Problem of State Filtering in Case of Partially Known System Matrices

Lenka Pavelková

***Abstract*— The linear state-space model with uniform innovations (LU model) proposed in previous author's work is extended here. The states and parameters of LU model are estimated under hard physical bounds. The estimation of the innovation boundaries is also included. Maximum a posteriori probability estimation reduces to the linear programming. The on-line estimation is running within a sliding window.**

Compared to the original model, we consider that model matrices can be time-variant. Also, offset terms are included. We present the problem of the joint parameter and state estimation, i.e., the state filtering with unknown model matrices. The ambiguity in the state estimates can be substantially decreased by partial knowledge of some entries in the model matrices. The simple example illustrates this approach.

I. INTRODUCTION

The real system is often modelled by a state space model. Here, the subtasks of parameter estimation and of the filtration (state estimation) arise. The innovations of state evolution as well as observation model are often supposed to have normal distribution. Kalman filtering (KF) [4] is then the first-option estimation method. The main advantage of the KF is the simplicity but its use is restricted by assumed knowledge of the parameters including innovation covariances. So the various extension of KF are used like the extended KF, the iterated KF and the unscented KF. All these extensions are compared in [6].

The above mentioned model deals with normally distributed innovations. The Gaussian distribution has unbounded support. This fact can often be accepted as a reasonable approximation of reality, which is mostly bounded. In some case, however, this assumption is unrealistic, e.g., in modelling of transportation systems (for encountered problems see e.g. [7]) or do not fit subsequent processing, for instance, robust control design [3]. Then, techniques similar to those dealing with unknown-but-bounded equation errors are used, see references in [8]. They often intentionally give up stochastic interpretation of the innovations and develop and analyze various algorithms of a min-max type, cf. [9]. The unknown parameters (or states) lie then within the bounded set. The complexity of this set is very high so approximation is needed to obtain recursively feasible solution. The unknown-but-bounded approaches [10], [11] face this problem by a recursive construction of a simple (typically outer) approximation of the bounded set (support of the distribution describing knowledge about estimated quantities). The approximation by an ellipsoid is described in [10]. An alternative way is an approximation by a multivariate box. This methodology is used in [11]. It brings simplicity to the subsequent use, as it provides very simple description of uncertainty. This solution can be extended so that approximated area consists of an union of non-overlapping boxes. This algorithm is described in [12].

The min-max type algorithms are definitely useful but the related decision-making tasks are unnecessarily difficult because of the broken connection to the established statistical tools.

The linear state model with uniform innovations (LU model) introduced by the author in [1] proposes an alternative to the above mentioned approaches. It keeps simultaneously the advantages of the probabilistic approach and the simplicity of the estimation algorithm.

By the LU model, the state and output innovations are considered to have the uniform distribution. Here, we extend the original LU model and perform its estimation.

The paper is concerned with the problem of quality of the state estimates. The simulated example shows that partial knowledge of the model matrices decreases the mean error of the estimates.

L. Pavelková is with Institute of Information Theory and Automation, Czech Academy of Science, Czech republic pavelkov@utia.cas.cz

II. LIST OF THE NOTATION

$\equiv$	equality by definition
$\propto$	equality up to a constant factor (proportionality)
x^*	a set of x-values, $x \in x^*$
$\mathring{x}$	the number of members in the countable set x^*
x^ℓ	the length of the vector x; vectors are always columns
u_t, y_t	known input and observed output of the system, respectively; the subscript $t \in t^* \subset \{0, 1, 2, \ldots\}$ labels discrete time
x_t	unobserved state
d_t	the data record at time t; $d_t = (y_t, u_t)$
$x^{k:l}$	the ordered sequence $(x_k, x_{k+1}, \ldots, x_l)$, $1 \leq k \leq l$; $x^{k:l} \equiv [x'_k, x'_{k+1}, \ldots, x'_l]'$
$'$	transposition
$\underline{x}$, $\overline{x}$	lower and upper bound on x, respectively (they are used entry-wise)
$\hat{\Theta}$	point estimate of the parameter Θ based on the available data
$\hat{x}_t$	point estimate of the state x_t based on the available data
${}^{x}r$	a quantity r with a non-numerical superscript x
$f(\cdot\|\cdot)$	probability density functions (pdf); respective pdfs are distinguished by the argument names; no formal

distinction is made between a random variable, its realization and an argument of the pdf

III. LINEAR STATE MODEL WITH UNIFORM INNOVATIONS

A. Model description

The considered system is modelled by the following state (1) and observation (2) equations

$$x_t = {}^cA_t x_{t-1} + {}^cB_t u_t + {}^cF_t + {}^xe_t \quad (1)$$

$$y_t = {}^cC_t x_t + {}^cD_t u_t + {}^cG_t + {}^ye_t, \quad (2)$$

where

x_t, u_t, y_t are state, input and output vectors respectively;

