PERIODICA POLYTECHNICA SER. EL. ENG. VOL. 33, NO. 4, PP. 165-182 (1989)

Maximum Likelihood Estimation of the Parameters of Linear Systems[1]

J. SCHOUKENS, R. PINTELON AND J. RENNEBOOG

Department ELEC
Vrije Universiteit Brussel

Received: June 15, 1988; Revised: July 24, 1989.

Abstract

A method is presented which estimates the parameters of Linear Systems (LS), modelled by their transfer function, using a very efficient iteration algorithm. The estimator is an error in variables method and takes into account the noise on the input and output measurements. During the estimation process, an approximation of the Cramer-Rao lower bound on the covariance matrix of the estimates is derived and the 'mean' model error is discussed.

Keywords: Parameter estimation, transfer function, mean model error, complex approximation.

Introduction

In a lot of engineering problems considerable effort is spent to find mathematical models for the studied systems. An important class of systems consists of the linear, concentrated parameter, time invariant systems which can be modelled by a transfer function. For these systems the modelling problem is reduced to the optimal determination of the order and the value of the parameters in the transfer function. Identification theory offers a theoretical framework to solve this problem. In the literature, a broad class of estimators is described, for example the Bayes estimator, Maximum Likelihood estimator, Markov estimator and the Least Squares (EYKHOFF, 1974; SORENSON, 1980), allowing to estimate the value of a set of parameters starting from disturbed measurements. In comparing the estimators, the Bayes estimator allows to include a priori information about the value of the unknown parameters by giving their *a priori* probability density function. However, this information is not available in most situations. The Least Squares uses no *a priori* information, but has the weakest properties.

The estimation method, described here, belongs to the class of Maximum Likelihood Estimators (MLE). To use this estimator it is necessary

[1]This work has been supported by the National Fund for Scientific Research, Belgium.

to know a priori the probability density function (PDF) of the noise on the measurements. In this article, a method will be described which estimates the parameters of a transfer function starting from the measured Fourier coefficients of the input and the output which can be both disturbed by Gaussian noise. This is not a severe restriction because in a lot of applications the Fourier coefficients are obtained as the result of a Discrete Fourier Transform. In this situation frequency domain noise can be proven to be approximately Gaussian distributed, independently of the noise distribution in the time domain.

An important choice one has to make before starting the estimation process, is the model that is used to describe the Device Under Test (DUT). The method presented here is intended to study continuous time systems. In the literature these systems are usually approximated by discrete time systems (e.g. ARMA models). This results in difference equations to describe the DUT which are more adapted to digital computers than differential equations. To get useful results it is necessary to choose a sampling rate which is much higher than the highest frequency occurring in the measured signal (SPRIET and VANSTEENKISTE, 1982), while, on the other hand, the sampling frequency must not be chosen too high so that stability problems are avoided during the estimation step. In many problems, the approximation of a continuous system with difference equations is acceptable, but to accurately describe the DUT, it is necessary to use a continuous model.

Our method is presented here as an estimator for continuous systems. However, the reader can notice that the technique is also applicable for discrete systems by a simple change of variables ($j\omega \longrightarrow e^{-j\omega T}$).

An important contribution of this method is that it incorporates noise disturbances on the output AND the input signals. It is obvious that this is an evident choice for a lot of problems, however, in the section giving an overview of existing methods it will be pointed out that commonly used methods assume that the noise on the input measurements is zero.

A Model of the Measurement System

An a priori noise analysis of the measurement system is made to obtain the PDF of the noise on the measurements, which is necessary to construct an MLE.

A.) Study of the Measurement System in the Time Domain

In this article, single input, single output (SISO) systems are studied. The measurements are made with a two-channel digitizer which is the heart of the measurement system, schematically represented in *Figure 1.* A multisine, which is the sum of harmonically related sinusoids, is used as the input signal. The phases are selected in a way to obtain a signal with a minimal crest factor (the crest factor is the ratio of the peak value of the signal to its effective value) so that a maximal signal to noise ratio is obtained (SCHROEDER, 1979). These signals are very well suited to inject energy at the optimal frequencies only, resulting in a maximum accuracy (e.g. a bandpass, a lowpass, a highpass spectrum can be easily realized). The measured input and output signals are given by the data sequences $x_m(k)$ and $y_m(k), k = 1, \ldots, N$, respectively. These measurements are corrupted by noise, as shown in *Figure 1.* Between the measured and exact signals the following relations hold:

$$\begin{aligned} x_m(k) &= x(kT) + n_1(k) = x(kT) + n_x(k), \\ y_m(k) &= y(kT) + n_2(k) + n_3(k) = y(kT) + n_y(k), \end{aligned} \tag{1}$$

with T being the sample period, k the k^{th} sample point and $n_x(k)$ and $n_y(k)$ being the errors on the input and output measurement, respectively.

The noise on the measurements is modelled by 3 sources:

n_1, n_2 : the measurement noise on the input and output, e.g. digitizing noise

n_3 : the noise created by the DUT and the environment.

The following assumptions are made concerning the noise.

$$E[n_x(k)n_y(l)] = 0 \qquad \forall k, l, \tag{2}$$

$$E[n_x(k)n_x(l)] = \rho_{nx}(k-l) \qquad \forall k, l, \tag{3}$$

$$E[n_y(k)n_y(l)] = \rho_{ny}(k-l) \qquad \forall k, l, \tag{4}$$

with $E[...]$ the mathematical expectation. This means that the input and output noise are not correlated. No assumptions are made concerning the mean value of the noise because only the DC component, which is not 7used in the estimation procedure, is affected by it.

