Approximate Recursive Bayesian Estimation of State Space Model with Uniform Noise

Lenka Pavelková[1] and Ladislav Jirsa[1]
[1]*Institute of Information Theory and Automation, The Czech Academy of Sciences, Pod Vodárenskou věží 4, Prague Czech Republic*
{pavelkov, jirsa}@utia.cas.cz

Keywords: State estimation, State-space models, Linear systems, Bounded noise, Probabilistic models, Approximate estimation, Recursive estimation

Abstract: This paper proposes a recursive algorithm for the state estimation of a linear stochastic state space model. A model with discrete-time inputs, outputs and states is considered. The model matrices are supposed to be known. A noise of the involved model is described by a uniform distribution. The states are estimated using Bayesian approach. Without using an approximation, the complexity of the posterior probability density function (pdf) increases with time. The paper proposes an approximation of this complex pdf so that a feasible support of the posterior pdf is kept during the estimation. The state estimation consists of two stages, namely the time and data update including the mentioned approximation. The behaviour of the proposed algorithm is illustrated by simulations and compared with other methods.

1 INTRODUCTION

A state space model is frequently used for a description of real systems. The unobserved states are estimated using measured data, i. e., system inputs and outputs, as well as modelled dependencies among particular states. Uncertainties of a state evolution model and of an observation model are often supposed to have normal distribution. Then, the states are standardly estimated by means of Kalman filtering (KF) (Jazwinski, 1970) and its extensions. However, the unbounded support of the Gaussian distribution can cause difficulties if the estimated quantity is physically bounded as, for instance, it may give unreasonable negative estimates of naturally non-negative variable. There are several ways to deal with this case.

In the KF framework, the state estimates can be projected onto the constraint surface via quadratic programming (Fletcher, 2000) or the Gaussian distribution is truncated (Simon and Simon, 2010).

A robust recursive Kalman-like algorithm for the state estimation of linear models with disturbances bounded by ellipsoids is proposed in (Becis-Aubry et al., 2008). The proposed algorithm consists of two steps: time updating and observation updating.

A zonotopic Kalman filter (ZKF) is proposed in (Combastel, 2015). Discrete-time LTV/LPV systems with state and measurement uncertainties are considered. ZKF computes minimal zonotopic sets enclosing all the admissible states. Explicit links between the zonotopic set-membership and the stochastic paradigms for Kalman filtering are given.

The papers (Lang et al., 2007) and (Shao et al., 2010) investigate constrained Bayesian state estimation problems by using a particle filter (PF) approach. In these papers, algorithms are proposed. inequality constraints are imposed by accept/reject steps in the algorithms. The Monte-Carlo methods require, however, a huge amount of samples to obtain acceptable results.

In the paper (Dabbene et al., 2014), a rapprochement between the stochastic and worst-case system identification viewpoints is presented. There, the so-called worst-case radius of information is decreased at the expense of a given probabilistic "risk". A case of uniformly distributed noise is supposed. A trade-off curve is constructed which shows how the radius of information decreases as a function of the accuracy.

In the paper (Chisci et al., 1996), the problem of recursively estimating the state of a discrete-time linear dynamical system subject to bounded disturbances is addressed. An approach based on minimum-volume bounding parallelotopes is introduced and an algorithm of polynomial complexity is derived. The estimates are intended in a "set-membership" sense.

In (Pavelková and Kárný, 2014), joint parameter and state estimation is proposed for linear state-space model with uniform state and output noises. The proposed approximate Bayesian estimator provides the maximum a posteriori estimate enriched by information on its precision without demanding the user to tune noise covariances.

Inspired by (Combastel, 2015) and (Chisci et al., 1996), we extend our previous results of an approximate Bayesian estimation for regressive model (Pavelková and Jirsa, 2017) to the state space model and propose an estimator that provides a state estimate of a linear state-space model with a uniform noise. We use Bayesian filtration applied to uniform pdfs. The approach is probabilistic, the method explicitly operates on pdfs using the general theory. The simple recursive algorithm gives a probabilistic estimate that is kept in a given class of functions. The approximate posterior probability function has a parallelotopic support.

