Subspace-Based Multivariable System Identification from Frequency Response Data

Tomas McKelvey, Hüseyin Akçay, and Lennart Ljung, *Fellow, IEEE*

***Abstract*—Two noniterative subspace-based algorithms which identify linear, time-invariant MIMO (multi-input/multioutput) systems from frequency response data are presented. The algorithms are related to the recent time-domain subspace identification techniques. The first algorithm uses equidistantly, in frequency, spaced data and is strongly consistent under weak noise assumptions. The second algorithm uses arbitrary frequency spacing and is strongly consistent under more restrictive noise assumptions. Promising results are obtained when the algorithms are applied to real frequency data originating from a large flexible structure.**

I. Introduction

LINEAR systems are most often characterized in the frequency domain. The properties of a closed-loop system can, for single-input, single-output (SISO) systems, very accurately and intuitively be determined by studying the frequency response function. The classical lead-lag compensator design is done entirely by shaping the Nyquist plot or the Bode plot of the open-loop system. From this perspective, it is quite natural to also consider performing system identification in the frequency domain, i.e., determining low order, linear models given samples of the frequency response. In the classical identification literature [30], [50], direct frequency domain identification has received little attention. However, recently the interest for frequency domain techniques, following the classical stochastic approach, has increased [47].

Motivated from robust control, Helmicki *et al.* [21] formulated system identification in the frequency domain, also known as $\mathcal{H}_\infty$ identification. Within this framework, a number of identification algorithms have emerged, see, e.g., [18], [39], and the references therein. Model validation in this framework has also been considered [43], [49]. Most model structures considered were of finite impulse response (FIR) type, and the objective was to obtain hard error bounds. This had several drawbacks in serious applications, especially in the identification of lightly damped systems, as illustrated clearly by the ARC testbed example in [15]. This leads to the conclusion that model structures, capable of representing finite-dimensional rational transfer functions, are needed.

Besides the need for a rational multivariable model of the system to be controlled, the modern tools for synthesis of robust controllers need accurate and automatic methods for synthesis of multivariable parametric models given frequency response data. Particularly, in the D-K iteration algorithm [14], solving the μ-synthesis problem, parametric models are identified from frequency data. The identified systems are used for dynamic scaling when determining the μ-norm.

Although all representations of a finite-dimensional linear system are equivalent, the state-space models stand out as the natural way of representing multivariable systems. Hence, most implementations of control synthesis tools use state-space representations.

Recently, identification and control of large flexible structures have received considerable attention [15], [28], [29], [6], [24], [23]. This type of system is also frequently encountered in the modal analysis area of mechanical engineering. Typically such systems are lightly damped and quite often, as in the system analysis and control design of mechanical structures, high-order models with many inputs and outputs are needed.

For structural design purposes, the finite element method provides accurate enough models. Then static and dynamic tests on the structure are performed to refine the finite element model. However, this traditional approach to model development may not be accurate enough if the intended use of the analytical model is to design a control system, since most modern multivariable control design techniques are based on state-space models of the systems. A direct method is then to realize the model from the experimental results.

If time-domain measurements are available, a great number of algorithms exist in the literature. These algorithms can be classified either as iterative or noniterative. Among the iterative algorithms, we find the prediction error methods [30], [50], and among the noniterative we find the more recent subspace based algorithms [13], [56], [54]. Noniterative methods do not involve nonlinear parametric optimization. In particular, subspace-based algorithms deliver state-space models without the need for an explicit parameterization of the model set. Essentially, there is no difference between multi-input, multi-output (MIMO) system identification and SISO system identification for a subspace-based algorithm. The algorithms also deliver estimated models in a state-space basis, wherein the transfer function is insensitive to small perturbations in the matrix elements. This leads to the ability to identify high-order systems. In [57] and [12], subspace-based algorithms were analyzed with respect to consistency, and asymptotic expressions for the quality of the estimates were derived in [57].

In practice, information about a system is often characterized in terms of the frequency response of the system

Manuscript received September 23, 1994; revised November 8, 1995. Recommended by Associate Editor, A. Vicino. This work was supported in part by the Swedish Research Council for Engineering Sciences (TFR).

T. McKelvey and L. Ljung are with the Department of Electrical Engineering, Linköping University, S-581 83 Linköping, Sweden.

H. Akçay was with the Department of Electrical Engineering, Linköping University, S-581 83 Linköping, Sweden. He is now with Tübitak, Marmara Research Center, Mathematics Division.

Publisher Item Identifier S 0018-9286(96)05343-3.

at some discrete set of frequencies. If the excitation of the system is well-designed, e.g., periodic input or stepped-sine, each transfer function measurement, compiled from a large number of time-domain measurements, is of high quality. Data originating from different experiments can easily be combined in the frequency domain. See [1], [5], [41], and [47] for a discussion on the data acquisition and the statistical properties of the frequency response data.

The problem of fitting a real-rational model to a given frequency response has been addressed by many authors, e.g., [47], [31], and [41]. In the traditional way, a system is modeled as a fraction of two polynomials with real coefficients, and a nonlinear least-squares fit to the frequency response data is sought. The solution to this nonlinear parametric optimization problem is obtained by iterative, numerical search. For certain noise models, such identification methods can be interpreted as statistical maximum-likelihood estimators and, as such, they are the frequency domain counterpart to the well-known time domain prediction error methods [30]. In an early result [26], and later refined in [46], a sequence of linear least-squares problems called SK-iterations are solved. However, SK-iterations are not guaranteed to terminate at the global minimum as indicated in [58]. A second drawback is the sensitivity of the poles and zeros of the system to polynomial factoring, if the system order is high. This deficiency can be reduced by introducing other parameterizations, e.g., orthogonal polynomials [11], [1], [44], the zero-pole-gain form, or the related RPM-parameterization [40].

On the other hand, the algorithm in [24] is noniterative and based on the famous Ho and Kalman realization algorithm [22] or Kung's smoothed version [25]. In [24], the impulse response coefficients of the system, also called Markov parameters, are estimated applying the inverse discrete Fourier transform (IDFT) to the frequency response data. The realization algorithms [22], [25] find a minimal state-space realization given a finite sequence of Markov parameters. In [24], a recursive scheme to calculate the estimates of the Markov parameters is proposed. Their approach is exact only if the impulse response dies out completely within the number of given frequency points, in other words if the system has a finite impulse response and therefore for lightly damped systems yields very poor estimates. To perform the inverse DFT, the frequency data must also be uniformly spaced. A similar approach is described in [20], where the approximate Markov parameters obtained from the IDFT are taken as the finite impulse response of a system and then, as a second step, the model reduction technique of balanced truncation [37] is applied. In [6], Bayard suggests first fitting a high-order rational model to the data using the SK-iterations and then calculating Markov parameters of the high-order model. Next, a reduced-order model in the state-space form is obtained utilizing the realization algorithm of Ho and Kalman [22].

A new frequency domain approach proposed by Liu and coworkers [28] is a frequency domain counterpart of the time-domain subspace methods by De Moor and Vandewalle [13] and Liu and Skelton [29]. Their approach does not require the data to be uniformly spaced in frequencies and also offers some frequency weighting capabilities.

In this paper we will introduce two frequency domain identification algorithms and provide stochastic analysis regarding their consistency properties, i.e., if the estimates will converge to the true transfer function as the number of data tends to infinity. The two algorithms share some common features:

- Given samples of the frequency function, minimal MIMO state-space models are delivered by the algorithms.
- A key step is the extraction of a low-dimensional subspace by the use of a truncated singular value decomposition of a noisy data matrix.
- The algorithms are noniterative.
- They are strongly consistent, i.e., when data are noisy, the estimated transfer function will converge to the correct one when the number of data tends to infinity.
- They are *correct*, i.e., any finite-dimensional rational transfer function will exactly be estimated, given a finite number of data (depending on the model order).

The first algorithm is based on the classical state-space realization algorithms [59], [25]. The algorithm combines the IDFT and a modified realization algorithm. The second algorithm presented is the frequency domain equivalence to the time-domain projection method [13], [56] and is related to the frequency domain algorithm [28]. This algorithm is applicable to data with arbitrary frequency spacing but is only strongly consistent if the noise covariance function is known.

We will now outline the contents of this paper. In Section II, we formulate the problem solved in the paper. A novel identification algorithm applicable when the frequency data are equidistant samples of the frequency function is presented in Section III. Analysis shows correctness and strong consistency for the algorithm. In Section IV, a second algorithm is presented which takes data with an arbitrary frequency spacing. Correctness is shown, and if the noise covariance function is known the algorithm is shown to be consistent. The relation between the two algorithms is discussed and an illuminating example illustrates the consistency properties of the algorithms. Some practical aspects are discussed in Section V. If data are given as samples of a continuous time transfer function, the bilinear transformation can be applied to convert the problem to an equivalent discrete-time identification problem. Section VI describes in some detail how this is done. In Section VII an identification application is considered with measured data from a flexible structure. The example clearly indicates that the subspace methods are competitive compared to the iterative nonlinear least-squares methods. Section VIII contains the conclusions.

II. Problem Formulation

Assume that G is a stable, MIMO, linear time-invariant, discrete-time system with input–output properties characterized by the impulse response coefficients g_k through the equation

$$y(t) = \sum_{k=0}^{\infty} g_k u(t-k) \tag{1}$$

where $y(t) \in \mathbb{R}^p$, $u(t) \in \mathbb{R}^m$, and $g_k \in \mathbb{R}^{p \times m}$. We also assume that the system is of finite order n and can thus be described

by a state-space model

$$x(t+1) = Ax(t) + Bu(t)$$
$$y(t) = Cx(t) + Du(t) \tag{2}$$

where $y(t) \in \mathbb{R}^p$, $u(t) \in \mathbb{R}^m$, and $x(t) \in \mathbb{R}^n$. The state-space model (2) has the impulse response

$$g_k = \begin{cases} D, & k = 0 \\ CA^{k-1}B, & k > 0. \end{cases} \tag{3}$$

The frequency response of (1) is calculated as

$$G(e^{j\omega}) = \sum_{k=0}^{\infty} g_k e^{-j\omega k}, \qquad 0 \le \omega \le \pi \tag{4}$$

which for the state-space model (2) can be written as

$$G(e^{j\omega}) = C(e^{j\omega}I - A)^{-1}B + D. \tag{5}$$

In (4), $j = \sqrt{-1}$ is the imaginary unit. If the state-space realization (2) has a minimal McMillan degree, i.e., the transfer function $G(e^{j\omega})$ cannot be described by a model with fewer variables, the extended observability matrix

$$\mathcal{O} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{q-1} \end{bmatrix} \in \mathbb{R}^{qp \times n} \tag{6}$$

and the extended controllability matrix

$$\mathcal{C} = [B \quad AB \quad \cdots \quad A^{r-1}B] \in \mathbb{R}^{n \times rm} \tag{7}$$

both have full rank n for all values $r, q \ge n$.

