

IDENTIFICATION WITH INTERRUPTIONS AS AN ANTI-BURSTING DEVICE

MIROSLAV KÁRNÝ

Behaviour known as 'bursting' of an estimator is well known and often found in applications of self-tuning regulators using a forgetting factor. This phenomenon can be described as short-term closed-loop instability, similar to initial behaviour of the regulator with bad prior information about the controlled system.

Bursting can be explained simply and this explanation leads to ways in which it can be removed.

One commonly used technique based on switching off the identification is the subject of this paper. Bursting is described from the Bayesian point of view, the identification with interruptions is designed and the simple stopping rule is given. The rule results from testing properly selected hypotheses under simplified assumptions. The behaviour of the algorithm is illustrated by a simple example.

1. INTRODUCTION

The majority of self-tuning regulators is primarily designed for systems with unknown but constant parameters. Adaptivity, i.e. ability to track at least slow parameter changes, is mainly achieved through exponential forgetting.

However, when such a scheme is used, short-term instability, called bursting, may occur [1], [2].

This phenomenon is now well understood: a linear constant regulator, which is the stationary point of the assumed type of self-tuner, generates actions which are not sufficiently exciting. It implies that a hyperplane of equivalent (equally probable) point estimates appears in the parameter space. The estimation becomes numerically unstable and this implies regulator instability.

The appearance of the bursting phenomenon is the probable the longer the regulator is (almost) constant, especially when closed loop behaviour is rather 'quiet'. But for a successfully controlling regulator the identification becomes unnecessary since the system model is sufficiently accurate. This observation results in one commonly

used strategy to suppress bursting: the identification is switched off, when it is unnecessary and switched on only when needed.

The present paper describes the simple decision rule for identification switching based on testing of a properly selected hypothesis. Identification and the explanation of exponential forgetting as well as hypothesis testing are Bayesian.

2. IDENTIFICATION OF SYSTEMS WITH SLOWLY VARYING PARAMETERS

Let us assume that the stochastic controlled system is given. The finite-dimensional vectors of data $d_{(t)}$ are measured at discrete time moments, labeled by $t = 1, 2, 3, \dots$. The directly manipulated variable are the system inputs $u_{(t)}$ and the rest of data, sampled within the sampling period indexed by t , is collected into v -vector of system outputs $y_{(t)}$

$$(2.1) \quad d_{(t)} = (y_{(t)}, u_{(t)})$$

The following short-hand notation will be used

$$(2.2) \quad d^{(t)} = \{d_{(1)}, d_{(2)}, \dots, d_{(t)}\}$$

The symbol $p(\alpha | \beta)$ will denote the probability density function of α conditioned on β .

The fact that some data are omitted in identification will be formally represented by using statistic

$$(2.3) \quad \omega(\cdot) : d^{(t)} \rightarrow \omega^{(t)}$$

The relation between input and output will be described by system model of the form

$$(2.4) \quad p(y_{(t)} | d^{(t-1)}, u_{(t)}, \Theta_{(t)}) = \Psi(y_{(t)}, z_{(t)}, \Theta_{(t)})$$

where $\Theta_{(t)}$ is the finite dimensional, generally time-dependent unknown parameter vector. Data enter the model through the known ρ -dimensional statistic

$$(2.5) \quad z(\cdot) : d^{(t-1)}, u_{(t)} \rightarrow z_{(t)}$$

The functions z, Ψ are assumed to be time invariant.

The symbol \propto used below means equality up to the factor independent of parameter Θ . Moreover the short-hand notation will be used

$$(2.6) \quad p(\Theta | \omega^{(t-1)}) = p_{t|t-1}(\Theta) \quad \text{for } \Theta = \Theta_{(t)}; \quad i = 0, 1$$

The regulator is supposed to satisfy natural conditions of control [4]

$$(2.7) \quad p(u_{(t)} | d^{(t-1)}, \Theta^{(t)}) = p(u_{(t)} | d^{(t-1)})$$

Assuming (2.7) the Bayes formula results in the following updating of conditional probability density function of unknown parameters by new data item

$$(2.8) \quad p_{t|t}(\theta) \propto p(y_{(t)} | d^{(t-1)}, u_{(t)}, \theta) p_{t|t-1}(\theta)$$