Throughout the paper, the following notation will be used: z_t is the value of a column vector z at a discrete-time instant $t \in t^\star \equiv \{1,2,\ldots,\bar{t}\}$; $z_{t;i}$ is the i-th entry of z_t; ℓ_z is the length of the vector z; $\underline{z}$ and $\overline{z}$ are lower and upper bounds on z, respectively; $\equiv$ means equality by the definition, $\propto$ means equality up to a constant factor. The symbol $f(\cdot|\cdot)$ denotes a conditional probability density function (pdf); names of arguments distinguish respective pdfs; no formal distinction is made between a random variable, its realisation and an argument of the pdf. $\mathcal{U}_z(\underline{z},\overline{z})$ denotes the uniform pdf of z with support $[\underline{z},\overline{z}]$.

2 ADDRESSED PROBLEM

A controlled system can be described by the set of ℓ_y-dimensional observable outputs y_t, of ℓ_u-dimensional system inputs u_t, and ℓ_x-dimensional unobservable system states x_t, $t \in t^\star \equiv \{1,2,\ldots,\bar{t}\}$. The input-output pair is called data, i.e. $d_t = (u_t, y_t)$.

In the considered Bayesian set up (Kárný et al., 2005), the system is modelled by pdfs. Using the chain rule and considering the independent input sequence $u_0, u_1, \ldots, u_{\bar{t}-1}$, the joint pdf of all involved variables, $f(d_1,\ldots,d_{\bar{t}},x_0,\ldots,x_{\bar{t}})$, can be factorised to the product of factors (1). Note that we use a different factorisation comparing to (Kárný et al., 2005). There, the time evolution model has the form $f(x_t|x_{t-1},u_t)$.

We factorise the joint pdf as follows

$$f(d_1,\ldots,d_{\bar{t}},x_0,\ldots,x_{\bar{t}}) = \tag{1}$$
$$= \underbrace{f(x_0)}_{\text{prior pdf}} \prod_{t=1}^{\bar{t}} \underbrace{f(u_t)}_{\text{input generator}} \times$$
$$\times \prod_{t=1}^{\bar{t}} \underbrace{f(y_t|x_t)}_{\text{observation model}} \underbrace{f(x_t|x_{t-1},u_{t-1})}_{\text{time evolution model}}.$$

The resulting form assumes that (i) state x_t satisfies Markov property and (ii) no direct relationship between input and output exists in the observation model.

The Bayesian state estimation or filtering (Kárný et al., 2005) consists in the evolution of the posterior pdf $f(x_t|d(t))$, $d(t) \equiv \{d_1, d_2, \ldots, d_t\}$ is a sequence of observed data records $d_t = (y_t, u_t)$, $t \in t^\star$, $d_0 \equiv u_0$. The evolution of $f(x_t|d(t))$ is described by the recursion that starts from the prior pdf $f(x_0|d(0)) \equiv f(x_0)$:

- Time update

$$f(x_t|d(t-1)) = \int_{x^\star_{t-1}} f(x_t|u_{t-1},x_{t-1}) f(x_{t-1}|d(t-1))\,\mathrm{d}x_{t-1} \tag{2}$$

that reflects the time evolution $x_{t-1} \to x_t$ and

- Data update

$$f(x_t|d(t)) = \frac{f(y_t|x_t) f(x_t|d(t-1))}{f(y_t|d(t-1))} =$$
$$= \frac{f(y_t|x_t) f(x_t|d(t-1))}{\int_{x^\star_{t-1}} f(y_t|x_t) f(x_t|d(t-1))\mathrm{d}x_{t-1}} \tag{3}$$

that incorporates information about data d_t.

Here, we focus on a linear model state space model in the form

$$\begin{aligned} x_t &= \underbrace{Ax_{t-1} + Bu_{t-1}}_{\tilde{x}_t} + \nu_t, \quad \nu_t \sim \mathcal{U}_\nu(-\rho,\rho) \\ y_t &= \underbrace{Cx_t}_{\tilde{y}_t} + n_t, \qquad\qquad n_t \sim \mathcal{U}_n(-r,r) \end{aligned} \tag{4}$$

where A, B, C are the known model matrices of appropriate dimensions, $\nu_t \in (-\rho,\rho)$ is the uniform state noise, $n_t \in (-r,r)$ is the uniform output noise.