In a lot of applications, the digitizing noise is the most important noise contribution. In this case, n_x and n_y can be approximately modelled as white noise sources, and (3) and (4) reduces to

$$\begin{aligned} E[n_x(k)n_x(l)] &= \sigma_x^2 \delta_{kl} \qquad \forall k, l, \\ E[n_y(k)n_y(l)] &= \sigma_y^2 \delta_{kl} \qquad \forall k, l, \end{aligned} \tag{5}$$

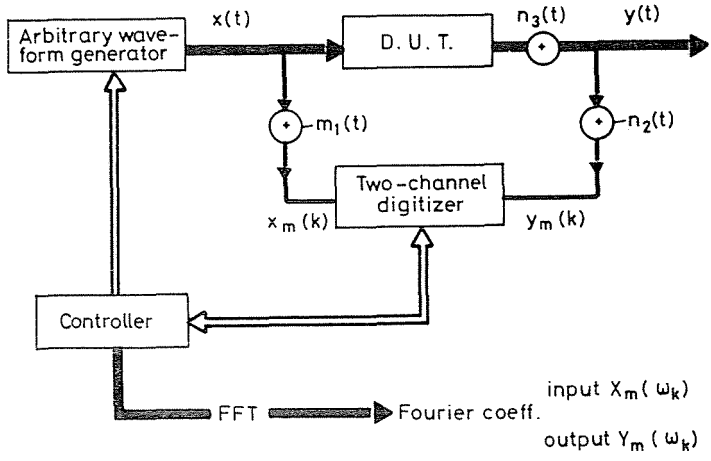

Fig. 1. Schematic presentation of the measurement system

with σ_x and σ_y being the standard deviations of the noises in the time domain. In (KOLLÁR, 1986) the conditions on the sampling rate to get white digitizing noise are studied.

B.) Modelling the Noise in the Frequency Domain

The spectra of the measured sequences, x_m and y_m, are calculated using the Discrete Fourier Transform (DFT). To avoid leakage effects, the generator is synchronized to the digitizer, otherwise the Interpolated Fast Fourier Transform can be used (RENDERS, SCHOUKENS and VILAIN, 1984). The results of the DFT will also be corrupted by noise, due to the time domain noise on the measurements x_m and y_m. For a given radian frequency ω_k, the following results are derived.

$$\begin{aligned} \mathbf{X}_m(\omega_k) &= \mathfrak{F}[x_m]_k + \mathfrak{F}[n_x]_k = \mathbf{X}(\omega_k) + \boldsymbol{\nu}_x(\omega_k), \\ \mathbf{Y}_m(\omega_k) &= \mathfrak{F}[y_m]_k + \mathfrak{F}[n_y]_k = \mathbf{Y}(\omega_k) + \boldsymbol{\nu}_y(\omega_k), \end{aligned} \tag{6}$$

with $\mathfrak{F}$ being the Fourier transform, ω_k the k^{th} component, $\mathbf{X}_m, \mathbf{Y}_m, \mathbf{X}, \mathbf{Y}$, $\boldsymbol{\nu}_x, \boldsymbol{\nu}_y$ complex vectors with both real and imaginary parts.

To construct an MLE, the knowledge of the PDF of $\boldsymbol{\nu}_x(\omega_k)$ and $\boldsymbol{\nu}_y(\omega_k)$ is necessary. SCHOUKENS and RENNEBOOG (1986) have shown that the PDF of the noise $\boldsymbol{\nu}_x, \boldsymbol{\nu}_y$ is given by a Gaussian distribution which is completely characterized by the mean and the covariance matrix of the noise.

Let us define

$$\boldsymbol{z}' = (\boldsymbol{\nu}'_{xR}, \boldsymbol{\nu}'_{yI}) = (\nu_{y1R} \ldots \nu_{yNR} \nu_{y1I} \ldots \nu_{yNR} \nu_{x1R} \ldots \nu_{xNR} \nu_{x1I} \ldots \nu_{xN_I}), \tag{7}$$

with N the number of Fourier coefficients, $\boldsymbol{\nu}_{xR}$ and $\boldsymbol{\nu}_{xI}$ being the real and imaginary part of $\boldsymbol{\nu}_x$, and $\mathbf{z}'$ being the transpose of $\mathbf{z}$. The covariance matrix of the noise is given by

$$\boldsymbol{\Sigma}_z = E[\mathbf{z}\ \mathbf{z}']. \tag{8}$$

From (SCHOUKENS and RENNEBOOG, 1986) it follows that $\boldsymbol{\Sigma}_z$ can be approximated by a diagonal matrix, even if the sampled noise sequences are highly correlated (the nondiagonal terms are negligible, compared to the diagonal terms). The diagonal terms are given by

$$\sigma^2_{\nu xk} = E[\nu_{xkR}\nu_{xkR}] = E[\nu_{xkI}\nu_{xkI}]. \tag{9}$$

$\sigma^2_{\nu yk}$ is defined similarly.

The values $\sigma^2_{\nu xk}$ and $\sigma^2_{\nu yk}$ can also be determined from a direct noise analysis.