${}^cA_t$, ${}^cB_t$, ${}^cF_t$, ${}^cC_t$, ${}^cD_t$, ${}^cG_t$ are model matrices of appropriate dimensions; they are sums of the form

$${}^cA_t = A_t + {}^eA, \; {}^cB_t = B_t + {}^eB, \text{ etc., where} \quad (3)$$

A_t contains known, generally time-variant, entries of ${}^cA_t$, ${}^eA$ contains unknown time-invariant entries of ${}^cA_t$ and zeros (similarly for other system matrices);

the unknown entries are collected into the "coefficient part" θ of the unknown parameter Θ (5);

${}^xe_t$, ${}^ye_t$ are the vectors of the state and output innovations respectively; they are assumed to be zero mean with constant variances, mutually conditionally independent and identically distributed.

The innovation are assumed to have uniform distribution

$$f\left({}^xe_t\right) = \mathcal{U}\left(0, {}^xr\right), \quad f\left({}^ye_t\right) = \mathcal{U}\left(0, {}^yr\right) \quad (4)$$

where $\mathcal{U}(\mu, r)$ is uniform pdf on the box with the center μ and half-width of the support interval equal to r.

To collect all estimated parameters, we denote

$$\Theta \equiv \left[\theta', {}^xr', {}^yr'\right]', \; \theta \equiv [\mathrm{col}({}^eA)', \mathrm{col}({}^eB)', \quad (5)$$

$$\mathrm{col}({}^eF)', \mathrm{col}({}^eC)', \mathrm{col}({}^eD)', \mathrm{col}({}^eG)']',$$

where

$\mathrm{col}(M)$ is an operator that converts matrix M with α rows and β columns into a column vector of the length $\alpha\beta$.

Equations (1) and (2) together with the assumptions (4) define the linear uniform state-space model (LU model).

We assume that generator of inputs $u^{1:\mathring{t}} \equiv [u'_1, \ldots, u'_{\mathring{t}}]'$ meets natural conditions of control [2], i.e., it uses explicitly neither state values nor unknown parameters. Further, we suppose that the initial state x_0 and parameter Θ are uniformly distributed on the set $\mathcal{S}0$ defined by the inequalities

$$\mathcal{S}0 = \left\{\underline{x}_0 \le x_0 \le \overline{x}_0, \, \underline{\Theta} \le \Theta \le \overline{\Theta}\right\}. \quad (6)$$

The state x_0 and parameters Θ are assumed a priori mutually independent, hence

$$f\left(x_0, {}^xr, {}^yr, \theta\right) = f\left(x_0\right) f\left({}^xr\right) f\left({}^yr\right) f\left(\theta\right).$$

Possible restrictions on the state values are in the form

$$\mathcal{S}2 = \{\underline{x} \le x_t \le \overline{x}\}, t \in t^* = \{1, 2, \ldots, \mathring{t}\}. \quad (7)$$

Then, the joint pdf of data $d^{1:\mathring{t}}$, $d_t = (y_t, u_t)$, the state trajectory $x^{0:\mathring{t}}$ and parameter Θ of the LU model is

$$f\left(d^{1:\mathring{t}}, x^{0:\mathring{t}}, \Theta\right) \propto \prod_{i=1}^{x^\ell} ({}^xr_i)^{-\mathring{t}} \prod_{j=1}^{y^\ell} ({}^yr_j)^{-\mathring{t}} \chi(\mathcal{S}) f\left(\Theta\right) \quad (8)$$

where

x^ℓ, y^ℓ is the size of the state and output vector, respectively, $\chi(\mathcal{S})$ is the indicator of the support $\mathcal{S}$.

The convex set $\mathcal{S}$ is given as follows

$$\mathcal{S} = \mathcal{S}0 \cap \mathcal{S}1 \cap \mathcal{S}2. \quad (9)$$

with $\mathcal{S}0$ given by (6) and $\mathcal{S}2$ by (7). The set $\mathcal{S}1$ is specified by inequalities

$$\begin{aligned} -{}^xr &\le x_t - {}^cA_t x_{t-1} - {}^cB_t u_t - {}^cF_t &\le {}^xr \\ -{}^yr &\le y_t - {}^cC_t x_t - {}^cD_t u_t - {}^cG_t &\le {}^yr \end{aligned} \quad (10)$$

with $t \in t^* = \{1, 2, \ldots, \mathring{t}\}$. Note that the inequalities (10) follow from the (1) – (3) with lower and upper noise bounds given by (4).