Equivalently, using pdf notation

$$\begin{aligned} f(x_t|u_{t-1},x_{t-1}) &= \mathcal{U}_x(\tilde{x}_t - \rho, \tilde{x}_t + \rho) \\ f(y_t|x_t) &= \mathcal{U}_y(\tilde{y}_t - r, \tilde{y}_t + r). \end{aligned} \tag{5}$$

State estimation of (5) according to (2) and (3) leads to a very complex form of posterior pdf. This paper proposes an approximate Bayesian state estimation of the linear state space model with uniform noise (LSU model) (4) where the posterior pdf is uniformly distributed on a parallelotopic support.

3 ALGORITHMIC SOLUTION

Here, the approximate state estimation of model (5) is proposed. The presented algorithm provides the evolution of the approximate posterior pdf $f(x_t|d(t))$. The proposed algorithm needs a knowledge about the noise bounds ρ and r in (5). These bounds are generally unknown. To obtain their estimates, the algorithm as proposed by author in (Pavelková and Kárný, 2014) can be used. It provides the point estimates of respective noise bounds of (5).

3.1 Approximate time update

3.1.1 Exact computation

The time update according to (2) starts at $t = 1$ with $f(x_{t-1}|d(t-1)) = f(x_0)$. We suppose that the prior pdf $f(x_0)$ is uniform on an orthotopic support,

$$f(x_0) = \mathcal{U}_{x_0}(\underline{x}_0, \overline{x}_0). \qquad (6)$$

In the next steps, without approximation, the prior pdf $f(x_{t-1}|d(t-1))$ would be generally non-uniform with a polytopic support. The below proposed double approximation keeps the uniform orthotopic form of $f(x_{t-1}|d(t-1))$, i.e. $f(x_{t-1}|d(t-1)) = \mathcal{U}_{x_{t-1}}(\underline{x}_{t-1}, \overline{x}_{t-1})$. Then, (2) gives

$$f(x_t|d(t-1)) =$$

$$= \int_{x^\star_{t-1}} \mathcal{U}_{x_t}(\tilde{x}_{t-1} - \rho, \tilde{x}_{t-1} + \rho)\, \mathcal{U}_{x_{t-1}}(\underline{x}_{t-1}, \overline{x}_{t-1})\, \mathrm{d}x_{t-1} =$$

$$= \frac{1}{|\det(A)|} \prod_{i=1}^{\ell_x} \frac{1}{2\rho_i(\overline{x}_{t-1;i} - \underline{x}_{t-1;i})} \times \qquad (7)$$

$$\times \prod_{i=1}^{\ell_x} \Big(\big[(x_{t;i} - B_i u_{t-1} + \rho_i)\chi(x_{t;i} < B_i u_{t-1} + \overline{m}_{t;i} - \rho_i) +$$

$$+ \overline{m}_{t;i}\chi(x_{t;i} \geq B_i u_{t-1} + \overline{m}_{t;i} - \rho_i) \big] -$$

$$- \big[\underline{m}_{t;i}\chi(x_t; i \leq B_i u_{t-1} + \underline{m}_{t;i} + \rho_i) +$$

$$+ (x_{t;i} - B_i u_{t-1} - \rho_i)\chi(x_{t;i} > B_i u_{t-1} + \underline{m}_{t;i} + \rho_i) \big] \Big) \times$$

$$\times \underbrace{\prod_{i=1}^{\ell_x} \chi(\underline{m}_{t;i} + B_i u_{t-1} - \rho_i \leq x_{t;i} \leq \overline{m}_{t;i} + B_i u_{t-1} + \rho_i)}_{\text{Cutting according to the conditions given by state evolution model.}},$$

where

$$\underline{m}_{t;i} = \sum_{j=1}^{\ell_x} \min(A_{ij}\underline{x}_{t-1;j}, A_{ij}\overline{x}_{t-1;j}), \qquad (8)$$

$$\overline{m}_{t;i} = \sum_{j=1}^{\ell_x} \max(A_{ij}\underline{x}_{t-1;j}, A_{ij}\overline{x}_{t-1;j}),$$

A_{ij} means the term on the i-th row and the j-th column of A. The resulting pdf (7) is trapezoidal.

