Approximate Bayesian Prediction Using State Space Model with Uniform Noise

Ladislav Jirsa[1], Lenka Kuklišová Pavelková[1], and Anthony Quinn[1 2]

[1] Institute of Information Theory and Automation, Czech Academy of Sciences, Prague, Czech republic,
{jirsa,pavelkov}@utia.cas.cz,
WWW home page: http://www.utia.cas.cz

[2] Department of Electronic and Electrical Engineering, University of Dublin, Ireland,
aquinn@tdc.ie

Abstract. This paper proposes a one-step-ahead Bayesian output predictor for the linear stochastic state space model with uniformly distributed state and output noises. A model with discrete-time inputs, outputs and states is considered. The model matrices and noise parameters are supposed to be known. Unknown states are estimated using Bayesian approach. A complex polytopic support of posterior probability density function (pdf) is approximated by a parallelotopic set. The state estimation consists of two stages, namely the time and data update including the mentioned approximation. The output prediction is performed as an inter-step between the time update and the data update. The behaviour of the proposed algorithm is illustrated by simulations and compared with Kalman filter.

Keywords: stochastic state space model, observation prediction, Bayesian state estimation, uniform noise

1 Introduction

A range of decision-making tasks, such as online prediction, fault detection and feedback control, need an adequate model of the considered system that provides a prediction of the system behaviour. A linear stochastic state space model is often used for this purpose. The states of this model are often unmeasurable. Then, to obtain the required prediction, an estimation of states of this model has to be solved first.

Bayesian filters form an effective tool for solving the state estimation. If the random disturbances entering the model are assumed to be Gaussian, then, fast and efficient estimation algorithms are based on Kalman filter (KF) [1].

However, the involved noise is often bounded in practice. To cope with this problem, the state estimates are projected onto the constraint surface [2] or the Gaussian distribution is truncated [3]. Nevertheless, these techniques in conjunction with the system model having unbounded support respect constraints within the estimation but not within the modelling.

Sequential Monte-Carlo sampling methods cope with constraints in stochastic models very well. The constraints are respected within the accept/reject steps of the algorithm, see e.g. [4]. These methods require, however, a huge amount of samples to obtain results with an acceptable precision.

Giving up the probabilistic approach, the techniques dealing with "unknown-but-bounded error" can be used for the state estimation. A set containing the estimates is of a very high complexity and has to be approximated, see e.g. [5], [6]. However, without the probabilistic interpretation, solution of related decision-making tasks is unnecessarily difficult because a rich set of statistical tools is not at disposal.

A beneficial interconnection of the probabilistic approach and the "unknown-but-bounded errors" estimation approach is presented e.g. in [8, 9]. In [8], an explicit bridge between the set-membership and the stochastic paradigms for Kalman filtering is presented and a zonotopic Kalman filter is proposed. In [9], an approximate Kalman-like Bayesian filter for the linear state space model with uniformly distributed noise (LSU model) is proposed which uses a parallelotopic approximation of the posterior density support.

Having the state estimates, the required prediction can be obtained directly from these estimates in a deterministic way, see e.g. [10]. An adaptive point output predictor is proposed in [11]. Examples of interval predictors are presented in [12, 13, 14].

We aim to enrich the class of predictors with a Bayesian predictor for models with bounded noise. We extend our previous research concerning state estimation of LSU model [9] and introduce a respective one-step-ahead predictor. The approach is probabilistic, the method explicitly operates on pdfs using the general theory. The simple approximate algorithm provides a probabilistic predictor that is kept in a given class of functions.

The paper is organised as follows. The Section 2 gives a brief introduction into the used uniform distribution and a geometric interpretation of involved supports. The Section 3 provides basics of the Bayesian approach and introduces the LSU model and the approximate state estimation of this model. The Section 4 presents main results, namely approximate output predictor for the above mentioned LSU model. The algorithmic summary is given in Section 5. In the Section 6, illustrative experiments are presented that demonstrate the predictor performance and quality in comparison with a predictor based on KF. The Section 7 concludes the paper.

Throughout the paper, the following notation will be used: x_t is the value of a column vector x at a discrete-time instant $t \in t^\star \equiv \{1, 2, \ldots, \overline{t}\}$; $x_{t;i}$ is the i-th entry of x_t; $\underline{x}$ and $\overline{x}$ are lower and upper bounds on x, 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.

2 Multidimensional uniform distribution and geometrical interpretation of its support

This section introduces the uniform pdf used further in the text together with theoretical and technical reasons and principles underlying the involved approximations. A brief discussion of the approximation properties concludes the section.

2.1 Definition of uniform pdf

Characteristic (or indicator) function $\chi_{x^\star}(x)$ on a set $x^\star$ is defined this way: if $x \in x^\star$, then $\chi_{x^\star}(x) = 1$, otherwise 0. Let $x \in \mathbb{R}^n$ and $Mx : \mathbb{R}^n \to \mathbb{R}^m$ be a continuous linear mapping. For $x^\star = \{x \,|\, a \leq Mx \leq b\}$, where $a, b \in \mathrm{R}^m$, M is a $(m \times n)$ matrix of rank n, we use the notation $\chi_{x^\star}(x) \equiv \chi(a \leq Mx \leq b)$. Vector inequality is considered by items.

