ESTIMATION WITH KNOWN NOISE MODEL

R. Pintelon and **J. Schoukens**
Department ELEC, Vrije Universiteit Brussel, Belgium

Keywords: estimators, frequency domain, discrete-time, continuous-time

Contents

Summary

This chapter gives an overview of frequency domain identification methods for single input, single output systems. Estimators such as the (weighted) linear least squares, the nonlinear least squares, the (weighted) total least squares, the maximum likelihood, and the subspace algorithms are discussed in detail.

The interrelation between the different approaches is highlighted through a study of the (equivalent) cost functions. Special attention is also paid to global minimizers that try to approximate the maximum likelihood estimator. The properties of the different approaches are illustrated by means of an "on-line" simulation example. A theoretical toolbox is provided that allows us to verify quickly the basic properties of the estimators through some elementary manipulations of the cost function.

1. Introduction

In this chapter we handle the identification of the plant model assuming that the noise model is known exactly. We give an overview of frequency domain identification methods for single input single output systems (Section 3). Afterwards, the particularities of systems with time delay (Section 5.1), systems in feedback (Section 5.2) and high order systems (Section 5.3) are discussed.

A second-order system $G(s,\theta) = 1/(1+s+s^2)$ is used as "on-line" illustration in Section 3. Figure 1 shows the true transfer function and the simulated noisy frequency response data (see Appendix for more information concerning the generation of the simulation data).

Before starting with the overview, we discuss the type of data (experiments) we can handle (Section 1.1), introduce some notations for the parametric plant models (Section 1.2) and present the general form of the identification algorithms (Section 2.1). Section 2.3, quick tools to analyze estimators, makes it possible to reveal some basic properties of estimators through a simple analysis of the cost function.

Combined with Section 2.4, which gives the general asymptotic properties of estimators minimizing a cost function that is quadratic-in-the-measurements, it allows an easy verification and understanding of the properties of the different estimators described in Section 3.

To simplify the notations we limit the discussion in this chapter to (broad band) periodic excitations. The results are, however, also valid for broad band random excitations. Indeed, the non-periodicity of the excitation signal simply results in an additional polynomial in the transfer function model (see *Frequency Domain System Identification*, Section 3.2.1).

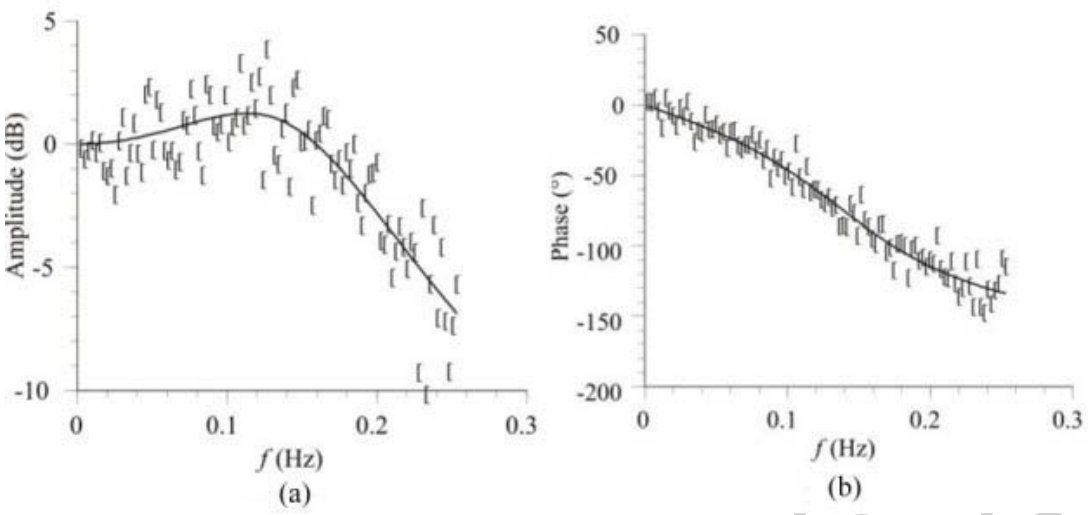

Figure 1. Second-order example $G(s,\theta) = 1/(1+s+s^2)$: true transfer function (solid line) and simulated noisy data (dots).

