— CDC 85

C 85

Discarding data to perform more accurate system identification.*

P. Carrette[†], G. Bastin, Y. Genin and M. Gevers

CESAME, Avenue G. Lemaître 4, B-1348 Louvain-la-Neuve, Belgium
E-mail : carrette@auto.ucl.ac.be

Abstract

We present results concerning the parameter estimates obtained by prediction error methods in the case of system input signals that are insufficiently rich. Such input signals are typical of industrial measurements where occasional stepwise reference changes occur. Using singular value decomposition techniques, we propose a new data selection criterion that discards the poorly informative data in order to decrease the total mean square error (MSE) of the estimated parameters.

1 Introduction

In this note, we consider an identification problem whose objective is to estimate the input-output (I/O) dynamics of a single input single output (SISO) ARMAX system by the use of an ARX model structure whose I/O dynamics is able to represent that of the system exactly. The system and model can then be written :

$$\begin{aligned} y &= \Phi\theta_0 + \mu \\ \hat{y}(\theta) &= \Phi\theta \end{aligned} \tag{1}$$

where $y, \hat{y}(\theta) \in \mathcal{R}^N$ are the system and model output sequences, respectively, $\Phi \in \mathcal{R}^{N\times n}$ stands for the regressor matrix containing delayed input and output sequences while $\mu \in \mathcal{R}^N$ denotes a gaussian colored noise disturbance. Finally, $\theta \in \mathcal{R}^n$ is the model parameter vector that is to approximate the system parameter vector θ_0.

Under this modeling framework, it has been shown in [1] that it is useful to eliminate data from the data set used for identification because such data increase the bias errors of the estimated parameters more than they decrease their variance errors, implying an increase of the total parameter MSE. This essentially happens when the information content of the input signal is poor. In [1], we have also proposed a heuristic data selection criterion based on the minimal amount of information about the I/O system dynamics carried out by the data. More precisely, the I/O information amount of a data pair is measured by the contribution of this data pair to the evolution of the singular values of the regressor matrix, associated to the model output prediction. A data pair is discarded from the original data set when the associated I/O information amount turns to be smaller than some threshold value. The major drawback of this selection criterion is that this threshold value must be chosen *a priori*.

The purpose of this note is to propose and analyse a new dynamic data selection criterion whose performance is more robust than that obtained with a fixed threshold value. This new selection criterion is based upon a monotonic increase, along the selected data set, of the mean information amount about the I/O system dynamics. This induces a decrease of the bias of the estimated parameters which also implies an improvement of the accuracy of the model parameter estimation along the data set.

2 Statistics of the estimated parameters

The parameter estimation approach used in this paper is the classical least squares (LS) estimate of the linear model (1) : it consists of minimizing the mean square of the model prediction errors, i.e. $\epsilon(\theta) := y - \hat{y}(\theta)$, over all possible values of the parameter vector θ.

By use of singular value decomposition techniques (SVD, see [4]), we can split the Φ matrix into $\Phi = U\Sigma V^T$ where $\Sigma = \text{diag}(\sigma_1, \cdots, \sigma_n)$ is the singular value matrix with $\sigma_i^2 = \lambda_i(\Phi^T\Phi) > 0$ while U and V are left-orthogonal matrices of appropriate dimensions. With the help of the eigen-regressor matrix $\Phi_V := \Phi V$, the LS solution can then be expressed in terms of the eigen-parameter vector $\hat{\theta}_V(N) := V^T\hat{\theta}(N)$ (i.e. $= \Phi_V^+ y$)

$$\hat{\theta}_V = \theta_{0V} + \Sigma^{-1}U^T\mu \tag{2}$$

with $\theta_{0V} = V^T\theta_0$. The original vector solution can be recovered as $\hat{\theta}(N) = V\hat{\theta}_V(N)$ which is seen to consist of n independent linear combinations of the optimal eigen-parameter vector $\hat{\theta}_V(N)$.

