Parallel Estimation Respecting Constraints of Parametric Models of Cold Rolling

Pavel Ettler * **Miroslav Kárný** **

* *COMPUREG Plzeň, s.r.o., 306 34 Plzeň, Czech Republic*
(e-mail: ettler@compureg.cz)
** *Institute of Information Theory and Automation AV ČR*
182 08 Praha 8, Czech Republic
(e-mail: school@utia.cas.cz)

Abstract:
Model-based predictors and controllers frequently depend on efficient recursive estimation of model parameters. Similarly often, there are known hard bounds on parameter values. Adaptive control applied for rolling mills represents a typical example of such case. While common estimation algorithms are elaborated enough to be utilized in industrial practice, it is difficult to find implementation of bounded estimation, which is both formally consistent and suitable for reliable applications. Solution offered in this paper is based on simultaneous run of two or more proven estimators different in applied process models. Both simulated and real data examples are provided.

Keywords: parameter estimation, process model, identification, steel industry

1. INTRODUCTION

Model-based prediction and control have their firm position among advanced control techniques and have been utilized for demanding industrial adaptive controllers for decades - see e.g. Ettler (1986) for early application of an adaptive AGC (Automatic Gauge Control). It can be based on recursive Bayesian parameter estimation representing theoretically consistent treatment of uncertainty – see e.g. Berger (1985). In special case – under the assumption of normal probability distribution of all processed quantities and when restricted to the linear normal autoregressive model with external variables (ARX) – it leads to the very efficient and numerically robust algorithm (see e.g. Bierman (1977)).

Industrial applications often work with a mathematical model, the structure of which is based, at least to some extent, on a physical model of the process (so called *gray box modelling*, e.g. Bohlin (1991)). Recursive estimation can track changes of parameter caused either by a real change of some physical parameter of the process or compensating imperfect matching of the process and its model. For unrestricted estimation, parameter estimates may occur in regions, which are formally correct but physically unreasonable. Then, especially in the case of an abrupt change of some physical parameter of the process, behavior of a corresponding predictor or controller can become undesirable. Application of constraints to the parameter estimates could be beneficial in such cases. Another motivation for avoiding undesirable parameter estimates has arisen in correspondence with an idea of model mixing (Ettler and Andrýsek (2007)).

Existence of bounded intervals for acceptable model parameters can be well respected within Bayesian framework (Berger (1985)). It suffices to restrict support of their prior distribution to this range. For recursive use, this possibility has been elaborated in Kárný (1982) for Gaussian parametric model and in Kárný and Pavelková (2007) for uniform parametric model. Neither of those solutions has been established as a practical tool. Thus, it makes sense to search for a suboptimal solution that preserve key features of these "clean" solutions. This paper provides such a solution.

Section 2 summarizes used modelling and estimation basics, Section 3 outlines the basic idea and a conceptually simple solution including a simulated example. Solution is further elaborated in Section 4 to provide more efficient algorithm. Section 5 compares the solution with a single bounded estimation. Results for experiments on rolling mill data are shown in Section 6, preceding concluding remarks in Section 7.

2. PROCESS MODEL AND PARAMETER ESTIMATION

At discrete time moment $t = 1, 2, \ldots,$ the considered process model relates scalar output $y(t)$ to a finite "influential" past collected into the regression vector

$$d(t) = [y(t-1),\ \ldots,\ y(t-t_y),\\ u_1(t),\ \ldots,\ u_1(t-t_u),\\ v(t),\ \ldots,\ v(t-t_v)]'.$$

It contains t_y samples of past outputs y, $t_u + 1$ samples of the optional input u and $t_v + 1$ samples of measured disturbance v. Both inputs and disturbance can be multidimensional. The stochastic relation is parameterized by unknown finite-dimensional parameters $\Theta(t)$

and fully described by probability density function (pdf) $f(y(t)|\text{past}, \Theta(t))$.