Uniform pdf $\mathcal{U}_x(a \leq Mx \leq b)$ of a random variable x is a product of the characteristic function $\chi_{x^\star}(x)$ and a normalising constant K, which is a reciprocal value of measure of the set $x^\star$, i.e.,

$$\mathcal{U}_x(a \leq Mx \leq b) = K\chi_x(a \leq Mx \leq b), \tag{1}$$

The set $x^\star$ is called support of $\mathcal{U}_x$. Because $\mathcal{U}_x$ is a positive constant within its support, and zero outside, the support plays a dominant role in definition of the uniform pdf. Alternatively, the normalising constant can be omitted, i.e. $\mathcal{U}_x(a \leq Mx \leq b) \propto \chi_x(a \leq Mx \leq b)$.

If M is the identity matrix, then a simplifying notation can be used,

$$\mathcal{U}_x(a \leq x \leq b) \equiv \mathcal{U}_x(a, b). \tag{2}$$

2.2 Examples of supports

For $x \in \mathbb{R}^n$, where $n \geq 2$, the support $x^\star$ of uniform pdf (1) can be a complex set, depending on the shape of M. Here, M is a matrix $(m \times n)$ of rank n, $m \geq n$, so that the measure K^{-1} of $x^\star$ is finite. The set $x^\star$ is convex, bounded by a finite number (up to $2m$) of hyperplanes (faces).

In the paper, the following three types of sets are applied

Convex polytope $m > n$, computation of a general polytope volume K^{-1} is a complex task [15].

Parallelotope $m = n$, M is a non-diagonal matrix, $K^{-1} = \left|\det\left[M^{-1}\mathrm{diag}(b - a)\right]\right|$, see [9].

Orthotope $m = n$, M is a diagonal matrix (as a special case, M can be identity matrix), $K^{-1} = \prod_{i=1}^{n} (b_i - a_i)/|M_{ii}|$. The orthotope represents a multidimensional interval (a box).

An illustrative example of two-dimensional sets is given in Fig. 1. Note that the respective sets are described by the corresponding system of linear inequalities in the characteristic function (1).

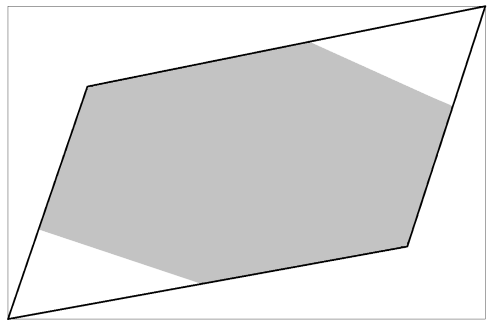

Fig. 1. An example of two-dimensional convex polytope (shaded area), parallelotope (thick line) and orthotope (thin line).

2.3 Approximation of a complex pdf by a uniform pdf with a parallelotopic support

Using a uniform pdf within the Bayesian estimation scheme typically induces non-uniform distribution on a geometrically difficult support. There are two issues that has to be considered:

- finite sufficient statistic does not exist because the uniform pdf is not a member of the exponential family [16]; this property calls for approximation of the support,
- sum of two independent uniformly distributed random quantities is not uniformly distributed (unlike of normally distributed), i.e., the uniform pdf is not closed under summing [17]; this property calls for approximation of the function.

Approximation of the support With increasing number of data processed, the size m of the matrix M in (1) increases, too. The support represented by the polytope becomes more complex as a consequence.

To avoid problems with storing the data, volume evaluation, analytical and computational complexity, memory requirements etc., it is practical to approximate the polytope by a tightly circumscribing parallelotope (m=n) with minimum volume [18].

Parallelotope can be tightly circumscribed by an orthotope in order to get the bounds of the parallelotope. Anyway, the paralellotope describes uncertainty more exactly than the orthotope as it less differs from the original polytope.

Both these approximations are also used for recursiveness: a support is "spoilt" by a processing step it enters. The approximation projects it back to the initial class, i.e. the function is kept, the support is changed.

Approximation of the function Sum of two random variables corresponds to convolution of their pdfs. Convolution of two uniform pdfs is a trapezoidal pdf [17]. Each subsequent convolution increases functional complexity of the resulting pdf.

To preserve the class of the function and computational complexity, we approximate trapezoidal pdf by a uniform pdf by minimising Kullback-Leibler divergence of these pdfs [9]. The result is a uniform pdf on the support of the trapezoid, i.e. the support is kept, the function is changed.

Discussion on the approximations The approximations mentioned above enable both analytical and computational efficiency. The tasks become tractable and they can be formulated recursively.

On the other hand, the approximations are local and propagation of their errors is not under control. Also, both support and function approximations increase uncertainty of the respective random quantities. However, this property can be interpreted as a form of implicit forgetting, useful in adaptive estimation.

3 Approximate Bayesian estimation of uniform state space model

This section introduces basics of the Bayesian filtering, describes the involved stochastic linear state space model with a uniform noise and presents the approximate state estimation of this model.

3.1 Bayesian filtering

In the considered Bayesian set up [19], system is described by the following pdfs:

$$\begin{aligned} \text{prior pdf:} &\quad f\left(x_0\right) \\ \text{observation model:} &\quad f\left(y_t | x_t\right) \\ \text{time evolution model:} &\quad f\left(x_t | x_{t-1}, u_{t-1}\right) \end{aligned} \tag{3}$$

where y_t is a scalar observable output, u_t is a system input, and x_t is an ℓ-dimensional unobservable system state, $t \in t^\star \equiv \{1, 2, \ldots, \bar{t}\}$.