Provided an excitation assumption about the regressor matrix Φ is satisfied (see [1]), then the first two probability moments of the eigen-parameter $\theta_{Vi}(N)$, using N data, are, respectively :

$$E\{\hat{\theta}_{Vi}(N) - \theta_{0Vi}\} \approx \alpha_i N \frac{\sigma^2}{\sigma_i^2(N)} \tag{3}$$

$$\text{Var}(\hat{\theta}_{Vi}(N)) \approx \frac{[\|\tilde{\Phi}_{uV}^{(i)}(N)\|_2^2 + \kappa_i N\sigma^2]\sigma^2}{\sigma_i^4(N)} \tag{4}$$

with

$$\sigma_i^2(N) \approx \|\Phi_{uV}^{(i)}(N)\|_2^2 + \eta_i N\sigma^2 \tag{5}$$

where α_i, η_i and κ_i are appropriate constants depending on $V^{(i)}$ and on the system disturbance, $\Phi_{uV}^{(i)}(N)$ denotes the input-dependent part of the i-th column of the eigen-regressor matrix Φ_V containing N rows and $\tilde{\Phi}_{uV}^{(i)}$ stands for a filtered version of $\Phi_{uV}^{(i)}(N)$.

Using the bias and the variance expressions from above, we may define a global measure of the accuracy of the estimation procedure in terms of the mean square error (MSE) of the estimated parameter vector, i.e. $MSE(N) := E\{\|\hat{\theta}(N) - \theta_0\|_2^2\}$. Assuming that we have a singular value $\sigma_{i_{\min}}$ significantly

*This paper presents research results of the Belgian Program on Interuniversity Poles of Attraction, initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture. The scientific responsability rests with its authors.

†Supported by a grant from IRSIA

smaller than the others (i.e. $\sigma_{i_{\min}} \ll \sigma_i$ with $i \neq i_{\min}$), the MSE is approximated by :

$$MSE(N) \approx \text{Bias}^2(\hat{\theta}_{Vi_{\min}}(N)) + \text{Var}(\hat{\theta}_{Vi_{\min}}(N)) \quad (6)$$

with $\hat{\theta}_{Vi_{\min}}$, is the *most poorly* estimated eigen-parameter (for it has the largest bias and variance).
After a detailed analysis of the contribution, i.e. $\Delta MSE(N)$, of the N-th regressor row, i.e. $\phi(N)$, to the expression (6) (see [2]), we can draw the following conclusions concerning the ability of that row to increase the estimation accuracy. If the contribution of the bias to $MSE(N-1)$ compared to that of the variance is negligible, then the new $\phi(N)$ unconditionally achieves a smaller value for the MSE. In the opposite case of negligible variance contribution to $MSE(N-1)$, then the MSE evolution is subjected to the increase of $\sigma^2_{i_{\min}}(N)/N$ due to the new $\phi(N)$. This can be viewed as a condition on the minimal amount of I/O information to be brought by the input-dependent part of the new $\phi(N)$: $|\phi^{(i)}_{uV}(N)|^2$ from (5).

3 Dynamic data selection criterion

With these results, we can proceed to a data selection procedure in order to ensure that any additional regressor row used for the parameter estimation contributes to a decrease of the MSE of the parameters. We then propose the following dynamic data selection criterion for an *a priori* given value of N_c :

Discard the N-th row of $\Phi(N)$ if

$$|\phi(N)V^{(i_{\min})}|^2 \ < \ I_{N \geq N_c} \frac{\sigma^2_{i_{\min}}(N-1)}{N-1}$$

where I_X denotes the Boolean indicator . Actually, N_c stands for the minimal number of regressor rows to be used in the estimation. This is to ensure a sufficient variance decrease of the parameter estimates through the first N_c rows of $\Phi(N)$ in order to end up with a possibly negligible contribution to the $MSE(N_c)$.

