# ESTIMATION OF ARX MODEL WITH UNIFORM NOISE – ALGORITHMS AND EXAMPLES

**Lenka Pavelková, Miroslav Kárný**

*Institute of Information Theory and Automation, Prague*
*Czech republic*

Abstract: Autoregressive model with exogenous inputs (ARX) is a widely used black-box type model. The output noise is usually supposed to have the Gaussian distribution with zero mean value. This model is good algorithmically processed but unbounded support of the normal distribution can cause problems in some applications. Here the output noise is assumed to have uniform distribution. The posterior probability density function (pdf) of this model is described by the system of the inequalities whose number is time increasing. To obtain recursively feasible estimation of uniform ARX model the approximation of the exact pdf is required. The algorithmic solution of this problem is presented here.

Keywords: ARX model, Bayesian estimation, uniform distribution, on line parameter estimation

## 1. INTRODUCTION

Recursive estimation (Peterka, 1981) is in the heart of a range adaptive decision-making units that include predictors, advising and fault-detecting systems as well as adaptive controllers. Sufficiency of simple, often black-box, locally valid models is the major advantage adaptive systems. It makes them relatively universal and keeps costs of their implementation low. Autoregressive model with exogenous inputs (ARX) is an important representative of this model class. Its innovations, stochastic unobserved stimulus of the model, are white, zero mean and have time-invariant variance. Mostly, the innovations are assumed to be normal. It singles out least squares as the adequate estimation procedure. Light tails of the normal distribution imply that unbounded support of normal distribution can often be accepted as a reasonable approximation of reality, which is mostly bounded. In some case, however, this assumption is unrealistic, e.g. in modelling of transportation systems, or do not fit subsequent processing, for instance, robust control design. Then, techniques similar to those dealing with unknown-but-bounded equation errors are used. The paper accepts the assumption that innovations are bounded but stays within the standard, here Bayesian, estimation setup (Peterka, 1981) by assuming their uniform distribution. The posterior probability density function (pdf) is described and then approximated by a pdf of the same type but having a fixed-dimensional statistic.

## 2. NOTION AND NOTATION

$'$ denotes transposition; $\mathring{x}$ means the number of entries in a vector $x$; $I$ denotes unit matrix; the subscript in round brackets defines dimensions of the matrix; $\equiv$ is equality by definition; $X^*$ denotes a set of $X$-values; $\chi_x(x^*)$ is the indicator of a set $x^*$ at the point $x$; $f(\cdot|\cdot)$ denotes probability density function (pdf); $t$ labels discrete-time moments, $t \in t^* \equiv \{1, 2, \ldots, \}$;

$d_t = (y_t, u_t)$ is the data record at time $t$ consisting of an observed system output $y_t$ and of an optional system input $u_t$; $X(t)$ denotes the sequence $(X_1, \ldots, X_t)$, $X \in \{d, y, u\}$. Names of arguments distinguish respective pdfs. No formal distinction is made between a random variable, its realization and an argument of a pdf. Integrals used are always definite and multivariate ones. The integration domain coincides with the support of the pdf in its argument.

## 3. ESTIMATION OF ARX MODEL WITH UNIFORM NOISE

The parameterized model of the system with a single output $y_t$

$$\begin{aligned} & f(y_t|u_t, d(t-1), \Theta) = f(y_t|\psi_t, \Theta) \equiv \mathcal{U}_{y_t}(\psi'\theta, r) \\ \equiv & \frac{\chi_{y_t}(-r \le y_t - \psi_t'\theta \le r)}{2r} = \frac{\chi_{y_t}(-r \le \Psi_t[1, -\theta']' \le r)}{2r} \end{aligned} \tag{1}$$

is the pdf that describes the ARX model with uniform innovations. In it,

$\psi_t$ is a column regression vector made of past observed data $d(t-1)$ and the current system input $u_t$; often, the state in the phase form is considered, i.e., $\psi_t' \equiv [u_t', d_{t-1}', \ldots, d_{t-\partial}', 1]$ with the model order $\partial \ge 0$,

$\Psi_t' \equiv [y_t, \psi_t']$ combines the modelled output and regression vector in the single data vector; for the state in the phase form $\Psi_t' = [d_t', d_{t-1}', \ldots, d_{t-\partial}', 1]$,

$\theta$ is a column vector of regression coefficients,

$r > 0$ is a positive scalar half-width of the range of innovations $e_t \equiv \Psi_t'[1, -\theta']'$,

$\Theta \equiv (\theta, r)$ labels unknown parameters of the model,

$\mathcal{U}_y(\mu, r)$ is a uniform pdf of $y$ given by expectation $\mu$ and half-width $r > 0$,

$\chi_x(x^*)$ is an indicator function of the set $x^*$ evaluated at value $x$; it equals 1 if $x \in x^*$ and it is zero otherwise.

