SWAPPING BASED JOINT ESTIMATION OF UNIFORM STATE MODEL

L. Pavelková

Institute of Information Theory and Automation
Department of Adaptive Systems
P.O.Box 18, 182 08 Prague 8, Czech Republic

ABSTRACT

The paper presents an algorithm for the on-line joint parameter and state estimation of the state model whose innovations are uniformly distributed. We use a Bayesian approach and evaluate a maximum a posteriori probability (MAP) estimates in discrete time instants. As the model innovations have a bounded support, the searched estimates lie within a set that is described by the system of inequations. In consequence, the problem of MAP estimation can be easily converted to the problem of linear programming. A joint state and parameter estimation is performed as the alternating subtasks of state filtration and parameter estimation. The resulting estimation algorithm is applied to the traffic data.

Index Terms— state model, Bayesian learning, state filtration, parameter estimation, uniform innovations

1. INTRODUCTION

The state space model is frequently used for the description of a real system. Here, the subtasks of the state filtration and parameter estimation arise. The innovations of state evolution as well as observation model are often supposed to have normal distribution and the problem is then solved by means of a Kalman filtering, see e.g. [1]. The unbounded support of the Gaussian distribution of innovations can cause difficulties in some cases, where the estimated quantities are physically restricted, e.g., in modelling of transportation systems [2].

Then, techniques similar to those dealing with unknown-but-bounded equation errors are used. They often intentionally give up stochastic interpretation of the innovations and develop and analyze various algorithms of a min-max type, cf. [3]. 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 [4], [5].

We aim to keep the tools of the probabilistic approach together with the characteristics of unknown-but-bounded errors techniques. The introduced state uniform (SU) model fulfills these requirements. We suppose that the system is described by the state model with uniformly distributed innovations. To obtain on-line estimates of the states and parameters, we use the Bayesian approach [6] and stem from the principle moving horizon estimation [7]. The theoretical background of the SU model is founded in the authors PhD thesis [8]. There, the algorithm for on-line joint state and parameter estimation uses the linearization around the newest estimates from previous step. Here, we summarize achieved results and propose another way how a joint parameter and state estimation can be performed.

Throughout, we use the following notation

$\equiv$	equality by definition
$\propto$	equality up to a constant factor (proportionality)
z^*	a set of z-values, $z \in z^*$
z^ℓ	the length of the column vector z
z_t	value of quantity z in discrete time instant t; $t \in t^* \subset \{0, 1, 2, \dots T\}$
$z^{k:l}$	the ordered sequence; $z^{k:l} \equiv [z'_k, z'_{k+1}, \dots, z'_l]'$, $0 \le k \le l$
$'$	transposition
$\underline{z}$, $\overline{z}$	lower and upper bound on z, respectively; they are used entry-wise
$f(\cdot\|\cdot)$	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

Note that integrals used are always definite and multivariate ones. The integration domain coincides with the support of the pdf in its argument. Note that vectors are always columns.

2. BASICS OF BAYESIAN LEARNING

In Bayesian view [6], [9], the system is described by probability density functions (pdfs). The quantities describing the system consist generally of observable outputs $y^{1:T}$, optional inputs $u^{1:T}$ and internal quantities $X^{0:T}$ that are never observed directly. $X^{0:T}$ consists of system states $x^{0:T}$ and/or a time invariant parameter Θ, i.e., $X^{0:T} \subset \left\{x^{0:T}, \Theta\right\}$. The collection of the outputs and inputs is called data and denoted

This work was supported partially by the grant GAČR 102/07/1596 and by the Research Center DAR - MŠMT 1M0572.

