EXAMPLES OF STATE AND PARAMETER ESTIMATION FOR LINEAR MODEL WITH UNIFORM INNOVATIONS

Lenka Pavelková

Department of Adaptive Systems
Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Prague, CZECH REPUBLIC
E-mail: pavelkov@utia.cas.cz

Keywords: State space model, uniform distribution, parameter estimation, filtration

1. INTRODUCTION

State estimation is an important subtask of a range decision making problems. Kalman filtering (KF) (Jazwinski, 1970) is the first-option method for its addressing. However, still there is no well-established methodology of selecting innovation covariances. Also, it is difficult to combine KF with hard restrictions on state ranges. Both these drawbacks can be avoided by assuming that the model innovations are uniform.

In this contribution, state-space model with uniformly distributed innovations is introduced and the Bayesian state estimation proposed, (Peterka, 1981). This extends parameter estimation of the controlled autoregressive model treated in (Kárný and Pavelková, 2005 - submitted). Similarly as in the latter case, the off-line evaluation of the maximum a posteriori probability (MAP) estimate of unknowns in the linear state-space model with uniform innovations reduces to linear programming (LP). The solution provides either estimates of the noise boundary and parameters or of the noise boundary and states.

The on-line estimation is obtained by applying LP on the sliding window, i.e., by considering only the fixed amount, say $0 < \partial$, of the newest last data and states items.

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. Simulation studies help to get an insight on the potential and restrictions of these heuristic method. This contribution shares the experimentally gained experience with both these solutions of the joint state and parameter estimation.

2. MODEL DESCRIPTION

We consider the standard linear state-space model

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

known in connection with from Kalman filtering theory. In it, x, u, y are unobserved state, known input and observed output of the system, respectively. They are real column vectors. The subscript $t \in \{0, 1, 2, \ldots\}$ labels discrete time. The involved time-invariant matrices A, B, C, D have appropriate dimensions. The model parameters A, B, C, D are collected into parameters Θ.

Unlike in the KF case, the distributions of vector innovations ${}^{x}e_t$ and ${}^{y}e_t$ are assumed to be uniform

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

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

Equations (1) together with the assumptions (2) define the linear uniform state-space model (LU).

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 $u^{1:t} \equiv [u_1', \ldots, u_t']'$ meets natural conditions of control (Peterka, 1981). They formalize assumption 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

$$f\left(d^{1:t}, x^{1:t} \middle| x_0,\, {}^{x}r,\, {}^{y}r, \Theta\right) \propto \prod_{i=1}^{n} {}^{x}r_i^{-t} \prod_{j=1}^{m} {}^{y}r_j^{-t} \chi(\mathcal{S}). \tag{3}$$

$\chi(\mathcal{S})$ 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,

$$\begin{aligned} -\,{}^{x}r &\leq x_\tau - Ax_{\tau-1} - Bu_\tau \leq {}^{x}r \\ -\,{}^{y}r &\leq y_\tau - Cx_\tau - Du_\tau \leq {}^{y}r, \end{aligned} \tag{4}$$

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

$$\mathcal{S}_0 = \left\{\underline{x}_0 \leq x_0 \leq \overline{x}_0, \quad 0 < {}^{x}r \leq {}^{x}\overline{r}, \quad 0 < {}^{y}r \leq {}^{y}\overline{r}\right\}. \tag{5}$$

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

For unknown Θ, the uniform prior pdf $f\left(x_0,\, {}^{x}r,\, {}^{y}r, \Theta\right)$ is chosen on the set (5) extended by conditions $\underline{\Theta} \leq \Theta \leq \overline{\Theta}$.

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*.

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.

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

where known matrices and vectors $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\underline{X}$, $\overline{X}$ 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:

$$X = \begin{bmatrix} x^{t:0} \\ {}^{x}r \\ {}^{y}r \end{bmatrix}. \tag{7}$$

Where $x^{t:0} \equiv [x_t', x_{t-1}', \ldots, u_0']'$ The matrices $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\underline{X}$, $\overline{X}$ 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.

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

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, $\mathcal{L}_{col}(M) \equiv [\mathbf{0}_{(\alpha,col)}, M]$.

$\mathrm{col}(M)$ 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:

$$
\begin{aligned}
\mathcal{C}' &\equiv [\mathbf{0}'_{((t+1)\mathring{x},1)}, \mathbf{1}'_{(\mathring{x}+\mathring{y})}], && (8)\\
\mathcal{A} &= \begin{bmatrix} \mathcal{A}11 & \mathcal{A}12 \\ \mathcal{A}21 & \mathcal{A}22 \end{bmatrix}, \quad \mathcal{B} = \begin{bmatrix} \mathcal{B}1' & \mathcal{B}2' \end{bmatrix}', \text{ with} \\
\mathcal{A}11 &= \mathcal{R}_{\mathring{x}}(I_{(t)} \otimes K \otimes I_{(\mathring{x})}) - \mathcal{L}_{\mathring{x}}(I_{(t)} \otimes K \otimes A), && (9)\\
\mathcal{A}12 &= -\mathbf{1}_{(2t)} \otimes \mathcal{R}_{\mathring{y}}(I_{(\mathring{x})}), \\
\mathcal{A}21 &= \mathcal{R}_{\mathring{x}}(I_{(t)} \otimes K \otimes C), \\
\mathcal{A}22 &= -\mathbf{1}_{(2t)} \otimes \mathcal{L}_{\mathring{x}}(I_{(\mathring{y})}), \\
\mathcal{B}1 &= \left[I_{(t)} \otimes K \otimes B\right] u^{t:1}, \\
\mathcal{B}2 &= -\left[I_{(t)} \otimes K \otimes D\right] u^{t:1} + \left[I_{(t)} \otimes K \otimes I_{(\mathring{y})}\right] y^{t:1}.
\end{aligned}
$$

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

$$
\underline{X} = \begin{bmatrix} -\infty \times \mathbf{1}_{(2t\mathring{x},1)} \\ \underline{x}_0 \\ \mathbf{0}_{(\mathring{x},1)} \\ \mathbf{0}_{(\mathring{y},1)} \end{bmatrix}, \quad \overline{X} = \begin{bmatrix} \infty \times \mathbf{1}_{(2t\mathring{x},1)} \\ \overline{x}_0 \\ {}^{x}\overline{r} \\ {}^{y}\overline{r} \end{bmatrix}. \qquad (10)
$$

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:

$$
X \equiv [\mathrm{col}(A)', \mathrm{col}(B)', \mathrm{col}(C)', \mathrm{col}(D)', {}^{x}r', {}^{y}r']'. \qquad (11)
$$

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

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

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 (Milanese and Belforte, 1982; Polyak *et al.*, 2004) 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 $x^{t:t-\partial} \equiv [x_t', \ldots, x_{t-\partial}']'$ on the intersection of sets $\mathcal{S}$ and $\mathcal{S}_0$ considered for $\tau = t-\partial, \ldots, t$. This approximates the limited-memory filter of Jazwinski (Jazwinski, 1970) 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 $x^{t:t-\partial}$ using technique from Section 3.3, with parameters Θ fixed at their last point estimates. The resulting estimates of states, $\hat{x}^{t:t-\partial}$ 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 Note - it is practically important that the estimates of the noise bounds can be very inaccurate.

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:

$$Ax_\tau \approx \hat{A}_t x_\tau + A\hat{x}_{\tau|t} - \hat{A}_t\hat{x}_{\tau|t}, \tau \in \{t-\partial, t+1-\partial \ldots, t\}. \tag{13}$$

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 be transformed in the standard form of LP (6). The exact assignments are omitted for brevity. The resulting algorithms has two principal distinctions from extended KF. First 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). Second 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.

5. ILLUSTRATIVE EXAMPLE

Consider a single-input single-output LU system (1) with two-dimensional state. The model parameters are

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

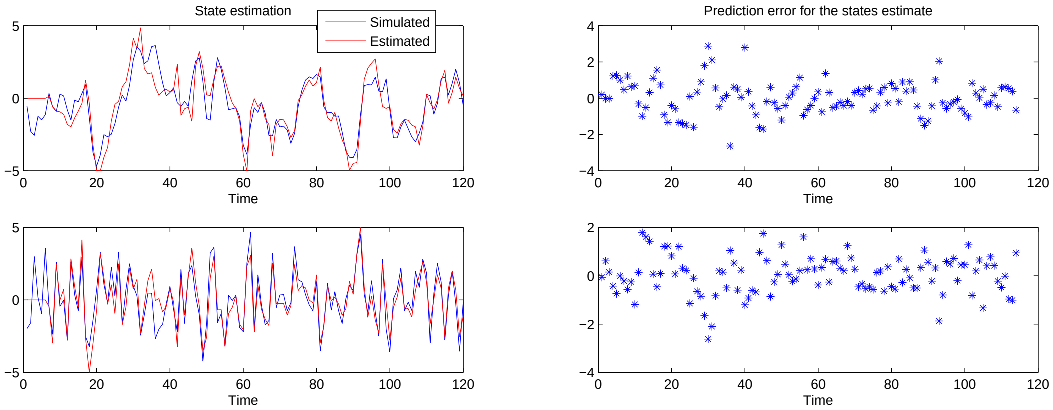

Fig. 1: Estimated state and estimate error

and noise half-widths (2)

$$^{x}r = 0.3 \times \mathbf{1}_{(2)}, \ \ ^{y}r = 0.3. \tag{15}$$

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) and expansion-based joint parameter and state estimation (Section 4.2) were 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 $\overline{A}_{2,2} = 0.3$, and $\underline{A}_{2,2} = -0.3$; (ii) upper bounds on half-widths are set to $^{x}\overline{r} = \, ^{y}\overline{r} = 1$, and are automatically extended when LP fails, see note in Section 3.2; and (iii) bounds ± 5 on all entries of the window, $x^{t-\partial:t}$