Under natural conditions of control (Peterka, 1981), which assume inputs conditionally independent of unknown parameters $f(u_t|d(t-1), \Theta) = f(u_t|d(t-1))$, the likelihood function assigned to the uniform ARX model has the form

$$\mathcal{L}(d(t), \Theta) \;=\; \frac{1}{r^t}\chi_r(r \ge 0)\chi_\Theta(-\mathbf{1}_t r \le W_t[-1, \theta']' \le \mathbf{1}_t r), \tag{2}$$

where $\mathbf{1}_t$ is the column vector consisting of $t$ units and $W_t' \equiv [\Psi_1, \ldots, \Psi_t]$. The matrix $W_t$ is assumed to be known, which requires knowledge of the initial regression vector $\psi_1$. For the state in the phase form, it is equivalent to the knowledge of $u_1, d_0, \ldots, d_{-\partial+1}$. The likelihood form (2) hints the reproducing prior pdf

$$f(\Theta) \propto \frac{1}{r^{\nu_0}}\chi_r(\bar{r} \ge r \ge 0)\chi_\Theta(-\mathbf{1}_{\nu_0} r \le W_0[-1, \theta']' \le \mathbf{1}_{\nu_0} r), \tag{3}$$

where $\infty > \bar{r} > 0$ is a sure upper bound on the half-width $r$ and $W_0$ can be interpreted as storage of data vectors (fictitiously) measured before the time $t = 1$ and $\nu_0$ is their number. The proper prior pdf is obtained for full-rank $W_0$ and $\nu_0 \ge \mathring{\Psi} + 1$.

For the prior pdf (3), the posterior pdf has the same form

$$
\begin{aligned}
f(\Theta|d(t)) &\propto \frac{1}{r^{\nu_t}}\chi_r(\overline{r} \geq r \geq 0)\chi_\Theta(-\mathbf{1}_{\nu_t} r \leq W_t[-1,\theta']' \leq \mathbf{1}_{\nu_t} r), \text{ with} \\
\nu_t &= \nu_{t-1}+1,\ \nu_0 \geq \mathring{\Psi}+1 \text{ is chosen a priori} \\
W_t' &= \left[W_{t-1}', \Psi_t\right],\ \text{full-rank } W_0 \text{ is chosen a priori.}
\end{aligned}
\tag{4}
$$

Applicability of the formula (4) is limited as the dimension $(\nu_t, \mathring{\Psi})$ of the matrix $W_t$ increases with the number $t$ of processed data $d(t)$, which also increases complexity of the convex support defined by them. This makes us search for an approximate solution, especially when considering the recursive estimation.

## 4. APPROXIMATE RECURSIVELY FEASIBLE ESTIMATION

The approximation needed for on line estimation exploits *Kullback-Leibler (KL) divergence* $\mathcal{D}\left(\tilde{f}||f\right)$ (Kullback and Leibler, 1951) that measures the proximity of a pair of pdfs $\tilde{f},\ f$ acting on a set $X^*$. It is defined as follows

$$
\mathcal{D}\left(\tilde{f}||f\right) \equiv \int \tilde{f}(X)\ln\left(\frac{\tilde{f}(X)}{f(X)}\right)dX. \tag{5}
$$

The KL divergence has the following properties of interest

$$
\begin{aligned}
&\mathcal{D}(\tilde{f}||f) \geq 0,\ \mathcal{D}(\tilde{f}||f) = 0 \text{ iff } \tilde{f} = f \text{ almost everywhere on } X^* \\
&\mathcal{D}(\tilde{f}||f) = \infty \text{ if the set on which } f(x)=0 \text{ and } \tilde{f}(x) > 0 \text{ has non-zero volume.}
\end{aligned}
\tag{6}
$$

To obtain approximate, recursively feasible, estimation of the uniform ARX model we have to approximate the exact posterior pdf by a pdf determined by a statistic whose finite dimension does not increase with the increasing number of data. Here, an approximation by a convex set is proposed. This provides a simple description of uncertainties suitable for robust prediction and control design.

