the full LDL^T decomposition after permutations performed in the steps 3i), 3iii) of the Algorithm.

As this problem seems to us to be of independent interest (the task appears in some classification problems), the solution will be given separately.

4. THE PERMUTATION OF VARIABLES AND LDL^T DECOMPOSITION

We have given LDL^T decomposition and the quadratic form

$$\psi(x) = x^T LDL^T x \tag{1}$$

The vector x splits in two subvectors

$$x = \begin{bmatrix} {}^1x \\ {}^2x \end{bmatrix} \begin{matrix} \} k \\ \} \varrho \end{matrix} \tag{2}$$

We solve the following problems:

a) To remove the variable x_m, $m \leqq k$, from ${}^1x$. For this purpose we exchange

$$x_m \leftrightarrow x_{m+1} \tag{3}$$

then we transform the destroyed LDL^T decomposition into right form, symbolically we perform transformation $\mathscr{R}$. We repeat procedure up to k-th entry, i.e.

$$x_m \leftrightarrow x_{m+1}\, \mathscr{R}\, x_m \leftrightarrow x_{m+2}\, \mathscr{R} \ldots x_m \leftrightarrow x_k\, \mathscr{R} \tag{4}$$

then we set $k = k - 1$.

b) To transfer l-th variable, $l > k$, from ${}^2x$ to ${}^1x$. With notations of the problem a) we proceed

$$x_l \leftrightarrow x_{l-1}\, \mathscr{R}\, x_l \leftrightarrow x_{l-2}\, \mathscr{R} \ldots x_l \leftrightarrow x_{k+1}\, \mathscr{R} \tag{5}$$

and then we increase k to $k + 1$.

The solutions of the both problems require effective computation

$$x_i \leftrightarrow x_{i+1}\, \mathscr{R} \tag{6}$$

Exchanging the i-th and $(i + 1)$-st variable we exchange the i-th and $(i + 1)$-st rows in the matrix L.

This permutation results in the following forms of the i-th and $(i + 1)$-st columns

$$\begin{matrix} i-1\{ \\ 1\{ \\ 1\{ \\ \\ \end{matrix} \begin{bmatrix} 0 \\ a \\ 1 \\ {}_1h \end{bmatrix} \Big\} \varrho - i + 1 \qquad \begin{matrix} i-1\{ \\ 1\{ \\ 1\{ \\ \\ \end{matrix} \begin{bmatrix} 0 \\ 1 \\ 0 \\ {}_2h \end{bmatrix} \Big\} \varrho - i + 1 \tag{7}$$

The new i-th and $(i + 1)$-st columns must have the form

$$\begin{matrix} i-1\{ \\ 1\{ \\ 1\{ \\ \\ \end{matrix} \begin{bmatrix} 0 \\ 1 \\ b \\ {}_1f \end{bmatrix} \Big\} \varrho - i + 1 \qquad \begin{matrix} i-1\{ \\ 1\{ \\ 1\{ \\ \\ \end{matrix} \begin{bmatrix} 0 \\ 0 \\ 1 \\ {}_2f \end{bmatrix} \Big\} \varrho - i + 1 \tag{8}$$

The original i-th and $(i + 1)$-st terms of the diagonal matrix D, say q_1, q_2, we transform in r_1, r_2

The following identity must be valid

$$\begin{bmatrix} a & 1 \\ 1 & 0 \\ {}_1h & {}_2h \end{bmatrix} \begin{bmatrix} q_1 & 0 \\ 0 & q_2 \end{bmatrix} \begin{bmatrix} a & 1 & {}_1h^T \\ 1 & 0 & {}_2h^T \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ b & 1 \\ {}_1f & {}_2f \end{bmatrix} \begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix} \begin{bmatrix} 1 & b & {}_1f^T \\ 0 & 1 & {}_2f^T \end{bmatrix} \tag{9}$$

The equation (9) is solved by

$$\begin{aligned} r_1 &= a^2 q_1 + q_2 \\ b &= \frac{a q_1}{r_1} \\ r_2 &= \frac{q_1 q_2}{r_1} \\ {}_1f &= {}_1hb + {}_2h \frac{q_2}{r_1} \\ {}_2f &= {}_1h - {}_2ha \end{aligned} \tag{10}$$

i.e. the transformation (6) requires aproximately $3(\varrho - i - 1)$ multiplications and the sequence of the type (4) or (5) between variables $j_1 < j_2$ needs $3 \sum_{i=j_1}^{j_2-1} (\varrho - i - 1)$ multiplications. This quantity is bounded by $3/2\varrho^2$ multiplications.

5. ILLUSTRATIVE EXAMPLE

To illustrate the behavior of the resulting algorithm we give rather simple example of parameter identification of a controlled system in the open loop.