This relation represents the objective identification step. Parameters changes are assumed to be slow, i.e. the successive parameter values do not differ too much. Such a rough description is insufficient to determine $p_{t+1|t}(\theta)$ for a given $p_{t|t}(\theta)$. It is, however, apparent that in the time interval $(t, t + 1)$ the values of parameters may change and there are no data for correction of $p_{t|t}(\theta)$, i.e. uncertainty about $\theta_{(t+1)}$ for given $\omega^{(t)}$ is greater than about $\theta_{(t)}$. This fact can be respected by defining

$$(2.10) \quad p_{t+1|t}(\theta) \propto [p_{t|t}(\theta)]^{\varphi_{(t)}}$$

where factor $\varphi_{(t)}$ (generally depending on data $\omega^{(t)}$) fulfils inequalities

$$(2.11) \quad 0 < \varphi_{(t)} \leq 1$$

Note that for $\varphi_{(t)} \leq 1$ the operation (2.10) means flattening of $p_{t|t}(\theta)$ to obtain $p_{t+1|t}(\theta)$. Hence the exponent $\varphi_{(t)}$ (known as the factor of exponential forgetting) reflects the increase of uncertainty about the unknown parameters due to the possible variations of their values.

The predictor, i.e. system outer model (without unknown parameters) is an estimate of $\Psi(y, z, \theta_{(t)})$ for any fixed pair y, z . The reasonable one is

$$(2.12) \quad \hat{\Psi}_{t|t-1}(y, z) = \int \Psi(y, z, \theta) p_{t|t-1}(\theta) d\theta$$

i.e. conditional expectation of $\Psi(y, z, \theta)$ with respect to $\theta_{(t)}$ having probability density function $p_{t|t-1}(\theta)$.

Let us note that for $y = y_{(t)}$, $z = z_{(t)}$ and $\omega^{(t-1)} = d^{(t-1)}$ i.e. when no past data are omitted in the condition, the following identity is valid:

$$(2.13) \quad \hat{\Psi}_{t|t-1}(y_{(t)}, z_{(t)}) = p(y_{(t)} | d^{(t-1)}, u_{(t)})$$

If the distribution given by $p_{t|t-1}(\theta)$ is sufficiently concentrated, then the following approximation can be used:

$$(2.14) \quad \hat{\Psi}_{t|t-1}(y, z) \approx \Psi(y, z, \hat{\theta}_{(t-1)})$$

where $\hat{\theta}_{(t-1)}$ is a point estimate of $\theta_{(t)}$ based on the statistic $\omega^{(t-1)}$, for instance the value of θ for which $p_{t|t-1}(\theta)$ reaches its maximum.

Now we specialize the theory to the normal multivariate regression model [4], [5]

$$(2.15) \quad p(y_{(t)} | d^{(t-1)}, u_{(t)}, \theta_{(t)}) = N(\hat{y}_{(t)}, \Omega_{(t)}^{-1})$$

where the conditional mean $\hat{y}_{(t)}$ can be expressed as follows

$$(2.16) \quad \hat{y}_{(t)} = P_{(t)}^T z_{(t)}$$

The matrix of the regression coefficients $P_{(t)}$ and the precision matrix $\Omega_{(t)}$ (inverse of the covariance matrix) form the set of unknown parameters

$$(2.17) \quad \Theta_{(t)} = \{P_{(t)}, \Omega_{(t)}\}, \quad \Omega > 0$$

If the conjugate prior distribution, which preserves the functional form under transformation (2.8), is used and the case $\omega^{(t)} = I^{(t)}$ is assumed, then

$$(2.18) \quad p_{t|t-1}(\Theta) \propto |\Omega|^{g_{(t-1)}/2} \exp \left\{ -\frac{1}{2} \text{tr} \left(\begin{bmatrix} -I \\ P \end{bmatrix}^T V_{(t-1)} \begin{bmatrix} -I \\ P \end{bmatrix} \Omega \right) \right\}$$