3.1.2 Approximation

We propose an approximation of the original distribution (7) by a uniform distribution. In (Bernardo, 1979), it is shown that an optimal approximation (in a Bayesian sense) of a pdf by another pdf is achieved by minimisation of Kullback-Leibler divergence (KLD) (Kullback and Leibler, 1951) of these two pdfs. If the approximate pdf is uniform, it keeps the support of the original pdf, see proof in Appendix A.1. Then,

$$f(x_t|d(t-1)) \approx$$

$$\approx \prod_{i=1}^{\ell_x} \frac{\chi(B_i u_{t-1} + \underline{m}_{t;i} - \rho_i \leq x_{t;i} \leq B_i u_{t-1} + \overline{m}_{t;i} + \rho_i)}{\overline{m}_{t;i} - \underline{m}_{t;i} + 2\rho_i} =$$

$$= \prod_{i=1}^{\ell_x} \mathcal{U}_{x_{t;i}}(B_i u_{t-1} + \underline{m}_{t;i} - \rho_i, B_i u_{t-1} + \overline{m}_{t;i} + \rho_i) =$$

$$= \mathcal{U}_{x_t}(B u_{t-1} + \underline{m}_t - \rho, B u_{t-1} + \overline{m}_t + \rho), \qquad (9)$$

where $\underline{m}_t = [\underline{m}_{t;1}, \ldots, \underline{m}_{t;\ell_x}]'$, $\overline{m}_t = [\overline{m}_{t;1}, \ldots, \overline{m}_{t;\ell_x}]'$, $i = 1, \ldots, \ell_x$ are defined by (8).

3.2 Approximate data update

3.2.1 Exact computation

In this step, both u_t and y_t are included. After the data update according to (3), we obtain a posterior pdf with a support in the form of polytope. This polytope results from the intersection of an orthotope obtained during time update and strips given by new data. For details see Appendix A.2. It holds

$$f(x_t|d(t)) = \frac{1}{I_t} \mathcal{U}_{y_t}(Cx_t - r, Cx_t + r) \times$$

$$\times \mathcal{U}_{x_t}(B u_{t-1} + \underline{m}_t - \rho, B u_{t-1} + \overline{m}_t + \rho) \qquad (10)$$

with

$$I_t = \int_{x^\star_t} \mathcal{U}_y(Cx_t - r, Cx_t + r) \times$$

$$\times \mathcal{U}_{x_t}(B u_{t-1} + \underline{m}_t - \rho, B u_{t-1} + \overline{m}_t + \rho)\mathrm{d}x_t.$$

3.2.2 Approximation

We propose an approximation of (10) by a uniform distribution on a parallelotope. For this purpose, we adapt the algorithm from (Vicino and Zappa, 1996). It holds

$$f(x_t|d(t)) \propto$$

$$\propto \chi(B u_{t-1} + \underline{m}_t - \rho \leq x_t \leq B u_{t-1} + \overline{m}_t + \rho) \times$$

$$\times \chi(Cx_t - r \leq y_t \leq Cx_t + r) =$$

$$= \prod_{i=1}^{\ell_x} \chi(B_i u_{t-1} + \underline{m}_{t;i} - \rho_i \leq x_{t;i} \leq B_i u_{t-1} + \overline{m}_{t;i} + \rho_i) \times$$

$$\times \prod_{j=1}^{\ell_y} \chi(y_{t;j} - r_j \leq C_j x_t \leq y_{t;j} + r_j) \qquad (11)$$

where C_j is j-th row of the matrix C. The approximated pdf has the form (see Appendix A.2)

$$f(x_t|d(t)) \approx K_t \chi(\underline{\underline{x}}_t \leq M_t x_t \leq \overline{\overline{x}}_t), \qquad (12)$$

where K_t is a normalising constant.

The resulting posterior pdf (12) has a uniform distribution on a parallelotopic support. The time update (7) in the next step assumes pdf with an orthotopic support, i.e. $f(x_t|d(t)) = \mathcal{U}_{x_t}(\underline{x}_t, \overline{x}_t)$. Therefore, we use "orthotopic" bounds of x_t. These bounds are obtained by circumscription of the parallelotope $\underline{\underline{x}}_t \leq M_t x_t \leq \overline{\overline{x}}_t$ in (12), see Appendix A.3. Then

$$\underline{x}_t \leq x_t \leq \overline{x}_t. \qquad (13)$$

In this way, the recursion is closed and the obtained orthotopic bounds (13) can be used in the next time update step (7) for the computation of the terms $\underline{m}$ and $\overline{m}$ (8).

