PERGAMON

Automatica 38 (2002) 1295–1311

automatica

www.elsevier.com/locate/automatica

Frequency-domain subspace system identification using non-parametric noise models ☆

R. Pintelon *

Dept. ELEC, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

Received 6 March 2001; received in revised form 4 September 2001; accepted 12 February 2002

Abstract

In the general case of non-uniformly spaced frequency-domain data and/or arbitrarily coloured disturbing noise, the frequency-domain subspace identification algorithms described in McKelvey, Akçay, and Ljung (IEEE Trans. Automatic Control 41(7) (1996) 960) and Van Overschee and De Moor (Signal Processing 52(2) (1996) 179) are consistent only if the covariance matrix of the disturbing noise is known. This paper studies the asymptotic properties (strong convergence, convergence rate, asymptotic normality, strong consistency and loss in efficiency) of these algorithms when the true noise covariance matrix is replaced by the sample noise covariance matrix obtained from a small number of independent repeated experiments. As an additional result the strong convergence (in case of model errors), the convergence rate and the asymptotic normality of the subspace algorithms with known noise covariance matrix follows. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Subspace identification; Frequency domain; Parameter estimation; Non-parametric noise model

1. Introduction

In McKelvey, Akçay, and Ljung (1996) and Van Overschee and De Moor (1996) frequency-domain subspace algorithms have been developed for discrete-time and continuous-time models, respectively. These identification methods have proven to be very effective in solving real life problems such as, for example, modal analysis (McKelvey et al., 1996), modelling of power transformers (Akcay, Islam, & Ninness, 1999), flight flutter analysis and modelling of synchronous machines (Pintelon & Schoukens, 2001a). In general (non-uniformly spaced frequency-domain data and/or arbitrarily coloured disturbing noise), these algorithms are consistent only if the covariance matrix of the disturbing noise is known. Therefore, instrumental variable-based versions have been developed which are consistent without requiring the knowledge of the noise covariance matrix (see McKelvey (1997) for discrete-time models and Yang & Sanada (2000) for continuous-time models). In this paper we opt for another approach and propose to replace the true noise covariance matrix in the subspace algorithms by the sample noise covariance matrix obtained from a small number of independent, repeated experiments.

The main contribution of the paper is to study the stochastic properties of the frequency-domain subspace algorithms described in McKelvey et al. (1996) and Van Overschee and De Moor (1996), where the true noise covariance matrix is replaced by the sample noise covariance matrix obtained from a small number of independent repeated experiments. As a second contribution, the strong convergence in the presence of model errors, the convergence rate and the asymptotic normality of the subspace algorithms using the known covariance matrix follows (McKelvey et al. (1996) and Van Overschee & De Moor (1996) handle the strong consistency only). In order to simplify the notations, the theory is developed for single input–single output systems. Generalization of the results to multivariable systems is straightforward.

The paper is organized as follows. Section 2 develops the basic model equations used by the frequency-domain subspace algorithms. To fully understand the subspace algorithms using the sample noise covariance matrix (sample subspace), a detailed description of the algorithms using the true noise covariance matrix (subspace) is instrumental.

☆ This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor H. Hjalmarsson under the direction of Editor Torsten Soederstroem.

* Tel.: +32-2-629.29.44; fax: +32-2-629.28.50.

E-mail address: rik.pintelon@vub.ac.be (R. Pintelon).

PII: S0005-1098(02)00036-5

Also the proofs of the asymptotic properties of the sample subspace estimates rely strongly on those of the subspace estimates. Therefore, the subspace algorithms are analysed first in Section 3 and next the sample subspace algorithms in Section 4. The assumptions commonly made in subspace identification are that the input is exactly known and that the system is proper. What to do if these assumptions are not met is discussed in Section 5. Section 6 illustrates the theory on simulation and real measurement examples.

2. Model equations

2.1. Plant model

Consider a proper, n_ath order single input–single output system. The relation between the input $u(t)$ and the output $y(t)$ can be written under state-space representation form as, respectively,

$$\begin{aligned} \frac{\mathrm{d}x(t)}{\mathrm{d}t} &= Ax(t) + Bu(t), \\ y(t) &= Cx(t) + Du(t) \end{aligned} \tag{1}$$

for continuous-time systems, and

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t) + Du(t) \end{aligned} \tag{2}$$

for discrete-time systems, where $x(t) \in \mathbb{R}^{n_a}$ is the state vector (Kailath, 1980). The frequency-domain subspace algorithms estimate the parameters $A \in \mathbb{R}^{n_a \times n_a}$, $B \in \mathbb{R}^{n_a \times 1}$, $C \in \mathbb{R}^{1 \times n_a}$ and $D \in \mathbb{R}$ from a transformed version of the state-space equations (1) and (2). These are constructed as follows. Assume that the input is periodic and that an integer number of periods of the steady-state response is observed. The discrete Fourier transform (DFT) of (1) and (2) then becomes

$$\begin{aligned} \xi_k X(k) &= AX(k) + BU(k), \\ Y(k) &= CX(k) + DU(k) \end{aligned} \tag{3}$$

with $Z(k)$, $Z = U, Y, X$, the DFT of $z(t)$, $z = u, y, x$

$$Z(k) = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} z(t) \mathrm{e}^{-\mathrm{j}2\pi kt/N} \tag{4}$$

and where $\xi_k = z_k = \exp(\mathrm{j}2\pi k/N)$ for discrete-time systems, and $\xi_k = s_k = \mathrm{j}2\pi f_s k/N$, with f_s the sampling frequency, for continuous-time systems (z and s are, respectively, the Z-transform and Laplace transform variables). Recursive use of the second and the first equation of (3) gives

$$\begin{aligned} \xi_k^p Y(k) &= \xi_k^{p-1}(C\xi_k X(k) + D\xi_k U(k)) \\ &= \xi_k^{p-1}(CAX(k) + CBU(k) + D\xi_k U(k)) \\ &\vdots \\ &= CA^p X(k) + (CA^{p-1}B + CA^{p-2}B\xi_k \\ &\quad + \cdots + CB\xi_k^{p-1} + D\xi_k^p)U(k). \end{aligned} \tag{5}$$

Writing the last equation of (5) for $p = 0, 1, \ldots, r-1$ on top of each other gives

$$W_r(k)Y(k) = O_r X(k) + S_r W_r(k) U(k) \tag{6}$$

with

$$W_r(k) = \begin{bmatrix} 1 \\ \xi_k \\ \cdots \\ \xi_k^{r-1} \end{bmatrix}, \quad O_r = \begin{bmatrix} C \\ CA \\ \cdots \\ CA^{r-1} \end{bmatrix},$$

and

$$S_r = \begin{bmatrix} D & 0 & \cdots & 0 & 0 \\ CB & D & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ CA^{r-2}B & CA^{r-3}B & \cdots & CB & D \end{bmatrix}. \tag{7}$$

Collecting (6) for $k = 1, 2, \ldots, F$ gives

$$\mathbf{Y} = O_r \mathbf{X} + S_r \mathbf{U} \tag{8}$$

with

$$\begin{aligned} \mathbf{Y} &= [W_r(1)Y(1) \quad W_r(2)Y(2) \quad \ldots \quad W_r(F)Y(F)], \\ \mathbf{U} &= [W_r(1)U(1) \quad W_r(2)U(2) \quad \ldots \quad W_r(F)U(F)], \\ \mathbf{X} &= [X(1) \quad X(2) \quad \ldots \quad X(F)]. \end{aligned} \tag{9}$$

The complex data matrices $\mathbf{Y}$ and $\mathbf{U}$ have r rows and F columns. $\mathbf{X}$ is a complex n_a by F matrix, and O_r and S_r are, respectively, real r by n_a and r by r matrices. Eq. (8) is converted in a real set of equations as

$$\mathbf{Y}^{\mathrm{re}} = O_r \mathbf{X}^{\mathrm{re}} + S_r \mathbf{U}^{\mathrm{re}}, \tag{10}$$

where $(\,)^{\mathrm{re}}$ locates the real and imaginary parts beside each other, for example,

$$\mathbf{Y}^{\mathrm{re}} = [\mathrm{Re}(\mathbf{Y}) \quad \mathrm{Im}(\mathbf{Y})]. \tag{11}$$

Eq. (10) with r larger than the model order n_a is the basic model used in frequency-domain subspace identification.

The extended observability matrix O_r has the following shift property:

$$O_{r[1:r-1,:]} A = O_{r[2:r,:]}, \tag{12}$$

which will be used in the identification procedure. O_r is not unique since it depends on the choice of the state variables. Indeed, replacing (A, B, C, D, X) by $(T^{-1}AT, T^{-1}B, CT, D, T^{-1}X)$, with T an invertible matrix, in the state-space equations (3), does not change the input–output transfer function

$$G(\xi) = C(\xi I_{n_a} - A)^{-1} B + D, \tag{13}$$

but does change O_r to $O_r T$. Note that $O_r \mathbf{X}$ and S_r in model equation (8) are invariant w.r.t. the invertible transformation T.

Since A, B and C are not unique, one may wonder how the quality of the estimates $\hat{A}$, $\hat{B}$ and $\hat{C}$ can be evaluated. This is possible by referring the estimates $\hat{A}$, $\hat{B}$ and $\hat{C}$ to one particular (true or noisy) state-space realization A, C (see Jansson, 2000)

$$\begin{aligned} &\hat{A}_T = T^{-1}\hat{A}T, \quad \hat{B}_T = T^{-1}\hat{B} \\ &\text{and} \quad \hat{C}_T = \hat{C}T \ \text{ with } T = \hat{O}_r^{+} O_r, \end{aligned} \tag{14}$$

where $+$ is the Moore–Penrose pseudo-inverse ($\hat{O}_r^{+} = (\hat{O}_r^{\mathrm{T}}\hat{O}_r)^{-1}\hat{O}_r^{\mathrm{T}}$), and where $O_r, \hat{O}_r$ are defined as in (7) using, respectively, (A, C) and $(\hat{A}, \hat{C})$. Note that applying a similarity transformation P to $\hat{A}$, $\hat{B}$ and $\hat{C}$ does not change $\hat{A}_T$, $\hat{B}_T$ and $\hat{C}_T$. Hence, it is possible to calculate the sample mean and sample covariance matrices of $\hat{A}_T$, $\hat{B}_T$ and $\hat{C}_T$. To simplify the notations the subscript T will be dropped in the sequel of the paper.

For identifiability purposes, it will be assumed that the state-space realization (3) is *observable*, $\mathrm{rank}(O_r) = n_a$ for any $r \geqslant n_a$, and *controllable*, $\mathrm{rank}([B \ AB \ \dots \ A^{q-1}B]) = n_a$ for any $q \geqslant n_a$.

2.2. Noise model

The theory is developed assuming that the input is exactly known and that the output is observed with errors

$$\begin{aligned} &U(k) = U_0(k), \\ &Y(k) = Y_0(k) + N_Y(k) \end{aligned} \tag{15}$$

with $U_0(k)$, $Y_0(k)$ the true input and output DFT spectra, and $N_Y(k)$ the noise errors. $N_Y(k)$ has zero mean, $E\{N_Y(k)\} = 0$, variance $\sigma_Y^2(k) = \mathrm{var}(N_Y(k)) = E\{|N_Y(k)|^2\}$, and is independent of $U_0(k)$ and $Y_0(k)$. In Section 5 is discussed as to what to do if the input observations are also noisy. For noisy output DFT spectra $Y(k)$, model (10) becomes

$$\mathbf{Y}^{\mathrm{re}} = O_r \mathbf{X}^{\mathrm{re}} + S_r \mathbf{U}^{\mathrm{re}} + \mathbf{N}_{\mathbf{Y}}^{\mathrm{re}}, \tag{16}$$

where $\mathbf{N_Y}$ has the same structure as $\mathbf{Y}$ in (9).

3. Subspace algorithms using the true noise covariance matrix

3.1. Subspace algorithms

Subspace identification algorithms are basically a three-step procedure. First, an estimate $\hat{O}_r$ of the extended observability matrix is obtained using model (16). This is the numerically most difficult step and consists mainly of eliminating the term depending on the input and reducing the noise influence. Next, $\hat{A}$ and $\hat{C}$ are found as the least-squares solution of the overdetermined set of Eqs. (12) and as the first row of $\hat{O}_r$ (see (7)), respectively. Finally, $\hat{B}$ and $\hat{D}$ are found as the linear least-squares solution of

$$\begin{aligned} &V_{\mathrm{SUB}}(B, D, \hat{A}, \hat{C}, Z) \\ &\quad = \sum_{k=1}^{F} W^2(k) |Y(k) - [\hat{C}(\xi_k I_{n_a} - \hat{A})^{-1}B + D]U(k)|^2 \end{aligned} \tag{17}$$

with Z being a vector containing the input–output data, and where $W(k)$ is a well chosen real weighting function.

We present three algorithms, one for discrete-time systems ($\xi = z$), based on McKelvey et al. (1996), and two for continuous-time systems ($\xi = s$), based on Yang and Sanada (2000) and Van Overschee and De Moor (1996), respectively. The numerically efficient implementation of these algorithms is due to Verhaegen (1994).

Algorithm 1 (Subspace algorithm for discrete-time systems).

(1) Estimate O_r given the data $Y(k)$, $U(k)$ and the noise variance $\sigma_Y^2(k)$:

(1a) *Initialization*: Choose a value of $r > n_a$ and form the following matrices

$$Z = \begin{bmatrix} \mathbf{U}^{\mathrm{re}} \\ \mathbf{Y}^{\mathrm{re}} \end{bmatrix} \quad \text{and} \quad C_{\mathbf{Y}} = \mathrm{Re}(\mathbf{C}\mathbf{C}^{H}) \tag{18}$$

with $\mathbf{C} = [W_r(1)\sigma_Y(1) \ W_r(2)\sigma_Y(2) \ \dots \ W_r(F)\sigma_Y(F)]$ and where $\mathbf{U}$ and $\mathbf{Y}$ are defined in (9), and $W_r(k)$ as in (7) with $\xi = z$.

(1b) *Elimination of the input term in* (16): Calculate the QR-factorization of Z^{T}. $Z^{\mathrm{T}} = QR$ or $Z = R^{\mathrm{T}}Q^{\mathrm{T}}$.

$$Z = \begin{bmatrix} R_{11}^{\mathrm{T}} & 0 \\ R_{12}^{\mathrm{T}} & R_{22}^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} Q_1^{\mathrm{T}} \\ Q_2^{\mathrm{T}} \end{bmatrix},$$

where R_{ij} are $r \times r$ blocks of upper triangular matrix R.

(1c) *Reduction of the noise influence in* (16): Calculate the singular value decomposition of $C_{\mathbf{Y}}^{-1/2} R_{22}^{\mathrm{T}}$,

$$C_{\mathbf{Y}}^{-1/2} R_{22}^{\mathrm{T}} = U \Sigma V^{\mathrm{T}},$$

where $C_{\mathbf{Y}}^{1/2}$ is a square root of $C_{\mathbf{Y}}$, and estimate O_r as

$$\hat{O}_r = C_{\mathbf{Y}}^{1/2} U_{[:,1:n_a]}.$$

(2) *Estimate A and C, given the estimate $\hat{O}_r$*: solve the shift property (12) in least-squares sense and select the first row of $\hat{O}_r$

$$\hat{A} = \hat{O}_{r[1:r-1,:]}^{+} \hat{O}_{r[2:r,:]} \quad \text{and} \quad \hat{C} = \hat{O}_{r[1,:]}. \tag{19}$$

(3) *Estimate B and D, given the estimates $\hat{A}$ and $\hat{C}$*: minimize (17) w.r.t. B and D with $W(k) = 1/\sigma_Y(k)$.

