On Frequency-Domain Identification of Linear Systems

István Kollár, *Member, IEEE*

Abstract—**The maximum-likelihood estimation of the parameters of linear systems and the properties of the estimator (Estimator for Linear Systems, ELiS) have been described in a paper by R. Pintelon and J. Schoukens ("Robust identification of transfer functions in s- and z-domain," IEEE Transactions on Instrumentation and Measurement, vol. 39, pp. 565–573, Aug. 1990). The mathematics used in the development of the method and the proofs are rather involved. However, several statements can be understood in heuristic terms.**

This paper discusses the complex-domain description of the method, which results in much simpler expressions. The method is also compared to other formations, giving more insight into the properties of the estimate. It turns out that robustness is at least partly due to the least-squares formulation.

Derivations are avoided where possible, and intuitive explanations are given instead.

I. Introduction

A general model for the frequency-domain identification of linear systems is shown in Fig. 1. The system is represented by its transfer function $H(\Omega)$, where $\Omega = s = j\omega = j2\pi f$ in the Laplace domain, or $\Omega = z^{-1} = \exp(-j\omega T_s)$ in the z-domain, respectively, and H is a rational form, eventually extended by a delay term:

$$H(\Omega) = e^{-j\omega T_d} \frac{b_0\Omega^0 + b_1\Omega^1 + \cdots + b_n\Omega^n}{a_0\Omega^0 + a_1\Omega^1 + \cdots + a_m\Omega^m}. \quad (1)$$

The excitation signal is represented by a Fourier series with complex amplitudes X_k at angular frequencies ω_k. The response of the system is $Y_k = H(\Omega_k)X_k$. Both the measured input and output complex amplitudes are corrupted by noises a_x and a_y, respectively (errors-in-variables model), which are usually assumed to be Gaussian, uncorrelated between input and output, and also uncorrelated between different frequency points.

Using the above assumptions, a maximum-likelihood estimate of the system parameters is developed in real terms in [1]–[4], and the asymptotic properties are proven in [3] and [4]. It is claimed that the algorithm converged to a good solution in every case studied (for systems with no delay), and that its properties are robust with respect to the noise distribution. That is, it is consistent, asymptotically normally distributed, and the asymptotic covariance matrix can be given in terms of the covariance of the noise, using the expression of the Cramér–Rao lower bound, calculated for the case of Gaussian noise. It is also stated that a major advantage is that the estimate takes input *and* output noises into consideration, and this often provides superiority over, e.g., the prediction error method. In what follows, we are going to discuss these statements, give heuristic explanations and simple formulations, and explore the relationships with other methods.

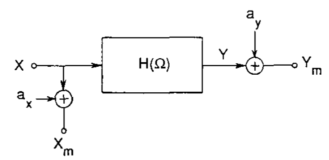

Fig. 1. The model to be investigated.

The purpose of this paper is threefold: first, to give a complex-domain formulation of the estimator, a more easily understandable one than the real-term form in [3] and [4]; second, to give some heuristic explanations of its properties on the basis of general principles, without using involved mathematics; and third, to discuss its relations to other models and situate it in a more general framework.

The paper is organized as follows. Section II presents the basics of the algorithm ELiS (Estimator for Linear Systems) in complex notations, resulting in rather simple expressions. In Section III, ELiS is presented as a least-squares estimate, which is an explanation for the robustness with respect to the noise distribution. Section IV gives the variance expressions of the estimate. Section V discusses the intuitive reasons why ELiS generally finds the global minimum of the cost function. Section VI presents alternative formulations of the estimation problem (different ways for the consideration of the noises), and shows that these lead to approximately the same cost function. Finally, Section VII discusses some possibilities of model validation, and examines the influence of the so-called nuisance parameters.

II. Formulation in Complex Notations

Let us investigate the model of Fig. 1. The time-domain measurements are transformed into the frequency domain, and the calculated complex input and output amplitudes at selected angular frequencies ω_k, $k = 1, \cdots,$

Manuscript received July 27, 1990; revised June 25, 1992.
The author was on leave at the Vrije Universiteit Brussels, Brussels, Belgium. He is with the Department of Measurement and Instrument Engineering, Technical University of Budapest, H-1521 Budapest, Hungary.
IEEE Log Number 9203658.

0018-9456/93$3.00

F are X_{mk} and Y_{mk}, respectively. The unknown parameters are the unknown coefficients of the numerator and the denominator of the transfer function (1), collected in the vector $\boldsymbol{P}$; further, the exact complex input and output amplitudes (vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$) are also unknown.