The relations (2.8), (2.9) reduce to the recursions

$$(2.19) \quad g_{(t)} = \varphi_{(t)}(g_{(t-1)} + 1) \quad g_{(0)} \geq 0$$

$$(2.20) \quad V_{(t)} = \varphi_{(t)}(V_{(t-1)} + \bar{z}_{(t)}\bar{z}_{(t)}^T) \quad V_{(0)} > 0^*$$

where

$$(2.21) \quad \bar{z} = \begin{bmatrix} y \\ z \end{bmatrix} \begin{matrix} v \\ \varrho \end{matrix}$$

If the matrix V is partitioned

$$(2.22) \quad V = \begin{bmatrix} V_y & V_{zy}^T \\ V_{zy} & V_z \end{bmatrix} \begin{matrix} v \\ \varrho \end{matrix}$$

then quantities known from least squares are defined

$$(2.23) \quad \hat{R} = (V_y - V_{zy}^T V_z^{-1} V_{zy}) / (\varrho - \varrho)$$

$$(2.24) \quad \hat{P} = V_z^{-1} V_{zy}$$

It can be shown that the system outer model has Student's form. It can be and will be approximated by the normal probability density function

$$(2.25) \quad \hat{P}_{t|t-1}(y, z) = N(\hat{P}_{(t-1)}^T z, \hat{R}_{(t-1)}(1 + \xi_{(t-1)}(z)))$$

where

$$(2.26) \quad \xi_{(t)}(z) = z^T V_{z(t)}^{-1} z$$

3. IDENTIFICATION WITH INTERRUPTIONS

The bursting is connected with the suboptimal strategy of control based on the enforced separation of identification and control. With this strategy, the occurrence of bursting can be influenced either by suitable selection of model structure [6] (some parameter being fixed) or by switching off the estimation. The second method

* $V > 0$ means that V is positive definite matrix.

is considered in this paper because it seems to be more suitable for a general (explicit) design of controllers.

Identification with interruptions will now be described, for the decision rule see the next section.

The main purpose of identification in the assumed control situation is to determine the relation between the past observed system history $d^{(t-1)}$, the present input $u_{(t)}$ and the next system output $y_{(t)}$,

$$(3.1) \quad d^{(t-1)}, u_{(t)} \rightarrow y_{(t)}$$

Uncertainty of this relation has two parts:

- objective - Nobody, with the same observing ability and the same prior information, is able to relax this part.
- subjective - This reflects personal uncertainty, which can in principle be removed.

To be sure that the model is sufficiently accurate, i.e. the subjective part of the uncertainty can be neglected, the parameters describing the objective uncertainty (Ω) have to be permanently updated.

The algorithm for identification of a part of the unknown parameter vector $\Theta = (\Theta_1, \Theta_2)$, say Θ_2 , can be described by the following relations:

$$(3.2) \quad p_{t+1|t}(\Theta_1, \Theta_2) = p_{t+1|t}(\Theta_1 | \Theta_2) p_{t+1|t}(\Theta_2)$$

The distribution of Θ_1 conditioned on Θ_2 is forced to be unchanged

$$(3.3) \quad p_{t+1|t}(\Theta_1 | \Theta_2) = p_{t|t-1}(\Theta_1 | \Theta_2)$$

Identification of Θ_2 is described by a combination of (2.8) and (2.10)

$$(3.4) \quad p_{t+1|t}(\Theta_2) \propto [p(y_{(t)} | d^{(t-1)}, u_{(t)}, \Theta_2) p_{t|t-1}(\Theta_2)]^{\rho_{(t)}}$$

The model needed is given by

$$(3.5) \quad p(y_{(t)} | d^{(t-1)}, u_{(t)}, \Theta_2) = \int p(y_{(t)} | d^{(t-1)}, u_{(t)}, \Theta) p_{t|t-1}(\Theta_1 | \Theta_2) d\Theta_1$$

Similar computations as in the case of full identification [5] give, for the regression model

$$(3.6) \quad p_{t|t-1}(\Theta) \propto \propto |\Omega|^{\vartheta_{(t-1)}/2} \exp \left\{ -\frac{1}{2} \text{tr} \left[((P - \hat{P}_{(t-1)})^T V_{z(t-1)} (P - \hat{P}_{(t-1)}) + (\hat{R}_{(t-1)} (\vartheta_{(t-1)} - \varrho)) \Omega) \right] \right\}$$

This is one of the alternative forms of (2.18). The sufficient statistic (ϑ, V) is here replaced by the equivalent one $(\vartheta, V_z, \hat{R}, \hat{P})$.

