INTERNATIONAL JOURNAL OF ADAPTIVE CONTROL AND SIGNAL PROCESSING
Int. J. Adapt. Control Signal Process. (2013)
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/acs.2438

State and parameter estimation of state-space model with entry-wise correlated uniform noise

Lenka Pavelková*,† and Miroslav Kárný

Institute of Information Theory and Automation, Pod vodárenskou věží 4, Prague, Czech Republic

SUMMARY

Joint parameter and state estimation is proposed for linear state-space model with uniform, entry-wise correlated, state and output noises (*LSU model* for short). The adopted Bayesian modelling and approximate estimation produce an estimator that (a) provides the maximum a posteriori estimate enriched by information on its precision, (b) respects correlated noise entries without demanding the user to tune noise covariances, and (c) respects bounded nature of real-life variables. Copyright © 2013 John Wiley & Sons, Ltd.

Received 11 January 2013; Revised 19 September 2013; Accepted 25 September 2013

KEY WORDS: state-space models; uncertain dynamic systems; bounded noise; filtering problems; estimation algorithms

1. INTRODUCTION

Online prediction, fault detection and quality check, advanced signal processing, feedback control, and others need an adequate model of the considered system. Description of the expected behaviour of the modelled system can often be constructed from first principles when including (often directly unobserved) states. The non-negligible stochastic part causing departures from the expectation is also to be modelled as its presence calls for feedback [1]. In generic case, grey-box state-space models [2] are obtained, which have the white-box, black-box, and input–output models as their special cases. Their exploitation needs estimators of both the model state and parameters. The extent of the research dealing with their various aspects (concepts, formal solutions, algorithms, approximations, analysis, and applications, concerning both general and specialised cases) is enormous (cf. references in [3]). The current paper focuses on online estimators, which (a) provide parameter and state estimates, including their precisions, (b) respect strong correlations of stochastic disturbances, especially of the state noise, and (c) cope with strictly bounded variables with incompletely known bounds.

These demands have arisen from practical needs of traffic control [4–6] where the reliable control and/or prediction requires information about the estimation quality. There, bounded variables are obviously involved such as a queue length at signalised intersection [4], a car position on the road [5], or a maximal permitted drive speed [6]. Strongly correlated errors occur whenever the adopted model is obtained by the linearisation [5, 6]. To our best knowledge, no ready solutions meet the demands (a)–(c) as the further discussion indicates.

Kalman filtering (KF) and its extensions (e.g. [7]) meet the demand (a) under the assumption that the employed linear state-space model has normally distributed state and observation noises.

*Correspondence to: Lenka Pavelková, Institute of Information Theory and Automation, Pod vodárenskou věží 4, Prague, Czech Republic.

†E-mail: pavelkov@utia.cas.cz

KF, however, faces significant, extensively counteracted, troubles with respect to both (b) and (c) as seen in the following representative samples of existing solutions.

As for requirements (a) and (b), KF and its extensions work well when the noise covariances are well chosen. Then, they also provide adequate information on the estimate precision. The covariances are predominantly taken as design parameters of KF as their estimation represents a highly nonlinear problem [8, 9]. The number of covariance entries grow quadratically with the state and output dimensions, which soon makes their experimental tuning infeasible. The number of opted entries can be decreased by a nontrivial parsimonious parametrisation [10] or made linearly growing in a factorised version of the state-space model [11, 12], but the sensitivity to covariances choice persists, and neither (a) nor (b) are met.

The requirement (c) can be fulfilled by a projection of the estimates onto the constraint surface via quadratic programming [13]. Another way is proposed by the probability density function (pdf) truncation approach [14] where the pdf of the state estimate is computed by the standard KF and then truncated at the constraint edges. The constrained state estimate is equal to the mean of the truncated pdf. The induced high-computational demands limit the degree of meeting (c). More importantly, techniques employing a projection or truncation in conjunction with the system model having unbounded support and light tails *respect the constraints during estimation but not during modelling*. Consequently, the posterior pdf forming the outcome of the Bayesian treatment is a worse estimator than necessary. Indeed, this pdf is a product of prior pdf on parameters and initial state, and likelihood function as a product of pdfs describing the state-space model with observed data inserted. The system model with the constrained states has a complex state and parameter-dependent normalising factor entering the likelihood. This factor is neglected by the discussed techniques.

The complexity of the support and form of the exact posterior pdf corresponding to the adequate modelling of boundedness limits feasibility of the Bayesian estimation. Nevertheless, it is desirable to address it as there is a broad range of problems in which explicit inclusion of constraints into the system model significantly improves the estimation quality.

The correct modelling of state boundedness and elaboration of the corresponding joint estimator providing both point estimates and information on their precisions are the main methodological contributions of the current paper. The proposed estimator considers uniform noises, but it can surely be extended to other models with bounded noises.

Sequential Monte Carlo sampling alias particle filtering [15] is a serious competitor to our treatment. Particle filtering avoids linearisation and provides a range of algorithms coping with (b) and (c). In [16], a constrained particle-filtering algorithm based on acceptance/rejection and optimisation strategies is proposed. Simulations show the efficacy of the proposed method in constraints handling and its robustness against poor prior information. Good results are obtained if computationally intensive simulations can be performed. Our semi-analytical treatment is able to decrease this load, and so it suits the situations where such simulations cannot be used.

The methods assuming unknown-but-bounded errors of state-space equations [17, 18] represent a significant group coping well with the unknown bounded states. The approach has been equipped with recursively updated ellipsoid [19] or boxes [20] approximating complex constraint sets arising in it. In a related work [21], interval analysis and set inversion are employed for estimation characterising the accuracy of the parameter estimates.

These methods give up a stochastic interpretation of the disturbances. They lose the rich Bayesian methodology [22]. Practically, they give up the whole arsenal of ready tools ranging from knowledge elicitation [23], to structure estimation [24], to aim elicitation [25]. This loss is pointless, as a careful comparison of stochastic estimation theory and unknown-but-bounded approach [26] reveals that the latter provides results having exact stochastic equivalents.

A general stochastic filtering applied to models with uniformly distributed noise is in [4]. It leads to linear-programming-type algorithms coping efficiently with constraints and *relaxing the need to supply external information on ranges of the noises in the model*. The resulting filter copes well with missing data [5]. However, it provides a point estimate only, namely a maximum a posteriori probability (MAP) estimate. Moreover, it neglects correlations between noise entries. The

DOI: 10.1002/acs

current paper removes these drawbacks and obtains the estimator, which meets the requirements (a)–(c).

After formalising the addressed problem (Section 2) and summarising the used theory (Section 3), the complete LSU model specifying the posterior pdf on unknown variables is described in Section 4. Then, online MAP estimation is presented in Section 5.1. It is performed on a sliding window [27] and provides multiple point estimates of the same unknown variable as a by-product. This enables us to obtain the estimate precision as presented in Section 5.2. Behaviour of the overall algorithm is illustrated and discussed in Section 6. Section 7 summarises algorithmic and methodological achievements and lists open problems.

Throughout, $\equiv$ means definition by equality. Boldface fonts are reserved for column-oriented vectors and matrices. Calligraphic fonts are used for their compound versions. The transposition is marked $'$. $\mathbf{z}^\star$ denotes a set of $\mathbf{z}$-values. $\mathbf{z}_t$ is the value of $\mathbf{z}$ at discrete-time instant $t \in t^\star \equiv \{1, 2, \dots, T\}$, $T < \infty$. $z_{t;i}$ is the i-th entry of $\mathbf{z}_t$. $\ell_{\mathbf{z}}$ is the length of the vector $\mathbf{z}$. The ordered sequence $\mathbf{z}^{k:l} \equiv (\mathbf{z}_k, \mathbf{z}_{k-1}, \dots, \mathbf{z}_l)$, $0 \leqslant l \leqslant k$ (it is void for $k < l$). $\underline{\mathbf{z}}$, $\overline{\mathbf{z}}$ are lower and upper bounds on $\mathbf{z}$, respectively. The inequalities $\overline{\mathbf{z}} \geqslant \mathbf{z} \geqslant \underline{\mathbf{z}}$ are meant entry-wise. $\chi(\mathbf{z}; \mathbf{z}^\star)$ is the set indicator, which equals 1 if $\mathbf{z} \in \mathbf{z}^\star$ and 0 otherwise. $\hat{\mathbf{z}}_t$ is a point estimate of $\mathbf{z}_t$. $\mathbf{z}_{(\alpha)}$ is the vector $\mathbf{z}$ of the length α. $\mathbf{M}_{(\alpha,\beta)}$ denotes the matrix $\mathbf{M}$ with α rows and β columns. $\mathbf{I}_{(\alpha)}$ is the square identity matrix of the order α. M_{ij} is the entry of the matrix $\mathbf{M}$ in the i-th row and j-th column. The symbol f denotes pdf distinguished by the argument names. No formal distinction is made among a random variable, its realisation, and a pdf argument.

2. ADDRESSED ESTIMATION PROBLEM

The considered system is modelled by the linear discrete-time state (1) and observation (2) equations

$$\mathbf{E}\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{B}\mathbf{u}_t + \mathbf{w}_t \tag{1}$$

$$\mathbf{F}\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{e}_t \tag{2}$$

$$t \in \{\tau - \min(\Delta, \tau - 1), \dots, \tau\} \tag{3}$$

$$\tau \in \tau^\star \equiv \{1, \dots, T\},\ 1 \leqslant \Delta \leqslant T - 1.$$

The structured time set (3) prepares the estimation on a moving window of the length $\Delta + 1$. The vector $\mathbf{y}_t$ is an $\ell_{\mathbf{y}}$-dimensional output, and $\mathbf{u}_t$ is an optional $\ell_{\mathbf{u}}$-dimensional input. Data records $\mathbf{d}_t = (\mathbf{y}_t, \mathbf{u}_t)$ contain the observed variables and $\ell_{\mathbf{x}}$-dimensional state $\mathbf{x}_t$ unobserved ones, $t \in t^\star$. Matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{E}$, and $\mathbf{F}$ have dimensions making (1) and (2) meaningful. $\mathbf{w}_t$ and $\mathbf{e}_t$ are vectors of the state and output noises, of sizes $\ell_{\mathbf{x}}$ and $\ell_{\mathbf{y}}$, respectively. They are assumed to be zero mean identically distributed, conditionally independent of past, having uncorrelated entries and constant variances. $\mathbf{E}$ and $\mathbf{F}$ are upper triangular matrices with unit diagonal: their presence models correlation of noise entries.