The basic equations can be written as

$$Y_k = H(\Omega_k, \boldsymbol{P})X_k, \qquad k = 1, 2, \cdots, F \tag{2a}$$

for the exact values, and

$$Y_{mk} = H(\Omega_k, \boldsymbol{P})(X_{mk} - a_{xk}) + a_{yk}, \qquad k = 1, 2, \cdots, F \tag{2b}$$

for the noisy observations, using $Y_{mk} = Y_k + a_{yk}$ and $X_{mk} = X_k + a_{xk}$.

In order to obtain a maximum-likelihood estimate of the unknown quantities, the conditional probability density function of the noises is needed. Assuming that the noise on the complex amplitudes is Gaussian and uncorrelated (which is usually at least approximately true after the DFT), its joint probability density function can be written as

$$\begin{aligned} p(\boldsymbol{a}_x, \boldsymbol{a}_y) &= \prod_{k=1}^{F} \frac{1}{2\pi\sigma_{xk}^2} \exp\left(-\frac{a_{Rxk}^2 + a_{Ixk}^2}{2\sigma_{xk}^2}\right) \prod_{k=1}^{F} \frac{1}{2\pi\sigma_{yk}^2} \cdot \exp\left(-\frac{a_{Ryk}^2 + a_{Iyk}^2}{2\sigma_{yk}^2}\right) \\ &= \prod_{k=1}^{F} \frac{1}{2\pi\sigma_{xk}^2} \exp\left(-\frac{a_{xk}\overline{a_{xk}}}{2\sigma_{xk}^2}\right) \prod_{k=1}^{F} \frac{1}{2\pi\sigma_{yk}^2} \cdot \exp\left(-\frac{a_{yk}\overline{a_{yk}}}{2\sigma_{yk}^2}\right) \end{aligned} \tag{3}$$

where a_{Rxk}, a_{Ixk}, a_{Ryk}, and a_{Iyk} are the real and imaginary parts of the input and the output noise amplitudes, respectively; σ_{xk} and σ_{yk} are the corresponding standard deviations; $\overline{a}$ is the complex conjugate of a; and $\boldsymbol{a}_x$ and $\boldsymbol{a}_y$ denote the vectors formed of a_{xk} and a_{yk}, respectively.

By expressing the noise variables in (3) in terms of X_{mk} and X_k ($a_{xk} = X_{mk} - X_k$, $a_{Rxk} = X_{Rmk} - X_{Rk}$, etc.), and using the assumption that σ_{xk} and σ_{yk} are known from a preceding noise analysis, the log-likelihood function is obtained:

$$\ln(L(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{P})) = \text{const} - \sum_{k=1}^{F}\left(\frac{(X_{mk} - X_k)(\overline{X_{mk} - X_k})}{2\sigma_{xk}^2}\right) - \sum_{k=1}^{F}\left(\frac{(Y_{mk} - Y_k)(\overline{Y_{mk} - Y_k})}{2\sigma_{yk}^2}\right). \tag{4}$$

where $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{P}$ are not independent of each other since (2a) must be fulfilled.

The maximization of (4) is equivalent to the minimization of the cost function

$$C(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{P}) = \sum_{k=1}^{F}\left(\frac{(X_{mk} - X_k)(\overline{X_{mk} - X_k})}{2\sigma_{xk}^2}\right) + \sum_{k=1}^{F}\left(\frac{(Y_{mk} - Y_k)(\overline{Y_{mk} - Y_k})}{2\sigma_{yk}^2}\right) \tag{5}$$

subject to the constraints

$$Y_k = H(\Omega_k, \boldsymbol{P})X_k, \qquad k = 1, 2, \cdots, F. \tag{6}$$

The constraints can be substituted into (5) in order to eliminate $\boldsymbol{Y}$. The result is a *nonlinear weighted least-squares* problem. Since, generally, we are not interested in $\boldsymbol{X}$ either, a better approach is to use the Lagrange-multiplier technique to eliminate both $\boldsymbol{X}$ and $\boldsymbol{Y}$. Fortunately, $\boldsymbol{X}$, $\boldsymbol{Y}$, and the multipliers can indeed be eliminated, and the following expression of the cost function is obtained for minimization:

$$C'(\boldsymbol{P}) = \sum_{k=1}^{F} \frac{|e^{-j\omega_k T_d} N(\Omega_k, \boldsymbol{P})X_{mk} - D(\Omega_k, \boldsymbol{P})Y_{mk}|^2}{2\sigma_{yk}^2 |D(\Omega_k, \boldsymbol{P})|^2 + 2\sigma_{xk}^2 |N(\Omega_k, \boldsymbol{P})|^2} \tag{7}$$

where $N(\Omega, \boldsymbol{P})$ and $D(\Omega, \boldsymbol{P})$ are the numerator and denominator of the transfer function, respectively.

