

Chapter 5

Identification of the Active Vibration Control Systems—The Bases

5.1 Introduction

To design an active control one needs the dynamical model of the compensator systems (from the control to be applied to the measurement of the residual acceleration or force).¹ Model identification from experimental data is a well established methodology [1, 2]. Identification of dynamic systems is an experimental approach for determining a dynamic model of a system. It includes four steps:

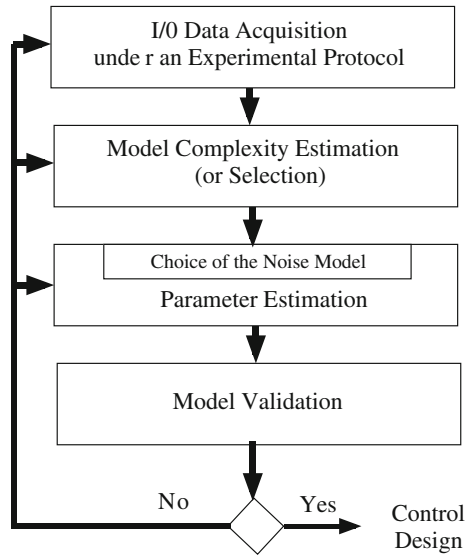
- (1) Input–output data acquisition under an experimental protocol.
- (2) Estimation of the model complexity (structure).
- (3) Estimation of the model parameters.
- (4) Validation of the identified model (structure of the model and values of the parameters).

A complete identification operation must comprise the four stages indicated above. System identification should be viewed as an iterative process as illustrated in Fig. 5.1 which has as objective to obtain a model which passes the model validation test and then can be used safely for control design.

The typical input excitation sequence is a PRBS (pseudorandom binary sequence). The type of model which will be identified is a discrete-time parametric model allowing later to directly design a control algorithm straightforwardly implementable on a computer. Model validation is the final key point. The estimation of the model parameters is done in a noisy environment. It is important to emphasize that it does not exist one single algorithm that can provide in all the cases a good model (i.e., which passes the model validation tests). Therefore, the appropriate algorithm which allows to obtain a model which passes the validation tests has to be used.

¹Linear feedback regulator design will require also the model of the disturbance. Linear feedforward compensator design will require in addition a model of the primary path. Design of adaptive regulators or of feedforward compensators require only the model of the secondary path.

Fig. 5.1 The identification methodology



In what follows, we would like to summarize some of the basic facts in system identification. For a detailed coverage of the subject, see [1, 3].

Figure 5.2 shows the principle of parameter estimation of a discrete-time model. An adjustable model of the discretized plant is built. Its parameters are driven by a parameter adaptation algorithm such that the prediction error (the difference between the true output and the predicted output by the model) is minimized in the sense of a certain criterion.

Several assumptions are implicitly made when one uses this approach

- the order of the discrete-time model representing the system is known;
- in the absence of noise the adaptation algorithm will drive the prediction error towards zero;

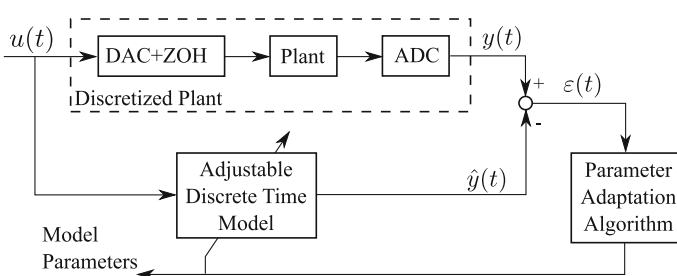


Fig. 5.2 Principle of model parameters estimation

- in the presence of noise, the estimated parameters will be asymptotically unbiased² (i.e., the noise does not influence asymptotically the precision of the parameter estimation); and
- the input to the system (the testing signal) is such that null prediction error implies null parameter errors (persistent excitation property).

The various steps indicated in Fig. 5.1 tries to assure that the parameter estimation algorithm will provide the good parameter estimates.

5.2 Input–Output Data Acquisition and Preprocessing

5.2.1 *Input–Output Data Acquisition Under an Experimental Protocol*

The experimental protocol should assure persistent excitation for the number of parameters to be estimated. It can be shown (see Chap. 4, Sect. 4.5 and [1]) that for identifying $2n$ parameters, the excitation signal should contain at least $n + 1$ sinusoids of distinct frequencies. To go beyond this constraints one usually uses Pseudorandom Binary Sequences (PRBS) since they contain a large number of sinusoids with energy equally distributed over the frequency domain. In addition the magnitude of the signal is constant allowing and easy selection with respect to the magnitude constraints on the plant input.

5.2.2 *Pseudorandom Binary Sequences (PRBS)*

Pseudorandom binary sequences are sequences of rectangular pulses, modulated in width, that approximate a discrete-time white noise and thus have a spectral content *rich* in frequencies. They owe their name *pseudo random* to the fact that they are characterized by a *sequence length* within which the variations in pulse width vary randomly, but that over a large time horizon, they are periodic, the period being defined by the length of the sequence. In the practice of system identification, one generally uses just one complete sequence and we should examine the properties of such a sequence.

