

TOWARDS BAYESIAN FILTERING ON RESTRICTED SUPPORT

Lenka Pavelková, Miroslav Kárný, Václav Šmídl

Institute of Information Theory and Automation, Prague, Czech Republic

ABSTRACT

Linear state-space model with uniformly distributed innovations is considered. Its state and parameters are estimated under hard physical bounds. Off-line maximum a posteriori probability estimation reduces to linear programming. No approximation is required for sole estimation of either model parameters or states. The noise bounds are estimated in both cases. The algorithm is extended to: (i) on-line mode by estimating within a sliding window, and (ii) joint state and parameter estimation. This approach may be used as a starting point for full Bayesian treatment of distributions with restricted support.

1. INTRODUCTION

Statistically inclined research concentrates naturally on stochastic features of the modeled systems and evaluated estimates. However, practically oriented engineers are more concerned with bounds on parameters and states. Respecting uncertain physical bounds is a challenging problem. A range of approaches to it has been published, including unknown-but-bounded methodology [1] or elliptical approximations [2] to name a few. They often intentionally avoid statistical aspects in order to focus on bounds. This dominating focus restricts their use. Here, we respect both uncertainty and physical bounds by exploiting Bayesian approach. Hard physical bounds are modeled by restricted support of the corresponding joint probability density function (pdf).

We show that off-line evaluation of maximum a posteriori probability (MAP) estimate of linear state-space model with uniform distributions of innovations reduces to linear programming (LP). This observation is a starting point for extensions of the approach for more complex models. As a first step, we propose the use of sliding-window for on-line estimation. By swapping between state and parameter estimations, joint parameter and state estimation is obtained. The use of Taylor expansion for approximation of products of unknowns solves also the joint parameter and state estimation.

Throughout, \equiv means equality by definition; \hat{x} length of the vector x ; vectors are always columns; $'$ denotes transposition; x , u , y are unobserved state, known input and observed

output of the system, respectively; entries of the state, input and output are real, i.e. $x \in \mathbb{R}^{\hat{x}}$, $u \in \mathbb{R}^{\hat{u}}$, $y \in \mathbb{R}^{\hat{y}}$; the subscript $t \in \{0, 1, 2, \dots\}$ labels discrete time; the sequence of states $(x_t, x_{t-1}, \dots, x_k)$, $k \leq t$ is ordered into the vector $x^{k:t} \equiv [x'_t, x'_{t-1}, \dots, x'_k]'$; non-numerical subscripts are placed to the left of the basic symbol; \underline{x} , \bar{x} are lower and upper bound on x , respectively (they are used entry-wise).

2. LINEAR UNIFORM STATE-SPACE MODEL

We consider the standard linear state-space model

$$x_t = Ax_{t-1} + Bu_t + {}^x e_t, \quad y_t = Cx_t + Du_t + {}^y e_t, \quad (1)$$

known from Kalman filtering (KF) theory [3]. The involved time-invariant matrices A , B , C , D have appropriate dimensions. Unlike in the KF case, the distributions of vector innovations ${}^x e_t$ and ${}^y e_t$ are assumed to be uniform

$$f({}^x e_t) = \mathcal{U}(0, {}^x r), \quad f({}^y e_t) = \mathcal{U}(0, {}^y r). \quad (2)$$

Here, $\mathcal{U}(\mu, {}^x r)$ denotes uniform pdf on the box with the center μ and half-width of the support interval ${}^x r$.

The unknown entries of A , B , C , D are collected into Θ .

Equations (1) together with the assumptions (2) define the linear uniform state-space model (LU). This model complements its classical Gaussian counterpart and provides the following advantages: (i) it respects natural bounds on stochastic disturbances, (ii) it allows estimation of the innovation range (unlike KF), and (iii) it allows—without excessive computational demands—to respect hard, physically justified, prior bounds on model parameters and states. Moreover, the presence of finite hard bounds makes the approximate extensions of basic estimation algorithms (such as joint parameter and state estimation) more robust.