For white noise sequences the expressions reduce to

$$\sigma^2_{\nu xk} = 2\frac{\sigma^2_{nx}}{N} \qquad \text{and} \qquad \sigma^2_{\nu yk} = 2\frac{\sigma^2_{ny}}{N}. \tag{10}$$

Definitions

To simplify the notations in the second part of this paper, it is necessary to systematically introduce many vectors. The following conventions are used (a . stands for an appropriate letter)

$..m$: measured value
$..e$: estimated value
$.R.$: real part
$\mathbf{X}, \boldsymbol{\nu}_x$: input Fourier coeff.
$.I.$: imaginary part
$\mathbf{Y}, \boldsymbol{\nu}_y$: output Fourier coeff.

e.g.

$\mathbf{X}_I$: the imaginary part of the exact Fourier coefficients of the input vector,

$\mathbf{X}_{Im}$: the imaginary part of the measured Fourier coefficients of the input vector.

Let us define also the following 'global' variables.

$$\mathbf{Z}' = (\mathbf{X}'_R \mathbf{X}'_I \mathbf{Y}'_R \mathbf{Y}'_I) \tag{11}$$

and define $\mathbf{Z}_e$ and $\mathbf{Z}_m$ in a similar way.

Model of the Device Under Test

To give a mathematical description of the Device Under Test (DUT), a model is chosen which gives the relation between the input signal and the output signal. We will use the Fourier coefficients as the basic measurements. As a consequence, the model should be defined in the frequency domain. One should notice that this approach also allows to start the estimator from direct measurements of the transfer function instead of the Fourier coefficients.

The complex model function $\mathbf{F}$ is defined for the exact Fourier coefficients $\mathbf{X}$ and $\mathbf{Y}$.

$$\mathbf{Y} = \mathbf{F}(\mathbf{X},\mathbf{p}) = \mathbf{F}_R(\mathbf{X},\mathbf{p}) + j\mathbf{F}_I(\mathbf{X},\mathbf{p}) \tag{12}$$

with $\mathbf{F}_R$ the real part, $\mathbf{F}_I$ the imaginary part of $\mathbf{F}$, and with $\mathbf{p}$ the unknown model parameters.

For nonlinear systems it is very difficult to reduce the complexity of these expressions. However, for linear systems, (12) becomes simpler as the k^{th} output Fourier coefficient depends only on the k^{th} input Fourier coefficient. Both are related by a complex function: the transfer function which depends upon the frequency.

$$\mathbf{Y}(\omega_k) = \frac{\mathbf{N}(\mathbf{p},\omega_k)}{\mathbf{D}(\mathbf{p},\omega_k)}\mathbf{X}(\omega_k) = \frac{\mathbf{N}_R(\mathbf{p},\omega_k)+j\mathbf{N}_I(\mathbf{p},\omega_k)}{\mathbf{D}_R(\mathbf{p},\omega_k)+j\mathbf{D}_I(\mathbf{p},\omega_k)}\mathbf{X}(\omega_k). \tag{13}$$

It has to be emphasized that the choice of the model order is extremely important. It should be chosen as simple as possible but must describe all the important contributions.

Formulation of the Maximum Likelihood Estimator

A.) Determination of the Likelihood Function

The first step in constructing an MLE is the determination of the likelihood function which is the conditional probability density function to realize a measurement $\mathbf{X}_m, \mathbf{Y}_m$ given the parameters **p, X, Y.** It can be remarked here that the input and output Fourier coefficients are also regarded as unknown parameters. During the determination of the likelihood function, it is presumed that the parameters satisfy the model equations (12).

$$\mathbf{Y} = \mathbf{F}(\mathbf{X},\mathbf{p}). \tag{15}$$

The likelihood function gives the probability of realizing the measured data given a set of parameters. This probability is given by

$$\mathrm{P}(\text{noise}) = \mathrm{P}(\mathbf{X}_m - \mathbf{X}, \mathbf{Y}_m - \mathbf{Y} \mid \mathbf{X}, \mathbf{Y}, \mathbf{p}). \tag{16}$$

The right hand side gives the conditional probability of realizing an experiment with result $\mathbf{X}_m, \mathbf{Y}_m$, given the parameter values **X, Y, p.** From the previous noise analysis we know that the noise in the frequency domain is approximated by a normal distribution with zero mean and a diagonal matrix.

The likelihood function becomes

$$\mathrm{P}(\mathbf{Z}_m \mid \mathbf{p}, \mathbf{Z}) = \frac{1}{\sqrt{(2\pi)^M \det \boldsymbol{\Sigma}_z}} e^{-\frac{(\mathbf{Z}_m - \mathrm{Z})' \boldsymbol{\Sigma}_z^{-1} (\mathbf{Z}_m - \mathbf{Z})}{2}}, \tag{17}$$

where

$$M = 4N.$$

B.) The Maximum Likelihood Estimator

The MLE is given by the values of $\mathbf{Z} = \mathbf{Z}_e$ and $\mathbf{p} = \mathbf{p}_e$ which maximize the likelihood function. Note that the covariance matrix $\boldsymbol{\Sigma}_z$ is not a function of the parameters **p** and **Z.**

In practice, the log likelihood function, defined as the logarithm of the likelihood function, is used to maximize the likelihood function. This expression is maximized by minimizing the cost function K defined as

$$K = (\mathbf{Z}_m - \mathbf{Z})' \boldsymbol{\Sigma}_z^{-1} (\mathbf{Z}_m - \mathbf{Z}). \tag{18}$$

It can be remarked here that the MLE is reduced to a weighted least squares method. During this minimization, the model equations (15) have to be satisfied. This can be done using the Lagrange formulation or by substitution of $\mathbf{Y}$ by $\mathbf{F(X,p)}$, i.e.