The problem is then as follows.

Given: Noise-corrupted M samples of the frequency response function

$$G_k = G(e^{j\omega_k}) + n_k, \qquad k = 1, \cdots, M. \tag{8}$$

Find: An identification algorithm which maps data G_k to a finite-dimensional transfer function $\hat{G}^M(e^{j\omega})$ such that with probability one (w.p. 1)

$$\lim_{M \to \infty} \|\hat{G}^M - G\|_\infty = 0 \tag{9}$$

where $\|X\|_\infty := \sup_{0 \le \omega \le \pi} \sigma_1(X(e^{j\omega}))$, and σ_1 denotes the largest singular value. Algorithms with this property are called *strongly consistent.*

We also require the algorithms to produce the true model if the noise is zero ($n_k = 0$) given a finite amount of data M, i.e., there exists some $M_0 < \infty$ such that

$$\|\hat{G}^M - G\|_\infty = 0, \qquad \text{for all } M > M_0. \tag{10}$$

Identification algorithms which satisfy (10) are called *correct.* In the next two sections we present two algorithms which have these properties. Strong consistency is a most natural requirement for any useful algorithm. As the amount of data increases, the estimate should improve and asymptotically the correct model should be obtained. In practice any algorithm will have to use a finite amount of data. Then correctness of an algorithm becomes important. This is particularly important for lightly damped systems and is further discussed in the next section.

In the problem formulation only discrete-time systems are discussed. By use of the bilinear transformation and discrete-time identification algorithms, continuous time models can be derived from samples of the frequency response of a continuous time system. We will return to this issue in Section VI. For the case when the true systems are infinite-dimensional, nice extensions have been reported in [36].

III. Uniformly Spaced Data

This section is devoted to the case of uniformly spaced data. Assume that $M + 1$ frequency response data G_k on a set of uniformly spaced frequencies

$$\omega_k = \frac{\pi k}{M}, \qquad k = 0, \cdots, M$$

are given. If the impulse response coefficients (3) are given, well-known realization algorithms can be used to obtain a state-space realization [22], [59], [25], [23]. The algorithm presented in this section is closely related to these results, but uses the coefficients of the IDFT from samples of the frequency response function.

Since G is a transfer function with a real valued impulse response (1), frequency response data on $[0, \pi]$ can be extended to $[\pi, 2\pi]$ by taking the complex conjugate of the given data G_k which forms the first step of the identification algorithm.

Algorithm 1:

1) Extend the transfer function samples to the full unit circle

$$G_{M+k} := G^*_{M-k}, \qquad k = 1, \cdots, M-1 \tag{11}$$

where $(\cdot)^*$ denotes complex conjugate.

2) Let $\hat{h}_i$ be defined by the 2M-point IDFT

$$\hat{h}_i := \frac{1}{2M} \sum_{k=0}^{2M-1} G_k e^{j2\pi ik/2M}, \qquad i = 0, \cdots, 2M-1. \tag{12}$$

3) Let the block Hankel matrix $\hat{H}$ be defined as

$$\hat{H} := \begin{bmatrix} \hat{h}_1 & \hat{h}_2 & \cdots & \hat{h}_r \\ \hat{h}_2 & \hat{h}_3 & \cdots & \hat{h}_{r+1} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{h}_q & \hat{h}_{q+1} & \cdots & \hat{h}_{q+t-1} \end{bmatrix} \in \mathbb{R}^{qp \times rm} \tag{13}$$

with number of block rows $q > n$ and block columns $r \ge n$. The dimension of $\hat{H}$ is bounded by $q + r \le 2M$.

4) Calculate the singular value decomposition (SVD) of the Hankel matrix

$$\hat{H} = \hat{U}\hat{\Sigma}\hat{V}^T.$$

5) Determine the system order n by inspecting the singular values and partition the SVD such that $\hat{\Sigma}_s$ contains the n largest singular values

$$\hat{H} = [\hat{U}_s \quad \hat{U}_o] \begin{bmatrix} \hat{\Sigma}_s & 0 \\ 0 & \hat{\Sigma}_o \end{bmatrix} \begin{bmatrix} \hat{V}^T \\ \hat{V}_o^T \end{bmatrix}. \tag{14}$$

6) Determine the system matrices $\hat{A}$ and $\hat{C}$ as

$$\hat{A} = (J_1\hat{U}_s)^{\dagger}J_2\hat{U}_s \tag{15}$$

$$\hat{C} = J_3\hat{U}_s \tag{16}$$

where

$$J_1 = [I_{(q-1)p} \quad 0_{(q-1)p\times p}] \tag{17}$$

$$J_2 = [0_{(q-1)p\times p} \quad I_{(q-1)p}] \tag{18}$$

$$J_3 = [I_p \quad 0_{p\times(q-1)p}] \tag{19}$$

and I_i denotes the $i\times i$ identify matrix, $0_{i\times j}$ the $i\times j$ zero matrix, and $X^{\dagger} = (X^TX)^{-1}X^T$ the Moore–Penrose pseudoinverse of the full column rank matrix X.

7) Solve a least-squares problem to determine $\hat{B}$ and $\hat{D}$

$$\hat{B}, \hat{D} = \arg\min_{\substack{B\in\mathbb{R}^{n\times m}\\ D\in\mathbb{R}^{p\times m}}} \sum_{k=0}^{M} \|G_k - D - \hat{C}(e^{j\omega_k}I - \hat{A})^{-1}B\|_F^2 \tag{20}$$

where $\|X\|_F = \sum_k\sum_s |x_{ks}|^2$ denotes the Frobenius norm.

8) The estimated transfer function is defined as

$$\hat{G}^M(z) = \hat{D} + \hat{C}(zI - \hat{A})^{-1}\hat{B}. \tag{21}$$

Notice that B and D appear linearly in the transfer function for fixed A and C. Hence, the optimization (20) has an analytical solution. The exact conditions when the least-squares problem in Step 7 has a unique solution are given by the following result.

Lemma 1: Let $A \in \mathbb{R}^{n\times n}$, $C \in \mathbb{R}^{p\times n}$, $M \geq n$ and define

$$\mathcal{X} = \begin{bmatrix} C(z_0I - A)^{-1} & I_p \\ C(z_1I - A)^{-1} & I_p \\ \vdots & \vdots \\ C(z_MI - A)^{-1} & I_p \end{bmatrix} \in \mathbb{R}^{(M+1)p\times(n+p)} \tag{22}$$

with distinct z_i ($z_i \neq z_j$, $i \neq j$) and z_i's do not coincide with any of the eigenvalues of A. Then

$$\operatorname{rank}\mathcal{X} = n + p \Leftrightarrow (A, C) \text{ is observable.}$$

Proof: $\mathcal{X}$ is rank deficient if and only if there exists $\begin{bmatrix} B \\ D \end{bmatrix} \neq 0$ such that

$$\mathcal{X}\begin{bmatrix} B \\ D \end{bmatrix} = 0 \Leftrightarrow D + C(z_iI - A)^{-1}B = 0, \qquad i = 0, \cdots M.$$

This can only be true if either: i) $G(z) = D + C(zI - A)^{-1}B$ has $M + 1$ zeros at z_i, or ii) $G(z) \equiv 0$ for all z. Case i) is impossible since the system is of order n and thus has at most n zeros. This implies $G(z) \equiv 0$. Recall that (A, C) is a nonobservable pair if and only if it is possible to find a pair $(B, D) \neq 0$ such that $G(z) \equiv 0$. □

The conditions in the lemma are naturally met when data are uniformly spaced in frequencies and we use (A, C) which are observable and A stable. Uniformly spaced frequencies imply distinctness, and stability of A implies that the eigenvalues of A do not coincide with the frequency argument $z_i = e^{j\omega_i}$.

A dual result which we need later is as follows.

Lemma 2: Let $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, and define

$$\mathcal{X} = [(z_1I - A)^{-1}B \quad (z_2I - A)^{-1}B \cdots (z_nI - A)^{-1}B] \tag{23}$$

distinct $z_i (z_i \neq z_j, i \neq j)$ and z_i's do not coincide with any of the eigenvalues of A. Then

$$\operatorname{rank}\mathcal{X} = n \Leftrightarrow (A, B) \text{ is controllable.}$$

Proof: The proof is along the same lines as above by assuming the existence of a row vector C such that $C\mathcal{X} = 0$ and noticing that the strictly proper system $C(zI - A)^{-1}B$ can at most have $n - 1$ zeros. □

Since we are interested in state-space realizations with real valued matrices, we will separate complex matrices into their real and imaginary parts. The following lemma shows that the full rank properties of a complex matrix are transferred to the compound matrix constructed from the real and imaginary parts. Let $\operatorname{Re}X$ and $\operatorname{Im}X$ denote the real and imaginary part of X, respectively.

Lemma 3: Let $Z \in \mathbb{C}^{n\times m}$, $n > m$. Then

$$Z \text{ has full rank} \Leftrightarrow \begin{bmatrix} \operatorname{Re}Z \\ \operatorname{Im}Z \end{bmatrix} \text{ has full rank.}$$

Proof: Let $Z = X + jY$ where X and Y are real matrices. Z has full rank and is equivalent in that there exists no complex vector $z = x + jy$, where x and y are real valued vectors such that $Zz = 0$. This is equivalent to saying that there exist no solutions x and y to the equations $Xx - Yy = 0$ and $Xy + Yx = 0$ which in turn is equivalent to saying that there exists no vector $v = [x^T\ y^T]^T$ such that

$$\begin{bmatrix} X & -Y \\ Y & X \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = Qv = 0.$$

It is the equivalent to saying that Q has full rank which implies that all columns in Q are linearly independent. □

We are now ready to verify that Algorithm 1 is correct, i.e., satisfies property (10).