2.1 Linear Gaussian ARX parametric model

We assume specifically that the modelled process can be at least approximately described by a linear auto-regressive model with external variables (ARX)

$$y(t) = P'(t)\ d(t) + e(t), \tag{1}$$

where $'$ denotes transposition, $P(t)$ is vector of m unknown parameters

$$P(t) = [p_1(t),\ p_2(t),\ \ldots,\ p_m(t)]' \tag{2}$$

and $e(t)$ denotes zero-mean normally distributed noise with unknown variance $r(t)$. This defines normal ARX model with unknown parameter $\Theta(t) = (P(t), r(r))$, i.e., $f(y(t)|\text{past}, \Theta(t))$ = normal pdf with expectation $P'(t)\ d(t)$ and variance $r(t)$.

2.2 Bayesian parameter estimation

Parameter estimation is a classical problem addressed in two main streams: frequentist approach and Bayesian one. There exists long-lasted controversy between them, which can hardly be resolved generally. In decision making tasks, among which prediction and control belong, the Bayesian direction is to be definitely preferred (DeGroot (1970); Kárný et al. (2005)) as it guides how to cope with uncertianties within the finite time. The need to choose a prior distribution is often taken as the main disadvantage of the Bayesian approach. At the practical level, however, its explicit consideration is quite helpful because there is always a prior knowledge about the inspected technical problem. Lack of tools of translating prior knowledge into the prior distribution is the true problem behind. However, it has been overcome at least for the considered class of problems (Kárný et al. (2003); O'Hagan et al. (2006)).

A deeper discussion is out of scope of this paper so that we just state: we adopt the Bayesian paradigm (see Berger (1985)) that treats unknown parameters as random variables. Full information about them is contained in posterior pdf $f(\Theta(t)|\text{observed past})$. It can be shown (Kárný et al. (2005)) that, for time invariant parameters, the posterior pdf has a fixed functional (Gauss-inverse-Wishart) form determined by a symmetric positive definite (extended) information matrix $V(t)$. By using forgetting (e.g. Kulhavý and Zarrop (1993)), this form can be preserved even for varying parameters.

Propagation of the information matrix $V(t)$ can be formally expressed as

$$V(t) = \lambda V(t-1) + [y(t),\ d(t)]'\ [y(t),\ d(t)], \tag{3}$$

where $\lambda < 1$ is a forgetting factor. This propagation is formally equivalent with recursive least squares (LS). For instance, with splitting $V = \begin{bmatrix} V_{0,0} & v' \\ v & \mathcal{V} \end{bmatrix}$

$$\hat{P}(t) = \mathcal{V}^{-1} v = \text{LS estimate of } P(t).$$

Note that to ensure numerical stability of the identification algorithm under all circumstances, propagation of the information matrix is realized in the form of matrix factorization of some kind, e.g. $V^{-1} = LDL'$ where L, D are lower triangular and diagonal matrices respectively.

3. BASIC IDEA

For the sake of simplicity, let us omit time indexing in most of the following notation, so that e.g. $P(t) \doteq P$ and $d(t) \doteq d$. We distinguish parameter estimates $\hat{P}, \hat{p}_i$ from their true but unknown values P, p_i.

Let us introduce extreme values $\tilde{p}_i$ for each estimate $\hat{p}_i$ except for the absolute term $\hat{p}_m$, i.e., $i \in I \equiv \{1, \ldots, m-1\}$ such that

$$\tilde{p}_i = \begin{cases} p_{\min\, i} & \text{if } \hat{p}_i \le p_{\min\, i} \\ p_{\max\, i} & \text{if } \hat{p}_i \ge p_{\max\, i} \end{cases}, \tag{4}$$

where $p_{\min\, i}, p_{\max\, i}$ are given parameter minima and maxima respectively. Then, having parameter constraints in mind, omitting the additive noise and using parameter estimates instead of their unknown values we can rewrite equation 1 into the form

$$y - \sum_{i_r} \tilde{p}_{i_r}\ d_{i_r} = \sum_{i_u} \hat{p}_{i_u}\ d_{i_u} + \hat{p}_m\ , \tag{5}$$

where indices $i_r \in I_r \subseteq I$ correspond to parameters for which restriction was applied, $i_u \in I_u \subseteq I$ belong to unrestricted parameters and it holds $i_r \neq i_u$ for each i_r, i_u.