The PRBS are generated by means of shift registers with feedback (implemented in hardware or software).³ The maximum length of a sequence is $L = 2^N - 1$, where N is the number of cells of the shift register.

²The parameter estimation error induced by the measurement noise is called “bias”.

³Routines for generating PRBS can be downloaded from the websites: <http://www.landau-adaptivecontrol.org> and <http://www.gipsa-lab.grenoble-inp.fr/~ioandore.landau/identificationandcontrol/>.

Fig. 5.3 Generation of a PRBS of length $2^5 - 1 = 31$ sampling periods

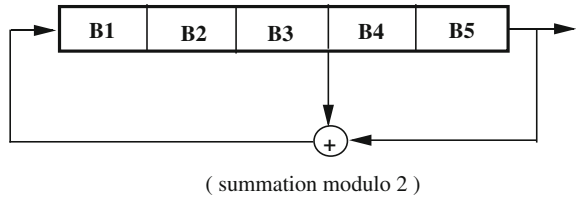


Table 5.1 Generation of maximum length PRBS

Number of cells N	Sequence length $L = 2^N - 1$	Bits added B_i and B_j
5	31	3 and 5
6	63	5 and 6
7	127	4 and 7
8	255	2, 3, 4 and 8
9	511	5 and 9
10	1023	7 and 10

Figure 5.3 presents the generation of a PRBS of length $31 = 2^5 - 1$ obtained by means of a $N = 5$ -cells shift register. Note that at least one of the N cells of the shift register should have an initial logic value different from zero (one generally takes all the initial values of the N cells equal to the logic value 1).

Table 5.1 gives the structure enabling maximum length PRBS to be generated for different numbers of cells. Note also a very important characteristic element of the PRBS: *the maximum duration of a PRBS impulse is equal to N (number of cells)*. This property is to be considered when choosing a PRBS for system identification.⁴

In order to cover the entire frequency spectrum generated by a particular PRBS, the length of a test must be at least equal to the length of the sequence. In a large number of cases, the duration of the test L is chosen equal to the length of the sequence. Through the use of a frequency divider for the clock frequency of the PRBS, it is possible to shape the energy distribution in the frequency domain. This is why, in a large number of practical situations, a submultiple of the sampling frequency is chosen as the clock frequency for the PRBS. Note that dividing the clock frequency of the PRBS will reduce the frequency range corresponding to a constant spectral density in the high frequencies while augmenting the spectral density in the low frequencies. In general, this will not affect the quality of identification, either because in many cases when this solution is considered, the plant to be identified has a low band pass or because the effect or the reduction of the signal/noise ratio at high frequencies can be compensated by the use of appropriate identification techniques; however, it is recommended to choose $p \leq 4$ where p is the frequency divider.

⁴Functions *prbs.m* and *prbs.c* available on the websites: <http://www.landau-adaptivecontrol.org> and [http://www.gipsa-lab.grenoble-inp.fr/ ioandore.landau/identificationandcontrol/](http://www.gipsa-lab.grenoble-inp.fr/ioandore.landau/identificationandcontrol/) allow to generate PRBS of various lengths and magnitudes.

Fig. 5.4 Spectral density of a PRBS sequence,
a $N = 8, p = 1,$
b $N = 8, p = 2,$
c $N = 8, p = 3$

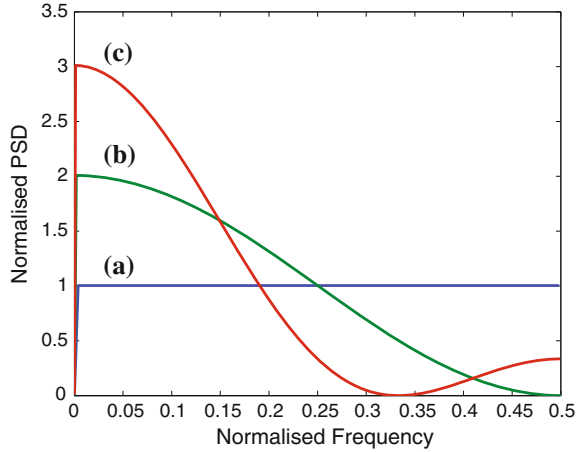


Figure 5.4 shows the spectral density of PRBS sequences generated with $N = 8$ for $p = 1, 2, 3$. As one can see, the energy of the spectrum is reduced in the high frequencies and augmented in the lower frequencies. Furthermore, for $p = 3$ a hole occurs at $f_s/3$.

Until now, we have been concerned only with the choice of the length and clock frequency of the PRBS; however, the magnitude of the PRBS must also be considered. Although the magnitude of the PRBS may be very low, it should lead to output variations larger than the residual noise level. If the signal/noise ratio is too low, the length of the test must be augmented in order to obtain a satisfactory parameter estimation.