The exact posterior pdf (4) motivates the form of the approximate pdf

$$
f(\Theta|d(t-1)) = \frac{\frac{1}{r^{\nu_{t-1}}}\chi_r(\overline{r} \geq r \geq 0)\chi_\Theta(-\mathbf{1}_{\mathring{V}} r \leq V_{t-1}[-1,\theta']' \leq \mathbf{1}_{\mathring{V}} r)}{\mathcal{I}(V_{t-1},\nu_{t-1})} \tag{7}
$$

$$
\mathcal{I}(V,\nu) \equiv \int \frac{1}{r^\nu}\chi_r(\overline{r} \geq r \geq 0)\chi_\Theta(-\mathbf{1}_{\mathring{V}} r \leq V[-1,\theta']' \leq \mathbf{1}_{\mathring{V}} r)\, d\Theta. \tag{8}
$$

These pdfs are determined by full-rank matrices $V_{t-1}$ with a fixed, finite first dimension $\mathring{V} \geq \mathring{\Psi}$. After Bayesian updating by the data vector $\Psi_t$, the (approximate) posterior pdf becomes

$$
\tilde{f}(\Theta|d(t)) = \frac{\frac{\chi_r(\overline{r} \geq r \geq 0)}{r^{\nu_{t-1}+1}}\chi_\Theta\left(-\mathbf{1}_{\mathring{V}+1} r \leq [V_{t-1}', \Psi_t]'[-1,\theta']' \leq \mathbf{1}_{\mathring{V}+1} r\right)}{\mathcal{I}\left([V_{t-1}', \Psi_t]', \nu_{t-1}+1\right)}. \tag{9}
$$

We search for the approximate posterior pdf $f(\Theta|d(t))$ in the form (7), with $t$ replacing $t-1$, which minimizes the KL divergence $\mathcal{D}(\tilde{f}(\Theta|d(t)||f(\Theta|d(t)))$. Thus, we apply the projection-based approximation of the Bayes rule (Andrýsek, 2004).

The KL divergence will be finite iff the support of $\tilde{f}$ will be included in the support of the constructed approximate pdf. This can be easily fulfilled by restricting $V_t'$ to be a sub-matrix of

$[V'_{t-1}, \Psi_t]$. We adopt this restriction. Then, the constructed pdf $f(\Theta|d(t))$ is constant function of $\theta$ on the integration domain in the formula (5) defining the KL divergence. The function optimized over the $(\mathring{V}, \mathring{\Psi})$ sub-matrices $V_t$ of the $(\mathring{V}+1, \mathring{\Psi})$ matrix $\left[V'_{t-1}, \Psi_t\right]'$ becomes $\ln[\mathcal{I}(V_t, \nu_t)]$ with $\nu_t = \nu_{t-1} + 1$. For the optimal choice of $V_t$, we have to inspect dependence of $\mathcal{I}(V, \nu)$ on $V$. The substitution $x' = [-1/r, \theta'/r]$ in (8) and a straightforward bounding from above give

$$\begin{aligned} \mathcal{I}(V,\nu) &= \int (-x_1)^{\nu-\mathring{\Psi}-1}\chi_{x_1}(x_1 \le -1/\bar{r})\chi_x(-\mathbf{1}_{\mathring{V}} \le Vx \le \mathbf{1}_{\mathring{V}})\,dx \le \frac{\mathcal{V}(V)}{\hat{r}^{\nu}} \\ \hat{r} &\equiv \arg\min_r \{\Theta:\ \bar{r} \ge r \ge 0, -r\mathbf{1}_{\mathring{V}} \le V[-1,\theta']' \le r\mathbf{1}_{\mathring{V}}\} \\ \mathcal{V}(V) &\equiv \int \chi_x(-\mathbf{1} \le Vx \le \mathbf{1})\,dx. \end{aligned} \tag{10}$$

The exact evaluation of $\mathcal{I}(V, \nu)$ is hard. Thus, instead of minimizing the KL divergence, which is its increasing function of $\mathcal{I}(V, \nu)$, we minimize the upper bound in (10) over $(\mathring{V}, \mathring{\Psi})$ sub-matrices of $[V'_{t-1}, \Psi_t]'$. The complete argument $\hat{\Theta} \equiv (\hat{\theta}, \hat{r}) \equiv (\hat{\theta}(V), \hat{r}(V)) \equiv \hat{\Theta}(V)$ of the minima (10) is found by linear programming (LP). It serves also as a point estimate of $\Theta$.

In summary, the set $V_t^*$ of candidates for $V_t$ contains $V_{t-1}$ and ${}^iV \equiv V_{t-1}, i = 1, \ldots, \mathring{V}$, with $i$th row replaced by $\Psi_t$. The best option $V_t$ minimizes the upper bound (10), i.e. $\mathcal{V}(V_t)/\hat{r}^{\nu_t}(V_t) \le \mathcal{V}(V)/\hat{r}^{\nu_t}(V), \forall V \in V_t^*$. It remains to find how the volume $\mathcal{V}(V)$ depends on $V$. It is in detail described in (Kárný and Pavelková, 2005 - submitted)

## 5. ALGORITHM

Obtained algorithm is summarized here.