If the noises $\mathbf{w}_t$, $\mathbf{e}_t$ are Gaussian and parameters, including noise covariances, known, the Bayesian state estimation reduces to the KF [29]. The paper departs from this standard and assumes that the state $\mathbf{w}_t$ and the observation $\mathbf{e}_t$ noises are uniformly distributed on a multivariate box with the centre at zero ($\mathbf{0}$) and with unknown half-widths $\mathbf{q}$ and $\mathbf{r}$ of the support intervals, respectively. Denoting the uniform pdf $\mathsf{U}_{\text{modelled variable}}(\text{mean}, \text{half-width})$, the adopted assumptions imply

$$\begin{aligned} f\left(\mathbf{w}_t | \mathbf{d}^{t-1:1}, \mathbf{u}_t, \mathbf{x}^{t-1:0}, \boldsymbol{\Theta}\right) &= f\left(\mathbf{w}_t | \mathbf{q}\right) = \mathsf{U}_{\mathbf{w}_t}\left(\mathbf{0}, \mathbf{q}\right) \\ f\left(\mathbf{e}_t | \mathbf{d}^{t-1:1}, \mathbf{u}_t, \mathbf{x}^{t-1:0}, \boldsymbol{\Theta}\right) &= f\left(\mathbf{e}_t | \mathbf{r}\right) = \mathsf{U}_{\mathbf{e}_t}\left(\mathbf{0}, \mathbf{r}\right), \end{aligned} \tag{4}$$

where $\boldsymbol{\Theta}$ denotes the collection of time-invariant unknown parameters determining the parametric LSU model (1), (2), and (4). Section 4.1 provides the remaining details of the considered parametrisation.

Addressed problem: *A recursive, online implementable, estimator of the parameters and states of the LSU model is searched for. Point estimates and information on their precisions are needed.*

DOI: 10.1002/acs

3. PRELIMINARIES

In the considered Bayesian set-up [11, 22, 28], the system is modelled by pdfs. The chain rule factorises the joint pdf of all involved variables as follows:

$$f\left(\mathbf{d}^{T:1}, \mathbf{x}^{T:0}, \mathbf{\Theta}\right) = \underbrace{f\left(\mathbf{x}_0, \mathbf{\Theta}\right)}_{\text{prior pdf}} \prod_{t \in t^\star} \underbrace{f\left(\mathbf{u}_t \left|\mathbf{d}^{t-1:1}\right.\right)}_{\text{controller}} \times \prod_{t \in t^\star} \underbrace{f\left(\mathbf{y}_t | \mathbf{x}_t, \mathbf{u}_t, \mathbf{\Theta}\right)}_{\text{observation model}} \times \underbrace{f\left(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{u}_t, \mathbf{\Theta}\right)}_{\text{state evolution model}}. \tag{5}$$

This form assumes that $\mathbf{\Theta}$ and $\mathbf{x}^{T:0}$ are unknown to the controller, that is, it accepts the natural conditions of control [29]. The sequence $\mathbf{d}^{T:1}$ of data records $\mathbf{d}_t = (\mathbf{y}_t, \mathbf{u}_t)$, $t \in t^\star$, is sequentially observed. Standard Bayesian estimation on a window of a fixed length $\Delta + 1$ works with the data $\mathbf{d}^{\tau:\tau-\Delta}$ and internals $\mathbf{X}_\tau$, collecting both unobserved states and unknown parameters,

$$\mathbf{X}_\tau \equiv [\mathbf{x}'_\tau, \ldots, \mathbf{x}'_{\tau-\Delta}, \mathbf{x}'_{\tau-\Delta-1}, \mathbf{\Theta}']', \tag{6}$$

where τ is defined in (3). The estimation reduces to evaluation of characteristics of the *posterior pdf* $f(\mathbf{X}_\tau|\mathbf{d}^{\tau:\tau-\Delta})$ of $\mathbf{X}_\tau$ conditioned on the observed data $\mathbf{d}^{\tau:\tau-\Delta}$. The Bayes rule provides the formal solution

$$f\left(\mathbf{X}_\tau\middle|\,\mathbf{d}^{\tau:\tau-\Delta}\right) = \frac{\mathcal{L}\left(\mathbf{X}_\tau, \mathbf{d}^{\tau:\tau-\Delta}\right) f\left(\mathbf{x}_{\tau-\Delta-1}, \mathbf{\Theta}\right)}{\int_{\mathbf{X}_\tau^\star} \mathcal{L}\left(\mathbf{X}_\tau, \mathbf{d}^{\tau:\tau-\Delta}\right) f\left(\mathbf{x}_{\tau-\Delta-1}, \mathbf{\Theta}\right) \mathrm{d}\mathbf{X}_\tau}$$
$$\mathcal{L}\left(\mathbf{X}_\tau, \mathbf{d}^{\tau:\tau-\Delta}\right) \equiv \prod_{t=\tau-\Delta}^{\tau} f\left(\mathbf{y}_t | \mathbf{x}_t, \mathbf{u}_t, \mathbf{\Theta}\right) f\left(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{u}_t, \mathbf{\Theta}\right) \chi(\mathbf{X}_\tau; \mathbf{X}_\tau^\star). \tag{7}$$

$\mathcal{L}(\cdot,\cdot)$ is the likelihood function—a function of internals with data fixed at observed values. The set $\mathbf{X}_\tau^\star$ delimiting possible values of internals is specified in (18).

The posterior pdf can be evaluated without knowing the controller, which is canceled in (7). A design of the Bayesian estimator requires a specification of the prior pdf, observation, and state evolution models. The LSU model discussed in Section 4 provides a nontrivial specification example.

Note that truncation of Gaussian distribution, an exact respecting of bounds on states in KF framework, would lead to the same set $\mathbf{X}_\tau^\star$. However, the likelihood would be much more complex because of the dependence of the normalisation factor of the system-describing pdf on internals.

Evaluation of the normalisation integral in (7) as well as of moments of the posterior pdf (7) is mostly too complex for online estimation and filtering. This makes us consider a MAP estimate $\hat{\mathbf{X}}_\tau$ of internals $\mathbf{X}_\tau$, [28],

$$\hat{\mathbf{X}}_\tau = \arg\max_{\mathbf{X}_\tau^\star} \mathcal{L}\left(\mathbf{X}_\tau, d^{\tau:\tau-\Delta}\right) f\left(\mathbf{x}_{\tau-\Delta-1}, \mathbf{\Theta}\right), \tag{8}$$

which does not need them. Section 5.2 adds information on the precision of this point estimate.

The observation and state evolution models are decisive modelling elements. It is advantageous to apply the chain rule also to their traditional vector-describing form (5)

$$f\left(\mathbf{y}_t | \mathbf{x}_t, \mathbf{u}_t, \mathbf{\Theta}\right) = \prod_{i=1}^{\ell_y} f\left(y_{t;i} | \mathbf{x}_t, \boldsymbol{\psi}_t, \mathbf{\Theta}\right), \tag{9}$$

where $\boldsymbol{\psi}_t$ is the regression vector of a finite length $\ell_{\boldsymbol{\psi}}$, $\boldsymbol{\psi}_t \equiv [y_{t;i+1}, \ldots, y_{t;\ell_y}, \mathbf{u}'_t]'$. The pdf $f\left(y_{t;i} | \mathbf{x}_t, \boldsymbol{\psi}_t, \mathbf{\Theta}\right)$ is called factor of the observation model, *observation factor* for short. An observation factor models a scalar entry of the output without neglecting correlations with others. Its functional form, regression vector, and parameters generally differ for different output entries. This

DOI: 10.1002/acs

flexibility and the scalar-wise modelling of the uncertainty are the main gains of the factorised form (9). Analogically, the state evolution model factorises

$$f\left(\mathbf{x}_t|\mathbf{x}_{t-1},\mathbf{u}_t,\mathbf{\Theta}\right)=\prod_{i=1}^{\ell_{\mathbf{x}}} f\left(x_{t;i}|\mathbf{x}_{t-1},\boldsymbol{\zeta}_t,\mathbf{\Theta}\right), \tag{10}$$

where the finite-dimensional $\boldsymbol{\zeta}_t \equiv [x_{t;i+1},\ldots,x_{t;\ell_{\mathbf{x}}},\mathbf{u}_t']'$. The consideration of factors of the state evolution model, *state factors* for short $f\left(x_{t;i}|\mathbf{x}_t,\boldsymbol{\zeta}_t,\mathbf{\Theta}\right)$, also needs scalar modelling of uncertainty and enhances modelling flexibility with respect to the functional form and parametrisation.

4. DETAILED DESCRIPTION OF THE LSU MODEL

This section thoroughly describes the LSU model outlined in Section 2.