1.1. Frequency Domain Data

The identification starts from measured input/output (DFT) spectra $U(k)$, $Y(k)$,

$$\begin{aligned} Y(k) &= Y_0(k) + N_Y(k) \\ U(k) &= U_0(k) + N_U(k) \end{aligned} \qquad (1)$$

with $U_0(k)$, $Y_0(k)$ the true unknown values, or from a measured frequency response function (FRF) $G(\Omega_k)$,

$$G(\Omega_k) = G_0(\Omega_k) + N_G(k) \qquad (2)$$

with $G_0(\Omega_k)$ the true unknown value, at a set of F frequencies $\Omega_k, k = 1,2,\ldots,F$, which may be a (sub)set of the DFT frequencies. Note that (2) is a special case of (1) with $Y(k) = G(\Omega_k)$ and $U(k) = 1$. The $2F$ complex valued vector Z contains the measured input/output (DFT) spectra

$$Z^T = [Z^T(1)Z^T(2)\ldots Z^T(F)] \text{ with } Z^T(k) = [Y(k)U(k)],\ k = 1,2,\ldots,F \qquad (3)$$

It is related to the true values by $Z = Z_0 + N_Z$, where the disturbing noise N_Z has zero mean and is independent of Z_0.

The frequency domain data (1), (2) can be obtained via time domain or frequency domain experiments. In a *time domain experiment* a broad band periodic excitation is applied to the plant and the steady state response is observed over an integer number of periods. N samples of the input and output signals are measured. These N input/output samples are transformed to the frequency domain using the discrete Fourier

transform. $F \leq N/2+1$ DFT frequencies of the input and output DFT spectra are used for the identification. In a *frequency domain experiment* a single sine excitation is applied to the plant and the input/output spectra of the steady state response are measured at the excited frequency. This experiment is repeated at F different frequencies. For example, high frequency network analyzers (micro-wave measurements) and impedance analyzers follow this measurement procedure. Also most dynamic signal analyzers have such a measurement mode. For both experiments the frequency domain errors $N_U(k)$ and $N_Y(k)$ in (1) are related to the noise sources in Fig. 2 as

$$N_Y(k) = N_g(k)G_0(\Omega_k) + M_Y(k) + N_P(k)$$
$$N_U(k) = N_g(k) + M_U(k) \qquad (4)$$