3. OFF-LINE ESTIMATION

Here, the joint posterior pdf of states and parameters is derived. Then, its maximization is converted into a standard formulation of linear programming (LP).

3.1. Posterior pdf

We assume that the generator of the inputs $u^{1:t} \equiv [u'_t, \dots, u'_1]'$ meets natural conditions of control [4]. They formalize as-

sumption that information about unknown quantities for generating u_t can only be extracted from the observed data $d^{1:t-1}$, where $d_t = (y_t, u_t)$. Then, for a given initial state x_0 , half-widths x_r , y_r and parameters Θ , the joint pdf of data and the state trajectory $x^{1:t}$ of the LU model is

$$f(d^{1:t}, x^{1:t} | x_0, x_r, y_r, \Theta) \propto \prod_{i=1}^n x_{r_i}^{-t} \prod_{j=1}^m y_{r_j}^{-t} \chi(\mathcal{S}). \quad (3)$$

$\chi(\mathcal{S})$ is the indicator of the support \mathcal{S} ; \propto denotes equality up to a constant factor. The convex set \mathcal{S} is given by inequalities,

$$\begin{aligned} -x_r &\leq x_\tau - Ax_{\tau-1} - Bu_\tau \leq x_r \\ -y_r &\leq y_\tau - Cx_\tau - Du_\tau \leq y_r, \end{aligned} \quad (4)$$

where $\tau = 1, 2, \dots, t$. Bayesian estimation of x_0 , x_r , y_r requires to complement the conditional pdf (3) by a *prior pdf* $f(x_0, x_r, y_r | \Theta)$. For known Θ , it can be chosen as uniform pdf on support \mathcal{S}_0 defined by inequalities

$$\mathcal{S}_0 = \{\underline{x}_0 \leq x_0 \leq \bar{x}_0, \quad 0 < x_r \leq \bar{x}_r, \quad 0 < y_r \leq \bar{y}_r\}. \quad (5)$$

The bounds \underline{x}_0 , \bar{x}_0 etc. determine support of the prior pdf.

For unknown Θ , the uniform prior pdf $f(x_0, x_r, y_r, \Theta)$ is chosen on the set (5) extended by conditions $\underline{\Theta} \leq \Theta \leq \bar{\Theta}$.

For fixed observations, $d^{1:t}$, and uniform prior (5), the expression (3)—on support $\mathcal{S} \cap \mathcal{S}_0$ —is proportional to *posterior pdf*. Due to the power $^{-t}$, it is sharply peaked at lower bounds on x_r and y_r . Moreover, the number of vertices of the support is proportional to the number of data. The proportionality factor may be large for realistic systems. Consequently, evaluation of moments of this pdf is computationally demanding. This motivates our focus on MAP estimation of all unknowns.

3.2. MAP estimation via LP

Without loss of generality, we assume that elements of x_r and y_r are (significantly) smaller than 1. Under this assumption, the negative logarithm of the posterior pdf can be approximated by sum of elements of x_r and y_r on the convex, linearly restricted set $\mathcal{S} \cap \mathcal{S}_0$. If the inequalities (4) are linear in the unknowns, the MAP estimation is equivalent to the problem of linear programming (LP) and can be solved by any of the available algorithms. This condition is satisfied if either (i) parameters Θ , or (ii) states $x^{1:t}$, are known. Note that convexity of the set $\mathcal{S} \cap \mathcal{S}_0$ is determined by choice of the prior bounds (5). LP will fail if these are chosen too restrictive.

In this Section, we derive solutions to both cases mentioned above, i.e. (i) estimation of states $x^{1:t}$, and x_r , y_r , given Θ , and (ii) estimation of parameters Θ and x_r , y_r , given the state $x^{1:t}$. Solutions are presented in the standard form of linear programming used by Matlab function `linprog`, i.e.

$$\begin{aligned} &\text{Find a vector } X \text{ such that } J \equiv C'X \rightarrow \min \\ &\text{while } \mathcal{A}X \leq \mathcal{B}, \underline{X} \leq X \leq \bar{X}, \end{aligned} \quad (6)$$

where known matrices and vectors \mathcal{A} , \mathcal{B} , \mathcal{C} , \underline{X} , \bar{X} will be derived for each case.