4.1. Parametrisation of the LSU Model

The model matrices **A**, **B**, **C**, **E**, and **F** in the state-space model (1)–(4) are expressed as sums

$$\mathbf{E}={}^{k}\mathbf{E}+{}^{u}\mathbf{E},$$

where ${}^{k}\mathbf{E}$ contains known entries of **E** and zeros, and ${}^{u}\mathbf{E}$ includes unknown entries of **E** and zeros. A similar decomposition applies to all model matrices. The unknown entries are collected into the column vector $\boldsymbol{\theta}$, forming a 'coefficient part' of the unknown parameters $\mathbf{\Theta}$ (12), as follows

$$\boldsymbol{\theta}\equiv\left[\mathrm{col}({}^{u}\mathbf{A})',\mathrm{col}({}^{u}\mathbf{B})',\mathrm{col}({}^{u}\mathbf{C})',\mathrm{col}({}^{u}\mathbf{E})',\mathrm{col}({}^{u}\mathbf{F})\right]'. \tag{11}$$

There, the mapping col(**Z**) transforms the nonzero entries of the matrix **Z** into a column vector.

The full collection of the estimated parameters is

$$\mathbf{\Theta}\equiv\left[\boldsymbol{\theta}',\mathbf{q}',\mathbf{r}'\right]', \tag{12}$$

where $\boldsymbol{\theta}$ is given by (11) and the half-widths **q**, **r** are defined in (4).

4.2. Prior information and constrained support of the state evolution model

Physically motivated prior information usually has the assumed form of possible ranges of internals. For t defined in (3), it reads

$$\underline{\mathbf{x}}_t\leqslant\mathbf{x}_t\leqslant\overline{\mathbf{x}}_t,\ \underline{\boldsymbol{\theta}}\leqslant\boldsymbol{\theta}\leqslant\overline{\boldsymbol{\theta}},\ \ \mathbf{0}<\mathbf{q}\leqslant\overline{\mathbf{q}},\ \mathbf{0}<\mathbf{r}\leqslant\overline{\mathbf{r}}\Rightarrow\begin{bmatrix}\underline{\boldsymbol{\theta}}\\ \mathbf{0}\\ \mathbf{0}\end{bmatrix}=\underline{\mathbf{\Theta}}\leqslant\mathbf{\Theta}\leqslant\overline{\mathbf{\Theta}}=\begin{bmatrix}\overline{\boldsymbol{\theta}}\\ \overline{\mathbf{q}}\\ \overline{\mathbf{r}}\end{bmatrix}. \tag{13}$$

The user specifies window length $\Delta+1$ (Sections 6 and 7), and surely met boundaries $\underline{\mathbf{\Theta}}$, $\overline{\mathbf{\Theta}}$, as well as the bounds restricting the model support $\underline{\mathbf{x}}_t\leqslant\mathbf{x}_t\leqslant\overline{\mathbf{x}}_t$ (typically time invariant). The constraint on the initial state $\mathbf{x}_{\tau-\Delta-1}$ is modified by the information propagating over time (Section 5).

4.3. Probabilistic representation of the LSU model

The state (1) and observation (2) equations serve for an explicit specification of the pdfs describing the LSU model. Separation of the unit diagonal $\mathbf{I}_{(\ell_{\mathbf{x}})}$ of the matrix **E**, $\mathbf{E}=\mathbf{I}_{(\ell_{\mathbf{x}})}+\mathbf{\Lambda}$ gives

$$\mathbf{x}_t=-\mathbf{\Lambda}\mathbf{x}_t+\mathbf{A}\mathbf{x}_{t-1}+\mathbf{B}\mathbf{u}_t+\mathbf{w}_t.$$

Thus, the individual entries $x_{t;i}$, $i=1,\ldots,\ell_{\mathbf{x}}$, of the state vector $\mathbf{x}_t$ evolve as follows

$$x_{t;i}=\underbrace{-\sum_{j=i+1}^{\ell_{\mathbf{x}}}\Lambda_{ij}x_{t;j}+\sum_{j=1}^{\ell_{\mathbf{x}}}A_{ij}x_{t-1;j}+\sum_{j=1}^{\ell_{\mathbf{u}}}B_{ij}u_{t;j}}_{\tilde{x}_{t;i}}+w_{t;i}. \tag{14}$$

DOI: 10.1002/acs

A similar decomposition of the matrix $\mathbf{F} = \mathbf{I}_{(\ell_\mathbf{y})} + \mathbf{\Pi}$ gives

$$\mathbf{y}_t = -\mathbf{\Pi}\mathbf{y}_t + \mathbf{C}\mathbf{x}_t + \mathbf{e}_t,$$

and the individual output entries $y_{t;i}$, $i = 1, \ldots, \ell_\mathbf{y}$, evolve as follows

$$y_{t;i} = \underbrace{-\sum_{j=i+1}^{\ell_\mathbf{y}} \Pi_{ij}\, y_{t;j} + \sum_{j=1}^{\ell_\mathbf{x}} C_{ij}\, x_{t;j}}_{\tilde{y}_{t;i}} + e_{t;i}. \tag{15}$$

Recall the nonzero matrices $\mathbf{\Lambda}$ and $\mathbf{\Pi}$ model nontrivial *correlations of noise entries*. Equations (14) and (15) together with the assumptions (4) provide the state factors in (10) corresponding to (14)

$$f\left(x_{t;i} | \mathbf{x}_{t-1}, \boldsymbol{\zeta}_t, \mathbf{\Theta}\right) = \mathsf{U}_{x_{t;i}}(\tilde{x}_{t;i}, q_i). \tag{16}$$

The observation factors in (9) corresponding to (15) are

$$f\left(y_{t;i} | \mathbf{x}_t, \boldsymbol{\psi}_t, \mathbf{\Theta}\right) = \mathsf{U}_{y_{t;i}}(\tilde{y}_{t;i}, r_i).$$

The prior information, enriched by the knowledge of the system model support, is given by (13). States $\mathbf{x}_{\tau-\Delta-1}$ and parameters $\mathbf{\Theta}$ are assumed *a priori* mutually independent. The joint pdf of data $\mathbf{d}^{\tau:\tau-\Delta}$ and internals $\mathbf{X}_\tau$ (6), $\tau \in \tau^\star$, with $\mathbf{\Theta}$ (12), is

$$\begin{aligned} f\left(\mathbf{d}^{\tau:\tau-\Delta}, \mathbf{X}_\tau\right) &= \left(\prod_{i=1}^{\ell_\mathbf{x}} \frac{1}{2\,q_i} \prod_{j=1}^{\ell_\mathbf{y}} \frac{1}{2\,r_j}\right)^{\Delta+1} \prod_{j=1}^{\ell_\mathbf{\Theta}} \frac{1}{\bar{\mathbf{\Theta}}_j - \underline{\mathbf{\Theta}}_j} \\ &\times \prod_{t=\tau-\Delta}^{\tau} \left(f\left(\mathbf{u}_t \middle| \mathbf{d}^{t-1:1}\right) \prod_{i=1}^{\ell_\mathbf{x}} \frac{1}{\bar{x}_{t;i} - \underline{x}_{t;i}} \right) \chi(\mathbf{X}_\tau; \mathbf{X}_\tau^\star). \end{aligned} \tag{17}$$

The indicator $\chi(\mathbf{X}_\tau; \mathbf{X}_\tau^\star)$ restricts the support of this pdf to the set $\mathbf{X}_\tau^\star$ (18) delimited by the observed data $\mathbf{d}^{\tau:\tau-\Delta} = \left(\mathbf{u}^{\tau:\tau-\Delta}, \mathbf{y}^{\tau:\tau-\Delta}\right)$ and by the system model support. The set $\mathbf{X}_\tau^\star$ contains such internals $\mathbf{X}_\tau$ (6) for which the noise terms in (1) and (2) are within the multivariate box defined by (4) and (13), that is, for t, τ specified by (3),

$$\begin{aligned} \mathbf{X}_\tau^\star = \big\{ \mathbf{X}_\tau :\ &(\mathbf{x}_t, \mathbf{\Theta}) \text{ meeting (13) and} \\ &|x_{t;i} - \tilde{x}_{t;i}| \leqslant q_i, |y_{t;j} - \tilde{y}_{t;j}| \leqslant r_j,\ i = 1, \ldots, \ell_\mathbf{x},\ j = 1, \ldots, \ell_\mathbf{y} \big\}. \end{aligned} \tag{18}$$

5. SOLUTION TO THE ESTIMATION PROBLEM

This section forms the technical core of the paper and provides the solution to the addressed estimation problem. Section 5.1 describes the MAP estimate of internals, whereas Appendix A complements technical details. Section 5.2 evaluates precision of the estimate.

