IDENTIFICATION OF TIME VARYING STOCHASTIC SYSTEMS

David A. Maluf [1,2] and *Laszlo Gerencser* [2]

[1] Computer Research Institute of Montreal, 1801 McGill College, Montreal, Canada.
[2] Department of Electrical Engineering, McGill University, 3470 University Street, Canada.

1 Abstract

This paper presents and discusses the simulation of time stochastic ARMA systems and the development of an off line identification based on a prediction error method combined with exponential forgetting. The simulation involves different stages, starting with the simulation of the Gaussian white noise as an input to the original system and, the simulation with a combined signal and Gaussian white noise (ARMAX model). Simulation of time variation is also based on Gaussian white noise for a higher complex error variation. Stability tests on the time system are achieved and appropriate measures are taken to improve the feasibility and stability of the system. This paper is intended to present the simulation of mathematical models based on completely observed input-output dynamical systems.

Keywords
Stochastic systems, ARMAX process, time varying systems.

2 Introduction

In the analysis and design of control systems, it is of high importance to have a mathematical model of a given system that describes the system dynamics as completely and accurately as possible. System descriptions can be achieved through parametric models which are accepted as mathematical models.

Limiting the work to Linear Time Varying (LTV) system description, the question can be bounded to the identification of the parameters of a dynamical input-output observed system. The identification can be achieved through the estimation of those parameters. A general technique in handling time systems is to simulate the system with Gaussian white noise and to assume the system is in a similar environment. Even though this assumption complicates the estimation process, it provides an excellent opportunity to test the Auto Regression and the Gauss-Newton minimization algorithms for high complex dynamical outputs.

System identification tool box is derived by L. Ljung, [2]. The tool box was designed for Linear Time Invariant (LTI) which does not satisfy the time variation as an initial requirement. Furthermore, the tool box ARMAX provides equal weights, which also contradict the need for exponential forgetting where recent data are given more weight. For this paper, most of those functions were rewritten to satisfy the initial requirements. This simulation works only for multiple-input-single-output problems.

2.1 Theoretical Approach

Least Square Estimation

Assume that for some N input-output observations, the estimation problem can be written as finding the parameters of

$$Y = X\Theta + \epsilon \tag{1}$$

where Y is the observed output and X the input. Thus the definition of the cost function can be written as

$$(Y - X\Theta)^T(Y - X\Theta) \quad min_{\Theta.} \tag{2}$$

The gradient of the former equation with respect to Θ will minimize the cost function, thus

$$\frac{\partial}{\partial\Theta}(Y - X\Theta)^T(Y - X\Theta) = 0 \tag{3}$$

and the solution will be of the type

$$X^T X\hat{\Theta} = X^T Y. \tag{4}$$

In the case $X^T X$ is nonsingular, there exists a unique solution for the estimation, and the parameters can be rewritten as

$$\hat{\Theta} = (X^T X)^{-1} X^T Y. \tag{5}$$

It is of great interest to notice that the second gradient of the cost function will provide the variance of the estimation.

Elements of State Analysis of Estimation

As mentioned before, the estimation problem can be written under the form

$$Y = X\Theta + \epsilon, \tag{6}$$

and, by linear regression, we have $E\epsilon\epsilon^T = I$ and taking $\hat{\Theta}$ as the Least Square of Θ^*, then

$$E(\hat{\Theta} - \Theta^*)(\hat{\Theta} - \Theta^*)^T = (X^T X)^{-1}. \tag{7}$$

Analyzing $V_N(\Theta)$ for the Auto Regression,

$$V_N = \frac{1}{2}\sum \epsilon^2(\Theta). \tag{8}$$

This approach provides linear weights for the estimation and therefore it does not apply completely for the Linear Time Varying systems.

Maximum Likelihood for Gaussian Processes

An Auto Regressive Moving Average (ARMA) process can be written in term of its parameters as

$$A^* y = C^* e \tag{9}$$

where A^* is a stable polynomial and e is a Gaussian white noise function $N(0, \sigma^2)$ for $(n = .. - 1, 0, 1, 2..)$. If we assume that $e_n = 0$ for $n < 0$, then $y_n = 0$ for $n < 0$. To apply the Maximum Likelihood (ML), it is necessary to compute the density function of

$$(y_1, y_2, ..., y_N) = T(e_1, e_2, ...e_N)^T \tag{10}$$

where T is a lower triangular matrix with diagonal $(1, 1.., 1)$ and its determinant equal to one. Thus we can say that T maps e to y under the $y = Te$ transformation.