Note that in a large number of applications, the significant increase in the PRBS level may be undesirable in view of the nonlinear character of the plants to be identified (we are concerned with the identification of a linear model around an operating point).

5.2.3 Data Preprocessing

The first aspect is that one works with centred data (variations of the real data) so that the first operation is the centering of the input/output data by subtracting their mean value.

When identifying the compensator system in active vibration control systems, one has to take into account the double differentiator behaviour. This means that a part of the model is known and we should identify only the unknown part. To do this, the input applied to the real system is filtered by a double discrete-time differentiator filter

$$(1 - q^{-1})^2 = 1 - 2q^{-1} + q^{-2} \tag{5.1}$$

This new input/output sequence is then centred and used together with the measured output data for identifying the unknown part of the model. After the unknown part of the model will be identified, the double differentiator should be included in the final model (the two transfer operators will be multiplied).

5.3 Model Order Estimation from Data

It is extremely important to be able to estimate the order of the system from input/output data since it is hard from physical reasoning to get a reliable estimation of the order of the system. To introduce the problem of order estimation from data, we will start with an example. Assume that the plant model can be described by:

$$y(t) = -a_1 y(t-1) + b_1 u(t-1) \quad (5.2)$$

and that the data are noise free. The order of this model is $n = n_A = n_B = 1$.

Question: Is there any way to test from data if the order assumption is correct? To do so, construct the following matrix:

$$\begin{bmatrix} y(t) & \vdots & y(t-1) & u(t-1) \\ y(t-1) & \vdots & y(t-2) & u(t-2) \\ y(t-2) & \vdots & y(t-3) & u(t-3) \end{bmatrix} = \begin{bmatrix} Y(t) & \vdots & R(1) \end{bmatrix} \quad (5.3)$$

Clearly, if the order of the model given in Eq. (5.2) is correct, the vector $Y(t)$ will be a linear combination of the columns of $R(1)$ ($Y(t) = R(1)\theta$ with $\theta^T = [-a_1, b_1]$) and the rank of the matrix will be 2 (instead of 3). If the plant model is of order 2 or higher, the matrix in (5.3) will be full rank. Of course, this procedure can be extended for testing the order of a model by testing the rank of the matrix $[Y(t), R(\hat{n})]$ where:

$$R(\hat{n}) = [Y(t-1), U(t-1), Y(t-2), U(t-2) \dots Y(t-\hat{n}), U(t-\hat{n})], \quad (5.4)$$

$$Y^T(t) = [y(t), y(t-1) \dots], \quad U^T(t) = [u(t), u(t-1) \dots]. \quad (5.5)$$

Unfortunately, as a consequence of the presence of noise, this procedure cannot directly be applied in practice.

A more practical approach results from the observation that the rank test problem can be approached by the searching of $\hat{\theta}$ which minimizes the following criterion for an estimated value of the order \hat{n} .

$$V_{LS}(\hat{n}, N) = \min_{\hat{\theta}} \frac{1}{N} \|Y(t) - R(\hat{n})\hat{\theta}\|^2 \quad (5.6)$$

where N is the number of samples. But this criterion is nothing else than an equivalent formulation of the least squares [4]. If the conditions for unbiased estimation using least squares are fulfilled, (5.6) is an efficient way for assessing the order of the model since $V_{LS}(\hat{n}) - V_{LS}(\hat{n} + 1) \rightarrow 0$ when $\hat{n} \geq n$.

In the meantime, the objective of the identification is to estimate lower order models (parsimony principle) and therefore, it is reasonable to add in the criterion (5.6) a term which penalizes the complexity of the model. Therefore, the *penalized criterion* for order estimation will take the form:

$$J_{LS}(\hat{n}, N) = V_{LS}(\hat{n}, N) + S(\hat{n}, N) \quad (5.7)$$

where typically

$$S(\hat{n}, N) = 2\hat{n}X(N) \quad (5.8)$$

and V_{LS} represents the *non penalized criterion*. $X(N)$ in (5.8) is a function that decreases with N . For example, in the so called $BIC_{LS}(\hat{n}, N)$ criterion, $X(N) = \frac{\log N}{N}$ (other choices are possible—see [3–5]) and the order \hat{n} is selected as the one which minimizes J_{LS} given by (5.7). Unfortunately, the results are unsatisfactory in practice because in the majority of situations, the conditions for unbiased parameter estimation using least squares are not fulfilled.

