Preprints of the 16th IFAC Symposium on System Identification
The International Federation of Automatic Control
Brussels, Belgium. July 11-13, 2012

Approximate Bayesian Recursive Estimation of Linear Model with Uniform Noise ⋆

Lenka Pavelková * **Miroslav Kárný** *

* *Institute of Information Theory and Automation, Prague, Czech Republic (Tel: +420 2 6605 2337; e-mail: pavelkov@utia.cas.cz).*

Abstract: Recursive estimation forms core of adaptive prediction and control. Dynamic exponential family is the only but narrow class of parametric models that allows exact Bayesian estimation. The paper provides an approximate estimation of important autoregressive model with exogenous variables (ARX) and *uniform* noise. This model reflects well physical nature of modelled system: majority of signals, noise and estimated parameters are bounded. Unlike former solutions, the paper proposes an algorithm that provides a full (approximate) posterior probability density function (pdf) of unknown parameters. Behaviour of the designed algorithm is illustrated by simulations.

Keywords: Parameter estimation; autoregressive models; bounded noise; probabilistic models; model approximation; recursive estimation.

1. INTRODUCTION

Adaptive systems, Astrom and Wittenmark [1989], computer intensive single-pass data processing, Hand et al. [2001], and various practical applications, Alaei et al. [2010], strongly rely on recursive estimation. Its exact version is rarely feasible, Daum [2005], and either general or tailored approximation techniques are developed. This paper belongs to the second group and proposes recursive estimation of parameters determining autoregressive model with exogenous variables (ARX) and *uniform* noise. The posterior pdf has a simple form but complexity of its convex parameter-dependent support quickly grows with the number of processed data. It can be limited by windowing them but even in this case maximum a posteriori estimate is at most evaluated using linear programming, Kárný and Pavelková [2007]. Circumscription of the support by ellipsoids provides a richer description of the posterior pdf, Polyak el al. [2004]. This approximation can be rather poor in practically important transient estimation period. This observation led to proposition to evolve circumscribing boxes Bemporad et al. [2004], which, however, can suffer the same problem. This brief paper proposes approximation by a regular polytope for which evaluation characteristics of the (approximate) posterior pdf is computationally cheap. Theoretical analysis of its asymptotic properties is yet incomplete so that its behaviour is illustrated by simulation.

Section 2 formalises the problem. Section 3 provides algorithmic solution. Illustrating example is in Section 4. Section 5 contains closing comments.

2. ADDRESSED PROBLEM

The considered parametric model of the system is a probability density function (pdf) with the bounded support delimited by the indicator function χ_{y_t}(set of y_t)

$$\mathsf{f}_{y_t}(\Theta) = 0.5\Theta_n \chi_{y_t}(-1 \leq \Psi_t'\Theta \leq 1) \tag{1}$$

describing a real system output y_t in discrete time $t \in \{1, 2, \ldots\}$ depending on the past measured outputs $y_1, \ldots, y_{t-1}$, on the past and current exogenous variables $x_1, \ldots, x_t$ and on unknown parameter $\Theta \in \Theta^\star$ [1] which is n-dimensional vector. The support of $\mathsf{f}_{y_t}(\Theta)$ is a polytope in n-dimensional real space. Ψ_t is n-dimensional data vector $\Psi_t' = [\psi_t', y_t]$, where $'$ denotes transposition and ψ_t is $n-1$ dimensional regression vector constructed in a known, recursively implementable, way from $y_1, \ldots, y_{t-1}$, $x_1, \ldots, x_t$ and known initial condition Ψ_0. Positive parameter entry $\Theta_n = 1/r$ is inversion of the half-width of the pdf support r delimited by the set indicator function χ_{y_t}. The initial $n-1$ entries of the vector parameter $\Theta' = [-\theta'\Theta_n, \Theta_n]$ are regression coefficients θ normalised by Θ_n. Note that an equivalent description of the model is

$$y_t = \psi_t'\theta + e_t \tag{2}$$

with e_t uniformly distributed on $[-r, r]$, i.e., $e_t \sim \mathcal{U}(-r, r)$.