5.1. Online point estimation of internals

We focus on point estimation and evaluate a series of MAP estimates $\hat{\mathbf{X}}_\tau$ (8) of the unknown $\mathbf{X}_\tau$ (6), $\tau \in \tau^\star$. With the joint pdf (17), it holds

$$\hat{\mathbf{X}}_\tau = \arg\max_{\mathbf{X}_\tau \in \mathbf{X}_\tau^\star} \left(\prod_{i=1}^{\ell_\mathbf{x}} \frac{1}{2\,q_i} \prod_{i=1}^{\ell_\mathbf{y}} \frac{1}{2\,r_i}\right)^{\Delta+1}. \tag{19}$$

DOI: 10.1002/acs

The support $\mathbf{X}_\tau^\star$ (18) defines the maximisation domain. The maximisation is performed as the minimisation of a negative logarithm of (19) divided by $\Delta + 1$ while omitting a positive constant that does not influence the result. Then, the MAP estimation converts into

$$\hat{\mathbf{X}}_\tau = \arg\min_{\mathbf{X}_\tau \in \mathbf{X}_\tau^\star} \left(\sum_{i=1}^{\ell_\mathbf{x}} \log(q_i) + \sum_{i=1}^{\ell_\mathbf{y}} \log(r_i) \right). \tag{20}$$

A simplified form of the MAP estimation is obtained by linearising the objective function in (20) via the first-order Taylor expansion

$$\hat{\mathbf{X}}_\tau = \arg\min_{\mathbf{X}_\tau \in \mathbf{X}_\tau^\star} \left(\sum_{i=1}^{\ell_\mathbf{x}} q_i + \sum_{i=1}^{\ell_\mathbf{y}} r_i \right). \tag{21}$$

To perform recursive estimation, we need to know the pdf (17) in each time step. In the transition from $f\left(\mathbf{d}^{\tau:\tau-\Delta}, \mathbf{X}_\tau\right)$ to $f\left(\mathbf{d}^{\tau+1:\tau-\Delta+1}, \mathbf{X}_{\tau+1}\right)$, we have to (i) add a new constraint concerning the state at time $t = \tau + 1$ and (ii) replace the discarded constraint concerning the state at time $t = (\tau + 1) - \Delta - 1$ by a relevant part of the prior information (13) concerning the new initial state, that is, $\mathbf{x}_{\tau-\Delta}$. The latter reduces to the replacement of $\underline{\mathbf{x}}_{\tau-\Delta-1} \leqslant \mathbf{x}_{\tau-\Delta-1} \leqslant \overline{\mathbf{x}}_{\tau-\Delta-1}$ by $\underline{\mathbf{x}}_{\tau-\Delta} \leqslant \mathbf{x}_{\tau-\Delta} \leqslant \overline{\mathbf{x}}_{\tau-\Delta}$. The latter interval is chosen as an intersection of the interval estimate of $\mathbf{x}_{\tau-\Delta}$ described in Section 5.2 and a priori physically determined bounds on this state.

In summary, the MAP estimate $\hat{\mathbf{X}}_\tau$ of internals $\mathbf{X}_\tau$ is the minimiser of (20) or (21) on the set $\mathbf{X}_\tau^\star$ (18). In the following, special cases of interest are discussed in detail.

5.1.1. Estimation of state and noise boundaries (void $\boldsymbol{\theta}$). For completely known model matrices, the problem (21) reduces to linear programming (LP) problem

$$\begin{gathered} \text{Find a vector } \mathbf{X}_\tau, \quad \tau \in \tau^\star, \text{ such that} \\ J \equiv \mathcal{C}'\mathbf{X}_\tau = \sum_{i=1}^{\ell_\mathbf{x}} q_j + \sum_{j=1}^{\ell_\mathbf{y}} r_j \rightarrow \min \\ \text{while } \mathcal{A}_\tau \mathbf{X}_\tau \leqslant \mathcal{B}_\tau,\ \underline{\mathbf{X}}_\tau \leqslant \mathbf{X}_\tau \leqslant \overline{\mathbf{X}}_\tau, \end{gathered} \tag{22}$$

$$\mathbf{X}_\tau = [\mathbf{x}'_\tau, \ldots, \mathbf{x}'_{\tau-\Delta}, \mathbf{x}'_{\tau-\Delta-1}, \mathbf{q}', \mathbf{r}']',\ \tau \in \tau^\star. \tag{23}$$

$\mathcal{C}' \equiv [\mathbf{0}'_{(\ell_{\mathbf{X}_\text{fi}}-\ell_\mathbf{x}-\ell_\mathbf{y})}, \mathbf{1}'_{(\ell_\mathbf{x}+\ell_\mathbf{y})}]$, consists of the vectors of zeros and ones of the indicated lengths. $\mathcal{A}_\tau$ and $\mathcal{B}_\tau$ are the known matrix and vector, respectively, reflecting the inequalities (18) describing $\mathbf{X}_\tau^\star$. $\underline{\mathbf{X}}_\tau$, $\overline{\mathbf{X}}_\tau$ are known vectors also implied by (18). A construction of $\mathcal{A}_\tau$, $\mathcal{B}_\tau$, $\underline{\mathbf{X}}_\tau$ and $\overline{\mathbf{X}}_\tau$ is thoroughly described in Appendix A.1.

5.1.2. Estimation of state and all noise characteristics. In general, the MAP estimation (20) becomes a nonlinear programming problem [30] due to the presence of products of internals. The adopted simplest solution, similar to the extended KF [7], linearises these products around current estimates and reduces the MAP estimation to the LP problem (22). In detail, we deal with the particular case of $\boldsymbol{\theta} = [\text{col}({}^u\mathbf{E})', \text{col}({}^u\mathbf{F})']'$ (cf. (11)),

$$\mathbf{X}_\tau = [\mathbf{x}'_\tau, \ldots, \mathbf{x}'_{\tau-\Delta}, \mathbf{x}'_{\tau-\Delta-1}, \text{col}({}^u\mathbf{E})', \text{col}({}^u\mathbf{F})', \mathbf{q}', \mathbf{r}']', \tag{24}$$

with states and the *full noise description* estimated. This special case is sufficient for reproduction of the results of the illustrative example (Section 6). It also guides in a straightforward solution to the general case (11).

DOI: 10.1002/acs

Rearranging the inequalities in (18), so that all entries of $\mathbf{X}_\tau$ are on the left-hand side, gives, for $t = \tau, \tau-1, \ldots, \tau-\Delta$, $i = 1, \ldots, \ell_{\mathbf{x}}$, $k = 1, \ldots, \ell_{\mathbf{y}}$,

$$\begin{aligned} x_{t;i} + \sum_{j=i+1}^{\ell_{\mathbf{x}}} \Lambda_{ij} x_{t;j} - \sum_{j=1}^{\ell_{\mathbf{x}}} A_{ij} x_{t-1;j} - q_i &\leqslant \sum_{j=1}^{\ell_{\mathbf{u}}} B_{ij} u_{t;j} \\ -x_{t;i} - \sum_{j=i+1}^{\ell_{\mathbf{x}}} \Lambda_{ij} x_{t;j} + \sum_{j=1}^{\ell_{\mathbf{x}}} A_{ij} x_{t-1;j} - q_i &\leqslant -\sum_{j=1}^{\ell_{\mathbf{u}}} B_{ij} u_{t;j} \\ -\sum_{j=k+1}^{\ell_{\mathbf{y}}} \Pi_{kj} y_{t;j} + \sum_{j=1}^{\ell_{\mathbf{x}}} C_{kj} x_{t;j} - r_k &\leqslant y_{t;k} \\ +\sum_{j=k+1}^{\ell_{\mathbf{y}}} \Pi_{kj} y_{t;j} - \sum_{j=1}^{\ell_{\mathbf{x}}} C_{kj} x_{t;j} - r_k &\leqslant -y_{t;k}. \end{aligned} \tag{25}$$

The product terms $\Lambda_{ij} x_{t;j}$ in (25), which make the problem nonlinear, are linearised using their latest estimates (indices omitted)

$$\mathbf{\Lambda x} = (\mathbf{\Lambda} - \hat{\mathbf{\Lambda}})(\mathbf{x} - \hat{\mathbf{x}}) + \mathbf{\Lambda}\hat{\mathbf{x}} - \hat{\mathbf{\Lambda}}\hat{\mathbf{x}} + \hat{\mathbf{\Lambda}}\mathbf{x} \approx \mathbf{\Lambda}\hat{\mathbf{x}} - \hat{\mathbf{\Lambda}}\hat{\mathbf{x}} + \hat{\mathbf{\Lambda}}\mathbf{x}, \tag{26}$$

where $\hat{\mathbf{\Lambda}}$, $\hat{\mathbf{x}}$ are the newest available estimates of $\mathbf{\Lambda}$ and $\mathbf{x}$, respectively. The first equality follows from simple algebraic operations; the approximation neglects the quadratic term. The LP problem (22) is obtained if the term $(\mathbf{\Lambda} - \hat{\mathbf{\Lambda}})(\mathbf{x} - \hat{\mathbf{x}})$ can indeed be neglected. The explicit form of the involved matrices and vectors is in Appendix A.2.

5.2. *Precision of state estimates*

The MAP estimation (20) or (21) provides the point estimate $\hat{\mathbf{X}}$ of the unknown $\mathbf{X}$. The individual states are, however, estimated several times during recursive estimation. This offers a possibility to extract information on precision of the state estimates. Here, a Bayesian solution is proposed that is inspired by ensemble filtering [31].

At time t, $(n+1)$ MAP estimates $\hat{\mathbf{x}}_{t-n|t-k}$ of $\mathbf{x}_{t-n}$, based on data up to time $t-k$, $k = 0, 1, \ldots, n$, are available for $2 < n \leqslant \min(t, \Delta)$. Fixing n and considering i-th entry $x_{t-n;i}$ of $\mathbf{x}_{t-n}$, we search for an interval estimate of $x_{t-n;i}$ while its estimates form the 'data' $\hat{x}_{t-n|t-k;i} \in [x_{t-n;i} - \rho_{t-n;i}, x_{t-n;i} + \rho_{t-n;i}]$, $\rho_{t-n;i} > 0$ is unknown. Maximum entropy principle [32] recommends to model the 'data' by the pdf $f(\hat{x}_{t-n|t-n;i}, \ldots, \hat{x}_{t-n|t;i} | x_{t-n;i}, \rho_{t-n;i})$ of the form

$$\begin{aligned} f(\hat{x}_{t-n|t-n;i}, \ldots, \hat{x}_{t-n|t;i} | x_{t-n;i}, \rho_{t-n;i}) &= \prod_{k=0}^{n} \mathsf{U}_{\hat{x}_{t-n|t-k;i}}(x_{t-n;i}, \rho_{t-n;i}) \\ &= \frac{\chi(x_{t-n;i}; -\rho_{t-n;i} \leqslant \underline{s}_{t-n;i} - x_{t-n;i} < \overline{s}_{t-n;i} - x_{t-n;i} \leqslant \rho_{t-n;i})}{(2\rho_{t-n;i})^{n+1}} \\ \underline{s}_{t-n;i} = \min_{k=0,\ldots,n} \hat{x}_{t-n|t-k;i}, \ \overline{s}_{t-n;i} &= \max_{k=0,\ldots,n} \hat{x}_{t-n|t-k;i}. \end{aligned} \tag{27}$$