3.3. Estimation of the state and the noise bounds

In the case of known parameters Θ , the unknowns are the state $x^{1:t}$ and the noise bounds x_r , y_r . Hence, the vector X of (6) is defined as follows:

$$X = \begin{bmatrix} x^{0:t} \\ x_r \\ y_r \end{bmatrix}. \quad (7)$$

The matrices \mathcal{A} , \mathcal{B} , \mathcal{C} , \underline{X} , \bar{X} will be defined using the following conventions:

$M_{(\alpha, \beta)}$ is a matrix with α rows and β columns.

$I_{(\alpha)}$ is the square identity matrix of the order α

$\mathbf{0}_{(\alpha, \beta)}$ is zero matrix of given dimensions.

$K \equiv [-1 \ 1]'$ is a repeatedly used vector.

$\mathbf{1}_{(\alpha)}$, $\mathbf{0}_{(\alpha)}$ are column vectors of ones, and zeros, respectively, both of length α .

$$\text{Kronecker product } G_{(\alpha, \beta)} \otimes H \equiv \begin{bmatrix} G_{11}H & \dots & G_{1\beta}H \\ \vdots & & \vdots \\ G_{\alpha 1}H & \dots & G_{\alpha\beta}H \end{bmatrix}.$$

Operator $\mathcal{R}_{col}(M)$ extends a matrix $M_{(\alpha, \beta)}$ by the zero matrix $\mathbf{0}_{(\alpha, col)}$ from the right, $\mathcal{R}_{col}(M) \equiv [M, \mathbf{0}_{(\alpha, col)}]$.

Operator $\mathcal{L}_{col}(M)$ extends a matrix $M_{(\alpha, \beta)}$ by the zero matrix $\mathbf{0}_{(\alpha, col)}$ from the left, $\mathcal{L}_{col}(M) \equiv [\mathbf{0}_{(\alpha, col)}, M]$.

$\text{col}(M)$ stacks the rows of the matrix M into a column vector.

Using these definitions, the set (4) can be written in the form of (6) as follows:

$$C' \equiv [\mathbf{0}'_{((t+1)\hat{x}, 1)}, \mathbf{1}'_{(\hat{x}+\hat{y})}], \quad (8)$$

$$\mathcal{A} = \begin{bmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} \\ \mathcal{A}_{21} & \mathcal{A}_{22} \end{bmatrix}, \quad \mathcal{B} = [\mathcal{B}_1' \quad \mathcal{B}_2']', \quad \text{with}$$

$$\mathcal{A}_{11} = \mathcal{R}_{\hat{x}}(I_{(t)} \otimes K \otimes I_{(\hat{x})}) - \mathcal{L}_{\hat{x}}(I_{(t)} \otimes K \otimes A), \quad (9)$$

$$\mathcal{A}_{12} = -\mathbf{1}_{(2t)} \otimes \mathcal{R}_{\hat{y}}(I_{(\hat{x})}),$$

$$\mathcal{A}_{21} = \mathcal{R}_{\hat{x}}(I_{(t)} \otimes K \otimes C),$$

$$\mathcal{A}_{22} = -\mathbf{1}_{(2t)} \otimes \mathcal{L}_{\hat{x}}(I_{(\hat{y})}),$$

$$\mathcal{B}_1 = [I_{(t)} \otimes K \otimes B] u^{1:t},$$

$$\mathcal{B}_2 = -[I_{(t)} \otimes K \otimes D] u^{1:t} + [I_{(t)} \otimes K \otimes I_{(\hat{y})}] y^{1:t}.$$