3.3 Point estimates

State point estimate corresponds to the centre of support parallelotope which is identical to the centre of circumscribing orthotope (Coxeter, 1973). Therefore,

$$\hat{x}_t = \frac{\underline{x}_t + \overline{x}_t}{2}. \qquad (14)$$

Note: Approximation of the posterior pdf (12) on the parallelotopic support by (13) on the orthotopic support, that enters the next time step as the prior pdf preserving the point estimate (14), increases the pdf support, which increases the state uncertainty and plays the role of forgetting.

3.4 Algorithmic summary

Here, the state estimation of model (5) considering known ρ, r is summarised.

Initialisation:

- Choose final time $\bar{t} > 0$, set initial time $t = 0$
- Determine $\underline{x}_0$, $\overline{x}_0$, u_0

On-line

(i) Set $t = t + 1$

(ii) Compute $\underline{m}_t$, $\overline{m}_t$ according to (8)

(iii) Perform data update according to (11)

(iv) Approximate the set $x_t^\star$ by a parallelotope to obtain the form (12) — successive intersections of parallelotope (orthotope) with the strips given by individual rows of output equation and the following approximation by parallelotope (the algorithm in (Vicino and Zappa, 1996))

(v) Compute $\underline{x}_t$, $\overline{x}_t$ (13)

(vi) Compute the point estimate $\hat{x}_t$ (14)

(vii) If $t < \bar{t}$, go to (i)

4 ILLUSTRATIVE EXAMPLE

In this section, the simulative experiments demonstrate the proposed algorithm properties. The algorithm is also compared with the zonotopic Kalman filter (Combastel, 2015), outlined in Section 1, as a similar method with adjustable geometrical complexity.

4.1 Experiment setup

The matrices of the state space model (4) are set as

$$A = \begin{bmatrix} 1.0 & -0.5 & 0.2 \\ 0.5 & 0.1 & 0.0 \\ 0.3 & 0.0 & -0.1 \end{bmatrix}, \quad B = \begin{bmatrix} 0.1 \\ 0.6 \\ 0.3 \end{bmatrix},$$

$$C = \begin{bmatrix} 1.0 & 0.0 & 0.5 \\ 0.0 & 1.0 & 0.5 \end{bmatrix}, \quad \begin{bmatrix} \rho \\ r \end{bmatrix} = \begin{bmatrix} 0.1 \\ 0.3 \end{bmatrix}. \qquad (15)$$

Input is randomly generated as $u_t \sim \mathcal{N}(0,1)$. Length of data sequences $\bar{t} = 100$.

4.2 Results

Example of simulated and estimated states together with the orthotopic bounds (13) is shown in Figure 1, that is zoomed to demonstrate a typical behaviour. Figure 2 shows sensitivity of the algorithm to values of state (ρ) and output (r) noise parameters, examined on the state $x_{t;2}$. Finally, Figure 3 shows dependence of execution time on the state dimension.

The proposed algorithm (called LSU) was compared to estimation on a moving window (Pavelková and Kárný, 2014) (WIN, previously developed by the author) and the zonotopic Kalman filter (Combastel, 2015) (ZKF), see Table 1. Median as a statistic shows the difference more significantly than the mean. Value of q denotes the order of zonotope (for $q = 3 \equiv \ell_x$, a zonotope of order ℓ_x in an ℓ_x-dimensional space is a parallelotope), t runs from 1 to $\bar{t}$.

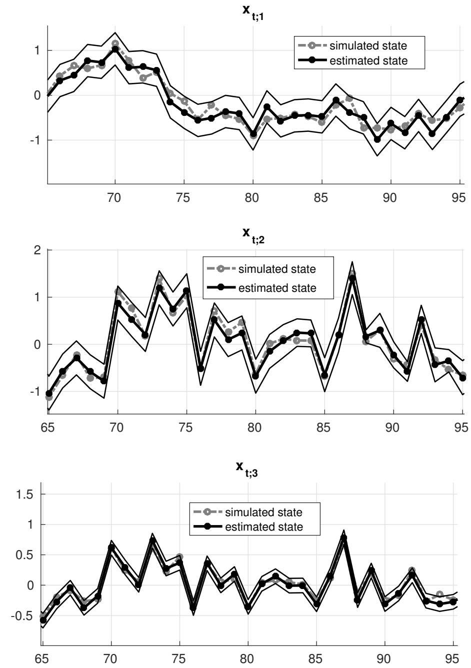

Figure 1: Simulated (dash-dot grey) vs. estimated (solid black) states $x_{t;1}$, $x_{t;2}$ and $x_{t;3}$ with estimated state bounds.

