# ON THE APPROXIMATION OF TIME-VARYING STOCHASTIC SYSTEMS

R. Genesio and R. Pomé CENS - Politecnico di Torino Istituto Elettrotecnico Nazionale Galileo Ferraris

Torino, Italy

### Abstract

A linear time-varying stochastic system described in terms of input-output data corrupted by noise is given and an optimal, time-invariant, low-order approximating model is required. After the problem statement, the paper introduces an input-independent criterion and then considers the problem of its evaluation from the available data. A procedure is developed in order to obtain in closed form the upper bound, corresponding to a given level of probability, of the error functional. Finally, the minimization of this quantity leads to the optimal model parameters and to the approximation measure.

### 1. - Introduction

The problem of modelling a high-order linear differential system by means of a low-order reduced model has been recently studied by several authors both for the analysis and for the design of control systems. (A complete set of references on this subject is given in [1].) Most of the proposed techniques attempt to approx imate typical responses of the system (impulse or step responses, transfer function, z-transfer function, time-moments, etc.) in terms of the corresponding responses of the model (see for example [2] [3] [4]). An alternate approach is gi ven by the uniform approximation methods which essentially minimize a worst case error (see for example [5] [6] [7]). Usually the proposed procedures concern linear time-invariant systems and a relatively small attention has been devoted to the case where a linear time-varying system has to be modelled by a linear time--invariant one of reduced order [5] [8] [9]. In these references the given system is assumed to be described in exact form by its state equations.

This paper considers the approximation by a linear time-invariant model of low-order of a time-varying stochastic linear system from input-output data corrupted by noise. This problem, which is of some interest in communication field  $\boxed{10}$ ,  $\boxed{11}$ ,  $\boxed{12}$ ,  $\boxed{13}$ , is approached from the point of view of the uniform approximation. After the problem statement a modelling criterion is defined and a method

721

is given in order to determine the optimal model from the real data which describe the system.

### 2. - Problem statement

Consider the linear, time-varying, dynamic system described by the relation

$$y(t) = \int_{0}^{\infty} g(t, s) u(t-s) ds \qquad (2.1)$$

where u and y represent input and output respectively. The system is stochastic since g is a real-valued stochastic process.

In the measurement situation the input system u is exactly determinable while the observed output is given by

$$z(t) = y(t) + d(t)$$
 (2.2)

i.e. the true output of the system is corrupted by an additive random noise d.

It is desired to determine a linear time-invariant model of low-order described by the input-output relation

$$\hat{\mathbf{y}}(t) = \int_{0}^{\infty} \hat{\mathbf{g}}(s) \, \mathbf{u}(t-s) \, \mathrm{d}s \qquad (2.3)$$

(where  $\hat{g}$  is a deterministic function) which gives the "best" approximation of the above system according to a certain criterion (see section 3).

To this end some assumptions are made:

# a) <u>on the system</u>

g is a real-valued, Gaussian, stochastic process for which the following conditions hold:

$$g(t, s) = 0$$
 for  $s < 0$  (2.4)

except on a set of sample functions of g of probability zero;

$$E \iint_{\Lambda} g(t, s)^2 dt ds < \infty$$
 (2.5)

(E means expected value) for any bounded measurable set  $\Lambda$  in the plane;

$$E g(t, s) = g_0(s)$$
 for  $s \ge 0$  (2.6)

and

$$E\left[g(t, s) g(t', s')\right] = \varphi_{g}(t-t'; s, s') + g_{O}(s) g_{O}(s')$$
(2.7)

The impulse response go is unknown. It is stable, sufficiently smooth and a time

 $T_s$  is known at which  $g_0(t)$  can be considered negligible. The system correlation function  $\varphi_g(t;s,s')$  is known.

### b) on the model

The model impulse response  $\hat{g}$  is defined by a low-dimension parameter vector p, i.e.  $\hat{g}(s) = \hat{g}[s(p)]$ .

# c) <u>on the noise</u>

The noise d is a zero-mean, Gaussian, stationary, stochastic process having a known correlation function  $\varphi(t)$ .

# d) on the measurements

Samples  $u_M$  of the input and  $z_M$  of the corresponding measurable output of the system are given, observed on a time interval of suitable length  $T_M$ .