Similarly, the set \mathcal{S}_0 (5) is represented by the following assignments:

$$\underline{X} = \begin{bmatrix} -\infty \times \mathbf{1}_{(2t\hat{x}, 1)} \\ \underline{x}_0 \\ \mathbf{0}_{(\hat{x}, 1)} \\ \mathbf{0}_{(\hat{y}, 1)} \end{bmatrix}, \quad \bar{X} = \begin{bmatrix} \infty \times \mathbf{1}_{(2t\hat{x}, 1)} \\ \bar{x}_0 \\ \bar{x}_r \\ \bar{y}_r \end{bmatrix}. \quad (10)$$

3.4. Estimation of the parameters and the noise bounds

In the case of known state trajectory $x^{1:t}$, the unknowns are parameters A, B, C, D and half-widths x_r, y_r . This case may arise in situations with directly measurable state. Moreover, these results will be needed for joint estimation of state and unknown parameters which will be addressed in the next Section. The unknowns form the vector X of the standard LP form (6) as follows:

$$X \equiv [\text{col}(A)', \text{col}(B)', \text{col}(C)', \text{col}(D)', x_r', y_r']'. \quad (11)$$

Using the introduced conventions, the following assignments transform the set \mathcal{S} (4) into the standard form (6):

$$\begin{aligned} C &\equiv [\mathbf{0}'_{(\hat{x}\hat{x}+\hat{x}\hat{u}+\hat{x}\hat{y}+\hat{u}\hat{y},1)}, \mathbf{1}'_{(\hat{x}+\hat{y},1)}]', \\ A &\equiv \begin{bmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} & \mathcal{A}_{13} \\ \mathcal{A}_{21} & \mathcal{A}_{22} & \mathcal{A}_{23} \end{bmatrix}, \quad B = [\mathcal{B}_1' \quad \mathcal{B}_2']', \\ \mathcal{A}_{11} &\equiv \begin{bmatrix} I_{(\hat{x})} \otimes K \otimes x'_{t-1} & I_{(\hat{x})} \otimes K \otimes u'_t \\ \vdots & \vdots \\ I_{(\hat{x})} \otimes K \otimes x'_0 & I_{(\hat{x})} \otimes K \otimes u'_1 \end{bmatrix}, \quad (12) \\ \mathcal{A}_{12} &\equiv \mathbf{0}_{(2t\hat{x},\hat{x}\hat{y}+\hat{y}\hat{u})}, \quad \mathcal{A}_{13} \equiv -\mathbf{1}_{(2t)} \otimes \mathcal{R}_{\hat{y}}(I_{(\hat{x})}), \\ \mathcal{A}_{21} &\equiv \mathbf{0}_{(2t\hat{y},\hat{x}^2+\hat{x}\hat{u})}, \quad \mathcal{A}_{23} \equiv -\mathbf{1}_{(2t)} \otimes \mathcal{L}_{\hat{x}}(I_{(\hat{y})}), \\ \mathcal{A}_{22} &\equiv \begin{bmatrix} I_{(\hat{y})} \otimes K \otimes x'_t & I_{(\hat{y})} \otimes K \otimes u'_t \\ \vdots & \vdots \\ I_{(\hat{y})} \otimes K \otimes x'_1 & I_{(\hat{y})} \otimes K \otimes u'_1 \end{bmatrix}, \\ \mathcal{B}_1 &= x^{1:t} \otimes K, \quad \mathcal{B}_2 = [I_{(t)} \otimes K \otimes I_{(\hat{y})}] y^{1:t}. \end{aligned}$$

4. ON-LINE ESTIMATION

Standard Bayesian *filtering and smoothing with a fixed lag* $\partial \geq 0$ integrates out from the posterior pdf the superfluous state $x_{t-\partial-1}$ in each time step, t . However, with increasing t , this operation yields increasingly complex support of the posterior pdf and soon becomes intractable. The unknown-but-bounded approaches [1, 2] face this problem by a recursive construction of simple (typically outer) approximation of the support. In order to avoid these approximations, we propose to use a sliding window of length ∂ and apply LP in order to find MAP estimate of the states $x^{t-\partial:t} \equiv [x'_t, \dots, x'_{t-\partial}]'$ on the intersection of sets \mathcal{S} and \mathcal{S}_0 considered for $\tau = t - \partial, \dots, t$. This approximates the limited-memory filter of Jazwinski [3] and provides an attractive alternative to forgetting. In this context, we relax the assumptions of previous Section, i.e. the necessary knowledge of either the state, or parameters Θ . However, this relaxation violates the assumptions of LP and further approximations are needed to restore tractability. In this Section, we outline two possible approaches (i) heuristically motivated technique based on swapping of techniques from Sections 3.3 and 3.4, and (ii) linearization of the inequalities around the last point estimates.