The recursions (2.19), (2.20) are transformed, when *full identification* is performed, into the form

$$(3.7) \quad V_{z(t)} = \varphi_{(t)} (V_{z(t-1)} + z_{(t)} z_{(t)}^T)$$

$$(3.8) \quad \hat{P}_{(t)} = \hat{P}_{(t-1)} + \frac{V_{z(t-1)}^{-1} z_{(t)} \hat{e}_{(t)}^T}{\sigma_{(t)}}$$

$$(3.9) \quad \vartheta_{(t)} = \varphi_{(t)}(\vartheta_{(t-1)} + 1)$$

$$(3.10) \quad \hat{e}_{(t)} = y_{(t)} - \hat{P}_{(t-1)}^T z_{(t)}$$

$$(3.11) \quad \sigma_{(t)} = 1 + z_{(t)}^T V_{z(t-1)}^{-1} z_{(t)}$$

$$(3.12) \quad \hat{K}_{(t)} = \varphi_{(t)} \frac{\vartheta_{(t-1)} - \varrho}{\vartheta_{(t)} - \varrho} \left[\hat{K}_{(t-1)} + \frac{\hat{e}_{(t)} \hat{e}_{(t)}^T}{(\vartheta_{(t-1)} - \varrho) \sigma_{(t)}} \right]$$

The *partial identification* of only $\Omega (= \Theta_2)$ for uncorrected distribution of $P (= \Theta_1)$ is given by (3.6) with

$$(3.13) \quad V_{z(t)} = V_{z(t-1)}$$

$$(3.14) \quad \hat{P}_{(t)} = \hat{P}_{(t-1)}$$

$$(3.15) \quad \vartheta_{(t)} = \varphi_{(t)}(\vartheta_{(t-1)} + 1) + (1 - \varphi_{(t)}) \varrho$$

This means that the pause in updating \hat{P} and V_z influences only $\vartheta_{(t)}$ a little.

4. DECISION RULE FOR IDENTIFICATION SWITCHING

The last but the most important problem is to find the rule for switching between partial and full identification. Loosely, we may ask: 'Is the information about the system sufficient?' In order to solve this problem we have to formalize this question.

Generally the question is connected with closed loop behaviour. However, in the given context, when separation of identification and control is enforced, it seems to be reasonable to restrict the problem to the question: 'Is the estimate of the parameters sufficiently accurate? Is the obtained system model sufficiently accurate?' These questions can be formalized in the testing of the following hypotheses:

H_0 : the system description based on the old estimate of parameters (approximation (2.14)) is sufficient, i.e. the predictive probability distribution for the next output $y_{(t)}$ can be modelled as

$$(4.1) \quad p(y_{(t)} | \omega^{(t-1)}, u_{(t)}, H_0) = \Psi(y_{(t)}, z_{(t)}, \hat{\Theta}_{(t-1)})$$

H_1 : the approximation (2.14) is too rough and the model of the type (2.12) has to be used

$$(4.2) \quad p(y_{(t)} | \omega^{(t-1)}, u_{(t)}, H_1) = \int \Psi(y_{(t)}, z_{(t)}, \Theta) p_{t|t-1}(\Theta) d\Theta$$

The possible decisions are

α_0 : accept H_0 , switch off identification

α_1 : accept H_1 , use the new data item $d_{(t)}$ to correct $p_{t|t-1}(\Theta)$.