Let us further create so many models as is the number of J possible combinations of vectors I_r, I_u meeting abovementioned conditions. To avoid confusion let us introduce model numbering and additional indices $j = 0, \ldots, J$ for parameters ${}^j\hat{p}_{i_u}$ while $j = 0$ stands for the original (full) model. Thus for the j-th model we get

$$y - \sum_{i_r}^{j} \tilde{p}_{i_r}\ d_{i_r} = \sum_{i_u} {}^j\hat{p}_{i_u}\ d_{i_u} + {}^j\hat{p}_m\ , \tag{6}$$

Note that $\tilde{p}_{i_r}$ are constants and therefore for a limitary case when $I_r \equiv I$ and I_u is an empty vector, parameter $\hat{p}_m$ remains the only variable allowing to preserve equality of (6).

Basic identification algorithm can be described as follows:

- Initialize J recursive estimators and let them run in parallel;
- For each estimation step
  (1) Compute parameter estimates for each estimator;
  (2) Set $j = 0$ and pre-select estimates of the full model as the estimation output, i.e. $\hat{P} = {}^0\hat{P}$;
  (3) `If` the estimates $\hat{P}$ lie inside bounds provide them as the estimation output and wait for the next step;
  (4) `Otherwise` set j to the estimator for which the restricted parameters occur on the left hand side of (6), compose the estimation output $\hat{P}$ from ${}^j\tilde{p}_{i_r}$ and ${}^j\hat{p}_{i_u}$ and go to the estimation step (3).

3.1 Illustrative Example

To illustrate the idea let us consider a simple model with three parameters

$$y(t) = [-0.9, -0.8, 1] \begin{bmatrix} y(t-1) + c \\ u(t-1) \\ 1 \end{bmatrix} + e(t), \qquad (7)$$

for which $t \in \{1, \ldots, t_n\}$, $t_n = 5000$, $e(t)$ is a zero-mean normally distributed noise with variance $r = 0.1$ and c is a term representing an abrupt change of process behavior for time $t > t_n/3$

$$c = \begin{cases} 0 \text{ if } t \leq t_n/3 \\ 5 \text{ otherwise} \end{cases} \qquad (8)$$

$u(t)$ is randomly generated, normally distributed sequence of process input. Output of the simulated system can be seen in Fig. 1.

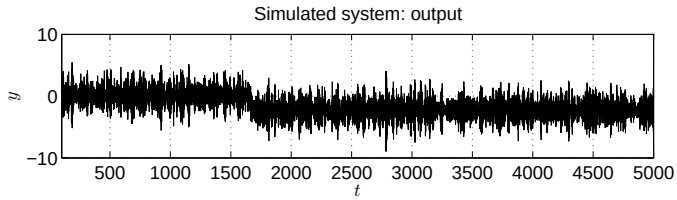

Fig. 1. Illustrative example: system output y. Abrupt change in the process behavior can be observed.