In [5, 6], it is proposed to replace the matrix $R(\hat{n})$ by an instrumental variable matrix $Z(\hat{n})$ whose elements will not be correlated with the measurement noise. Such an instrumental matrix $Z(\hat{n})$ can be obtained by replacing in the matrix $R(\hat{n})$, the columns $Y(t - 1)$, $Y(t - 2)$, $Y(t - 3)$ by delayed version of $U(t - L - i)$, i.e., where $L > n$:

$$Z(\hat{n}) = [U(t - L - 1), U(t - 1), U(t - L - 2), U(t - 2) \dots] \quad (5.9)$$

and therefore, the following criterion is used for the order estimation:

$$J_{IV}(\hat{n}, N) = \min_{\hat{\theta}} \frac{1}{N} \|Y(t) - Z(\hat{n})\hat{\theta}\|^2 + \frac{2\hat{n}\log N}{N} \quad (5.10)$$

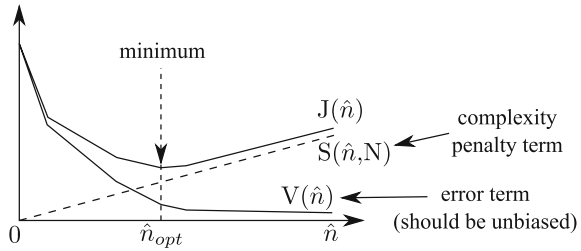
and

$$\hat{n} = \min_{\hat{n}} J_{IV}(\hat{n}). \quad (5.11)$$

A typical curve of the evolution of the criterion (5.10) as a function of \hat{n} is shown in Fig. 5.5.

It is shown in [5] that using this criterion a consistent estimate of the order \hat{n} is obtained under mild noise conditions (i.e., $\lim_{N \rightarrow \infty} Pr(\hat{n} = n) = 1$ where Pr denotes the probability). Comparisons with other order estimation criteria are also provided in this reference.

Fig. 5.5 Evaluation of the criterion for order estimation



Once an estimated order \hat{n} is selected, one can apply a similar procedure to estimate $\hat{n}_A, \hat{n} - \hat{d}, \hat{n}_B + \hat{d}$, from which \hat{n}_A, \hat{n}_B and \hat{d} are obtained.⁵

5.4 Parameter Estimation Algorithms

The algorithms which will be used for parameter estimation will depend on the assumptions made on the noise disturbing the measurements, assumptions which have to be confirmed by the model validation.

It is important to emphasize that no one single *plant + noise* structure can describe all the situations encountered in practice. Furthermore, there is no a unique parameter estimation algorithm that may be used with all possible plant + noise structures such that the estimated parameters are always unbiased. The most typical structures for *plant + noise* are shown in Fig. 5.6.

The various “plant + noise” models shown in Fig. 5.6 can be described by:

$$y(t) = \frac{q^{-d}B(q^{-1})}{A(q^{-1})}u(t) + \eta(t) \tag{5.12}$$

For structure S1 one has:

$$\eta(t) = \frac{1}{A(q^{-1})}e(t) \tag{5.13}$$

where $e(t)$ is a discrete-time Gaussian white noise (zero mean and standard deviation σ).

For structure S2 one has:

$$\eta(t) = w(t) \tag{5.14}$$

a centred noise of finite power and uncorrelated with the input $u(t)$.

⁵Routines corresponding to this method in MATLAB (*estorderiv.m*) and Scilab (*estorderiv.sci*) can be downloaded from the websites: <http://www.landau-adaptivecontrol.org> and <http://www.gipsa-lab.grenoble-inp.fr/ioandore.landau/identificationandcontrol/>.

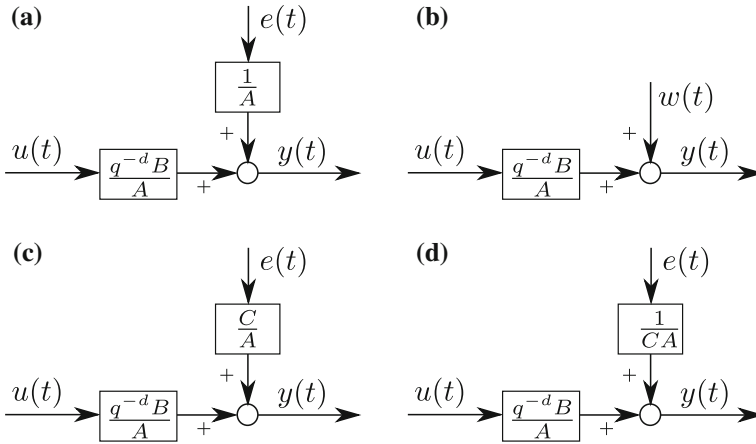


Fig. 5.6 Structure of the “plant + noise” models. **a S1:** $A(q^{-1})y(t) = q^{-d}B(q^{-1})u(t) + e(t)$. **b S2:** $A(q^{-1})y(t) = q^{-d}B(q^{-1})u(t) + A(q^{-1})w(t)$. **c S3:** $A(q^{-1})y(t) = q^{-d}B(q^{-1})u(t) + C(q^{-1})e(t)$. **d S4:** $A(q^{-1})y(t) = q^{-d}B(q^{-1})u(t) + (1/C(q^{-1}))e(t)$