# 3. - Approximation criterion

Let  $\mathbf{u}$  be the set of the inputs u of system and model and let  $\mathbf{u}$  be formed by the square integrable functions on a certain time interval  $[O, T_A]$ . According to the system assumptions also the corresponding outputs y of the system are square integrable (with probability one) and therefore two norms can be introduced

$$\| u \|^{2} = \int_{O}^{T_{A}} u^{2}(t) dt$$
 (3.1)

$$\|\hat{y} - y\|^2 = E \int_{O}^{T_A} [\hat{y}(t) - y(t)]^2 dt$$
 (3.2)

Now, for every u, it is well known that  $\overline{14}$ 

$$\|\hat{y} - y\|^2 \leq K^2 \|\|u\|^2$$
 (3.3)

where  $K^2$  represents the Hilbert-Schmidt norm of the difference between model and system, i.e.

$$K^{2} = E \int_{O}^{T} A \int_{O}^{t} \left\{ \hat{g} \left[ s(p) \right] - g(t, s) \right\}^{2} dt ds \qquad (3.4)$$

Observe that  $K^2$  represents an upper bound for the usual norm of the difference between model and system induced by definitions (3.1) and (3.2).

At this point, according to relation (3.3), an input independent approximation criterion is introduced and the optimal time-invariant model of low order for the given system is defined as that which corresponds to the value  $p_0$  of p minimizing the

quantity (3.4).

Finally, in order to carry out this operation it can be remarked that the value  $p_{O}$  to be determined also corresponds to the minimum of the expression

$$\int_{O}^{T_{A}} (T_{A} - s) \left\{ \hat{g} [s(p)] - g_{O}(s) \right\}^{2} ds \qquad (3.5)$$

which can be easily obtained from (3.4) and which differs from this of a known value depending on the system correlation function  $\varphi_g$  (see section 2).

# 4. - Model determination

The evaluation of the index (3.4) (or (3.5)) and then of the optimal model is not directly possible since  $\hat{g}$  can be easily computed for any fixed p but the impulse r<u>e</u> sponse  $g_0$  is unknown. Therefore, this function must be estimated from the available input-output data represented by samples  $u_M$  and  $z_M$  (see section 2).

According to (2.1) the relation (2.2) can be expressed in the form

$$z(t) = \int_{0}^{\infty} g_{0}(s) u(t-s)ds + \int_{0}^{\infty} [g(t, s) - g_{0}(s)] u(t-s)ds + d(t)$$
(4.1)

and letting

$$y_{v}(t) = \int_{0}^{\infty} [g(t, s) - g_{0}(s)] u(t-s) ds$$
 (4.2)

as

$$z(t) = \int_{0}^{\infty} g_{0}(s) u(t-s)ds + y_{v}(t) + d(t)$$
 (4.3)

Now, by choosing a suitable sampling time  $T_c$  such that the functions  $g_o$  (see section 2) and  $u_M$  can be sufficiently approximated from their samples and by introducing the following matrix notations (the symbol T means transpose)

$$\mathbf{Z}^{\mathrm{T}} = \left[ \mathbf{z}_{\mathrm{M}}^{\mathrm{(O)}} \quad \mathbf{z}_{\mathrm{M}}^{\mathrm{(T_{c})}} \dots \mathbf{z}_{\mathrm{M}}^{\mathrm{[(m-1)T_{c}]}} \right]$$
(4.4)

$$\mathbf{G}_{\mathbf{O}}^{\mathrm{T}} = \begin{bmatrix} \mathbf{g}_{\mathbf{O}}(\mathrm{O}) & \mathbf{g}_{\mathbf{O}}(\mathrm{T}_{\mathrm{C}}) & \dots & \mathbf{g}_{\mathrm{O}} \begin{bmatrix} (n-1)\mathrm{T}_{\mathrm{C}} \end{bmatrix} \end{bmatrix} \mathbf{\gamma}$$
(4.5)

$$Y_{v} = \begin{bmatrix} y_{v}(O) & y_{v}(T_{c}) & \dots & y_{v} \begin{bmatrix} (m-1)T_{c} \end{bmatrix} \end{bmatrix}$$
(4.6)

$$D = \left[d (O) \quad d(T_c) \dots d \left[(m-1)T_c\right]\right]$$
(4.7)

$$U = \begin{bmatrix} u_{M}(O) & O & \cdot & O \\ u_{M}(T_{c}) & u_{M}(O) & \cdot & O \\ \cdot & \cdot & \cdot & \cdot \\ u_{M} \left[ (n-1)T_{c} \right] u_{M} \left[ (n-2)T_{c} \right] \cdot & u_{M}(O) \\ \cdot & \cdot & \cdot \\ u_{M} \left[ (m-1)T_{c} \right] u_{M} \left[ (m-2)T_{c} \right] \cdot & u_{M} \left[ (m-n)T_{c} \right] \end{bmatrix}$$
(4.8)

the relation (4.3), which corresponds to the observed data, becomes in discrete form

$$Z = U G_{O} + Y_{v} + D$$
 (4.9)