Under natural conditions of control, Peterka [1981], stating that parameters are unknown to x_t generator, Bayes rule, Bernardo and Smith [1994], provides the posterior pdf $\mathsf{f}_\Theta(V_t, \nu_t, L_t, U_t, m_t)$ of Θ conditioned on data up to and including time t which has the form

$$\mathsf{f}_\Theta(V_t, \nu_t, L_t, U_t, m_t) = \frac{\Theta_n^{\nu_t} \chi_\Theta\left(L_t \leq V_t\Theta \leq U_t\right)}{\mathsf{J}(V_t, \nu_t, L_t, U_t, m_t)} \tag{3}$$

where V_t is a (m_t, n) matrix constructed from Ψ vectors, see below, ν_t is a scalar corresponding to degrees of free-

⋆ This work was partially supported by the Research Centre DAR - MŠMT 1M0572 and project TA01030123.

[1] $\Theta^\star$ means a set of Θ

dom, L_t and U_t are lower and upper bounds, respectively, they are vectors of the length m_t.

It holds for a conjugate prior $\mathsf{f}_\Theta(V_0, \nu_0, L_0, U_0, m_0)$ determined by the optional (m_0, n) matrix V_0, ν_0 and m_0-vectors L_0 and U_0. They have to guarantee finiteness of the normalisation integral

$$\mathsf{J}(V_t, \nu_t, L_t, U_t, m_t) = \int_{\Theta^\star} \Theta_n^{\nu_t} \chi_\Theta \left(L_t \le V_t\Theta \le U_t\right) \mathrm{d}\Theta \tag{4}$$

The statistics V_t, ν_t, L_t, U_t, m_t update "recursively"

$$V_t = \begin{bmatrix} V_{t-1} \\ \Psi_t' \end{bmatrix}, \; L_t = \begin{bmatrix} L_{t-1} \\ -1 \end{bmatrix}, \; \nu_t = \nu_{t-1} + 1, \tag{5}$$
$$U_t = \begin{bmatrix} U_{t-1} \\ 1 \end{bmatrix}, \; m_t = m_{t-1} + 1.$$

The quote marks at the word "recursively" stress that this "recursion" cannot run permanently – with increasing t the dimension m_t grows permanently together with the complexity of the support of the posterior pdf. Thus, an approximate posterior pdf $\hat{\mathsf{f}}_\Theta$ is to be constructed. The pdf $\hat{\mathsf{f}}_\Theta$ is a projection of $\mathsf{f}_\Theta(V_t, \nu_t, L_t, U_t, m_t)$ (3) on a properly selected set $\hat{\mathsf{f}}^\star_\Theta$ of feasible pdfs. In Bernardo [1979], it was shown that the pdf ${}^{O}\hat{\mathsf{f}}_\Theta \in \hat{\mathsf{f}}^\star_\Theta$ approximating *optimally* the exact pdf f_Θ is to be a minimiser of the Kullback-Leibler divergence $\mathsf{D}(\mathsf{f}_\Theta||\hat{\mathsf{f}}_\Theta)$, Kullback and Leibler [1951],

$$
\begin{aligned}
{}^{O}\hat{\mathsf{f}}_\Theta \in \mathrm{Arg}\min_{\hat{\mathsf{f}} \in \hat{\mathsf{f}}^\star_\Theta} \mathsf{D}(\mathsf{f}_\Theta||\hat{\mathsf{f}}_\Theta) = \mathrm{Arg}\min_{\hat{\mathsf{f}} \in \hat{\mathsf{f}}^\star_\Theta} \int_{\Theta^\star} \mathsf{f}_\Theta(V_t, \nu_t, L_t, U_t, m_t) \\
\times \ln\left(\frac{\mathsf{f}_\Theta(V_t, \nu_t, L_t, U_t, m_t)}{\hat{\mathsf{f}}_\Theta}\right) \mathrm{d}\Theta.
\end{aligned} \tag{6}
$$