Since (7) is a sum of quadratic terms (although nonlinear in $\boldsymbol{P}$ because of the denominators), $C'(\boldsymbol{P})$ can be minimized using powerful numerical techniques developed for nonlinear least-squares problems (Newton–Gauss method or Levenberg–Marquardt method). This can be done in complex terms (e.g., in Matlab; see [5]), paying attention to maintain that the elements of $\boldsymbol{P}$ are real or, alternatively, (7) can be written as a sum of squared real terms. In order to obtain start values for the iteration, the sum of the numerators in (7) can be minimized, which is an ordinary linear least-squares problem, having a unique solution, and can be readily solved by standard procedures.

In the following sections, we will refer to the above-described estimator as ELiS.

III. ELS as a Least-Squares Estimator

It is clear from the above description that *normality* of the frequency-domain noises ($\boldsymbol{a}_x$ and $\boldsymbol{a}_y$) was only used for the derivation of (5), which is a weighted least-squares cost function. It is generally true that for independent additive Gaussian noise with known variance, the maximum-likelihood estimate is nothing other than a (maybe weighted) LS estimate. Thus, ELiS, the maximum-likelihood estimate for the Gaussian case, is also a *least-squares* estimate. Consequently, the general properties of LS estimates will apply to it, independently of the noise distribution, and in the Gaussian case, those of the maximum-likelihood estimates as well. Indeed, LS estimates are, under rather mild conditions, consistent and asymptotically normally distributed, independently of the noise distribution, as was also stated for ELiS [3]–[4]. The heuristics behind this are quite simple: if all of the random errors have contributions in the same order of magnitude to the uncertainty of the estimate, for a very large number of observations, a linear approximation of the weak dependence may be used. Consequently, the error of the estimate is influenced by a large number of additive, small impacts, and the law of large numbers and the central limit theorem can be employed.

Thus, the robustness of the estimate with respect to the noise distribution may be understood in terms of the LS solution.

Exact conditions of the asymptotic properties of LS estimates for certain problems can be found in the literature. In [6]–[8], there are similar statements as in [3] and [4] concerning the asymptotic properties of the *prediction error method*. Also, there are papers on asymptotic properties of general nonlinear LS estimates (e.g., [9]). Unfortunately, although these are closely related to the case of ELiS, there are significant differences. The prediction error model operates in the time domain, assuming no input noise. Reference [9] deals with the model

$$y_k = f_k(\theta) + e_k, \qquad k = 1, 2, \cdots \tag{8}$$

where θ is the vector of unknown parameters, and the errors are independent, identically distributed random variables. This model is rather general. However, in contrast to (8), the number of parameters in (5) increases with every new measurement (cf. X_k). Thus, with respect to the asymptotic properties, (5) *is not a regular nonlinear least-squares problem*.

Similar but simpler models than that given by (2b) (e.g., just one equation or equations linear in the parameters) are discussed in [10, ch. 29: pp. 399–443]. In its terms, (2a) is called a *functional relationship*, and (2b) a *structural relationship*. X_k and Y_k are *nuisance parameters*, which appear in the model, although we are not interested in them.

Considering the reduced-order formulation given by (7), there is no apparent relation to (8). As we are going to see in Section VI, by the transfer of the input noise to the output, applying an appropriate transformation, a very close, but not identical, model can be obtained.

Seemingly, the case of ELiS is somewhat more general than those already investigated in the statistical literature. Moreover, the model investigated in [3] and [4] and the proofs can be further generalized to a class of problems wider than just the estimation of transfer functions. ELiS is, in this context, an application of a general theory of the LS estimation of structural relationships, yet to be elaborated.

IV. Covariance Matrix of the Estimates

The quality of estimates can be well characterized by their variances. There is a lower bound of the variances of estimates: this is the so-called Cramér–Rao lower bound (CRLB). This bound approximately gives the variances (or better stated, the covariance matrix of the estimates) in the case of maximum-likelihood estimation if the number of observations is sufficiently high and mild conditions are fulfilled. In our case, a rule of thumb is to provide that F is at least the double of the number of unknown parameters.