To complete the probabilistic description of the discussed variables, we need to define the prior pdf $f(x_{t-n;i}, \rho_{t-n;i})$. The ranges of these variables can be specified from the results obtained on the previous window. For presentation simplicity, we do not use this information. Without such information, the maximum entropy principle leads to the improper uniform pdf on the widest possible domains $x_{t-n;i} \in (-\infty, \infty)$, $\rho_{t-n;i} \in (0, \infty)$ with independent $x_{t-n;i}$ and $\rho_{t-n;i}$. With this choice, the likelihood function in (27) becomes proportional to the posterior pdf on $x_{t-n;i}$, $\rho_{t-n;i}$ determined by the sufficient statistics $\underline{s}_{t-n;i}, \overline{s}_{t-n;i}$ equivalent to

$$\mu_{t-n;i} = \frac{\overline{s}_{t-n;i} + \underline{s}_{t-n;i}}{2}, \ \sigma_{t-n;i} = \frac{\overline{s}_{t-n;i} - \underline{s}_{t-n;i}}{2}. \tag{28}$$

DOI: 10.1002/acs

The interval estimate of $x_{t-n;i}$ is delimited by boundaries

$$E[x_{t-n;i} \pm \kappa \rho_{t-n;i} | \underline{s}_{t-n;i}, \overline{s}_{t-n;i}, n], \ \kappa > 0, \tag{29}$$

where $E[\bullet|\underline{s}_{t-n;i}, \overline{s}_{t-n;i}, n]$ is the expectation determined by the discussed posterior pdf. The optional κ controls the probability assigned to the interval estimate.

To simplify reading, we suppress the fixed subscript $_{t-n;i}$ for variables in integrations. First, we need to evaluate the normalisation integral of the posterior pdf. It reads

$$\begin{aligned} J_{n+1} &= \frac{1}{2^{n+1}} \int_{-\infty}^{\infty} \int_{\max(0,\overline{s}-x,x-\underline{s})}^{\infty} \rho^{-n-1} \mathrm{d}\rho \mathrm{d}x \\ &= \frac{1}{2^{n+1}} \left(\int_{-\infty}^{\mu} \int_{\overline{s}-x}^{\infty} \rho^{-n-1} \mathrm{d}\rho \mathrm{d}x + \int_{\mu}^{\infty} \int_{x-\underline{s}}^{\infty} \rho^{-n-1} \mathrm{d}\rho \mathrm{d}x \right) = \frac{1}{2^n} \frac{1}{n(n-1)} \frac{1}{\sigma_{t-n;i}^{n-1}}. \end{aligned} \tag{30}$$

With it, the expectation of $\rho_{t-n;i}$ has the form

$$E[\rho_{t-n;i} | \underline{s}_{t-n;i}, \overline{s}_{t-n;i}, n] = E[\rho_{t-n;i} | \mu_{t-n;i}, \sigma_{t-n;i}, n] = 0.5 J_n / J_{n+1} = \frac{n}{n-2} \sigma_{t-n;i}, \tag{31}$$

and the expectation of $x_{t-n;i}$ becomes

$$E[x_{t-n;i} | \underline{s}_{t-n;i}, \overline{s}_{t-n;i}, n] = E[x_{t-n;i} | \mu_{t-n;i}, \sigma_{t-n;i}, n] = K / J_{n+1} = \mu_{t-n;i}.$$

These formulae exploit the definitions (28), the normalisation integral (30), and the equality

$$K \equiv \frac{\int_{-\infty}^{\mu} x \int_{\overline{s}-x}^{\infty} \rho^{-n-1} \mathrm{d}\rho \mathrm{d}x + \int_{\mu}^{\infty} x \int_{x-\underline{s}}^{\infty} \rho^{-n-1} \mathrm{d}\rho \mathrm{d}x}{2^{n+1}} = \frac{1}{2^n} \frac{1}{n(n-1)} \frac{\mu_{t-n;i}}{\sigma_{t-n;i}^{n-1}}.$$

6. ILLUSTRATIVE EXPERIMENTS

This section provides an example on which the functionality of the proposed estimator is illustrated. It also illustrates validity of the claims made with respect to KF-based estimators. The reader interested in the motivating practical application is referred to works [4–6, 33]. A detailed study including the innovative features of the proposed estimator will be published in an independent paper.

6.1. Simulation setup

Experiments were performed with the simulated system of the form (1), (2), and (4) given by

$$\mathbf{A} = \begin{bmatrix} 1 & -0.5 & 0.2 \\ 0.5 & 0.1 & 0 \\ 0.3 & 0 & -0.1 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 0.1 \\ 0.6 \\ 0.3 \end{bmatrix}, \mathbf{C} = \begin{bmatrix} 1 & 0 & 0.5 \\ 0 & 1 & 0.5 \end{bmatrix}, \mathbf{D} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\mathbf{E} = \begin{bmatrix} 1 & 0.5 & 0.25 \\ 0 & 1 & 0.3 \\ 0 & 0 & 1 \end{bmatrix}, \mathbf{F} = \begin{bmatrix} 1 & 0.4 \\ 0 & 1 \end{bmatrix}, \mathbf{q} = \begin{bmatrix} 0.1 & 0.1 & 0.1 \end{bmatrix}', \mathbf{r} = \begin{bmatrix} 0.3 & 0.3 \end{bmatrix}'.$$

The correlation-expressing entries $\boldsymbol{\theta} \equiv [E_{1,2},\ E_{1,3},\ E_{2,3},\ F_{1,2}]'$ are taken as unknown; for details, see Section 5.1.2. Thus, the estimated internals are $\mathbf{X}_\tau = [\mathbf{x}'_\tau, \ldots, \mathbf{x}'_{\tau-\Delta}, \mathbf{x}'_{\tau-\Delta-1}, \boldsymbol{\theta}', \mathbf{q}', \mathbf{r}']'$.

The system was stimulated in open loop by independent identically distributed inputs u_τ having uniform distribution on the interval $[-1, 1]$. The experiments ran for $\tau \in \tau^\star = \{1, 2, \ldots, T\}$, $T = 2500$. The state constraints $\underline{\mathbf{x}} \leqslant \mathbf{x}_t \leqslant \overline{\mathbf{x}}$ were applied, see (13), with $\underline{\mathbf{x}} = [-0.80; -0.61; -0.40]$, $\overline{\mathbf{x}} = [0.76; 0.66; 0.44]$. They were attained in less than 30% cases.

The memory length Δ varied in order to demonstrate its influence on the estimation quality and evaluation time.

Then, the results obtained by the LSU estimator were compared with those based on the truncated KF. In this case, we changed the simulated system by setting $\mathbf{E} = \mathbf{I}_{(3)}$, $\mathbf{F} = \mathbf{I}_{(2)}$ to have fully comparable results. KF ran with noise dispersions matching the simulated ones and then with incorrectly chosen, three times higher, values.

DOI: 10.1002/acs

Additional estimation runs with LSU models differing in the state dimension $\ell_{\mathbf{x}}$ were also made to inspect dependence of the evaluation time on problem size.

The evaluations ran in MATLAB interpreter so that time demands have had a relative meaning only. The function 'linprog' from MATLAB optimisation toolbox solved the LP problem (22). Naturally, any general-purpose LP programming code could be used. The freely available implementation found at the link [34] was used for runs with the truncated KF.

6.2. *Performance criteria*

The estimation quality and comparison of the LSU estimator with the truncated KF exploits the following criteria.

The absolute prediction error $\delta(y_{\tau;i})$ and its mean $\bar{\delta}(y_i)$ are defined as follows:

$$\delta(y_{\tau;i}) \equiv |y_{\tau;i} - \hat{y}_{\tau;i}|,\ \tau \in \tau^\star,\ \bar{\delta}(y_i) = \frac{1}{T-\Delta} \sum_{\tau=\Delta+1}^{T} \delta(y_{\tau;i}),\ i = 1, 2, \qquad (32)$$

where y_τ is the simulated output and $\hat{y}_\tau$ is its prediction based on estimates gained from data observed up to time $\tau - 1$.

The absolute error of the state estimates $\delta(x)$ and its mean $\bar{\delta}(x)$ are defined similarly

$$\delta(x_{\tau;j}) \equiv |x_{\tau;j} - \hat{x}_{\tau;j}|,\ \tau \in \tau^\star,\ \bar{\delta}(x_j) = \frac{1}{T-\Delta} \sum_{\tau=\Delta+1}^{T} \delta(x_{\tau;j}),\ j = 1, 2, 3, \qquad (33)$$

where $\mathbf{x}_\tau$ is the simulated state and $\hat{\mathbf{x}}_\tau$ is its estimate based on data measured up to time $\tau - 1$.

The evaluation complexity is quantified via a mean computational time t_c per estimation step.