For the model with three parameters $m = 3$ we get a set of $J = 4$ model equations

$$\begin{array}{lllll} \text{model 0}: y & & = {}^0\hat{p}_1\, d_1 + {}^0\hat{p}_2\, d_2 + {}^0\hat{p}_3\, d_3 \\ \text{model 1}: y - \tilde{p}_1\, d_1 & & = {}^1\hat{p}_2\, d_2 + {}^1\hat{p}_3\, d_3 \\ \text{model 2}: y & - \tilde{p}_2\, d_2 = {}^2\hat{p}_1\, d_1 & + {}^2\hat{p}_3\, d_3 \\ \text{model 3}: y - \tilde{p}_1\, d_1 & - \tilde{p}_2\, d_2 = & {}^3\hat{p}_3\, d_3 \end{array}$$

Let us define parameter boundaries

$$\begin{array}{ll} p_{\min 1} = -1.00 & p_{\max 1} = -0.888 \\ p_{\min 2} = -0.82 & p_{\max 2} = -0.797 \end{array} \qquad (9)$$

Fig. 2 shows time progress of restricted and unrestricted parameter estimates (thick and thin lines respectively). Constraints for parameters $\hat{p}_1, \hat{p}_2$ were set close to their true values, which forced the algorithm to project the whole change of the process behavior into the change of the absolute term - parameter $\hat{p}_3$. Such smooth transient to a new equilibrium could be hardly achieved by plain tuning of a single estimator.

For this special simulated case the process and its model match exactly and the constraints could be set closely to their true values. This resulted in improvement of the output prediction $\hat{y}$ against ${}^0\hat{y}$ as can bee seen on comparison of prediction errors in Fig. 3. Decrease of the prediction error is generally not expected for real processes, reasonable behavior of parameter estimates remains as the main criterion instead.

4. ELABORATING THE SOLUTION

Utilization of J parallel estimators is technically simple but it is connected with two drawbacks:

- Number of estimators can be too high for systems with more parameters. This is not the main obstacle for applications in industry as the models have rarely more than five parameters and nowadays industrial computers are powerful enough.
- Trajectories of relevant parameter estimates of particular estimators can differ significantly for real situations. Then, switching among models due to application of constraints could result in excessive changes in parameters in some cases. This might cause problems for time-delayed systems.

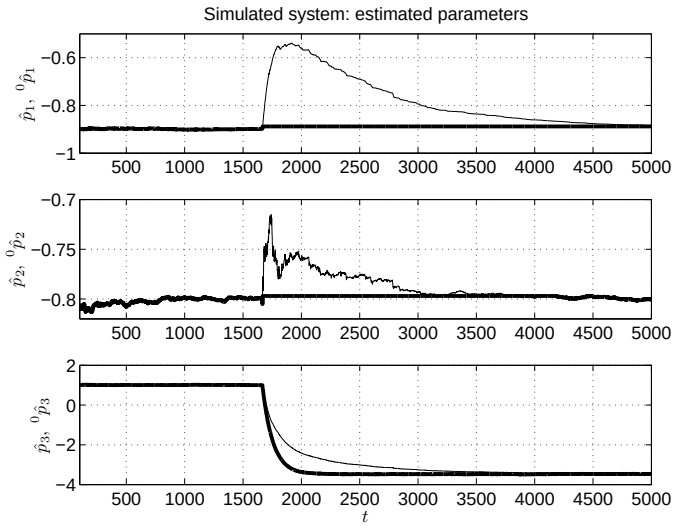

Fig. 2. Illustrative example: estimated parameters. Thick line corresponds to restricted parameter estimates $\hat{p}_i$ while the thin line stands for unrestricted estimates ${}^0\hat{p}_i$ of the full estimator.

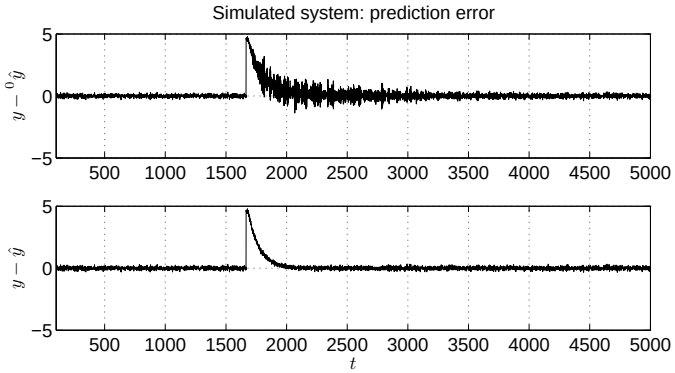

Fig. 3. Illustrative example: application of parameter constraints helped to minimize prediction error for this special simulated case. Situations for unrestricted and bounded estimations are shown respectively.

