Approximate On-line Estimation of Uniform State Model with Application on Traffic Data

Lenka Pavelková*

* *Department of Adaptive Systems, Institute of information Theory and Automation, Prague, Czech republic,* pavelkov@utia.cas.cz

Keywords : state model; uniform innovations; parameter and state estimation

Abstract : The model predictive control is an advanced method of process control. The controller relies on the dynamic model of the process obtained by the system identification. The control quality depends strongly on the quality of the used model.

The linear models are often employed for the approximated description of the nonlinear controlled system due to the simplicity of the identification algorithms. However, this models have only limited validity. The nonlinear models describe the system generally much better than the linear ones but their identification is a nontrivial task.

Here, a model that is both easily identifiable and sufficiently precise is presented. The state model with uniform innovations (SU model) proposes an alternative to the standardly used linear state-space model with normal innovations that leads to the Kalman filter. By the SU model, the state and output innovations are considered to have the uniform distribution. This assumption implements the nonlinearity into the originally linear system. The system states and parameters are estimated on-line with fixed memory on the sliding window. The sliding window as the alternative of the forgetting allows to catch the slow parameter changes. The MAP estimation of the SU model reduces to the linear eventually convex programming.

The main advantages of the proposed model are the simplicity of the estimation algorithm and the possibility to estimate both the parameters and states including the innovation boundaries.

The paper is concerned with the problem of the approximative solution of the on-line joint state and parameter estimation.

1 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

Int. Workshop on Assessment and Future Directions of NMPC
Pavia, Italy, September 5-9, 2008

covariances. So the various extension of KF are used like the extended KF, the iterated extended KF and the unscented KF, see e.g. [7].

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.

The state model with uniform innovations (SU model) introduced by author in [2] proposes an alternative to the above mentioned model. By the SU model, the state and output innovations are considered to have the uniform distribution.

In order to use the SU model for the model predictive control [1], the model identification should run continuously (on-line). The Bayesian parameter and state estimation [3] is performed on the sliding window. That keeps the computational feasibility in the reasonable ranges and at the same time it allows to catch the slow parameter changes.

The exact solution of the on-line estimation with memory length is too complex. Therefore, it is necessary to propose the approximate solution of this problem. The paper is concerned with the problem of the approximative on-line joint state and parameter estimation of the SU model.

2 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
x_t, u_t, y_t	unobserved state, known input and observed output of the system, respectively
	the subscript $t \in t^* \subset \{0, 1, 2, \ldots\}$ labels discrete time
d_t	the data record at time t; $d_t = (y_t, u_t)$
$x^{k:l}$	the ordered sequence $(x_k, x_{k+1}, \ldots, x_l)$, $1 \le k \le l$; $x^{k:l} \equiv [x'_k, x'_{k+1}, \ldots, x'_l]'$;
$'$	transposition
$\underline{x}$, $\overline{x}$	lower and upper bound on x, respectively (they are used entry-wise)
${}^x r$	the quantity r with non-numerical superscript x;
$f(\cdot\vert\cdot)$	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 a pdf

3 State model with uniform innovations

3.1 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

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

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

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

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

$$ \Theta \equiv \left[\theta,\ {}^xr,\ {}^yr\right],\ \theta \equiv \left[{}^eA,\ {}^eB,\ {}^eF,\ {}^eC,\ {}^eD,\ {}^eG\right]. \tag{5}$$

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

We assume that generator of inputs $u^{1:\mathring{t}} \equiv [u_1', \ldots, u_{\mathring{t}}']'$ meets natural conditions of control [3], 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

$$ \mathcal{S}_0 = \left\{\underline{x}_0 \le x_0 \le \overline{x}_0, \quad 0 < {}^xr \le {}^x\overline{r}, \quad 0 < {}^yr \le {}^y\overline{r}, \quad \underline{\theta} \le \theta \le \overline{\theta}\right\}. \tag{6}$$

They are assumed a priori mutually independent, hence

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

A possible restrictions on the state values are in the form

$$ \mathcal{S}_x = \{\underline{x} \le x_t \le \overline{x}\}, t \in t^*. \tag{7}$$