The mean volume V of the interval LSU estimate (31) is evaluated for the largest $n = \Delta$ and $\kappa = 1$. Its KF counterpart V_{KF} is the mean of ellipsoids volumes determined by covariance matrices P_τ resulting from KF

$$V \equiv \left(\frac{2\Delta}{\Delta-2}\right)^3 \frac{\sum_{\tau=\Delta+1}^{T} \Pi_{j=1}^{3} \sigma_{\tau-\Delta;j}}{T-\Delta},\quad V_{KF} \equiv \frac{\sum_{\tau=\Delta+1}^{T} |P_\tau|}{T-\Delta}. \qquad (34)$$

The confidence N in the LSU interval estimate is expressed by

$$N \equiv (\text{the number of } \mathbf{x}_\tau \notin \text{the box with ends (29) for } n = \Delta) \times \frac{100}{T-\Delta},\quad \tau = \Delta+1, \ldots, T, \qquad (35)$$

that is, by the portion of the realised states $\mathbf{x}_t$ outside of the constructed LSU interval estimate for the specified κ and $\tau \in (\Delta + 1, T)$. Similarly, the confidence N_{KF} in the KF estimate is expressed by

$$N_{KF} \equiv \left(\text{the number of } \mathbf{x}_\tau \notin (\mathbf{x}_\tau - \hat{\mathbf{x}}_\tau)' P_\tau^{-1} (\mathbf{x}_\tau - \hat{\mathbf{x}}_\tau) \leqslant \kappa^3\right) \times \frac{100}{T-\Delta},\quad \tau = \Delta+1, \ldots, T. \qquad (36)$$

Note that the omission of the initial Δ values in KF-related characteristics makes the results comparable with their LSU counterparts.

6.3. *Results*

Table I summarises the numerical indicators of the LSU filtering quality as a function of the memory length Δ; see definition (3).

Figures are appropriately zoomed to provide an insight into a typical behaviour, which is otherwise lost when displaying the whole estimation progress. Figure 1, shows estimates of the unknown F entry as a function of time for a pair Δ values. Figure 2 displays the interval estimates of the third state entry as a function of time. Additionally, Figure 3 shows dependence of the mean computational time t_c per estimation step on the size of the estimated state.

Tables II and III provide comparison of the proposed LSU estimator with the truncated KF (for uncorrelated noise entries).

DOI: 10.1002/acs

Table I. Performance of the LSU estimator as a function of the window lengths Δ.

Δ	t_c [s]	$V \times 10^{-3}$	N [%] $\kappa = 1.0$	$\kappa = 1.5$	$\kappa = 2.0$
5	0.03	17	34	12	5
10	0.05	8	46	16	7
15	0.09	6	51	17	6
20	0.17	6	53	17	6
25	0.28	5	53	17	6
30	0.42	5	53	16	6
35	0.62	5	53	16	5
40	0.87	5	53	16	5
45	1.14	5	53	16	5
50	1.53	5	54	16	5

t_c is the computation time per estimation step; V is the volume of the interval estimate (34); N expresses estimator reliability; see (35).

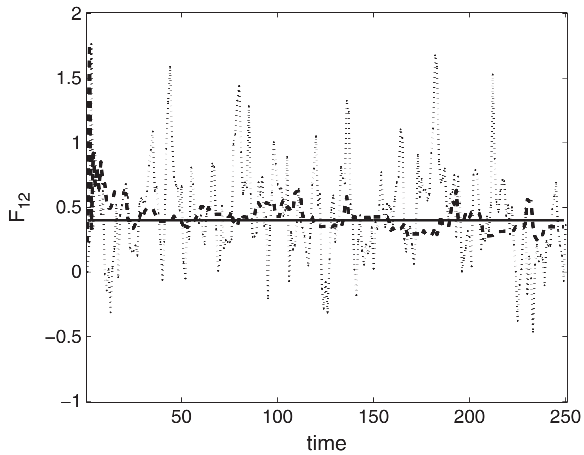

Figure 1. The point estimates $\hat{F}_{12}$ as a function of time (marked by **...** for $\Delta = 5$ and by **- -** for $\Delta = 25$) and true value F_{12} (marked by —).

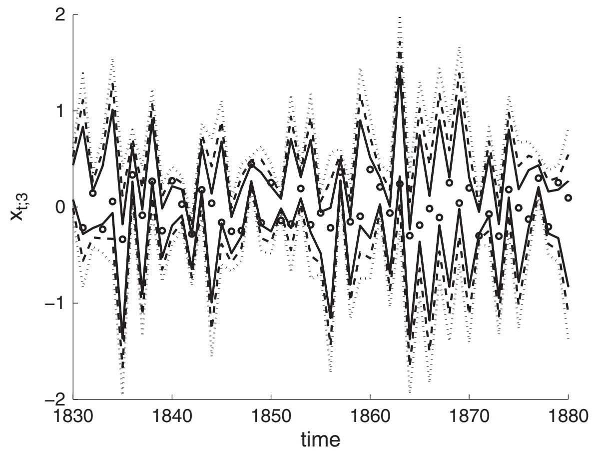

Figure 2. A zoomed part of the interval estimates (29) of $x_{t;3}$ as a function of time; **o** marks the true values of $x_{t;3}$, **–** bounds the interval estimates for $\kappa = 1$, **- -** bounds the interval estimates for $\kappa = 1.5$, **...** bounds the interval estimates for $\kappa = 2$. These intervals cover the min-max range $[\underline{s}, \overline{s}]$; see (27).

DOI: 10.1002/acs

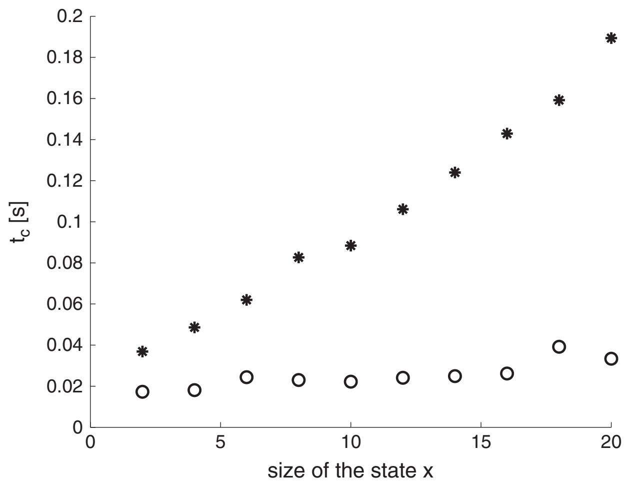

Figure 3. Dependence of the mean computation time t_c of the LSU estimator on the size of the estimated state $\ell_{\mathbf{x}}$ ($\ell_{\mathbf{y}} = 1$, $\ell_{\mathbf{u}} = 1$ are fixed), * holds for $\Delta = 25$, **o** for $\Delta = 5$.

Table II. Comparison of the LSU estimator and the truncated KF according to the mean absolute state estimation errors $\bar{\delta}(x_i)$, $i = 1, 2, 3$, (33), and mean absolute prediction errors $\bar{\delta}(y_j)$, $j = 1, 2$, (32).

Estimator	State estimation errors $\bar{\delta}(x_1)$	$\bar{\delta}(x_2)$	$\bar{\delta}(x_3)$	Prediction errors $\bar{\delta}(y_1)$	$\bar{\delta}(y_2)$
LSU, $\Delta = 5$	0.15	0.10	0.06	0.10	0.10
LSU, $\Delta = 25$	0.09	0.07	0.05	0.13	0.11
KF with well-adjusted covariances	0.07	0.06	0.05	0.13	0.15
KF with incorrect covariances	0.10	0.11	0.08	0.50	0.51

Table III. Comparison of confidence interval volumes V, V_{KF} and reliability N, N_{KF} of the LSU estimator and truncated KF (cf. (34) and (35)).

		N [%]		
Estimator	V	$\kappa = 1.0$	$\kappa = 1.5$	$\kappa = 2.0$
LSU, $\Delta = 5$	1.7×10^{-2}	30	11	4
LSU, $\Delta = 25$	5.6×10^{-3}	49	14	5
		N_{KF} [%]		
Estimator	V_{KF}	$\kappa = 1.0$	$\kappa = 1.5$	$\kappa = 2.0$
KF with well-adjusted covariances	7.1×10^{-7}	67	9	1
KF with incorrect covariances	7.1×10^{-4}	10	1	0

6.4. Discussion

The presented as well as a range of nonpresented experiments confirmed the expected properties of the proposed algorithm.

- Quality of the LSU estimator increases with memory length Δ up to saturation. The improvement is paid by increasing computational time (cf. Table I).
- The volume of the interval estimate also decreases with the memory length up to saturation. When stabilised, the reliability of estimates expressed by N stabilises, too. Then, the scaling factor κ qualitatively behaves as the multiple of standard deviation used for practical determination of quartiles of normal distribution (cf. Table I and also Figure 2).

DOI: 10.1002/acs

- The point estimates of parameters move around true values. They do not stabilise completely, as the finite memory-based estimation acts as a sort of forgetting and make the estimator permanently sensitive to parameter changes (cf. Figure 1).
- The computational complexity is well under control and allows treatment of problems of real-life complexity (Figure 3).

The comparison of the LSU estimator and the truncated KF also confirms the expectations.

- Our estimator with a sufficiently long memory is competitive with the truncated KF *having well adjusted variances* of state and measurement noise, both with respect to state estimation and prediction (cf. Table II).
- The truncated KF becomes worse than the LSU estimator in both their roles if the noise variances are improperly adjusted.
- The truncated KF provides much smaller confidence regions than the LSU estimator. Here, the effect of the finite memory is visible. Outcomes of both estimators are relatively reliable. Our limited experience indicates that the combined effect of truncation and improperly tuned variances of KF is quite complex and hardly predictable. In this respect, the LSU exhibits a bit worse but robustly and predictably varying quality (cf. Table III).

7. CONCLUSION

In the paper, a joint estimator of parameters and states of a linear state-space model with entry-wise correlated uniform noises is proposed.

Practically, the estimator meets conditions allowing its wide simple use for complex multivariate systems. It (a) avoids the difficult tuning of noise covariances inhibiting the efficient use of the standard KF, while tuning the estimator optional parameters is simple and robust to their choice; (b) provides recursively both point estimates and information on their precision; (c) incorporates simply hard bounds on the estimated variables, which decreases the ambiguity of the estimates as the inspected model becomes set smaller because of the exploitation of the available information on possible ranges of estimated variables; (d) fits to robust control applications that deal with polytope-type descriptions of uncertainty; and (e) copes simply with missing data [5].