978-1-4244-2710-9/09/$25.00 © 2009 IEEE

$d^{1:T}$, i.e. $d_t = (y_t, u_t)$, $t \in t^* = \{1, \ldots, T\}$. The joint pdf

$$f(d^{1:T}, X^{0:T}) \tag{1}$$

describing both observed and internal quantities can be decomposed onto a product of the following elements [6]:
- observation model $\{f(y_t|u_t, d^{1:t-1}, X_t)\}_{t\in t^*}$
- time evolution model $\left\{f(X_t|u_t, d^{1:t-1}, X_{t-1})\right\}_{t\in t^*}$
- controller $\{f(u_t|d^{1:t-1}) \equiv f(u_t|d^{1:t-1}, X^{1:t-1})\}_{t\in t^*}$; here the validity of the natural conditions of control [6] is supposed i.e., $X^{1:t-1}$ is unknown to the controller
- prior pdf $f(X_0) \equiv f(X_0|d_0)$

The evolution of the pdf $f(X_t|d^{1:t-1})$, called Bayesian filtering of unknown internal quantities X_t, is described by the following recursion that starts from the prior pdf $f(X_0)$

$$\begin{aligned} f(X_t|d^{1:t}) &\propto f(y_t|u_t, d^{1:t-1}, X_t) f(X_t|u_t, d^{1:t-1}) \qquad (2)\\ &= f(y_t|u_t, d^{1:t-1}, X_t) \\ &\times \int f(X_t|u_t, d^{1:t-1}, X_{t-1}) f(X_{t-1}|d^{1:t-1})\, dX_{t-1} \end{aligned}$$

The described Bayesian filtering combines prior information in $f(X_0)$, theoretical knowledge described by $f(y_t|u_t, d^{1:t-1}, X_t)$, $f(X_t|u_t, d^{1:t-1}, X_{t-1})$ and observed data $d^{1:t} = (y^{1:t}, u^{1:t})$ by using deductive rules of the calculus with pdfs. Note that a controller is missing in (2); it doesn't depend on the internal quantity X_t and therefore plays no role in its estimation. We don't need to know controller strategy but only the generated input values.

3. STATE MODEL WITH UNIFORM INNOVATIONS

3.1. Model description

The proposed model is composed of the following state (3) and observation (4) equations

$$x_t = A_t x_{t-1} + B_t u_t + F_t + w_t \tag{3}$$
$$y_t = C_t x_t + D_t u_t + G_t + e_t, \tag{4}$$

where
x_t, u_t, y_t are unobserved state, known input and observed output vectors respectively;
A_t, B_t, F_t, C_t, D_t, G_t are model matrices of appropriate dimensions; the known entries of the model matrices can be time variant; the unknown entries are supposed to be time invariant on considered estimation horizon and (collected into the column vector θ) they form a part of estimated Θ;
w_t, e_t are the vectors of the state and output innovations respectively, they are assumed to have uniform distribution

$$f(w_t) = \mathcal{U}(0, p), \quad f(e_t) = \mathcal{U}(0, r) \tag{5}$$

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

To collect all estimated parameters, we denote

$$\Theta \equiv [\theta', p', r']', \tag{6}$$

where θ contains the unknown entries of model matrices converted into a column vector.

Further, we suppose that x_0 and Θ are mutually independent and uniformly distributed on the set $\mathcal{S}0$

$$\mathcal{S}0 = \left\{\underline{x}_0 \leq x_0 \leq \overline{x}_0,\, \underline{\Theta} \leq \Theta \leq \overline{\Theta}\right\}. \tag{7}$$

Possible restrictions on the state values are given by

$$\mathcal{S}2 = \{\underline{x} \leq x_t \leq \overline{x}\}, t \in t^* = \{1, 2, \ldots, T\}. \tag{8}$$

Equations (3) and (4) together with the assumptions (5) and (7), (8) define the state uniform model (SU model).

The joint pdf (1) of data $d^{1:T}$, the state trajectory $x^{0:T}$ and parameter Θ of the SU model takes the form (see [8] for detail description)

$$f\left(d^{1:T}, x^{0:T}, \Theta\right) \propto \left[\prod_{i=1}^{x^\ell}(p_i) \prod_{j=1}^{y^\ell}(r_j)\right]^{-T} \chi(\mathcal{S}) \tag{9}$$

where $\chi(\mathcal{S})$ is the indicator of the support $\mathcal{S}$,

$$\mathcal{S} = \mathcal{S}0 \cap \mathcal{S}1 \cap \mathcal{S}2 \tag{10}$$

with $\mathcal{S}0$ given by (7) and $\mathcal{S}2$ by (8); $\mathcal{S}1$ is specified by (3) and (4) with lower and upper innovations bounds given by (5).