Taking a sequence of observations $(y_1, y_2, ..., y_N)$, and assuming the parameter Θ, the computation of $f(y_1, y_2, ..., y_N, \Theta)$ maximizes along Θ. The computation for the density of Θ is similar to the computation of ϵ_n from

$$C\epsilon = Ay \tag{11}$$

where

$$\epsilon_n = y_n = 0 \quad n \leq 0 \tag{12}$$

for initial conditions. As a conclusion Maximum Likelihood is equivalent to minimizing

$$V_N(\Theta) = \sum \epsilon_i^2(\Theta) \tag{13}$$

which is defined as a cost function for the estimation.

2.2 Auto Regression Approach

Auto Regressive Processes

Assume the following process

$$y_t + a_1 y_{t-1} + ... + a_p y_{t-p} = e_t. \tag{14}$$

The definition of the shift operator q^{-1} is such that

$$q^{-1}(y)_t = y_{t-1} \tag{15}$$
$$q^{-2}(y)_t = y_{t-2}$$
$$......$$

which can be written as

$$(1 + a_1 q^{-1} + ... + a_p q^{-p})(y)_t = e_t \tag{16}$$

or

$$A(q^{-1})y = e \tag{17}$$

and can be formally written as

$$y = \frac{1}{Aq^{-1}} e. \tag{18}$$

The possible meaning of $\frac{1}{Aq^{-1}}$ can be defined as

$$\frac{1}{Aq^{-1}} = \sum \frac{\sigma_i}{(1 - \lambda_i z^{-1})} \tag{19}$$

which is possible when $\lambda_i s$ are simple roots of $A(z^{-1}) = 0$. Finally the definition of the regressive process will be

$$y_{it} = \frac{\sigma_i}{1 - \lambda_i q^{-i}} e. \tag{20}$$

2.3 Gauss Newton Minimization

The Gauss-Newton direction usually is defined as

$$gn = (\Psi^T \Psi)^{-1} \Psi^T e \tag{21}$$

where Ψ is the gradient matrix of the predictions with respect to the parameters, and e is the residual vector.

The regression iterative search yields a local minimum, still it is not guaranteed as a global minimum. To solve the local global problem, the selection of an initial condition for the minimization will settle that question. As a start, for the selection of the initial conditions, it is necessary that the estimated (C, A) should be stable. Computing

$$\epsilon_{\Theta n}(\Theta) = \frac{\partial}{\partial} \epsilon_n(\Theta), \quad 1 \leq n. \tag{22}$$

3 Time Variant Systems

For time systems, the computation of the maximum possible variation over a period (t_1, t_2) for the time parameters can be bounded and written as

$$sup_t |\Theta_t| < \delta. \tag{23}$$

Assuming that δ is small relatively and for such values, the system can be called a "slowly" system.

For continuous time, the cost function can be written as

$$V_T = \int_0^T e^{-\lambda(Ts)} |d\epsilon_T(\Theta)|^e \tag{24}$$

The gradient of the cost function will be in such case

$$\frac{\partial}{\partial \Theta} V_T = \int_0^T e^{-\lambda(T-s)} \lambda \epsilon_{\Theta s}(\Theta) d\epsilon s(\Theta). \tag{25}$$

It can be shown that the cost function is made in two parts, thus

$$V_{\theta \ldots \Theta_T}(\Theta) - W_{\Theta\Theta}(\Theta, \Theta_t^*) \tag{26}$$

where the first function depends on the number of derivatives and the second is related to the time invariancy of the system. Limiting the study to discrete "slowly" time systems, and assuming that Θ^* is "slowly" time , the ARMA process can be written

$$y_n = (H(\Theta^*)e)_n \quad |\Theta_{n+1}^* - \Theta^*| < \delta. \tag{27}$$

For special processes

$$y_n = \Theta_n^* + e_n, \tag{28}$$

then if $\Theta_n^* = \Theta^*$ then the estimated parameter would be

$$\hat{\Theta}_N = \frac{1}{N} \sum_1^N y_n \tag{29}$$

which is the Least Square (LSQ) Estimation of Θ. When Θ^* is truly time varying, then the implementation of forgetting past informations and the estimator would change to

$$\hat{\Theta}_N = \sum_1^N (1 - \lambda^i) \lambda y_{N-i} \tag{30}$$

where $0 < \lambda < 1$ is a necessary condition. The selection of λ, which can be called the forgetting factor, can be related to the order of the cube of the path behavior of the estimator process and,

$$\frac{1}{T} \sum_{t=1}^T |\hat{\Theta}_t - \Theta_t^*|^2 = O(\delta^{\frac{1}{3}}). \tag{31}$$