Theorem 1: Let G be a stable discrete-time system of order n represented by (2), and let $G_k = G(e^{j\pi k/M})$, $k = 0, \cdots, M$. Let the transfer function $\hat{G}^M(e^{j\omega})$ be given by Algorithm 1 with $q > n$ and $r \geq n$. Then for all $M > n$

$$\|\hat{G}^M - G\|_\infty = 0.$$

Proof: Denote by $\mathcal{O}$ and $\mathcal{C}$ the extended observability (6) and controllability (7) matrices from the system realization (A, B, C, D). Let $\rho(A)$ denote the spectral radius of A, i.e., $\rho(A) := \max\{|\lambda|: \lambda \in \lambda(A)\}$, where $\lambda(A)$ denotes the set of eigenvalues of A. Since G is a stable transfer function, it can be represented by the following Taylor series:

$$G(z) = D + C(zI - A)^{-1}B = D + \sum_{k=1}^{\infty} CA^{k-1}Bz^{-k}$$

$$= \sum_{k=0}^{\infty} g_kz^{-k}, \qquad \rho(A) < |z|.$$

Notice that $\hat{h}_k$ defined by (12) for $k > 0$ can be written as

$$\hat{h}_k = \frac{1}{2M}\sum_{s=0}^{2M-1}\sum_{i=0}^{\infty} g_i e^{j2\pi s(k-i)/2M} = \sum_{i=0}^{\infty} g_{k+2iM}$$
$$= CA^{k-1}\left(\sum_{i=0}^{\infty} A^{2iM}\right)B = CA^{k-1}(I - A^{2M})^{-1}B$$

and therefore $\hat{H}$ can be factored as

$$\hat{H} = \mathcal{O}(I - A^{2M})^{-1}\mathcal{C}. \tag{24}$$

From the dimensions of the factors in (24), it is clear that $\hat{H}$ has a maximum rank of n. Furthermore, since the system is stable $\rho(A) < 1$, $(I - A^{2M})$ is always of rank n. Minimality of the system also implies that both $\mathcal{C}$ and $\mathcal{O}$ are of rank n, and hence also $\hat{H}$, if $r \geq n$ and $q > n$. In (14), then $\hat{\Sigma}_o = 0$ which means that the column range spaces of $\hat{H}$, $\mathcal{O}$, and $\hat{U}_s$ are equal. An extended observability matrix for some realization of G is then given by $\hat{\mathcal{O}} = \hat{U}_s$ since there exists a nonsingular $n \times n$ matrix S such that

$$\mathcal{O} = \hat{U}_s = \mathcal{O}S.$$

$\hat{U}_s$ is thus an extended observability matrix for a state-space realization $(\hat{A}, \hat{B}, \hat{C}, \hat{D})$ which is similar to the original realization (A, B, C, D) from (2). This proves (16). From the structure of

$$J_1\hat{U}_s = \begin{bmatrix} \hat{C} \\ \hat{C}\hat{A} \\ \vdots \\ \hat{C}\hat{A}^{q-2} \end{bmatrix}, \quad J_2\hat{U}_s = \begin{bmatrix} \hat{C}\hat{A} \\ \hat{C}\hat{A}^2 \\ \vdots \\ \hat{C}\hat{A}^{q-1} \end{bmatrix}$$

we notice that $J_1\hat{U}_s\hat{A} = J_2\hat{U}_s$ which proves (15) since J_1U_s has full rank. Hence, $\hat{A}$ and $\hat{C}$ are related to the original realization as

$$\hat{A} = S^{-1}AS, \quad \hat{C} = CS.$$

The equality between the state-space model and the frequency response can be written

$$\mathcal{X}\begin{bmatrix} B \\ D \end{bmatrix} = \hat{\mathcal{G}}$$

where $\mathcal{X}$ is given by (22) and

$$\hat{\mathcal{G}} = [G_0^T \quad G_1^T \quad \cdots \quad G_M^T]^T.$$

Since G is assumed to be stable $(\rho(A) < 1)$ and minimal, (A, C) is an observable pair and

$$\operatorname{rank}\mathcal{X} = \operatorname{rank}\begin{bmatrix} \operatorname{Re}\mathcal{X} \\ \operatorname{Im}\mathcal{X} \end{bmatrix} = n + p$$

according to Lemma 1 and Lemma 3. Hence, $\hat{B}$ and $\hat{D}$ are well-defined and, by separating the real imaginary parts, given by

$$\begin{bmatrix} \hat{B} \\ \hat{D} \end{bmatrix} = \begin{bmatrix} \operatorname{Re}\hat{\mathcal{X}} \\ \operatorname{Im}\hat{\mathcal{X}} \end{bmatrix}^{\dagger}\begin{bmatrix} \operatorname{Re}\hat{\mathcal{G}} \\ \operatorname{Im}\hat{\mathcal{G}} \end{bmatrix} = \begin{bmatrix} S^{-1} & 0 \\ 0 & I_p \end{bmatrix}\begin{bmatrix} \operatorname{Re}\mathcal{X} \\ \operatorname{Im}\mathcal{X} \end{bmatrix}^{\dagger}\begin{bmatrix} \operatorname{Re}\hat{\mathcal{G}} \\ \operatorname{Im}\hat{\mathcal{G}} \end{bmatrix}$$
$$= \begin{bmatrix} S^{-1} & 0 \\ 0 & I_p \end{bmatrix}\begin{bmatrix} B \\ D \end{bmatrix}$$

where $\hat{\mathcal{X}}$ is derived as in (22) using $\hat{A}$ and $\hat{C}$.

The estimated transfer function is thus

$$\hat{G}^M(z) = \hat{D} + \hat{C}(zI - \hat{A})^{-1}\hat{B}$$
$$= D + CS(zI - S^{-1}AS)^{-1}S^{-1}B = G(z).$$

Letting $q = n + 1$, $r = n$, $M = n + 1$, we satisfy the size condition of the Hankel matrix (13) and the condition in Lemma 1. □

This result is by no means unique, and the same result is achieved by many identification algorithms such as an ARX model fitting to frequency data in a least-squares sense which is known as Levy's method [26]. The result is merely stated to make it clear that finite-dimensional rational transfer functions are exactly recovered, given a finite number of data. This is in contrast to algorithms wherein the estimated model parameters depend *linearly* on the measured data. Examples of such approaches are least-squares optimization of finite impulse response models or other types of model structures using orthogonal basis functions. For such models, the model order has to increase to infinity to exactly capture a finite-dimensional rational transfer function. The same deficiency is also present in the frequency domain identification algorithms [20], [24]. For a finite data set, the difference between algorithms which satisfy (10) and the ones which do not becomes more pronounced as the poles of the system move toward the unit circle. Hence, when identifying systems with very little damping, it is important to use an algorithm which is correct.

From (12), notice that

$$\lim_{M\to\infty}\hat{h}_k = \int_0^{2\pi} G(e^{j2\pi\theta})e^{j2\pi k\theta}d\theta = g_k$$

where g_k from (3) is the kth impulse response coefficient of G. Thus, for noise-free data, $\hat{H}$ tends to a limit as M tends to infinity and equals to the Hankel matrix of the impulse response which is used in Kung's realization algorithm [25].

The state-space realization given by Algorithm 1 using noise-free data is balanced in the sense that the q-block row observability matrix $\mathcal{O}$ and the r-block column controllability matrix $\mathcal{C}$ satisfy

$$\hat{\mathcal{O}}^T\hat{\mathcal{O}} = I, \quad \hat{\mathcal{C}}\hat{\mathcal{C}}^T = (I - \hat{A}^{2M})\hat{\Sigma}_s^2(I - \hat{A}^{2M})^T \tag{25}$$

where $\hat{\Sigma}_s$ is given by (14). As M, q, and r jointly tend to infinity, the products of (25) will converge to the observability and controllability Gramians, respectively, and the diagonal elements of $\hat{\Sigma}_s$ will converge to the Hankel singular values of the system. These facts are important when studying the properties of the algorithm when applied to data from infinite-dimensional systems [36].

A. Consistency Analysis

If we adopt the view on the noise n_k in (8) as being stochastic variables, the estimated transfer function $\hat{G}^M(z)$ will also be a stochastic variable. The aim of this section is to show that Algorithm 1 is strongly consistent (9).

Let us assume that the noise term n_k is a zero mean complex random variable with covariance

$$E\begin{bmatrix} \operatorname{Re} n_k \\ \operatorname{Im} n_k \end{bmatrix}[\operatorname{Re} n_s^T \quad \operatorname{Im} n_s^T] = \begin{bmatrix} \frac{1}{2}R_k & 0 \\ 0 & \frac{1}{2}R_k \end{bmatrix}\delta_{ks} \tag{26}$$

and hence

$$E\{n_k n_s^H\} = R_k \delta_{ks}. \tag{27}$$

Here E denotes the expectation operator, $(\cdot)^H$ the complex conjugate transpose, and δ_{ks} is the Kronecker delta. From (26) we see that the real and imaginary parts of n_k are assumed to be independent. Furthermore, assume that the covariance function is uniformly bounded

$$R_k \leq R. \tag{28}$$

Notice that (27) implies that the noise terms for different frequencies are independent. For more information on these types of complex noise models, see [9] and [47]. The assumptions are rather weak and, for example, valid asymptotically if the frequency response is obtaincd as the empirical transfer function estimate and the time-domain noise signal is colored, see [30].

In the noisy case, the Hankel matrix is given by

$$\hat{H} = H + \Delta H \tag{29}$$

where $\hat{H}$ is given by (13). Here H denotes the noise-free Hankel matrix originating from the true system, and ΔH represents the Hankel matrix of the noise part. In general, $\hat{H} \in \mathbb{R}^{qp \times rm}$ will be of full rank $[= \min(qp, mr)]$ due to the perturbation matrix ΔH. If the largest singular value of ΔH is significantly smaller than the nth singular value of H, the n left singular vectors $\hat{U}_s$ corresponding to the n largest singular values of $\hat{H}$ will be close to the unperturbed counterparts, and the estimated system will be close to the true system. The SVD of the identification algorithm will thus have a noise threshold, and when the noise level increases over this level, the resulting estimates will not be reliable since the singular vectors in $\hat{U}_s$ might change places.

The consistency proof will be performed in two stages. First, we will show that the mapping from the data to the perturbation of the estimated A and C matrices is a locally continuous function of the Hankel matrix perturbation ΔH. Second, we show that ΔH tends to zero w.p. 1 and that the estimates of B and D converge to their correct values, respectively.

Let the SVD's of the unperturbed Hankel matrix be given as

$$H = [U_s \quad U_o] \begin{bmatrix} \Sigma_s & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_s^T \\ V_o^T \end{bmatrix} \tag{30}$$

where Σ_s has n positive singular values on the diagonal in a nonincreasing order. Since the system is assumed to have order n, H has rank n and the nth singular value of H, denoted by $\sigma_n(H)$, will be some positive number.

Lemma 4: Let G be a system of order n. Assume G_k is corrupted by noise such that $\hat{H} = H + \Delta H$ with $\|\Delta H\|_F \leq \epsilon$, and H is the corresponding noise-free Hankel matrix. Furthermore, let (A, C) be the state-space matrices of G in a realization such that $\mathcal{O} = U_s$ in (30), and let q, r, M satisfy the bounds given in Theorem 1. Let $\hat{A}$ and $\hat{C}$ be given by Algorithm 1. Then there exist constants c, $c' > 0$

$$0 < \epsilon_0 < \sigma_n(H)/4 \tag{31}$$

and a nonsingular matrix $T \in \mathbb{R}^{n \times n}$ (depending on ϵ) such that

$$\|T\hat{A}T^{-1} - A\|_F \leq c\epsilon \tag{32}$$

$$\|\hat{C}T^{-1} - C\|_F \leq c'\epsilon \tag{33}$$

for all $\epsilon \leq \epsilon_0$.