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 SU model is

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

where
$x^\ell,\ y^\ell$ 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

$$ \mathcal{S} = \mathcal{S}_0 \cap \mathcal{S}_\Theta \cap \mathcal{S}_x. \tag{9}$$

with $\mathcal{S}_0$ given by (6) and $\mathcal{S}_x$ by (7). The set $\mathcal{S}_\Theta$ is specified by inequalities

$$\begin{array}{lll} -{}^{x}r & \leq x_t - (A_t + {}^{e}A)x_{t-1} - (B_t + {}^{e}B)u_t - (F_t + {}^{e}F) & \leq {}^{x}r \\ -{}^{y}r & \leq y_t - (C_t + {}^{e}C)x_t \quad - (D_t + {}^{e}D)u_t - (G_t + {}^{e}G) & \leq {}^{y}r \end{array} \tag{10}$$

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

3.2 General estimation problem

For the purpose of estimation, the joint pdf (8) can be rewritten as follows

$$f\left(d^{1:\mathring{t}}, x^{0:\mathring{t}}, \Theta\right) = f(D, X) = f(D|\, X)\, f(X) \tag{11}$$

The estimated (unknown) quantities from (8) are collected into X. The remaining (known) quantities are collected into D. Generally, D contains known elements from the set $\left\{d^{\mathring{t}}, x^{0:\mathring{t}}, \theta\right\}$ and X contains unknown elements from the set $\left\{x^{0:\mathring{t}}, \theta, {}^{x}r, {}^{y}r\right\}$. By construction, it holds $D \cap X = \emptyset$ and $D \cup X = \{d^{1:\mathring{t}}, x^{0:\mathring{t}}, \theta, {}^{x}r, {}^{y}r\}$.

Using the notation (11), we can formulate the estimation problems as follows. The posterior pdf $f(X|\, D)$ is searched for on the basis of the prior pdf $f(X)$ and of the pdf describing known quantities $f(D|\, X)$.

According to the Bayes rule, the required posterior pdf $f(X|\, D)$ is proportional to $f(D|\, X)\, f(X)$ on support $\mathcal{S}$ defined by (9). The number of vertices of the support is proportional to the number of data. This is a large number for realistic situations. Consequently, evaluation of moments of this pdf is computationally demanding. This is why we evaluate the maximum a posteriori probability (MAP) estimate $\hat{X}_{MAP}$ [5] of the unknown X.

Taking the negative logarithm of the posterior pdf and applying the approximation $\ln(r) \approx r - 1$, $0 < r \leq 2$, we get

$$\hat{X}_{MAP} = \arg\min_{X \in \mathcal{S}} \left(\sum_{i=1}^{x^\ell} {}^{x}r_i + \sum_{j=1}^{y^\ell} {}^{y}r_j \right). \tag{12}$$

where x^ℓ, y^ℓ is the size of the state and output vector and $\mathcal{S}$ is given by (9).

3.3 On-line joint parameter and state estimation

The real-time (on-line) estimation provides the state and parameter estimates in each time step. Standard Bayesian learning [3] with a fixed lag $\partial \geq 0$ 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.

The Bayesian on-line estimation with restricted memory evolves the join pdf

$$f\left(d^{t-\partial:t}, x^{t-\partial:t}, \Theta\right), t \in t^* = \{1, 2, \ldots, \mathring{t}\}, \tag{13}$$

where integer $1 < \partial \ll \mathring{t}$ means memory length.

For $t \geq \partial + 2$ holds

$$f\left(d^{t-\partial:t}, x^{t-\partial:t}, \Theta\right) \propto \prod_{i=1}^{x^{\ell}} ({}^{x}r_i)^{-\partial+1} \prod_{j=1}^{y^{\ell}} ({}^{y}r_j)^{-\partial} \chi({}^{\partial}\tilde{\mathcal{S}})$$

$$\times f(\Theta)\, g(x_{t-\partial}|u_{t-1}, d^{t-\partial:t-2}, \Theta) \tag{14}$$

where g represents non-uniform pdf that arises after integration. In the time instant t, g is power function containing the power up to t. With increasing t, the estimation becomes intractable because of increasing complexity of the support of the posterior pdf. Therefore, an approximation is proposed.

3.4 Approximate estimation

The factor g in (14) is to be approximated by a pdf $\tilde{f}(x_{t-\partial}|u_{t-1}, d^{t-\partial:t-2}, \Theta)$, $t \in t^*$ so that it holds for measured data $d^{t-\partial:t-2}, u_{t-1}$,

$$\frac{g(x_{t-\partial}|u_{t-1}, d^{t-\partial:t-2}, \Theta)}{\tilde{f}(x_{t-\partial}|u_{t-1}, d^{t-\partial:t-2}, \Theta)} \approx 1,\ \ x \in x^*, \Theta \in \Theta^*. \tag{15}$$