3.1 Simulation and Identification

Identifying models from data involves decision making on the part of the person in search of models, as well as some complex and demanding computations to furnish the bases for these decisions. Typically, one has to go through several loops revising previous decisions and searching for initial conditions. Interactive computerized simulation is thus the tool to approach system identification in practice. It is also a nice way to test the huge theoretical background and obtaining its maximum use.

The following is the presentation of a tool box that provided the initial tool for the simulation of time stochastic systems. It is not necessary to work with a specific system, but for the purpose of testing and simulating, it is necessary to select a practical model and try to manipulate it as much as possible. It is also obvious that to start the identification process, observed data is necessary. Let us first define the system

$$A_t y = B_t u + C_t e \tag{32}$$

where the A is time and is stable. Let us assume also that C is stable. Simulating the former system with a specific input forms a set of input-output observations for a length N, whereas simulating the system with no input ends up with a set of output observations only. Finally let us assume that e is Gaussian white noise.

Proceeding now to the identification, it is necessary to start with either set of observations. To be able to identify exactly the simulated system, it is also advisable to provide the structure and the number of delays of the simulated system. The identification process starts by finding an initial estimate and afterwards weighing the prediction error exponentially. From that point we can start the iterative search, minimizing the cost function with respect to a tolerance.

4 L. Ljung Identification Toolbox

The system identification tool box presented by L. Ljung is an extension to Matlab. The system identification tool box contains the basic tools for building mathematical models of dynamical systems, based on input-output observations.

Parametric models.

A major condition for estimating the parameters of a system, the input-output observations have to be represented by a finite dimensional linear time invariant model. The general parametric model structure is given by

$$A(q)y(t) = \frac{B(q)}{F(q)}u(t-nk) + \frac{C(q)}{D(q)}e(t) \qquad (33)$$

where A, B, C, D are polynomials in the delay q^{-1} such as

$$A(q) = 1 + a_1 q^{-1} + a_2 q^{-2} + ... + a_N q^{-N}, \quad (34)$$

where nk is the number of delays from input to output and $e(t)$ is Gaussian white noise with a variance λ. Equation 33 provides virtually a complete structure for a black box where all the parameters can be identified separately.

The system identification tool box implements four different methods: Instrumental Variable method (IV), Least-Square method (LSQ), Maximum Likelihood (ML) and Spectral Analysis (SPA). These four methods are implemented non-recursively and can be classified under two categories. The first three methods are based on the parametric model whereas the fourth method estimates the frequency functions directly without the use of the parametric models as set by equation 33.

The system identification has different ways to represent a system model. The *theta* format which is related to equation 33 carries all the polynomial orders, delays, parameters and their estimated covariances in a single matrix. The *polynomial* format is defined by the five standard A, B, C, D, F matrix structures as set by equation 33.

Ideal simulation

The system identification tool box is equipped to simulate a system with selected inputs and noise. It is not necessary to mention again that the simulation is in a discrete domain for LTI systems. Nevertheless, a time system can be simulated by invoking the LTI ideal simulation at every sample of input with new parameters. Such a process requires passing as arguments the final conditions of the ideal simulation of every sample as initial conditions to the next sample ideal simulation.

Model Validation

The system identification tool box has intended functions for solving the problem of getting the best model structure that will fit the given input-output observations. All these functions were obtained by heuristic methods. The residual analysis provides that the residuals associated with the given observations and a model should be white and independent of the input for the model to be a "correct" description of the system. The correlation function of the noise and the cross correlation between the noise and input should not exceed a given boundary of confidence intervals in order for the model to be acceptable.