Proof. See McKelvey et al. (1996). □

Step 1b is due to Verhaegen (1994) and calculates in a numerical stable way the elimination of the term $S_r\mathbf{U}^{\mathrm{re}}$ in

(16): $R_{22}^{\mathrm{T}} = \mathbf{Y}^{\mathrm{re}}\Pi$, where

$$\mathbf{Y}^{\mathrm{re}}\Pi = O_r\mathbf{X}^{\mathrm{re}}\Pi + \mathbf{N}_{\mathbf{Y}}^{\mathrm{re}}\Pi \quad \text{with}$$
$$\Pi = I_{2F} - \mathbf{U}^{\mathrm{re\,T}}(\mathbf{U}^{\mathrm{re}}\mathbf{U}^{\mathrm{re\,T}})^{-1}\mathbf{U}^{\mathrm{re}}. \tag{20}$$

Proof. See Verhaegen, 1994 and Appendix B. □

One could use Algorithm 1 with $\xi = s$ for continuous-time systems. This works reasonably well for small values of r. However, for larger values, the data matrix Z (18) in Algorithm 1 becomes ill-conditioned, resulting in poor estimates. The condition number of Z can be improved by using a bilinear frequency transformation $w = (s/\omega_{\mathrm{scale}} - 1)/(s/\omega_{\mathrm{scale}} + 1)$ with $\omega_{\mathrm{scale}} > 0$ a frequency scaling (Yang & Sanada, 2000). Algorithm 1 with $\xi = w$ gives then estimates of A, B, C and D in the w domain. The original A, B, C and D parameters are recovered via a bilinear state-space transformation (Yang & Sanada, 2000). In Van Overschee and De Moor (1996), the ill-conditioning problem is solved by introducing two scalar orthogonal polynomial bases which orthogonalize, respectively, the first r rows of Z and the last r rows of Z. It can be shown that there are no other two scalar polynomial bases which result in a smaller condition number of Z (Rolain, Pintelon, Xu, & Vold, 1995). The final algorithm is also a three-step procedure. First, a generalized extended observability matrix $\hat{O}_{r\perp}$ is estimated. This matrix has a generalized shift structure which is used to estimate A. Next, $\hat{A}$ and $\hat{C}$ are estimated using $\hat{O}_{r\perp}$. Finally, $\hat{B}$ and $\hat{D}$ are the linear least-squares solution of (17).

Algorithm 2 (Subspace algorithms for continuous-time systems)

1. Algorithm 1 with $\xi = (s/\omega_{\mathrm{scale}} - 1)/(s/\omega_{\mathrm{scale}} + 1)$, followed by a bilinear state-space transformation of the estimated A, B, C and D parameters: see Yang and Sanada (2000) for more details.
2. Algorithm 1 with $\xi = s$, and where the first r rows (input data) and the last r rows (output data) of Z (18) are orthogonalized using two scalar orthogonal bases in s: see Van Overschee and De Moor (1996) for the original algorithm, and Pintelon and Schoukens (2001) for an improved numerical implementation.

3.2. Assumptions

The following assumptions are made to prove the asymptotic properties for $F \to \infty$ of the subspace algorithms. They are grouped according to the property to be shown.

(i) For the *strong convergence*, it is assumed that:

Assumption 1 (Disturbing noise). The frequency-domain error $N_Y(k)$ is an independent (over k) zero mean random variable with uniformly bounded absolute moments of order 4. $N_Y(k)$ is independent of the true excitation $U(k)$.

The independence assumption of $N_Y(k)$ is fulfilled if the frequency-domain data is obtained from F single sine experiments. It is asymptotically fulfilled for the measurement time approaching infinity, if the data are obtained from a time-domain experiment with coloured disturbing noise, for example, filtered white noise (Brillinger, 1981).

Assumption 2 (Covariance matrix disturbing noise). The variance $\sigma_Y^2(k) = \mathrm{var}(N_Y(k))$ of $N_Y(k)$ is known.

Assumption 3 (Persistence of excitation). There exists an F_0 such that for any $F \geqslant F_0$, ∞ included, $\mathbf{U}^{\mathrm{re}}\mathbf{U}^{\mathrm{re\,T}}/F \geqslant cI_r$ with $0 < c < \infty$ and c independent of F.

Assumption 4 (Distinct frequencies). The number of distinct frequencies in the band $[f_{\min}, f_{\max}]$ increases as an $O(F)$.

Assumption 5 (identifiability conditions). There exists an F_0 such that for any $F \geqslant F_0$, ∞ included, (i) $\mathrm{rank}(\mathbf{Y}_0^{\mathrm{re}}\Pi) \geqslant n_a$, and (ii) $E\{V_{\mathrm{SUB}}(B,D,A,C,Z)\}$ has a unique global minimum w.r.t. (B,D) for any (A,C) in a closed neighbourhood of the solution $(\tilde{A},\tilde{C})$ of the noiseless problem $(Y(k), \sigma_Y^2(k) \to Y_0(k), \sigma_Y^2(k)$ in steps 1 and 3 of Algorithms 1 and 2).

(ii) For the *stochastic convergence rate*, it is assumed in addition to Assumptions 1–5 that:

Assumption 6 (Simple eigenvalues). There exists an F_0 such that for any $F \geqslant F_0$, ∞ included, the non-zero singular values of the r by $2F$ matrix $\mathbf{Y}_0^{\mathrm{re}}\Pi$, with Π the idempotent matrix defined in (20), are simple eigenvalues of $\mathbf{Y}_0^{\mathrm{re}}\Pi\mathbf{Y}_0^{\mathrm{re\,T}}$.

(iii) For the *asymptotic normality*, it is assumed in addition to Assumptions 1–6 that:

Assumption 7 (Disturbing noise). The disturbing noise $N_Y(k)$ has uniformly bounded absolute moments of order $4 + \varepsilon$ with $\varepsilon > 0$.

(iv) For the *deterministic convergence rate*, it is assumed in addition to Assumptions 1–6 that: To study the convergence of the solution $\tilde{\theta}$ of the noiseless problem $(Y(k), \sigma_Y^2(k) \to Y_0(k), \sigma_Y^2(k)$ in steps 1 and 3 of Algorithms 1 and 2), the strategy of adding new frequencies to the data must be defined. This information is needed because the model errors depend on the power spectrum of the excitation.

Assumption 8 (Strategy of adding frequencies). As $F \to \infty$, the frequencies f_k cover the frequency interval $[f_{\min}, f_{\max}]$ with a density function $n(f)$ defined as

$$n(f) = \lim_{\Delta f \to 0} \lim_{F \to \infty} \frac{N_F(f + \Delta f) - N_F(f)}{F\,\Delta f}, \tag{21}$$

where $N_F(f)$ is the number of frequencies in the interval $[0, f]$ when the total number of frequencies is F. The

density $n(f)$ is continuous with bounded second-order derivative w.r.t. f in $[f_{\min}, f_{\max}]$ except at a finite number of frequencies.

Special cases are a uniform ($n(f)$ independent of f) or a logarithmic ($n(f)$ is proportional to f^{-1}) distribution of the number frequencies in $[f_{\min}, f_{\max}]$.

Assumption 9 (Signal power spectra). The true signal (cross-)power spectra $|U(f)|^2$, $|Y_0(f)|^2$ and $Y_0(f)\bar{U}(f)$ have bounded second-order derivatives w.r.t. the frequency f in the frequency band $[f_{\min}, f_{\max}]$, except at a finite number of frequencies.

(v) For the *strong consistency*, it is assumed in addition to Assumptions 1–4 that:

Assumption 10 (True plant model). The true plant model can be written under the form (3) where (A, C) is observable and (A, B) is controllable.

3.3. Asymptotic properties

Define the $n_a^2 + 2n_a + 1$ by 1 vector θ as

$$\theta = [\mathrm{vec}^{\mathrm{T}}(A) \quad B^{\mathrm{T}} \quad C \quad D]^{\mathrm{T}}, \tag{22}$$

where vec(A) stacks the columns of the matrix A on top of each other. Define furthermore $\hat{\theta}$ as the noisy subspace estimate (outcome Algorithms 1 and 2), $\tilde{\theta}$ as the solution of the noiseless problem ($Y(k), \sigma_Y^2(k) \to Y_0(k), \sigma_Y^2(k)$ in Algorithms 1 and 2), and θ_* the solution of the noiseless problem with an infinite amount of data ($Y(k), \sigma_Y^2(k) \to Y_0(k), \sigma_Y^2(k)$ and $F = \infty$ in Algorithms 1 and 2). The following parameter vectors are also introduced:

$$\alpha = [\mathrm{vec}^{\mathrm{T}}(A) \quad C]^{\mathrm{T}} \quad \text{and} \quad \beta = [B^{\mathrm{T}} \quad D]^{\mathrm{T}}, \tag{23}$$

where the estimated β depends on the estimated α: $\beta = \beta(\alpha)$ (see step 3 of Algorithms 1 and 2). The properties of the estimate $\hat{\theta}$ are established in the following theorem (O_{p} stands for O in probability, see also the introduction of Appendix A):

Theorem 1 (Asymptotic properties $\hat{\theta}$). *Under the assumptions of Section* 3.2 *the estimate* $\hat{\theta}$ *obtained using Algorithm* 1 *or* 2 *has the following asymptotic* ($F \to \infty$) *properties*:

(1) Stochastic convergence: $\hat{\theta}$ *converges strongly to* $\tilde{\theta}$ (*Assumptions* 1–5).
(2) Stochastic convergence rate: $\hat{\theta}$ *converges in probability at the rate* $O_{\mathrm{p}}(F^{-1/2})$ *to* $\tilde{\theta}$:
$$\sqrt{F}(\hat{\alpha} - \tilde{\alpha}) = \sqrt{F}\delta_\alpha + o_{\mathrm{p}}(F^0)$$
and
$$\sqrt{F}(\hat{\beta} - \tilde{\beta}) = \sqrt{F}\delta_\beta + O_{\mathrm{p}}(F^{-1/2}),$$
where $\sqrt{F}\delta_\alpha = O_{\mathrm{p}}(F^0)$ *and* $\sqrt{F}\delta_\beta = O_{\mathrm{p}}(F^0)$ (*Assumptions* 1–6).
(3) Asymptotic normality: $\sqrt{F}(\hat{\theta} - \tilde{\theta})$ *converges in law to a zero mean Gaussian random variable* (*Assumptions* 1–7).
(4) Deterministic convergence: $\tilde{\theta}$ *converges at the rate* $O(F^{-2})$ *to* θ_*, (*Assumptions* 1–6 *and* 8, 9).
(5) Consistency: $\hat{\theta}$ *is strongly consistent*; *replace* $\tilde{\theta}$ *by* θ_0 *in properties* 1–3, *and Assumption* 5 *by Assumption* 10.

Proof. See Appendix B. □

4. Subspace algorithms using the sample noise covariance matrix

4.1. Sample subspace algorithms

The noise variance $\sigma_Y^2(k)$ was assumed to be known exactly, and under these conditions, the properties of Algorithms 1 and 2 were studied. In practice, this information is not available, but should be extracted from the experimental data. In this section, the exact noise variance is replaced by its sample value. This is only possible if independent, repeated experiments are available. The M experiments are processed and their DFT spectra

$$U^{[l]}(k), Y^{[l]}(k), \quad l = 1, \ldots, M \text{ and } k = 1, \ldots, F \tag{24}$$

are calculated. The sample mean $\hat{Y}(k)$ and sample variance $\hat{\sigma}_Y^2(k)$ are obtained directly from these measurements,

$$\hat{\sigma}_Y^2(k) = \frac{1}{M-1}\sum_{l=1}^{M} |Y^{[l]}(k) - \hat{Y}(k)|^2$$

with

$$\hat{Y}(k) = \frac{1}{M}\sum_{l=1}^{M} Y^{[l]}(k). \tag{25}$$

Replacing $Y(k)$ and $\sigma_Y^2(k)$ by $\hat{Y}(k)$ and $\hat{\sigma}_Y^2(k)/M$ in Algorithms 1 and 2 defines the sample subspace algorithms.

4.2. Assumptions

In addition to Section 3.2, the following assumptions are necessary to study the asymptotic behaviour ($F \to \infty$) of the sample subspace algorithms. First, it is required that M independent repeated experiments are available. Next, an assumption about the disturbing errors of the lth experiment is made.

Assumption 11 (M independent repeated experiments). The measured input/output DFT spectra $U^{[l]}(k)$, $Y^{[l]}(k)$, $k = 1, 2, \ldots, F$ and $l = 1, 2, \ldots, M$, satisfy

$$Y^{[l]}(k) = Y_0(k) + N_Y^{[l]}(k) \quad \text{and} \quad U^{[l]}(k) = U_0(k), \tag{26}$$

where the true deterministic values $U_0(k)$, $Y_0(k)$ are independent of l, and where the disturbing errors $N_Y^{[l]}(k)$ are independent over l.

A practical solution for obtaining M independent repeated experiments consists of applying periodic excitations to the plant and observing M consecutive periods of the steady-state response. The DFT spectra (24) of each period are considered as independent experiments from one period to the other as formalized in Assumption 11. This is only approximately true since some correlation exists between neighbouring periods. Since the correlation of filtered white noise decays exponentially, the correlation between two neighbouring periods disappears inversely proportional to the length of the period. In practice, it can be neglected if the period length is large compared with the correlation length of the noise.

Assumption 12 (Zero mean normally distributed errors). The noise $N_Y^{[l]}(k)$ is an independent (over the frequency k), circular complex normally distributed random variable with zero mean and variance $\sigma_Y^2(k)$.

The normality assumption is necessary to ensure that the sample mean $\hat{Y}(k)$ and the sample variance $\hat{\sigma}_Y^2(k)$ are independent random variables (Anderson, 1958). It is asymptotically (measurement time approaching infinity) fulfilled for coloured disturbing noise (Brillinger, 1981).

4.3. Asymptotic properties

Define $\hat{\theta}_S$ as the noisy sample subspace estimate (outcome Algorithms 1 and 2, where $Y(k), \sigma_Y^2(k)$ are replaced by $\hat{Y}(k), \hat{\sigma}_Y^2(k)/M$); $\tilde{\alpha}_S$ as the result of steps 1 and 2 of Algorithms 1 and 2, where $Y(k), \sigma_Y^2(k)$ are replaced by $Y_0(k), \sigma_Y^2(k)/M$; $\tilde{\beta}_S$ the minimizer of the expected value of (17), where $Y(k), \sigma_Y^2(k)$, $\hat{\alpha}$ are replaced by $\hat{Y}(k), \hat{\sigma}_Y^2(k)$, $\tilde{\alpha}_S$

$$\tilde{\beta}_S = \arg\min_{\beta} E\left\{\sum_{k=1}^{F} \hat{\sigma}_Y^{-2}(k)|\hat{Y}(k) - [\tilde{C}_S(\xi_k I_{n_a} - \tilde{A}_S)^{-1}B + D]U(k)|^2\right\}; \quad (27)$$

and α_{*S}, β_{*S} the limit values for $F \to \infty$ of $\tilde{\alpha}_S$, $\tilde{\beta}_S$, respectively. The properties of $\hat{\theta}_S$ are established in the following theorem:

Theorem 2 (Asymptotic properties $\hat{\theta}_S$). *Under the assumptions of Sections* 3.2 *and* 4.2, *the estimate* $\hat{\theta}_S$ *obtained using Algorithm* 1 *or* 2, *where* $Y(k)$, $\sigma_Y^2(k)$ *are replaced by* $\hat{Y}(k), \hat{\sigma}_Y^2(k)/M$, *has the following asymptotic* ($F \to \infty$) *properties*:

(1) Stochastic convergence: $\hat{\theta}_S$ *converges strongly to* $\tilde{\theta}_S$, *with* $M \geqslant 2$ *for* $\hat{\alpha}_S$ *and* $M \geqslant 4$ *for* $\hat{\beta}_S$ (*Assumptions* 1–5, 11(M) *and* 12).