For structure S3 one has

$$\eta(t) = \frac{C(q^{-1})}{A(q^{-1})} e(t) \tag{5.15}$$

and for structure S4 one has:

$$\eta(t) = \frac{1}{C(q^{-1})A(q^{-1})} e(t) \tag{5.16}$$

Based on the experience of the authors in identifying active vibration control systems, one can say that in most of the situations they are represented correctly by ARMAX models (structure S3). Therefore, most likely, algorithms for estimating parameters for ARMAX models will provide good results (should be confirmed by model validation). The simplest and in general most efficient algorithms for identifying active vibration control systems are the “recursive extended least squares” and the “output error with extend predictor model”.⁶ Details on these two type of algorithms will be given next. Nevertheless, there is no guarantee that ARMAX representation is the good one for all possible configuration which can be encountered in practice. Therefore one has to be prepared to use also other parameter estimation algorithms if the validation of the identified models using the above mentioned algorithms fails.⁷

⁶Routines for these algorithms can be downloaded from the websites: <http://www.landau-adaptivecontrol.org> and <http://www.gipsa-lab.grenoble-inp.fr/~ioandore.landau/identificationandcontrol/>.

⁷The interactive stand alone software iReg (<http://tudor-bogdan.airimitoiaie.name/ireg.html>) provides parameter estimations algorithms for all the mentioned “plant + noise” structures as well as

All the recursive parameter estimation algorithms use the same parameter adaptation algorithm:

$$\hat{\theta}(t + 1) = \hat{\theta}(t) + F(t)\Phi(t)v(t + 1) \tag{5.17}$$

$$F(t + 1)^{-1} = \lambda_1(t)F(t)^{-1} + \lambda_2(t)\Phi(t)\Phi^T(t) \tag{5.18}$$

$$\begin{aligned} 0 < \lambda_1(t) \leq 1; \quad 0 \leq \lambda_2(t) < 2; \quad F(0) > 0 \\ F^{-1}(t) > \alpha F^{-1}(0); \quad 0 < \alpha < \infty \\ v(t + 1) = \frac{v^\circ(t + 1)}{1 + \Phi^T(t)F(t)\Phi(t)} \end{aligned} \tag{5.19}$$

What changes from an identification algorithm to another is:

- the structure of the adjustable predictor;
- how the adaptation error is generated;
- how the regressor vector is generated;
- how the adaptation error is generated; and
- the size of the adjustable parameter vector (the number of parameters).

The various options for the selection of the time profile of the adaptation gain $F(t)$ in (5.19) have been discussed in Sect.4.3.4. For system identification of a linear time invariant models, a decreasing adaptation gain type algorithm will be used or an algorithm with variable forgetting factor. We will present next the “recursive extended least squares” and the “output error with extended predictor”.

5.4.1 Recursive Extended Least Squares (RELS)

This method has been developed in order to identify without bias *plant + noise* models of the form (ARMAX model):

$$A(q^{-1})y(t) = q^{-d}B(q^{-1})u(t) + C(q^{-1})e(t) \tag{5.20}$$

The idea is to simultaneously identify the plant model and the noise model, in order to obtain a prediction (adaptation) error which is asymptotically white.

The model generating the data can be expressed as:

$$\begin{aligned} y(t + 1) &= -A^*(q^{-1})y(t) + B^*(q^{-1})u(t - d) + C^*(q^{-1})e(t) + e(t + 1) \\ &= \theta^T\phi_0(t) + e(t + 1) \end{aligned} \tag{5.21}$$

(Footnote 7 continued)

an automated identification procedure covering all the stages of system identification. It has been extensively used for identification of active vibration control systems.

where:

$$\theta^T = [a_1 \dots a_{n_A}, b_1 \dots b_{n_B}, c_1 \dots c_{n_C}] \quad (5.22)$$

$$\begin{aligned} \phi_0^T(t) = & [-y(t) \dots -y(t - n_A + 1), u(t - d) \dots u(t - d - n_B + 1), \\ & e(t) \dots e(t - n_C + 1)] \end{aligned} \quad (5.23)$$

Assume that the parameters are known and construct a predictor that will give a white prediction error:

$$\hat{y}(t + 1) = -A^*(q^{-1})y(t) + B^*(q^{-1})u(t - d) + C^*(q^{-1})e(t) \quad (5.24)$$

Furthermore, this predictor will minimize $E\{[y(t + 1) - \hat{y}(t + 1)]^2\}$ [1].

The prediction error, in the case of known parameters, is given by:

$$\varepsilon(t + 1) = y(t + 1) - \hat{y}(t + 1) = e(t + 1) \quad (5.25)$$

This allows rewriting Eq. (5.24) in the form:

$$\hat{y}(t + 1) = -A^*(q^{-1})y(t) + B^*(q^{-1})u(t - d) + C^*(q^{-1})\varepsilon(t) \quad (5.26)$$

Subtracting now (5.26) from (5.21), one gets:

$$\varepsilon(t + 1) = -C^*(q^{-1})[\varepsilon(t) - e(t)] + e(t) \quad (5.27)$$

i.e.,

$$C(q^{-1})[\varepsilon(t + 1) - e(t + 1)] = 0 \quad (5.28)$$

Since $C(q^{-1})$ is an asymptotically stable polynomial, it results that $\varepsilon(t + 1)$ will become white asymptotically.