Methodologically, the work opens a way to state and parameter estimators fitting to models with bounded (not necessarily uniform) noises. Also, the proposed Bayesian counterpart of ensemble filters is applicable to a range of moving-window estimators. Even for the proposed estimator, it opens a way to a systematic choice of the proper window length Δ. At present, it is selected experimentally under the classical guiding: (i) select the longest Δ you can computationally afford or (ii) stop the increase of Δ if the observed improvement of the prediction quality is negligible. The ensemble-type processing will, however, allow to make the choice via classical Bayesian structure estimation [35] and its simplified variants originating in Akaike's work [36]. The claimed advantages of the stochastic interpretation of noises are well seen on this important problem. Future work will include the following:

- an exploitation of the information on the estimate precision for estimating a fine model structure as well as the memory length via Bayesian hypotheses testing [29];
- selection or development of specialised linear or nonlinear programming codes, which exploit the fact that good prior guesses are available from preceding time steps for a majority of the estimated internals;
- a use of the LSU model in marginalised particle filters;
- an exploitation of the LSU models as factors in models considering different distributions and their dynamic mixtures [37]; and
- extensive tests on variety of real data.

Despite the significant extent of these open problems, the LSU model is sufficiently matured to be used in difficult problems requiring recursive estimation [38].

DOI: 10.1002/acs

APPENDIX A: ON-LINE ESTIMATION OF INTERNALS IN DETAIL

A.1. Estimation of state and noise boundaries: void $\boldsymbol{\theta}$

Here, the arrays involved in (22) for void $\boldsymbol{\theta}$, that is, $\mathbf{X}_\tau$ (23), are presented. After rearranging (1) and (2) so that entries of $\mathbf{X}_\tau$ are on the left-hand side, it holds, for $t = \tau - \Delta, \ldots, \tau$,

$$\begin{aligned} \mathbf{E}\mathbf{x}_t - \mathbf{A}\mathbf{x}_{t-1} - \mathbf{q} &\leqslant \mathbf{B}\mathbf{u}_t \\ -\mathbf{E}\mathbf{x}_t + \mathbf{A}\mathbf{x}_{t-1} - \mathbf{q} &\leqslant -\mathbf{B}\mathbf{u}_t \\ \mathbf{C}\mathbf{x}_t - \mathbf{r} &\leqslant \mathbf{F}\mathbf{y}_t \\ -\mathbf{C}\mathbf{x}_t - \mathbf{r} &\leqslant -\mathbf{F}\mathbf{y}_t. \end{aligned}$$

The matrices $\mathcal{A}_\tau$ and $\mathcal{B}_\tau$ have the following form:

$$\mathcal{A}_\tau = \begin{bmatrix} \mathcal{A}11 & \mathcal{A}12 \\ \mathcal{A}21 & \mathcal{A}22 \end{bmatrix}, \ \mathcal{B}_\tau = \begin{bmatrix} \mathcal{B}1_\tau \\ \mathcal{B}2_\tau \end{bmatrix}, \text{ with}$$

$$\mathcal{A}11 = \begin{bmatrix} \mathbf{E} & -\mathbf{A} & \mathbf{0}_{(\ell_x,\ell_x)} & \ldots & \mathbf{0}_{(\ell_x,\ell_x)} & \mathbf{0}_{(\ell_x,\ell_x)} \\ \mathbf{0}_{(\ell_x,\ell_x)} & \mathbf{E} & -\mathbf{A} & \ldots & \mathbf{0}_{(\ell_x,\ell_x)} & \mathbf{0}_{(\ell_x,\ell_x)} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_{(\ell_x,\ell_x)} & \mathbf{0}_{(\ell_x,\ell_x)} & \mathbf{0}_{(\ell_x,\ell_x)} & \ldots & \mathbf{E} & -\mathbf{A} \end{bmatrix} \otimes \mathbf{K}$$

$$\mathcal{A}12 = \mathbf{1}_{(2(\Delta+1))} \otimes \left[-\mathbf{I}_{(\ell_x)} \ \mathbf{0}_{(\ell_x,\ell_y)}\right]$$

$$\mathcal{A}21 = \begin{bmatrix} \mathbf{C} & \ldots & \mathbf{0}_{(\ell_y,\ell_x)} & \mathbf{0}_{(\ell_y,\ell_x)} & \mathbf{0}_{(\ell_y,\ell_x)} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \mathbf{0}_{(\ell_y,\ell_x)} & \ldots & \mathbf{0}_{(\ell_y,\ell_x)} & \mathbf{C} & \mathbf{0}_{(\ell_y,\ell_x)} \end{bmatrix} \otimes \mathbf{K}$$

$$\mathcal{A}22 = \mathbf{1}_{(2(\Delta+1))} \otimes \left[\mathbf{0}_{(\ell_y,\ell_x)} \ -\mathbf{I}_{(\ell_y)}\right]$$

$$\mathcal{B}1_\tau = \begin{bmatrix} \mathbf{B}u_\tau & \ldots & \mathbf{B}\mathbf{u}_{\tau-\Delta} \end{bmatrix}' \otimes \mathbf{K}, \ \mathcal{B}2_\tau = \begin{bmatrix} \mathbf{F}y_\tau & \ldots & \mathbf{F}y_{\tau-\Delta} \end{bmatrix}' \otimes \mathbf{K}$$

where $\otimes$ denotes Kronecker product, and $\mathbf{K} \equiv [1 \ -1]'$,
$\mathbf{0}_{(\alpha,\beta)}$ is the zero matrix of the indicated dimensions. Note that $\mathcal{A}_\tau$ is time invariant.

Prior information on $\mathbf{X}_\tau$ reflecting constraints (18) is in the form $\underline{\mathbf{X}}_\tau \leqslant \mathbf{X}_\tau \leqslant \overline{\mathbf{X}}_\tau$ with

$$\begin{aligned} \underline{\mathbf{X}}'_\tau &= \begin{bmatrix} \underline{\mathbf{x}}'_\tau, \ldots, \underline{\mathbf{x}}'_{\tau-\Delta-1}, & \mathbf{0}'_{(\ell_x,1)} & \mathbf{0}'_{(\ell_y,1)} \end{bmatrix}' \\ \overline{\mathbf{X}}'_\tau &= \begin{bmatrix} \overline{\mathbf{x}}'_\tau, \ldots, \overline{\mathbf{x}}'_{\tau-\Delta-1}, & \overline{\mathbf{q}}', & \overline{\mathbf{r}}' \end{bmatrix}'. \end{aligned}$$

A.2. Estimation of state, noise boundaries, and correlating matrices $\mathbf{E}$ *and* $\mathbf{F}$

Here, the matrices and vectors involved in (22) for $\mathbf{X}_\tau$ (24) after linearisation (26) are presented. Because of the linearisation, the majority of matrices and vectors become truly time and data dependent.

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

$$\mathcal{A}11 = \begin{bmatrix} \hat{\mathbf{E}}_\tau & -\mathbf{A} & \mathbf{0}_{(\ell_x,\ell_x)} & \ldots & \mathbf{0}_{(\ell_x,\ell_x)} & \mathbf{0}_{(\ell_x,\ell_x)} \\ \mathbf{0}_{(\ell_x,\ell_x)} & \hat{\mathbf{E}}_\tau & -\mathbf{A} & \ldots & \mathbf{0}_{(\ell_x,\ell_x)} & \mathbf{0}_{(\ell_x,\ell_x)} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_{(\ell_x,\ell_x)} & \mathbf{0}_{(\ell_x,\ell_x)} & \mathbf{0}_{(\ell_x,\ell_x)} & \ldots & \hat{\mathbf{E}}_\tau & -\mathbf{A} \end{bmatrix} \otimes \mathbf{K},$$

DOI: 10.1002/acs

where $\hat{\mathbf{E}}_\tau = \mathbf{I}_{(\ell_x)} + \hat{\mathbf{\Lambda}}_\tau$; see (26).

$$\mathcal{A}12 = \begin{bmatrix} \hat{\mathbf{X}}_\tau^{(1)} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{X}}_\tau^{(2)} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \hat{\mathbf{X}}_\tau^{(\ell_x-1)} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \hat{\mathbf{X}}_{\tau-\Delta}^{(1)} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{X}}_{\tau-\Delta}^{(2)} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \hat{\mathbf{X}}_{\tau-\Delta}^{(\ell_x-1)} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \end{bmatrix} \otimes \mathbf{K},$$

where $\hat{\mathbf{X}}_t^{(i)} = [\hat{x}_{t;i+1}, \dots, \hat{x}_{t;\ell_x}]$, $t = \tau, \dots, \tau - \Delta$, $i = 1, \dots, \ell_x - 1$ (cf. (10)),

$$\mathcal{A}13 = \mathbf{0}_{(2\ell_x(\Delta+1), \ell_y - 1 + \ell_y - 2 + \dots + \ell_y - \ell_y + 1)}, \quad \mathcal{A}14 = \mathbf{1}_{(2(\Delta+1))} \otimes \left[-\mathbf{I}_{(\ell_x)} \ \mathbf{0}_{(\ell_x, \ell_y)} \right],$$

$$\mathcal{A}21 = \begin{bmatrix} \mathbf{C} & \dots & \mathbf{0}_{(\ell_y,\ell_x)} & \mathbf{0}_{(\ell_y,\ell_x)} & \mathbf{0}_{(\ell_y,\ell_x)} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \mathbf{0}_{(\ell_y,\ell_x)} & \dots & \mathbf{0}_{(\ell_y,\ell_x)} & \mathbf{C} & \mathbf{0}_{(\ell_y,\ell_x)} \end{bmatrix} \otimes \mathbf{K},$$