The latter drawback motivates us for further considerations. Evaluation of the full model estimation is obviously needful. However, the estimators, which are currently not selected to provide active estimates, are running uselessly for the given estimation step. Modification of the basic idea now consists in reduction of number of concurrently running estimators to two: the full one and the estimator the model of which reflects actual need for parameter restriction(s). Then the problem can be narrowed to proper initialization of the estimator, which was selected to be started.

Initialization of an estimator consists in proper initial setting of its information matrix V. To do that we can extract relevant elements from ${}^0V$ of the full estimator to initialize smaller matrix ${}^jV$ of the estimator, which was selected to be started in current step. Note that we speak here about the matrix V for the sake of simplicity while its factorized form is used in a real application instead.

As an example let us consider the system with three parameters again. Let us assume that for the current step two estimators should be active:

$$\begin{array}{ll} \text{model } 0: y & = {}^0\hat{p}_1\, d_1 + {}^0\hat{p}_2\, d_2 + {}^0\hat{p}_3\, d_3 \\ \text{model } 1: y - \tilde{p}_1\, d_1 = & \quad {}^1\hat{p}_2\, d_2 + {}^1\hat{p}_3\, d_3 \end{array} \tag{10}$$

Estimator 0 has been already running while estimator 1 was newly selected and has to be initialized. We choose appropriate elements of the matrix ${}^0V$ to initialize ${}^1V$:

$$\begin{array}{l} {}^0V_{0,0} \\ {}^0V_{0,1}\ {}^0V_{1,1} \\ {}^0V_{0,2}\ {}^0V_{1,2}\ {}^0V_{2,2} \\ {}^0V_{0,3}\ {}^0V_{1,3}\ {}^0V_{2,3}\ {}^0V_{3,3} \end{array}$$

$$\Downarrow \tag{11}$$

$$\begin{array}{lll} {}^0V_{0,0} & & \\ - & - & \\ {}^0V_{0,2} & - & {}^0V_{2,2} \\ {}^0V_{0,3} & - & {}^0V_{2,3}\ {}^0V_{3,3} \end{array} \quad \Rightarrow \quad \begin{array}{l} {}^1V_{0,0} \\ {}^1V_{0,1}\ {}^1V_{1,1} \\ {}^1V_{0,2}\ {}^1V_{1,2}\ {}^1V_{2,2} \end{array}$$

Modified algorithm can be summarized as follows:

- Initialize basic recursive full estimator and start its run;
- For each estimation step:
  (1) Compute parameter estimates for full estimator and for a second estimator j if it is activated;
  (2) Pre-select estimates of selected model for estimation output, i.e. $\hat{P} = {}^0\hat{P}$ or $\hat{P} = {}^j\hat{P}$;
  (3) `If` the estimates $\hat{P}$ lie inside bounds provide them as estimation output and wait for the next step;
  (4) `Otherwise` set j for (possibly another) estimator according to applied parameter constraints. `If` the selected estimator was inactive in preceding step initialize its ${}^jV$ based on the information matrix of the estimator whose estimates were used last time;
  (5) Execute the j-th estimator and set $\hat{P} = {}^j\hat{P}$ and go to estimation step (3).

Results of the modified algorithm are very similar to the basic algorithm when tested on the above-introduced simulated system – see Fig. 2.