B. On-line joint parameter and state estimation

The Bayesian learning, i.e., parameter estimation and state filtration [2], consist in the evaluation of the posterior probability density function (pdf) describing the unknown quantity. This task is too complex because of the time-increasing complexity of the support of the posterior pdf. Therefore, we evaluate the maximum a posteriori probability (MAP) estimate [5] of the unknown quantity.

The real-time (on-line) estimation provides the state and/or parameter estimates in each time step. Standard Bayesian learning with a fixed lag $\partial \ge 0$ works with the data $d^{t-\partial:t}$ and states $x^{t-\partial:t}$. The superfluous state $x_{t-\partial-1}$ and data item $d_{t-\partial-1}$ are integrated out from the posterior pdf in every time step t. This integration induces non-uniform terms in the posterior pdf. We approximate these terms in order to preserve the uniformity. The approximate joint pdf takes the form

$$\tilde{f}(d^{t-\partial:t}, x^{t-\partial:t}, \Theta) \quad (11)$$

$$\propto \prod_{i=1}^{x^\ell} ({}^xr_i)^{-(\partial+1)} \prod_{j=1}^{y^\ell} ({}^yr_j)^{-(\partial+1)} \chi(\tilde{\mathcal{S}}_t) f\left(\Theta\right)$$

with $t \in t^* = \{\partial + 1, \ldots, \mathring{t}\}$, $1 < \partial \ll \mathring{t}$

where

$\chi(\tilde{\mathcal{S}}_t)$ is the indicator of the support $\tilde{\mathcal{S}}_t$.

The time variant convex set

$$\tilde{\mathcal{S}}_t = \tilde{\mathcal{S}}0_t \cap \tilde{\mathcal{S}}1_t \cap \tilde{\mathcal{S}}2_t \quad (12)$$

stems from the original set $\mathcal{S}$ (9). It holds for the individual sub-matrices, $\tau \in \{t - \partial, \ldots, t\}$, $t \in \{\partial + 1, \ldots, \mathring{t}\}$,

$$\tilde{\mathcal{S}}0_t = \left\{x_{t-\partial-1} = \hat{x}_{t-\partial-1}, \, \underline{\Theta} \le \Theta \le \overline{\Theta}\right\}, \quad (13)$$

$$\tilde{\mathcal{S}}1_t = \quad (14)$$

$$
\begin{aligned}
\{-{}^{x}r &\leq x_\tau - {}^{c}A_\tau x_{\tau-1} - {}^{c}B_\tau u_\tau - {}^{c}F_\tau &\leq {}^{x}r,\\
-{}^{y}r &\leq y_\tau - {}^{c}C_\tau x_\tau \quad - {}^{c}D_\tau u_\tau - {}^{c}G_\tau &\leq {}^{y}r\},
\end{aligned}
$$

$$
\tilde{\mathcal{S}}2_t = \{\underline{x} \leq x_\tau \leq \overline{x}\}. \tag{15}
$$

The on-line joint estimation consist in the evaluation of the pdf

$$
f\left(x^{t-\partial:t}, \Theta \middle| d^{t-\partial:t}\right),\ t \in \{\partial+1, \ldots, \mathring{t}\},\ 1 < \partial \ll \mathring{t}. \tag{16}
$$

The corresponding MAP estimate is obtained as follows. Taking the negative logarithm of the posterior pdf and applying the approximation $\ln(r) \approx r - 1,\ 0 < r \leq 2$, we get

$$
\hat{X}_{MAP} = \text{arg min}_{X_t \in \tilde{\mathcal{S}}_t} \left(\sum_{i=1}^{x^\ell} {}^{x}r_i + \sum_{j=1}^{y^\ell} {}^{y}r_j \right) \tag{17}
$$

where
X_t contains estimated quantities, i.e., states and parameters, reorganized into the column vector,
$\tilde{\mathcal{S}}_t$ is given by (12).