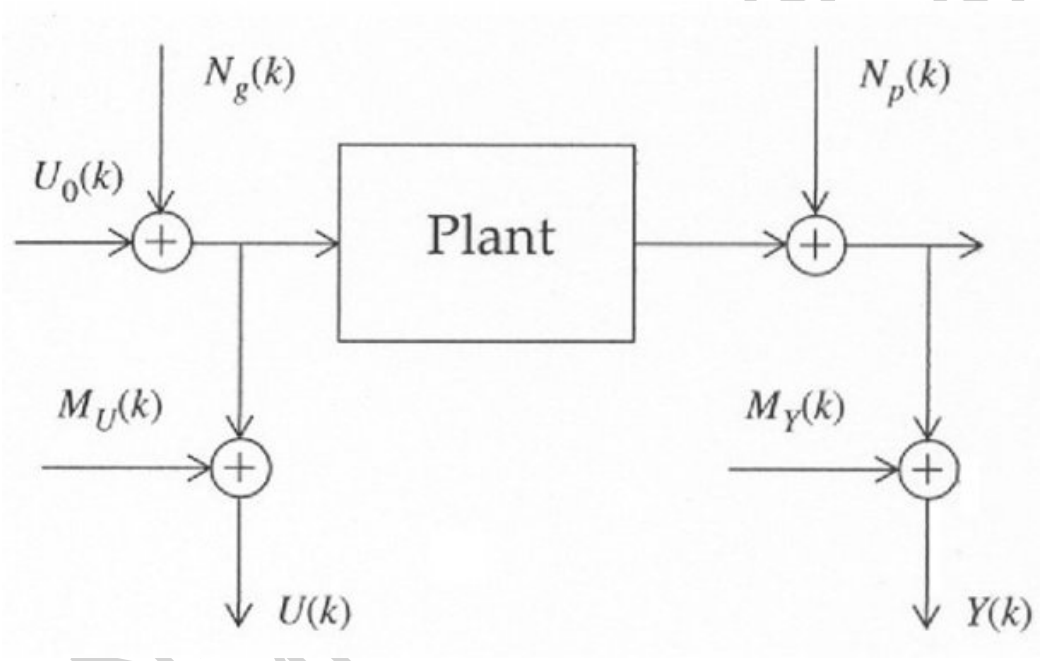

Figure 2. Measurement of a plant using a time or frequency domain experiment. $N_g(k)$ is the generator noise, $M_U(k)$ and $M_Y(k)$ are the input and output measurement errors, and $N_p(k)$ is the process noise.

with $G_0(\Omega_k)$ the plant transfer function.

Due to the imperfections of the measurement devices it is recommended not to use measurements at DC and in the neighborhood of the Nyquist frequency. Indeed, acquisition units mostly introduce DC offset errors and anti-alias protection is mostly only guaranteed up to about 80% of the Nyquist frequency.

The measurements can also be the result of a linearization of a nonlinear system at an operating point. This will introduce DC-values in the input and output signals which are not compatible with the linear model, and hence should be removed.

Ideally we would like to know the properties of an estimator for each finite value of F. Except for the (weighted) linear least squares, this is only possible for "sufficiently large", values of F. To analyze the stochastic properties of the estimators for F "sufficiently large" we will make a mental experiment where the number of frequencies F tends to infinity.

For a frequency domain experiment this implies that the number of single sine measurements F tends to infinity, while for a time domain experiment this implies that the number of measured time domain samples N tends to infinity. Note that we do not consider time domain experiments $(N \to \infty)$ with periodic signals containing a fixed number (independent of N) of frequencies F. Indeed, for such experiments the signal-to-noise ratio tends to infinity as $N \to \infty$ at the excited DFT frequencies, and hence all the estimators considered in Section 3 would be consistent in a trivial manner. For broad band multisine excitations the signal-to-noise ratio per spectral line remains an $O(N^0)$ so that consistency is a non-trivial issue.

1.2. Plant Model

For periodic excitations the true input and output DFT spectra are related by

$$Y_0(k) = G(\Omega_k, \theta_0)U_0(k) \text{ with } G(\Omega, \theta) = \frac{\sum_{r=0}^{n_b} b_r \Omega^r}{\sum_{r}^{n_a} a_r \Omega^r} \text{ and } \theta = [a_0, \ldots, a_{n_a}, b_0, \ldots, b_{n_b}]^T \quad (5)$$

where $\Omega = z^{-1}$ for discrete-time systems, $\Omega = s$ for continuous-time system, $\Omega = \sqrt{s}$ for diffusion phenomena and $\Omega = \tanh(\tau_R s)$ for commensurate micro-wave filters. In practice the true input/output DFT spectra $U_0(k)$, $Y_0(k)$ and the true model parameters θ_0 are unknown so that $Y(k) - G(\Omega_k, \theta)U(k) \neq 0$. For any parameterization given in Section 3.2.1 of *Frequency Domain System Identification* (rational form, partial fraction expansion and state space representation) we can use the *output error* which is the difference between the observed output $Y(k)$ and the modeled output $Y(\Omega_k, \theta)$

$$Y(k) - Y(\Omega_k, \theta) \text{ with } Y(\Omega_k, \theta) = G(\Omega_k, \theta)U(k) \quad (6)$$

For the *rational form* (5), it is convenient also to introduce the *equation error* $e(\Omega_k, \theta, Z(k))$ which is the difference $Y(k) - G(\Omega_k, \theta)U(k)$ multiplied by $A(\Omega_k, \theta)$

$$e(\Omega_k, \theta, Z(k)) = A(\Omega_k, \theta)Y(k) - B(\Omega_k, \theta)U(k) \quad (7)$$

Unless mentioned otherwise, we will assume that the parameterization of the plant model is identifiable. It implies that the parameter vector θ contains only the free parameters of the model; for example, all the numerator and denominator coefficients of the rational form $G(\Omega, \theta)$ in (5) except $a_0 = 1$. Note, however, that from a numerical

point of view it is often better to use the full overparameterized form in combination with dedicated numerical methods (Pintelon and Schoukens, 2001).