Admissible decision functions, i.e. acceptable rules selecting decisions α_i , are

$$(4.3) \quad \alpha(\cdot) : \omega^{(t-1)}, d_{(t)} \rightarrow \{\alpha_0, \alpha_1\}$$

To select the best of them, the losses connected with possible combinations of α_i, H_j have to be specified. The simplest loss function will be used here

$$(4.4) \quad L(\alpha_i, H_j) = (1 - \delta_{ij}) l, \quad l > 0, \quad i, j = 0, 1$$

where δ is Kronecker symbol. This function penalizes both of the possible discrepancies by the same weight l . The optimal decision rule minimizes expected loss

$$(4.5) \quad E[L(\alpha(\cdot), H_{(t)})] \quad t = 1, 2, 3, \dots$$

The solution of the task (4.1) to (4.5) has known structure, which can be verified by direct inspection

$$(4.6) \quad \alpha(\cdot) = \alpha_0 \quad \text{iff} \quad \frac{P(y_{(t)} | \omega^{(t-1)}, u_{(t)}, H_0)}{P(y_{(t)} | \omega^{(t-1)}, u_{(t)}, H_1)} \geq \gamma_{(t)}$$

where

$$(4.7) \quad \gamma_{(t)} = \frac{p_{t|t-1}(H_1)}{p_{t|t-1}(H_0)}$$

The rather simplifying assumption will be used that $\gamma_{(t)}$ is known constant with value near unity

$$(4.8) \quad \gamma_{(t)} \approx 1$$

It means that the probabilities of occurrence H_0, H_1 are almost the same independently of the past observed history.

Writing the rule (4.6) in the logarithmic form for the regression model with normal approximation to Student's distribution (2.25) we have

$$(4.9) \quad \alpha(\cdot) = \alpha_0 \quad \text{iff} \quad \hat{e}_{(t)}^T \hat{R}_{(t-1)}^{-1} \hat{e}_{(t)} \leq v \sigma_{(t)} \frac{\ln(\sigma_{(t)})}{\sigma_{(t)} - 1} + \frac{\sigma_{(t)} \ln(\gamma_{(t)})}{\sigma_{(t)} - 1}$$

Assuming (4.8) and assuming stationarity which is of our main interest, the following behaviour can be expected

$$(4.10) \quad \sigma_{(t)} \rightarrow 1 \quad (\sigma_{(t)} > 1)$$

then the final simplified version of the rule (4.6) is obtained

$$(4.11) \quad \alpha(\cdot) = \alpha_0 \quad \text{iff} \quad \frac{1}{v} \hat{e}_{(t)}^T \hat{R}_{(t-1)}^{-1} \hat{e}_{(t)} \leq \text{constant} \approx 1$$

Finally let us notice:

a) Assuming the case

$$(4.12) \quad p(H_1 | \cdot) \gg p(H_0 | \cdot)$$

which occurs e.g. when identification starts, it is reasonable to expect $\sigma_{(t)}$ to be of the same magnitude as $\gamma_{(t)}$. In this case the approximation (4.8) is also reasonable.

b) If

$$(4.13) \quad \hat{R}_{(t)} = r_{(t)}I \quad r_{(t)} > 0$$

then the rule (4.11) has the form

$$(4.14) \quad \alpha(\cdot) = \alpha_0 \quad \text{iff} \quad \frac{\hat{e}_{(t)}^T \hat{e}_{(t)}}{vr_{(t-1)}} = \frac{\hat{e}_{(t)}^T \hat{e}_{(t)}}{\text{tr}(\hat{R}_{(t-1)})} \leq \text{constant}$$

which can be successfully used even if (4.13) is valid only approximately.

c) The rule (4.11) has the following interesting interpretation: The uncertainty of prediction under hypothesis H_0 is always less than that under hypothesis H_1 . The less uncertain hypothesis H_0 is confirmed only when highly probable output occurs (see Fig. 1).

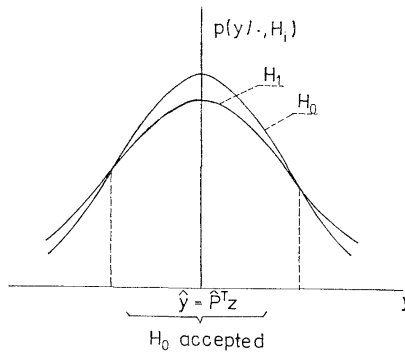


Fig. 1. Illustration of the proposed decision rule.