To solve this problem, we use the method of the linear programming (LP) [5],

$$
\begin{aligned}
&\text{Find a vector } X_t \text{ such that } J \equiv \mathcal{C}' X_t\\
&= \sum_{i=1}^{x^\ell} {}^{x}r_i + \sum_{j=1}^{y^\ell} {}^{y}r_j \ \rightarrow \min
\end{aligned} \tag{18}
$$

$$
\text{while } \mathcal{A}_t X_t \leq \mathcal{B}_t,\ \underline{X}_t \leq X_t \leq \overline{X}_t, t \in t^* \tag{19}
$$

where
$\mathcal{C}' \equiv [\mathbf{0}'_{(X^\ell - x^\ell - y^\ell)}, \mathbf{1}'_{(x^\ell + y^\ell)}]$, $\mathcal{C}^\ell = X^\ell$; $\mathbf{0}_{(len)}$, $\mathbf{1}_{(len)}$ are the vectors of zeros and ones, respectively, both of the length len;
$\mathcal{A}_t$ and $\mathcal{B}_t$ are known matrix and vector, respectively; they result from the inequalities describing the set $\tilde{\mathcal{S}}1_t$ (14);
$\underline{X}_t$, $\overline{X}_t$ are known vectors; they stem from the sets $\tilde{\mathcal{S}}0_t$ (13) and $\tilde{\mathcal{S}}2_t$ (15).

In the case of the joint estimation of the parameters and states, the conditions of the linearity are not fulfilled because of the terms ${}^{e}Ax_{t-1}$ in (1) and ${}^{e}Cx_t$ in (2). We linearize these terms in the following way

$$
\begin{aligned}
{}^{e}Ax_{t-1} &= ({}^{e}A - \hat{A})x_{t-1} + \hat{A}x_{t-1}\\
&= ({}^{e}A - \hat{A})(x_{t-1} - \hat{x}_{t-1})\\
&+ ({}^{e}A - \hat{A})\hat{x}_{t-1} + \hat{A}x_{t-1}\\
&\approx {}^{e}A\hat{x}_{t-1} - \hat{A}\hat{x}_{t-1} + \hat{A}x_{t-1}, t \in t^*,
\end{aligned} \tag{20}
$$

where $\hat{A}$, $\hat{x}_{t-1}$ are the newest available estimates of the parameter ${}^{e}A$ and state x_{t-1}, respectively. It is supposed that the mean of $({}^{e}A - \hat{A})(x_{t-1} - \hat{x}_{t-1}) \approx 0$

Using similar expansion for ${}^{e}Cx_t$, $t \in t^*$, we get

$$
{}^{e}Cx_t \approx {}^{e}C\hat{x}_t - \hat{C}\hat{x}_t + \hat{C}x_t \tag{21}
$$

Then, the resulting inequalities for LP (18) are in the following form, $\tau \in \{t-\partial, \ldots, t\}$, $t \in \{\partial+1, \ldots, \mathring{t}\}$,

$$
\begin{aligned}
&x_\tau - {}^{e}A\hat{x}_{\tau-1} - \hat{A}x_{\tau-1} - A_\tau x_{\tau-1} - {}^{e}Bu_\tau - {}^{e}F - {}^{x}r\\
&\quad \leq -\hat{A}\hat{x}_{\tau-1} + B_\tau u_\tau + F_\tau\\[1em]
&-x_\tau + {}^{e}A\hat{x}_{\tau-1} + \hat{A}x_{\tau-1} + A_\tau x_{\tau-1} + {}^{e}Bu_\tau + {}^{e}F - {}^{x}r\\
&\quad \leq \hat{A}\hat{x}_{\tau-1} - B_\tau u_\tau - F_\tau\\[1em]
&{}^{e}C\hat{x}_\tau + \hat{C}x_\tau + C_\tau x_\tau + {}^{e}Du_\tau + {}^{e}G - {}^{y}r\\
&\quad \leq +y_\tau + \hat{C}\hat{x}_\tau - D_\tau u_\tau - G_\tau\\[1em]
&-{}^{e}C\hat{x}_\tau - \hat{C}x_\tau - C_\tau x_\tau - {}^{e}Du_\tau - {}^{e}G - {}^{y}r\\
&\quad \leq -y_\tau - \hat{C}\hat{x}_\tau + D_\tau u_\tau + G_\tau
\end{aligned} \tag{22}
$$

$$
\begin{aligned}
&\underline{x} \leq x_\tau \leq \overline{x}\\
&0 \leq {}^{x}r \leq {}^{x}\overline{r},\ 0 \leq {}^{x}r \leq {}^{y}\overline{r}\\
&x_{t-\partial-1} = \hat{x}_{t-\partial-1},\ \underline{\theta} \leq \theta \leq \overline{\theta}
\end{aligned}
$$

$\hat{x}_{t-\partial-1}$ is the estimate of $x_{t-\partial-1}$ from the previous step.