Proof: Let

$$\begin{bmatrix} U_s^T \\ U_o^T \end{bmatrix} \Delta H [V_s \quad V_o] = \begin{bmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{bmatrix} = E. \tag{34}$$

From (34) we notice that $\|E\|_F = \|\Delta H\|_F \leq \epsilon$ and also, $\|E_{ij}\|_2 \leq E_{ij}\|_F$ since the spectral norm is upper bounded by the Frobenius norm. Therefore, it is clear that

$$\begin{aligned} \delta &= \sigma_n(H) - \|E_{11}\|_2 - \|E_{22}\|_2 \\ &\geq \sigma_n(H) - \|E_{11}\|_F - \|E_{22}\|_F \\ &\geq \sigma_n(H) - 2\epsilon < \tfrac{1}{2}\sigma_n(H) > 0 \end{aligned} \tag{35}$$

by assumption (31). We also have

$$\frac{\|E_{21}\, E_{12}^T\|_F}{\delta} \leq \frac{\epsilon}{\frac{1}{2}\sigma_n(H)} \leq \frac{1}{2}.$$

Then by [17, Th. 8.3.5] there exists a matrix P satisfying

$$\|P\|_F \leq \frac{2\epsilon}{\delta} \leq \frac{4\epsilon}{\sigma_n(H)}$$

such that range$(\hat{U}_s)$ and range$(U_s + U_o P)$ are equal. Since the range spaces are equal and U_s is of full rank, there exists a unique nonsingular matrix T such that

$$\hat{U}_s = (U_s + U_o P)T \tag{36}$$

holds.

Let (A, B, C, D) be the realization of G such that $\mathcal{O} = U_s$, where U_s is given by (30). From (15) we obtain

$$\hat{A} = (J_s(U_s + U_o P)T)^{\dagger} J_2(U_s + U_o P)T$$

which can be written

$$T\hat{A}T^{-1} = (J_1(U_s + U_o P))^{\dagger} J_2(U_s + U_o P)$$

where we used the fact that $(XT)^{\dagger} = T^{-1}(X)^{\dagger}$. Since U_s is an extended observability matrix of an nth-order minimal system, $J_1 U_s$ is of full rank. The matrix $J_1(U_s + U_o P)$ will thus also be of full rank for all sufficiently small values of ϵ. From [17, Th. 5.3.1] we obtain for some constants c, $\epsilon_0 > 0$

$$\|T\hat{A}T^{-1} - A\|_F \leq c\epsilon, \qquad \forall \epsilon < \epsilon_0. \tag{37}$$

The matrix $\hat{C}$ from (16) can be expressed as

$$\hat{C} = J_3(U_s + U_o P)T$$

which directly gives a constant $c' > 0$, such that

$$\|\hat{C}T^{-1} - C\|_F \leq c'\epsilon, \qquad \forall \epsilon < \epsilon_0. \tag{38}$$

□

Lemma 4 shows that $\hat{A}$ and $\hat{C}$ (when suitably transformed) are continuous functions of H in the neighborhood of A and C. Hence, if $\Delta H \to 0$ w.p. 1, then $\hat{A} \to A$ and $\hat{C} \to C$, w.p. 1.

After establishing this continuity result we can state the main consistency theorem.

Theorem 2: Let G be a stable linear system of order n and let G_k be given by (8) where n_k satisfies the assumptions (26) and (28). Let $\hat{G}^M(z)$ denote the transfer function obtained by Algorithm 1 with $q > n$ and $r \geq n$ using $M+1$ data points. Then, w.p. 1

$$\lim_{M\to\infty} \|\hat{G}^M - G\|_\infty = 0. \tag{39}$$

Proof: Let (A, B, C, D) be the realization of G such that $\mathcal{O} = U_s$ where U_s is given by (30). From (12) and (13) we see that the elements of the Hankel matrix $\Delta H = \hat{H} - H$ are given by

$$[\Delta H]_{i,j} = \Delta h_{i+j-1} \tag{40}$$

with

$$\Delta h_k = \frac{1}{2M} \sum_{i=0}^{2M-1} n_i e^{j2\pi ki/(2M)}. \tag{41}$$

These are all sample mean values of zero mean independent random variables with a common bound on the second moments. Applying [10, Th. 5.1.2], we directly obtain

$$\lim_{M\to\infty} \Delta h_k = 0, \qquad \text{w.p. } 1$$

which implies that

$$\lim_{M\to\infty} \|\Delta H\|_F = 0, \qquad \text{w.p. } 1 \tag{42}$$

since ΔH has a fixed dimension $qp \times rm$. By applying Lemma 4, we obtain $T\hat{A}T^{-1} \to A$ and $\hat{C}T^{-1} \to C$ w.p. 1 for some nonsingular matrix T.

The least-squares estimate of B and D in Step 7 of the algorithm is given by

$$\begin{bmatrix} \hat{B} \\ \hat{D} \end{bmatrix} = \begin{bmatrix} \operatorname{Re}\hat{\mathcal{X}} \\ \operatorname{Im}\hat{\mathcal{X}} \end{bmatrix}^\dagger \begin{bmatrix} \operatorname{Re}\hat{\mathcal{G}} \\ \operatorname{Im}\hat{\mathcal{G}} \end{bmatrix}$$

where

$$\hat{\mathcal{X}} = \begin{bmatrix} \hat{C}(z_0 I - \hat{A}) & I_p \\ \hat{C}(z_1 I - \hat{A}) & I_p \\ \vdots & \vdots \\ \hat{C}(z_M I - \hat{A}) & I_p \end{bmatrix}, \quad \hat{\mathcal{G}} = \begin{bmatrix} G_0 \\ G_1 \\ \vdots \\ G_M \end{bmatrix}.$$

To simplify notation let us write

$$\begin{aligned} \hat{E} &= \begin{bmatrix} T & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \hat{B} \\ \hat{D} \end{bmatrix} \\ &= \left(\begin{bmatrix} \operatorname{Re}\hat{\mathcal{X}} \\ \operatorname{Im}\hat{\mathcal{X}} \end{bmatrix} \begin{bmatrix} T^{-1} & 0 \\ 0 & I \end{bmatrix} \right)^\dagger \begin{bmatrix} \operatorname{Re}\hat{\mathcal{G}} \\ \operatorname{Im}\hat{\mathcal{G}} \end{bmatrix} = \hat{\boldsymbol{X}}^\dagger \hat{\boldsymbol{G}} \end{aligned}$$

and let $E = [B^T C^T]^T$, $\boldsymbol{X}$, and $\boldsymbol{G}$ denote the noise-free counterparts. The estimation error is then given by

$$\begin{aligned} \|\hat{E} - E\|_F &= \|\hat{\boldsymbol{X}}^\dagger \hat{\boldsymbol{G}} - \boldsymbol{X}^\dagger \boldsymbol{G}\|_F \\ &\leq \|(\hat{\boldsymbol{X}}^\dagger - \boldsymbol{X}^\dagger)\boldsymbol{G}\|_F + \|\hat{\boldsymbol{X}}^\dagger(\hat{\boldsymbol{G}} - G)\|_F. \end{aligned}$$

Using the result of Lemma 4 and simple calculations reveals that $\|\hat{\boldsymbol{X}}^\dagger - X^\dagger\|_F = O(M^{-1/2}\epsilon)$ for some M large enough. The second term can be expressed as

$$\hat{\boldsymbol{X}}^\dagger(\hat{\boldsymbol{G}} - \boldsymbol{G}) = \frac{1}{M+1} \sum_{k=0}^{M} c_R(k) \operatorname{Re} n_k + c_I(k) \operatorname{Im} n_k$$

for some bounded matrices $c_R(k)$ and $c_I(k)$ since all elements in $\hat{\boldsymbol{X}}^\dagger$ are bounded and the $2(M+1)$ blocks in $\hat{\boldsymbol{G}} - \boldsymbol{G}$ are $\operatorname{Re} n_k$ and $\operatorname{Im} n_k$, $k = 0, \cdots, M-1$. The sum converges to zero w.p. 1 as $M \to \infty$ since each term is independent, of zero mean, and with bounded variance [10]. Thus, using (42) and Lemma 4, we obtain, w.p. 1

$$\|\hat{E} - E\|_F = \left\| \begin{bmatrix} T & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \hat{B} \\ \hat{D} \end{bmatrix} - \begin{bmatrix} B \\ D \end{bmatrix} \right\|_F \to 0$$

$$\|T\hat{A}T^{-1} - A\|_F \to 0$$

$$\|\hat{C}T^{-1} - C\|_F \to 0$$

as $M \to \infty$ and the result follows. □

B. Alternative Ways of Calculating A

Step 6 of Algorithm 1 (15) is based on the relation

$$J_1 \hat{U}_s \hat{A} = J_2 \hat{U}_s$$

which exactly holds in the noise-free case. Recall the block shift structure of the extended observability matrix (6). In the noisy case, the expression (15) is the solution to

$$\hat{A} = \arg \min_{A \in \mathbb{R}^{n\times n}} \|J_1 \hat{U}_s A - J_2 \hat{U}_s\|_F.$$

Implicit in this solution is the noise model

$$J_1 \hat{U}_s \hat{A} = J_s \hat{U}_s + N$$

and $\hat{A}$ minimizes the Frobenius norm of N. This noise model is, however, not consistent with the original equation since both $J_1\hat{U}_s$ and $J_2\hat{U}_s$ contain errors. With this more correct view we obtain the error model

$$(J_1 \hat{U}_s + N_1)\hat{A} = J_2 \hat{U}_s + N_2$$

and the total least-squares (TLS) method can be applied [17]

$$\hat{A} = \arg \min_{\substack{A \in \mathbb{R}^{n\times n} \\ (J_1\hat{U}_s + N_1)A = J_2\hat{U}_s + N_2}} \|[N_1 \quad N_2]\|_F. \tag{43}$$

The TLS solution can be obtained by an SVD [17]. The TLS technique for calculating A is also found in the signal processing algorithm ESPRIT [45]. In [53] an overview of these methods is given. Practical experience shows similar performance for the least-squares solutions (15) and the total least squares (43) when applied to noisy data.

By applying the array signal processing technique of weighted subspace fitting, the poles of the system, or equivalently the eigenvalues of the matrix A, can be optimally calculated given the statistical properties of the constructed extended observability matrix. In [38] and [52] these ideas are

exploited in a system identification context. A disadvantage with weighted subspace fitting is the introduction of a nonlinear optimization step which has to be solved by iterative search.

C. Summary

In this section we have analyzed the noniterative frequency domain state-space identification algorithm described by Algorithm 1. If the frequency data are noise free and generated by an nth-order system, we have shown that only $n+2$ equidistant frequency samples are required to exactly recover the true system. Asymptotic stochastic analysis shows that the algorithm is consistent if each measurement is perturbed by an independent stochastic noise term with bounded covariance function.