We assume that (i) state x_t satisfies Markov property, (ii) no direct relationship between input and output exists in the observation model, and (iii) the inputs consist of a known sequence $u_0, u_1, \ldots, u_{\bar{t}-1}$.

The Bayesian state estimation or filtering [19] consists in the evolution of the posterior pdf $f(x_t|d(t))$ where $d(t) \equiv \{d_1,\, d_2,\, \ldots,\, d_t\}$ is a sequence of observed data records $d_t = (y_t, u_t)$, $t \in t^\star$. The evolution of $f(x_t|d(t))$ is described by the two-steps recursion that starts from the prior pdf $f(x_0)$:

- Time update

$$f(x_t|d(t-1)) = \int_{x^\star_{t-1}} f(x_t|u_{t-1}, x_{t-1}) f(x_{t-1}|d(t-1))\, \mathrm{d}x_{t-1}, \tag{4}$$

– Data update

$$f(x_t|d(t)) = \frac{f(y_t|x_t)f(x_t|d(t-1))}{\int_{x_t^\star} f(y_t|x_t)f(x_t|d(t-1))\mathrm{d}x_t} = \frac{f(y_t|x_t)f(x_t|d(t-1))}{f(y_t|d(t-1))}. \quad (5)$$

The denominator in (5) represents the Bayesian output predictor.

3.2 Linear state space model with uniform noise

We introduce a linear state space model with a uniform noise (LSU model) in the form

$$\begin{array}{ll} x_t = Ax_{t-1} + Bu_{t-1} + \nu_t, & \nu_t \sim \mathcal{U}_\nu(-\rho, \rho) \\ y_t = Cx_t + n_t, & n_t \sim \mathcal{U}_n(-r, r) \end{array} \quad (6)$$

where A, B, C are the known model matrices/vectors of appropriate dimensions, $\nu_t \in (-\rho, \rho)$ is the uniform state noise with known parameter ρ, $n_t \in (-r, r)$ is the uniform observation noise with known parameter r. Equivalently, using above mentioned pdf notation (3)

$$\begin{aligned} f(x_t|u_{t-1}, x_{t-1}) &= \mathcal{U}_x(\underbrace{Ax_{t-1} + Bu_{t-1}}_{\tilde{x}_t} -\rho, \underbrace{Ax_{t-1} + Bu_{t-1}}_{\tilde{x}_t} +\rho) \\ f(y_t|x_t) &= \mathcal{U}_y(\underbrace{Cx_t}_{\tilde{y}_t} -r, \underbrace{Cx_t}_{\tilde{y}_t} +r). \end{aligned} \quad (7)$$

State estimation of LSU model (7) according to (4) and (5) leads to a very complex form of posterior pdf. In [9], an approximate Bayesian state estimation of this model is proposed. The presented algorithm provides the evolution of the approximate posterior pdf $f(x_t|d(t))$ that is uniformly distributed on a parallelotopic support.

3.3 Approximate state estimation of LSU model

Approximate time update The time update according to (4) starts at the time $t = 1$ with $f(x_{t-1}|d(t-1)) = f(x_0) = \mathcal{U}_{x_0}(\underline{x}_0, \overline{x}_0)$, i.e., $f(x_0)$ is uniformly distributed on an orthotopic support. In next steps, without approximation, the pdf $f(x_{t-1}|d(t-1))$ would be non-uniform and having a polytopic support. A below described double approximation as proposed in [9] keeps the uniform orthotopic form of $f(x_{t-1}|d(t-1))$, i.e. $f(x_{t-1}|d(t-1)) = \mathcal{U}_{x_{t-1}}(\underline{x}_{t-1}, \overline{x}_{t-1})$, $t \in t^\star$. Then, according to (4),

$$f(x_t|d(t-1)) = \frac{1}{|\det(A)|} \prod_{i=1}^{\ell} \frac{1}{2\rho_i(\overline{x}_{t-1;i} - \underline{x}_{t-1;i})} \times \quad (8)$$

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

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

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

where

$$\underline{m}_{t;i} = \sum_{j=1}^{\ell} \min(A_{ij}\underline{x}_{t-1;j} + B_i u_{t-1}, A_{ij}\overline{x}_{t-1;j} + B_i u_{t-1}), \tag{9}$$

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

A_{ij} means the term on the i-th row and the j-th column of A. The resulting pdf (8) is trapezoidal [17].

In [9], the following approximation (based on a minimising of Kullback-Leibler divergence of two pdfs) of the original distribution (8) by a uniform distribution is proposed

$$f(x_t|d(t-1)) \approx \prod_{i=1}^{\ell} \frac{\chi(\underline{m}_{t;i} - \rho_i \leq x_{t;i} \leq \overline{m}_{t;i} + \rho_i)}{\overline{m}_{t;i} - \underline{m}_{t;i} + 2\rho_i} =$$

$$= \prod_{i=1}^{\ell} \mathcal{U}_{x_{t;i}}(\underline{m}_{t;i} - \rho_i, \overline{m}_{t;i} + \rho_i) = \mathcal{U}_{x_t}(\underline{m}_t - \rho, \overline{m}_t + \rho), \tag{10}$$

where $\underline{m}_t = [\underline{m}_{t;1}, \ldots, \underline{m}_{t;\ell}]'$, $\overline{m}_t = [\overline{m}_{t;1}, \ldots, \overline{m}_{t;\ell}]'$, the vectors entries are defined by (9).