Note that the estimates based on the data up to time $t-1$, i.e., $d^{t-\partial-1:t-1}$ are used for LP performed in the time t. Therefore, we have only the estimates from $\hat{x}_{t-\partial}$ to $\hat{x}_{t-1}$ at disposal. The missing estimate $\hat{x}_t$ is replaced by its prediction, i.e.,

$$
\hat{x}_t = (A_t + \hat{A})\hat{x}_{t-1} + (B_t + \hat{B})u_t + (F_t + \hat{F}). \tag{23}
$$

Then, $\mathcal{A}_t$, $\mathcal{B}_t$ are in the form

$$
\mathcal{A}_t = \begin{bmatrix} \mathcal{A}11_t & \mathcal{A}12_t & \mathcal{A}13_t & \mathcal{A}14_t \\ \mathcal{A}21_t & \mathcal{A}22_t & \mathcal{A}23_t & \mathcal{A}24_t \end{bmatrix},\ \mathcal{B}_t = \begin{bmatrix} \mathcal{B}1_t \\ \mathcal{B}2_t \end{bmatrix}
$$

with

$$
\mathcal{A}11_t
$$

$$
= \begin{bmatrix}
I_{(x^\ell,x^\ell)} & -\hat{A} - A_t & \mathbf{0}_{(x^\ell,x^\ell)} & \ldots & \mathbf{0}_{(x^\ell,x^\ell)}\\
-I_{(x^\ell,x^\ell)} & \hat{A} + A_t & \mathbf{0}_{(x^\ell,x^\ell)} & \ldots & \mathbf{0}_{(x^\ell,x^\ell)}\\
\mathbf{0}_{(x^\ell,x^\ell)} & I_{(x^\ell,x^\ell)} & -\hat{A} - A_{t-1} & \ldots & \mathbf{0}_{(x^\ell,x^\ell)}\\
\mathbf{0}_{(x^\ell,x^\ell)} & -I_{(x^\ell,x^\ell)} & \hat{A} + A_{t-1} & \ldots & \mathbf{0}_{(x^\ell,x^\ell)}\\
\vdots & \vdots & \vdots & \ddots & \vdots\\
\mathbf{0}_{(x^\ell,x^\ell)} & \mathbf{0}_{(x^\ell,x^\ell)} & \ldots & \ldots & I_{(x^\ell,x^\ell)}\\
\mathbf{0}_{(x^\ell,x^\ell)} & \mathbf{0}_{(x^\ell,x^\ell)} & \ldots & \ldots & -I_{(x^\ell,x^\ell)}
\end{bmatrix},
$$

$$
\mathcal{A}12_t \equiv \begin{bmatrix}
-SEL_{{}^{e}A,\hat{x}_{t-1}} & -SEL_{{}^{e}B,u_t} & -SEL_{{}^{e}F,1}\\
SEL_{{}^{e}A,\hat{x}_{t-1}} & SEL_{{}^{e}B,u_t} & SEL_{{}^{e}F,1}\\
\vdots & \vdots & \vdots\\
-SEL_{{}^{e}A,\hat{x}_{t-\partial-1}} & -SEL_{{}^{e}B,u_{t-\partial}} & -SEL_{{}^{e}F,1}\\
SEL_{{}^{e}A,\hat{x}_{t-\partial-1}} & SEL_{{}^{e}B,u_{t-\partial}} & SEL_{{}^{e}F,1}
\end{bmatrix},
$$

where

$$SEL_{M;v_t} = \begin{bmatrix} sel_M(v_t,1) & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & sel_M(v_t,v^\ell) \end{bmatrix}, \quad (24)$$

$sel_M(v_t,i)$ creates a reduced vector (scalar) $\tilde{v}_t$ from $v_t \in \{x_t, u_t, 1\}$, $t \in t^*$; it selects entries of v_t with indeces corresponding to the non-zero columns on the i-th row of the matrix (vector) M, $M \in \{{}^eA,\ {}^eB,\ {}^eF,\ {}^eC,\ {}^eD,\ {}^eG\}$; note that if all entries on i-th row are equal to zero, then $\tilde{v}_t$ is an "empty" vector.

Note that $\mathcal{A}12_t$ has $2(\partial+1)x^\ell$ rows and m columns, the upper bound on m is $x^\ell(x^\ell+u^\ell+1)$,

$\mathcal{A}13_t = \mathbf{0}_{(2(\partial+1)x^\ell,n)}$, maximal size of n is $y^\ell(x^\ell+u^\ell+1)$.