5. COMPARISON WITH SINGLE ESTIMATION WITH APPLIED CONSTRAINTS

The proposed algorithm allows to cope with constraints applied to relatively abruptly changing parameters. In real systems it may also happen that parameters vary very slowly and estimates reach the bounds due to random drifts of the mode of the posterior situation. This case was inspected in Kárný (1982), showing that time evolution of the extended information is to be completely uninfluenced by presence of bounds. The point estimate of regression coefficients is just found as the maximizer of the posterior pdf over the set given by constraints. The resulting estimate is used in prediction but it is *not* fed into subsequent data updating. Algorithmically, a simple quadratic programming is solved.

This variant is considered here for comparison with the proposed algorithm as it is a priori unclear, which of the reasons of crossing the constraints is decisive. Fig. 4 compares both approaches on our testing example. Parameter estimates for the single bounded estimation are depicted as ${}^S\hat{p}_i$. While ${}^S\hat{p}_1$ and ${}^0\hat{p}_1$ are almost equal and are not plotted here, ${}^S\hat{p}_2$ and ${}^S\hat{p}_3$ are less favourable than corresponding estimates from the proposed estimation algorithm.

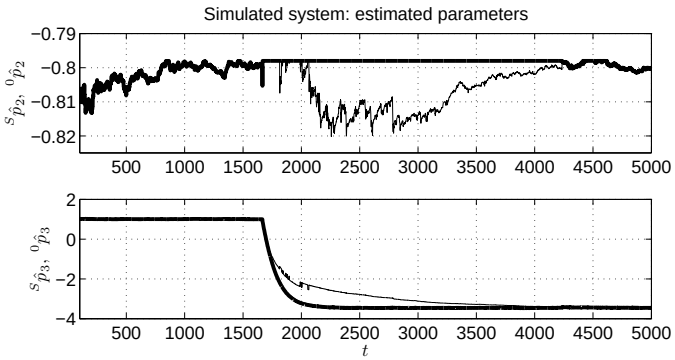

Fig. 4. Comparison of the second and third parameter estimates for single and parallel bounded estimations (thin and thick lines respectively).

6. EXPERIMENTS ON ROLLING MILL DATA

The modified algorithm was tested on real data from a reversing cold rolling mill. We selected two simplified models from the variety of models, which are being employed in the field of automatic gauge control. For both examples the system output was represented by deviation of the output strip thickness h_2. Process time is now derived from movement of the processed strip with distance between adjacent samples 0.08 m.

6.1 Real example I

Let us explore how the parameter estimation copes with the situation when the thickness measurement is temporarily deteriorated by the strip jittering – see Fig. 5.

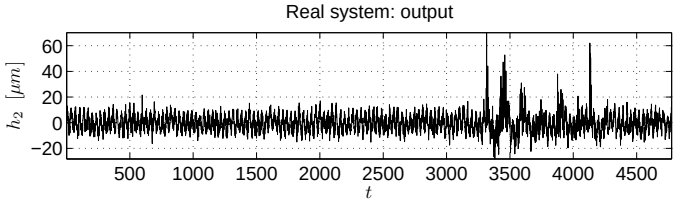

Fig. 5. Real system output – deviation of the output strip thickness h_2 measured on a cold rolling mill. Increased deviation was caused by inaccurate contact measurement of a jittering strip.

Let us consider a simplified model coming from the well known *gaugemeter principle* (see e.g. Ettler and Andrýsek (2007) for details) the regression vector of which has the form

$$d(t) = [z, f(F), 1]' \,, \tag{12}$$

where z stands for the uncompensated rolling gap and $f(F)$ denotes a function of the rolling force.

Parameter boundaries were set to

$$\begin{array}{ll} p_{\min 1} = -1.00 & p_{\max 1} = -0.02 \\ p_{\min 2} = -100.00 & p_{\max 2} = 0.00 \end{array} \tag{13}$$

Estimation results can be seen in Fig. 6. Problems with measurement of the output caused the first unrestricted parameter (thin line) to change its sign, which is physically inadmissible. Constraint applied by the restricted estimation (thick line) kept the negative sign of the first parameter influencing harmlessly time progress of the second parameter.