$$K = \left[\begin{bmatrix}\mathbf{X}_m\\ \mathbf{Y}_m\end{bmatrix} - \begin{bmatrix}\mathbf{X}_e\\ \mathbf{F}(\mathbf{X}_e,\mathbf{p}_e)\end{bmatrix}\right]' \boldsymbol{\Sigma}_z^{-1} \left[\begin{bmatrix}\mathbf{X}_m\\ \mathbf{Y}_m\end{bmatrix} - \begin{bmatrix}\mathbf{X}_e\\ \mathbf{F}(\mathbf{X}_e,\mathbf{p}_e)\end{bmatrix}\right]. \quad (19)$$

The complexity of this cost function can be reduced considerably for linear systems by eliminating the estimated Fourier coefficients in the problem. This results in a higher convergence rate and reduces the required computer power. To simplify the problem, the minimization of equation (19) is formulated using Lagrange multipliers and a new cost function, K_1, is defined as:

$$\begin{aligned} K_1 &= \tfrac{1}{2}(\mathbf{Z}_m - \mathbf{Z}_e)'\boldsymbol{\Sigma}_z^{-1}(\mathbf{Z}_m - \mathbf{Z}_e)\\ &- \sum_{k=1}^{n}\{L_{R_k}[\mathbf{D}(\mathbf{p},\omega_k)\mathbf{Y}_e(\omega_k) - \mathbf{N}(\mathbf{p},\omega_k)\mathbf{X}_e(\omega_k)]\}_{\text{real}}\\ &- \sum_{k=1}^{n}\{L_{I_k}[\mathbf{D}(\mathbf{p},\omega_k)\mathbf{Y}_e(\omega_k) - \mathbf{N}(\mathbf{p},\omega_k)\mathbf{X}_e(\omega_k)]\}_{\text{imag}}, \end{aligned} \quad (20)$$

with $\mathbf{D}(\mathbf{p},\omega_k)$ and $\mathbf{N}(\mathbf{p},\omega_k)$ the denominator and numerator of the transfer function as defined in (14) and L_{R_k} and L_{I_k} the Lagrange multipliers.

The minimum of K_1 is found by solving the following set of nonlinear equations:

$$\frac{\partial K_1}{\partial \mathbf{Z}_e} = 0 \quad \text{(a)} \quad 1\frac{\partial K_1}{\partial \mathbf{L}_R} = 0 \quad \text{(b)} \quad 1\frac{\partial K_1}{\partial \mathbf{L}_I} = 0 \quad \text{(c)} \quad (21)$$

Using the second and third set of equations, it is possible to eliminate $\mathbf{X}_e$ and $\mathbf{Y}_e$. Finally the cost function is reduced to

$$K_1 = \sum_{k=1}^{N} \frac{|\,\mathbf{N}(\mathbf{p}_e,\omega_k)\mathbf{X}_m(\omega_k) - \mathbf{D}(\mathbf{p}_e,\omega_k)\mathbf{Y}_m(\omega_k)\,|^2}{\sigma_{\nu xk}^2\,|\mathbf{N}(\mathbf{p}_e,\omega_k)|^2 + \sigma_{\nu yk}^2\,|\mathbf{D}(\mathbf{p}_e,\omega_k)|^2}. \quad (22.)$$

This cost function depends only upon the model parameters $\mathbf{p}_e$ and the measured input and output Fourier coefficients. The minimization is again a nonlinear problem, due to the denominator of (22). The Gauss–Newton iteration procedure is chosen again, but in comparison to the problem for nonlinear systems, only the vector $\mathbf{p}_e$ is estimated which results in an important reduction in the number of unknown parameters. To find start

values, the cost function (22) is minimized in the first iteration step setting the denominator equal to one. So the minimization procedure is reduced to a linear least squares problem and the start values for $\mathbf{p}_e$ are found as the solution of a set of linear equations. In the following iteration steps, also the denominator is considered to be a function of $\mathbf{p}_e$ during the calculation of the Jacobian.

Study of the Uncertainty on the Estimates (Calculation of the Cramer–Rao lower bound)

After the estimation of the parameters, it is also necessary to have an idea of the uncertainty on the estimates which is described by the covariance matrix of the parameters.

$$\Sigma_{\theta_e} = E[(\boldsymbol{\theta}_e - \boldsymbol{\theta})(\boldsymbol{\theta}_e - \boldsymbol{\theta})' \mid \boldsymbol{\theta}]. \tag{23}$$

It was proven that there exists a lower bound on $\Sigma_{\boldsymbol{\theta}}$ for a given set of measurements (Cramer–Rao lower bound, EYKHOFF, 1974):

$$\Sigma_{\boldsymbol{\theta}} > \mathbf{J}^{-1} \text{ with } \mathbf{J} = E\left[\left(\frac{\partial}{\partial \boldsymbol{\theta}} \ln f(\mathbf{Z}_m \mid \boldsymbol{\theta})\right)\left(\frac{\partial}{\partial \boldsymbol{\theta}} \ln f(\mathbf{Z}_m \mid \boldsymbol{\theta})\right)' \mid \boldsymbol{\theta}_e\right], \tag{24}$$

where $\mathbf{J}$ is the Fisher information matrix. Using (20) and (21) it is found that