$$\mathcal{A}14_t \equiv \mathbf{1}_{(2(\partial+1),1)} \otimes \left[-I_{(x^\ell)}\ \mathbf{0}_{(x^\ell,y^\ell)}\right],$$

where $\otimes$ denotes Kronecker product,

$$\mathcal{A}21_t = \begin{bmatrix} \hat{C}+C_t & \mathbf{0}_{(y^\ell,x^\ell)} & \dots & \mathbf{0}_{(y^\ell,x^\ell)} \\ -\hat{C}-C_t & \mathbf{0}_{(y^\ell,x^\ell)} & \dots & \mathbf{0}_{(y^\ell,x^\ell)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{(y^\ell,x^\ell)} & \mathbf{0}_{(y^\ell,x^\ell)} & \dots & \hat{C}+C_{t-\partial} \\ \mathbf{0}_{(y^\ell,x^\ell)} & \mathbf{0}_{(y^\ell,x^\ell)} & \dots & -\hat{C}-C_{t-\partial} \end{bmatrix},$$

$$\mathcal{A}22_t = \mathbf{0}_{2(\partial+1)y^\ell,m},\ m \text{ is defined by } \mathcal{A}11,$$

$$\mathcal{A}23_t \equiv \begin{bmatrix} SEL_{{}^eC,\hat{x}_t} & SEL_{{}^eD,u_t} & SEL_{{}^eG,1} \\ -SEL_{{}^eC,\hat{x}_t} & -SEL_{{}^eD,u_t} & -SEL_{{}^eG,1} \\ \vdots & \vdots & \vdots \\ SEL_{{}^eC,\hat{x}_{t-\partial}} & SEL_{{}^eD,u_{t-\partial}} & SEL_{{}^eG,1} \\ -SEL_{{}^eC,\hat{x}_{t-\partial}} & -SEL_{{}^eD,u_{t-\partial 1}} & -SEL_{{}^eG,1} \end{bmatrix},$$

where $sel_M(v_t,i)$ is defined by (24), $\hat{x}_t$ is obtained as the prediction (23). Note that $\mathcal{A}23_t$ has $2(\partial+1)y^\ell$ rows and n columns with n defined by $\mathcal{A}12_t$.

$$\mathcal{A}24_t = \mathbf{1}_{(2(\partial+1),1)} \otimes \left[\mathbf{0}_{(y^\ell,x^\ell)}\ \ -I_{(y^\ell)}\right],$$

$$\mathcal{B}1_t = \begin{bmatrix} -\hat{A}\hat{x}_{t-1}+B_t u_t+F_t \\ +\hat{A}\hat{x}_{t-1}-B_t u_t-F_t \\ \vdots \\ -\hat{A}\hat{x}_{t-\partial}+B_{t-\partial+1}u_{t-\partial+1}+F_{t-\partial+1} \\ +\hat{A}\hat{x}_{t-\partial}-B_{t-\partial+1}u_{t-\partial+1}-F_{t-\partial+1} \\ A_{t-\partial}\hat{x}_{t-\partial-1}+B_{t-\partial}u_{t-\partial}+F_{t-\partial} \\ -A_{t-\partial}\hat{x}_{t-\partial-1}-B_{t-\partial}u_{t-\partial}-F_{t-\partial} \end{bmatrix},$$

where $\hat{x}_t$ is obtained as the prediction (23),

$$\mathcal{B}2_t = \begin{bmatrix} \hat{C}\hat{x}_t+y_t-D_t u_t-G_t \\ -\hat{C}\hat{x}_t-y_t+D_t u_t+G_t \\ \vdots \\ \hat{C}\hat{x}_{t-\partial}+y_{t-\partial}-D_{t-\partial}u_{t-\partial}-G_{t-\partial} \\ -\hat{C}\hat{x}_{t-\partial}-y_{t-\partial}+D_{t-\partial}u_{t-\partial}+G_{t-\partial} \end{bmatrix}.$$

The resulting algorithm has two principal distinctions from the extended KF: (i) the algorithm updates estimates on the whole window of the length ∂ and (ii) the realistic hard bounds on the estimated quantities reduce the ambiguity of the model arising from estimating a product of two unknowns.

IV. ILLUSTRATIVE EXAMPLE

A. Simulated model

The two state system with scalar input and output and uniform noise is simulated. The model is described by the LU model (1), (2) and (4), $t \in t^* = \{1,\dots,\mathring{t}\}$, with

$${}^cA_t = \begin{bmatrix} 1 & 0.5 \\ -0.5 & 0 \end{bmatrix},\ {}^cB_t = \begin{bmatrix} 1 \\ 3 \end{bmatrix},\ {}^cF_t = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

$${}^cC_t = \begin{bmatrix} 1 & 1 \end{bmatrix},\ {}^cD_t = 0,\ {}^cG_t = 1,$$

$${}^xr = \begin{bmatrix} 10^{-1} \\ 10^{-1} \end{bmatrix},\ {}^yr = 10^{-1}.$$

Note that all model matrices are time invariant in this example. Therefore, we omit their time indexes for the simplicity.