Akaike's Final Prediction Error Criterion. The theory behind this criterion is that, in a set of different models, the model with the smallest value of FPE

$$FPE = \frac{1+\frac{n}{N}}{1-\frac{n}{N}} + V \qquad (35)$$

where

$$V = inf \frac{1}{2} \sum_{i=1}^{N} \epsilon_i^2(\Theta, \Theta^*) \qquad (36)$$

for an equal weighted cost function and

$$V = inf\frac{1}{2}\sum_{i=1}^{N}\lambda^{N-i}\epsilon_i^2(\Theta,\Theta^*)\lambda \qquad (37)$$

for a normalized exponentially- weighted cost function.

Stability

In identifying the parameters of a system, the estimation of the parameters may diverge in a non-stable state. The system identification tool box provided a tool to check the stability and provide the necessary modifications to the system to ensure a complete, stable estimated system. This tool is used at every iteration in the auto regression.

5 Simulation of Time Varying Systems

This section explains the simulation of a general system in discrete time trying to approach a realistic behavior as much as possible.

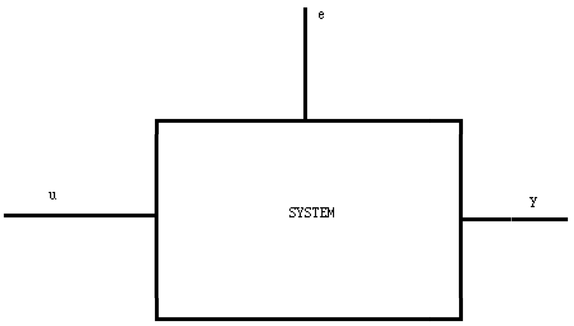

Figure 1: System peripherals, inputs, noise, output.

Considering the system in Figure 1, the basic input output configuration can be depicted. To start the simulation and generate the observations y_t, the definition of the initial system at the initial time is essential. Let A, B, C define the system structure such that (A, C) are stable and A is third order time variant and let δ_t be a Gaussian white noise assigned as error in the parameters variation with respect to the initial system parameters of (n-1) order

$$A_{t+1} = A_t + \delta_t. \qquad (38)$$

Such an equation might drift the system into an unstable state. Stabilization is needed to be able to approach a realistic operating system, and thus the new system will be processed and stabilized at every time t or,

$$A_t = stabilize(A_t). \qquad (39)$$

Finally, let u_t be an input sine signal and e_t a Gaussian white noise. The simulation of the former system will result in the generation of the output y_t which is collected at every sample.

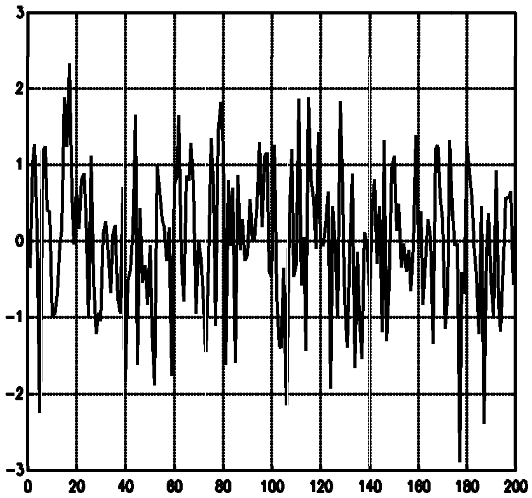

Figure 2: Input signals; sine and white noise.

In Figure 2, a discrete sine signal and a normalized Gaussian white noise are shown over a period of 200 samples. It obvious that input noise has tremendously affected the input signal and is not recognizable as a sine signal.

The simulation process is done through filtering the system with a general tapped delay line filter structure. As mentioned in the previous section, the identification tool box "IDSIM" is designed for (LTI) systems which is not the case when A is time varying. The solution for this problem is to simulate the system with "IDSIM" at every sample and pass the final conditions and the time parameters as arguments for the next sample simulation as the initial conditions and the new system structure.

In the simulation, the time variation error results in the outputs random behavior. Although specific time variations such as periodic or decaying variations can be depicted from the output error, in general cases the white time variation will provide a good simulation.

6 Identification of ARMAX Stochastic Processes

Now that the input was corrupted with noise and the system varies with time, the identification of the stochastic process can be presented. From what was presented in the previous section, only the observations of the input u and output y will be used to identify the system.