We propose for $t \geq \partial+2$ the approximation by the „cutt off" the superfluous old states with $g(x_{t-\partial}|u_{t-1}, d^{t-\partial:t-2}, \Theta)$ from (14) replaced by the product

$$f\left(x_{t-\partial}\right| x_{t-\partial-1}, u_{t-\partial}, \Theta) \prod_{\tau=t-\partial}^{t-1} f\left(u_\tau \middle| d^{t-\partial:\tau-1}\right),$$

with $x_{t-\partial-1} = \hat{x}_{t-\partial-1}$; $\hat{x}_{t-\partial-1}$ is the point estimate of $x_{t-\partial-1}$ from the previous estimation step. Then holds

$$f\left(d^{t-\partial:t}, x^{t-\partial:t}, \Theta\right) \propto \prod_{i=1}^{x^{\ell}} ({}^{x}r_i)^{-\partial+1} \prod_{j=1}^{y^{\ell}} ({}^{y}r_j)^{-\partial+1} \chi({}^{\partial}\mathcal{S}) f(\Theta)\,. \tag{16}$$

where

$$ {}^{\partial}\mathcal{S} = {}^{\partial}\mathcal{S}_0 \cap {}^{\partial}\mathcal{S}_\Theta \cap {}^{\partial}\mathcal{S}_x \tag{17}$$

${}^{\partial}\mathcal{S}_0 = \left\{x_{t-\partial-1} = \hat{x}_{t-\partial-1},\quad 0 < {}^{x}r \leq {}^{x}\overline{r},\quad 0 < {}^{y}r \leq {}^{y}\overline{r},\quad \underline{\theta} \leq \theta \leq \overline{\theta}\right\}$, cf. (6),

${}^{\partial}\mathcal{S}_x = \mathcal{S}_x$ (7),

${}^{\partial}\mathcal{S}_\Theta$ is given by the inequalities

$$\begin{array}{lll} -{}^{x}r & \leq x_\tau - (A_\tau + {}^{e}A)x_{\tau-1} - (B_\tau + {}^{e}B)u_\tau - (F_\tau + {}^{e}F) & \leq {}^{x}r \\ -{}^{y}r & \leq y_\tau - (C_\tau + {}^{e}C)x_\tau \quad - (D_\tau + {}^{e}D)u_\tau - (G_\tau + {}^{e}G) & \leq {}^{y}r \end{array}$$

with $\tau \in \{t-\partial, t-\partial+1, \ldots, t\}$, $t \in t^* = \{\partial+1, \partial+2, \ldots, \mathring{t}\}$, cf. (10).

Using the notation (11), the estimation problem consists now in the searching for the posterior pdf $f(X|\, D)$ on the basis of the prior pdf $f(X)$ and of the pdf describing observed quantities $f(D|\, X)$ with $D = d^{t-\partial:t}$, $X = (x^{t-\partial:t}, \theta, {}^{x}r, {}^{y}r)$, $t \in t^*$, $\partial > 0$.

The unknown estimated quantity is

$$X = \left[(x^{t:t-\partial})', col(\,^{e}A)', col(\,^{e}B)', col(\,^{e}F)', col(\,^{e}C)', col(\,^{e}D)', col(\,^{e}G)',\,^{x}r',\,^{y}r'\right]' \tag{18}$$

where $t \in t^* = \{\partial + 1, \ldots, \mathring{t}\}$ and col(M) stacks the non-zero rows of the matrix M into a column vector.

The MAP estimate (12) takes now the following form

$$\hat{X}_{MAP} = \arg\,\min_{X \in \,^{\partial}\mathcal{S}} \left(\sum_{i=1}^{x^{\ell}} {}^{x}r_i + \sum_{j=1}^{y^{\ell}} {}^{y}r_j \right). \tag{19}$$

The MAP estimate can be obtained by the method of the convex programming [6]. We are searching for the minimum of the function in (19) on the set ${}^{\partial}\mathcal{S}$ (17). Our estimation algorithms are running in the Matlab environment (see www.mathworks.com) using its optimization toolbox.