The input is simulated as a random, uniformly distributed, signal from the interval $[-1,1]$. The input values are independent for different time moments. The data set consists of $\mathring{t} = 500$ data pairs (inputs and outputs).

B. Evaluation of experiments

To evaluate the quality of the estimation, the mean error (ME) of the output predictions and ME of the state or parameter estimates are used, respectively. Generally, the ME of the quantity E is computed entry-wise in the following way

$$ME_E = \frac{1}{\mathring{t}} \sum_{t=1}^{\mathring{t}} |E_t - R_t| \quad (25)$$

where E_t are the predicted outputs, state estimates and parameter estimates, respectively; R_t are the true values of the outputs and states, respectively; $\mathring{t}$ is the number of the samples. In the case of the parameter estimation, $R_t = R,\ t \in \{1,2,...,\mathring{t}\}$, $R \in \{{}^eA,\ {}^eB,\ {}^eF,\ {}^eC,\ {}^eD,\ {}^eG\}$.

C. Joint parameter and state estimation

Here, the states x_t, $t \in t^*$, the model matrices ${}^eA$, ${}^eB$, ${}^eF$, ${}^eC$, ${}^eD$, ${}^eG$ and the innovation boundaries ${}^xr$, ${}^yr$ are estimated. The estimation algorithm is running on-line with various memory lengths ∂. Two cases are considered.

[a] All model matrices are completely estimated, i.e., ${}^{e}A = {}^{c}A$, ${}^{e}B = {}^{c}B$, etc. The restrictions on the entries of the model matrices are chosen so that all entries of the estimated model matrices should be within the interval given by

$$ {}^{c}M(i,j) - K_1 \leq {}^{e}M(i,j) \leq {}^{c}M(i,j) + K_2, $$

where ${}^{c}M(i,j)$ and ${}^{e}M(i,j)$, $i \in \{1, 2, \ldots, m\}$, $j \in \{1, 2, \ldots, n\}$, are i-th rows and j-th column entry of the simulated model matrix ${}^{c}M$ and the estimated model matrix ${}^{e}M$, respectively, both of the size (m, n); K_1, K_2 are positive scalars.

[b] The 2nd row and 2nd column entry of the model matrix ${}^{c}A$ and the vector ${}^{c}F$ are supposed to be known. The above mentioned restrictions are used on the remaining entries.

The comparison of the ME of the state estimates for cases **[a]** and **[b]** and various memory lengths ∂ is on Figure 1. We can see that the ME of the state estimates is smaller in the case of the partially known model matrices.

The comparison of the ME of model parameters is on Figure 2. The parameter ${}^{c}G$ was chosen as a representative because of the most remarkable difference between the case **[a]** and **[b]**.

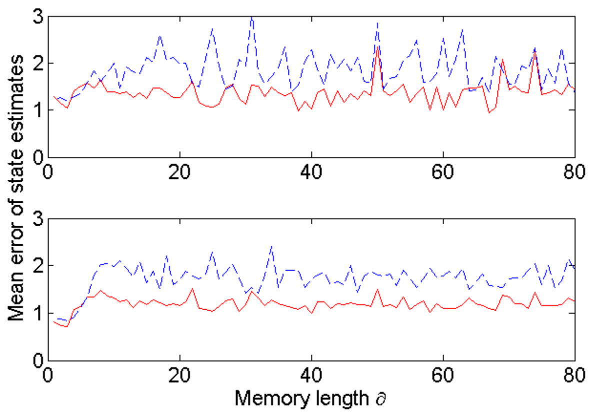

Fig. 1. The mean errors of the state estimates depending on memory length ∂ for all model matrices unknown (dashed line) and partially known model matrices (solid line)

D. Discussion of the results

The experiment with the joint parameter and state estimation supports the intuitive expectation that knowledge of some model matrix entries improves the quality of the estimation measured by the ME of estimate errors.

Moreover, partial knowledge of the model matrices decreases or even remove the ambiguity of the state and parameter estimates caused by the products ${}^{e}A\, x_{t-1}$ and ${}^{e}C\, x_t$ in the model equations (1) and (2), respectively. This fact is important in the real data application. There, every small piece of the information about model matrices can improve the quality of the estimation.

In the case of the LU model, the incorporation of our knowledge into the model is easy. It consists in the adding of the constraints, i.e., additional inequalities and equalities, into the linear programming.