(2) Stochastic convergence rate: $\hat{\theta}_S$ *converges in probability at the rate* $O_p(F^{-1/2})$ *to* $\tilde{\theta}_S$:

$$\sqrt{F}(\hat{\alpha}_S - \tilde{\alpha}_S) = \sqrt{F}\delta_{\alpha_S} + o_p(F^0)$$

and

$$\sqrt{F}(\hat{\beta}_S - \tilde{\beta}_S) = \sqrt{F}\delta_{\beta_S} + O_p(F^{-1/2}),$$

where $\sqrt{F}\delta_{\alpha_S} = O_p(F^0)$ *and* $\sqrt{F}\delta_{\beta_S} = O_p(F^0)$, *and with* $M \geqslant 2$ *for* $\hat{\alpha}_S$ *and* $M \geqslant 4$ *for* $\hat{\beta}_S$ (*Assumptions* 1–6, 11(M) *and* 12).

(3) Asymptotic normality: $\sqrt{F}(\hat{\theta}_S - \tilde{\theta}_S)$ *converges in law to a zero mean Gaussian random variable*, *with* $M \geqslant 2$ *for* $\hat{\alpha}_S$ *and* $M \geqslant 4$ *for* $\hat{\beta}_S$ (*Assumptions* 1–7, 11(M) *and* 12).

(4) Deterministic convergence: $\tilde{\theta}_S$ *converges at the rate* $O(F^{-2})$ *to* θ_{*S}, *with* $M \geqslant 2$ *for* $\tilde{\alpha}_S$ *and* $M \geqslant 4$ *for* $\tilde{\beta}_S$ (*Assumptions* 1–6, 8, 9, 11(M) *and* 12).

(5) Consistency: $\hat{\theta}_S$ *is strongly consistent*; *replace* $\tilde{\theta}_S$ *by* θ_0 *in properties* 1–3, *and Assumption* 5 *by Assumption* 10.

Proof. See Appendix C. □

Although the sample variance (25) is a very poor estimate of the variance for small values of M (e.g., for $M = 10$ the 95% confidence bound is $0.3 \leqslant \hat{\sigma}_Y^2(k)/\sigma_Y^2(k) \leqslant 2.1$), it follows that, respectively, $M \geqslant 2$ and $M \geqslant 4$ experiments are sufficient for, respectively, $\hat{\alpha}_S$ and $\hat{\beta}_S$, to have similar asymptotic properties as $\hat{\alpha}$ and $\hat{\beta}$. The basic reason for this is the averaging over the frequencies of $\hat{\sigma}_Y^2(k)$ and $1/\hat{\sigma}_Y^2(k)$ in steps 1 and 3 of Algorithms 1 and 2, respectively. Note the similarity between on the one hand the sample total least-squares estimates (Pintelon, Guillaume, Vandersteen, & Rolain, 1998) and $\hat{\alpha}_S$, and on the other hand the sample maximum likelihood estimates (Schoukens, Pintelon, Vandersteen, & Guillaume, 1997) and $\hat{\beta}_S$, which also requires that, respectively, $M \geqslant 2$ and $M \geqslant 4$.

4.4. Comparison between the subspace and sample subspace estimates

The relationship between the subspace $\hat{\theta}$ and the sample subspace $\hat{\theta}_S$ estimates is established in the following theorem:

Theorem 3 (Relationship between $\hat{\theta}_S$ and $\hat{\theta}$). *Under the conditions of Theorems* 1 *and* 2, *the estimates based on the true and the sample variances are related to each other by*

(1) *in case of model errors* $\hat{\theta}$ *and* $\hat{\theta}_S$ *converge to the same noiseless solutions*: $\tilde{\theta} = \tilde{\theta}_S$ *and* $\theta_* = \theta_{*S}$ (*Assumptions* 1–6, 8, 9, 11(M) *and* 12),

(2) *in the absence of model errors* ($\tilde{\theta} = \tilde{\theta}_S = \tilde{\theta}_0$),

for $M \geqslant 2$: $\delta_{\alpha_S} = \delta_\alpha \Rightarrow \mathrm{Cov}(\sqrt{F}\delta_{\alpha_S}) = \mathrm{Cov}(\sqrt{F}\delta_\alpha)$,

$$\text{for } M \geqslant 4: \begin{cases} \mathrm{Cov}(\sqrt{F}\delta_{\alpha_S}, \sqrt{F}\delta_{\beta_S}(\hat{\alpha}_S)) \\ \quad = \mathrm{Cov}(\sqrt{F}\delta_\alpha, \sqrt{F}\delta_\beta(\hat{\alpha})), \\ \mathrm{Cov}(\sqrt{F}\delta_{\beta_S}(\hat{\alpha}_S)) = \mathrm{Cov}(\sqrt{F}\delta_\beta(\hat{\alpha})) \\ \quad + \dfrac{1}{M-3}\mathrm{Cov}(\sqrt{F}\delta_\beta(\alpha_0)) \\ \quad + O(F^{-1}), \end{cases}$$

where $\mathrm{Cov}(\sqrt{F}\delta_\beta(\hat{\alpha}))$ *and* $\mathrm{Cov}(\sqrt{F}\delta_\beta(\alpha_0))$ *are the covariance matrices when, respectively,* α *is estimated and* α *is known* (*Assumptions* 1–4, 6 *and* 10–12).

Proof. See Appendix D. □

It follows that there is no increase in uncertainty for the estimated poles (depend on A only), while there is a small increase in uncertainty for the estimated zeroes (depend on A, B, C and D). Using the results of Theorem 3, it is possible to quantify the increase in uncertainty on the estimated transfer function.

Corollary (Uncertainty estimated transfer function). *Under the conditions of Theorems* 1 *and* 2, *transfer function* (13), *evaluated in the estimated model parameters, can be written as*

$$\sqrt{F}(G(\xi,\hat{\psi}) - G(\xi,\tilde{\psi})) = \sqrt{F}\delta_G(\xi,\hat{\psi}) + O_p(F^{-1/2}),$$

$$\delta_G(\xi,\hat{\psi}) = \frac{\partial G(\xi,\psi)}{\partial\tilde{\psi}}\,\delta_\psi$$

with $\psi = \theta$ *or* θ_S, *and* $\sqrt{F}\delta_G(\xi,\hat{\psi}) = O_p(F^0)$ (*Assumptions* 1–6, 8, 9, 11($M = 4$) *and* 12). *In the absence of model errors* ($\tilde{\theta} = \theta_0$) *the increase in uncertainty of* $G(\xi,\hat{\theta}_S)$ *w.r.t.* $G(\xi,\hat{\theta})$ *is given by*

$$\mathrm{var}(\sqrt{F}\delta_G(\xi,\hat{\theta}_S)) = \mathrm{var}(\sqrt{F}\delta_G(\xi,\hat{\theta})) + \frac{1}{M-3}\mathrm{var}(\sqrt{F}\delta_G(\xi,\alpha_0,\hat{\beta})) + O(F^{-1}), \tag{28}$$

where $\mathrm{var}(\sqrt{F}\delta_G(\xi,\hat{\theta}))$ *and* $\mathrm{var}(\sqrt{F}\delta_G(\xi,\alpha_0,\hat{\beta}))$ *are the variances when, respectively,* α *is estimated and* α *is known* (*Assumptions* 1–4, 6 *and* 10–12).

Proof. See Appendix E. □

From (28) it can be seen that $\mathrm{var}(\sqrt{F}\delta_G(\xi,\hat{\theta}_S))$ is asymptotically ($F \to \infty$) bounded by

$$\mathrm{var}(\sqrt{F}\delta_G(\xi,\hat{\theta})) \leqslant \mathrm{var}(\sqrt{F}\delta_G(\xi,\hat{\theta}_S)) \leqslant \frac{M-2}{M-3}\max(\mathrm{var}(\sqrt{F}\delta_G(\xi,\hat{\theta})), \mathrm{var}(\sqrt{F}\delta_G(\xi,\alpha_0,\hat{\beta}))). \tag{29}$$

If $\mathrm{var}(\sqrt{F}\delta_G(\xi,\hat{\theta})) \geqslant \mathrm{var}(\sqrt{F}\delta_G(\xi,\alpha_0,\hat{\beta}))$, which is the generic case, then (29) simplifies to

$$1 \leqslant \frac{\mathrm{std}(\sqrt{F}\delta_G(\xi,\hat{\theta}_S))}{\mathrm{std}(\sqrt{F}\delta_G(\xi,\hat{\theta}))} \leqslant \sqrt{\frac{M-2}{M-3}}. \tag{30}$$

Hence, replacing the true variance by the sample variance increases the uncertainty of the estimated transfer function by a factor $\sqrt{(M-2)/(M-3)}$, which is close to one. For example, for $M = 4, 5, 6$ and 10, the standard deviation increases by 41%, 22%, 15% and 7%, respectively.

Note that in case of model errors (a true linear model does not exist or does not belong to the considered model set), the increase in uncertainty can be much larger than predicted by (29).

5. Practical remarks

- If the input is observed with errors, $U(k) = U_0(k) + N_U(k)$, then the (sample) subspace Algorithms 1 and 2 generate biased estimates. The bias can almost completely be removed by replacing the data $\hat{U}(k)$, $\hat{Y}(k)$ and $\hat{\sigma}^2_Y(k)$, by 1, $\hat{G}(\xi_k) = \hat{Y}(k)/\hat{U}(k)$ and $\hat{\sigma}^2_G(k)$, where

$$\hat{\sigma}^2_G(k) = |\hat{G}(\xi_k)|^2\left(\frac{\hat{\sigma}^2_Y(k)}{|\hat{Y}(k)|^2} + \frac{\hat{\sigma}^2_U(k)}{|\hat{U}(k)|^2} - 2\mathrm{Re}\left(\frac{\hat{\sigma}^2_{YU}(k)}{\hat{Y}(k)\bar{\hat{U}}(k)}\right)\right) \tag{31}$$

with $\hat{\sigma}^2_U(k)$ the sample variance of the input DFT spectra, and $\hat{\sigma}^2_{YU}(k)$ the sample covariance of the input and output DFT spectra

$$\hat{\sigma}^2_{YU}(k) = \frac{1}{M-1}\sum_{l=1}^{M}(Y^{[l]}(k) - \hat{Y}(k)) \times \overline{(U^{[l]}(k) - \hat{U}(k))}. \tag{32}$$

Indeed, for circular complex normally distributed errors $\hat{N}_U(k)$ and $\hat{N}_Y(k)$, the relative bias

$$b(k) = |(E\{\hat{G}(\xi_k)\} - G_0(\xi_k))/G_0(\xi_k)| \tag{33}$$

with G_0 the true plant transfer function, is smaller than 1×10^{-4} for input and output signal-to-noise ratios $|U_0(k)|/\sigma_U(k)$, $|Y_0(k)|/\sigma_Y(k)$ larger than 10 dB (Pintelon & Schoukens, 2001b). To avoid the technical difficulty that the moments of order 2 and higher of $\hat{G}(\xi_k) = \hat{Y}(k)/\hat{U}(k)$ do not exist (Guillaume, Kollár, & Pintelon, 1996), large, highly improbable values of $\hat{G}(\xi_k)$ are excluded. Define thereto the truncated ratio $\underline{\hat{G}}(\xi_k)$ as

$$\underline{\hat{G}}(\xi_k) = \begin{cases} \hat{Y}(k)/\hat{U}(k), & |\hat{U}(k)/U_0(k)| \geqslant L, \\ 0, & |\hat{U}(k)/U_0(k)| < L \end{cases} \tag{34}$$

with L an arbitrarily small number. Note that this is exactly what is done in practice: if the ratio $\hat{Y}(k)/\hat{U}(k)$ is unacceptably large, then it is rejected. For input signal-to-noise ratios larger than 10 dB and $L = 1 \times 10^{-3}$, the change in bias of $\hat{\underline{G}}(\xi_k)$ w.r.t. $G(\xi_k)$ is negligible and the variance of the truncated estimate is in good approximation given by the (31) (Guillaume et al., 1996). From a practical point of view, it can be concluded that the noise on $\hat{\underline{G}}(\xi_k)$ has zero mean with existing higher order moments and, hence, satisfies the assumptions of Theorems 1 and 2.

- If the required transfer function models is improper, $n_a < n_b$, then the role of the input and output are interchanged: $\hat{U}(k), \hat{Y}(k)$ and $\hat{\sigma}_Y^2(k)$ are replaced by 1, $\hat{G}(\xi_k) = \hat{U}(k)/\hat{Y}(k)$ and $\hat{\sigma}_G^2(k)$ (31).
- Although the subspace estimates $\hat{\theta}$ and $\hat{\theta}_S$ are strongly consistent ($F \to \infty$) for any $r > n_a$, with r independent of F, the finite sample properties of $\hat{\theta}$ and $\hat{\theta}_S$ strongly depend on the choice of r. For example, values of r close to n_a+1 usually result in poor estimates. An appropriate choice of r is, therefore, recommended. A possible solution consists in choosing r such that

$$V_{\mathrm{SUB}}(\hat{C}_S, \hat{D}_S, \hat{A}_S, \hat{C}_S, Z) = \sum_{k=1}^{F} \frac{|\hat{Y}(k) - G(\xi_k, \hat{\theta}_S)U(k)|^2}{\hat{\sigma}_Y^2(k)} \tag{35}$$

is minimal. This optimization requires an exhaustive search for all $r > n_a$ values (the cost function (35) is a craggy function of r, with a lot of peaks and dips). In practice, we limit the search to the interval $[1.5n_a, 6n_a]$. However, sometimes it may be necessary to go beyond the upper limit $6n_a$ to find the optimum (see Section 6, modelling of a synchronous motor). It also turns out that the optimal value of r strongly depends on the plant and the noise characteristics.

6. Simulation and real measurement examples

6.1. Simulation examples

Two simulation examples are shown. The first illustrates Theorem 1 in the absence of model errors and the Corollary to Theorem 3, while the second illustrates Theorem 3 in the presence of model errors. Each Monte-Carlo simulation is performed with N_{r} runs, M independent repeated experiments, and F frequencies. For each run, three estimates are calculated: (i) the sample subspace estimates $\hat{\theta}_S$, (ii) the subspace estimates $\hat{\theta}$, and (iii) the subspace estimates $\hat{\beta}(\alpha_0)$ when α is known exactly. The sample mean $\langle G(\xi, \hat{\psi}) \rangle$ and the sample variance of each set of N_{r} transfer function estimates $G(\xi, \hat{\psi})$ is calculated, where $\hat{\psi} = \hat{\theta}_S$, $\hat{\theta}$ and $[\alpha_0^{\mathrm{T}} \hat{\beta}^{\mathrm{T}}(\alpha_0)]^{\mathrm{T}}$ for sets (i), (ii) and (iii), respectively. The sample variances of the last two sets make it possible to estimate the right-hand side of (28).