### Initial mode

- Select the structure of the ARX model, i.e., structure of data vectors $\Psi_t$.
- Select the dimension $\mathring{V} \ge \mathring{\Psi}$ of the statistic $V$.
- Select lower and upper bounds on the estimated parameters and construct the matrix $V_0$.
- Collect the data up to the moment when $\Psi_t$ complement $V_t$ to its full chosen dimension $\mathring{V}$.
- Fill initial data into the first regression vector $\psi_1$, choose $\nu_0$ and set $t = 0$.

### Recursive mode

- Set $t = t + 1$, acquire data $d_t$ and create the data vector $\Psi_t = [y_t, \psi_t']'$.
- Set $\nu_t = \nu_{t-1} + 1$.
- Update by $\Psi_t$ the matrix $V_{t-1}$ to the matrix $V_t$ of the same dimension as follows.

If $\mathring{V}_{t-1} < \mathring{V}$ set $V_t' = [V'_{t-1}, \Psi_t]$
otherwise
– Evaluate $H = \left(V'_{t-1}V_{t-1}\right)^{-0.5}$
– Compute the vector $F = H\Psi_t$, the matrix $G = HV'_{t-1}$ with $i$th column ${}^iG$ and evaluate the scalar $\gamma = 1 + F'F$.
– For $i = 1, \ldots, \mathring{V}$, evaluate $a_i \equiv \gamma(1 - {}^iG'\,{}^iG) + ({}^iG'F)^2$ and find $k$ such that $|a_k| \ge |a_i|$.
– Set $V_t = V_{t-1}$ if $|a_k| \le 1$ else replace $k$th row of $V_{t-1}$ by $\Psi_t'$ to get $V_t$.

- Preserve the point estimate $\hat{\Theta}_t \equiv \hat{\Theta}_{t-1}$ of parameters $\Theta$ if $V_t = V_{t-1}$ otherwise update it

$$\hat{\Theta}_t = \arg\min_{r\in[\underline{r},\bar{r}]} \{\Theta:\ -\mathbf{1}_{\mathring{V}}r \le V_t[-1,\theta']'\mathbf{1}_{\mathring{V}}r\} \tag{11}$$

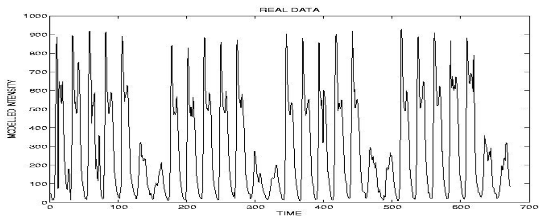
Increase appropriately $\bar{r}$ if the above LP fails.

- Go to the beginning of Recursive mode while data are available.

6. ILLUSTRATIVE EXAMPLE

The research reported here has been motivated by practical problems in estimation of urban-traffic models. This section illustrates basic features of the proposed algorithm on prediction of real transportation data. In a particular sub-problem, the counts of cars crossing a cross-road per hour were recorded. To get the required predictor, logarithms of these counts are modelled by fourth order uniform AR model. Predictions and prediction errors are expressed in original counts.

The processed data are on Figure 1 together with prediction errors obtained for $\mathring{V} = 7\times\mathring{\Psi} = 42$. These errors are typical.



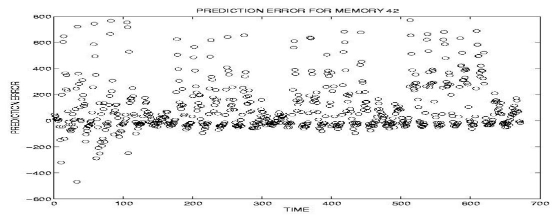




Fig. 1: Transportation data   Prediction errors, $\mathring{V} = 7 \times \mathring{\Psi} = 42$

Typical trajectories of parameter estimates $\hat{\theta}$ and $\hat{r}$ for $\mathring{V} = 42$ are in Figure 2. Character of