The key obstacle is that the minimiser is to be found *without* storing the approximated pdf. This is rarely possible Kulhavý [1990a], Kulhavý [1993b] but it can be approximately done in the studied case if we select the following set of approximating pdfs

$$\hat{\mathsf{f}}^\star_\Theta = \left\{ \hat{\mathsf{f}}_\Theta(v, \nu, l, u, m) = \frac{\Theta_n^{\nu} \chi_\Theta \left(l \le v\Theta \le u\right)}{\mathsf{J}(v, \nu, l, u, m)} \right. \tag{7}$$
$$\left. \mathsf{J}(v, \nu, l, u, m) \text{ given by (4) and a fixed } m \text{ and } \nu = \nu_t \right\}.$$

Theorem 1. (Almost Recursive Feasibility). The optimal approximate pdf minimising (6) over (7) is given by $\nu = \nu_t$, chosen m and optimum triple

$${}^{O}(v, l, u) \in \mathrm{Arg} \min_{(v,l,u) \in (v,l,u)^\star} \mathsf{J}(v, \nu_t, l, u, m), \tag{8}$$

where the set $(v, l, u)^\star$ guarantees that the support of $\hat{\mathsf{f}}_\Theta(v, \nu_t, l, u, m)$ circumscribes the support of $\mathsf{f}_\Theta(V_t, \nu_t, L_t, U_t, m_t)$.

Proof. Inserting the approximated and approximating pdfs into the definition, we get

$$\mathsf{D}(\mathsf{f}_\Theta||\hat{\mathsf{f}}_\Theta) = \ln(\mathsf{J}(v, \nu_t, l, u, m)) - \ln(\mathsf{J}(V_t, \nu_t, L_t, U_t, m_t))$$
$$+ \int_{\Theta^\star} \mathsf{f}_\Theta \ln\left(\frac{\chi_\Theta(\text{support of } \mathsf{f}_\Theta)}{\chi_\Theta(\text{support of } \hat{\mathsf{f}}_\Theta)}\right) \mathrm{d}\Theta.$$

The first term is increasing function of $\mathsf{J}(v, \nu_t, l, u, m)$. The second term does not depend on the optimised triple (v, l, u) and the last term is zero for all optimised triples guaranteeing that the support of $\hat{\mathsf{f}}_\Theta$ circumscribes the support of $\mathsf{f}_\Theta(V_t, \nu_t, L_t, U_t, m_t)$. □

This guides us directly to the recursive construction of the approximate pdf. Knowing that

$$S_{t-1} = \{\Theta :\, \Theta_n > 0,\, L_{t-1} \le V_{t-1}\Theta \le U_{t-1}\} \tag{9}$$
$$\subset \{\Theta :\, \Theta_n > 0,\, l_{t-1} \le v_{t-1}\Theta \le u_{t-1}\} = \hat{S}_{t-1},$$

we know that, see (5),

$$S_t = \{\Theta : \Theta_n > 0,\, L_t \le V_t\Theta \le U_t\} \tag{10}$$
$$\subset \left\{\Theta : \Theta_n > 0 \begin{bmatrix} l_{t-1} \\ -1 \end{bmatrix} \le \begin{bmatrix} v_{t-1} \\ \Psi_t' \end{bmatrix} \Theta \le \begin{bmatrix} u_{t-1} \\ 1 \end{bmatrix} \right\} = \tilde{S}_t.$$

Thus, triples (v, l, u) guaranteeing that the set

$$\hat{S}_t = \{\Theta :\, \Theta_n > 0,\, l \le v\Theta \le u\} \tag{11}$$

includes the set $\tilde{S}_t$ on the right hand side of (10) guarantee the circumscribing. By minimising over them the function $\mathsf{J}(v, \nu_t, l, u, m)$ (8), we get suboptimal approximating pdf. Its sub-optimality follows from the fact that some optimality candidates were dropped.

3. ALGORITHMIC SOLUTION

Here, we propose an algorithm for the recursive updating of the matrix v. We restrict ourselves to square matrices v, i.e., $m = n$. The meaningful options with $m > n$ are discarded pragmatically: analytical evaluation of the optimised J (8) is known to be computationally difficult, Gelfand and Dey [1994]. The v_{t-1} is supposed to be an upper triangular matrix with unit diagonal and the proposed construction preserves this form for v_t.