Of course,  $\gamma$  represents a known normalizing factor ,

$$n = T_s/T_c - 1, \quad m = T_M/T_c - 1$$
 (4.10)

(see section 2). It is assumed m  $\ge$  n and u<sub>M</sub>(O)  $\ne$  O. The minimum variance unbiased estimate (Gauss-Markov) of the unknown vector G<sub>O</sub> is  $\boxed{15}$ 

$$\overline{G}_{O} = (U^{T}C^{-1}U)^{-1} U^{T}C^{-1}Z$$
(4.11)

where G is the covariance matrix of  $Y_v^+D$ , that is, in the case of g and d not corr<u>e</u> lated

$$C = E \left[ Y_{v} Y_{v}^{T} \right] + E \left[ D D^{T} \right]$$
(4.12)

According to the assumptions of section 2, this matrix can be evaluated from the knowledge of  $u_{\rm M}$  and  $\varphi_{\rm g}$  for the first term and from the knowledge of  $\varphi_{\rm d}$  in a very direct form for the second term.

The estimate error

$$\boldsymbol{\xi} = \overline{G}_{O} - G_{O} \qquad (4.13)$$

is a Gaussian random vector having zero mean value and covariance matrix

$$Q = (U^{T} C^{-1} U)^{-1}$$
 (4.14)

The computation of the optimal model is obviously based on the knowledge of the estimate  $\overline{G}_{O}$ .

In fact, coming back to relation (3.5) and letting

$$\hat{\mathbf{G}}^{\mathrm{T}} = \left[ \hat{\mathbf{g}}(\mathbf{O}) \ \hat{\mathbf{g}}(\mathbf{T}_{\mathbf{C}}) \ \dots \ \hat{\mathbf{g}} \ \left[ \mathbf{n-1} \right]^{\mathrm{T}} \mathbf{Y}$$

$$(4.15)$$

this approximation index can be written as

$$\begin{bmatrix} \hat{\mathbf{G}}(\mathbf{p}) - \mathbf{G}_{\mathbf{O}} \end{bmatrix}^{\mathrm{T}} \mathbb{W} \begin{bmatrix} \hat{\mathbf{G}}(\mathbf{p}) - \mathbf{G}_{\mathbf{O}} \end{bmatrix}$$
(4.16)

where W is a diagonal nxn positive definite weighting matrix depending on the weighting function  $(T_{\Delta} - s)$  of (3.5).

According to (4.13), if one defines

$$e(p) = \hat{G}(p) - \overline{G}_{O}$$
 (4.17)

(4.16) becomes

$$\left[e(p) + \boldsymbol{\varepsilon}\right]^{\mathrm{T}} \mathbb{W} \left[e(p) + \boldsymbol{\varepsilon}\right]$$
(4.18)

which has to be minimized with respect to p.

Indeed the expression (4.18) is a random variable with known statistics. Therefore it seems to be reasonable, by following a worst case criterion, to assume that the optimal model which is possible to determine is sought by minimizing (always with respect to p)

$$\max_{\boldsymbol{\xi}} (e + \boldsymbol{\xi})^{\mathrm{T}} W (e + \boldsymbol{\xi})$$
(4.19)

where the vector  $\boldsymbol{\varepsilon}$  lies in the confidence region which corresponds to a given level of probability P. This region (see above) is obviously defined by the relation

$$\boldsymbol{\varepsilon}^{\mathrm{T}} \mathbf{Q}^{-1} \boldsymbol{\varepsilon} \leq \mathbf{F}(\mathbf{P}) \tag{4.20}$$

where F is a known function of P.

The problem of finding the quantity (4.19) with the constraint (4.20) is now taken into account.