The adaptive version of this predictor is as follows. The *a priori* adjustable predictor will take the form:

$$\hat{y}^\circ(t + 1) = -\hat{A}^*(q^{-1}, t)y(t) + \hat{B}^*(q^{-1}, t)u(t) + \hat{C}^*(q^{-1}, t)\varepsilon(t) = \hat{\theta}^T(t)\phi(t) \quad (5.29)$$

in which:

$$\hat{\theta}^T = [\hat{a}_1(t) \dots \hat{a}_{n_A}(t), \hat{b}_1(t) \dots \hat{b}_{n_B}(t), \hat{c}_1(t) \dots \hat{c}_{n_C}(t)] \quad (5.30)$$

$$\begin{aligned} \phi^T(t) = & [-y(t) \dots -y(t - n_A + 1), u(t - d) \dots u(t - d - n_B + 1), \\ & \varepsilon(t) \dots \varepsilon(t - n_C + 1)] \end{aligned} \quad (5.31)$$

The *a posteriori* adjustable predictor will be given by:

$$\hat{y}(t + 1) = \hat{\theta}^T(t + 1)\phi(t) \quad (5.32)$$

The *a posteriori* prediction error $\varepsilon(t)$ which enters in the observation vector of the predictor is given by

$$\varepsilon(t) = y(t) - \hat{y}(t) \quad (5.33)$$

(where $\hat{y}(t)$ is now the *a posteriori* output of the adjustable predictor) and the *a priori* prediction error is given by:

$$\varepsilon^\circ(t+1) = y(t+1) - \hat{y}^\circ(t+1) \quad (5.34)$$

The *a posteriori* prediction equation is obtained subtracting (5.32) from (5.21) and observing that (5.21) can be alternatively expressed as:

$$y(t+1) = \theta^T \phi(t) - C^*(q^{-1})\varepsilon(t) + C(q^{-1})e(t) \quad (5.35)$$

(by adding and subtracting the term $\pm C^*(q^{-1})\varepsilon(t)$). One obtains:

$$\varepsilon(t+1) = -C^*(q^{-1})\varepsilon(t) + \left[\theta - \hat{\theta}(t+1) \right]^T \phi(t) + C(q^{-1})e(t) \quad (5.36)$$

from which it results that:

$$\varepsilon(t+1) = \frac{1}{C(q^{-1})} \left[\theta - \hat{\theta}(t+1) \right]^T \phi(t) + e(t) \quad (5.37)$$

In the deterministic case $C(q^{-1}) = 1$, $e(t) \equiv 0$ and it can be seen that (5.37) has the appropriate format corresponding to Theorem 4.1 given in Chap. 4. One immediately concludes, using the PAA given in (5.17) through (5.19), with $\Phi(t) = \phi(t)$, $v(t) = \varepsilon(t)$, and $v^\circ(t) = \varepsilon^\circ(t)$ that, in the deterministic case, global asymptotic stability is assured without any positive real condition. In stochastic environment, either using ODE or martingales, it can be shown [7] that the convergence is assured provided that (sufficient condition):

$$H'(z^{-1}) = \frac{1}{C(z^{-1})} - \frac{\lambda_2}{2} \quad (5.38)$$

is a strictly positive real transfer function for $2 > \lambda_2 \geq \max_t \lambda_2(t)$.

5.4.2 Output Error with Extended Prediction Model (XOLOE)

This algorithm can be used for identification of *plant + noise* models of the ARMAX form. It has been developed initially with the aim to remove the *positive real condition* required by the *output error* algorithm. It turns out that the XOLOE can be interpreted

as a variant of the ELS. To see this, consider the *a priori* output of the adjustable predictor for ELS (5.29), which can be rewritten as follows by adding and subtracting the term $\pm \hat{A}^*(q^{-1}, t)\hat{y}(t)$:

$$\begin{aligned} \hat{y}^\circ(t+1) &= -\hat{A}^*(q^{-1}, t)\hat{y}(t) + \hat{B}^*(q^{-1}, t)u(t-d) \\ &\quad + \left[\hat{C}^*(q^{-1}, t)\varepsilon(t) - \hat{A}^*(q^{-1}, t)[y(t) - \hat{y}(t)] \right] \end{aligned} \quad (5.39)$$

Defining:

$$\hat{H}^*(q^{-1}, t) = \hat{C}^*(q^{-1}, t) - \hat{A}^*(q^{-1}, t) = \hat{h}_1(t) + q^{-1}\hat{h}_2(t) + \dots$$

with:

$$\hat{h}_i(t) = \hat{c}_i(t) - \hat{a}_i(t); \quad i = 1, 2, \dots, \max(n_A, n_C)$$