Before providing a simple construction of $\hat{S}_t$ (11) covering $\tilde{S}_t$ in (10), we evaluate the normalisation factor (4) and moments of Θ. It can be done analytically. The splitting with the separated last column

$$v = \begin{bmatrix} v_\psi & v_y \\ 0 & 1 \end{bmatrix} \tag{12}$$

helps us to express them.

Theorem 2. (Normalisation and Expectation of Posterior Pdf). The normalisation factor (4) of the pdf (7) has the form

$$\mathsf{J}(v, \nu, l, u, n) = \prod_{i=1}^{n-1} (u_i - l_i) \frac{u_n^{\nu+1} - (\max(l_n, 0))^{\nu+1}}{\nu + 1} \tag{13}$$

and the expected values are

$$\begin{bmatrix} \hat{\theta} \\ \hat{r} \end{bmatrix} = \mathsf{E}\left(\begin{bmatrix} \theta \\ r \end{bmatrix} \middle| v, \nu, l, u, n \right) \tag{14}$$
$$\hat{\theta} = \hat{r} v_\psi^{-1} [\mathsf{I}_{n-1}, 0] \frac{u + l}{2} - v_\psi^{-1} v_y,$$
$$[\mathsf{I}_{n-1}, 0] \text{ denote } n-1 \text{ rows of unit } n \text{ matrix}$$
$$\hat{r} = \frac{\nu + 1}{\nu} \frac{1 - \gamma^\nu}{1 - \gamma^{\nu+1}} u_n^{-1}, \; \gamma = \frac{\max(l_n, 0)}{u_n}.$$

Proof. The straightforward integration uses Fubini theorem and the substitution $x = v\Theta$, which has unit Jacobian and leaves $x_n = \Theta_n$,

$$\mathsf{J}(v, \nu, l, u, n) = \int_{\Theta^\star} \Theta_n^{\nu} \chi(l \le v\Theta \le u)\, \mathrm{d}\Theta$$
$$= \prod_{i=1}^{n-1} (u_i - l_i) \frac{u_n^{\nu+1} - (\max(l_n, 0))^{\nu+1}}{\nu + 1}.$$

Introducing $\gamma = \frac{\max(l_n,0)}{u_n} < 1$, we get

$$\hat{r} = \frac{\mathsf{J}(v,\nu-1,l,u,n)}{\mathsf{J}(v,\nu,l,u,n)} = \frac{\nu+1}{\nu}\frac{1-\gamma^{\nu}}{1-\gamma^{\nu+1}}u_n^{-1}.$$

The expectation of $\theta = \Theta/\Theta_n$ is obtained via the substitution as above and by exploiting that $v^{-1} = \begin{bmatrix} v_\psi^{-1} & -v_\psi^{-1}v_y \\ 0 & 1 \end{bmatrix}$ and using the first $n-1$ rows of n-dimensional unit matrix $[\mathsf{I}_{n-1}, 0]$

$$\begin{aligned}\hat{\theta} &= \frac{\int_{\Theta^\star}[\mathsf{I}_{n-1},0]\Theta\Theta_n^{\nu-1}\chi(-l \le v\Theta \le u)\,\mathrm{d}\Theta}{\mathsf{J}(v,\nu,l,u,n)} \\ &= \frac{\int_{x^\star}\left[\, v_\psi^{-1}, -v_\psi^{-1}v_y \,\right] x x_n^{\nu-1}\chi(-l \le x \le u)\,\mathrm{d}x}{\mathsf{J}(v,\nu,l,u,n)} \\ &= \hat{r}v_\psi^{-1}[\mathsf{I}_{n-1},0]\frac{u+l}{2} - v_\psi^{-1}v_y.\end{aligned}$$

□

The result provides immediately point output prediction as $\hat{\theta}'\psi$. Higher moments of parameters can be evaluated similarly as in the theorem and serve for evaluation of moments of the predicted output.

The form of the normalisation factor (13) confirms (intuitively obvious) optimal choice of circumscribing set as the set having small differences of individual lower and upper bounds.