4 Illustrative example

The designed algorithm is applied on estimation of the queue lengths that form on the arms of the controlled four-arm intersection. The intersection is described by the following quantities

name	notation	unit	description
queue length	$\xi_{t;i}$	u.c.	number of the cars before the TL (for the i-th arm)
occupancy	$O_{t;i}$	%	relative time of detector activation
input intensity	$I_{t;i}$	u.c/per.	amount of cars passing through the input detector
passage	$P_{t;i}$	u.c/per.	amount of cars passing from arm i into the intersection space
output intensity	$Y_{t;i}$	u.c/per.	amount of cars passing through the output detector
saturated flow	S_i	u.c/per.	saturated flow - max. amount of cars that can go through the arm i of the intersection
turning rate	α_{ji}	%	ratio of cars that from direction j turn to the direction i
green time	$z_{t;i}$	%	ratio of the "green" time and period (TL cycle time)
-	κ_i, β_i, λ_i	-	constants describing linear relation between queue length and occupancy

where
u.c. means unit car
t is time index
i, j denote i-th and j-th arm of the intersection, respectively
TL means the traffic light
per. is period

The data were obtained by the Aimsun simulator (see www.aimsun.com). For the intersection specification, the SU model equations (1) and (2) are used. Here, the system output consist of the occupancies O_t and output intensities Y_t on the individual intersection arms for the whole workweek (five days). Total amount of the data entries is about 5.000; the data are sampled with the period of 90 second. The green time z_t is the system input. The queue length ξ_t is the state that is estimated together with the parameters κ, β, λ. The known part of the model matrices are composed from S_i, α_{ji}, I_t.

The estimation was running with the various memory length ∂. For the result evaluation, the mean error (ME) of the state estimates is used

$$ME = \frac{1}{\mathring{t}} \sum_{t=1}^{\mathring{t}} |\hat{x}_t - x_t|$$

where $\hat{x}_t$ is estimated state, x_t is the simulated state, $\mathring{t}$ is the number of the data samples. The ME for the particular intersection arms are on the Fig. 1

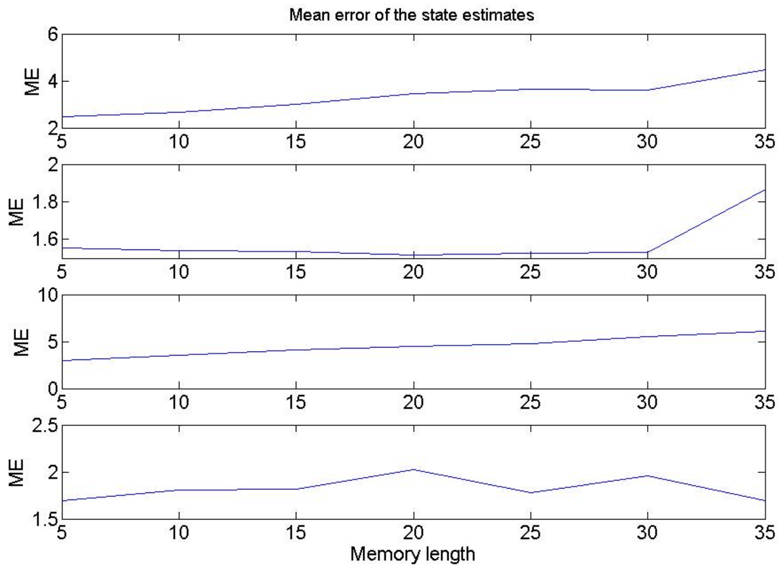

Figure 1: Mean errors of queue length estimates depending on ∂

The experiment shows that the best results are reached with the small ∂. We suppose that it is caused by the continuously changing character of traffic flow. Therefore the shorter memory length is ∂ is suitable for the estimation.

5 Conclusion

5.1 Achieved results

The proposed approach provides the following advantages: (i) it allows estimation of the innovation range and (ii) it allows (without excessive computational demands) to respect ,,naturally" hard, physically given, prior bounds on model parameters and states, (iii) 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) (iv) it enables parameter tracking through the memory length ∂, (v) it provides an easy entry of of the partial knowledge on the parameters and (vi) it opens a way for Bayesian filtering of non-linear systems.

5.2 Future research

The following research aims to improve the quality of the approximation. Instead of the proposed ,,cutt off" approximation, the non-uniform pdf $g(x_{t-\partial}|u_{t-1}, d^{t-\partial:t-2}, \Theta)$ will be substitute the uniform one $\tilde{f}(x_{t-\partial}|u_{t-1}, d^{t-\partial:t-2}, \Theta)$ so that g and $\tilde{f}$ will have the same support.

Further, the SU model forms the starting point for the non-uniform models with restricted support.

Acknowledgements: The paper has been supported by the Research Center DAR, MŠMT 1M0572.

References

[1] E. F. Camacho and C. Bordons *Model Predictive Control.* Springer, 2004.

[2] 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.

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

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

[5] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag, 1985.

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

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