The system is determined by

$$\hat{y}_{(t)} = P^T z_{(t)} = [2,\ 0.98,\ -1.98] \begin{bmatrix} u_{(t)} \\ y_{(t-1)} \\ u_{(t-1)} \end{bmatrix}, \quad \Omega = 1 \tag{1}$$

and the white-noise input is normally distributed

$$u_{(t)} \propto N(0, 1) \tag{2}$$

The estimates of $P_2 = 0.98$ are drawn in the Fig. 1 for unbounded and bounded case in which

$$\bar{P} = \begin{bmatrix} 0 \\ 0 \\ -2 \end{bmatrix}, \quad \bar{\bar{P}} = \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix} \tag{3}$$

Because the true parameter values are in the range given both estimates coincide "asymptotically". But transient identification part is influenced substantially as is

apparent from the sample dispersion of prediction errors which is 5.6 for unbounded and 1.2 for bounded case.

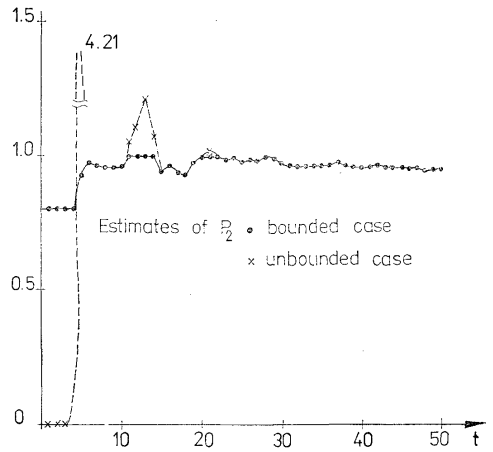

Fig. 1.

The feature that the activity of any bound influences all estimates is illustrated in Fig. 2 which contains the estimates of P_2 when bounds on P_3 were changed to the unrealistic ones ($\hat{P}_1 \approx 2.1$, $\hat{P}_3 = 0$)

(4) $$0 \leqq P_3 \leqq 2$$

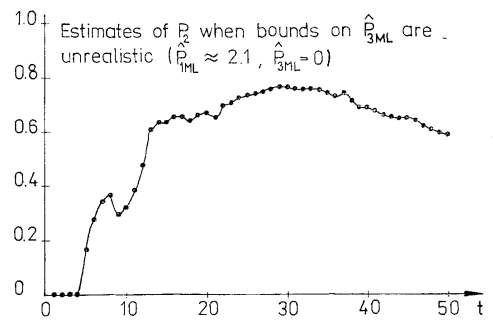

Fig. 2.

6. CONCLUSIONS

The Bayesian solution of the problem of the recursive parameter estimation in a given interval was presented. Practical importance of this problem was stressed and simulation experiments support expected favourable influence on the transient identification behavior.

Moreover at least the following interesting features of the problem assumed can be found:

- It separates clearly the experience accumulation (computation of sufficient statistic) and decision making (selection of parameter estimates).
- It shows that the information contained in the data must not be destroyed by any prior information.
- It demands the use of the special type of mathematical programming in almost standard identification problem. This combination requires some modifications of the fundamental algorithm [1], which result in still effective but more stable computation.

  As the by-product the re-computation of the LDL^T decomposition after permutations was solved in the unified and effective manner. This problem can be met also in computations connected with the system classification [9].
- It clarifies that the activity of any bound changes the other entries of point estimate.
- It is useful also when computing time is bounded (this may appear in adaptive control), because of its monotonicity computation can be stopped before given number of iterations.

At the end our main practical conclusion has to be repeated: any reasonable (may be heuristic) solution must respect the splitting of experience accumulation and decision making.

(Received June 5, 1981.)

REFERENCES

[1] R. Fletcher, M. P. Jackson: Minimization of a quadratic function of many variables subject only to lower and upper bounds. J. Inst. Math. Appl. *14* (1976), 159—174.

[2] 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).

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

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

[5] M. Kárný: Probabilistic Identification and Self-Reproducing Forms of Distribution Functions (in Czech). Ph. D. Dissertation, Institute of Information Theory and Automation of the Czechoslovak Academy of Sciences, Prague 1976.

[6] L. Ljung: Analysis of vector stochastic algorithms. IEEE Trans. Autom. Control *AC-22* (1977), 551—579.

[7] G. K. Kel'mans, A. S. Poznyak, A. V. Chernitser: Adaptive locally optimal control. Internat. J. Systems Sci. *12* (1981), 235—254.
[8] M. Kárný, J. Böhm, A. Halousková: Identification and Control of Cold Rolling System (in Czech). Internal Report o.p. Škoda Válcovny 1980.
[9] M. Kárný: Automated structure determination of regression model (to appear).

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