The algorithm providing $\hat{S}_t$ (11) is based on the following simple theorem.

Theorem 3. (Orthogonal Reduction of Weight Entry). Let us have a pair of inequalities for n-dimensional real vector z, given by scalar lower $\tilde{L},\tilde{l}$ and upper $\tilde{U},\tilde{u}$ bounds and weighting n-vectors $\tilde{W}$, $\tilde{w}$, respectively

$$\{z : \tilde{L} \le \tilde{W}'z \le \tilde{U},\ \tilde{l} \le \tilde{w}'z \le \tilde{u}\}, \tag{15}$$

where $\tilde{W}_1 = 1$ and $\tilde{w}_1 \ge 0$. Let us define

$$a = \frac{1}{1+\tilde{w}_1^2},\ b = \tilde{w}_1 a = \frac{\tilde{w}_1}{1+\tilde{w}_1^2} \ge 0. \tag{16}$$

Then, the set (15) is included in the set

$$\begin{aligned}&\{z : \hat{L} \le \hat{W}'z \le \hat{U},\ \hat{l} \le \hat{w}'z \le \hat{u}\} && (17)\\ &\hat{W} = a\tilde{W} + b\tilde{w},\ \hat{L} = a\tilde{L} + b\tilde{l},\ \hat{U} = a\tilde{U} + b\tilde{u} &&\\ &\hat{w} = -b\tilde{W} + a\tilde{w},\ \hat{l} = -b\tilde{U} + a\tilde{l},\ \hat{u} = -b\tilde{L} + a\tilde{u},\ \Rightarrow &&\\ &\hat{W}_1 = 1 \ \text{ and } \hat{w}_1 = 0. && (18)\end{aligned}$$

Proof. For $w_1 = 0$, both sets are identical. Let us consider $w_1 > 0$. Then, the parameters $\hat{W}$, $\hat{L}$, $\hat{U}$ determine inequality that arisen as linear combination of the original inequalities with positive weights a, b. The parameters $\hat{w}$, $\hat{l}$, $\hat{u}$ determine inequality that arisen as linear combination of the original inequalities with the negative weight $-b$ and positive weight a. The weights were chosen so that (18) holds. □

With this simple result, it suffices run a cycle over all rows of v_{t-1}, l_{t-1}, u_{t-1} in the role $\tilde{W}$, $\tilde{L}$, $\tilde{U}$ whereas $\tilde{w}$, $\tilde{l}$, $\tilde{u}$ starts with the values $\Psi_t, -1, 1$, see (10) and is gradually zeroed, see (18). Non-negativity of $\tilde{w}_1$ is simply reached by multiplying the corresponding inequalities by signum of w_1. This procedure corresponds with gradual orthogonal (not orthonormal) rotations known in connection with factorised least-squares, Rontogiannis and Theodoridis [1998].

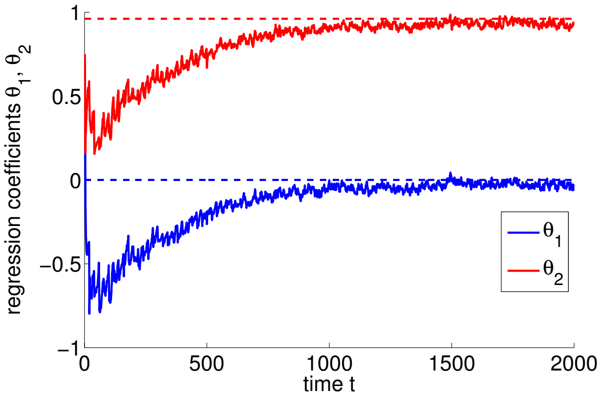

Fig. 1. Time course of regression-coefficients estimates θ_1 and θ_2 (full lines) with true values (dashed lines)

4. ILLUSTRATIVE EXAMPLE

This section numerically illustrates behaviour of the resulting algorithm on ARX model (2) which has two poles ± 0.9604 and static gain approximately one. It is described by the following equation with $\psi = [y_{t-1}, y_{t-2}, x_t, x_{t-1}]'$, $\theta = [0,\ 0.9604,\ 1,\ -0.96]'$, $t \in t^\star = \{1, 2, \ldots, 10^5\}$

$$y_t = 0y_{t-1} + 0.9604y_{t-2} + 1x_t - 0.96x_{t-1} + e_t, \tag{19}$$

where noise terms e_t are uniformly distributed with half-width $r = 0.05$, i.e., $e_t \sim \mathcal{U}(-0.05, 0.05)$. The system is stimulated by deterministic bi-level signal.