$$\frac{\partial}{\partial \boldsymbol{\theta}} \ln f(\mathbf{Z}_m \mid \boldsymbol{\theta}_e) = -\frac{\partial \boldsymbol{\zeta}'}{\partial \boldsymbol{\theta}} \Sigma_z^{-1} \boldsymbol{\zeta} = \frac{\partial \mathbf{Z}_e'}{\partial \boldsymbol{\theta}} \Sigma_z^{-1} \boldsymbol{\zeta}, \tag{25}$$

with

$$\boldsymbol{\zeta} = \mathbf{Z} - \mathbf{Z}_e.$$

Substitution in (24) results in

$$\mathbf{J}_e = E[(\frac{\partial \mathbf{Z}_e'}{\partial \boldsymbol{\theta}} \Sigma_z^{-1} \boldsymbol{\zeta})(\boldsymbol{\zeta}' \Sigma_z^{-1} \frac{\partial \mathbf{Z}_e}{\partial \boldsymbol{\theta}'}) \mid \boldsymbol{\theta}_e], \tag{26}$$

which reduces to

$$\mathbf{J}_e = \Psi \Sigma_z^{-1} \Psi' \quad \text{with} \quad \Psi = \frac{\partial \mathbf{Z}_e}{\partial \boldsymbol{\theta}'} \quad \text{and} \quad E[\boldsymbol{\zeta}\boldsymbol{\zeta}'] = E[\mathbf{z}\ \mathbf{z}']. \tag{27}$$

In (24), $\mathbf{J}$ is changed to $\mathbf{J}_e$, to indicate that $\mathbf{J}_e$ is calculated in terms of $\boldsymbol{\theta}_e$ instead of $\boldsymbol{\theta}$. However, if $\boldsymbol{\theta}_e$ is a good estimate of $\boldsymbol{\theta}$, the inverse of $\mathbf{J}_e$ can be used as an estimate of the Cramer–Rao lower bound. From experiments and simulations, it was found that the uncertainty on the estimates usually reaches the lower bound. From these results it can be concluded that an estimate of the uncertainty on the parameter estimations is found as an extra result of the estimation process.

Study of the Model Errors for Linear Systems

In literature, a lot of attention is paid to structure selection, and more specifically to order estimation in linear single input single output (SISO) systems. Order selection for noisy systems is a complex problem because a proposed model is built on noisy observations and not on exact theoretical knowledge. In this case the notion of desired model order is a more questionable one, due to the fact that the model is something that is constructed as an image of an unknown process. It need not cover all aspects of the process itself, so that the model may very well be of a lower complexity. Therefore, a class of models of interest has to be specified, and within this class the most suitable member has to be found, according to a certain predefined criterion. A lot of methods are developed to make this choice in a proper way. During the selection process a choice has to be made between the goodness of the fit and the parsimony of the model. The AIC, BIC, FPE etc. are techniques which where described in literature in the previous decade. These methods minimize a cost function based on the loss function of the fit and the number of parameters used in the model. The selected structure has to minimize this cost function. In (SCHOUKENS et al, 1984) the influence of model errors on the loss function (mean, variance) are studied for linear SISO. The MLE is used to estimate the parameters of the system, modelled by a transfer function. Starting from the value of the cost function, an idea about the model errors will be given. Using this knowledge it can be decided if it makes sense to improve a model (e.g. the model errors become very small) and if two models can be distinguished, starting from the observations. There are also some guide-lines given to combine multiple experiments in an optimal way to get the maximal information about the model structure. In this section only the results of this paper are reported.

It is shown that there are two contributions to the expected value of the cost function

$$E[K] = K_{\text{model}} + K_{\text{noise}} \tag{30}$$

with

$$K_{\text{noise}} = N - \frac{n_p}{2} \quad \text{and} \quad n_p \quad \text{the number of unknown parameters}$$

and

$$K_{\text{model}} \cong \frac{1}{2} \sum_{k=1}^{N} \frac{\left| t_k^2 \mid \mathbf{X}(\omega_k) \right|^2}{\mid T_{ek} \mid^2 \sigma_{\nu xk}^2 + \sigma_{\nu yk}^2}.$$

In this expression the following notations are used:

T_{ek} the exact value of the transfer function at frequency f_k,

t_k the model error on the transfer function.

In a lot of situations, the used input spectrum is flat, and in this case, a 'mean model error' can be defined as

$$t^2_{\text{mean}} = \frac{2K_{\text{model}}}{N \mid A \mid^2} \{ \mid T_e \mid^2 \sigma_a^2 + \sigma_b^2 \}. \tag{31}$$

Using this result, it is possible to get an idea about the value of the model errors, starting from the knowledge of the value of the cost function.

It was also shown that the optimal experiment strategy is choosing a minimal number of frequencies and averaging the experimental data before processing them. However, care should be taken by decreasing the number of frequencies because the number should be high enough to keep the asymptotic properties of the estimator. Also, the possibility of undetected modelling errors is increasing if the number of frequencies becomes too low.

Examples

The proposed method has been applied in many simulations and for numerous measurements. The parameter estimates were unbiased, even for very high noise levels and the covariance matrix of the estimates coincides with the matrix given by (28).