In the first simulation, we take the fourth-order discrete-time system and the second-order discrete-time noise model of McKelvey et al. (1996), namely, (3) with $\xi = z$ and

$$A = \begin{bmatrix} 0.8876 & 0.4494 & 0 & 0 \\ -0.4494 & 0.7978 & 0 & 0 \\ 0 & 0 & -0.6129 & 0.064516 \\ 0 & 0 & -6.4516 & -0.7419 \end{bmatrix},$$

$$B = \begin{bmatrix} 0.2247 \\ 0.8989 \\ 0.0323 \\ 0.1290 \end{bmatrix}, \quad C = \begin{bmatrix} 0.4719 \\ 0.1124 \\ 9.6774 \\ 1.6129 \end{bmatrix}^{\mathrm{T}},$$

$$D = 0.9626 \tag{36}$$

for the plant model, and

$$A_n = \begin{bmatrix} 0.6296 & 0.0741 \\ -7.4074 & 0.4815 \end{bmatrix}, \quad B_n = \begin{bmatrix} 0.037 \\ 0.7407 \end{bmatrix},$$

$$C_n = \begin{bmatrix} 1.6300 \\ 0.0740 \end{bmatrix}^{\mathrm{T}}, \quad D_n = 0.2 \tag{37}$$

for the noise model. The input–output spectra are generated at $F = 999$ frequencies $f_k/f_s = k/2000$, $k = 1, 2, \ldots, F$, with f_s the sampling frequency. Circular complex normally distributed noise $N_Y(k)$ with zero mean and variance $\sigma_Y^2(k) = |H_0(z_k)|^2$ is added to the output spectrum $Y_0(k)$, where $H_0(z)$ is the monic transfer function representation of (37). The results of the Monte-Carlo simulation with $N_{\mathrm{r}} = 1 \times 10^4$, $M = 7$, and $n_a = 4$, $r = 8$ in Algorithm 1, are shown in Fig. 1. It follows that the sample subspace estimates are consistent (see Fig. 1a), and that the increase in uncertainty is predicted very well by (28) (see Fig. 1b, taking into account that the 95% confidence region of the sample standard deviation equals $0.99 \leqslant \hat{\sigma}/\sigma \leqslant 1.01$ for $N_{\mathrm{r}} = 1 \times 10^4$).

The second simulation is exactly the same as the first except that a first-order model is estimated ($n_a = 1$ in step 1c of Algorithm 1). The results are shown in Fig. 2. It follows that the sample subspace and subspace estimates converge to the same solution θ_* (see Fig. 2a), and that the increase in uncertainty is much larger than that predicted by (28) (see Fig. 2b, (28) is valid in the absence of model errors only).

6.2. Real measurement example

Fig. 3 shows the measured q-axis impedance of a 3.4 MW synchronous motor. The measurements were carried out using a multisine excitation of 1000 A consisting of $F = 100$ frequencies logarithmically spaced in the band [12 mHz, 12 Hz]. The sample mean and sample variances were obtained by analysing $M = 30$ periods of the input and output signals. Note the particularly large dynamic range

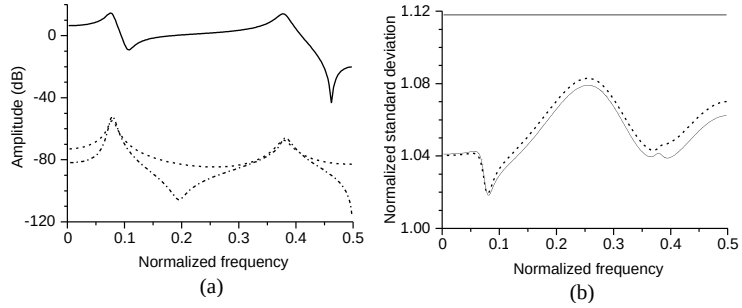

Fig. 1. Simulation results of the fourth-order discrete-time plant—no model errors and $M = 7$: (a) the true plant $G_0(z)$ (solid line) and the difference $G_0(z) - \langle G(Z, \hat{\theta}_S)\rangle$ (dash-dot line) together with its 95% uncertainty bound (dashed line); (b) ratio $\mathrm{std}(\delta_G(z, \hat{\theta}_S))/\mathrm{std}(\delta_G(z, \hat{\theta}))$: the sample value (solid line), the theoretical value (28) (dashed line) and the upper bound (30) (bold horizontal line).

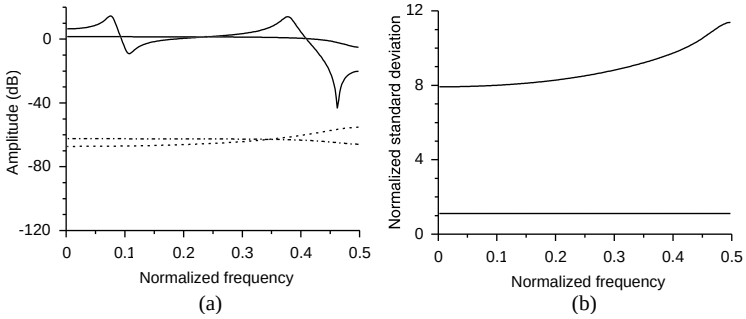

Fig. 2. Simulation results of the fourth-order discrete-time plant—model errors and $M = 7$: (a) the true plant $G_0(z)$ (solid line), the estimated first-order model (bold line) and the difference $\langle G(z, \hat{\theta})\rangle - \langle G(z, \hat{\theta}_S)\rangle$ (dash-dot line) together with its 95% uncertainty bound (dashed line); (b) ratio $\mathrm{std}(\delta_G(z, \hat{\theta}_S))/\mathrm{std}(\delta_G(z, \hat{\theta}))$: the sample value (solid line) and the upper bound (30) (bold horizontal line).

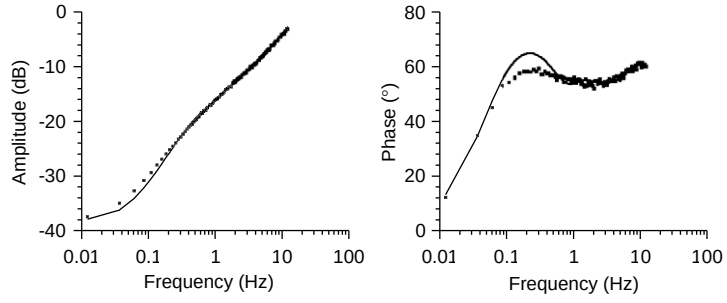

Fig. 3. Comparison between the measured (dots) and modelled (solid line) q-axis impedance of a synchronous machine. The model $n_b = 4$ over $n_a \leqslant 4$ is calculated with $r = 61$ in subspace Algorithm 2.

in both the amplitude and frequency band. For physical reasons, the impedance should be modelled with a rational form in s of order n_b over $n_a = n_b - 1$. Since the model is improper, the role of the input and the output should be interchanged in subspace Algorithm 2.2 (see Section 5), so that at the end a model of order n_b over $n_a \leqslant n_b$ is estimated. The influence of the parameter r in the sample subspace Algorithm 2.2 on the cost function (35) is shown in Fig. 4 for a model of order $n_b = 4$ over $n_a \leqslant 4$. It can be seen that several local minima exist and that $r = 61$ gives the smallest value of (35) (Fig. 3 shows the model estimated with $r = 61$). It also follows from Fig. 4 that it can be dangerous to estimate the model order for a fixed value of r in the subspace algorithms (see, for example, step 1c of Algorithm 1).

7. Conclusion

In this paper, the stochastic properties of frequency-domain subspace algorithms using the noise covariance matrix have been analysed. It turns out that replacing the true noise covariance matrix by the sample noise covariance matrix, obtained from a small number of independent experiments, does almost not affect the stochastic properties of the estimated model parameters. The strong convergence, strong consistency, convergence rate and asymptotic normality are preserved, and in the absence of model errors the uncertainty increases slightly only. In the presence of model errors the increase in uncertainty can be (very) large. The sample subspace estimates of A, C require at least two independent repeated experiments, while at least four are needed for B, D.

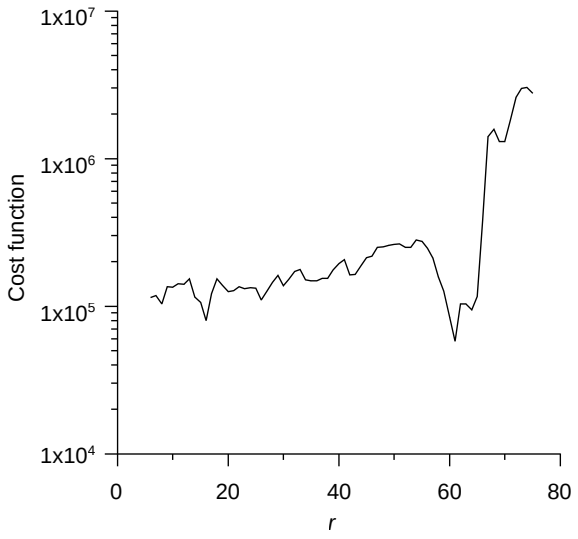

Fig. 4. Modelling of the q-axis impedance of a synchronous machine: cost function (35) as a function of the parameter r in subspace Algorithm 2 for model $n_b = 4$ over $n_a \leqslant 4$.

Acknowledgements

This work is sponsored by the Fund for Scientific Research (FWO-Vlaanderen), the Flemish Government (GOA-IMMI) and the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy programming (IUAP 4/2).

Appendix A. Lemmas used in the proofs of Theorems 1 and 2

The first three lemmas of this appendix study the asymptotic ($F \to \infty$) properties of the vector function $f_F(\hat{\theta}, Z) \in \mathbb{R}^{n_f}$, where $\hat{\theta} \in \mathbb{R}^{n_\theta}$ is a stochastic vector, $Z \in \mathbb{R}^F$ are the noisy observations, and $n_f = \dim(f_F)$ and $n_\theta = \dim(\theta)$ are independent of F. The convergence, the convergence rate and the asymptotic distribution function are analysed. The abbreviations w.p. 1 and in prob. respectively, stand, for convergence with probability one (almost sure convergence) and convergence in probability. Similarly, $O_{\text{a.s.}}(F^k)$, a.s.lim and $O_{\text{p}}(F^k)$, plim stand for $O(F^k)$, lim w.p. 1 and in prob., respectively.

Lemma 1 (Strong convergence). *Let $Z \in \mathbb{R}^F$ be a stochastic variable. Let $f_F(\theta, Z) \in \mathbb{R}^{n_f}$ be a continuous function of θ in Θ_{r}, a compact subset of $\mathbb{R}^{n_\theta}$, with $n_f = \dim(f_F)$ and $n_\theta = \dim(\theta)$ independent of F. If*

(1) *$f_F(\theta, Z)$ converges uniformly w.p. 1 to $f(\theta)$ in Θ_{r},*
(2) *$\hat{\theta}$ converges w.p. 1 to θ_*, an interior point of Θ_{r},*

then $f_F(\hat{\theta}, Z)$ converges w.p. 1 to $f(\theta_*)$.

Proof. Consider those stochastic realizations of Z for which $f_F(\theta, Z)$ converges uniformly to $f(\theta)$ in Θ_{r} and $\hat{\theta}$ converges to θ_*. Due to the almost sure convergence, these realizations have probability measure one. Choose an arbitrary $\varepsilon > 0$ and construct the set $\Theta_\varepsilon = \{\theta \mid \|\theta - \theta_*\|_2 < \varepsilon\} \subset \Theta_{\text{r}}$. Since for any of the considered realizations Z, $f_F(\theta, Z)$ converges uniformly to $f(\theta)$ in Θ_{r} and $\hat{\theta}$ converges to θ_*, there exists, for any $\delta > 0$, an F_0 independent of θ, such that for any $F \geqslant F_0$

$$\hat{\theta} \in \Theta_\varepsilon \text{ and } \|f_F(\theta, Z) - f(\theta)\|_2 \leqslant \delta/2 \quad \text{for any } \theta \in \Theta_{\text{r}}. \tag{A.1}$$

The function $f(\theta)$ is continuous in Θ_{r} since it is the limit of a uniformly convergent sequence of continuous functions (see Kaplan, 1993, Theorem 31, Remark 2). Hence, there exists an ε such that

$$\|f(\theta) - f(\theta_*)\|_2 \leqslant \delta/2 \quad \text{for any } \theta \in \Theta_\varepsilon. \tag{A.2}$$

Combining (A.1) and (A.2) shows that for any $\delta > 0$, there exists an ε and an F_0 such that for any $F \geqslant F_0$

$$\|f_F(\hat{\theta}, Z) - f(\theta_*)\|_2 \leqslant \|f_F(\hat{\theta}, Z) - f(\hat{\theta})\|_2 + \|f(\hat{\theta}) - f(\theta_*)\|_2 \leqslant \delta. \tag{A.3}$$

Making δ arbitrarily small and noting that the considered realizations Z occur w.p. 1, reveals directly that $\text{a.s.lim}_{F\to\infty}(f_F(\hat{\theta}, Z) - f(\theta_*)) = 0$ or $f_F(\hat{\theta}, Z) = f(\theta_*) + o_{\text{a.s.}}(F^0)$. □

Lemma 2 (Convergence rate). *Consider the conditions of Lemma* 1. *If in addition*

(1) *$f_F(\theta, Z)$ converges uniformly at the rate $O_{\text{p}}(F^{-m})$ to $f(\theta)$ in Θ_{r},*
(2) *$f'_F(\theta, Z) = \partial f_F(\theta, Z)/\partial\theta$ is a continuous function in Θ_{r} with $\|f'_F(\theta, Z)\|_2 \leqslant O_{\text{a.s.}}(F^0)$ uniformly in Θ_{r},*
(3) *$\hat{\theta}$ converges at the rate $O_{\text{p}}(F^{-m})$ to θ_*.*

then $f_F(\hat{\theta}, Z)$ converges w.p. 1 to $f(\theta_*)$ at the rate $O_{\text{p}}(F^{-m})$.

Proof. Note that the conditions of Lemma 1 are satisfied so that only the convergence rate must be proven. Applying the mean value theorem (Kaplan, 1993) to $f_F(\theta, Z)$ at the points $\hat{\theta}, \theta_*$ gives

$$f_F(\hat{\theta}, Z) = f_F(\theta_*, Z) + f'_F(\widehat{\theta}, Z)(\hat{\theta} - \theta_*) \tag{A.4}$$

with $\widehat{\theta}$ a point on the straight line connecting $\hat{\theta}$ to θ_* ($\widehat{\theta} = t\hat{\theta} + (1-t)\theta_*$ with $t \in [0, 1]$). $\widehat{\theta}$ converges w.p. 1 to θ_* since

$$\underset{F\to\infty}{\text{a.s.lim}}(\widehat{\theta} - \theta_*) = (\lim_{F\to\infty} t)\,\underset{F\to\infty}{\text{a.s.lim}}(\hat{\theta} - \theta_*) = 0. \tag{A.5}$$

Consider the realizations Z for which $\widehat{\theta}$ converges to θ_* and $\|f'_F(\theta, Z)\|_2 \leqslant O(F^0)$ uniformly in Θ_{r} (condition 2). For these realizations, there is an F_0 such that for any $F \geqslant F_0$,

$\widehat{\theta} \in \Theta_r$ and, hence, $\|f'_F(\widehat{\theta}, Z)\|_2 \leqslant O(F^0)$. Since these realizations occur w.p. 1, we have $\|f'_F(\widehat{\theta}, Z)\|_2 \leqslant O_{a.s.}(F^0)$ and, hence, $\|f'_F(\widehat{\theta}, Z)\|_2 \leqslant O_p(F^0)$ (convergence w.p. 1 implies convergence in prob., see Lukacs, 1975). Putting this result in (A.4), taking into account that $f_F(\theta_*, Z) = f(\theta_*) + O_p(F^{-m})$ (condition 1), $\hat{\theta} = \theta_* + O_p(F^{-m})$ (condition 3), and that n_θ and n_f are F-independent integers, proves the lemma. □

Corollary to Lemma 2. *Consider the conditions of Lemma* 2 *with* $m = 1/2$. *If in addition*

(1) $f'_F(\theta, Z)$ *converges uniformly at the rate* $O_p(F^{-1/2})$ *to* $f'(\theta)$ *in* Θ_r,
(2) $g_{ij}(\theta, Z) = \partial^2 f_F(\theta, Z)/\partial\theta_{[i]}\partial\theta_{[j]}$ *is a continuous function in* Θ_r *with* $\|g_{ij}(\theta, Z)\|_2 \leqslant O_{a.s.}(F^0)$ *uniformly in* Θ_r,

then $f_F(\hat{\theta}, Z) = f(\theta_*) + \delta_f(Z) + O_p(F^{-1})$ *with* $\delta_f(Z) = f'_F(\theta_*, Z)(\hat{\theta} - \theta_*) = O_p(F^{-1/2})$.