Estimation results are in Figures 1–3, which indicate typical satisfactory behaviour of the proposed algorithm. The values of the regression coefficient $\theta_1 - \theta_4$ estimates converges to their true values. The value of the half width r estimate converges to a slightly lower value than a real one.

Courses of lower l and upper u bounds are depicted in Figures 4 and 5. The bounds $l_1 - l_3$, $u_1 - u_3$ oscillate around a certain level from the very beginning. The bounds l_4 and u_4 are growing initially and stabilize for $t \approx 2000$. The bound $l_5 = 0$ by definition while bound u_5 grows up relatively long and up to a relatively high value but it finally also stabilizes for $t \approx 20000$.

Samples of simulated data are in Figures 6 and 7. The estimation quality is confirmed by computing absolute prediction error

$$E_t = |y_t - (\hat{\theta}'\psi_t + \hat{e}_t)|$$

where $\hat{e}_t \sim \mathcal{U}(-\hat{r}, \hat{r})$. The time course of the absolute prediction error is in Figure 8. You can see that prediction error quickly decreases up to a certain minimal level.

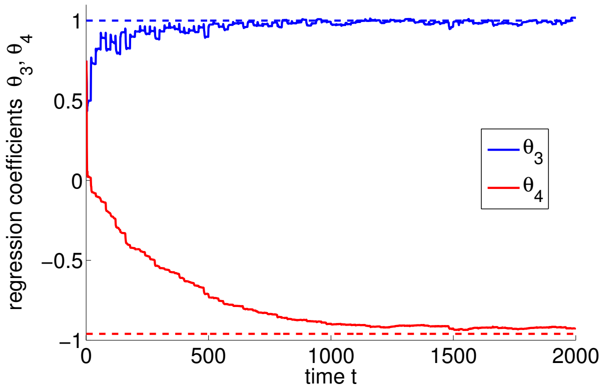

Fig. 2. Time course of regression-coefficients estimates θ_3 and θ_4 (full lines) with true values (dashed lines)

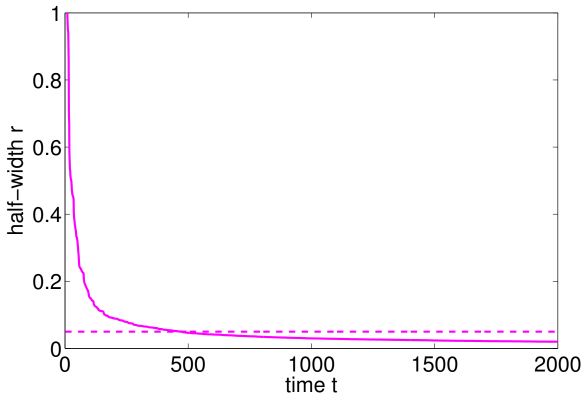

Fig. 3. Time course of half-widths estimates r (full line) with true value (dashed line)

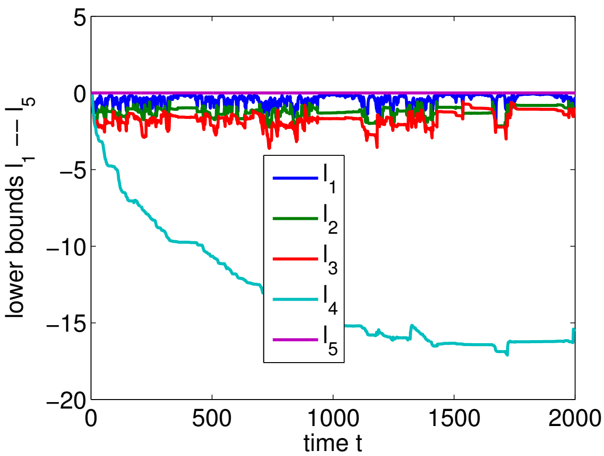

Fig. 4. Time course of lower bounds l in (11)

5. CONCLUDING REMARKS

The proposed recursive estimation of ARX model with uniform noise seems to be effective. Still there is a couple open problems and inevitable further steps. For instance:

- We conjecture that this construction guarantees quite tight circumscription but we have no definite analysis in this respect.