3.2. On-line state and parameter estimation

The real-time (on-line) estimation provides the state and/or parameter estimates in each time step. We use the principle of a moving horizon estimators [7] and perform the standard Bayesian filtering (2) on a sliding window of the length $\delta \geq 1$, i.e. we work with the data $d^{t-\delta:t}$. The internal quantity X_t, $t \in t^*$ consists either of the states $x^{t-\delta:t}$ or of the parameter Θ or of both of them. The superfluous state $x_{t-\delta-1}$ and data item $d_{t-\delta-1}$ are integrated out from the posterior pdf in every time step t. This integration induces non-uniform term in the posterior pdf. In the time instant $t \in t^*$, this term is described by a power function containing the powers up to t. With increasing t, the estimation becomes intractable because of increasing complexity of the support of the posterior pdf. We apply an approximation of the non-uniform term in each step [8]. Its principle consist in the replacing of the oldest state by its point estimate from the previous step. The approximate joint pdf takes then the form

$$f(d^{t-\delta:t}, x^{t-\delta:t}, \Theta) \propto \left[\prod_{i=1}^{x^\ell}(p_i) \prod_{j=1}^{y^\ell}(r_j)\right]^{-(\delta+1)} \chi(\tilde{\mathcal{S}}_t) \tag{11}$$

with $t \in t^* \equiv \{\delta+1,\dots,T\}$, $1 < \delta \le T$
where $\chi(\tilde{\mathcal{S}}_t)$ is the indicator of the support $\tilde{\mathcal{S}}_t$, $\tilde{\mathcal{S}}_t$ stems from (10) and holds

$$\tilde{\mathcal{S}}_t = \tilde{\mathcal{S}}0_t \cap \tilde{\mathcal{S}}1_t \cap \tilde{\mathcal{S}}2_t \tag{12}$$

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

$$\begin{aligned}\tilde{\mathcal{S}}1_t = \{ & -p \le x_\tau - {}^cA_\tau x_{\tau-1} - {}^cB_\tau u_\tau - {}^cF_\tau \le p, \\ & -r \le y_\tau - {}^cC_\tau x_\tau \ - {}^cD_\tau u_\tau - {}^cG_\tau \le r \},\end{aligned} \tag{14}$$

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

$\tau \in \{t-\delta,\dots,t\}$, $t \in \{\delta+1,\dots,T\}$.

The on-line estimation of $X_t \subset \{x^{t-\delta:t}, \Theta\}$ consist in the evaluation of the pdf

$$f\left(X_t \middle| d^{t-\delta:t}\right),\ t \in \{\delta+1,\dots,T\},\ 1 < \delta \le T, \tag{16}$$

We focus on a maximum a posteriori (MAP) estimation, see e.g. [10], that provides a point estimate of an unobserved quantity X_t. Here, the MAP estimate of X_t has the following form [8]

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

where $\tilde{\mathcal{S}}_t$ is given by (12); the inequalities in $\tilde{\mathcal{S}}1_t$ are reorganized so that terms containing X_t are on the left side.

To solve this problem, we use the method of the linear programming (LP) [11],

$$\begin{aligned}&\text{Find a vector } X_t \text{ such that } J \equiv \mathcal{C}' X_t \\ &= \sum_{i=1}^{x^\ell} p_i + \sum_{j=1}^{y^\ell} r_j \ \to \min \\ &\text{while } \mathcal{A}_t X_t \le \mathcal{B}_t,\ \underline{X}_t \le X_t \le \overline{X}_t, t \in t^*\end{aligned} \tag{18}$$

where
$\mathcal{C}' \equiv [\mathbf{0}'_{(X^\ell - x^\ell - y^\ell)}, \mathbf{1}'_{(x^\ell + y^\ell)}]$; $\mathbf{0}_{(len)}$, $\mathbf{1}_{(len)}$ 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).