Since the above estimator is a maximum-likelihood one in the Gaussian case, the covariance matrix can be approximately computed from the expression of the corresponding CRLB. In the non-Gaussian case—since the covariance depends first of all on the structure of the algorithm itself, and to a lesser extent on the distributions of the noises with small individual impacts—the covariance will be approximately the same.

Before giving the expressions of the CRLB, an important aspect requires discussion. From (1), it is clear that if all of the coefficients are to be estimated, the result is undetermined in a multiplier factor; thus, some additional constraints are to be imposed, e.g., one of the parameters can be fixed. Let us denote the vector of the estimated ("free") coefficients by $\boldsymbol{P}_f$, and let us maintain the constraints (6) during the derivations. In this case,

$$\operatorname{cov}\{[\boldsymbol{X}_{ML}, \boldsymbol{P}_{fML}]\} = \mathcal{E}\left\{\left(\frac{\partial C(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{P})}{\partial[\boldsymbol{X}, \boldsymbol{P}_f]}\right)^T \left(\frac{\partial C(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{P})}{\partial[\boldsymbol{X}, \boldsymbol{P}_f]}\right)\right\}^{-1} \tag{9}$$

where T denotes the transpose of the row vector, and $\boldsymbol{X}$ and $\boldsymbol{P}$ denote the true values.

Since we are interested in the covariance matrix of $\boldsymbol{P}_{fML}$ only, (9) can be reduced. After some calculations, the following expression is obtained:

$$\operatorname{cov}\{\boldsymbol{P}_{fML}\} = \mathcal{E}\left\{\left(\frac{\partial C'(\boldsymbol{P})}{\partial \boldsymbol{P}_f}\right)^T \left(\frac{\partial C'(\boldsymbol{P})}{\partial \boldsymbol{P}_f}\right)\right\}^{-1}. \tag{10}$$

Note that the measured quantities in (7) are substituted by the exact values.

It is easy to see that if (6) is fulfilled, C has the same derivative as C', so we also have an alternative form:

$$\operatorname{cov}\{\boldsymbol{P}_{fML}\} = \mathcal{E}\left\{\left(\frac{\partial C(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{P})}{\partial \boldsymbol{P}_f}\right)^T \left(\frac{\partial C(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{P})}{\partial \boldsymbol{P}_f}\right)\right\}^{-1}. \tag{11}$$

There are two important problems with these expressions.

1) They contain the true (unknown) values of the parameters: $\boldsymbol{P}$ in C' $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{P}$ and,

2) The expected value has to be evaluated with respect to the measurement noises.

In practice, the parameters $\boldsymbol{P}$ in C' can be replaced by their estimators: if the estimation is good, the introduced error will not be significant. We are going to treat this in some more detail in Section VII.

When $F >> 1$, the expected value is well approximated by the actual value of the matrix between the brackets, since the summations in (7) or (5) effectively decrease the relative variances.

V. Local Minima of the Cost Function

The cost function (7) is nonlinear; thus, there is no guarantee that it has no local minima in addition to the global one. Gradient-based methods are only capable of finding a local minimum in the direction their algorithm drives them. This may be problematic since the advantageous properties of ELiS are guaranteed only if the solution found is at least close to the global minimum. However, it is claimed in [1]–[4] that this is usually the case.

In my opinion, the power of ELiS lies at least partly in the fact that the *starting values of the iteration* are very well chosen. Let us consider the form of (7). It is similar to an ordinary linear least-squares problem (sum of the numerators), but *parameter-dependent weights* are introduced. Every LS formulation somehow minimizes the squares of the differences, and the weights "only modify" this minimization. Thus, the initial ordinary linear LS step—which finds the global minimum of the sum of the numerators of (7)—is a key to the good convergence properties of ELiS since it already brings us close to the desired solution.

Nevertheless, there may still be situations where the minimization procedure can be based on bad local minima. A solution may be a very thorough exploration of the cost function, which is rather hopeless for higher orders. Another possibility is the method proposed in [11]. Here, for the Gaussian case, a modified cost function (average of cost functions for different sets of measurements) is formed, and it is proven that this estimate has better properties than the one based on the LS cost function only—at the price of more calculations. A third possibility is to increase the number of frequency points when minimizing (7). This method is very straightforward, but one may quickly run out of memory, and also, the calculation time increases rapidly.