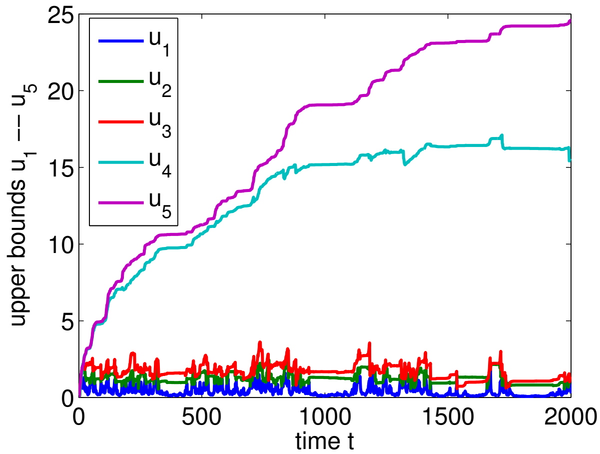

Fig. 5. Time course of upper bounds u in (11)

- The accumulation of errors caused by the use approximate prior pdf as the starting point for the application of Bayes rule is to be counteracted. In Kárný and Dedecius [2012], the problem was addressed generally and leads to a sort of forgetting technique. Its application to the considered case is desirable.
- The considered hard bound on the noise is realistic but at the same time sensitive to outlying observations. Inspection of counter-measures is inevitable.
- The available analytic description of the (approximate) posterior pdf allows Bayesian structure estimation. The corresponding efficient search within the extensive space of possible hypotheses is to be developed, Kárný and Kulhavý [1988].
- Currently, the uniformly distributed noise is considered. Further research will focus also on another distribution with restricted support, e.g. truncated Gaussian or triangular one, that are supposed to approximate the reality more precisely.

REFERENCES

H. Alaei, K. Salahshoor, and H. Alaei (2010). Model predictive control of distillation column based recursive parameter estimation method using HYSYS simulation. *Intelligent Computing and Cognitive Informatics, International Conference on*, 0: 308–311.

K. Astrom, and B. Wittenmark (1989). *Adaptive Control*. Massachusetts: Addison-Wesley.

A. Bemporad A., C. Filippi C. and F. Torrisi (2004). Inner and outer approximations of polytopes using boxes. *Computational Geometry*, 27:151–178.

A. Benveniste, M. Métivier, P. and Priouret (1990). *Adaptive Algorithms and Stochastic Approximations*. Berlin: Springer.

J. Berger (1985). *Statistical Decision Theory and Bayesian Analysis*. New York: Springer.

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

J.M. Bernardo, and A.F.M. Smith (1994). *Bayesian Theory*. Wiley.

F. Daum (2005). Nonlinear filters: beyond the Kalman filter. *Aerospace and Electronic Systems Magazine, IEEE*, 20(8): 57–69.

A.E. Gelfand and D.K. Dey (1994). Bayesian model choice - asymptotics and exact calculations. *Journal of the Royal Statistical Society Series B-Methodological*, 56(3):501–514.

D. Hand, H. Mannila, P. and Smyth (2001). *Principles of data mining*. MIT Press.

R. Horst, H. and Tuy (1996). *Global Optimization*. Springer. 727 pp.

M. Kárný and K. Dedecius (2012). Approximate Bayesian Recursive Estimation: On Approximation Errors. Technical report No. 2317. ÚTIA AV ČR, Prague 2012 http://library.utia.cas.cz/separaty/2012/AS/karny-approximate bayesian recursive estimation on approximation errors.pdf.