IV. Nonuniformly Spaced Data

In this section we will discuss an algorithm which is applicable for the case when samples of the frequency response

$$G_k = G(e^{j\omega_k}) + n_k, \qquad k = 1, \cdots, M$$

are given at arbitrary, distinct frequencies. The algorithm can be interpreted as a direct frequency domain formulation of the time-domain subspace algorithm [56] and has some connections with a frequency domain algorithm presented in [28]. The contribution of this section is to present how to incorporate covariance information into the algorithm and to perform stochastic analysis revealing the consistency properties. The resulting algorithm is consistent for a much larger class of noise sources in comparison with the algorithm in [28]. The algorithm and analysis are an extension of the results presented in [34] and [35]. A similar algorithm for the case of data from a continuous time system has also recently appeared [55].

We will first outline the algorithm and discuss why the noise covariance information has to be known *a priori* and how to incorporate it to yield consistent estimates. An algorithm will then be presented which uses the QR-factorization as an efficient implementation. Two theorems which summarize the discussion will be presented, and the relation between Algorithm 1, the algorithm in this section, and the algorithm [28] will be shown. The section is concluded by presenting an example illustrating the consistency properties of the algorithms.

A. The Algorithm

Consider the discrete Fourier transform (DFT) of the state-space equations (2)

$$e^{j\omega}X(\omega) = AX(\omega) + BU(\omega)$$

$$Y(\omega) = CX(\omega) + DU(\omega)$$

where $X(\omega)$, $U(\omega)$, and $Y(\omega)$ denote the transformed time-domain signals. If we denote by $X^i(\omega)$ the resulting state transform when $U(\omega) = e_i$, the unit vector with one in the ith position, and define the compound state matrix as

$$X^C(\omega) = [X^1(\omega) \quad X^2(\omega) \quad \cdots \quad X^m(\omega)] \in \mathbb{C}^{n\times m}$$

we can describe the transfer function in the state-space form

$$\begin{aligned} e^{j\omega}X^C(\omega) &= AX^C(\omega) + B \\ G(e^{j\omega}) &= CX^C(\omega) + D. \end{aligned} \tag{44}$$

By recursive use of (44) and (6) we obtain the relation

$$\begin{bmatrix} G(e^{j\omega}) \\ e^{j\omega}G(e^{j\omega}) \\ \vdots \\ e^{j(q-1)\omega}G(e^{j\omega}) \end{bmatrix} = \mathcal{O}X^C(\omega) + \Gamma \begin{bmatrix} I_m \\ e^{j\omega}I_m \\ \vdots \\ e^{j(q-1)\omega}I_m \end{bmatrix} \tag{45}$$

where Γ is the lower triangular block Toeplitz matrix with the impulse response defined as

$$\Gamma = \begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ CA^{q-2}B & CA^{q-3} & \cdots & D \end{bmatrix} \in \mathbb{R}^{pq\times mq}. \tag{46}$$

Let us form a matrix from the frequency response samples G_i as shown in (47) at the bottom of the page, and conformably, let $\boldsymbol{N}$ denote a matrix with the same structure as $\boldsymbol{G}$ but with n_i inserted instead of G_i. Furthermore, define the block Vandermonde matrix derived from the frequencies as (48) shown at the bottom of the page, where the subscript m indicates the size of the blocks. Finally, denote the ordered collection of all compound states as

$$\boldsymbol{X}^C = \frac{1}{\sqrt{M}}[X^C(\omega_1) \quad X^C(\omega_2) \quad \cdots \quad X^C(\omega_M)] \in \mathbb{C}^{n\times mM}. \tag{49}$$

$$\boldsymbol{G} = \frac{1}{\sqrt{M}} \begin{bmatrix} G_1 & G_2 & \cdots & G_m \\ e^{j\omega_1}G_1 & e^{j\omega_2}G_2 & \cdots & e^{j\omega_M}G_m \\ e^{j2\omega_1}G_1 & e^{j2\omega_2}G_2 & \cdots & e^{j2\omega_M}G_M \\ \vdots & \vdots & \ddots & \vdots \\ e^{j(q-1)\omega_1}G_1 & e^{j(q-1)\omega_2}G_2 & \cdots & e^{j(q-1)\omega_M}G_M \end{bmatrix} \in \mathbb{C}^{qp\times mM} \tag{47}$$

$$\boldsymbol{W}_m = \frac{1}{\sqrt{M}} \begin{bmatrix} I_m & I_m & \cdots & I_m \\ e^{j\omega_1}I_m & e^{j\omega_2}I_m & \cdots & e^{j\omega_M}I_m \\ e^{j2\omega_1}I_m & e^{j2\omega_2}I_m & \cdots & e^{j2\omega_M}I_m \\ \vdots & \vdots & \ddots & \vdots \\ e^{j(q-1)\omega_1}I_m & e^{j(q-1)\omega_2}I_m & \cdots & e^{j(q-1)\omega_M}I_m \end{bmatrix} \in \mathbb{C}^{qm\times mM} \tag{48}$$

From (45)–(49) we obtain the matrix equation

$$\boldsymbol{G} = \mathcal{O}\boldsymbol{X}^C + \Gamma \boldsymbol{W}_m + \boldsymbol{N}. \tag{50}$$

Since we assume the system to be minimal, (A, B) is a controllable pair and Lemma 2 shows that $\boldsymbol{X}^C$ has full rank n. The matrix product $\mathcal{O}\boldsymbol{X}^C$ is thus of rank n and has a range space equal to $\mathcal{O}$.

When the number of samples increases, the number of columns in $\boldsymbol{G}$ will increase. Since $\mathcal{O}$ is a real matrix, we are interested in the real range space and can convert (50) into a relation involving only real valued matrices

$$\underbrace{[\operatorname{Re}\boldsymbol{G} \quad \operatorname{Im}\boldsymbol{G}]}_{\mathcal{G}} = \mathcal{O}\underbrace{[\operatorname{Re}\boldsymbol{X}^C \quad \operatorname{Im}\boldsymbol{X}^C]}_{\mathcal{X}} + \Gamma\underbrace{[\operatorname{Re}\boldsymbol{W}_m \quad \operatorname{Im}\boldsymbol{W}_m]}_{\mathcal{W}} + \underbrace{[\operatorname{Re}\boldsymbol{N} \quad \operatorname{Im}\boldsymbol{N}]}_{\mathcal{N}}. \tag{51}$$

When $\boldsymbol{W}_m$ and $\boldsymbol{X}^C$ are matrices of full rank, the transposed version of Lemma 3 shows that $\mathcal{W}$ and $\mathcal{X}$ also have full rank.

Let us first consider the noise-free case which implies that $\mathcal{N} = 0$. If the term $\Gamma\mathcal{W}$ in (51) also were zero, the range space of $\mathcal{G}$ would equal the range space of $\mathcal{O}$ if $q > n$ and $M \geq n$ and a straightforward factorization as in Algorithm 1 would yield an extended observability matrix similar to $\mathcal{O}$. However, if the number of columns is sufficiently large, there exist matrices, when multiplied from right in (50), which cancel the term $\Gamma\mathcal{W}$ but retain the important information. One such matrix is the projection onto the nullspace of $\mathcal{W}$ and is given by

$$\mathcal{W}^\perp = I - \mathcal{W}^T(\mathcal{W}\mathcal{W}^T)^{-1}\mathcal{W} \tag{52}$$

and we obtain

$$\mathcal{G}\mathcal{W}^\perp = \mathcal{O}\mathcal{X}\mathcal{W}^\perp.$$

The range space of $\mathcal{G}\mathcal{W}^\perp$ equals the range space of $\mathcal{O}$ unless rank cancellations occur. A sufficient condition for the range spaces to be equal is that the intersection between the row spaces of $\mathcal{W}$ and $\mathcal{X}$ is empty. In most analyses of time-domain subspace methods this is stated as an assumption. However, here we can present sufficient conditions in terms of the data and the system.

Lemma 5: Let $M \geq q + n$, $\boldsymbol{W}_n$ and $\boldsymbol{X}^C$ be given by (48) and (49) with distinct frequencies ω_i such that $e^{j\omega_i} \notin \lambda(A)$. Then

$$\operatorname{rank}\begin{bmatrix}\boldsymbol{W}_m \\ \boldsymbol{X}^C\end{bmatrix} = qm + n$$

$$\Leftrightarrow$$

$$(A, B) \text{ is a controllable pair.}$$

Proof: Equation (44) implies that $X^C(\omega) = (e^{j\omega}I - A)^{-1}B$. $\begin{bmatrix}\boldsymbol{W}_m \\ \boldsymbol{X}^C\end{bmatrix}$ is rank deficient if and only if there exists a row vector

$$[D \quad E_1 \quad \cdots \quad E_{q-1} \quad C] \neq 0$$

such that

$$[D \quad E_1 \quad \cdots \quad E_{q-1} \quad C]\begin{bmatrix}\boldsymbol{W}_m \\ \boldsymbol{X}^C\end{bmatrix} = 0$$

$$\Leftrightarrow$$

$$H(e^{j\omega_k}) = 0, \quad k = 1, \cdots, M$$

where

$$H(z) = D + C(zI - A)^{-1}B + \sum_{k=1}^{q-1} E_k z^k = 0.$$

This can only be true if either: 1) Each element in the rational vector $H(z)$ has M zeros at z_k, or ii) $H(z) = 0$ for all z. Case i) is impossible since each element of $H(z)$ has at most $n + q - 1$ zeros. This implies case ii), and consequently $C(zI - A)^{-1}B \equiv 0$. Recall that (A, B) is a noncontrollable pair if and only if it is possible to find a vector $C \neq 0$ such that $C(zI - A)^{-1}B \equiv 0$. □

If the frequencies are distinct, the number of data satisfies $M \geq q+n$, and (A, B) is controllable, the row spaces of $\boldsymbol{W}_m$ and $\boldsymbol{X}^C$ do not intersect. By applying Lemma 3 we conclude that the two row spaces of $\mathcal{W}$ and $\mathcal{X}$ do not intersect and the range space of $\mathcal{G}\mathcal{W}^\perp$ coincides with the range space of $\mathcal{O}$. By using the singular value factorization of $\mathcal{G}\mathcal{W}^\perp$ and proceeding according to the Steps 5–7 in Algorithm 1, we obtain a state-space model which is similar to (A, B, C, D).

Let us now return to the normal case when $\mathcal{N} \neq 0$. After the projection we obtain the relation

$$\mathcal{G}\mathcal{W}^\perp = \mathcal{O}\mathcal{X}\mathcal{W}^\perp + \mathcal{N}\mathcal{W}^\perp. \tag{53}$$

In the generic case $\mathcal{G}\mathcal{W}^\perp$ will be of full rank $qp > n$, and some type of approximation is necessary to obtain a good estimate of the observability range space. By using the singular value decomposition of $\mathcal{G}\mathcal{W}^\perp$, the n left singular vectors, corresponding to the n largest singular values, form a strongly consistent estimate of the range space of $\mathcal{O}$, if two conditions hold (w.p. 1) [12]

i)

$$\lim_{M\to\infty} \underbrace{\mathcal{O}\mathcal{X}\mathcal{W}^\perp(\mathcal{N}\mathcal{W}^\perp)^T}_{S(M)} = 0 \tag{54}$$

ii)

$$\lim_{M\to\infty} \underbrace{\mathcal{N}\mathcal{W}^\perp(\mathcal{N}\mathcal{W}^\perp)^T}_{P(M)} = \alpha I \tag{55}$$

for some scalar α.