Note that in the case of the joint estimation of the parameters and states, the conditions of the linearity are not fulfilled because of the terms $A_t x_{t-1}$ in (3) and $C_t x_t$ in (4). In the authors thesis [8], these terms were linearized around their point estimates from the previous step using the Taylor expansion.

In the next subsection, we propose alternative method how the joint state and parameter estimates can be obtained.

3.3. Swapping based on-line joint estimation

The idea of the proposed approach is to perform the task of joint estimation as two subsequent steps for each discrete time instant. First, in the time $t \in t^*$, we estimate the state $x^{t-\delta:t}$ with parameter Θ fixed at its last point estimate $\hat{\Theta}$. Then, we use the resulting estimates of states, $\hat{x}^{t-\delta:t}$, to obtain new estimates $\hat{\Theta}$. Initial values of the parameter estimate can be found in off-line mode using, for instance, sampling methods [12].

4. ILLUSTRATIVE EXAMPLE

The proposed algorithm is applied on the traffic data. We estimate the length of the car queue that forms on an arm of the controlled intersection. The intersection is described by the following quantities [13] where $u.c.$ means unit car, $per.$ is sampling period, TL means the traffic light: (i) **known** - ratio of the "green time" of TL and sampling period z_t [%], maximum amount of cars that is able to pass through the intersection (saturated flow) $S\,[u.c/per.]$ (ii) **measured** on the traffic detectors - amount of cars passing through the input detector (input intensity) $I_t\,[u.c/per.]$, amount of cars passing through the output detectors (output intensity) $Y_t\,[u.c/per.]$, relative time of input detector activation (occupancy) O_t [%] (iii) **estimated** - number of the cars before the TL (queue length) $\xi_t\,[u.c.]$, constants κ, β, λ describing linear relation between ξ_t and O_t.

The SU model (3), (4) and (5) of considered arm of the intersection is constructed as follows. The system output $y_t = [Y_t; O_t]$, the input $u_t = z_t$, the state $x_t = [\xi_t; O_t]$, the estimated parameter $\Theta = [\kappa,\, \beta,\, \lambda,\, p,\, r]$. The model matrices are as follows

$$A_t = \begin{bmatrix} s_t & 0 \\ \kappa & \beta \end{bmatrix},\ B_t = \begin{bmatrix} -s_t\, S - (1-s_t)\, I_t \\ 0 \end{bmatrix},$$

$$F_t = \begin{bmatrix} I_t \\ \lambda_t \end{bmatrix},\ C_t = \begin{bmatrix} (1-s_t) & 0 \\ 0 & 1 \end{bmatrix},$$

$$D_t = \begin{bmatrix} s_t\, S + (1-s_t)\, I_t \\ 0 \end{bmatrix},\ G_t = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

$s_t = \left(1 + e^{S z_t - \hat{\xi}_t + I_t z_t}\right)^{-1} \in [0; 1]$ is a queue indicator ($s_t = 0$ for none or small queue); it is computed separately using $\hat{\xi}_t$, i.e., the estimate of ξ_t.

Note that we present only an estimation of one arm of the intersection for the sake of simplicity. In the case of estimation of the complete intersection, we have to include also turning rates because the cars turn into different output arms.

For the experiment, a one day data collection was used with sampling period 80 seconds, i.e. 1080 data entries. The estimation ran with the memory length $\delta = 5$.

In the example, initial values of κ, β, λ were given at the beginning of estimation. These parameters are not constant but "slightly varying" during the day depending on the traffic conditions. We suppose its permanence on the window δ.

These slow variations motivate us for the use of joint state and parameter estimation.