Approximate data update Performing the data update of (10) according to (5), we obtain a posterior pdf with a support in the form of polytope

$$f(x_t|d(t)) = \frac{1}{\mathcal{I}_t} \mathcal{U}_{y_t}(Cx_t - r, Cx_t + r)\mathcal{U}_{x_t}(\underline{m}_t - \rho, \overline{m}_t + \rho) \tag{11}$$

with

$$\mathcal{I}_t = \int_{x_t^\star} \mathcal{U}_y(Cx_t - r, Cx_t + r)\mathcal{U}_{x_t}(\underline{m}_t - \rho, \overline{m}_t + \rho)\mathrm{d}x_t.$$

It holds

$$f(x_t|d(t)) \propto \chi(\underline{m}_t - \rho \leq x_t \leq \overline{m}_t + \rho)\chi(Cx_t - r \leq y_t \leq Cx_t + r) =$$

$$= \chi\left(\begin{bmatrix} \underline{m}_t - \rho \\ y_t - r \end{bmatrix} \leq \begin{bmatrix} I \\ C \end{bmatrix} x_t \leq \begin{bmatrix} \overline{m}_t + \rho \\ y_t + r \end{bmatrix}\right) \tag{12}$$

where I is the $(\ell \times \ell)$ identity matrix. The uniform pdf (12) has a polytopic support. In [9], an approximation of (12) by a uniform pdf with a parallelotopic support is proposed. It has the form

$$f(x_t|d(t)) \approx K_t \chi(\underline{\underline{x}}_t \leq M_t x_t \leq \overline{\overline{x}}_t), \tag{13}$$

where K_t is a normalising constant.

The posterior pdf (13) is the result of state filtering in the time t.

Nevertheless, the time update (8) in the next step assumes pdf with an orthotopic support, i.e. $f(x_t|d(t)) = \mathcal{U}_{x_t}(\underline{x}_t, \overline{x}_t)$. The required bounds $\underline{x}_t$, $\overline{x}_t$ are obtained by circumscription of the parallelotope $\underline{\underline{x}}_t \leq M_t x_t \leq \overline{\overline{x}}_t$ in (13) by an orthotope, see Fig. 1 and computational details in [9]. Then

$$\underline{x}_t \leq x_t \leq \overline{x}_t. \tag{14}$$

In this way, the recursion is closed and the obtained orthotopic bounds (14) can be used in the next time update step (8) for the computation of the terms $\underline{m}$ and $\overline{m}$ (9).

Point estimates of states State point estimate corresponds to the centre of circumscribing orthotope [9]

$$\hat{x}_t = \frac{\underline{x}_t + \overline{x}_t}{2}. \tag{15}$$

4 Approximate predictor for scalar output and ℓ-dimensional state

This paper enriches the above described approximated state estimator in an output prediction. The proposed predictor corresponds to the denominator of (5). Its exact computation is a complex task. In this section, an approximated uniform predictor is proposed.

Assume y_t be a scalar. Note that a modelling of a scalar output is sufficient because the chain rule for pdfs [19] implies that the multivariate case can be treated using a collection of such models. Generally, state $x_t \in \mathbb{R}^\ell$ is a vector.

4.1 Predictive pdf

The data predictor of a linear state-space model is the denominator of (5), where $f(y_t|x_t)$ is given by (7) and $f(x_t|d(t-1))$ is the result of the approximate time update (10). Then,

$$f(y_t|d(t-1)) \propto \int_{x_t^\star} \chi(Cx_t - r \leq y_t \leq Cx_t + r)\, \chi(\underline{m}_t - \rho \leq x_t \leq \overline{m}_t + \rho)\, \mathrm{d}x_t, \tag{16}$$

where C is a matrix $(1 \times \ell)$, i.e. a row vector.

As we assume the data y_t be *unknown* yet, the formula (16) must be integrated within the state bounds from *time update* step (10), i.e. from $\underline{m}_t - \rho$ to $\overline{m}_t + \rho$. These bounds are implicitly guaranteed by the second term in (16), therefore they are not explicitly specified hereafter.

Considering $x_t = [x_{t;1}, \ldots, x_{t;\ell}]'$, the formula (16) is

$$f(y_t|d(t-1)) \propto \int \chi\left(y_t - r \leq \sum_{i=1}^{\ell} C_i x_{t;i} \leq y_t + r \right) \times \\ \times \prod_{j=1}^{\ell} \chi\left(\underline{m}_{t;j} - \rho_j \leq x_{t;j} \leq \overline{m}_{t;j} + \rho_j \right) \mathrm{d}x_{t;1} \ldots \mathrm{d}x_{t;\ell}. \quad (17)$$

Assuming $C_1 > 0$, the first term in (17) can be expressed as

$$\chi\left(y_t - r \leq \sum_{i=1}^{\ell} C_i x_{t;i} \leq y_t + r \right) = \\ = \chi\left(\frac{y_t - \sum_{i=2}^{\ell} C_i x_{t;i} - r}{C_1} \leq x_{t;1} \leq \frac{y_t - \sum_{i=2}^{\ell} C_i x_{t;i} + r}{C_1} \right) \equiv R_1(y_t|x_t).$$