Example 1 : Estimation of the transfer function of a bandpass filter

As a first example, the previous theoretical results are illustrated on a bandpass filter. The transfer function is given in *Figure 2.* The measurements were done using a 12-bit digitizer, the sampling frequency was 50 kHz and the number of measurement points in the time domain was 2048. A multisine, which is the sum of sinusoids with optimized phase to get a minimal peak factor, was used as input signal. The frequencies of the sinusoids were $f_k = k \cdot 48.8$ Hz, $k = 5, 6, \ldots 20$ and the amplitudes were chosen equal to each other. The full scale was 2 Volts for both channels. From a preliminary noise analysis it turned out that the noise was white with a standard deviation of $1.24 \cdot 10^{-5}$ V on the real/imaginary parts of the Fourier coefficients.

The model used to describe the device under test is given by

$$T(\Omega) = \frac{a_0 + a_1\Omega + \ldots + a_6\Omega^6}{1 + b_1\Omega + b_2\Omega^2 + \ldots + b_6\Omega^6}, \qquad \Omega = j\omega.$$

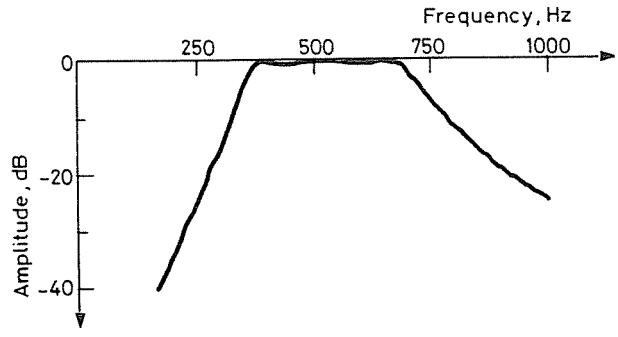

Fig. 2. Amplitude of the transfer function

25 experiments were done and the measurements were processed using 4 different models

— model 1: a_0, a_1, a_4, a_5, a_6 are set equal to zero,
— model 2: a_0, a_1, a_5, a_6 are set equal to zero,
— model 3: a_0, a_6 are set equal to zero,
— model 4: all parameters are estimated.

In *Table 1,* the results after processing 25 experiments are given. Because the measurements were done without locking the digitizer on the signal, it was not possible to average them before estimating. However, from the value of the cost function it is seen that the model error is dominant and from the section describing the device under test it results that both approaches have the same behaviour.

Table 1
Study of the behaviour of the cost function

model	cost function	mean model error in the passband(dB)
1	7700	0.0370
2	500	0.0095
3	387	0.0083
4	335	0.0078

This experiment shows that care should be taken during a model selection procedure. From *Table 1* it is seen that the model errors are dominant compared to the noise contribution, however, if the value of the cost function is transformed to a mean model error in the passband, it is seen that for model 2 the estimated error is about 0.02 dB, which is very small. From

this it can be concluded that the use of more complex models makes no more sense, because it is very difficult to make measurements where the systematic errors are not of the same order of magnitude. The estimated values of the transfer function for model 2 were also compared with direct measurements of the transfer function. It turned out that the differences in the passband were about 0.01 dB (*Figure 3*). The values of the parameters of the model 2 together with a statistical analysis (comparison of the uncertainty with the Cramer–Rao lower limit) can be found in (SCHOUKENS, 1985) and (PINTELON, 1988).

The run time to solve these problems is in the order of a few seconds on a Macintosh II computer (with 68020 processor and a 68881 coprocessor).

Example 2 : Black box modelling of the transfer characteristics of a loudspeaker

In a second example, it is shown that the method can also be used in black box modelling. The transfer function of a loudspeaker was measured, and a model was fitted to the data to describe the transfer characteristics of the loudspeaker (amplitude and phase). These results were used later on for digital correction of the loudspeaker (SCHOUKENS, 1985). For technical reasons the order of the correction filter was limited to 20. So the black box model was also selected to be a transfer function of 20^{th} order. As the results had to be used in a digital filter, a discrete model was selected and the estimation was done in the z-domain instead of the Laplace domain using an adapted estimation algorithm. 139 spectral lines were used, distributed as follows: 81 lines in (284.8 Hz, 10.05 kHz) and 58 lines in (10.21 kHz, 19.49 kHz). In *Figure 3* the result of the fit is shown.

Example 3: Analysis of a mechanical structure

In model analysis of mechanical systems, it is very important to know the resonance frequency and the damping of the different modes. This problem can also be formulated as the estimation of a transfer function on input (shaker) and output (acceleration sensor)

data. We used the algorithm to analyse the data of a mechanical structure. The excitation signal was a multisine consisting of 309 components with frequencies between 99.609 Hz to 400.39 Hz. The results are given in *Figure 4*. A transfer function of order 7 was used and after the estimation procedure the poles were calculated to know the resonance frequencies and the damping. The results for the two resonances of interest are: $f = 287.5$ Hz and damping $= 0.0030$; $f = 321.8$ Hz and damping $= 0.0022$.