Let us first investigate condition i). First note that $\mathcal{W}^\perp(\mathcal{W}^\perp)^T = \mathcal{W}^\perp$. Since $\mathcal{W}$ has full rank, all elements in $\mathcal{W}^\perp$ will be bounded and we can express the elements of $S(M)$ in (54) by

$$[S(M)]_{ij} = \frac{1}{M}\sum_{k=1}^{M} c_{ij}^R(k)\operatorname{Re} n_k + c_{ij}^I(k)\operatorname{Im} n_k$$

for some bounded real constants $c_{ij}^R(k)$, $c_{ij}^I(k)$. From the zero mean assumption and the boundedness of the moments of n_k and the limit result [10, Th. 5.1.2], we conclude that

$$\lim_{M\to\infty}[S(M)]_{ij} = 0, \qquad \text{w.p. } 1, \qquad \forall i, j.$$

By using the explicit expression of (52), $P(M)$ in (55) naturally divides into two terms

$$P(M) = \mathcal{N}\mathcal{N}^T + \mathcal{N}\mathcal{W}^T(\mathcal{W}\mathcal{W}^T)^{-1}\mathcal{W}\mathcal{N}^T. \tag{56}$$

By the assumption of distinct frequencies, $\mathcal{W}\mathcal{W}^T > cI$ for some constant c, and hence the elements of $(\mathcal{W}\mathcal{W}^T)^{-1}$ are bounded. The block elements of the second matrix in (56) are of the form

$$\begin{aligned}&[\mathcal{N}\mathcal{W}^T(\mathcal{W}\mathcal{W}^T)^{-1}\mathcal{W}\mathcal{N}^T]_{ij}\\ &= \frac{1}{M^2}\sum_{k=1}^{M}\sum_{l=1}^{M} c_{ij}^R(k,l)\operatorname{Re} n_k \operatorname{Re} n_l^T + c_{ij}^I(k,l)\operatorname{Im} n_k \operatorname{Im} n_l^T\\ &\quad + c_{ij}^{RI}(k,l)\operatorname{Re} n_k \operatorname{Im} n_l^T + c_{ij}^{IR}(k,l)\operatorname{Im} n_k \operatorname{Re} n_l^T\end{aligned}$$

for some constants $c_{ij}^R(k,l)$, $c_{ij}^I(k,l)$, $c_{ij}^{RI}(k,l)$, $c_{ij}^{IR}(k,l)$ which are bounded since $\mathcal{W}$ has full rank and bounded elements. The second moment of these block elements are bounded as

$$E\left\|\left[\frac{1}{M}\mathcal{N}\mathcal{W}^T(\mathcal{W}\mathcal{W}^T)^{-1}\mathcal{W}^T\mathcal{N}^T\right]_{ij}\right\|_F^2 \le \frac{c_3}{M^3} + \frac{c_4}{M^2}$$

for some bounded constants c_i, if we add the assumption that n_k have bounded fourth-order moments. By applying Chebyshev's inequality and the Borel–Cantelli lemma [10], we conclude that this term also converges to zero w.p. 1.

By recognizing that $\boldsymbol{N} = \boldsymbol{W}_p \operatorname{diag}(n_1, \cdots, n_M)$, the first term in (56) is simply

$$\begin{aligned}\mathcal{N}\mathcal{N}^T &= [\operatorname{Re}\boldsymbol{N} \quad \operatorname{Im}\boldsymbol{N}][\operatorname{Re}\boldsymbol{N} \quad \operatorname{Im}\boldsymbol{N}]^T\\ &= \operatorname{Re}(\boldsymbol{W}_p \operatorname{diag}(n_1 n_1^H, \cdots, n_m n_M^H)\boldsymbol{W}_p^H)\end{aligned} \tag{57}$$

with the expected value

$$E\mathcal{N}\mathcal{N}^T = \operatorname{Re}(\boldsymbol{W}_p \operatorname{diag}(R_1, \cdots, R_m)\boldsymbol{W}_p^H).$$

We will now establish that $\mathcal{N}\mathcal{N}^T$ converges to its expected value w.p. 1. Consider

$$\mathcal{N}\mathcal{N}^T - E\mathcal{N}\mathcal{N}^T \tag{58}$$

which is zero mean. Let

$$\epsilon_k = n_k n_k^H - R_k.$$

Then ϵ_k is zero mean and has bounded second moments, since n_k has bounded fourth moments. It is straightforward to see that the (i, j)th block element in the matrix (58) is given by

$$[\mathcal{N}\mathcal{N}^H - E\mathcal{N}\mathcal{N}^T]_{ij} = \frac{1}{M}\sum_{k=1}^{M} c_{i,j}(k)\epsilon_k \tag{59}$$

for some constants $c_{i,j}(k)$ which are bounded since all elements in $\boldsymbol{W}_p$ are bounded. The expression (59) is a sum of zero mean random variables with bounded second moments and we conclude as before that $\mathcal{N}\mathcal{N}^T \to E\mathcal{N}\mathcal{N}^T$ w.p. 1 as $M \to \infty$. We have established that $P(M)$ converges to its mean value as M tends to infinity. Sufficient conditions which guarantee that

$$E\,P(M) \to \alpha I$$

are that $R_k = \alpha I$ for all k and $\omega_k = 2\pi(k-1)/M$, $k = 1, \cdots, M$ [35]. However, if the covariance function is known, we can weigh the row space of $\mathcal{G}\mathcal{W}^\perp$ with a matrix $\boldsymbol{K} \in \mathbb{R}^{qp\times qp}$ satisfying

$$\boldsymbol{K}\boldsymbol{K}^T = \alpha \operatorname{Re}(\boldsymbol{W}_p \operatorname{diag}(R_1, R_2, \cdots, R_M)\boldsymbol{W}_p^H) \tag{60}$$

for some $\alpha > 0$. The matrix $\boldsymbol{K}$ can be found by a Cholesky factorization given the covariance data R_k. The weighted version of (53) then becomes

$$\boldsymbol{K}^{-1}\mathcal{G}\mathcal{W}^\perp = \boldsymbol{K}^{-1}\mathcal{O}\mathcal{X}\mathcal{W}^\perp + \boldsymbol{K}^{-1}\mathcal{N}\mathcal{W}^\perp. \tag{61}$$

This weighted version satisfies the requirement (55) since w.p. 1

$$\boldsymbol{K}^{-1}\mathcal{N}\mathcal{W}^\perp(\boldsymbol{K}^{-1}\mathcal{N}\mathcal{W}^\perp)^T \to \alpha^{-1}I$$

as $M \to \infty$. The n left singular vectors $\hat{U}_s$ corresponding to the n largest singular values of $\boldsymbol{K}^{-1}\mathcal{G}\mathcal{W}^\perp$ will form a strongly consistent estimate of $\boldsymbol{K}^{-1}\mathcal{O}$. The observability range space is then simply recovered by $\boldsymbol{K}\hat{U}_s$. By using the same continuity result as employed in the proof of Theorem 2, we conclude that the outlined algorithm is strongly consistent.

A numerically efficient way of forming $\mathcal{G}\mathcal{W}^\perp$ is to use the QR-factorization

$$\begin{bmatrix}\mathcal{W}\\ \mathcal{G}\end{bmatrix} = \begin{bmatrix}\boldsymbol{R}_{11} & 0\\ \boldsymbol{R}_{21} & \boldsymbol{R}_{22}\end{bmatrix}\begin{bmatrix}\boldsymbol{Q}_1^T\\ \boldsymbol{Q}_2^T\end{bmatrix}.$$

A simple derivation yields

$$\mathcal{G}\mathcal{W}^\perp = \boldsymbol{R}_{22}\boldsymbol{Q}_2^T$$

and it suffices to use $\boldsymbol{R}_{22}$, since $\boldsymbol{Q}_2^T$ is a matrix of full rank. Let us summarize the outlined method into the following algorithm.

Algorithm 2:

1) Given data G_k, ω_k, and R_k, form the matrices $\boldsymbol{G}$, $\boldsymbol{W}_m$, and $\boldsymbol{K}$.
2) Calculate the QR-factorization

$$\begin{bmatrix}\operatorname{Re}\boldsymbol{W}_m & \operatorname{Im}\boldsymbol{W}_m\\ \operatorname{Re}\boldsymbol{G} & \operatorname{Im}\boldsymbol{G}\end{bmatrix} = \begin{bmatrix}\boldsymbol{R}_{11} & 0\\ \boldsymbol{R}_{21} & \boldsymbol{R}_{22}\end{bmatrix}\begin{bmatrix}\boldsymbol{Q}_1^T\\ \boldsymbol{Q}_2^T\end{bmatrix}. \tag{62}$$

3) Calculate the SVD

$$\boldsymbol{K}^{-1}\boldsymbol{R}_{22} = \hat{U}\hat{\Sigma}\hat{V}^T. \tag{63}$$

4) Determine the system order n by inspecting the singular values and partition the SVD such that $\hat{\Sigma}_s$ contains the n largest singular values

$$\boldsymbol{K}^{-1}\boldsymbol{R}_{22} = [\hat{U}_s \quad \hat{U}_o]\begin{bmatrix}\hat{\Sigma}_s & 0\\ 0 & \hat{\Sigma}_o\end{bmatrix}\begin{bmatrix}\hat{V}_s^T\\ \hat{V}_o^T\end{bmatrix}. \tag{64}$$

5) Determine the system matrices $\hat{A}$ and $\hat{C}$ as

$$\hat{A} = (J_1 \boldsymbol{K}\hat{U}_s)^{\dagger} J_2 \boldsymbol{K}\hat{U}_s \tag{65}$$

$$\hat{C} = J_3 \boldsymbol{K}\hat{U}_s \tag{66}$$

where J_i are defined by (17)–(19).

6) Solve a least-squares problem to determine $\hat{B}$ and $\hat{D}$

$$\hat{B}, \hat{D} = \arg \min_{\substack{B \in \mathbb{R}^{n\times m} \\ D \in \mathbb{R}^{p\times m}}} \sum_{k=1}^{M} \| R_k^{-1/2} \cdot (G_k - D - \hat{C}(e^{g\omega_k} I - \hat{A})^{-1} B) \|_F^2. \tag{67}$$

7) The estimated transfer function is defined as

$$\hat{G}^M(z) = \hat{D} + \hat{C}(zI - \hat{A})^{-1}\hat{B}. \tag{68}$$

In the least-square solution in Step 6, we weigh with the inverse of the covariance factor to make optimal use of the covariance information. If $\hat{A}$ and $\hat{C}$ had not been estimated from data, (67) would be the minimum variance estimate or the best linear unbiased estimate (BLUE) of B and D. Of course, the same weighted least-squares step could also be used in Algorithm 1.