5. A SIMPLE EXAMPLE

To illustrate the behaviour of the proposed algorithm a first order minimum-phase system

$$\hat{y}_{(t)} = [b_0, a_1, b_1] z_{(t)} = [1.02, 0.98, -1.00] \begin{bmatrix} u_{(t)} \\ y_{(t-1)} \\ u_{(t-1)} \end{bmatrix}, \quad \Omega = 1$$

was controlled with a rather small forgetting factor (0.95). A one-stage-ahead output criterion was minimized under enforced separation of identification and control.

The point estimates of the b_0 are given in Figure 2. The runs (a) with and (b) without switching of identification are plotted. The rather dangerous changes of point estimates are apparent in case (b). The behaviour of the rest of the estimates is similar.

These different sequences of estimates result in very similar output behaviour giving the output sample dispersions 1.015 in case (a), 1.065 in case (b). The theoretical optimum is 1.000.

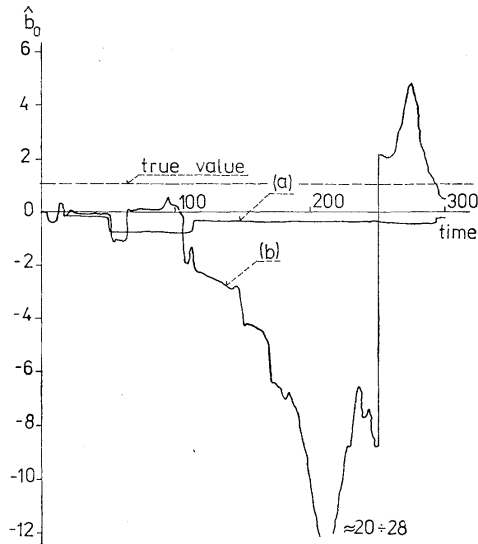


Fig. 2. Example of estimate behaviour. (a) — identification with switching (b) — identification without switching.

The displayed results illustrate the known fact that the quality of control depends on identification primarily through the quality of output predictions. Particular values of the point estimates used are not decisive.

6. CONCLUSION

The simplified hypotheses testing was used as the starting point for a proposition of an identification algorithm with switching. The objective of this algorithm was to prevent closed loop short-term instability known as 'bursting'.

The implicit requirement for simplicity of the resulting algorithm was fulfilled. It can be added simply to the existing identification algorithm without extra cost.

The algorithm has a strong intuitive appeal and has been successfully tested on a cold rolling system.

The practically interesting by – product for time – sharing computations seems to be the fact that only a relatively low number of samples of estimation is really needed for model correction, i.e. central-processor-unit time is saved.

(Received November 11, 1981.)

REFERENCES

- [1] S. Wittenmark: On Adaptive Control of Low Order Systems. Lund Institute of Technology, Div. of Aut. Control, Report 6918, 1969.
- [2] K. J. Aström, B. Westberg and B. Wittenmark: Self-Tuning Controllers Based on Pole Placement Design. Lund Institute of Technology, Dept. of Aut. Control, Report LUTFD 2 (TFRT – 3148), 1978.
- [3] V. Peterka: Subjective probability approach to real-time identification. Preprints of the 4th IFAC Symposium on Identification and System Parameter Estimation, Tbilisi, USSR, vol. 3, pp. 83–96, 1976.
- [4] V. Peterka: Bayesian system identification. *Automatica* 17 (1981), 1, 41–53.
- [5] V. Peterka: Experience accumulation for decision making in multivariate time series. *Problems Control Inform. Theory* 7 (1978), 3, 143–159.
- [6] B. Egardt: Stability of Model Reference Adaptive and Self-Tuning Regulators. Lund Institute of Technology, Dept. of Aut. Control, Report LUTFD 2 (TFRT – 1017), 1978.

Ing. Miroslav Kárný, CSc., Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation – Czechoslovak Academy of Sciences), Pod vodárenskou věží 4, 182 08 Praha 8, Czechoslovakia.