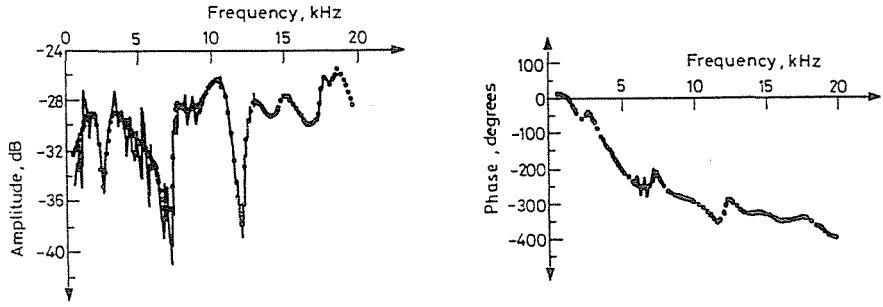

Fig. 3. Fit of loudspeaker characteristics
full line: measurements, •: model

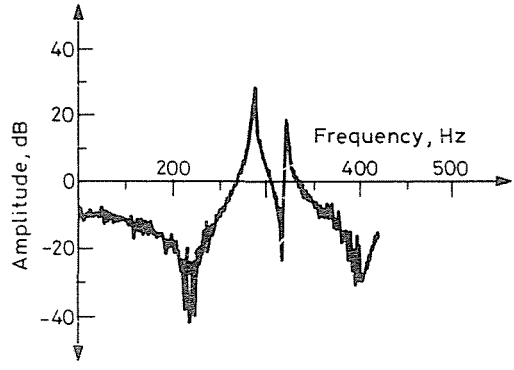

Fig. 4. Measured and estimated transfer function of a vibrating mechanical system

COMPARISON WITH EXISTING METHODS

Since the end of the fifties, a lot of effort has been applied to the development and study of the properties of identification methods for transfer functions. This investigation is today just as important (RAKE, 1980, LJUNG, 1985a, LJUNG, 1985b, THOMASETH et al, 1985). A wide variety of estimators have resulted, on the one hand the methods of LEVI (1959), SANATHANAN and KOERNER (1963), STROBEL (1960), LAWRENCE (1979),VAN DEN BOS (1974), the correlation method (LJUNG,1985b) and the empirical transfer function smoothing (LJUNG, 1985b) for the identification of dynamic continuous time systems in the frequency domain; while on the other hand the least squares (LS), generalized least squares (GLS),

instrumental variable (IV) (WONG, 1967), maximum likelihood (LJUNG, 1985a) and prediction error (LJUNG, 1985a) for the identification of dynamic discrete time systems. At present, most of the identification schemes available in the literature apply to difference equation models. The reason for this is that their representation is extremely well applicable for digital computers, and from a statistical point of view, noise handling in difference equations is less problematic than in differential equations. A disadvantage is that they only approximately describe continuous time models, which are the natural representation of almost all physical phenomena.

The IV method, which is commonly used to estimate discrete time transfer functions (STOICA, 1983; LJUNG, 1985a; LJUNG, 1985b), requires the construction of so-called 'instrumental time series', which are causally related to the components of the observed input and output, but independent of all the noncausally related components (noise) of these observations (WONG, 1967). Although the quality (accuracy) of the estimator depends mostly on the choice of these instrumental time series (SODERSTROM and STOICA, 1981; STOICA and SODERSTROM, 1983), no practical method can be found in the literature to calculate them. This problem is partially overcome by the refined instrumental variable (RIV) method (YOUNG and JAKEMAN, 1979; JAKEMAN and YOUNG, 1979) which generates the instrumental variables via a rather involved iterative algorithm.

The frequency domain methods of LEVI (1959), SANATHANAN and KOERNER (1963) and LAWRENCE (1979) (who proposed a better convergence alternative of SANATHAN's method) generate biased estimates (VAN DEN ENDEN et al, 1977), but no theoretical justification, from a statistical point of view, can be found in their work. STROBEL published a method (STROBEL, 1966) which is consistent under the restrictive assumption that the noise on the measurements, expressed in dB, must be Gaussian. The empirical transfer function estimate (ETFE) is proven to be consistent and closely related to the frequency analysis using the correlation method and to the time-domain prediction error methods (LJUNG,1985b).

We emphasize here that the methods described above (except VAN DEN BOS, 1974), employ only one disturbing noise source, called process noise or output noise (n_2 and n_3 in *Fig. 1*), in their model structure. However, there is no reason to neglect the input noise source in the model, since in electrical systems the input measurement noise is mostly of the same magnitude as the output measurement noise.

VAN DEN BOS proved the consistency of his method, for the number of time domain samples increasing to infinity, with the following assumptions (see VAN DEN BOS, 1974, pp. 79–87) :

1° the number of parameters to be estimated is less than or equal to twice the number of harmonics in the input and output signals (regularity condition normal equation),

2° the input and output signals are corrupted with stationary correlated noise. In other words, the proof is based on the fact that the noise on the Fourier coefficients tends to zero (in mean square sense) as the number of time domain samples increases (VAN DEN BOS, 1974, pp. 87). That is the main difference in the theoretical sphere from the method presented here.

From this brief review we conclude that there is no straightforward solution available in the literature to identify accurately the continuous time model shown in *Figure 1,* where the DUT is a dynamic, linear, time invariant system. This paper presents an original approach to fill some of these gaps and it is also the first time that a wide class of nonlinear systems can be practically identified.

Conclusion

In this paper an original method has been presented to estimate the parameters of linear and nonlinear systems. The method belongs to the class of MLE. Due to the general approach, the technique can be applied to a lot of problems using the same measurement setup. Due to the use of a time domain measurement which is transformed into the frequency domain, the necessary a priori information to build an MLE is easily derived without making difficult measurements and severe restrictions on the noise properties in the time domain. The covariance matrix of the parameters is obtained during the estimation process. It turned out that no extra calculations were needed to estimate the Cramer–Rao lower bound. From the experiments it was seen that this lower bound is reached which means that for the given measurements the estimates are those with the smallest uncertainty.

References

EYKHOFF, P. (1974): System Identification. Parameter and State Estimation. New York, John Wiley & Sons.