Then,

$$f(y_t|d(t-1)) \propto \int \underbrace{\left[\int R_1(y_t|x_t)\, \chi\left(\underline{m}_{t;1} - \rho_1 \leq x_{t;1} \leq \overline{m}_{t;1} + \rho_1 \right) \mathrm{d}x_{t;1} \right]}_{I_1(y_t)} \times \\ \times \prod_{j=2}^{\ell} \chi\left(\underline{m}_{t;j} - \rho_j \leq x_{t;j} \leq \overline{m}_{t;j} + \rho_j \right) \mathrm{d}x_{t;2} \ldots \mathrm{d}x_{t;\ell}. \quad (18)$$

Computing I_1, we get

$$I_1(y_t) = \max\left(J_1(y_t); 0 \right), \quad (19)$$

where

$$J_1(y_t) = \min\left(\overline{m}_{t;1} + \rho_1; \frac{y_t - \sum_{i=2}^{\ell} C_i x_{t;i} + r}{C_1} \right) - \max\left(\underline{m}_{t;1} - \rho_1; \frac{y_t - \sum_{i=2}^{\ell} C_i x_{t;i} - r}{C_1} \right), \quad (20)$$

which is proportional to a symmetric trapezoidal pdf of y_t. We will approximate it by a uniform pdf on its support. This is achieved by minimisation of Kullback-Leibler divergence of these pdfs [9]. The purpose is (i) to preserve the class of uniform pdfs, (ii) to keep the computation recursive and tractable.

$$I_1(y_t) \approx k_1 \chi\left(y_t - r^{\oplus(1)} \leq \sum_{i=2}^{\ell} C_i x_{t;i} \leq y_t + r^{\ominus(1)} \right), \quad (21)$$

where $k_1 = \left(r^{\oplus(1)} + r^{\ominus(1)}\right)^{-1}$. It can be shown that, according to the sign of C_1,

$$\begin{array}{lll} r^{\ominus(1)} = r - C_1(\underline{m}_{t;1} - \rho_1), & r^{\oplus(1)} = r + C_1(\overline{m}_{t;1} + \rho_1), & \text{if } C_1 > 0, \\ r^{\ominus(1)} = r - C_1(\overline{m}_{t;1} + \rho_1), & r^{\oplus(1)} = r + C_1(\underline{m}_{t;1} - \rho_1), & \text{if } C_1 < 0 \end{array} \tag{22}$$

(i.e. the terms in the parentheses are swapped). The formulae (22) were inferred for $C_1 \neq 0$ but they hold for $C_1 = 0$ as well.

Substituting (21) into (18), we get

$$f(y_t|d(t-1)) \approx K_1 \int \chi\left(y_t - r^{\oplus(1)} \leq \sum_{i=2}^{\ell} C_i x_{t;i} \leq y_t + r^{\ominus(1)}\right) \times \\ \times \prod_{j=2}^{\ell} \chi\left(\underline{m}_{t;j} - \rho_j \leq x_{t;j} \leq \overline{m}_{t;j} + \rho_j\right) \mathrm{d}x_{t;2} \ldots \mathrm{d}x_{t;\ell} \tag{23}$$

having the same form as (17) with a normalising constant K_1.

Applying the procedure (17)–(23) for integration over $x_{t;2}, \ldots, x_{t;\ell}$, we get the approximate uniform predictive pdf

$$f(y_t|d(t-1)) \approx K\chi\left(-r^{\ominus(\ell)} \leq y_t \leq r^{\oplus(\ell)}\right), \tag{24}$$

where $K = \left(r^{\oplus(\ell)} + r^{\ominus(\ell)}\right)^{-1}$. If we define vectors $s^{\ominus}$ and $s^{\oplus}$ so that

$$\begin{array}{lll} s_i^{\ominus} = \underline{m}_{t;i} - \rho_i, & s_i^{\oplus} = \overline{m}_{t;i} + \rho_i, & \text{if } C_i \geq 0, \\ s_i^{\ominus} = \overline{m}_{t;i} + \rho_i, & s_i^{\oplus} = \underline{m}_{t;i} - \rho_i, & \text{if } C_i < 0, \end{array} \tag{25}$$

then

$$r^{\ominus(\ell)} = r - \sum_{i=1}^{\ell} C_i s_i^{\ominus} \;=\; r - Cs^{\ominus} \;\equiv\; -\underline{y}_t, \tag{26}$$
$$r^{\oplus(\ell)} = r + \sum_{i=1}^{\ell} C_i s_i^{\oplus} \;=\; r + Cs^{\oplus} \;\equiv\; \overline{y}_t$$

and we get prediction bounds, i.e. $\chi\left(\underline{y}_t \leq y_t \leq \overline{y}_t\right)$.

This predictive pdf is conditioned by $\underline{m}_t$ and $\overline{m}_t$ considered as statistics, provided the parameters A and B be known.