Equation (5.39) can be rewritten as:

$$\hat{y}^\circ(t+1) = -\hat{A}^*(q^{-1}, t)\hat{y}(t) + \hat{B}^*(q^{-1}, t)u(t-d) + \hat{H}^*(q^{-1}, t)\varepsilon(t) \quad (5.40)$$

$$= \hat{\theta}^T(t)\phi(t) \quad (5.41)$$

where now:

$$\begin{aligned} \hat{\theta}^T(t) &= \left[\hat{a}_1(t) \dots \hat{a}_{n_A}(t), \hat{b}_1(t) \dots \hat{b}_{n_B}(t), \hat{h}_1(t) \dots \hat{h}_{n_H}(t) \right] \\ \phi^T(t) &= [-\hat{y}(t), \dots, \hat{y}(t - n_A + 1), u(t-d) \dots u(t-d - n_B + 1), \\ &\quad \varepsilon(t) \dots \varepsilon(t - n_C + 1)] \end{aligned}$$

Equation (5.40) corresponds to the adjustable predictor for the output error with extended prediction model. One immediately concludes, using the PAA given in (5.17)–(5.19), with $\Phi(t) = \phi(t)$, $\nu(t) = \varepsilon(t)$, and $\nu^\circ(t) = \varepsilon^\circ(t)$ (defined in Eqs. (5.33) and (5.34), respectively) that, in the deterministic case, global asymptotic stability is assured without any positive real condition. In the stochastic context, one has the (sufficient) convergence condition: $H'(z^{-1}) = \frac{1}{C(z^{-1})} - \frac{\lambda_2}{2}$ should be SPR ($2 > \lambda_2 \geq \max_t \lambda_2(t)$) similar to that for ELS.

Despite their similar asymptotic properties, there is a difference in the first n_A components of the vector $\phi(t)$. While the RELS algorithm uses the measurements $y(t)$, $y(t-1)$, \dots directly affected by the noise, the XOLOE algorithm uses the predicted *a posteriori* outputs $\hat{y}(t)$, $\hat{y}(t-1)$ which depend upon the noise only indirectly through the PAA. This explains why a better estimation is obtained with XOLOE than with RELS over short or medium time horizons (it removes the bias more quickly).

5.5 Validation of the Identified Models

The identification methods considered above (*recursive extended least squares* and *output error with extended predictor*) belongs to the class of methods based on the whitening of the residual errors, i.e., the identified ARMAX predictor is an optimal predictor if the residual error is a white noise. If the residual prediction error is a white noise sequence, in addition to obtaining unbiased parameter estimates, this also means that the identified model gives the best prediction for the plant output in the sense that it minimizes the variance of the prediction error. On the other hand, since the residual error is white and a white noise is not correlated with any other variable, then all the correlations between the input and the output of the plant are represented by the identified model and what remains unmodelled does not depend on the input.

The principle of the validation method is as follows:

- If the plant + noise structure chosen is correct, i.e., representative of reality.
- If an appropriate parameter estimation method for the structure chosen has been used.
- If the orders of the polynomials $A(q^{-1})$, $B(q^{-1})$, $C(q^{-1})$ and the value of d (delay) have been correctly chosen (the plant model is in the model set).

Then the prediction error $\varepsilon(t)$ asymptotically tends toward a white noise, which implies:

$$\lim_{t \rightarrow \infty} E\{\varepsilon(t)\varepsilon(t-i)\} = 0; \quad i = 1, 2, 3 \dots; -1, -2, -3 \dots$$

The validation method implements this principle.⁸ It is made up of several steps:

- (1) Creation of an I/O file for the identified model (using the same input sequence as for the system).
- (2) Creation of a residual prediction error file for the identified model.
- (3) *Whiteness* (uncorrelatedness) test on the residual prediction errors sequence.

5.5.1 Whiteness Test

Let $\{\varepsilon(t)\}$ be the centred sequence of the residual prediction errors (centred: measured value–mean value). One computes:

⁸Routines corresponding to this validation method in MATLAB and Scilab can be downloaded from the websites: <http://www.landau-adaptivecontrol.org> and <http://www.gipsa-lab.grenoble-inp.fr/ioandore.landau/identificationandcontrol/>.

$$R(0) = \frac{1}{N} \sum_{t=1}^N \varepsilon^2(t), \quad RN(0) = \frac{R(0)}{R(0)} = 1, \quad (5.42)$$

$$R(i) = \frac{1}{N} \sum_{t=1}^N \varepsilon(t)\varepsilon(t-i), \quad RN(i) = \frac{R(i)}{R(0)}, \quad i = 1, 2, 3, \dots, n_A, \dots \quad (5.43)$$

with $i_{\max} \geq n_A$ (degree of polynomial $A(q^{-1})$), which are estimations of the (normalized) autocorrelations.