As a first step, it is advisable to have a look at the spectral analysis of the observations. Moving to a higher step, the least square method finds an ARX model. In Figure 3, the bodeplot of the spectral analysis and the ARX model transfer functions can be compared with the suggested time varying ARMAX model. The modified ARMAX model is a set of complex steps derived from the ARMAX model presented in the identification tool box.

The modified ARMAX model starts by finding a first estimate from the least square method model ARX as presented above. Then we determine the initial C-estimate parameters by building a high order auto regression model of the residual of the initial estimate and stabilizing the estimated parameters and finally using the least square method with the new residuals as inputs. At this point, the initial estimated error can be computed and weighted exponentially as explained in the previous sections and the cost function will be obtained. Finally, there is the minimization loop where the gradient of the cost function is computed for a defined number of iterations or the gauss-newton minimization becomes smaller than a preset tolerance.

Running the observations through ARMAX routine which is presented by the identification tool box, some similarities and improvements can be observed among the spectral analysis of the modified ARMAX and ARMAX. Figure 4 shows the magnitude and phase bode plot for the three systems' transfer functions.

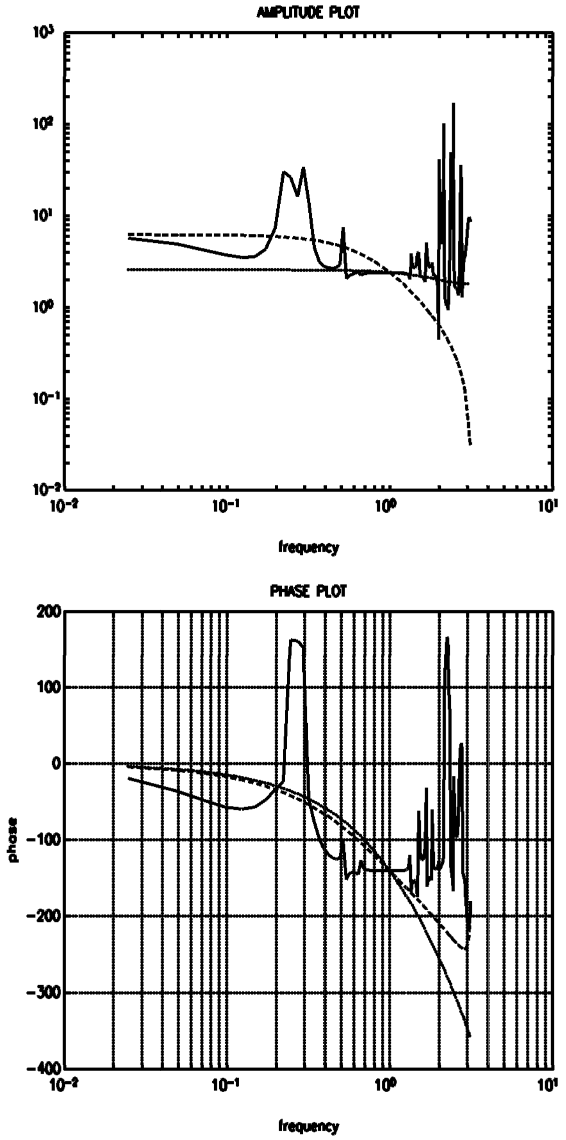

Figure 3: Bode plot of the spectral analysis, modified ARMAX and least square method transfer functions (dotted).

Comments on the shapes and plots of the modified ARMAX and the other plots are hard to be done, for the reason that the requirements and boundaries that the system follows are unknown. Normally once the characteristics of the system are known, the plots will be meaningful and their interpretation will be helpful. For Figure 5, the noise spectral analysis shows the differences between the mentioned systems' transfer functions.