Proof. Taylor series expansion of $f_F(\theta, Z)$ at the points $\hat{\theta}, \theta_*$ gives (Kaplan, 1993).

$$f_F(\hat{\theta}, Z) = f_F(\theta_*, Z) + f'_F(\theta_*, Z)(\hat{\theta} - \theta_*) + \frac{1}{2}\sum_{i,j} \times(\hat{\theta}_{[j]} - \theta_{*[j]}) \frac{\partial^2 f_F(\theta, Z)}{\partial\widehat{\theta}_{[i]}\,\partial\widehat{\theta}_{[j]}} (\hat{\theta}_{[i]} - \theta_{*[i]}), \quad \text{(A.6)}$$

with $\widehat{\theta}$ a point on the straight line connecting $\hat{\theta}$ to θ_* ($\widehat{\theta} = t\hat{\theta} + (1-t)\theta_*$ with $t \in [0, 1]$). Following the lines of the Proof of Lemma 2, it can easily be seen that the second and third terms on the right-hand side of (A.6) are an $O_p(F^{-1/2})$ and $O_p(F^{-1})$, respectively. □

Lemma 3 (Asymptotic normality). *Consider the conditions of Lemmas* 1 *and* 2 *with* $m = 1/2$. *If in addition*

(1) $f_F(\theta, Z)$ *is asymptotically normally distributed in* Θ_r,
(2) $f'_F(\theta, Z)$ *converges uniformly w.p.* 1 *to* $f'(\theta)$ *in* Θ_r.
(3) $\hat{\theta}$ *is asymptotically normally distributed,*
then $f_F(\hat{\theta}, Z)$ *converges w.p.* 1 *to* $f(\theta_*)$ *at the rate* $O_p(F^{-1/2})$, *and is asymptotically normally distributed.*

Proof. Only the asymptotic normality must be proven. Since $\widehat{\theta}$ converges w.p. 1 to θ_* (see Eq. (A.5)), it follows from condition 2 that $f'_F(\widehat{\theta}, Z) = f'(\theta_*) + o_{a.s.}(F^0)$ (proof. apply Lemma 1). Putting this result in (A.4), taking into account conditions 1 and 3 of Lemma 2 with $m = 1/2$, $f_F(\theta_*, Z) - f(\theta_*) = O_p(F^{-1/2})$ and $\hat{\theta} - \theta_* = O_p(F^{-1/2})$, gives

$$f_F(\hat{\theta}, Z) - f(\theta_*) = \delta_F(Z) + o_p(F^{-1/2}),$$

$$\delta_F(Z) = f_F(\theta_*, Z) - f(\theta_*) + f'(\theta_*)(\hat{\theta} - \theta_*), \quad \text{(A.7)}$$

where $\delta_F(Z) = O_p(F^{-1/2})$ is asymptotically normally distributed (a finite linear combination of asymptotically normally distributed random variables is asymptotically normally distributed). Multiplying (A.7) by $\sqrt{F}$ and taking the limit gives

$$\operatorname*{plim}_{F\to\infty} \sqrt{F}(f_F(\hat{\theta}, Z) - f(\theta_*) - \delta_F(Z)) = 0. \quad \text{(A.8)}$$

Since convergence in probability implies convergence in law (see Lukacs, 1975), it follows from (A.8) that $\sqrt{F}(f_F(\hat{\theta}, Z) - f(\theta_*))$ is asymptotically normally distributed. □

Lemma 4 (Properties of left singular values). *Let* u_k *be the left singular vector corresponding to the singular value* σ_k *of a real* $n \times F$ *matrix* A *with* $F > n$. *The non-linear map relating* u_k *and* σ_k *to* A *has the following properties*: (i) *it is a continuous function of* AA^T; (ii) *it is a continuous function of* AA^T *with continuous derivative if* $\lambda_k = \sigma_k^2$ *is a simple eigenvalue of* AA^T.

Proof. Property (i) is shown in Golub and Van Loan (1996). Property (ii) follows directly from the implicit function theorem (Kaplan, 1993). Indeed, λ_k is a solution of the polynomial equation $P(\lambda) = 0$ with $P(\lambda) = \det(\lambda I_F - AA^T)$ and I_F the $F \times F$ identity matrix. If λ_k is a simple eigenvalue of AA^T, then it is a simple root of $P(\lambda) = 0$ so that $P'(\lambda_k) \neq 0$, with $'$ the derivative w.r.t. λ. Since $P'(\lambda_k) \neq 0$ it follows from the implicit function theorem that $\lambda_k = \lambda_k(AA^T)$ is a continuous function with continuous derivative. Since u_k is the solution of $AA^T u_k = \lambda_k(AA^T)u_k$ with $\|u_k\|_2 = 1$, $u_k = u_k(AA^T)$ is also a continuous function with continuous derivative. □

Lemma 5 (Estimation range space—asymptotic properties). *Consider the real* $n \times F$ *matrix* $A = A_0 + N$, *with* $F > n$, A_0 *the deterministic part with* $\operatorname{rank}(A_0) = r_A \leqslant n$, *and* N *the zero mean noise contribution. Form the matrix* $B = C_N^{-1/2}A$, *with* $C_N^{1/2}$ *a square root of* $C_N = E\{NN^T\}$ ($C_N = C_N^{1/2}C_N^{T/2}$), *and calculate its singular value decomposition* $B = U\Sigma V^T$, *where* $\operatorname{diag}(\Sigma)$ *contains the ordered singular values* (*from large to small*).

(i) *If* NN^T/F *and* A_0N^T/F *converge w.p.* 1 *for* $F \to \infty$ *to* C_N/F *and* 0, *respectively, then* $(C_N/F)^{1/2}U_{[:,1:r_A]}$ *converges* ($F \to \infty$) *w.p.* 1 *to range* (A_0).
(ii) *If in addition to* (i), *the non-zero singular values of* B *are simple eigenvalues of* BB^T *for any* $F \geqslant F_0 > 0$ (*infinity included*) *and the convergence rates of* NN^T/F *and* A_0N^T/F *are an* $O_p(F^{-1/2})$, *then the convergence rate of* $(C_N/F)^{1/2}U_{[:,1:r_A]}$ *is an* $O_p(F^{-1/2})$.
(iii) *If in addition to* (i) *and* (ii), NN^T/F *and* A_0N^T/F *are asymptotically normally distributed, then* $(C_N/F)^{1/2}U_{[:,1:r_A]}$ *is asymptotically normally distributed.*

Proof. The matrix $U_{[:,1:r_A]}$ contains the $r_A = \operatorname{rank}(A_0)$ left to singular vectors of B corresponding to the r_A largest sin-

gular values of B. Hence, to study the asymptotic ($F \to \infty$) behaviour of $U_{[:,1:r_A]}$, it is sufficient to study the asymptotic ($F \to \infty$) behaviour of BB^{T}.

(i) Under the assumptions (i) of the lemma, it follows that for $F \to \infty$,

$$BB^{\mathrm{T}} \to E\{BB^{\mathrm{T}}\} = C_N^{-1/2} A_0 A_0^{\mathrm{T}} C_N^{-\mathrm{T}/2} + I_n \text{ w.p. 1.} \quad \text{(A.9)}$$

Using the singular value decomposition (SVD) $C_N^{-1/2} A_0 = U_0 \Sigma_0 V_0^T$, where $\mathrm{diag}(\Sigma_0)$ contains the ordered singular values (from large to small), the right-hand side of (A.9) can be written as

$$C_N^{-1/2} A_0 A_0^{\mathrm{T}} C_N^{-\mathrm{T}/2} + I_n = U_0 \Sigma_0^2 U_0^{\mathrm{T}} + I_n = U_0(\Sigma_0^2 + I_n) U_0^{\mathrm{T}}. \quad \text{(A.10)}$$

It follows from (A.9) and (A.10) that

$$\mathrm{SVD}(E\{BB^{\mathrm{T}}\}) = U_0(\Sigma_0^2 + I_n) U_0^{\mathrm{T}} \quad \text{with}$$

$$U_{0[:,1:r_A]} = \mathrm{range}(C_N^{-1/2} A_0). \quad \text{(A.11)}$$

Since the left singular values of B are continuous functions of BB^{T} (Lemma 4(i)), it follows from (A.9) and (A.11) that all the conditions of Lemma 1 with $f_F(\theta, Z) = f(\theta)$ and $\theta = \mathrm{vec}(BB^{\mathrm{T}})$ are satisfied. Hence, $U_{[:,1:r_A]}$ converges w.p. 1 to $U_{0[:,1:r_A]} = \mathrm{range}(C_N^{-1/2} A_0)$, which implies that $(C_N/F)^{1/2} U_{[:,1:r_A]}$ converges w.p. 1 to range (A_0).

(ii) If for any $F \geqslant F_0 > 0$, infinity included, the non-zero singular values of B are simple eigenvalues of BB^{T}, then the range space of B is for any $F \geqslant F_0 > 0$, infinity included, a continuous function of BB^{T} with continuous derivative (Lemma 4(ii)). Hence all the conditions of Lemma 2 with $f_F(\theta, Z) = f(\theta)$ and $\theta = \mathrm{vec}(BB^{\mathrm{T}})$ are satisfied so that $(C_N/F)^{1/2} U_{[:,1:r_A]}$ converges at the rate $O_{\mathrm{p}}(F^{-1/2})$ to range (A_0).

(iii) All the conditions of Lemma 3 with $f_F(\theta, Z) = f(\theta)$ and $\theta = \mathrm{vec}(BB^{\mathrm{T}}/F)$ are satisfied so that $(C_N/F)^{1/2} U_{[:,1:r_A]}$ is asymptotically normally distributed. □

Lemma 6 (Properties noise on data matrix). *Under Assumptions* 1–3,6 *and* 7, $\mathbf{Y}^{\mathrm{re}} \Pi$ *and* $C_{\mathbf{Y}}$ *satisfy all the conditions of Lemma* 5: *Assumptions* 1–3 *for Lemma* 5(i), *Assumptions* 1–3, *and* 6 *for Lemma* 5(ii), *and Assumptions* 1–3,6 *and* 7 *for Lemma* 5(iii).

Proof. The following sums are used in the analysis

$$\frac{1}{F} \mathbf{N}_{\mathbf{Y}}^{\mathrm{re}} \mathbf{V}^{\mathrm{re\,T}} = \mathrm{Re}\left(\frac{1}{F} \sum_{k=1}^{F} N_Y(k) \bar{V}(k) W_r(k) W_r^H(k) \right),$$

$$\frac{1}{F} \mathbf{N}_{\mathbf{Y}}^{\mathrm{re}} \mathbf{N}_{\mathbf{Y}}^{\mathrm{re\,T}} = \mathrm{Re}\left(\frac{1}{F} \sum_{k=1}^{F} |N_Y(k)|^2 W_r(k) W_r^H(k) \right), \quad \text{(A.12)}$$

where $\mathbf{V}, V$ equal $\mathbf{U}, U$ or $\mathbf{Y}_0, Y_0$.

Conditions of Lemma 5(i): Under Assumption 1, $N_Y(k)\bar{V}(k)$ and $|N_Y(k)|^2$ are independent (over k) random variables with uniformly bounded second-order moments. Applying the strong law of large numbers for independent random variables (Lukacs, 1975) to the sums in (A.12) shows that $\mathbf{N}_{\mathbf{Y}}^{\mathrm{re}} \mathbf{V}^{\mathrm{re\,T}}/F$ and $\mathbf{N}_{\mathbf{Y}}^{\mathrm{re}} \mathbf{N}_{\mathbf{Y}}^{\mathrm{re\,T}}/F$ converge w.p. 1 to their expected values 0 and $C_{\mathbf{Y}}/F$, respectively. Under Assumption 3, this implies that $\mathbf{N}_{\mathbf{Y}}^{\mathrm{re}} \Pi \mathbf{Y}_0^{\mathrm{re\,T}}/F$ and $\mathbf{N}_{\mathbf{Y}}^{\mathrm{re}} \Pi \mathbf{N}_{\mathbf{Y}}^{\mathrm{re\,T}}/F$, with Π defined in (20), converge w.p. 1 to 0 and $C_{\mathbf{Y}}/F$, respectively.

Conditions of Lemma 5(ii): Because $N_Y(k)\bar{V}(k)$ and $|N_Y(k)|^2$ are independent over the frequency k, it is easy to verify that under Assumption 1,

$$\mathrm{Cov}\left(\mathrm{vec}\left(\frac{1}{F} \mathbf{N}_{\mathbf{Y}}^{\mathrm{re}} \mathbf{V}^{\mathrm{re\,T}} \right) \right) = O(F^{-1}) \quad \text{and}$$

$$\mathrm{Cov}\left(\mathrm{vec}\left(\frac{1}{F} \mathbf{N}_{\mathbf{Y}}^{\mathrm{re}} \mathbf{N}_{\mathbf{Y}}^{\mathrm{re\,T}} \right) \right) = O(F^{-1}) \quad \text{(A.13)}$$

which shows that $\mathbf{N}_{\mathbf{Y}}^{\mathrm{re}} \mathbf{V}^{\mathrm{re\,T}}/F$ and $\mathbf{N}_{\mathbf{Y}}^{\mathrm{re}} \mathbf{N}_{\mathbf{Y}}^{\mathrm{re\,T}}/F$ converge in probability at the rate $O_{\mathrm{p}}(F^{-1/2})$ to their expected values 0 and $C_{\mathbf{Y}}/F$, respectively (Lukacs, 1975). Under Assumption 3, this implies that $\mathbf{N}_{\mathbf{Y}}^{\mathrm{re}} \Pi \mathbf{Y}_0^{\mathrm{re\,T}}/F$ and $\mathbf{N}_{\mathbf{Y}}^{\mathrm{re}} \Pi \mathbf{N}_{\mathbf{Y}}^{\mathrm{re\,T}}/F$ converge in probability at the rate $O_{\mathrm{p}}(F^{-1/2})$ to 0 and $C_{\mathbf{Y}}/F$, respectively.

Conditions of Lemma 5(iii): Using the results of part (ii) of this proof, it follows that

$$\mathbf{N}_{\mathbf{Y}}^{\mathrm{re}} \Pi \mathbf{N}_{\mathbf{Y}}^{\mathrm{re\,T}}/F = \mathbf{N}_{\mathbf{Y}}^{\mathrm{re}} \mathbf{N}_{\mathbf{Y}}^{\mathrm{re\,T}}/F + O_{\mathrm{p}}(F^{-1}), \quad \text{(A.14)}$$

where $(\mathbf{N}_{\mathbf{Y}}^{\mathrm{re}} \mathbf{N}_{\mathbf{Y}}^{\mathrm{re\,T}} - C_{\mathbf{Y}})/F = O_{\mathrm{p}}(F^{-1/2})$. Under Assumptions 1 and 7, the central limit theorem for independent random variables (Billingsley, 1995) can be applied to the sums in (A.12), which shows that $\mathbf{N}_{\mathbf{Y}}^{\mathrm{re}} \mathbf{V}^{\mathrm{re\,T}}/F$ and $\mathbf{N}_{\mathbf{Y}}^{\mathrm{re}} \mathbf{N}_{\mathbf{Y}}^{\mathrm{re\,T}}/F$ are asymptotically normally distributed. Under Assumption 3, it implies that $\mathbf{N}_{\mathbf{Y}}^{\mathrm{re}} \Pi \mathbf{Y}_0^{\mathrm{re\,T}}/F$ and $\mathbf{N}_{\mathbf{Y}}^{\mathrm{re}} \Pi \mathbf{N}_{\mathbf{Y}}^{\mathrm{re\,T}}/F$ (see Eq. (A.14)) are asymptotically normally distributed. □

Lemma 7 (Asymptotic behaviour $\tilde{O}_{\mathrm{r}}$). *Under Assumptions* 3,5(i),6,8 *and* 9, *the solution of the noiseless problem* $\tilde{O}_r$ $(Y(k), \sigma_Y^2(k) \to Y_0(k), \sigma_Y^2(k)$ *in step* 1 *of Algorithm* 1) *converges at the rate* $O(F^{-2})$ *to its limit value* O_{r*}.