Since the standard iteration algorithms usually converge to an efficient estimate from the above starting values [1]–[4], the above improvements are rarely necessary. According to practical experience, the size of the iteration steps may have to be restricted sometimes (e.g., by using the Levenberg–Marquardt algorithm) in order to avoid the iteration "jumping out" from the deepest valley.

VI. Considering the Input–Output Noise Model

A key feature of ELiS is the consideration of both input and output noises.

In this section, we will show that the input and output noises also can be taken into consideration in other frequency-domain schemes, even if sometimes only approximately, and the very same cost function can be obtained as in ELiS.

A. Equation Error Formulation

First, let us assume that the signal-to-noise ratio is large, that is, $X_m \approx X$ and $Y_m \approx Y$. It seems to be quite reasonable to minimize the following sum:

$$C_W(P) = \sum_{k=1}^{F} W_k |e_k|^2 = \sum_{k=1}^{F} W_k \left| e^{-j\omega_k T_d} \frac{N(\Omega_k, P_N)}{D(\Omega_k, P_D)} - \frac{Y_{mk}}{X_{mk}} \right|^2 \quad (12)$$

where the weights W_k are to be chosen as reciprocals of the variances of the corresponding terms (*minimum chi-square estimate*, [10, pp. 101–104]). Since

$$\text{var}\{X_{mk}\} = \mathcal{E}\{(X_{mk} - X_k)\overline{(X_{mk} - X_k)}\} = 2\sigma_{xk}^2 \quad \text{and} \quad \text{var}\{Y_{mk}\} = 2\sigma_{yk}^2 \quad (13)$$

using first-order approximations for small noise ($\sigma_{xk}^2 << |X_k|^2$),

$$\text{var}\{e_k\} = \text{var}\left\{\frac{Y_{mk}}{X_{mk}}\right\} \approx \left|\frac{Y_k}{X_k}\right|^2 \left(\frac{2\sigma_{yk}^2}{|Y_k|^2} + \frac{2\sigma_{xk}^2}{|X_k|^2}\right) = \frac{2}{|X_k|^2}\left(\sigma_{yk}^2 + \left|\frac{Y_k}{X_k}\right|^2 \sigma_{xk}^2\right) \approx \frac{2}{|X_{mk}|^2}\left(\sigma_{yk}^2 + \left|\frac{N(\Omega_k, P_N)}{D(\Omega_k, P_D)}\right|^2 \sigma_{xk}^2\right). \quad (14)$$

Substituting $W_k = 1/\text{var}\{e_k\}$ into (12), we obtain exactly (7).

B. Prediction Error Formulation

Another possibility is the following: let us use a prediction error model with all of the noises calculated with reference to the output, using an appropriate transformation! The minimum chi-square cost function is, in this case,

$$C_{pe}(P) = \sum_{k=1}^{F} W_k |e_k|^2 = \sum_{k=1}^{F} W_k \left| e^{-j\omega_k T_d} \frac{N(\Omega_k, P_N)}{D(\Omega_k, P_D)} X_{mk} - Y_{mk} \right|^2 \quad (15)$$

where the weights are again to be chosen as reciprocals of the corresponding variances:

$$\frac{1}{W_k} = 2\sigma_{yk}^2 + \left|\frac{N(\Omega_k, P_N)}{D(\Omega_k, P_D)}\right|^2 2\sigma_{xk}^2 \quad (16)$$

and the ELiS cost function is obtained again. It is remarkable that in the last case, no approximation was used. The referral of the input noise to the output can be interpreted as follows: we formulate an equivalent problem, where the system is excited not by the amplitudes X, but by the amplitudes X_m. This formulation is close to (8), and probably the proof of [9] can be extended to the case of ELiS. However, there is one very significant difference between the two models: in (15), the variances of the individual terms [see (16)] depend on the parameters to be estimated. Because of this very fact, the minimization of (15) is not equivalent anymore to the maximum-likelihood formulation: the equivalence of the ML and the LS problem was based on the assumption that σ_{xk}^2 and σ_{yk}^2 are known constants [see the constant term in (4)].

VII. Model Validation

The quickest and most straightforward way of checking if the identified model is valid is to examine the value of the cost function (7) for the estimated values. It is obvious from (15) that the cost function is χ^2-distributed, with

$2F - n_p$ degrees of freedom, where n_p is the number of estimated parameters; if the model is valid, the value of the cost function should not be larger than a test value, calculated from the χ^2-distribution. The χ^2-value can also be used for determining the excess of modeling errors [4].