The results of the swapping-based estimation are displayed on Figures 1–3. Trajectories of the simulated and estimated states and the estimation error are on Fig. 1. The simulated and predicted output and prediction error are on Fig. 2. The estimates of the matrix A and of the estimates of half-width ^{x}r, ^{y}r are on 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 well,
- quality of the joint swaped-based estimation depends strongly on the quality of initial estimates
- joint expansion-based estimation give good result for output prediction
- finite window serves as forgetting hence no convergence of parameters is to be expected
- window length influences the estimation quality
- the quality of state estimates may outperform the quality of parameter estimates (or vice versa) when estimated jointly

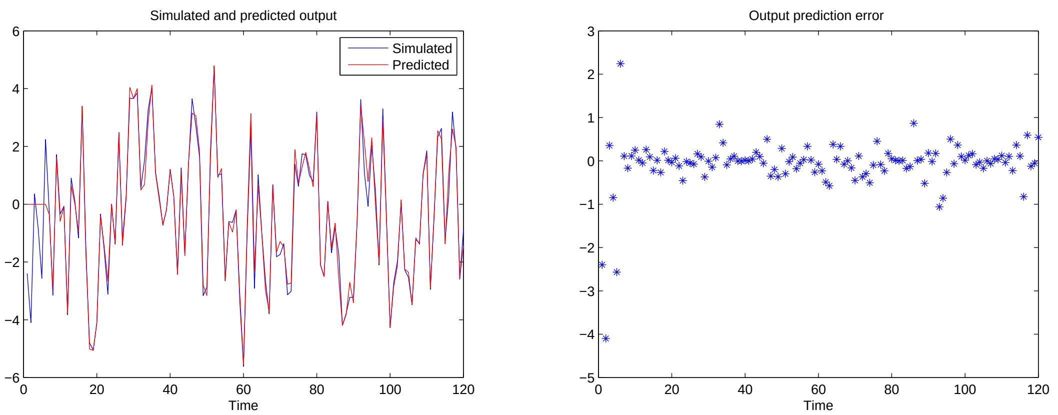

Fig. 2: Predicted output and prediction error

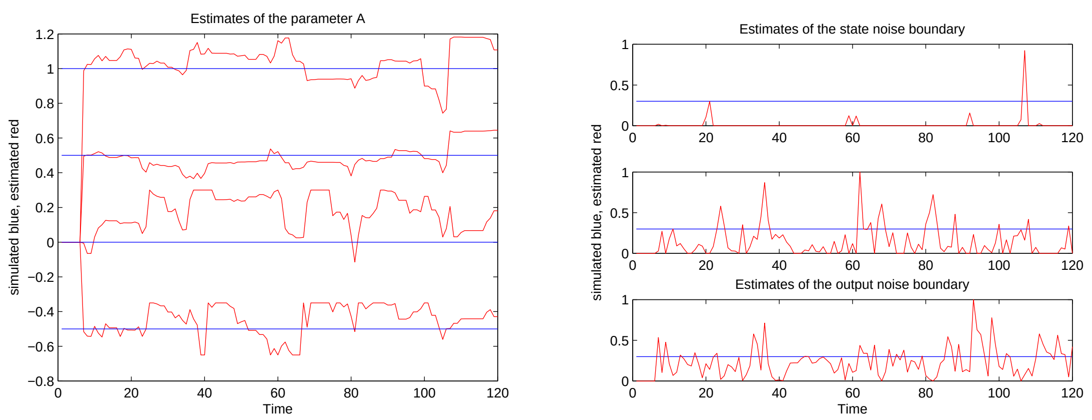

Fig. 3: Parameter A estimation and noise boundary estimation

6. CONCLUSIONS

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 main current contributions include feasible care about hard bounds of estimated quantities; joint estimation of parameters, state, and noise bounds; parameter tracking via windowing the joint estimation.

The current effort aims to improve the quality of the expansion based joint expansion.

ACKNOWLEDGEMENTS

This work was supported by Center DAR, MŠMT 1M0572, project Bayes MŠMT 2C06001 and project BADDYR AVČR 1ET 100 750 401.

REFERENCES

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

Kárný, M. and L. Pavelková (2005 - submitted). Projection-based Bayesian Recursive Estimation of ARX Model with Uniform Innovations. *Systems & Control Letters*.

Milanese, M. and G. Belforte (1982). Estimation theory and uncertainty intervals evaluation in presence of unknown but bounded errors linear families of models and estimators. *IEEE Transactions on Automatic Control* **27**(2), 408—-414.

Peterka, V. (1981). Bayesian system identification. In: *Trends and Progress in System Identification* (P. Eykhoff, Ed.). pp. 239–304. Pergamon Press. Oxford.

Polyak, B.T., S.A. Nazin, C. Durieu and E. Walter (2004). Ellipsoidal parameter or state estimation under model uncertainty. *Automatica* **40**(7), 1171–1179.