Proof. Under Assumptions 3, 8 and 9, it will be shown subsequently that $\mathbf{Y}_0^{\mathrm{re}} \Pi \mathbf{Y}_0^{\mathrm{re\,T}}/F$ converges at the rate $O(F^{-2})$ to its limit value. Using Lemma 4 and Assumptions 5(i) and 6, it follows that the conditions of Lemma 2, with $m = 2$, $f_F(\theta, Z) = f(\theta)$ and $\theta = \mathrm{vec}(\mathbf{Y}_0^{\mathrm{re}} \Pi \mathbf{Y}_0^{\mathrm{re\,T}}/F)$, are satisfied for the n_a left singular vectors corresponding to the n_a largest singular values of $\mathbf{Y}_0^{\mathrm{re}} \Pi \mathbf{Y}_0^{\mathrm{re\,T}}/F$. Hence, $\tilde{O}_r = O_{r^*} + O(F^{-2})$, with $\tilde{O}_r = \mathrm{range}(\mathbf{Y}_0^{\mathrm{re}} \Pi)$ and $O_{r^*} = \lim_{F \to \infty} \tilde{O}_r$.

The matrix $\mathbf{Y}_0^{\mathrm{re}} \Pi \mathbf{Y}_0^{\mathrm{re\,T}}/F$ is a function of the following sums:

$$\frac{1}{F} \mathbf{V}^{\mathrm{re}} \mathbf{V}^{\mathrm{re\,T}} = \mathrm{Re}\left(\frac{1}{F} \sum_{k=1}^{F} |V(k)|^2 W_r(k) W_r^H(k) \right),$$

$$\frac{1}{F} \mathbf{Y}_0^{\mathrm{re}} \mathbf{U}^{\mathrm{re\,T}} = \mathrm{Re}\left(\frac{1}{F} \sum_{k=1}^{F} Y_0(k) \bar{U}(k) W_r(k) W_r^H(k) \right), \quad \text{(A.15)}$$

where $\mathbf{V}, V$ equal $\mathbf{U}, U$ or $\mathbf{Y}_0, Y_0$. Under Assumptions 8 and 9, the Riemann sums (A.15) converge to, respectively,

$$\mathrm{Re}\left(\int_{f_{\min}}^{f_{\max}} |V(f)|^2 W_r(f) W_r^H(f) n(f)\,\mathrm{d}f\right),$$

$$\mathrm{Re}\left(\int_{f_{\min}}^{f_{\max}} Y_0(f)\bar{U}(f) W_r(f) W_r^H(f) n(f)\,\mathrm{d}f\right) \qquad \text{(A.16)}$$

at the rate $O(F^{-2})$ (Ralston & Rabinowitz, 1984). Applying Lemma 3 with $m=2$, $f_F(\theta, Z)=f(\theta)$ and

$$\theta = [\mathrm{vec}^{\mathrm{T}}(\mathbf{U}^{\mathrm{re}}\mathbf{U}^{\mathrm{re\,T}}/F)\ \ \mathrm{vec}^{\mathrm{T}}(\mathbf{Y}_0^{\mathrm{re}}\mathbf{Y}_0^{\mathrm{re\,T}}/F)\ \ \mathrm{vec}^{\mathrm{T}}(\mathbf{Y}_0^{\mathrm{re}}\mathbf{U}^{\mathrm{re\,T}}/F)]^{\mathrm{T}}$$

to $\mathbf{Y}_0^{\mathrm{re}}\Pi\mathbf{Y}_0^{\mathrm{re\,T}}/F$ shows that, under Assumption 3, $\mathbf{Y}_0^{\mathrm{re}}\Pi\mathbf{Y}^{\mathrm{re\,T}}/F$ converges at the rate $O(F^{-2})$ to its limit value. □

Lemma 8 (Properties sample variance). *Under Assumptions* 11 *and* 12, $\sigma_Y^2(k)/\hat{\sigma}_Y^2(k)$, *where* $\sigma_Y^2(k)$ *and* $\hat{\sigma}_Y^2(k)$ *are defined in*, *Assumption* 2 *and Eq.* (25), *respectively*, *is a Fisher's* $F(\infty, 2M-2)$ *distributed random variable whose rth order moment exists if* $M > r+1$. *The first and second order moments of* $\sigma_Y^2(k)/\hat{\sigma}_Y^2(k)$ *are*

$$E\{\sigma_Y^2(k)/\hat{\sigma}_Y^2(k)\} = \frac{M-1}{M-2},$$

$$E\{\sigma_Y^4(k)/\hat{\sigma}_Y^4(k)\} = \frac{(M-1)^2}{(M-2)(M-3)}.$$

Proof. See Stuart and Ord (1987). □

Appendix B. Proof of Theorem 1

The proof of the theorem is divided in three steps: first the properties of $\hat{O}_r$ are proven, next the properties of $\hat{A}$, $\hat{C}$, and finally, the properties of $\hat{B}$, $\hat{D}$. The discrete-time case (Algorithm 1) is handled first and next the continuous-time case (Algorithm 2). Some frequently used abbreviations and notations are explained in the introduction of Appendix A.

B.1. Estimate of the extended observability matrix

From step 1c of Algorithm 1, it follows that $\hat{O}_r = C_{\mathbf{Y}}^{1/2} U_{[:,1:n_a]}$, where $U_{[:,1:n_a]}$ contains the n_a left singular vectors corresponding to the n_a largest singular values of $C_{\mathbf{Y}}^{-1/2} R_{22}^{\mathrm{T}}$. Using the QR-factorization in step 1b, it can be easily found that $\mathbf{Y}^{\mathrm{re}}$ and Π in Eq. (20) can be written as $\mathbf{Y}^{\mathrm{re}} = R_{12}^{\mathrm{T}} Q_1^{\mathrm{T}} + R_{22}^{\mathrm{T}} Q_2^{\mathrm{T}}$ and $\Pi = I_{2F} - Q_1 Q_1^{\mathrm{T}}$, respectively. Since $Q_1^{\mathrm{T}} Q_1 = I_r$ and $Q_2^{\mathrm{T}} Q_1 = 0$, it follows that $\mathbf{Y}^{\mathrm{re}}\Pi = R_{22}^{\mathrm{T}} Q_2^{\mathrm{T}}$. Hence, the stochastic properties of $\hat{O}_r$ are determined by the asymptotic properties of the n_a left singular vectors corresponding to the n_a largest singular values of $C_{\mathbf{Y}}^{-1/2}\mathbf{Y}^{\mathrm{re}}\Pi$. Under the assumptions of Theorem 1, it is shown in Lemma 6 that $\mathbf{Y}^{\mathrm{re}}\Pi$ and $C_{\mathbf{Y}}$ satisfy all the conditions of Lemma 5. Hence, $(C_{\mathbf{Y}}/F)^{1/2} U_{[:,1:r_Y]}$, with $r_Y = \mathrm{rank}(\mathbf{Y}_0^{\mathrm{re}}\Pi)$, converges w.p. 1 to $\mathrm{range}(\mathbf{Y}_0^{\mathrm{re}}\Pi)$ at the rate $O_{\mathrm{p}}(F^{-1/2})$, and is asymptotically normally distributed. These results imply the following properties for $\hat{O}_r$: Assumption 5(i) guarantees that $\mathrm{rank}(\mathbf{Y}_0^{\mathrm{re}}\Pi) \geqslant n_a$. Hence, in the presence of model errors, $\hat{O}_r = (C_{\mathbf{Y}}/F)^{1/2} U_{[:,1:n_a]}$ converges strongly to the solution of the noiseless problem $\tilde{O}_r = (C_{\mathbf{Y}}/F)^{1/2} U_{0[:,1:n_a]}$, with $\mathrm{SVD}(C_{\mathbf{Y}}^{-1/2}\mathbf{Y}_0^{\mathrm{re}}\Pi) = U_0\Sigma_0 V_0^{\mathrm{T}}$. The convergence rate is an $O_{\mathrm{p}}(F^{-1/2})$, and $\hat{O}_r$ is asymptotically normally distributed. Under Assumptions 3, 5(i), 6, 8 and 9, it is shown in Lemma 7 that $\tilde{O}_r$ converges at the rate $O(F^{-2})$ to its limit value O_{r^*}. If Assumptions 4 (distinct frequencies) and 10 (no model errors) are satisfied, then it has been shown in McKelvey et al. (1996) that $\mathrm{rank}(\mathbf{Y}_0^{\mathrm{re}}\Pi) = n_a$ and $\tilde{O}_r = O_{r0}$ which proves the strong consistency of $\hat{O}_r$.

B.2. Estimates $\hat{A}, \hat{C}$

To prove that $\hat{\alpha}$ has the same asymptotic properties as $\hat{O}_r$, it is sufficient to notice that $\hat{A}, \hat{C}$ in (19) are continuous functions of $\hat{O}_r$ with continuous derivative, and to apply Lemmas 1–3 to $\hat{A}, \hat{C}$ with $f_F(\theta, Z) = f(\theta)$ and $\theta = \mathrm{vec}(O_r)$.

B.3. Estimates $\hat{B}, \hat{D}$

Minimization of cost function (17) w.r.t. $\beta = [B^{\mathrm{T}} D]^{\mathrm{T}}$ with $W(k) = 1/\sigma_Y(k)$ gives

$$\hat{\beta} = [\mathrm{Re}(H^H C_Y^{-1} H)]^{-1}\mathrm{Re}(H^H C_Y^{-1} Y), \qquad \text{(B.1)}$$

where $C_Y = \mathrm{diag}(\sigma_Y^2(1), \sigma_Y^2(2), \ldots, \sigma_Y^2(F))$, $Y_{[k]} = Y(k)$, and $H = H(\hat{\alpha}) = -\partial e(\beta, \hat{\alpha})/\partial\beta$ with $e(\beta, \hat{\alpha})$ an $F \times 1$ vector with kth entry

$$e_{[k]}(\beta, \hat{\alpha}) = Y(k) - [\hat{C}(\xi_k I_{n_a} - \hat{A})^{-1} B + D] U(k).$$

The asymptotic ($F \to \infty$) properties of $\hat{\beta}$ are shown as follows. First, Lemmas 1–3 are applied to the stochastic sums

$$H^H(\hat{\alpha}) C_Y^{-1} H(\hat{\alpha})/F = f_F(\hat{\theta})$$

$$\text{and} \quad H(\hat{\alpha})^H C_Y^{-1} Y/F = f_F(\hat{\theta}, Z) \qquad \text{(B.2)}$$

in (B.1), where $\theta = \alpha$ and $Z = Y$. Next, applying Lemmas 1–3 to $\hat{\beta}$ (B.1) with $f_F(\theta, Z) = f(\theta)$ and

$$\theta = [\mathrm{vec}^{\mathrm{T}}(H^H(\hat{\alpha}) C_Y^{-1} H(\hat{\alpha})/F) \quad \mathrm{vec}^{\mathrm{T}}(H(\hat{\alpha})^H C_Y^{-1} Y/F)]^{\mathrm{T}}$$

shows that $\hat{\beta}$ has exactly the same asymptotic properties as the stochastic sums (B.2). A similar reasoning is followed to study the deterministic convergence ($F \to \infty$) of the solution $\tilde{\beta}$

$$\tilde{\beta} = [\mathrm{Re}(H^H(\tilde{\alpha}) C_Y^{-1} H(\tilde{\alpha}))]^{-1}\mathrm{Re}(H^H(\tilde{\alpha}) C_Y^{-1} Y_0) \qquad \text{(B.3)}$$

of the noiseless problem.

(i) *Strong convergence* (*consistency*) *and convergence rate*: Under Assumptions 1–6, $\hat{\alpha}$ converges strongly to $\tilde{\alpha}$

at the rate $O_p(F^{-1/2})$ (see Section B.2). Under Assumption 1, $f_F(\theta,Z) = H(\alpha)^H C_Y^{-1} Y/F$ converges strongly to $E\{f_F(\theta,Z)\} = H(\alpha)^H C_Y^{-1} Y_0/F$ at the rate $O_p(F^{-1/2})$. The convergence is uniform in Θ_r, a closed and bounded neighbourhood of $\tilde{\theta}$. The first statement is proven by the strong law of large numbers (Lukacs, 1975), the second by $\mathrm{Cov}(\mathrm{vec}(f_F(\theta,Z))) = O(F^0)$ (Lukacs, 1975), and the third by the continuity of $f_F(\theta,Z)$ in Θ_r. All the conditions of Lemmas 1 and 2 are satisfied, so that $f_F(\hat{\theta})$ and $f_F(\hat{\theta},Z)$ in (B.2) converge strongly at the rate $O_p(F^{-1/2})$ to $f_F(\tilde{\theta})$ and $E\{f_F(\tilde{\theta},Z)\}$, respectively. Hence, $\hat{\beta}$ converges strongly at the rate $O_p(F^{-1/2})$ to $\tilde{\beta}$. Because $\hat{\beta}$ is a continuous function of the sums in (B.2) with continuous second-order derivatives, it is easy to verify that all the conditions of the Corollary to Lemma 2 are satisfied. Hence, the convergence rate of $\hat{\beta}$ can be refined as

$$\hat{\beta} = \tilde{\beta} + \delta_\beta + O_p(F^{-1})$$

with $\delta_\beta = O_p(F^{-1/2})$. Under Assumption 10 (no model errors), $\tilde{\alpha} = \alpha_0$ (see Section B.2) so that $\tilde{\beta} = \beta_0$, which proves the strong consistency of $\hat{\beta}$.

(ii) *Convergence* $\tilde{\beta}$: Under Assumptions 8 and 9, $\tilde{\alpha}$ converges to α_* at the rate $O(F^{-2})$ (see Section B.2). Under the same assumptions, the Riemann sums

$$H^H(\alpha)C_Y^{-1}H(\alpha)/F \quad \text{and} \quad H^H(\alpha)C_Y^{-1}Y_0/F$$

converge at the rate $O(F^{-2})$ to their limit values for any α in a closed and bounded neighbourhood of α_* (Ralston & Rabinowitz, 1984). Applying Lemma 2 to $\tilde{\beta}$ (B.3) with $m = 2$ shows that $\tilde{\beta} = \beta_* + O(F^{-2})$.

(iii) *Asymptotic normality*: Under Assumptions 1–7, $\hat{\alpha}$ converges strongly to $\tilde{\alpha}$ at the rate $O_p(F^{-1/2})$ and is asymptotically normally distributed (see Section B.2). Under Assumptions 1–7, the sum $f_F(\theta,Z)$ in (B.3) converges uniformly in Θ_r w.p. 1 and at the rate $O_p(F^{-1/2})$ to $E\{f_F(\theta,Z)\}$ (see part (i) of this section), and is asymptotically normally distributed (proof: see Billingsley, 1995, central limit theorem for independent random variables). Hence, all the conditions of Lemma 3 are satisfied for $f_F(\hat{\theta})$ and $f_F(\hat{\theta},Z)$ in (B.2), so that $f_F(\hat{\theta})$ and $f_F(\hat{\theta},Z)$ are asymptotically normally distributed. It follows directly that $\hat{\beta}$ is asymptotically normally distributed.