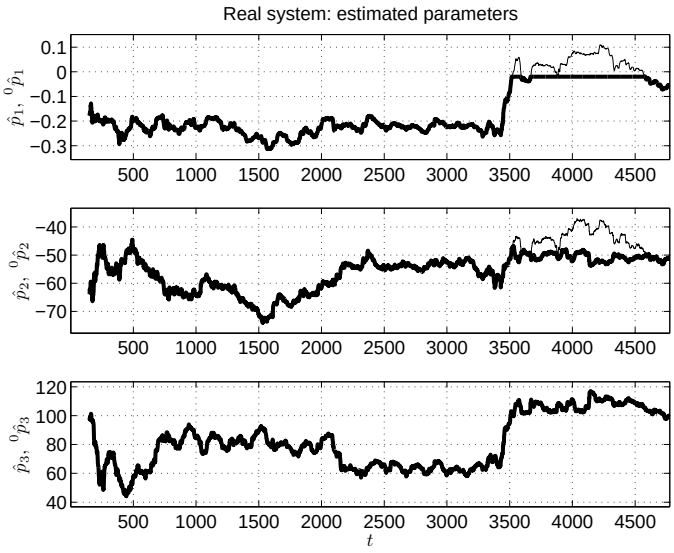

Fig. 6. Estimated parameters for the real system. Thick line corresponds to restricted parameter estimates $\hat{p}_i$ while the thin line stands for unrestricted estimates ${}^0\hat{p}_i$ of the full estimator.

6.2 Real example II

Another example depicts the situation when three metal strips were welded together to make processing on a rolling mill more effective. Particular strip parts can have slightly different physical properties, which should be reflected by changes of parameter estimates. However, measurement peaks caused by the welds – see Fig. 7 – can induce undesirable behavior of parameter estimates.

Let us consider another simplified model based for this one on the so called *mass-flow principle* (see e.g. Ettler and Andrýsek (2007) again) the regression vector of which has the form

$$d(t) = [v_r\ h_1, v_r, 1]' \,, \tag{14}$$

where h_1 stands for deviation of the input strip thickness and v_r denotes ratio of input and output strip speeds $v_r = v_1/v_2 < 1$.

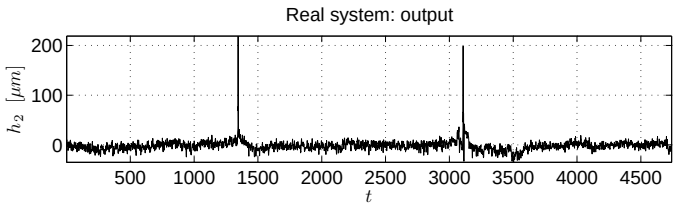

Fig. 7. Real system output – deviation of the output strip thickness h_2 measured on a cold rolling mill. Distinct peaks are caused by welds between adjacent metal strips.

Parameter boundaries were set to

$$\begin{array}{ll} p_{\min 1} = 0.01 & p_{\max 1} = 0.5 \\ p_{\min 2} = 0.00 & p_{\max 2} = 1500.0 \end{array} \tag{15}$$

Estimation results can be seen in Fig. 8. For the first part of the strip the first unrestricted parameter (thin line) tends to change its sign, which should be avoided. To the contrary, measurement peaks caused ${}^0\hat{p}_1$ to exceed the upper boundary. Thanks to the restricted estimation (thick line) undesirable effects were acceptably transferred to the second and third parameters.