4.2 Point prediction

Mean value of (24) is

$$\hat{y}_t \equiv \mathrm{E}\left[y_t|d(t-1)\right] = \frac{r^{\oplus(\ell)} - r^{\ominus(\ell)}}{2} = C \underbrace{\left(\frac{\overline{m}_t + \underline{m}_t}{2}\right)}_{\mathrm{E}[x_t|d(t-1)]}. \tag{27}$$

Remarks: (i) This formula is identical with the point one-step-ahead predictor for Kalman filter [1] (note that $(\overline{m}_t + \underline{m}_t)/2 = A\hat{x}_{t-1} + Bu_{t-1}$, see (9) and (15)). However, their results differ because of different time-updated estimates $\hat{x}_{t-1}$ supplied by each estimator. (ii) The point prediction is unaffected by the aproximation (21), because the approximated trapezoidal pdf is symmetric and its approximation by the uniform pdf preserves the mean value.

5 Algorithmic summary

Here, the proposed algorithm of observation prediction for model (7) is summarised. It is assumed, that model matrices A, B, C are known as well as the noise bounds r, ρ.

Initialisation:

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

On-line

(i) Set $t = t + 1$
(ii) Compute $\underline{m}_t$, $\overline{m}_t$ according to (9)
(iii) Compute predictor according to (24)
(iv) Get the point output prediction according to (27)
(v) Obtain new data u_t, y_t
(vi) Add successively single data strips according to (12) and approximate the obtained support by a parallelotope (for details see Appendix A.2 in [9]) to obtain the resulting form (13)
(vii) Compute $\underline{x}_t$, $\overline{x}_t$ (14)
(viii) Compute the point estimate $\hat{x}_t$ (15)
(ix) If $t < \overline{t}$, go to (i)

6 Experiments

In this section, the simulative experiments demonstrate the proposed algorithm properties. The algorithm is also compared with the Kalman filter (KF).

6.1 Experiment setup

The matrices of the state space model (6) are set as

$$A = \begin{bmatrix} 0.4 & -0.3 & 0.1 \\ -0.4 & 0.4 & 0.0 \\ 0.3 & 0.2 & 0.1 \end{bmatrix}, \quad B = \begin{bmatrix} 0.1 \\ 0.6 \\ 0.3 \end{bmatrix},$$
$$C = \begin{bmatrix} -1.0 & 0.9 & -0.5 \end{bmatrix}. \tag{28}$$

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

6.2 Results

We compare performance of the LSU predictor (24) with the KF predictor [20]. We examine a modellling mismatch of KF caused by a uniform noise pdf.

Prediction error and predictive interval We evaluated medians of output prediction errors $(\hat{y}_t - y_t)$ and half-widths of prediction intervals. For LSU predictor, half-width is defined as $(\overline{y}_t - \hat{y}_t) \equiv (\overline{y}_t - \underline{y}_t)/2$. For KF predictor, standard deviation σ_t of the predictive pdf was used instead, as a measure of variance. The prediction interval is then $\hat{y}_t \pm \text{half\_width}_t$, specific for each predictor.

Half-width of the predictive interval for the LSU predictor (24) is

$$\frac{\overline{y}_t - \underline{y}_t}{2} = |C|\,|A|\,\frac{\overline{x}_{t-1} - \underline{x}_{t-1}}{2} + |C|\rho + r, \tag{29}$$

absolute value of a matrix applies to its elements.

Standard deviation of predictive pdf of the KF predictor [20] is

$$\sigma_t = \sqrt{CA\ \mathrm{cov}(\hat{x}_{t-1})\, A'C' + CR_wC' + R_v}, \tag{30}$$

where R_w is a covariance of state noise and R_v is a covariance of output noise. Both the formulae have a similar structure, however, they combine different terms in a different way. The covariance matrices in KF (30) were chosen as $R_w = c\,\mathrm{diag}(\rho^2)$ and $R_v = c\,r^2$. The value of c plays a role of a "matching" parameter and will be discussed later. The noise bounds and covariances are fixed and known.

We examined the influence of noise bounds ρ and r on the predictive statistics mentioned above, $c = 2.7$. The results and comparison are in Table 1.

Table 1. Influence of noise parameters ρ and r on medians of prediction errors $(\hat{y}_t - y_t)$ and half-widths of prediction intervals $(\overline{y}_t - \hat{y}_t)$ for LSU and KF.

		prediction error		half-width	
ρ	r	LSU	KF	LSU	KF
0.001	0.001	−0.0001	−0.0001	0.0128	0.0031
0.001	0.01	−0.0005	−0.0001	0.0124	0.0168
0.001	0.1	0.0016	0.0038	0.1024	0.1644
0.01	0.001	−0.0009	−0.0008	0.0616	0.0242
0.01	0.01	−0.0015	−0.0013	0.1282	0.0309
0.01	0.1	−0.0046	−0.0008	0.1240	0.1679
0.1	0.001	−0.0088	−0.0037	0.4109	0.2411
0.1	0.01	−0.0090	−0.0078	0.6913	0.2419
0.1	0.1	−0.0147	−0.0128	1.2815	0.3086
2	2	−0.2949	−0.2561	25.3609	6.1710

Predictive interval and KF covariance matrices Further, we investigated the influence of the parameter c in R_w and R_v on KF predictive interval. Data were generated with $\rho = 0.1$ and $r = 0.3$ and the prediction by KF was run with various values of c. Then, the same experiment was done with $\rho = r = 2$. The results are in Table 2. The purpose of these experiments was, in context of mismatch between uniform noise and KF's assumption of normality, to examine influence of c on median(σ_t) and on the number of observed output values out of the predictive intervals, $\hat{y}_t \pm \sigma_t$ (denoted '$\notin$' in Table 2) and $\hat{y}_t \pm 2\sigma_t$ (denoted '$\notin_2$'). Setting for the matching the second moments, $n \sim \mathcal{U}_n(-r, r)$, var($n$) $= r^2/3$.