Let us summarize the discussion on the noise-free case into a theorem on correctness.

Theorem 3: Let G be a stable system of order n and G_k, $k = 1, \cdots, M$ be noise-free samples of the transfer function $G(e^{j\omega})$ at M distinct frequencies ω_k. Furthermore, let $q > n$, $M_0 \geq n + q$, and $\boldsymbol{K} \in \mathbb{R}^{qp\times qp}$ be any nonsingular matrix. Finally, let $\hat{G}^M(z)$ be given by Algorithm 2. Then

$$\|\hat{G}^M - G\|_\infty = 0$$

for all $M \geq M_0$.

The consistency properties in the preceeding discussion are summarized in the following result.

Theorem 4: Let G be a stable system of order n. Let $G_k = G(e^{j\omega_k}) + n_k$, $k = 1, \cdots, M$ be noisy samples of the transfer function at M distinct frequencies ω_k. Let n_k satisfy (26)–(28) and have bounded fourth-order moments. Furthermore let $q > n$, $\boldsymbol{K}$ be given by (60), and $\hat{G}^M(z)$ be given by Algorithm 2. Then, w.p. 1

$$\lim_{M\to\infty} |\hat{G}^M - G\|_\infty = 0.$$

The possibility to use nonequidistantly spaced samples in the algorithm has a price. For the algorithm to be consistent, the relative variance at each frequency has to be known *a priori*. This knowledge is necessary to derive the weighting matrix $\boldsymbol{K}$.

B. Relation Between the Algorithms

This section will show how the two presented algorithms, Algorithm 1 and Algorithm 2, are related to each other. To make any comparison meaningful, let us consider the case with data sampled equidistantly in frequency covering the full unit circle $\omega_k = 2\pi k/M$, $k = 0, \cdots, M-1$. Also let us assume the data are noise free, and we let $\boldsymbol{K} = I$.

In Algorithm 2 the observability range space is determined from

$$\boldsymbol{G}\boldsymbol{W}^{\perp} \in \mathbb{C}^{qp\times mM}$$

which is a matrix where the number of block columns grows as the number of data increases. In the case with equidistant frequencies, it is easy to derive another matrix $\tilde{\boldsymbol{W}}^{\perp}$ such that $\boldsymbol{W}\tilde{\boldsymbol{W}}^{\perp} = 0$. It is easy to verify that the matrix (69), shown at the bottom of the page, has this desired property. Notice that $\boldsymbol{G}\tilde{\boldsymbol{W}}^{\perp}$ is a matrix with fixed dimensions $qp \times rm$, independent of M, the number of data. Furthermore, the (i_1, i_2)th block entry of $\boldsymbol{G}\tilde{\boldsymbol{W}}^{\perp}$ is given by

$$\begin{aligned}[\boldsymbol{G}\tilde{\boldsymbol{W}}^{\perp}]_{i_1, i_2} &= \frac{1}{M}\sum_{k=0}^{M-1} G(\omega_k) e^{j2\pi k(i_2+i_1-1)/M} \\ &= \frac{1}{M}\sum_{k=0}^{M-1} G(e^{jk2\pi/M}) e^{j2\pi k(i_2+i_1-1)/M} \\ &= \hat{h}_{i_2+i_1-1}, \qquad i_1 = 1, \cdots, q,\ i_2 = 1, \cdots, r\end{aligned}$$

which we recognize from (12) as the M-point inverse DFT of the full frequency response data, and we directly see that

$$\hat{H} = \boldsymbol{G}\tilde{\boldsymbol{W}}^{\perp}$$

where $\hat{H}$ is defined by (13). Algorithm 2 with this particular choice of annihilator matrix (69) is thus equal to Algorithm 1. The key difference between Algorithm 1 and Algorithm 2 is what type of matrix is used to annihilate the influence of the term $\Gamma \boldsymbol{W}_m$. In the equidistantly spaced case, we use a matrix of size $mM \times mr$ in Algorithm 1, while Algorithm 2 uses the maximal rank annihilator which is of size $mM \times mM$. In the nonuniformly spaced case we cannot *a priori* derive a smaller matrix since then there is a risk that we not only annihilate $\boldsymbol{W}_m$ but also cancel some of the row space of $\boldsymbol{X}^C$.

The frequency domain method described in [28] is related to Algorithm 2 presented above. The algorithm in [28] uses the samples of the Fourier transform of the input and output signals $(U(\omega_k), Y(\omega_k))$ to determine data matrices corresponding to $\boldsymbol{G}$ and $\boldsymbol{W}_m$. If we consider the single-input case with uniform excitation for all frequencies, $U(\omega_k) \equiv 1$, and extend the Fourier transform samples $U(\omega_k)$ and $Y(\omega_k)$ with

$$\tilde{\boldsymbol{W}}^{\perp} = \frac{1}{\sqrt{M}} \begin{bmatrix} I_m & I_m & \cdots & I_m \\ e^{j2\pi/M} I_m & e^{j4\pi/M} I_m & \cdots & e^{j2\pi r/M} I_m \\ \vdots & \vdots & \ddots & \vdots \\ e^{j2\pi(M-1)/M} I_m & e^{j4\pi(M-1)/M} I_m & \cdots & e^{j2\pi r(M-1)/M} I_m \end{bmatrix} \in \mathbb{C}^{mM\times mr} \tag{69}$$

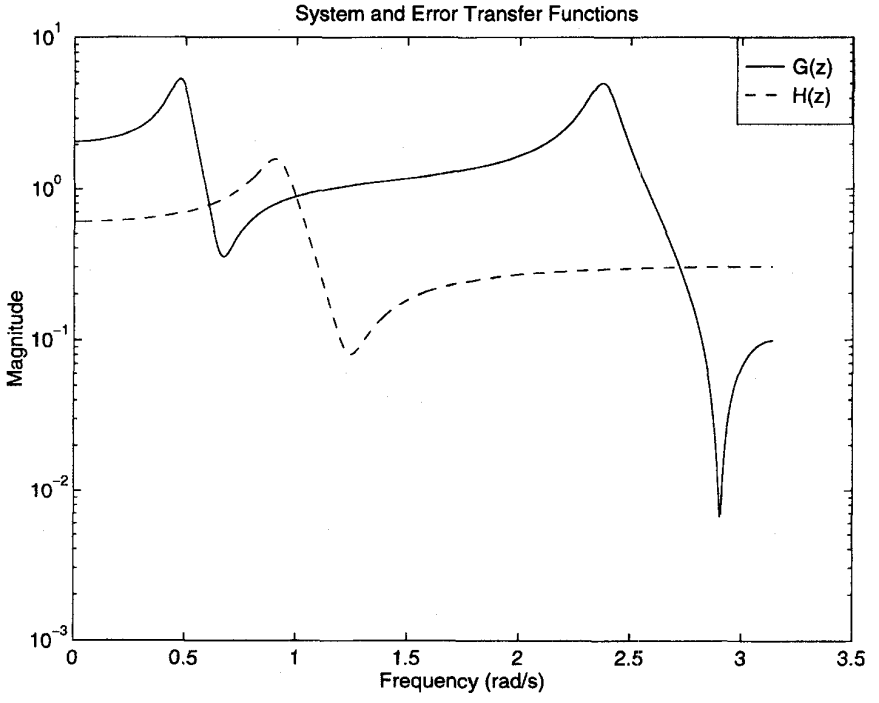

Fig. 1. Magnitude of system and error transfer functions.

their corresponding negative frequency values

$$U(-\omega_k) = U(\omega_k)^*, \quad Y(-\omega_k) = Y(\omega_k)^*$$

the data matrices in Algorithm 2 ($\boldsymbol{G}$ and $\boldsymbol{W}_m$) and the ones in the algorithm [28] will coincide. In their algorithm the observability range space is determined from the matrix

$$\boldsymbol{G}\boldsymbol{W}_m^{\perp}\boldsymbol{G}^H \in \mathbb{C}^{qp\times qp}. \tag{70}$$

If the columns in $\boldsymbol{G}$ and $\boldsymbol{W}_m$ are ordered in a particular way, (70) is a real matrix [28] and the extraction of the real and imaginary parts, as done in Algorithm 2, becomes superfluous. However, since the algorithms in [28] do not utilize any correction for the noise variances, it is only consistent if the frequencies are equidistantly spaced and the noise covariance function is constant and proportional to the identity matrix (see [35] and [33]).

C. Illustrative Example

This section describes an identification example based on simulated data. From the results of the example we will clearly see the difference between the two algorithms. Particularly, we will see the necessity to know the noise covariance function if Algorithm 2 is employed.

1) Experimental Setup: Let the true system $G(z) = D + C(zI - A)^{-1}B$ be a fourth-order system described by the state-space model

$$A = \begin{bmatrix} 0.8876 & 0.4494 & 0 & 0 \\ -0.4494 & 0.7978 & 0 & 0 \\ 0 & 0 & -0.6129 & 0.0645 \\ 0 & 0 & -6.4516 & -0.7419 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.2247 \\ 0.8989 \\ 0.0323 \\ 0.1290 \end{bmatrix}$$

$$C = [0.4719 \quad 0.1124 \quad 9.6774 \quad 1.6129]$$

$$D = 0.9626.$$

We assume $M + 1$ uniformly spaced experimental data are given as

$$G_k = G(z_k) + H(z_k)e_k, \quad z_k = e^{j\pi k/M}, \qquad k = 0, \cdots, M$$

where the noise term $H(z_k)e_k$ is composed of a noise transfer function $H(z)$, given by a second-order state-space model

$$H(z) = C_n + (zI - A_n)^{-1}B_n + D_n$$

with

$$A_n = \begin{bmatrix} 0.6296 & 0.0741 \\ -7.4074 & 0.4815 \end{bmatrix}, \quad B_n = \begin{bmatrix} 1.11 \\ 22.2 \end{bmatrix}$$

$$C_n = [1.6300 \quad 0.0740], \quad D_n = 0.356$$

and e_k being independent complex normally distributed random variables with unit variance. The variance for each frequency is given by $R_k = |H(z_k)|^2$. The magnitudes of $G(z)$ and $H(z)$ are depicted in Fig. 1.

In our study, we will consider four different identification algorithms:

A1: Algorithm 1;

A2: Algorithm 2 without knowledge of R_k;

A2wi: Algorithm 2 with knowledge of R_k;

LS: Levy's least-squares method [26].