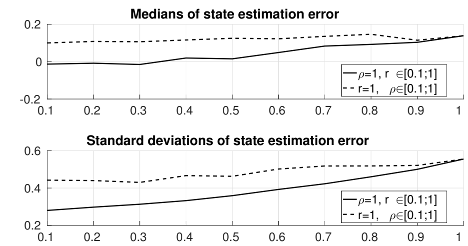

Figure 2: State estimation errors for $x_{t;2}$ depending on state noise ρ and output noise r

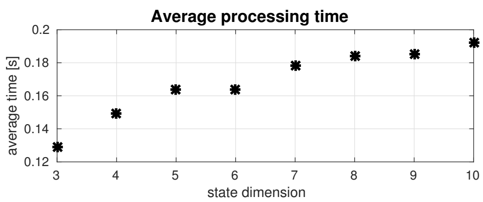

Figure 3: Dependence of average processing time on state dimension

4.3 Discussion

The presented LSU algorithm estimates unknown states as points and bounds (intervals), that contain the simulated values, see Figure 1 and Table 1.

As seen from Figure 2, the state estimation error of the LSU algorithm is more influenced by value of r than of ρ. The parameter r enters the data update (11). The first term in (11) is given by the time update (9), the second term represents the data (measurement) processing. The higher r is, the wider is the data interval in (11) and the more uncertainty (less information) the measurement has. Therefore, with increasing r, the state estimate is more influenced by the time update only and less corrected by the data.

Table 1: Characteristics of estimates, state 2

	LSU	WIN	q=15 ZKF	q=3 ZKF
$\sum_{t=1}^{\bar{t}} (\hat{x}_{t;2} - x_{t;2})^2$	1.770	11.810	1.980	4.810
median$(\hat{x}_{t;2} - x_{t;2})$	0.034	-0.058	0.001	-0.015
std$(\hat{x}_{t;2} - x_{t;2})$	0.133	0.399	0.141	0.220
median$(\hat{y}_{t;1} - y_{t;1})$	10^{-18}	0.032	0.022	-0.011
std$(\hat{y}_{t;1} - y_{t;1})$	10^{-16}	0.197	0.221	0.206
median$(\hat{y}_{t;2} - y_{t;2})$	-0.007	-0.069	-0.031	-0.040
std$(\hat{y}_{t;2} - y_{t;2})$	0.054	0.404	0.222	0.312
median$(\overline{x}_{t;2} - \hat{x}_{t;2})$	0.360	0.464	0.408	0.385
std$(\overline{x}_{t;2} - \hat{x}_{t;2})$	0.053	0.303	0.028	0.023
$\#(x_{t;2} \notin \langle \underline{x}_{t;2}; \overline{x}_{t;2} \rangle)$ [%]	0	9.6	0	7
execution time [s]	0.13	14.41	0.11	0.11

The computational complexity of the LSU algorithm, see Figure 3, is appropriately low and the algorithm is suitable for treating systems of a higher dimension.

According to Table 1, ZKF with higher zonotope order q estimates comparably to LSU but predicts worse and the interval width, represented by the bounds (13), is more conservative. Increasing q did not bring practical improvement. For parallelotopic order ($q = 3$), estimation error of ZKF is higher than for $q = 15$, bounds are tighter and 7 % of states are not contained inside. The WIN algorithm is about 100× slower than both LSU and ZKF and it has the greatest estimation error and amount of not-contained states. Execution time of LSU and ZKF is similar.

5 CONCLUDING REMARKS

An approximate Bayesian filtration algorithm for the state estimation of a linear state space model with uniform noise was proposed. Exact results of time and data update are approximated by a uniform pdf on orthotopic/parallelotopic support to prevent increasing of the computational complexity and to keep the posterior pdf in the given class.