Table 2. Influence of c on median(σ_t) and on observed output values out of the predictive intervals ($\notin$ and $\notin_2$) for KF.

	$\rho = 0.1$, $r = 0.3$			$\rho = 2$, $r = 2$		
c	$\notin$	$\notin_2$	median(σ_t)	$\notin$	$\notin_2$	median(σ_t)
0.33	40	3	0.2040	29	7	2.1683
0.6	17	0	0.2737	19	0	2.9091
1	6	0	0.3534	10	0	3.7556
1.5	2	0	0.4328	5	0	4.5996
2	0	0	0.4998	1	0	5.3112
2.7	0	0	0.5807	0	0	6.1710

This behaviour is also illustrated in Fig. 2.

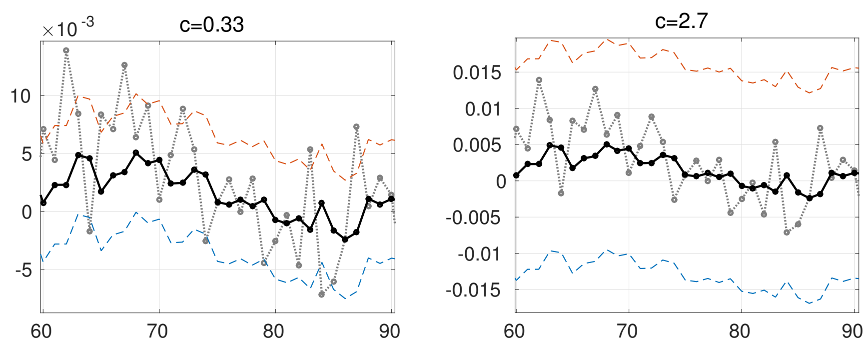

Fig. 2. Simulated (dotted grey) vs. KF-predicted (solid black) output y_t with prediction bounds (thin dashed), $\rho = 0.003$, $r = 0.007$. Left figure: c by moment matching ($c = 0.33$), right figure: empirical value $c = 2.7$.

Predictive interval and LSU preditor The LSU predictive interval contains all the observed output values.

Influence of input Variance s of input generator has no influence on values in Tables 1 and 2, i.e. the same results have been observed for the system without input, $s = 0$. However, presence of input influences the absolute values of output (excitation), as shown in Fig. 3. Note that the width of the prediction interval seems changed due to different scales on the vertical axes.

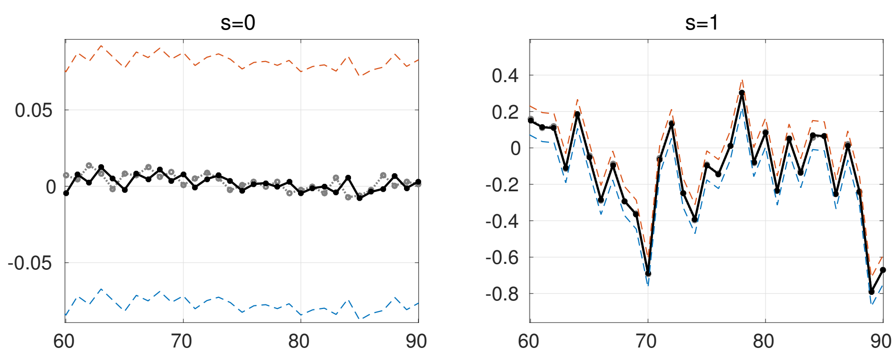

Fig. 3. Simulated (dotted grey) vs. LSU-predicted (solid black) output y_t with prediction bounds (thin dashed), $\rho = 0.003$, $r = 0.007$. Left figure: no input ($s = 0$), right figure: input with $s = 1$.

6.3 Discussion

The LSU predictive performance was examined, using simulated data with uniformly distributed noises, and compared to KF as a standard tool, but assuming normally distributed noises. This modelling difference was involved in the study as well.

The LSU predictor, in comparison to KF predictor, has moderately higher prediction errors, see Table 1 (standard deviations of prediction errors are in a similar relation as medians, not reported).

Setting up the KF noise covariances (30) meets the fact of different parametric models of noise. Therefore, a matching parameter c was introduced to interconnect ρ and r with R_w and R_v. If we matched second moments of normal and uniform pdfs, then $c = 1/3$. However, Table 2 shows that predictive interval $\hat{y}_t \pm \sigma_t$, constructed with such covariances, does not contain significant amount of observed outputs and the predictive interval $\hat{y}_t \pm 2\sigma_t$ does not contain all observed outputs as well. To include all the data in both predictive intervals, the parameter c was set empirically to 2.7 (see also Fig. 2), its square root equals 1.64. It means, that noise parameters for KF are set 1.64-times *higher* (in the sense of standard deviation) than the corresponding parameters used for simulation of the data. Therefore, uncertainty of the KF predictor was artificially increased to broaden the predictive interval, otherwise some data would be missed (Table 2, Fig. 2). This observation indicates that KF, with covariances set by $c \leq 1$, cannot process correctly data that are uniformly distributed.