1.3. Noise Model

The measured input/output (DFT) spectra $U(k)$, $Y(k)$, are related to the true unknown values $U_0(k)$, $Y_0(k)$ by

$$\begin{aligned} Y(k) &= Y_0(k) + N_Y(k) \\ U(k) &= U_0(k) + N_U(k) \end{aligned} \tag{8}$$

where $N_U(k)$, $N_Y(k)$ are circular complex distributed random variables with zero mean and second order moments

$$\sigma^2_{VW}(k) = E\{N_V(k)\bar{N}_W(k)\} \text{ and } E\{N_U(k)N_V(k)\} = 0 \text{ for } V,W = Y \text{ and/or } U \tag{9}$$

(Picinbono, 1993). Eq. (9) implies that the real and imaginary parts of the noise are uncorrelated and have equal variance. These conditions are met in practice for a time and frequency domain experiment (Brillinger, 1981). It is also assumed that the input/output disturbances $N_U(k)$, $N_Y(k)$ are independent of the true (unknown) excitation $U_0(k)$.

2. Estimation Algorithms – General

2.1. General Form of Cost Functions

Most algorithms discussed in this chapter minimize a "quadratic-like" cost function $V(\theta, Z)$

$$V(\theta, Z) = \varepsilon^H(\theta, Z)\varepsilon(\theta, Z) = \sum_{k=1}^{F} \left|\varepsilon(\Omega_k, \theta, Z(k))\right|^2 \tag{10}$$

where the residual $\varepsilon(\Omega_k, \theta, Z(k))$ is a weighted version of the output error $Y(k) - Y(\Omega_k, \theta)$ (6) or the equation error $e(\Omega_k, \theta, Z(k))$ (7). Note that $\varepsilon_{[k]}(\theta, Z) = \varepsilon(\Omega_k, \theta, Z(k))$ depends on the measurements at frequency Ω_k only. The minimizer of (10) is denoted by $\hat{\theta}(Z)$.

A first important subclass of (10) are the cost functions $V(\theta, Z)$ which are *quadratic-in-the-measurements* Z. For these cost functions the residual $\varepsilon(\theta, Z)$ is linear in $Z = Z_0 + N_Z$ and can be written as

$$\varepsilon(\theta, Z) = \varepsilon(\theta, Z_0) + \varepsilon(\theta, N_Z) \tag{11}$$

with $\varepsilon(\theta,0)=0$. Using (11), (10) becomes

$$V(\theta,Z)=V(\theta,Z_0)+V(\theta,N_Z)+2\,\mathrm{Re}(\varepsilon^H(\theta,Z_0)\varepsilon(\theta,N_Z)) \qquad (12)$$

Examples are the linear least squares (Section 3.1), the nonlinear least squares (Section 3.2), the total least squares (Section 3.3), and the maximum likelihood (Section 3.4) estimators.

A second important subclass of (10) is that of the cost functions $V(\theta,Z)=f(\theta,\eta(Z),Z)$ which depend on an initial guess $\eta(Z)$ of the model parameters,

$$V(\theta,Z)=\varepsilon^H(\theta,Z)\varepsilon(\theta,Z)=\sum_{k=1}^{F}\left|\varepsilon(\Omega_k,\theta,\eta(Z),Z(k))\right|^2=f(\theta,\eta(Z),Z) \qquad (13)$$

and which are quadratic-in-the-measurements Z when $\eta(Z)$ in (13) is replaced by a non-random vector η. The weighted linear least squares (Section 3.1) and bootstrapped total least squares (Section 3.5) belong to this class.

To study the asymptotic behavior of the identification algorithms it is convenient to scale the cost function with the number of frequencies, $V_F(\theta,Z)=V(\theta,Z)/F$. The expected value of the cost function $V_F(\theta)=E\{V_F(\theta,Z)\}$ and its minimizer $\tilde{\theta}(Z_0)$ play an important role in the convergence analysis of the estimate $\hat{\theta}(Z)$.