The simple and fast algorithm yields point estimates of the state and its bounds that contain the true value. Prediction error is the lowest of all the compared methods. In this sense, the algorithm performs better than those used for comparison.

Although the WIN algorithm shows the worst results, it provides (unlike the other methods) noise parameters estimates, too. Therefore, it can be used in conjunction with the proposed LSU algorithm in the case of unknown noise parameters.

The proposed estimator can be utilised either directly as it is or it can be used as a local filter e.g. in tasks of static merging within flexible parametric classes (Azizi and Quinn, 2018).

Future work focuses on extension of the proposed algorithm. The extension will include the simultaneous state and noise bounds estimation and involvement of other bounded distributions to generalise the class of used pdfs.

ACKNOWLEDGEMENTS

This research was partially supported by the grant GAČR 18-15970S.

REFERENCES

Azizi, S. and Quinn, A. (2018). Hierarchical Fully Probabilistic Design for Deliberator-Based Merging in Multiple Participant Systems. *IEEE Transactions on Systems Man Cybernetics-Systems*, 48(4):565–573.

Becis-Aubry, Y., Boutayeb, M., and Darouach, M. (2008). State estimation in the presence of bounded disturbances. *Automatica*, 44:1867–1873.

Bernardo, J. M. (1979). Expected information as expected utility. *The Annals of Statistics*, 7(3):686–690.

Chisci, L., Garulli, A., and Zappa, G. (1996). Recursive state bounding by parallelotopes. *AUTOMATICA*, 32(7):1049–1055.

Combastel, C. (2015). Zonotopes and Kalman observers: Gain optimality under distinct uncertainty paradigms and robust convergence. *Automatica*, 55:265 – 273.

Coxeter, H. S. M. (1973). *Regular polytopes*. Courier Corporation.

Dabbene, F., Sznaier, M., and Tempo, R. (2014). Probabilistic optimal estimation with uniformly distributed noise. *IEEE Transactions on Automatic Control*, 59(8):2113–2127.

Fletcher, R. (2000). *Practical Methods of Optimization*. John Wiley & Sons. ISBN: 0471494631.

Jazwinski, A. M. (1970). *Stochastic Processes and Filtering Theory*. Academic Press, New York.

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

Kullback, S. and Leibler, R. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22:79–87.

Lang, L., Chen, W., Bakshi, B. R., Goel, P. K., and Ungarala, S. (2007). Bayesian estimation via sequential Monte Carlo sampling – Constrained dynamic systems. *Automatica*, 43(9):1615–1622.

Pavelková, L. and Jirsa, L. (2017). Recursive Bayesian estimation of uniform autoregressive model using approximation by parallelotopes. *International Journal of Adaptive Control and Signal Processing*, 31(8):1184–1192.

Pavelková, L. and Kárný, M. (2014). State and parameter estimation of state-space model with entry-wise correlated uniform noise. *International Journal of Adaptive Control and Signal Processing*, 28(11):1189–1205. DOI: 10.1002/acs.2438.

Shao, X., Huang, B., and Lee, J. M. (2010). Constrained Bayesian state estimation - A comparative study and a new particle filter based approach. *Journal of Process Control*, 20(1):143–157.

Simon, D. and Simon, D. L. (2010). Constrained Kalman filtering via density function truncation for turbofan engine health estimation. *International Journal of Systems Science*, 41:159–171. www.informaworld.com/10.1080/00207720903042970.

Vicino, A. and Zappa, G. (1996). Sequential approximation of feasible parameter sets for identification with set membership uncertainty. *IEEE Transactions on Automatic Control*, 41(6):774–785.

A APPENDIX

A.1 Approximation of the trapezoidal pdf

According to (Bernardo, 1979), minimisation od Kullback-Liebler divergence (Kullback and Leibler, 1951) (KLD) gives, in a Bayesian sense, an optimal approximation of pdf. KLD of two pdfs, f_1 and f_2, equals $D(f_1||f_2) = \int f_1(x) \ln \frac{f_1(x)}{f_2(x)}\,\mathrm{d}x$.