Figure 1 shows the histograms of errors of queue length estimates E_t computed as the absolute value of difference between the real and estimated values of ξ_t; $E_t = |\xi_t - \hat{\xi}_t|$. Here, the swapping based joint estimation improve the accuracy of state estimates (compared to the "pure" state estimation that works constant parameter values). We have observed that this algorithm is suitable for "slightly varying" parameters because the sliding window of length δ works as a forgetting factor. The algorithm requires good initial estimates of parameter.

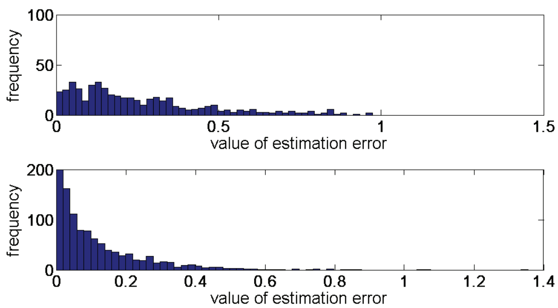

Fig. 1. Comparison of errors E_t for joint state and parameter estimation (down) and state estimation with parameter fixed at their initial values (up)

Note that entry of boundaries on states and parameters is very easy. It consist in modification of sets $\mathcal{S}0_t$ (13) and $\mathcal{S}2_t$ (15). This is equal to adding of inequalities (constraints) into the linear programming. Here, $\xi_t \geq 0$, $0 \leq O_t \leq 100$.

5. CONCLUDING REMARKS

The proposed approach provides the following advantages: (i) it allows to respect "naturally" hard, physically given, prior bounds on model parameters and states, (ii) it enables the joint estimation of parameters, state, and innovation ranges, (iii) it provides an easy entry of of the partial knowledge on the model matrices by the decomposition of model matrices into known and estimated parts, (iv) it updates estimates on the whole window of the length δ, (v) it enables parameter tracking through the memory length δ.

The following research aims at further improving of the estimates quality. Till now, we worked only with MAP estimates. We aim to refine our results by an estimation of the precision of these point estimates. The original MAP estimation task gives result $\hat{X}_t$ for the posterior pdf $f(X^{1:T}|d^{1:T})$. We aim to find the approximative uniform pdf $\tilde{f}(X^{1:T}|d^{1:T}) = \mathcal{U}_X\left(\hat{X}_t, R\right)$ so that the distance between f and $\tilde{f}$ is minimal.

6. REFERENCES

[1] M. S. Grewal and A. P. Andrews, *Kalman Filtering: Theory and Practice Using MATLAB*, Wiley, 2008.

[2] 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*, vol. 1831, pp. 47–56, 2003.

[3] 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, pp. 107–132, 1995.

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

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

[6] M. Kárný, J. Böhm, T. V. Guy, L. Jirsa, I. Nagy, P. Nedoma, and L. Tesař, *Optimized Bayesian Dynamic Advising: Theory and Algorithms*, Springer, London, 2005.

[7] C.V. Rao, J.B. Rawlings, and D.Q. Mayne, "Constrained state estimation for nonlinear discrete-time systems: stability and moving horizon approximations," *IEEE Transactions on Automatic Control*, vol. 48, no. 2, pp. 246– 258, 2003.

[8] L. Pavelková, *Estimation of Models with Uniform Innovations and its Application on Traffic Data*, Ph.D. thesis, Czech Technical University in Prague, Faculty of Transportation Sciences, December 2008, http://simu0292.utia.cas.cz/bibl/detail.php?BID=0323418.

[9] V. Peterka, "Bayesian system identification," in *Trends and Progress in System Identification*, P. Eykhoff, Ed., pp. 239–304. Pergamon Press, Oxford, 1981.

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

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

[12] P. Li, R. Goodall, and V. Kadirkamanathan, "Estimation of parameters in a linear state space model using a Rao-Blackwellised particle filter," *IEE Proceedings-control theory and applications*, vol. 151, pp. 728 – 738, 2004.

[13] J. Homolová and I. Nagy, "Traffic model of a microregion," in *Preprints of the 16th World Congress of the International Federation of Automatic Control - (Horáček, P.; Šimandl, M.; Zítek, P.)*, Prague, 2005.