Indeed, under some suitable assumptions (see Pintelon and Schoukens, 2001) it can be shown that $\hat{\theta}(Z)$ converges (in some stochastic sense) to $\tilde{\theta}(Z_0)$ (see also Section 2.4). Hence, the consistency (convergence to the true value θ_0) of $\hat{\theta}(Z)$ can be verified by simple analysis of $V_F(\theta)$. When model errors are present, then $\tilde{\theta}(Z_0)\neq\theta_0$ and $\tilde{\theta}(Z_0)$ will vary as the number of frequencies F increases.

The deterministic convergence of $\tilde{\theta}(Z_0)$ to some limit value θ_*, depends on the way data (frequencies) are added in the time or frequency domain experiment. The notations introduced are summarized in Table 1.

Cost function	$V_F(\theta,Z)=\begin{cases}V(\theta,Z)/F\\ f(\theta,\eta(Z),Z)/F\end{cases}$	$V_F(\theta,Z)=\begin{cases}E\{V(\theta,Z)/F\}\\ E\{f(\theta,\eta(Z),Z)/F\}\end{cases}$	$V_*(\theta)=\lim_{F\to\infty}V_F(\theta)$
Minimizer	$\hat{\theta}(Z)$	$\tilde{\theta}(Z_0)$	$\theta_*=\lim_{F\to\infty}\tilde{\theta}(Z_0)$

Table 1. Overview of notations frequently used. $\eta(Z)$ is an (initial) estimate of the model parameters and η_* is its limit value.

2.2. Minimization of Cost Functions

Often, a Newton-Gauss type algorithm is used to find the minimizer $\hat{\theta}(Z)$ of (10). Rewriting (10) as $V(\theta, Z) = \varepsilon_{\text{re}}^T(\theta, Z)\varepsilon_{\text{re}}(\theta, Z)$, where $(\,)_{\text{re}}$ stacks the real and imaginary parts on top of each other,

$$\varepsilon_{\text{re}}(\theta, Z) = \begin{bmatrix} \text{Re}(\varepsilon(\theta, Z)) \\ \text{Im}(\varepsilon(\theta, Z)) \end{bmatrix} \tag{14}$$

the i th iteration step of this algorithm is given by

$$J_{\text{re}}^T(\theta^{(i-1)}, Z) J_{\text{re}}(\theta^{(i-1)}, Z) \Delta\theta^{(i)} = -J_{\text{re}}^T(\theta^{(i-1)}, Z)\varepsilon_{\text{re}}(\theta^{(i-1)}, Z) \tag{15}$$

with $\Delta\theta^{(i)} = \theta^{(i)} - \theta^{(i-1)}$ and $J(\theta, Z) = \partial\varepsilon(\theta, Z)/\partial\theta$ the Jacobian of the vector $\varepsilon(\theta, Z)$. Using complex numbers, (15) can be written as

$$\text{Re}(J^H(\theta^{(i-1)}, Z) J(\theta^{(i-1)}, Z))\Delta\theta^{(i)} = -\text{Re}(J^H(\theta^{(i-1)}, Z)\varepsilon(\theta^{(i-1)}, Z)) \tag{16}$$

If the algorithm converges to the global minimum, then, $\hat{\theta}(Z) = \theta^{(\infty)}$. When identifying continuous-time systems in the s- and $\sqrt{s}$ -domains, it is indispensable to scale the frequency axis (and hence also the parameters) to guarantee the numerical stability of the normal equations (15). Without scaling, identification in the s- and $\sqrt{s}$ -domains is often impossible with the available computing precision, even for modest orders of the transfer function. Although the scale factor which minimizes the condition number of $J_{\text{re}}(\theta^{(i-1)}Z)$ is plant and model dependent, a good compromise is to use the arithmetic mean of the maximum and minimum angular frequencies in the frequency-band of interest: $\omega_{\text{scale}} = (\omega_{\max} + \omega_{\min})/2$. For example, the term $a_m s^m$ becomes after scaling $a_m \omega_{\text{scale}}^m (s/\omega_{\text{scale}})^m$ and $a_m \omega_{\text{scale}}^m$ is estimated. The numerical stability can still be improved by solving the overdetermined set of equations