4.1. Swapping-based joint estimation

The idea of this approach is to estimate the state $x^{1:t}$ using technique from Section 3.3, with parameters Θ fixed at their last point estimates. The resulting estimates of states, $\hat{x}^{t-\partial:t}$ are subsequently used in technique from Section 3.4 to obtain new estimates of the parameters Θ . Initial values of the estimates can be found in off-line mode using, for instance, a subspace method [5] or sampling methods, e.g. [6]. It is practically important that the estimates of the noise bounds can be very inaccurate. Performance of this approach will be illustrated on a simple example.

4.2. Expansion-based joint estimation

Linearization of non-linear equations at point estimates is common idea, used in various extensions of KF. It could be applied to inequalities (4) using approximations of the following kind:

$$Ax_\tau \approx \hat{A}_t x_\tau + A\hat{x}_{\tau|t} - \hat{A}_t \hat{x}_{\tau|t}, \quad \tau \in \{t - \partial, t + 1 - \partial, \dots, t\}. \quad (13)$$

where $\hat{A}_t, \hat{x}_{\tau|t}$ are newest available estimates of parameters and states, respectively. Using equivalent expansion for Cx_τ , the resulting inequalities can—once again—be transformed in the standard form of LP (6). The exact assignments are omitted for brevity. The resulting algorithm has two principal distinctions from extended KF; (i) the algorithm updates estimates of the whole window of length ∂ hence, more sophisticated approaches (such as moving average of point estimates) can be used to improve quality of the points of expansion $\hat{A}_t, \hat{x}_{\tau|t}$ in (13); and (ii) the realistic hard bounds on the estimated quantities reduce the ambiguity of the model (arising from estimating a product of two unknowns). From these distinctions we conjecture that the estimation is better conditioned and more robust than extended KF. Simulation experiments support this conjecture.

5. ILLUSTRATIVE EXAMPLE

Consider a single-input single-output LU system (1) with two-dimensional state, parameters

$$A = \begin{bmatrix} 1 & 0.5 \\ -0.5 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \quad C = \mathbf{1}'_{(2)}, \quad D = 0_{(1,1)}, \quad (14)$$

and noise half-widths (2)

$$x_r = 0.1 \times \mathbf{1}_{(2)}, \quad y_r = 0.1. \quad (15)$$

The system was driven by white zero-mean uniform noise with half-width 0.5 and 120 data samples were recorded.

The on-line swapping-based joint parameter and state estimation (Section 4.1) was used with window length $\partial = 5$, and prior distribution (5) restricted by the following bounds: (i) on individual entries of Θ , the bounds were set 30% above

and below the actual simulated value, with the exception of $A_{2,2} = 0$ which was set to $\bar{A}_{2,2} = 0.3$, and $\underline{A}_{2,2} = -0.3$; (ii) upper bounds on half-widths are set to $x_r = y_r = 1$, and are automatically extended when LP fails, see remark in Section 3.2; and (iii) bounds ± 5 on all entries of the window, $x^{t-\partial:t}$.

The results of estimation are displayed in Figures 1–3 via trajectories of the simulated states and their estimates (Fig. 1), estimates of the matrix A (Fig. 2), and of the estimates of half-width x_r , y_r (Fig. 3). The presented experiments serve for illustration only. Our current experience can be summarized as follows:

- Individual state or parameter estimation (Section 3) works very well. Occasionally, it is necessary to increase upper bounds on noise half-widths to achieve tractability.
- Performance of the joint parameter and state estimation strongly depends on the quality of initial estimates.
- The quality of state estimates may outperform the quality of parameter estimates (or vice versa) when estimated jointly. The example demonstrates this effect.
- The use of a finite window serves as forgetting, hence no convergence of the parameter estimates is to be expected.
- The chosen memory length $\partial = 5$ seems to have an optimal value: the estimation performance deteriorates when the used value deviates from it in both directions.