Another possibility is to examine the residuals. However, here we already run into problems since, for this, we would need good estimates of the *nuisance parameters* too. These estimates can also be calculated using the estimated values of the transfer function, but they are not consistent, i.e., the variance does not decrease to zero as F increases. It can even be shown that the estimates of the complex input and output amplitudes at a given frequency (X_k and Y_k) will be correlated, and the frequency dependence of the variances of the residuals will also differ from the noise spectra. The heuristic explanation is as follows. Since X_k and Y_k appear in just one term of (5) only, for the estimation of X_k and Y_k, we may minimize

$$C_k(X_k, Y_k) = \frac{(X_{mk} - X_k)\overline{(X_{mk} - X_k)}}{2\sigma_{xk}^2} + \frac{(Y_{mk} - Y_k)\overline{(Y_{mk} - Y_k)}}{2\sigma_{yk}^2} \quad (17)$$

with the constraint

$$Y_k = H(\Omega_k, \boldsymbol{P}_{ML})X_k. \quad (18)$$

This minimization will do nothing other than "distribute" the power of the noise samples between the input and output amplitudes in a balanced way, resulting in correlated residuals with finite variances.

However, there is one test that can be performed using the residuals: the residuals *referred to the output* (or in other words, the residuals of the cost function C') can be investigated. These residuals should be Gaussian, have zero mean and variance given by (16), and should be only slightly correlated at different frequencies (cf. the $2F - n_p$ degrees of freedom of the χ^2-variable). These properties can be tested, e.g., an excessive mean value (distortion) indicates deviations from the assumed model of the system.

The problem of not having good estimates of the nuisance parameters also arises when trying to evaluate the Cramér–Rao lower bound (11) since here we have to substitute the exact values by some approximations. Luckily enough, because we maintain (6), the covariance matrix is not very sensitive to small variations of the complex amplitudes; thus, the calculated CRLB will approximate the true covariance values well.

VIII. Summary

A complex-domain formulation of the LS estimation of the parameters of linear systems has been presented. It has been shown that several properties of the estimation can be well explained in heuristic terms. It has also been pointed out that on the one hand, the estimator belongs to a general class of structural relation estimation; on the other hand, the proofs given in [3] and [4] apply to a wider class of problems than just the estimation of transfer functions.

Acknowledgment

The author is very much indebted to J. Schoukens and R. Pintelon for fruitful discussions and suggestions concerning the ideas presented in this paper.

References

[1] J. Schoukens, R. Pintelon, and J. Renneboog, "A maximum likelihood estimator for linear and nonlinear systems—A practical application of estimation techniques in measurement problems," *IEEE Trans. Instrum. Meas.*, vol. 37, pp. 10–17, Jan. 1988.

[2] —, "Maximum likelihood estimation of the parameters of linear systems," *Periodica Polytechnica Ser. Elec. Eng.*, vol. 33, no. 4, pp. 165–182, 1989.

[3] R. Pintelon and J. Schoukens, "Robust identification of transfer functions in *s*- and *z*-domain," *IEEE Trans. Instrum. Meas.*, vol. 39, pp. 565–573, Aug. 1990.

[4] J. Schoukens and R. Pintelon, *Identification of Linear Systems—A Practical Guideline to Accurate Modelling.* Oxford, U.K.: Pergamon, 1991.

[5] I. Kollár, R. Pintelon, and J. Schoukens, "Frequency domain system identification toolbox for Matlab," in *Proc. 9th IFAC/IFORS Symp. Identification and Syst. Parameter Estimation* (Budapest, Hungary, July 1991), Cs. Bányász and L. Keviczky, Eds. Oxford, U.K.: Pergamon, 1991, pp. 1243–1246.

[6] G. C. Goodwin and R. L. Payne, *Dynamic System Identification.* New York: Academic, 1977.

[7] L. Ljung, *System Identification: Theory for the User.* Englewood Cliffs, NJ: Prentice-Hall, 1987.

[8] L. Ljung and P. E. Caines, "Asymptotic normality of prediction error estimation for approximate system models," *Stochastics*, vol. 3, pp. 29–46, 1979.

[9] R. I. Jennrich, "Asymptotic properties on non-linear least squares estimators," *Ann. Math. Statist.*, vol. 40, no. 2, pp. 633–643, 1969.

[10] M. Kendall and A. Stuart, *The Advanced Theory of Statistics, Vol. 2.* London, U.K.: Griffin, 1979.

[11] H. O. Hartley and A. Booker, "Nonlinear least squares estimation," *Ann. Math. Statist.*, vol. 36, pp. 638–650, 1965.