M. Kárný and R. Kulhavý (1988). Structure determination of regression-type models for adaptive prediction and control. In J.C. Spall, editor, *Bayesian Analysis of Time Series and Dynamic Models*. Marcel Dekker, New York. Chapter 12.

M. Kárný, and L. Pavelková (2007). Projection-based Bayesian recursive estimation of ARX model with uniform innovations. *Systems & Control Letters*, 56(9/10):646–655.

R. Kulhavý (1990a). A Bayes-closed approximation of recursive nonlinear estimation. *Int. J. Adaptive Control and Signal Processing*, 4: 271–285.

— (1990b). Recursive Bayesian estimation under memory limitations. *Kybernetika*, 26: 1–20.

— (1990c). Recursive nonlinear estimation: A geometric approach. *Automatica*, 26(3): 545–555.

R. Kulhavý (1993a). Can approximate Bayesian estimation be consistent with the ideal solution? In *Proc. of the 12th IFAC World Congress*, volume 4, 225–228. Sydney, Australia.

— (1993b). Implementation of Bayesian parameter estimation in adaptive control and signal processing. *The Statistician*, 42: 471–482.

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

H. Kushner (1967). *Stochastic stability and control*. Academic Press.

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

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

M. Rao (1987). *Measure Theory and Integration*. New York: John Wiley.

A.A. Rontogiannis and S. Theodoridis(1998). New fast QR decomposition least squares adaptive algorithms. *IEEE Transactions on Signal Processing*,46(8):2113–2121.

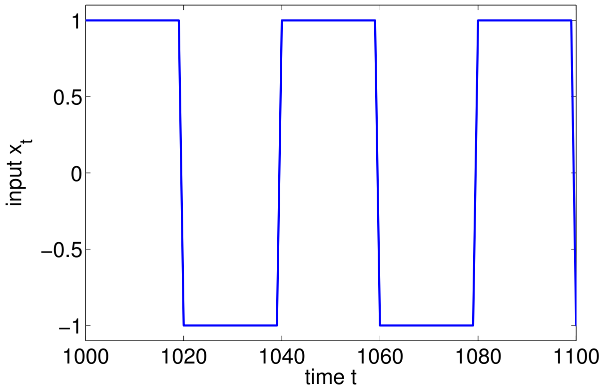

Fig. 6. Applied inputs x_t - detail

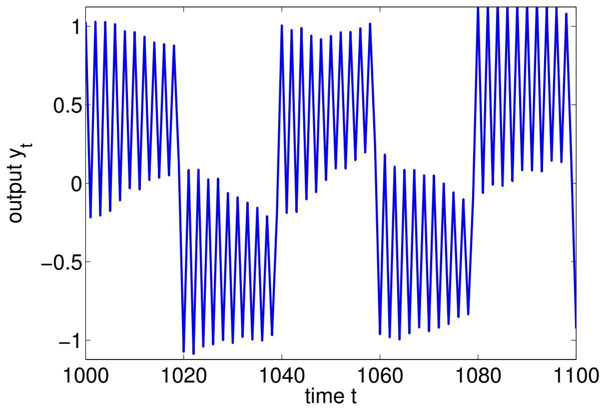

Fig. 7. Observed outputs y_t - detail

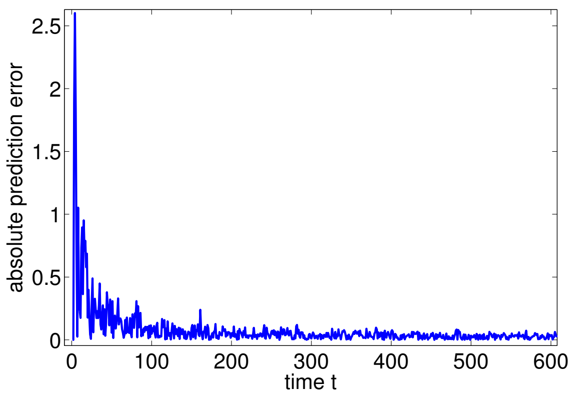

Fig. 8. Course of the absolute prediction error E_t