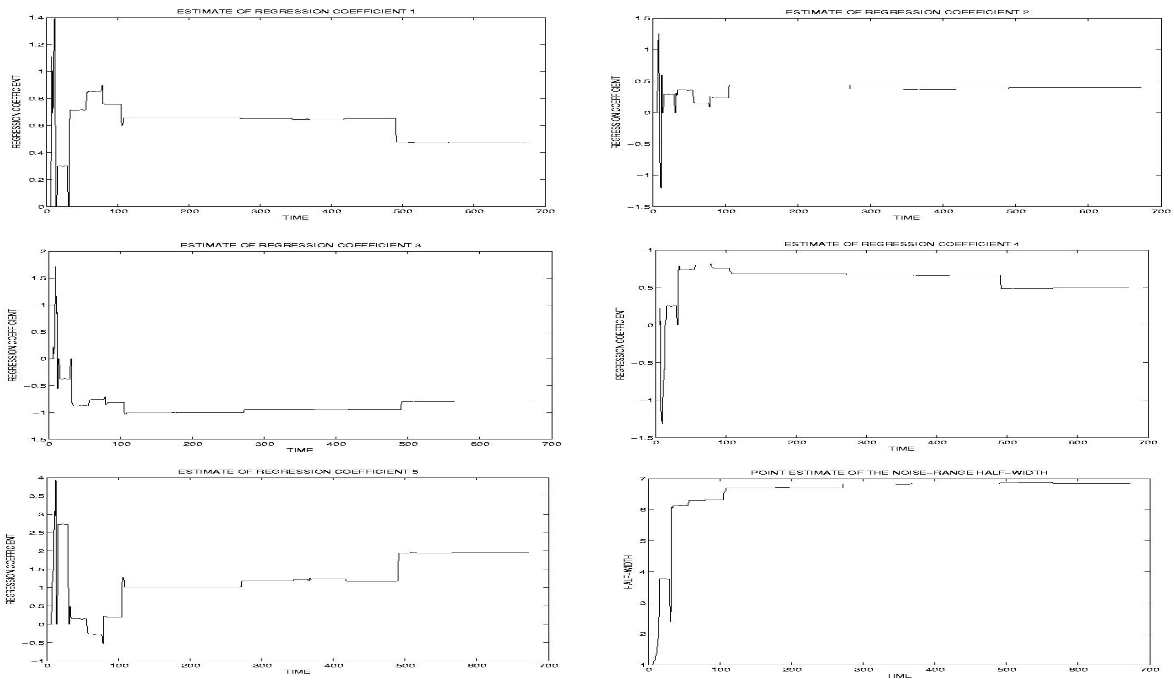



Fig. 2: Estimates $\hat{\theta}$ of $\theta$ for $\mathring{V} = 7 \times \mathring{\Psi} = 42$.

these trajectories is similar for tested dimensions $\mathring{V} = 12, 18, 24, 30, 36, 42, 60, 180$.

Table 1 containing elementary statistical evaluation complements the overall picture about properties of the algorithms. It is worth of noticing that the range 11-59% of informative data. The extension of the storage length $\mathring{V}$ essentially ceases to influence the result quality for $\mathring{V} > 24 = 4 \times \mathring{\Psi}$.

Table 1: Elementary sample statistics of estimation; $\hat{e} \equiv$ prediction error

| storage length $\mathring{V}$ | 12 | 24 | 42 | 180 |
|---|---|---|---|---|
| mean of $\hat{e}$ | 114 | 100 | 99 | 96 |
| minimum of $\hat{e}$ | -320 | -450 | -467 | -462 |
| maximum of $\hat{e}$ | 828 | 773 | 772 | 774 |
| standard deviation of $\hat{e}$ | 192 | 190 | 189 | 188 |
| ratio of standard deviations of $\hat{e}$ and data | 0.74 | 0.73 | 0.73 | 0.73 |
| portion of informative data [ % ] | 43 | 11 | 18 | 59 |
| elapsed time [s] | 4.2 | 11.1 | 22.6 | 702.4 |

## 7. CONCLUSIONS

The paper provides an approximation of the recursive Bayesian estimation of the ARX model with uniform noise. The proposed algorithm provides description of possible models by a convex support of the (approximate) posterior pdf. It respects hard bounds on all model parameters.

With respect to further research it offers related problems, like structure estimation. Up to now the model structure was supposed to be known.

## REFERENCES


Andrýsek, J. (2004). Approximate recursive Bayesian estimation of dynamic probabilistic mixtures. In: *Multiple Participant Decision Making* (J. Andrýsek, M. Kárný and J. Kracík, Eds.). Advanced Knowledge International. Adelaide. pp. 39–54.

Kárný, M. and L. Pavelková (2005 - submitted). Projection-based bayesian recursive estimation of arx model with uniform innovations. *Systems & Control Letters*.

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

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



Acknowledgment:
This research has been partially supported by Czech-Slovenian project MSMT 8-2006-06 ”Data-driven modelling for decision making support and process monitoring” and by grants GA ČR 102/03/0049 and AV ČR S1075351