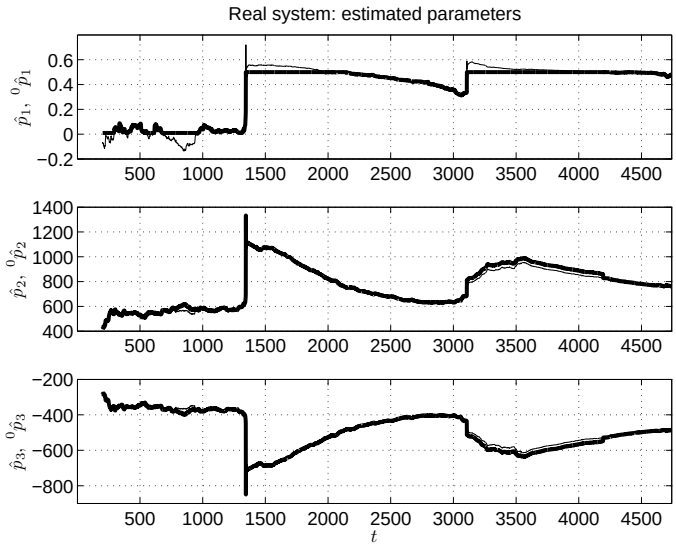

Fig. 8. Estimated parameters for the real system. Thick line corresponds to restricted parameter estimates $\hat{p}_i$ while the thin line stands for unrestricted estimates ${}^0\hat{p}_i$ of the full estimator.

7. CONCLUSION

The paper introduced two algorithms for bounded parameter estimation the latter of which is considered to be more suitable for industrial applications such as thickness control for cold rolling mills. Simulated and real data examples outlined potential benefits of algorithms.

There still remains a potential drawback of the latter algorithms the main idea of which lies in discontinuous execution of an alternative estimator. While the initialization of its information matrix is straightforward, switching back to the full estimator might cause abrupt change of parameters in some special case. Formal solution implies

modification of the ${}^0V$ matrix of the full estimator when the alternative estimator is being switched off. This remains as a motivation for future research.

ACKNOWLEDGEMENTS

This work was partially supported by grants from the Czech Ministry of Education, Youth and Sports: 7D09008 (ProBaSensor – Eurostars project) and 1M0572 (DAR – Research Centre Data-Algorithms-Decision Making).

REFERENCES

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

Bierman, G. (1977). *Factorization Methods for Discrete Sequential Estimation*. Academic Press, New York.

Bohlin, T. (1991). *Interactive System Identification: Prospects and Pitfalls*. Springer-Verlag, New York.

DeGroot, M. (1970). *Optimal Statistical Decisions*. McGraw-Hill, New York.

Ettler, P. (1986). An adaptive controller for Skoda twenty-roll cold rolling mills. In *Proceedings of 2nd IFAC Workshop on Adaptive Systems in Control and Signal Processing*, 277–280. Lund Institute of Technology, Lund, Sweden.

Ettler, P. and Andrýsek, J. (2007). Mixing models to improve gauge prediction for cold rolling mills. In *Proceedings of the 12th IFAC Symposium on Automation in Mining, Mineral and Metal Processing*. Québec, Canada.

Kárný, M. (1982). Recursive parameter estimation of regression model when the interval of possible values is given. *Kybernetika*, 18(2), 164–178.

Kárný, M., Böhm, J., Guy, T.V., Jirsa, L., Nagy, I., Nedoma, P., and Tesař, L. (2005). *Optimized Bayesian Dynamic Advising: Theory and Algorithms*. Springer, London.

Kárný, M., Nedoma, P., Khailova, N., and Pavelková, L. (2003). Prior information in structure estimation. *IEE Proceedings — Control Theory and Applications*, 150(6), 643–653.

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

Kulhavý, R. and Zarrop, M.B. (1993). On a general concept of forgetting. *International Journal of Control*, 58(4), 905–924.

O'Hagan, A., Buck, C., Daneshkhah, A., Eiser, J., Garthwaite, P., Jenkinson, D.J., Oakley, J., and Rakow, T. (2006). *Uncertain judgement: eliciting experts' probabilities*. John Wiley & Sons.