$$J_{\text{re}}(\theta^{(i-1)}, Z)\Delta\theta^{(i)} = -\varepsilon_{\text{re}}(\theta^{(i-1)}, Z) \tag{17}$$

instead of (15), for example, using the singular value decomposition or a QR-factorization (Golub and Van Loan, 1996). The convergence region of the Newton-Gauss algorithm can be enlarged by using a Levenberg-Marquardt version of (15) and (17) (see Fletcher, 1991 and Appendix).

2.3. Quick Tools to Analyze Estimators

The minimum we can expect from a "sound" estimator is that in the noiseless case we get the true answer (correctness property). In the noisy case we should get asymptotically $(F \to \infty)$, the true answer (consistency property) and hopefully a

"small" uncertainty (efficiency property). We may also wonder whether the estimates depend on the particular parameter constraint chosen ($a_0 = 1$, or $\|\theta\|_2^2 = 1$, or …), how fast the estimates converge, and what happens with the estimates if the true model does not belong to the considered model set. Some of the previously raised questions can easily be analyzed using the following quick tools. The first step in the analysis consists of calculating the (equivalent) cost function $V(\theta, Z)$ of the identification method. Next we verify the following:-

1. *correctness:* assuming that the true model belongs to the model set, the identification algorithm is correct if it produces the true model for a finite amount of noiseless ($N_Z = 0$) data. This is true if $V_F(\theta, Z_0)$ is minimal in the true model parameters θ_0. All the identification algorithms of Section 3 are correct.

2. *consistency:* the (equivalent) cost function minimized by most identification methods in this chapter is a quadratic function of the measurements Z. The expected value of such cost functions can be written as

$$V_F(\theta) = E\{V_F(\theta, Z)\} = V_F(\theta, Z_0) + E\{V_F(\theta, N_Z)\} \qquad (18)$$

(see (12), Z_0 is deterministic and N_Z is independent of Z_0). A necessary condition for consistency is that the limit of the expected value of the cost function $V_*(\theta)$ (see Table 1) is minimal in θ_0 (Pintelon and Schoukens, 2001). It follows from (18) that this condition is satisfied if $E\{V_F(\theta, N_Z)\}$ is a θ-independent constant. Hence, for correct methods we have then $\tilde{\theta}(Z_0) = \theta_0$. For cost functions of the form (13), we replace $\eta(Z)$ by its limit value η_* before taking the expected value of the cost function. The same analysis is then performed on $V_F(\theta) = E\{f_F(\theta, \eta_*, Z)\}$.

3. *convergence to the noiseless solution:* if model errors exist, for example, because of a wrong choice of the order of the numerator and/or denominator polynomials, or because a true linear lumped model simply does not exist, then $\hat{\theta}(Z)$ converges to $\tilde{\theta}(Z_0) \neq \theta_0$. Under some conditions, the value $\tilde{\theta}(Z_0)$ is independent of the noise level of the measurements. To verify this, we replace $\mathrm{Cov}(N_Z)$ by $\upsilon^2 \mathrm{Cov}(N_Z)$ in the cost function (18), with υ a real number. If this transforms $V_F(\theta, Z_0)$ into $f(\upsilon^2) V_F(\theta, Z_0)$ and if $E\{V_F(\theta, N_Z)\}$ is a θ-independent constant then $\tilde{\theta}(Z_0)$ and its limit value θ_* are independent of the noise level υ. This is true for any υ, and hence also for $\upsilon \to 0$ which defines, asymptotically, the noiseless solution. For cost functions of the form (13), the analysis is performed on $V_F(\theta) = E\{f_F(\theta, \eta_*, Z)\}$ and the same conclusions hold if η_*, the limit value of $\eta(Z)$, is independent of υ.