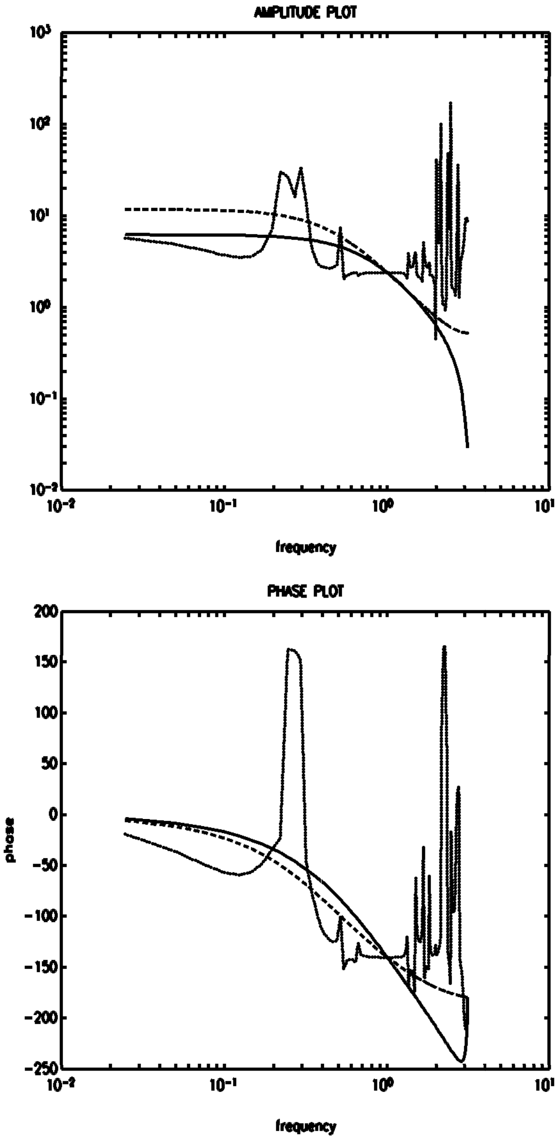

Figure 4: Magnitude and phase of the transfer function of the spectral analysis, modified ARMAX and ARMAX.

Once the system is identified, it is possible to simulate the expected output free of noise. Knowing the input observations, the simulation difference between the identified system using the modified ARMAX and the identified system using ARMAX in this case seems to degenerate into a periodic signal.

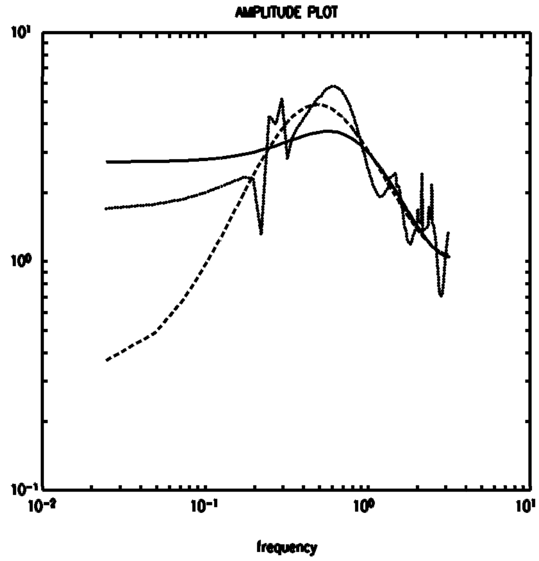

Figure 5: Magnitude and phase of the noise of the spectral analysis, modified ARMAX and ARMAX.

7 Conclusion

Off line simulation does not leave much choice of interpreting the behavior of the identification process as does the on line simulation. In the simulation presented in this paper, the noise input coefficient was increased to a certain limit to obtain a bad noise to signal ratio, and also the system parameters were increased to the upper limit around 2 percent drift, and hence is defined in this paper as a “slowly” time varying systems. To to be able to obtain the maximum effect of the simulation, and for the identification routine, the number of iterations and the tolerance were decreased to 10 and 0.001 for the minimization process to observe the behavior and path of the approximation.

Even though the time variation had a major effect on control system simulations, which can be seen by comparing the bode plots of the identified systems. The proposed time varying identification tries to predict more precisely the final system then the ARMAX identification.

8 REFERENCES

[1] Ogata, K., ” Discrete-Time Control Systems ”, Prentice Hall, 1987

[2] Ljung, L., " System Identification ", Prentice Hall, 1987

[3] Caines, P.E. " Linear Stochastic Systems ", John Wiley and Sons, inc. 1988

[4] Fomin, S.V., Sinai, Y.G., Cornfeld, I.P., " Ergodic Theory ", Springer Verlag, 1982

[5] Soderstrom, T., Stoica, P., "System Identification ", Prentice-Hall International, London, 1987

[6] Goodwin, G.C., Payne, R.L., " Dynamic System Identification : Experiment Design and Data Analysis ", Academic Press, 1977

[7] Eykhoff, P., "System Identification ", North-Holland, 1974