On the other hand, higher values of c improve matching KF to uniformly distributed data. With c increasing, variance of normal pdf increases, which, on finite support, resembles a “pseudo-uniform” distribution.

The LSU predictive interval (29) reflects the noise model and therefore naturally expresses bounds of actual knowledge on the system, however conservative it may appear, particularly for higher value of noises ρ and r (Table 1). Although the KF predictive interval given by (30) contains all the observed outputs for various combinations of noises, with c chosen for this purpose, it is narrower. The question is, whether the narrower predictive pdf given by (mismodelling) KF does not pretend artificial precision, while the corresponding knowledge is actually not at disposal.

7 Concluding remarks

We extended the LSU estimator of unknown states of a linear state space model with uniformly distributed noise [9] by one-step-ahead LSU predictor for scalar output. The limitation of the output dimension is not essentially critical. Using the chain rule, scalar random variables can be composed together into a vector a vice versa.

Because of properties of uniformly distributed pdf, approximation of pdf was necessary in the process of the predictor construction. However, this approximation does not bias the point prediction.

The LSU predictor was compared with KF (Kalman) predictor, together with predictive intervals computed pro each predictor using the corresponding formulae. It was shown that covariances of KF must be adjusted carefully to give meaningful predictive intervals. This effect pointed out a mismatch between uniform data and assumption of normal distribution at KF.

Predictive intervals given by LSU are more conservative than those by KF. Consequence of this observation should manifest itself in intended application of the LSU predictor for information sharing in sensor networks.

The literature reveals many contemporary examples of sensor networks with uniformly distributed innovations and/or observation processes [21, 22]. The LSU filtering algorithm published in [9], and the consistent predictor developed in this paper, will find important applications in such contexts. In particular, [23] has proposed that optimal probabilistic (i.e. Bayesian) knowledge transfer between interacting nodes can be accomplished optimally via the data predictor. However, results are available to date only for Kalman (normally modelled) nodes [20]. The progress in this paper will allow the extension of the Bayesian transfer learning technique in [23] to uniformly modelled processing nodes, significantly enriching the application range in networked knowledge processing. This research will be published shortly.

Bibliography

[1] A. M. Jazwinski. *Stochastic Processes and Filtering Theory.* Academic Press, New York, 1970.

[2] R. Fletcher. *Practical Methods of Optimization.* John Wiley & Sons, 2000.

[3] D. Simon and D. L. Simon. Constrained Kalman filtering via density function truncation for turbofan engine health estimation. *International Journal of Systems Science*, 41:159–171, 2010.

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

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

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

[7] L. Pavelková and M. Kárný. 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, 2014.

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

[9] L. Pavelková and L. Jirsa. Approximate recursive Bayesian estimation of state space model with uniform noise. In *Proceedings of the 15th International Conference on Informatics in Control, Automation and Robotics (ICINCO)*, pages 388–394, Porto, Portugal, 2018.

[10] K. Belda and L. Pavelková. Online Tuned Model Predictive Control for Robotic Systems with Bounded Noise. In *Proceedings of the 22nd IEEE International Conference on Methods and Models in Automation and Robotics (MMAR)*, pages 694–699, Miedzyzdroje, Poland, 2017.

[11] I. Mizumoto, Y. Fujimoto, and M. Ikejiri. Adaptive output predictor based adaptive predictive control with ASPR constraint. *Automatica*, 57:152–163, 2015.

[12] M. C. Campi, G. Calafiore, and S. Garatti. Interval predictor models: Identification and reliability. *Automatica*, 45(2):382–392, 2009.

[13] J. M. Bravo, T. Alamo, M. Vasallo, and M. E. Gegundez. A General Framework for Predictors Based on Bounding Techniques and Local Approximation. *IEEE Transaction on Automatic Control*, 62(7):3430–3435, JUL 2017.

[14] L. G. Crespo, S. P. Kenny, and D. P. Giesy. Staircase predictor models for reliability and risk analysis. *Structural Safety*, 75:35–44, 2018.

[15] J. Lawrence. Polytope volume computation. *Mathematics of Computation*, 57(195):259–271, 1991.

[16] J. M. Bernardo and A. F. M. Smith. *Bayesian Theory.* John Wiley & Sons, 1997.

[17] S. Kotz and J. R. Dorp. *Beyond beta: Other continuous families of distributions with bounded support and applications*. World Scientific, 2004.

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

[19] 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.

[20] C. Foley and A. Quinn. Fully probabilistic design for knowledge transfer in a pair of Kalman filters. *IEEE Signal Processing Letters*, 25:487–490, 2018.

[21] A. Goudjil, M. Pouliquen, E. Pigeon, O. Gehan, and B. Targui. Recursive output error identification algorithm for switched linear systems with bounded noise. *IFAC-PapersOnLine*, 50(1):14112–14117, 2017.

[22] J. He, X. Duan, P. Cheng, L. Shi, and L. Cai. Accurate clock synchronization in wireless sensor networks with bounded noise. *Automatica*, 81:350–358, 2017.

[23] A. Quinn, M. Kárný, and T. Guy. Fully probabilistic design of hierarchical bayesian models. *Information Sciences*, 369(1):532–547, 2016.