B.4. Continuous-time systems

The proof of Algorithm 2.1 follows directly from that of Algorithm 1. The main difference between Algorithms 2.2 and 1 is the Gram–Schmidt orthogonalization of the input and output data in (18). Hence, the proof of Algorithm 2.2 merely boils down to showing that (i) the coefficients γ_n in the Gram–Schmidt orthogonalization procedure converge strongly at the rate $O_p(F^{-1/2})$ to $\tilde{\gamma}_n$; (ii) the γ_n's are asymptotically normally distributed; and (iii) the $\tilde{\gamma}_n$'s converge at the rate $O(F^{-2})$ to their limit values γ_{n^*}. Under Assumptions 1 and 7–9, this can be easily shown using Lemmas 1–3, the strong law of large numbers for independent random variables (Lukacs, 1975), the central limit theorem for independent random variables (Billingsley, 1995), and the convergence rate of a Riemann sum to the corresponding Riemann integral (Ralston & Rabinowitz, 1984).

Appendix C. Proof of Theorem 2

The proof is divided into two steps: first the properties of $\hat{O}_{rS}$, $\hat{A}_S$ and $\hat{C}_S$ are proven, next those of $\hat{B}_S$ and $\hat{D}_S$. Only the discrete-time case (Algorithm 1) is handled. The proof of the continuous-time case (Algorithm 2) follows exactly the same lines. Some frequently used abbreviations and notations are defined in the introduction of Appendix A.

C.1. Estimates $\hat{O}_{rS}$, $\hat{A}_S$ and $\hat{C}_S$

Since $\hat{A}_S$ and $\hat{C}_S$ depend on $\hat{O}_{rS}$ only (see step 2 of Algorithm 1), it is sufficient to prove the properties for $\hat{O}_{rS}$. From the proof of the properties of $\hat{O}_r$ (see Section B.1) and Lemmas 1–3, it can be concluded that it is sufficient to study the asymptotic behaviour of $\hat{C}_{\hat{\mathbf{Y}}}$

$$\hat{C}_{\hat{\mathbf{Y}}}/F = \mathrm{Re}\left(\frac{1}{F}\sum_{k=1}^{F}(\hat{\sigma}_Y^2(k)/M)W_r(k)W_r^H(k)\right). \tag{C.1}$$

The stochastic sum (C.1) has the following asymptotic properties. (i) Under Assumptions 1 and 11 ($M \geqslant 2$), $\hat{C}_{\hat{\mathbf{Y}}}/F$ converges strongly to its expected value $C_{\hat{\mathbf{Y}}}/F = C_{\mathbf{Y}}/(MF)$ at the rate $O_p(F^{-1/2})$ (proof: the strong law of large numbers for independent random variables with $E\{\hat{\sigma}_Y^2(k)\} = \sigma_Y^2(k)$ for $M \geqslant 2$, and $\mathrm{Cov}(\mathrm{vec}(\hat{C}_{\hat{\mathbf{Y}}}/F)) = O(F^0)$, Lukacs, 1975). (ii) Under Assumptions 1, 7 and 11 ($M \geqslant 2$)$\hat{C}_{\hat{\mathbf{Y}}}/F$ is asymptotically normally distributed (proof: central limit theorem for independent random variables, Billingsley, 1995). (iii) Under Assumptions 8 and 9, the Riemann sum $C_{\hat{\mathbf{Y}}}/F$ converges at the rate $O_p(F^{-2})$ to

$$\mathrm{Re}\left(\int_{f_{\min}}^{f_{\max}}(\sigma_Y^2(f)/M)W_r(f)W_r^H(f)n(f)\mathrm{d}f\right)$$

(proof: Ralston & Rabinowitz, 1984). We conclude that all conditions of Lemmas 1–3 are fulfilled for $\hat{O}_{rS}$, so that $\hat{O}_{rS}$, with $M \geqslant 2$, has the same asymptotic properties as $\hat{O}_r$.

C.2. Estimates $\hat{B}_S$ and $\hat{D}_S$

Replacing $Y(k)$, $\sigma_Y^2(k)$ by $\hat{Y}(k)$, $\hat{\sigma}_Y^2(k)/M$ in (B.1) gives

$$\hat{\beta}_S = [\mathrm{Re}(H_S^H \hat{C}_{\hat{Y}}^{-1} H_S)]^{-1}\mathrm{Re}(H_S^H \hat{C}_{\hat{Y}}^{-1}\hat{Y}) \tag{C.2}$$

with $H_S = H(\hat{\alpha}_S)$ and $\hat{C}_{\hat{Y}} = M^{-1}\,\mathrm{diag}(\hat{\sigma}_Y^2(1),\ \hat{\sigma}_Y^2(2),\ldots,\ \hat{\sigma}_Y^2(F))$. The asymptotic properties of $\hat{\beta}_S$ are proven by showing that the stochastic sums $H_S^H \hat{C}_{\hat{Y}}^{-1} H_S$ and $H_S^H \hat{C}_{\hat{Y}}^{-1}\hat{Y}$ in (C.2) have the same asymptotic properties as $H^H C_Y^{-1} H$ and $H^H C_Y^{-1} Y$ in (B.1). Since under Assumptions 11 and 12, the

sample mean $\hat{Y}(k)$ and sample variance $\hat{\sigma}_Y^2(k)$ are independent random variables (Anderson, 1958), it is sufficient to prove the existence of some moments of $1/\hat{\sigma}_Y^2(k)$. The strong convergence (consistency), convergence rate, and asymptotic normality of the stochastic sums in (C.2) require the existence of the moments of $1/\hat{\sigma}_Y^2(k)$ of order $r = 2$, 2 and $2+\delta$, respectively (see strong law of large numbers and central limit theorem for independent random variables). From Lemma 8, it follows that these moments exist if $M \geqslant 4$.

Appendix D. Proof of Theorem 3

Some frequently used abbreviations and notations are explained in the introduction of Appendix A.

D.1. Solution of the noiseless problem

By definition it follows that $\tilde{\alpha}_S = \tilde{\alpha}$ and $\alpha_{*S} = \alpha_*$. Since under Assumptions 11 and 12 $\hat{Y}(k)$ and $\hat{\sigma}_Y^2(k)$ are independent random variables (Anderson, 1958), it follows from (25) and (27) that the noiseless solution $\tilde{\beta}_S$ is given by

$$\tilde{\beta}_S = [\mathrm{Re}(H^H(\tilde{\alpha}_S)E\{\hat{C}_{\hat{Y}}^{-1}\}H(\tilde{\alpha}_S))]^{-1} \times \mathrm{Re}(H^H(\tilde{\alpha}_S)E\{\hat{C}_{\hat{Y}}^{-1}\}Y_0) \quad \text{(D.1)}$$

(proof: follow the lines of Section B.3). Using $\tilde{\alpha}_S = \tilde{\alpha}$ and $E\{\sigma_Y^2(k)/\hat{\sigma}_Y^2(k)\} = (M-1)/(M-2)$ (Lemma 8), it can be concluded from (D.1) and (B.3) that $\tilde{\beta}_S = \tilde{\beta}$.

D.2. Asymptotic uncertainty of $\hat{\alpha}_S$ in the absence of model errors

To prove that the uncertainty of $\hat{\alpha}_S$ and $\hat{\alpha}$ are asymptotically ($F \to \infty$) the same ($\delta_{\alpha_S} = \delta_\alpha$). It is sufficient to show that the uncertainty of $\hat{O}_{rS}$ and $\hat{O}_r$ are asymptotically the same. Since $\hat{O}_{rS} = (\hat{C}_{\hat{\mathbf{Y}}}/F)^{1/2}U_{[:,1:n_a]}$, where $U_{[:,1:n_a]}$ contains the n_a left singular vectors corresponding to the n_a largest singular values of $\hat{C}_{\hat{\mathbf{Y}}}^{-1/2}\hat{\mathbf{Y}}^{\mathrm{re}}\Pi$, the noise on $\hat{C}_{\hat{\mathbf{Y}}}^{-1/2}\hat{\mathbf{Y}}^{\mathrm{re}}\Pi\,\hat{\mathbf{Y}}^{\mathrm{re\,T}}\hat{C}_{\hat{\mathbf{Y}}}^{-\mathrm{T}/2}$ is studied first ($\Pi^2 = \Pi$). Using $\hat{\mathbf{Y}}^{\mathrm{re}} = \mathbf{Y}_0^{\mathrm{re}} + \hat{\mathbf{N}}_{\mathbf{Y}}^{\mathrm{re}}$, $\mathbf{Y}_0^{\mathrm{re}}\Pi\,\hat{\mathbf{N}}_{\mathbf{Y}}^{\mathrm{re\,T}} = O_{\mathrm{p}}(F^{1/2})$ (Lemma 6) and (A.14) we find

$$\hat{\mathbf{Y}}^{\mathrm{re}}\Pi\,\hat{\mathbf{Y}}^{\mathrm{re\,T}} = (\mathbf{Y}_0^{\mathrm{re}}\Pi\mathbf{Y}_0^{\mathrm{re\,T}} + C_{\hat{\mathbf{Y}}}) + \mathbf{N} + O_{\mathrm{p}}(F^0),$$

$$\mathbf{N} = \mathbf{Y}_0^{\mathrm{re}}\Pi\hat{\mathbf{N}}_{\mathbf{Y}}^{\mathrm{re\,T}} + \hat{\mathbf{N}}_{\mathbf{Y}}^{\mathrm{re}}\Pi\,\mathbf{Y}_0^{\mathrm{re\,T}} + (\hat{\mathbf{N}}_{\mathbf{Y}}^{\mathrm{re}}\Pi\,\hat{\mathbf{N}}_{\mathbf{Y}}^{\mathrm{re\,T}} - C_{\hat{\mathbf{Y}}}) \quad \text{(D.2)}$$

with $\mathbf{N} = O_{\mathrm{p}}(F^{1/2})$. Because $\hat{C}_{\hat{\mathbf{Y}}}^{-1/2} = C_{\hat{\mathbf{Y}}}^{-1/2} + O_{\mathrm{p}}(F^{-1})$ with $C_{\hat{\mathbf{Y}}}^{-1/2} = O(F^{-1/2})$ (see Section C.1). (D.2) can be written as

$$\hat{C}_{\hat{\mathbf{Y}}}^{-1/2}\hat{\mathbf{Y}}^{\mathrm{re}}\Pi\hat{\mathbf{Y}}^{\mathrm{re\,T}}\hat{C}_{\hat{\mathbf{Y}}}^{-T/2} = \hat{C}_{\hat{\mathbf{Y}}}^{-1/2}(\mathbf{Y}_0^{\mathrm{re}}\Pi\mathbf{Y}_0^{\mathrm{re\,T}} + C_{\hat{\mathbf{Y}}})\hat{C}_{\hat{\mathbf{Y}}}^{-\mathrm{T}/2} + \mathbf{M} + O_{\mathrm{p}}(F^{-1}),$$

$$\mathbf{M} = C_{\hat{\mathbf{Y}}}^{-1/2}\mathbf{N}C_{\hat{\mathbf{Y}}}^{-\mathrm{T}/2} \quad \text{(D.3)}$$

with $\mathbf{M} = O_{\mathrm{p}}(F^{-1/2})$. Define the singular value decompositions

$$C_{\hat{\mathbf{Y}}}^{-1/2}\mathbf{Y}_0^{\mathrm{re}}\Pi\mathbf{Y}_0^{\mathrm{re\,T}}C_{\hat{\mathbf{Y}}}^{-\mathrm{T}/2} = U_0\Sigma_0U_0^{\mathrm{T}},$$

$$\hat{C}_{\hat{\mathbf{Y}}}^{-1/2}(\mathbf{Y}_0^{\mathrm{re}}\Pi\mathbf{Y}_0^{\mathrm{re\,T}} + C_{\hat{\mathbf{Y}}})\hat{C}_{\hat{\mathbf{Y}}}^{-\mathrm{T}/2} = U_1\Sigma_1U_1^{\mathrm{T}}, \quad \text{(D.4)}$$

where $\mathrm{diag}(\Sigma_0)$, $\mathrm{diag}(\Sigma_1) \in \mathbb{R}^r$ contain the ordered singular values (from large to small). The first term on the right-hand side of (D.3) can then be written as

$$\begin{aligned}\hat{C}_{\hat{\mathbf{Y}}}^{-1/2}&(\mathbf{Y}_0^{\mathrm{re}}\Pi\mathbf{Y}_0^{\mathrm{re\,T}} + C_{\hat{\mathbf{Y}}})\hat{C}_{\hat{\mathbf{Y}}}^{-\mathrm{T}/2}\\ &= \hat{C}_{\hat{\mathbf{Y}}}^{-1/2}C_{\hat{\mathbf{Y}}}^{1/2}(C_{\hat{\mathbf{Y}}}^{-1/2}\mathbf{Y}_0^{\mathrm{re}}\Pi\mathbf{Y}_0^{\mathrm{re\,T}}C_{\hat{\mathbf{Y}}}^{-\mathrm{T}/2} + I_r)C_{\hat{\mathbf{Y}}}^{\mathrm{T}/2}\hat{C}_{\hat{\mathbf{Y}}}^{-\mathrm{T}/2}\\ &= \hat{C}_{\hat{\mathbf{Y}}}^{-1/2}C_{\hat{\mathbf{Y}}}^{1/2}(U_0(\Sigma_0 + I_r)U_0^{\mathrm{T}})C_{\hat{\mathbf{Y}}}^{\mathrm{T}/2}\hat{C}_{\hat{\mathbf{Y}}}^{-\mathrm{T}/2}. \quad \text{(D.5)}\end{aligned}$$

Under Assumption 10 (no model errors), we have $\mathrm{rank}(\Sigma_0) = n_a$ and $U_{0[:,1:n_a]} = (C_{\hat{\mathbf{Y}}}/F)^{-1/2}O_{r0}$ so that

$$\hat{C}_{\hat{\mathbf{Y}}}^{-1/2}C_{\hat{\mathbf{Y}}}^{1/2}U_{0[:,1:n_a]} = (\hat{C}_{\hat{\mathbf{Y}}}/F)^{-1/2}O_{r0}. \quad \text{(D.6)}$$

Combining (D.4)–(D.6) gives

$$\begin{aligned}(\hat{C}_{\hat{\mathbf{Y}}}/F)^{1/2}U_{1[:,1:n_a]} &= (\hat{C}_{\hat{\mathbf{Y}}}/F)^{1/2}\mathrm{range}(\hat{C}_{\hat{\mathbf{Y}}}^{-1/2}C_{\hat{\mathbf{Y}}}^{1/2}U_{0[:,1:n_a]})\\ &= O_{r0}. \quad \text{(D.7)}\end{aligned}$$

Under Assumption 6, $\mathrm{range}(\hat{C}_{\hat{\mathbf{Y}}}^{-1/2}\hat{\mathbf{Y}}^{\mathrm{re}}\Pi)$ is a continuous function of $\hat{C}_{\hat{\mathbf{Y}}}^{-1/2}\hat{\mathbf{Y}}^{\mathrm{re}}\Pi\,\hat{\mathbf{Y}}^{\mathrm{re\,T}}\hat{C}_{\hat{\mathbf{Y}}}^{-\mathrm{T}/2}$ with continuous derivative (Lemma 4). Hence, it follows from (D.3), (D.4), (D.7) and the mean value theorem that

$$\hat{O}_{rS} = O_{r0} + (\hat{C}_{\hat{\mathbf{Y}}}/F)^{1/2}g(\mathbf{M}) + o_{\mathrm{p}}(F^{-1/2}), \quad \text{(D.8)}$$

where $g(\mathbf{M}) = O_{\mathrm{p}}(F^{-1/2})$ is a linear matrix function of $\mathrm{vec}(\mathbf{M})$. Since $\hat{C}_{\hat{\mathbf{Y}}}^{1/2} = C_{\hat{\mathbf{Y}}}^{1/2} + O_{\mathrm{p}}(F^0)$ with $C_{\hat{\mathbf{Y}}}^{1/2} = O(F^{1/2})$, (D.8) becomes

$$\hat{O}_{rS} = O_{r0} + (C_{\hat{\mathbf{Y}}}/F)^{1/2}g(\mathbf{M}) + o_{\mathrm{p}}(F^{-1/2}). \quad \text{(D.9)}$$

Making similar calculations for the estimate $\hat{O}_r$ based on the data $\hat{Y}(k)$, $\sigma_Y^2(k)/M$ gives

$$\hat{O}_r = O_{r0} + (C_{\hat{\mathbf{Y}}}/F)^{1/2}g(\mathbf{M}) + o_{\mathrm{p}}(F^{-1/2}), \quad \text{(D.10)}$$

where $g(\mathbf{M})$ is exactly the same function as in (D.9). It can be concluded from (D.9) and(D.10) that $\hat{O}_{rS}$ and $\hat{O}_r$ have asymptotically the same uncertainty.