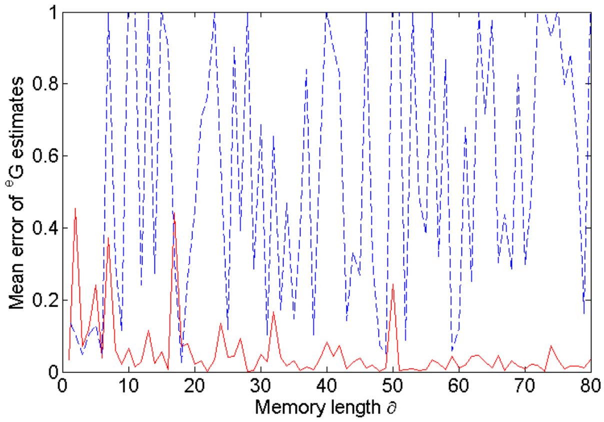

Fig. 2. The mean errors of ${}^{c}G$ estimate depending on memory length ∂ for all model matrices unknown (dashed line) and partially known model matrices (solid line)

V. CONCLUSIONS AND FUTURE WORKS

A. Achieved results

The proposed approach provides the following advantages:

- it enables the joint estimation of parameters, state, and innovation bounds whereas the realistic hard bounds on the estimated quantities reduce the ambiguity of the model (arising from estimating a product of two unknowns);
- it allows (without excessive computational demands) to respect „naturally" hard, physically given, prior bounds on model parameters and states;
- it provides an easy entry of of the partial knowledge on the parameters;
- it allows estimation of the innovation range.

B. Future research

We aim at further improving of the estimates quality. The possible ways are:

- using of the method of the non-linear mathematical programming;
- applying of a more precise approximation for the posterior pdf.

Till now, we worked only with MAP estimates. We aim to refine our results also by an estimation of the precision of these point estimates. The original MAP estimation task gives result $\hat{X}$ for the posterior pdf $f(X|D)$. Now, we aim to find the approximative uniform pdf $\tilde{f}(X|D) = \mathcal{U}_X\left(\hat{X}, {}^{X}R\right)$ so that the distance between f and $\tilde{f}$ is minimal. For this purpose, the Kullback-Leibler divergence [13] will be used that measures well proximity of a pair of pdfs.

VI. ACKNOWLEDGMENTS

The paper has been supported by the Research Center DAR, MŠMT 1M0572, project "Bayes", MŠMT 2C06001 and Czech-Slovenian project KONTAKT MEB090607.

REFERENCES

[1] L. Pavelková and M. Kárný and V. Šmídl, "Towards Bayesian filtering on restricted support," *Proceedings of NSSPW '06. Cambridge*, Nonlinear Statistical Signal Processing Workshop 2006. Cambridge (GB), 13.-15.09.2006.

[2] M. Kárný et al. *Optimized Bayesian Dynamic Advising: Theory and Algorithms*. Springer, 2005.

[3] M. Green and D. J. N. Limebeer, *Linear robust control*. Prentice Hall, 1995.

[4] B. D. O. Anderson and S. B. Moore, *Optimal Filtering*. Prentice Hall Ins., 1979.

[5] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[6] C. Antoniou and M. Ben-Akiva and H. Koutsopoulos, "Online calibration of traffic prediction models", *Traffic Flow Theory*, No. 1934, pages 235–245, 2005.

[7] K. K. Srinivasan and Z. Y. Guo, "Day-to-day evolution of network flows under departure time dynamics in commuter decisions", *Travel Demand and Land Use 2003 Transportation Research Record*, Vol. 1831, pages 47–56, 2003.

[8] J. Roll and A. Nazin and L. Ljung, "Nonlinear system identification via direct weight optimization", *Automatica*, Vol. 41, No. 3, pages 475–490, year = 2004.

[9] B. M. Ninness and G. C. Goodwin, "Rapprochment Between Bounded-Error and Stochastic Estimation Theory", *International Journal of Adaptive Control and Signal Processing*, Vol. 9, No. 1, pages 107–132, 1995.

[10] B. T. Polyak and S. A. Nazin and C. Durieu and E. Walter, "Ellipsoidal parameter or state estimation under model uncertainty", *Automatica*, Vol. 40, No. 7, pages 1171–1179, 2004.

[11] M. Milanese and G. Belforte, "Estimation theory and uncertainty intervals evaluation in presence of unknown but bounded errors Linear families of models and estimators", *IEEE Transactions on Automatic Control*, Vol. 27, No. 2, pages = 408-414, 1982.

[12] A. Bemporad and C. Filippi and F. Torrisi, "Inner and outer approximations of polytopes using boxes", *Computational Geometry*, Vol. 27, pages 151–178, 2004.

[13] S. Kullback and R. Leibler, "On information and sufficiency", *Annals of Mathematical Statistics*, Vol. 22, pages 79–87, 1951.