Let be

$$\mathbf{\xi} = Q^{1/2} \mathbf{\alpha}$$
,  $e = Q^{1/2} \mathbf{a}$  (4.21)

By substitution in (4.20) and (4.21) it is obtained

$$\max_{\boldsymbol{\alpha}} (\mathbf{a} + \boldsymbol{\alpha}) \mathbf{Q}^{1/2} \mathbf{W} \mathbf{Q}^{1/2} (\mathbf{a} + \boldsymbol{\alpha})$$
(4.22)

$$\boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\alpha} \leq \mathbf{F}(\mathbf{P}) \tag{4.23}$$

and introducing the lagrangian multiplier /u the necessary condition for determining (4.22) is written as

$$\alpha = - \left[ Q^{1/2} W Q^{1/2} - \mu^{u} I \right]^{-1} Q^{1/2} W Q^{1/2} e$$
(4.24)

The use of (4.24) in (4.23) in order to derive ju yields

$$a^{T} (Q^{1/2} W Q^{1/2}) [Q^{1/2} W Q^{1/2} - \mu I]^{-2} (Q^{1/2} W Q^{1/2}) a \leq F(P)$$
 (4.25)

The first member of this equation is a quadratic form which is defined by a matrix

whose eigenvectors are the same  $\overline{\lfloor 16 \rfloor}$  of the matrix  $(Q^{1/2} W Q^{1/2})$ . Therefore, the diagonalization of this quadratic form is possible by introducing the transformation

$$b = M^{-1}a$$
 (4.26)

where M is the normalized modal matrix of  $(Q^{1/2} W Q^{1/2})$  whose eigenvalues are indicated by  $\lambda_i$ . Then (4.25) can be rewritten in the form

$$\sum_{i=1}^{n} b_{i}^{2} \frac{\lambda_{i}^{2}}{(\lambda_{i} - \mu)^{2}} = F$$
(4.27)

where in terms of the original variable it is

$$b = M^{-1}Q^{-1/2}e$$
 (4.28)

and where it is taken into account that the solution of the problem belongs to the boundary of the region defined by (4.20).

Equation (4.27) has 2n solution for  $\mu$ . What is, among these, the value  $\mu_0$  which allows to determine the quantity (4.19)? A sufficient condition can be easily derived from the second derivative of the functional (modified with the introduction of the lagrangian multiplier). So, the matrix

$$\left[Q^{1/2} W Q^{1/2} - \int^{u I} \right]$$
 (4.29)

must be definite negative and hence the relation

$$u > \lambda_i$$
,  $i = 1, 2 \dots n$  (4.30)

must be satisfied. Therefore, according to the form of (4.27), /u<sub>o</sub> is usually the only solution of (4.27) greater than the maximum eigenvalue of  $(Q^{1/2} W Q^{1/2})$ . Observe that singular cases can arise when the condition

$$\sum_{i=1}^{n} b_{i}^{2} \frac{\lambda_{i}^{2}}{(\lambda_{i}^{-} \mu)^{2}} \leq F \qquad (4.31)$$

holds, due to the fact that the coefficient  $b_i$  corresponding to the maximum  $\lambda_i$  is equal to zero. The problem can still be solved but multiple solutions can happen  $\sqrt{17/2}$ .

From  $\mu_0$  it is possible to calculate the expression (4.19). In fact, by use of (4.24), expression (4.22) becomes

$$a^{T} \left[ I - Q^{1/2} W Q^{1/2} (Q^{1/2} W Q^{1/2} - \mu_{o} I)^{-1} Q^{1/2} W Q^{1/2} - [I - Q^{1/2} W Q^{1/2} (Q^{1/2} W Q^{1/2}) - \mu_{o} I)^{-1} \right] a$$
(4.32)

and this quadratic form can be diagonalized by introducing the transformation(4.26).

Finally, the quantity (4.19) is written in the form

$$\sum_{i=1}^{n} b_{i}^{2} \frac{\lambda_{i} \mu_{0}^{2}}{(\lambda_{i} - \mu_{0})^{2}}$$
(4.33)

and this represents the required result.