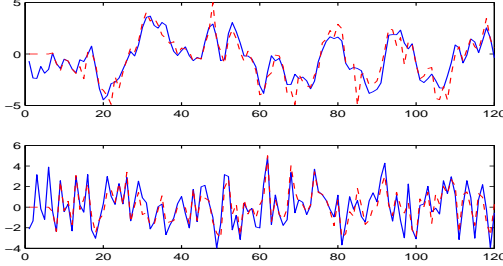


Fig. 1. MAP state estimates (dashed line) and simulated values (full line).

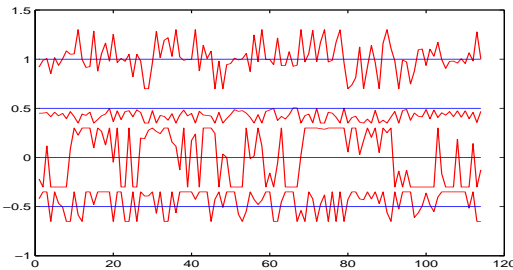


Fig. 2. MAP estimates of the entries of the matrix A (dashed line) and simulated values (full line)

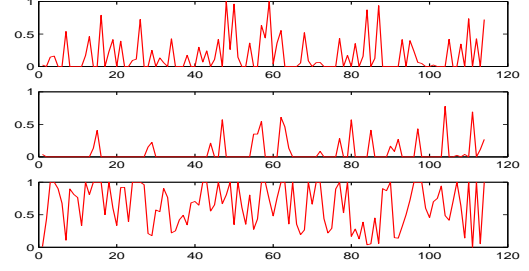


Fig. 3. MAP estimates of the half-widths x_r , (top and middle) and y_r (bottom), true values are $x_r = 0.1 \times \mathbf{1}_{(2)}$ and $y_r = 0.1$

6. CONCLUDING REMARKS

The proposed approach opens a way for on-line parameter and state estimation for a class of non-uniform distributions with restricted support as well as for Bayesian filtering of non-linear systems. The directly feasible cases are those in which linear programming is replaced by convex one. Moreover, the outer approximation of the support by ellipsoids or by union of boxes may serve as a good preliminary step for an efficient sampling-based estimation.

The main current contributions, we see, are: (i) introduction of LP as a complement to usual least-square type paradigm; (ii) feasible care about hard bounds of estimated quantities (iii) joint estimation of parameters, state, and noise bounds; (iv) parameter tracking via windowing the joint estimation; (v) feasibility of large-scale problems.

7. REFERENCES

- [1] M. Milanese and G. Belforte, “Estimation theory and uncertainty intervals evaluation in presence of unknown but bounded errors...,” *IEEE Transactions on Automatic Control*, vol. 27, no. 2, pp. 408–414, 1982.
- [2] B. Polyak, S. Nazin, C. Durieu, and E. Walter, “Ellipsoidal parameter or state estimation under model uncertainty,” *Automatica*, vol. 40, no. 7, pp. 1171–1179, 2004.
- [3] A. Jazwinski, *Stochastic Processes and Filtering Theory*. New York: Academic Press, 1970.
- [4] V. Peterka, “Bayesian system identification,” in *Trends and Progress in System Identification* (P. Eykhoff, ed.), pp. 239–304, Oxford: Pergamon Press, 1981.
- [5] S. Qin, W. Lin, and L. Ljung, “A novel subspace identification approach with enforced causal models,” *Automatica*, vol. 41, pp. 2043–2053, 2005.
- [6] P. Li, R. Goodall, and V. Kadiramanathan, “Estimation of parameters in a linear state space model using a Rao-Blackwellised particle filter,” *IEE Proceedings-control theory and applications*, vol. 151, pp. 728–738, 2004.