If the residual prediction error sequence is perfectly white (theoretical situation), and the number of samples is very large ($N \rightarrow \infty$), then $RN(0) = 1$, $RN(i) = 0$, $i \geq 1$.⁹

In real situations, however, this is never the case (i.e., $RN(i) \neq 0$; $i \geq 1$), since on the one hand, $\varepsilon(t)$ contains residual structural errors (order errors, nonlinear effects, non-Gaussian noises), and on the other hand, the number of samples may be relatively small in some cases. Also, it should be kept in mind that one always seeks to identify *good* simple models (with few parameters).

One considers as a practical validation criterion (extensively tested on applications):

$$RN(0) = 1; \quad |RN(i)| \leq \frac{2.17}{\sqrt{N}}; \quad i \geq 1 \quad (5.44)$$

where N is the number of samples.

This test has been defined taking into account the fact that for a white noise sequence $RN(i)$, $i \neq 0$ has an asymptotically Gaussian (normal) distribution with zero mean and standard deviation:

$$\sigma = \frac{1}{\sqrt{N}}$$

The confidence interval considered in (5.44) corresponds to a 3 % level of significance of the hypothesis test for Gaussian distribution.

If $RN(i)$ obeys the Gaussian distribution $(0, 1/\sqrt{N})$, there is only a probability of 1.5 % that $RN(i)$ is larger than $2.17/\sqrt{N}$, or that $RN(i)$ is smaller than $-2.17/\sqrt{N}$. Therefore, if a computed value $RN(i)$ falls outside the range of the confidence interval, the hypothesis $\varepsilon(t)$ and $\varepsilon(t-i)$ are independent should be rejected, i.e., $\{\varepsilon(t)\}$ is not a white noise sequence.

The following remarks are important:

- If several identified models have the same complexity (number of parameters), one chooses the model given by the methods that lead to the smallest $|RN(i)|$;
- A *too good* validation criterion indicates that model simplifications may be possible.

⁹Conversely, for Gaussian data, uncorrelation implies independence. In this case, $RN(i) = 0$, $i \geq 1$ implies independence between $\varepsilon(t)$, $\varepsilon(t-1)$, ..., i.e., the sequence of residuals $\{\varepsilon(t)\}$ is a Gaussian white noise.

- To a certain extent, taking into account the relative weight of various non-Gaussian and modelling errors (which increases with the number of samples), the validation criterion may be slightly tightened for small N and slightly relaxed for large N . Therefore, for simplicity's sake, one can consider as a basic practical numerical value for the validation criterion value:

$$|RN(i)| \leq 0.15; \quad i \geq 1.$$

Note also that a complete model validation implies, after the validation using the input/output sequence used for identification, a validation using a plant input/output sequence other than the one used for identification.

5.6 Concluding Remarks

Basic elements for the identification of discrete-time models for dynamical systems have been laid down in this chapter. The following facts have to be emphasized:

1. System identification includes four basic steps:
 - input/output data acquisition under an experimental protocol;
 - estimation or selection of the model complexity;
 - estimation of the model parameters; and
 - validation of the identified model (structure of the model and values of parameters).

This procedure has to be repeated (with appropriate changes at each step) if the validation of the model fails.
2. Recursive or off-line parameter estimation algorithms can be used for identification of the plant model parameters.
3. The various recursive parameter estimation algorithms use the same structure for the PAA. They differ from each other in the following ways:
 - structure of the adjustable predictor;
 - nature of the components of the observation vector; and
 - the way in which the adaptation error is generated.
4. The stochastic noises, which contaminate the measured output, may cause errors in the parameter estimates (bias). For a specific type of noise, appropriate recursive identification algorithms providing asymptotically unbiased estimates are available.
5. A unique *plant + noise* model structure that describes all the situations encountered in practice does not exist, nor is there a unique identification method providing satisfactory parameter estimates (unbiased estimates) in all situations.

5.7 Notes and References

A more detailed discussion of the subject following the same pathway can be found in [1]. The associated website: <http://www.gipsa-lab.grenoble-inp.fr/~ioandore.landau/identificationandcontrol/> provides MATLAB and scilab functions for system identification as well as simulated and real input/output data for training.

For a general coverage of system identification see [3, 4].

References

1. Landau I, Zito G (2005) Digital control systems - design, identification and implementation. Springer, London
2. Ljung L, Söderström T (1983) Theory and practice of recursive identification. The MIT Press, Cambridge
3. Ljung L (1999) System identification - theory for the user, 2nd edn. Prentice Hall, Englewood Cliffs
4. Soderstrom, T., Stoica, P.: System Identification. Prentice Hall (1989)
5. Duong HN, Landau ID (1996) An IV based criterion for model order selection. *Automatica* 32(6):909–914
6. Duong HN, Landau I (1994) On statistical properties of a test for model structure selection using the extended instrumental variable approach. *IEEE Trans Autom Control* 39(1):211–215. doi:[10.1109/9.273371](https://doi.org/10.1109/9.273371)
7. Landau ID, Lozano R, M'Saad M, Karimi A (2011) Adaptive control, 2nd edn. Springer, London