4. *dependence on the parameter constraint:* from a numerical point of view it is also handy that the estimate of the plant transfer function $G(\Omega_k, \hat{\theta}(Z))$ is independent of the particular parameter constraint chosen, for example, $a_i = 1$, or $b_j = 1$, or $\|\theta\|_2^2 = 1$ … Indeed, if we fix a zero coefficient to one then the normal equations (15) become ill-conditioned. To avoid this problem, it is better to use the constraint $\|\theta\|_2^2 = 1$. The estimated plant model $G(\Omega, \hat{\theta}(Z))$ is independent of the parameter constraint chosen if, for any $\lambda \neq 0$, $V_F(\lambda\theta, Z) = V_F(\theta, Z)$, with θ the full over-parameterized form (Pintelon and Schoukens, 2001).

5. *numerical reliability of the normal equations:* the Hessian of the expected value of the cost function has full rank in the true parameter values: $\text{rank}(V_F''(\theta_0)) = \dim(\theta) =$ number of free model parameters ($'$ is the derivative w.r.t. θ). If the Hessian is not of full rank then the cost function cannot be approximated by a quadratic function in the neighborhood of the solution θ_0. This is problematic for most of the nonlinear minimization algorithms.

-
-
-

Bibliography

Brillinger D. R. (1981). Time Series: Data Analysis and Theory. McGraw-Hill, New York. [Treats the properties of noise after a discrete Fourier transform].

Fletcher R. (1991). Practical Methods of Optimization (2nd ed.). Wiley, New York. [State of the art on nonlinear minimization algorithms].

Golub G. H., and C. F. Van Loan (1996). Matrix Computations (3rd ed.). John Hopkins University Press, Baltimore. [A comprehensive overview of modern numerical linear algebra].

Ljung L. (1999). System Identification: Theory for the User (second edition). Prentice Hall, Upper Saddle River. [State of the art on time domain system identification].

Picinbono B. (1993). Random Signals and Systems. Prentice-Hall, Englewood Cliffs. [Contains a discussion of complex random variables].

Pintelon R., and J. Schoukens (2001). System Identification: a Frequency Domain Approach. IEEE Press, Piscataway. [A comprehensive overview of frequency domain system identification].

Rolain Y., R. Pintelon, K. Q. Xu, and H. Vold (1995). Best conditioned parametric identification of transfer function models in the frequency domain. IEEE Trans. on Automatic Control, vol. AC-40, no. 11, pp. 1954-1960. [Frequency domain system identification using orthogonal polynomials].

Schoukens J., R. Pintelon, and T. Dobrowiecki (2001). Linear modelling in the presence of nonlinear distortions. Proceedings of the 18th IEEE Instrumentation and Measurement Technology Conference, Budapest, Hungary, May 21-24. [Discusses the influence of nonlinear distortions on linear transfer function modeling].

Van Barel M., and A. Bultheel (1994). Discrete linearized least-squares rational approximation on the unit circle. Journal of Computational and Applied Mathematics, vol. 50, pp. 545-563. [Presents a numerically efficient algorithm for calculating an orthogonal vector polynomial basis].

Van Huffel S., and J. Vandewalle (1991). The Total Least Squares Problem: Computational Aspects and Analysis. SIAM, Frontiers in Applied Mathematics, Philadelphia. [State of the art on total least squares].

Biographical Sketches

Rik Pintelon was born in Ghent, Belgium, on December 4, 1959. He received the degree of electrical engineer (burgerlijk ingenieur) in July 1982, the degree of doctor in applied sciences in January 1988, and the qualification to teach at university level (geaggregeerde voor het hoger onderwijs) in April 1994, all from the Vrije Universiteit Brussel (VUB), Brussels, Belgium. From October 1982 till September 2000 he was a researcher of the Fund for Scientific Research – Flanders at the VUB. Since October 2000 he is professor at the VUB in the Electrical Measurement Department (ELEC). His main research interests are in the field of parameter estimation / system identification, and signal processing.

Johan Schoukens was born in Belgium in 1957. He received the degree of engineer in 1980 and the degree of doctor in applied sciences in 1985, both from the Vrije Universiteit Brussel. The prime factors of his interest are in the filed of system identification for linear and nonlinear systems and growing tomatoes in his green house.