4 Simulation

The system under study is an ARMAX system with the following characteristics : the t-th row of Φ consists of $\phi(t) = [-y(t-1), u(t-1)]$, $\theta_0 = [-0.8, 0.5]^T$ and $\mu(t) = (1 + 0.8z^{-1} + 0.3z^{-2})e(t)$ with $e(t) = N(0, 0.01)$, a Gaussian white noise with z^{-1}, the delay operator.
A realisation of the identification data set $\{(u(t), y(t))\}$, consisting of data generated by a piecewise constant input signal, is given in Figure 1. The simulation results consist of the MSE of the estimated parameter vector $\hat{\theta}(N)$ as a function of the number of data used in the identification data set. These results are evaluated using Monte-Carlo simulations over 200 data set realisations. To begin with, we present in Figure 2 the MSE (i.e. $MSE(N)$ in (—)) of the estimated parameters achieved along the whole data set. We see that the $MSEi_{\min}(N)$ (– –), that is the MSE computed for the most poorly estimated eigen-parameter $\hat{\theta}_{Vi_{\min}}(N)$, almost exactly matches $MSE(N)$. We observe that the contribution of its bias (–·) is largely influenced by the step changes in the input data set while that of its variance (···) is slowly decreasing. We present in Figure 3 the MSE (i.e. $MSE(N)$ in (—)) of the parameters achieved after data selection with $N_c = 40$ as

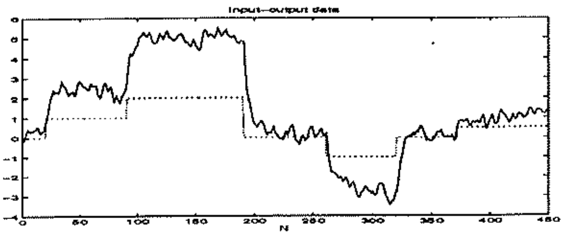

Figure 1: Input (···) and output (—) data from the simulated system.

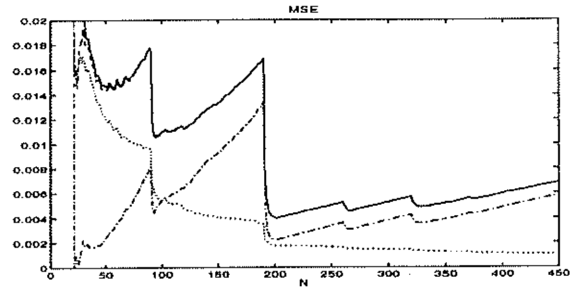

Figure 2: MSE (—) of the estimated parameters and its bias (–·) and variance (···) contributions.

a function of the number of data used in the original identification data set. We see that the contribution of the bias of the estimated parameters (–·) along the data set is reduced with respect to that in Figure 2 along the data set while that

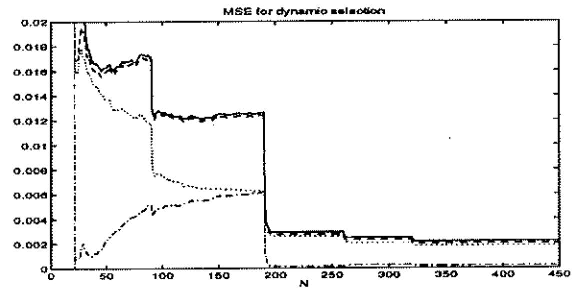

Figure 3: MSE (—) of the estimated parameters and its bias (–·) and variance (···) contributions achieved after dynamic data selection with $N_c = 40$.

of the variances (···) looks quite similar. Finally, it is worth mentioning that the achieved results are robust with respect to the N_c value (see [2]).

More details about the achieved results are found in [2] and will be given at the conference.

References

[1] P. Carrette, G. Bastin, Y. Genin and M. Gevers, *Discarding data may help in system identification!*, submitted.

[2] P. Carrette, G. Bastin, Y. Genin and M. Gevers, *Decrease of the MSE of the estimated parameters in system identification via data selection*, internal report.

[3] L. Ljung, *System Identification : theory for the user*, Prentice Hall, 1987.

[4] G.W. Stewart and J.G. Sun, *Matrix Perturbation Theory*, Computer science and Scientific computing, Academic Press, Inc. San Diego, 1990.