$$\mathcal{A}22 = \mathbf{0}_{(2\ell_y(\Delta+1), \ell_x - 1 + \ell_x - 2 + \dots \ell_x - \ell_x + 1)}, \quad \mathcal{A}24 = \mathbf{1}_{(2(\Delta+1))} \otimes \left[\mathbf{0}_{(\ell_y,\ell_x)} \ -\mathbf{I}_{(\ell_y)} \right],$$

$$\mathcal{A}23 = \begin{bmatrix} \mathbf{Y}_\tau^{(1)} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_\tau^{(2)} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{Y}_\tau^{(\ell_y-1)} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{Y}_{\tau-\Delta}^{(1)} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_{\tau-\Delta}^{(2)} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{Y}_{\tau-\Delta}^{(\ell_y-1)} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \end{bmatrix} \otimes \mathbf{K},$$

where $\mathbf{Y}_t^{(i)} = [y_{t;i+1}, \dots, y_{t;\ell_y}]$, $t = \tau, \tau - 1, \dots, \tau - \Delta$, $i = 1, \dots, \ell_y - 1$ (cf. (9)),

$$\mathcal{B}1_\tau = \begin{bmatrix} \mathbf{B}u_\tau + \hat{\Lambda}\hat{\mathbf{x}}_\tau \\ \vdots \\ \mathbf{B}\mathbf{u}_{\tau-\Delta} + \hat{\Lambda}\hat{\mathbf{x}}_{\tau-\Delta} \end{bmatrix} \otimes \mathbf{K}, \quad \mathcal{B}2_\tau = \begin{bmatrix} y_\tau & \dots & y_{\tau-\Delta} \end{bmatrix}' \otimes \mathbf{K}$$

where $\underline{\mathbf{X}}_\tau$, $\overline{\mathbf{X}}_\tau$ are known vectors; they stem from the constraints (18) and have the following form:

$$\underline{\mathbf{X}}'_\tau = \left[\ \underline{\mathbf{x}}'_\tau, \dots, \underline{\mathbf{x}}'_{\tau-\Delta-1}, \ \mathrm{col}(\underline{\mathbf{E}})', \ \mathrm{col}(\underline{\mathbf{F}})', \ \mathbf{0}'_{(\ell_x)}, \ \mathbf{0}'_{(\ell_y)} \ \right]$$
$$\overline{\mathbf{X}}'_\tau = \left[\ \overline{\mathbf{x}}'_\tau, \dots, \overline{\mathbf{x}}'_{\tau-\Delta-1}, \ \mathrm{col}(\overline{\mathbf{E}})', \ \mathrm{col}(\overline{\mathbf{F}})', \ \overline{\mathbf{q}}', \ \overline{\mathbf{r}}' \ \right].$$

DOI: 10.1002/acs

ACKNOWLEDGEMENTS

This research was partially supported by the grants TA ČR TA01030123 and GA ČR 13-13502S.

REFERENCES

1. Astrom K. *Introduction to Stochastic Control*. Academic Press: New York, 1970.
2. Bohlin T. *Interactive System Identification: Prospects and Pitfalls*. Springer-Verlag: New York, 1991.
3. Chen Z. Bayesian filtering: From Kalman filters to particle filters, and beyond, 2004. Manuscript. (Available from: http://www.damas.ift.ulaval.ca/_seminar/filesA11/10.1.1.107.7415.pdf). [Accessed on October 2013].
4. Pavelková L. Estimation of models with uniform innovations and its application on traffic data. *Ph.D. Thesis*, Czech Technical University in Prague, Faculty of Transportation Sciences, 2008.
5. Pavelková L. Nonlinear Bayesian state filtering with missing measurements and bounded noise: its application to vehicle position estimation. *Kybernetika* 2011; **47**(3):370–384.
6. Suzdaleva E, Nagy I, Pavelková L. Fuel consumption optimization: early experiments. *Preprints of the 16th IFAC Symposium on System Identification Sysid 2012*, Brussels, Belgium, 2012; 751–756.
7. Grewal MS, Andrews AP. *Kalman Filtering: Theory and Practice Using MATLAB*. Wiley: Hoboken, New Jersey, 2008.
8. Axelsson P, Orguner U, Gustafsson F, Norrlof M. ML estimation of process noise variance in dynamic systems. *Proceedings of 18th IFAC World Congress*, Milano, Italy, 2011; 5609–5614.
9. Matisko. P, Havlena V. Noise covariance estimation for Kalman filter tuning using Bayesian approach and Monte Carlo. *International Journal of Adaptive Control and Signal Processing* 2013. DOI: 10.1002/acs.2369.
10. Gaspari G, Cohn SE. Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society* 1999; **125**(554):723–757.
11. Kárný M, Böhm J, Guy TV, Jirsa L, Nagy I, Nedoma P, Tesař L. *Optimized Bayesian Dynamic Advising: Theory and Algorithms*. Springer: London, 2005.
12. Suzdaleva E, Nagy I. Recursive state estimation for hybrid systems. *Applied Mathematical Modelling* 2012; **36**(4):1347–1358.
13. Simon D. *Optimal State Estimation: Kalman, H-Infinity, and Nonlinear Approaches*. Wiley-Interscience, 2006. ISBN: 978-0471708582.
14. Simon D, Simon D. Constrained Kalman filtering via density function truncation for turbofan engine health estimation. *International Journal of Systems Science* 2010; **41**:159–171.
15. Doucet A, De Freitas N, Gordon N. *Sequential Monte Carlo Methods in Practice*. Springer: NY, USA, 2001.
16. Shao X, Huang B, Lee J. Constrained Bayesian state estimation – a comparative study and a new particle filter based approach. *Journal of Process Control* 2010; **20**(1):143–157.
17. Norton J. Special issue on bounded-error estimation: Issue 1. *International Journal of Adaptive Control and Signal Processing* 1994; **8**:1–118.
18. Norton J. Special issue on bounded-error estimation: Issue 2. *International Journal of Adaptive Control and Signal Processing* 1995; **9**:1–132.
19. Polyak B, Nazin S, Durieu C, Walter E. Ellipsoidal parameter or state estimation under model uncertainty. *Automatica* 2004; **40**(7):1171–1179.
20. Bemporad A, Filippi C, Torrisi F. Inner and outer approximations of polytopes using boxes. *Computational Geometry* 2004; **27**:151–178.
21. Reynet O, Jaulin L. A new interval-based method to characterize estimability. *International Journal of Adaptive Control and Signal Processing* 2011; **25**(3):288–294.
22. DeGroot M. *Optimal Statistical Decisions*. McGraw-Hill: New York, 1970.
23. Garthwaite P, Kadane J, O´Hagan A. Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association* 2005; **100**(470):680–700.
24. Kárný M, Nedoma P, Khailova N, Pavelková L. Prior information in structure estimation. *IEE Proceedings — Control Theory and Applications* 2003; **150**(6):643–653.
25. Kárný M, Guy T. Preference elicitation in fully probabilistic design of decision strategies. *Proceedings of the 49th IEEE Conference on Decision and Control*, Atlanta, 2010; 1–6.
26. Ninness B, Goodwin G. Rapprochement between bounded-error and stochastic estimation theory. *International Journal of Adaptive Control and Signal Processing* 1995; **9**:107–132.
27. Robertson DG, Lee JH, Rawlings JB. A moving horizon-based approach for least-squares estimation. *AIChE Journal* 1996; **42**(8):2209–2224.
28. Berger J. *Statistical Decision Theory and Bayesian Analysis*. Springer: New York, 1985.
29. Peterka V. Bayesian system identification. In *Trends and Progress in System Identification*, Eykhoff P (ed.). Pergamon Press: Oxford, 1981; 239–304.
30. Fletcher R. *Practical Methods of Optimization*. John Wiley & Sons: Great Britain, Padstow, Cornwall, 2000. ISBN: 0471494631.
31. Gillijns S, Mendoza OB, Chandrasekar J, Moor B, Bernstein DS, Ridley A. What is the ensemble Kalman filter and how well does it work? *Proceedings of the 2006 American Control Conference*, Minneapolis, Minnesota, USA, 2006; FrA08.1, DOI: 10.1109/ACC.2006.1657419.

32. Shore J, Johnson R. Axiomatic derivation of the principle of maximum entropy & the principle of minimum cross-entropy. *IEEE Transactions on Information Theory* 1980; **26**(1):26–37.
33. Pavelková L. State estimation with missing data and bounded uncertainty. *Technical Report 2296*, ÚTIA AV ČR, Praha, 2011. (Available from: http://www.utia.cz/node/631/0357970). [Accessed on October 2013].
34. Simon D. Kalman filtering with state constraints: a survey of linear and nonlinear algorithms. *IET Control Theory And Applications* 2010; **4**(8):1303–1318. Algorithms (Available form: http://academic.csuohio.edu/simond/ConstrKF). [Accessed on October 2013].
35. Berec L. Model structure identification: global and local views. Bayesian solution. *Ph.D. Thesis*, Czech Technical University, Faculty of Nuclear Sciences and Physical Engineering, Prague, 1998.
36. Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 1974; **19**:716–722.
37. Nagy I, Suzdaleva E, Kárný M, Mlynářová T. Bayesian estimation of dynamic finite mixtures. *International Journal of Adaptive Control and Signal Processing* 2011; **25**(9):765–787.
38. Pavelková L. Vehicle position estimation using GPS/CAN data based on nonlinear programming. *Proceedings of the 13th IASTED International Conference on Intelligent Systems and Control*, Cambridge, UK, 2011; 208–215. ISBN 978-0-88986-889-2.

DOI: 10.1002/acs