Denote the set describing the support $\chi^\star \equiv \mathrm{supp}(\chi)$. Given pdf on a bounded support, $f_1(x) = g(x)\chi_1(x)$, where $0 < g(x) < +\infty\ \forall x \in \chi_1^\star$, we search for the optimal approximation by a uniform pdf, $f_2(x) = I^{-1}\chi_2(x)$, where $I = \mathrm{vol}(\chi_2^\star)$. We look for $\hat{f}_2 = \arg\min_{f_2} D(f_1||f_2)$. Function arguments are omitted.

Here, $D(f_1||f_2) = \int g \ln \frac{Ig\chi_1}{\chi_2}\,\mathrm{d}x = \int_{\chi_1^\star} g \ln g\,\mathrm{d}x + \int_{\chi_1^\star} g \ln \frac{\chi_1}{\chi_2}\,\mathrm{d}x + \ln I \int_{\chi_1^\star} g\,\mathrm{d}x$. The first term is independent of f_2, the second term is finite (zero) if $\chi_1^\star \subset \chi_2^\star$. The third term depends on f_2 through I: the larger support of f_2, the higher I. Hence, to minimise KLD, we minimise the measure of $\chi_2^\star$ choosing $\chi_2 = \chi_1$, i.e. $\hat{f}_2$ be the uniform pdf on the support of f_1.

A.2 Approximation of the polytope by a parallelotope

The second term in (11) can be understood as ℓ_y data strips in x_t-space, $y_{t;i} - r_i \leq C_i x_t \leq y_{t;i} + r_i$, $i = 1, \ldots, \ell_y$. These strips intersect with ℓ_x strips given by the first term, which forms a parallelotope (actually orthotope). The intersection defines a polytope which is to be approximated by another parallelotope (12).

Theory and algorithm of the approximation is described in (Vicino and Zappa, 1996). Briefly: (i) One strip is added to the parallelotope and all these $\ell_x + 1$ strips are *tightened* to remove redundancy, i.e. narrowed and/or shifted, so that their intersection is unchanged. (ii) One strip of these $\ell_x + 1$ strips is discarded, so that the intersection of the remaining ℓ_x strips has minimal volume. (iii) The procedure is repeated for all ℓ_y strips in the first term of (11), which gives M_t, $\underline{\underline{x}}_t$ and $\bar{\bar{x}}_t$ in (12). Note that volume of the parallelotope equals $\left|\det\left[M_t^{-1}\mathrm{diag}(\bar{\bar{x}}_t - \underline{\underline{x}}_t)\right]\right|$.

A.3 Circumscription of a parallelotope by an orthotope

The parallelotope defined in (12) is circumscribed by an orthotope to get the bounds $\underline{x}_t$ and $\bar{x}_t$ in (13).

The parallelotope in the boundary form $\underline{\underline{x}}_t \leq M_t x_t \leq \bar{\bar{x}}_t$ can be equivalently written as $-\mathbf{1}_{(\ell_x)} \leq W_t x - c_t \leq \mathbf{1}_{(\ell_x)}$, where $\mathbf{1}_{(\ell_x)}$ is a unit vector of length ℓ_x, $W_{t;ij} = \frac{2}{\bar{\bar{x}}_{t;i} - \underline{\underline{x}}_{t;i}} M_{t;ij}$ and $c_{t;i} = \frac{\bar{\bar{x}}_{t;i} + \underline{\underline{x}}_{t;i}}{\bar{\bar{x}}_{t;i} - \underline{\underline{x}}_{t;i}}$. Defining $T_t = W_t^{-1}$ and $x_{ct} = T_t c_t$, we express the parallelotope in the direct form $x_t = x_{ct} + T_t k$, where the norm $||k||_\infty \leq 1$ and x_{ct} is the central point. Summing absolute values of the i^{th} row elements in T_t, we get the i^{th} coordinate of a vertex that is most distant from the centre in the i^{th} direction. It represents the half-width of a box (orthotope), in the i^{th} direction, tightly containing the parallelotope. Formally, $Q_{t;ii} = \sum_{j=1}^{\ell_x} |T_{t;ij}|$, where the diagonal matrix Q_t defines the circumscribing orthotope $x_t = x_{ct} + Q_t k$. The boundary form (13) is then $\underline{x}_t = x_{ct} - Q_t \mathbf{1}_{(\ell_x)} \leq x_t \leq x_{ct} + Q_t \mathbf{1}_{(\ell_x)} = \bar{x}_t$.