JAKEMAN, A. – YOUNG, P. (1979): Refined Instrumental Variable Method of Recursive Time Series Analysis. Part II: Multivariable Systems. *International Journal of. Control,* Vol. 29, No. 4, pp. 621–644.

KOLLÁR, I. (1986): The Noise Model of Quantization. *Proc. 1st International Symposium on the Measurement of Electrical Quantities,* Como, June 19–21, pp 125–129.

LAWRENCE, P. J. – ROGERS, .G.J. (1979): Sequential Transfer Function Synthesis as from Measured Data. *Proc. IEE,* Vol. 126, No. 1, pp. 104–106.

LEVI, E. C. (1959): Complex Curve Fitting. *IRE Transactions on Automatic Control,* pp. 37–43.

LJUNG, L. (1985a): Asymptotic Variance of Transfer Function Estimates Obtained by the Instrumental Variable Methods. *Proc. 7th IFAC Symposium on Identification and System Parameter Estimation,* York, U.K., July 3–7, Vol. 2, pp. 1341–1350.

LJUNG, L. (1985b): On the Estimation of Transfer Functions. *Proc. 7th IFAC Symposium on Identification and System Parameter Estimation,* York, U.K., July 3–7, Vol. 2, pp. 1653–1657.

PINTELON, R. (1988): Analysis and Application of a Maximum Likelihood Estimator for Linear Systems. Doct. Thesis, Vrije Universiteit Brussel, Department ELEC, Belgium.

RAKE, H. (1980): Step Response and Frequency Response Methods. *Automatica,* Vol.. 16, pp. 519–526.

RENDERS, H. – SCHOUKENS, J. – VILAIN, G. (1984): High Accuracy Spectrum Analysis of Sampled Discrete Frequency Signals by Analytical Leakage Compensation. *IEEE Trans. on Instrumentation and Measurement,* Vol. IM-33, No. 4, pp. 287–293.

SANATHANAN, C.K. – KOERNER, J. (1963): Transfer Function Synthesis as a Ratio of Two Complex Polynomials. *IEEE Transactions on Automatic Control,* pp. 56–58.

SCHOUKENS, J. (1985): Parameterestimatie in lineaire en niet-lineaire systemen met behulp van digitale tijdsdomein metingen. Doct. Thesis, Vrije Universiteit Brussels, Department ELEC, Belgium.

SCHOUKENS, J. – RENNEBOOG, J. (1986): Study of the Noise Influence on the Fourier Coefficients after a Discrete Fourier Transform. *IEEE Trans. on Instrumentation and Measurement,* Vol. IM-35, No. 3, pp. 278–286.

SCHOUKENS, J. – PINTELON, R. – RENNEBOOG, J. (1988): On the Limits of Order Estimation. *Preprints of the 8th IFAC Symposium on Identification and System Parameter Estimation,* August 27–31, Beijing, China, Vol. 2, pp. 1026–1031.

SCHROEDER, M. R. (1979): Synthesis of Low Peak-Factor Signals and Binary Sequences with Specified Fourier Amplitude Spectra. *International Journal of Control,* Vol. 30, No. 5, pp. 85–89.

SODERSTROM, T. – STOICA, P. (1981): Comparison of some instrumental variable methods D Consistency and Accuracy Aspects. *Automatica,* Vol. 17, pp. 101–115.

SORENSON, H. W. (1980): Parameter and State Estimation. Principles and Problems. New York, Marcel Dekker.

SPRIET, J.A. – VANSTEENKISTE G.C. (1982): Computer aided modelling and simulation. London, Academic Press.

STOICA, P. – SODERSTROM T.(1983): Optimal Instrumental Variable Estimation and Approximate Implementations. *IEEE Transactions on Automatic Control,* Vol. AC-28, No. 7, pp. 757–772.

STROBEL (1966): Continuous Linear Systems. *Proc. Third IFAC World Congress,* London, pp. 1.F1-1.F10.

THOMASETH, K. – YOUNG, P. – COBELLI, C. (1985): Recursive-Iterative Estimation of Multiple-Input Transfer Function Models. *Proc. 7th IFAC Symposium on Identification and System Parameter Estimation,* York, U.K., July 3–7, Vol. 2, pp. 1345–1350.

VAN DEN BOS, A. (1974): Estimation of Parameters of Linear Systems Using Periodic Test Signals. Doct. Thesis, Technische Hogeschool Delft, The Netherlands. Delftse Universitaire Pers.

VAN DEN ENDEN, A.W. – GROENENDAAL, G. C. – VAN DE ZEE, E. (1977): An Improved Complex Curve-Fitting Method. *Proceedings of the Conference on Computer Aided Design of Electronic, Microwave Circuits and Systems,* Hull (GB), pp. 53–58.

WONG, K.Y. – POLAK, E. (1967): Identification of Linear Discrete-Time Systems Using the Instrumental Variable Method. *IEEE Transactions on Automatic Control,* Vol. AC-12, No. 6, pp. 707–718.

YOUNG, P. - JAKEMAN A. (1979): Refined Instrumental Variable Methods of Recursive Time-Series Analysis, Part I: Single Input, Single Output Systems. *International Journal of Control,* Vol. 29, No. 1, pp. 1–30.

Address:

Prof. Johan Schoukens, Prof. Rik Pintelon, Prof. Jean Renneboog
Dienst ELEC
Vrije Universiteit Brussel
B-1050 Brussels, Pleinlaan 2.
Belgium