D.3. Asymptotic uncertainty of $\hat{\beta}_S$ in the absence of model errors

The estimate $\hat{\beta}_S$ (C.2) can be written as

$$\hat{\beta}_S = R_S^{-1}v_S \quad \text{(D.11)}$$

with $R_S = \mathrm{Re}(H_S^H\hat{C}_{\hat{Y}}^{-1}H_S)/F$, $v_S = \mathrm{Re}(H_S^H\hat{C}_{\hat{Y}}^{-1}\hat{Y})/F$. Under Assumptions 1–4, 6 and 10–12, R_S and v_S converge w.p. 1 and at the rate $O_{\mathrm{p}}(F^{-1/2})$ to R_{*S} and b_{*S}, respectively

$$R_S = R_{*S} + \delta_R,$$

$$v_S = v_{*S} + \delta_v \quad \text{(D.12)}$$

with $\delta_R = O_p(F^{-1/2})$, $\delta_v = O_p(F^{-1/2})$, and

$$R_{*S} = \frac{M-1}{M-2}\,\mathrm{Re}(H_0^H C_{\hat{Y}}^{-1} H_0)/F,$$

$$v_{*S} = \frac{M-1}{M-2}\,\mathrm{Re}(H_0^H C_{\hat{Y}}^{-1} Y_0)/F \tag{D.13}$$

with $H_0 = H(\alpha_0)$ (proof. use Lemma 2 with $\theta = [\alpha^T\ \mathrm{diag}^T(\hat{C}_{\hat{Y}})]^T$ and Lemma 8). Taylor series expansion of $\hat{\beta}_S$ w.r.t. R_S, v_S gives

$$\begin{aligned}\hat{\beta}_S &= (R_{*S} + \delta_R)^{-1}(v_{*S} + \delta_v)\\ &= R_{*S}^{-1} v_{*S} + R_{*S}^{-1}\delta_v - R_{*S}^{-1}\delta_R R_{*S}^{-1} v_{*S} + O_p(F^{-1}).\end{aligned} \tag{D.14}$$

Using $R_{*S}^{-1} v_{*S} = \beta_0$ and (D.12), we find

$$\begin{aligned}&R_{*S}^{-1}\delta_v - R_{*S}^{-1}\delta_R R_{*S}^{-1} v_{*S}\\ &\quad = \frac{M-2}{M-1}[\mathrm{Re}(H_0^H C_{\hat{Y}}^{-1} H_0/F)]^{-1}\\ &\qquad [\mathrm{Re}(H_S^H \hat{C}_{\hat{Y}}^{-1}(\delta_{H_S}\beta_0 + \hat{N}_Y)/F)]\end{aligned} \tag{D.15}$$

with $\delta_{H_S} = H_S - H_0$ and $\hat{N}_Y = \hat{Y} - Y_0$. Combining (D.14) and (D.15). using $\delta_{H_{S[i,j]}} = O_p(F^{-1/2})$, gives

$$\hat{\beta}_S(\hat{\alpha}_S) = \beta_0 + d_{\beta_S}(\hat{\alpha}_S) + O_p(F^{-1}),$$

$$\begin{aligned}d_{\beta_S}(\hat{\alpha}_S) = \frac{M-2}{M-1}\,&[\mathrm{Re}(H_0^H C_{\hat{Y}}^{-1} H_0)]^{-1}\\ &\times[\mathrm{Re}(H_0^H \hat{C}_{\hat{Y}}^{-1}(\delta_{H_S}\beta_0 + \hat{N}_Y))],\end{aligned} \tag{D.16}$$

where $d_{\beta_S}(\hat{\alpha}_S) = O_p(F^{-1/2})$. Following the same lines for $\hat{\beta}$ (B.1), where Y and C_Y are replaced by $\hat{Y}$ and $C_{\hat{Y}} = C_Y/M$, it is found that

$$\hat{\beta}(\hat{\alpha}) = \beta_0 + \delta_\beta(\hat{\alpha}) + O_p(F^{-1}),$$

$$\begin{aligned}\delta_\beta(\hat{\alpha}) = &[\mathrm{Re}(H_0^H C_{\hat{Y}}^{-1} H_0)]^{-1}\\ &\times[\mathrm{Re}(H_0^H C_{\hat{Y}}^{-1}(\delta_H\beta_0 + \hat{N}_Y))],\end{aligned} \tag{D.17}$$

where $\delta_H = H(\hat{\alpha}) - H_0$. If α_0 would have been known, then (D.17) reduces to

$$\hat{\beta}(\alpha_0) = \beta_0 + \delta_\beta(\alpha_0) + O_p(F^{-1}),$$

$$\delta_\beta(\alpha_0) = [\mathrm{Re}(H_0^H C_{\hat{Y}}^{-1} H_0)]^{-1}[\mathrm{Re}(H_0^H C_{\hat{Y}}^{-1}\hat{N}_Y)]. \tag{D.18}$$

Using $\delta_{H_S} = \delta_H + O_p(F^{-1})$ (proof: $\delta_{\alpha_S} = \delta_\alpha$, see Section D.2), (D.16) can be written as

$$\hat{\beta}_S(\hat{\alpha}_S) = \beta_0 + \delta_{\beta_S}(\hat{\alpha}_S) + O_p(F^{-1}),$$

$$\begin{aligned}\delta_{\beta_S}(\hat{\alpha}_S) = \frac{M-2}{M-1}&[\mathrm{Re}(H_0^H C_{\hat{Y}}^{-1} H_0)]^{-1}\\ &\times[\mathrm{Re}(H_0^H \hat{C}_{\hat{Y}}^{-1}(\delta_H\beta_0 + \hat{N}_Y))].\end{aligned} \tag{D.19}$$

In the sequel of this section, the covariance matrix of $\delta_{\beta_S}(\hat{\alpha}_S)$ is compared with that of $\delta_\beta(\hat{\alpha})$.

Taking into account (i) that the sample covariance $\hat{C}_{\hat{Y}}$ is independent of the sample mean $\hat{N}_Y$ for normally distributed errors (Anderson, 1958), (ii) that δ_H and $\hat{N}_Y$ are correlated, (iii) that $\hat{N}_Y$ is circular complex distributed ($E\{\hat{N}_Y\hat{N}_Y^T\} = 0$), and (iv) the results of Lemma 8, the covariances of $\delta_\beta(\hat{\alpha}), \delta_{\beta_S}(\hat{\alpha}_S)$ and $\delta_\beta(\alpha_0)$ are given by

$$\mathrm{Cov}(\delta_\beta(\hat{\alpha})) = P[2\,\mathrm{Re}(H_0^H \psi H_0)]P,$$

$$\begin{aligned}\mathrm{Cov}(\delta_{\beta_S}(\hat{\alpha}_S)) = &P[2\,\mathrm{Re}(H_0^H \psi H_0)]P\\ &+ \frac{1}{M-3}P[2\,\mathrm{Re}(H_0^H \psi_D H_0)]P,\end{aligned}$$

$$\mathrm{Cov}(\delta_\beta(\alpha_0)) = P \tag{D.20}$$

with $P = [2\,\mathrm{Re}(H_0^H C_{\hat{Y}}^{-1} H_0)]^{-1}$, $\psi = C_{\hat{Y}}^{-1}(C_{\hat{Y}} + Q)C_{\hat{Y}}^{-1}$, ψ_D a diagonal matrix satisfying $\psi_{D[k,k]} = \psi_{[k,k]}$, and $Q = E\{\delta_H\beta_0\beta_0^T\delta_H^H + \hat{N}_Y\beta_0^T\delta_H^H + \delta_H\beta_0\hat{N}_Y^H\}$. Using the asymptotic expression $(\delta_H)_{[i,j]} = f_{ij}(\hat{N}_Y) + O_p(F^{-1})$, with $E\{f_{ij}(\hat{N}_Y)\} = O(F^{-1/2})$ (δ_H depends on $\hat{N}_Y$ through $\hat{\alpha}$), it can easily be seen that $Q_{[i,j]} = O(F^{-1})$. Hence,

$$\begin{aligned}\psi_{D[k,k]} = \psi_{[k,k]} &= M^2\sigma_Y^{-4}(k)(\sigma_Y^2(k)/M + Q_{[k,k]})\\ &= M/\sigma_Y^2(k) + O(F^{-1})\end{aligned} \tag{D.21}$$

so that $2\,\mathrm{Re}(H_0^H\psi_D H_0) = P^{-1} + O(F^0)$, and

$$\begin{aligned}\mathrm{Cov}(\delta_{\beta_S}(\hat{\alpha}_S)) = \mathrm{Cov}(\delta_\beta(\hat{\alpha})) &+ \frac{1}{M-3}\mathrm{Cov}(\delta_\beta(\alpha_0))\\ &+ O(F^{-2})\end{aligned}$$

($P = O(F^{-1})$) which concludes the proof. □

D.4. Asymptotic covariance matrix between $\hat{\alpha}_S$ and $\hat{\beta}_S$ in the absence of model errors

The asymptotic covariance matrix $\mathrm{Cov}(\delta_{\alpha_S}, \delta_{\beta_S})$ between $\hat{\alpha}_S$ and $\hat{\beta}_S$ (D.19) is given by

$$\begin{aligned}E\{\delta_{\alpha_S}\delta_{\beta_S}^T\} &= \frac{M-2}{M-1}E\{\delta_\alpha[\mathrm{Re}((\delta_H\beta_0 + \hat{N}_Y)^H E\{\hat{C}_{\hat{Y}}^{-1}\}H_0)]\\ &\quad\times[\mathrm{Re}(H_0^H C_{\hat{Y}}^{-1} H_0)]^{-1}\}\\ &= E\{\delta_\alpha[\mathrm{Re}((\delta_H\beta_0 + \hat{N}_Y)^H C_{\hat{Y}}^{-1} H_0)]\\ &\quad\times[\mathrm{Re}(H_0^H C_{\hat{Y}}^{-1} H_0)]^{-1}\}\\ &= E\{\delta_\alpha\delta_\beta^T\}\end{aligned} \tag{D.22}$$

The first equality used $\delta_{\alpha_S} = \delta_\alpha$ (see Section D.2), and the independence of the sample covariance $\hat{C}_{\hat{Y}}$ and the sample mean $\hat{N}_Y$ for independent normally distributed errors (Anderson, 1958); the second equality $E\{\hat{\sigma}_Y^{-2}(k)\} = \sigma_Y^{-2}(k)(M-1)/(M-2)$ (Lemma 8); and the third equality (D.17).

Appendix E. Proof of the Corollary to Theorem 3

Apply the Corollary to Lemma 2 with $f_F(\theta, Z) = G(\xi, \theta)$ and $\theta = [\alpha^T \beta^T]^T$, using the results of Theorem 3. □

References

Akcay, H., Islam, S. M., & Ninness, B. (1999). Subspace-based identification of power transformer models from frequency response data. *IEEE Transactions on Instrumentation and Measurement*, *IM-48*(3), 700–704.

Anderson, T. W. (1958). *An introduction to multivariate statistical analysis*. New York: Wiley.

Billingsley, P. (1995). *Probability and measure*. New York: Wiley.

Brillinger, D. R. (1981). *Time series: Data analysis and theory*. New York: McGraw-Hill.

Golub, G. H., & Van Loan, C. F. (1996). *Matrix computations* (3rd ed.). Baltimore, MD: John Hopkins University Press.

Guillaume, P., Kollár, I., & Pintelon, R. (1996). Statistical analysis of nonparametric transfer function estimates. *IEEE Transactions on Instrumentation and Measurement*, *IM-45*(2), 594–600.

Jansson, M. (2000). Asymptotic variance analysis of subspace identification methods. *Preprints of the 12th IFAC symposium on system identification*, Santa Barbara, USA, June 21–23.

Kailath, T. (1980). *Linear systems*. Englewood Cliffs, NJ: Prentice-Hall.

Kaplan, W. (1993). *Advanced calculus*. Reading, MA: Addison-Wesley.

Lukacs, E. (1975). *Stochastic convergence*. New York: Academic Press.

McKelvey, T. (1997). Frequency domain system identification with instrumental variable based subspace algorithm. *Proceedings of the 1997 ASME design engineering technical conferences*, Sacramento, CA, September 14–17 (pp. 1–8).

McKelvey, T., Akçay, H., & Ljung, L. (1996). Subspace-based multivariable system identification from frequency response data. *IEEE Transactions on Automatic Control*, *41*(7), 960–979.

Pintelon, R., & Schoukens, J. (2001a). *System identification: A frequency domain approach*. Piscataway, NJ: IEEE Press.

Pintelon, R., & Schoukens, J. (2001b). Measurement of frequency response functions in the presence of correlated input/output errors. *IEEE Transactions on Instrumentation and Measurement*, *50*(6), 1753–1760.

Pintelon, R., Guillaume, P., Vandersteen, G., & Rolain, Y. (1998). Analyses, development and applications of TLS algorithms in frequency-domain system identification. *SIAM Journal of Matrix Analysis and Application*, *19*(4), 983–1004.

Ralston, A., & Rabinowitz, P. (1984). *A first course in numerical analysis*. New York: McGraw-Hill.

Rolain, Y., Pintelon, R., Xu, K. Q., & Vold, H. (1995). Best conditioned parametric identification of transfer function models in the frequency domain. *IEEE Transactions on Automatic Control*, *AC-40*(11), 1954–1960.

Schoukens, J., Pintelon, R., Vandersteen, G., & Guillaume, P. (1997). Frequency-domain system identification using non-parametric noise models estimated from a small number of data sets. *Automatica*, *33*(6), 1073–1086.

Stuart, A., & Ord, J. K. (1987). *Distribution theory Kendall's advanced theory of statistics*, Vol. 1. London: Charles Griffin.

Van Overschee, P., & De Moor, B. (1996). Continuous time–frequency domain subspace system identification. *Signal Processing*, *52*(2), 179–194.

Verhaegen, M. (1994). Identification of the deterministic part of MIMO state space models given in innovations form from input–output data. *Automatica*, *30*(1), 61–74.

Yang, Z. J., & Sanada, S. (2000). Frequency domain subspace identification with the aid of the w-operator. *Electrical Engineering in Japan*, *132*(1), 46–56.

Rik Pintelon was born in Gent, Belgium, on December 4, 1959. He received the degree of electrotechnical–mechanical engineer (burgerlijkingenieur) 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.