At this point the whole procedure of approximation can be summarized. From the real data  $u_{M}$  and  $z_{M}$ , the matrices U and Z are constructed and the estimate  $\overline{G}_{O}$  of the unknown vector  $G_{O}$  is evaluated according to (4.11). Then, using eigenvalues and eigenvector of  $Q^{1/2} W Q^{1/2}$  the solution  $u_{O}$  of equation (4.27) is found. Finally, the optimal approximating model is derived as that corresponding to the value of p which gives (see (4.33))

$$\min_{p} \sum_{i=1}^{n} b_{i}^{2} \frac{\lambda_{i} / u_{o}^{2}}{(\lambda_{i}^{-} / u_{o})^{2}}$$
(4.34)

Remark that from (4.34) and through (3.5) the value  $K^2(p_0)$  of (3.4) corresponding to  $p_0$  can be easily computed and therefore the method furnishes, together with the optimal model, also a measure of the reached approximation. It is possible to conclude that, for any square integrable signal u on [O,  $T_A$ ], on the basis of system knowledge which is possible to derive from the real data  $u_M$  and  $z_M$ , there is at least a given probability P that the condition

$$E \int_{O}^{T_{A}} \left[ \hat{y}_{O}(t) - y(t) \right]^{2} dt \leq K^{2}(p_{O}) \int_{O}^{T_{A}} u^{2}(t) dt \qquad (4.35)$$

holds ( $\hat{y}_0$  represents the optimal model output corresponding to u).

### 5. - Conclusions

The problem of modelling a linear, time-varying, stochastic system described by input-output data corrupted by noise has been considered. The assumption is made that the approximating model is time-invariant and has a simple structure described in terms of a parameter vector. An approximation criterion is introduc ed following a min-max approach and the optimal model is defined as that which minimizes the Hilbert-Schmidt norm of the difference between model and system. The evaluation of this performance index is not directly possible from the available data on the system: therefore a procedure is given in order to obtain an estimate of such an index and in order to determine its upper bound corresponding to a given level of probability. Finally, the minimization of this quantity, which has to be carried out by a numerical iterative procedure, leads to the optimal approximating model and furnishes a measure of the obtained approximation which results, of course, input-independent.

# References

- [1] R. Genesio and M. Milanese: "A note on derivation and use of reduced order models" - IEEE Trans. Automat. Contr., vol. AC-21, Feb. 1976.
- 2] C.F. Chen and L.S. Shieh: "A novel approach to linear model simplification" - Int. J. Contr., vol. 8, pp. 561-570, 1968.
- [3] M.R. Chidambara: "Two simple techniques for the simplification of large dynamic systems" Proc. JACC '69, pp. 669-674, 1969.
- [4] T.C. Hsia: "On the simplification of linear systems" IEEE Trans. Autom. Contr., vol; AC-17, pp. 372-374, 1972.
- 5] H. Heffes and P.E. Sarachik: "Uniform approximation of linear systems" -Bell Syst. Tech. J., vol. 48, pp. 209-231, 1969.
- <u>[6]</u> M. Milanese and A. Negro: "Uniform approximation of systems. A Banach space approach" J.O.T.A., vol. 12, pp. 203-217, 1973.
- [7] R.J.P. De Figuereido, A. Caprihan and A.N. Netrevali: "On optimal modeling of systems" - J.O.T.A., vol. 11, pp. 68-83, 1973.
- G.J. Bierman: "Weighted least squares stationary approximations to linear systems" IEEE Trans. Automat. Contr., vol. AC-17, pp. 232-234, 1972.
- [9] H. Nosrati and H.E. Meadows: "Modeling of linear time-varying systems by linear time-invariant systems of lower order" - IEEE Trans. Automat. Contr., vol. AC-18, pp. 50-52, 1973.
- [10] W.L. Root: "On the measurement and use of time-varying communication channels" Inform. and Contr., vol. 8, pp. 390-422, 1965.
- [11] I. Bar-David: "Estimation of linear weighting functions in Gaussian noise"
   IEEE Trans. Inform. Theory, vol. IT-14, pp. 288-293, 1968.
- [12] P.A. Bello: "Measurement of random time-variant linear channels" IEEE Trans. Inform. Theory, vol. IT-15, pp. 469-475, 1969.
- [13] E. Mosca: "A deterministic approach to a class of nonparametric system identification problems" IEEE Trans. Inform. Theory, vol. IT-17, pp. 686-696, 1971.
- [14] N. I. Akhiezer and I. M. Glazman: "Theory of Linear Operators in Hilbert Space", vol. I, F. Ungar, New York, 1966.
- /15 / R. Deutsch: "Estimation Theory", Prentice-Hall, Englewood Cliffs, 1965.
- [16] F.R. Gantmacher: "The Theory of Matrices", vol. I, Chelsea, New York, 1959.
- [17] R. Genesio and R. Pomé: "Identification of reduced models from noisy data" - Int. J. Contr., vol. 21, pp. 203-211, 1975.