Recall that Algorithm 1 does not use any explicit noise information. Levy's algorithm is included since it is also in

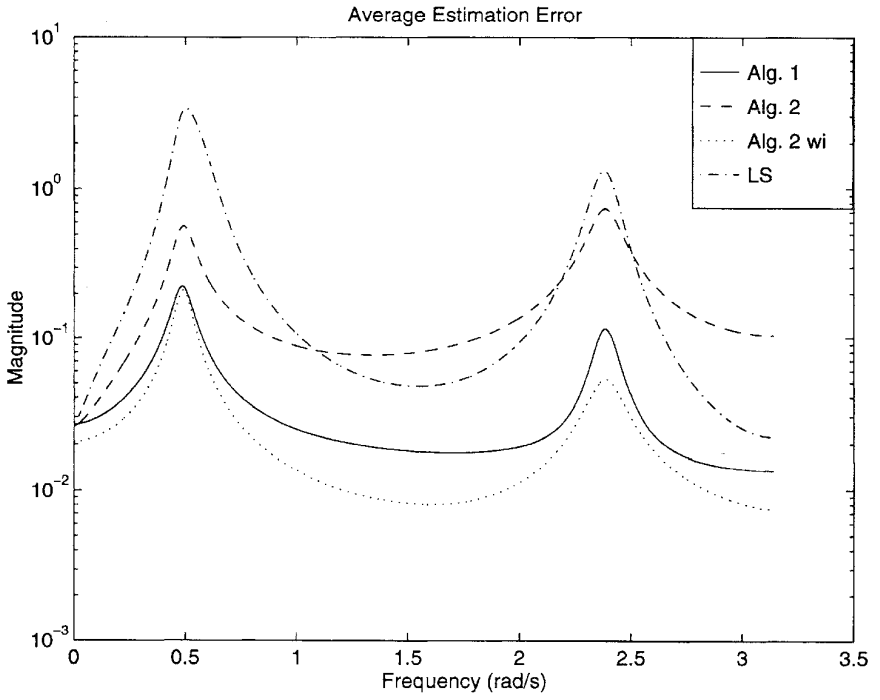

Fig. 2. Results from Monte Carlo simulations using data length $M = 1600$. The mean values of the magnitudes of the transfer function errors calculated over the 100 estimated models are shown for the four identification cases.

the class of noniterative methods which are correct. However, by common knowledge, it is not a consistent method except for a very particular noise model. The method is still a commonly used method mostly because of its simplicity.

To examine the consistency properties of the four cases we perform Monte Carlo simulations estimating the system, given the samples G_k, using different noise realizations of e_k, and making the frequency grid denser and denser, i.e., increasing M. Data lengths of 100, 200, 400, 800, and 1600 frequency samples will be used. For each data length, 100 different noise realizations are generated, and the algorithms estimate 100 models. To assess the quality of the resulting model both the supremum norm of the estimation error

$$\|\hat{G}^M(z) + G(z)\|_\infty$$

as well as the H_2 norm

$$\|\hat{G}^M(z) - G(z)\|_2^2 := \frac{1}{2\pi}\int_{-\pi}^{\pi} |\hat{G}^M(e^{j\omega}) - G(e^{j\omega})|^2 d\omega$$

are determined for each estimated model and averaged over the 100 estimated models.

2) Estimation Results: As expected from the analysis, Algorithm 1 and Algorithm 2 with covariance information yield the best performance. The simple least-squares algorithm has the worst performance. In Table I, the averaged maximum identification error is presented. The results indicate that Algorithm 2 without the covariance information and the LS algorithm are not consistent for this noise model which is predicted by the analysis. In Fig. 2, the error magnitudes, averaged over the 100 identification experiments, are shown for the four identification cases.

TABLE I
MONTE CARLO SIMULATIONS COMPARING THE FOUR ALGORITHMS. THE ESTIMATION ERROR DECREASES SIGNIFICANTLY AS THE NUMBER OF IDENTIFICATION DATA INCREASES FOR MODELS ESTIMATED WITH ALGORITHM 1 AND ALGORITHM 2 WHEN USING THE COVARIANCE INFORMATION (COLUMN A2wi). THE LEFT TABLE SHOWS THE AVERAGED SUPREMUM ERROR, AND THE RIGHT TABLE THE AVERAGED H_2 ERROR

	Average $\|\hat{G}^M - G\|_\infty$			
M	A1	A2	A2wi	LS
100	1.7340	1.2388	0.9207	3.3410
200	0.9708	0.9870	0.6059	3.3772
400	0.5631	0.8781	0.4472	3.3798
800	0.3659	0.8085	0.3211	3.3962
1600	0.2603	0.7651	0.2317	3.3977

	Average $\|\hat{G}^M - G\|_2^2$			
M	A1	A2	A2wi	LS
100	0.1580	0.1434	0.0365	0.5334
200	0.0541	0.0911	0.0163	0.5269
400	0.0200	0.0769	0.0088	0.5267
800	0.0078	0.0597	0.0044	0.5207
1600	0.0037	0.0560	0.0023	0.5192

V. PRACTICAL ASPECTS

A. Estimating Models of Different Orders

When facing a practical identification problem many models of different orders are estimated and compared to find a suitable "best" model. In the presented algorithms most of the computational effort lies in the SVD factorization (14) for Algorithm 1 and in the QR-factorization and the SVD for Algorithm 2. Given the factorization (14), all models of order less than q are easily obtained from the rest of the algorithms by letting n range from 1 to $q - 1$. Hence, the choice of appropriate model order can easily be accomplished by direct

comparison of a wide range of models with different orders at a rather low computational cost.

B. Guaranteeing Stability

Many times, stability is a most desirable feature of the estimated model. A stable A (all eigenvalues inside the unit circle) can be guaranteed by the following procedure [32]:

$$\hat{A} = \hat{U}_s^{\dagger} \begin{bmatrix} J_2 \hat{U}_s \\ 0_{p \times n} \end{bmatrix}.$$

The price paid is that the method will not yield the true A matrix even for he noise-free case unless the true system has a finite impulse response or if $q \to \infty$.

We would like to suggest a different approach to guarantee stability by adding an extra projection Step 5b) after (15). In this step all unstable eigenvalues are projected into the unit circle. The idea can be implemented in the following way.

- Transform $\hat{A}$ to the complex Schur form with the eigenvalues λ_i on the diagonal.
- Project any diagonal elements (eigenvalues) satisfying $1 < |\lambda_i| \leq 2$ into the unit disc by $\lambda_i := \lambda_i \left(\frac{2}{|\lambda_i|} - 1 \right)$. Eigenvalues with magnitude $|\lambda_i| > 2$ are set to zero. Eigenvalues on the unit circle can be moved into the unit disc by changing the magnitude of the eigenvalue to $1 - \epsilon$ for some small positive ϵ, i.e., $\lambda_i := \lambda_i (1 - \epsilon)$.
- Finally transform $\hat{A}$ back to its original form before proceeding further to determine $\hat{B}$ and $\hat{D}$.

This way of imposing stability does not suffer from consistency problems when identifying stable systems since only unstable eigenvalues are affected. A second advantage is that the magnitude of the frequency response is approximately unchanged by the projection.

VI. Identification of Continuous-Time Models

Most real world processes subject to modeling are of continuous-time character. However, measured input and output signals of the process are almost without exception in sampled form. If a discrete-time model is sought, the modeling is straightforward since the measured data are in sampled form. If, on the other hand, a continuous-time model is desired, two options are available. Either we assume that the input signal is piecewise constant between the sampling instants, which is known as the zero order hold (ZOH) assumption, or we assume that the input signal is band limited (BL) such that the continuous-time signals exactly can be reconstructed from the given samples.

To estimate continuous-time domain models under the ZOH assumption is straightforward if the following assumptions hold [42]:

- The input signal $u(t)$ is constant between the sampling instants.
- The continuous-time system is proper, i.e., $\lim_{\omega \to \infty} G(j\omega) < \infty$.
- The magnitudes of the imaginary parts of the poles and zeros of the system are less than the Nyquist frequency (π/T).

A discrete-time model can be estimated using the sampled data, and the continuous-time model is obtained by inverse ZOH-sampling of the discrete-time system. These observations hold both for time domain as well as frequency domain methods.

The BL assumption, as already mentioned, means that the input and output signals have no power above the Nyquist frequency. The estimation problem for general excitation signals should be performed in the time domain and is rather involved [48]. For the case of periodic excitation, the modeling is considerably simplified by considering identification in the frequency domain [42]. Under the BL assumption the DFT of the input and output signals equal their continuous-time transforms and the relation

$$G(j\omega_k) = \frac{Y(\omega_k)}{U(\omega_k)}$$

holds exactly. A continuous-time transfer function $G(s)$ can thus directly be fitted to the frequency data.

It is well known that the parameter estimation problem is better conditioned for discrete-time transfer functions since powers of $e^{j\omega}$ form a natural orthogonal basis [7]. By use of the bilinear transformation, the continuous-time identification problem can be solved in the discrete domain without introducing any approximation errors since the supremum norm $\| \cdot \|_{\infty}$ is invariant under the bilinear transformation.

A. The Bilinear Transformation

The bilinear transformation maps the complex values in the s domain to the z domain as

$$s = \frac{2(z-1)}{T(z+1)}$$

with the inverse

$$z = \frac{2 + sT}{2 - sT}.$$

The parameter T is a parameter in which the user is free to specify under constraint that $2/T$ is not a pole of the continuous-time system [2] and can be seen as a sort of sampling period.

If the continuous-time transfer function is given by $G(s)$, the bilinear transformation gives the discrete-time transfer function

$$G\left(\frac{2(z-1)}{T(z+1)} \right) = G^d(z).$$

The bilinear transformation maps poles and zeros in the left half-plane into the unit circle, while the right half-plane is mapped to the complement of the unit disc. The poles and zeros on imaginary axis are mapped onto the unit circle. Stability properties are thus preserved.

The important feature of the bilinear transformation is that the frequency response is invariant if we prewarp the frequency scale. Let the continuous-time transfer function be evaluated at $j\omega_k$ and let the bilinear transformed discrete-time transfer function be evaluated at $e^{j\omega_k^d}$, then it holds that [3]

$$G(j\omega_k) = G\left(\frac{2(e^{j\omega_k^d} - 1)}{T(e^{j\omega_k^d} + 1)} \right) = G^d(e^{j\omega_k^d})$$

if

$$\tan(\omega_k^d/2) = \omega_k T/2.$$

Hence, given samples of a continuous-time transfer function G_k at frequencies ω_k, the samples of the corresponding bilinear transformed discrete-time transfer function can be obtained simply as

$$G_k^d = G_k, \qquad k = 1, \cdots, M \tag{71}$$

$$\omega_k^d = 2\,\text{atan}\left(\frac{\omega_k T}{2}\right), \qquad k = 1, \cdots, M \tag{72}$$

where atan denotes the inverse of tan. After the discrete-time transfer function is estimated, the continuous-time transfer function is obtained through the inverse map

$$G(s) = G^d\left(\frac{2+sT}{2-sT}\right).$$

If samples of the input and output Fourier transforms are the primary data, the same approach can be applied.

State-Space Models: For state-space models, the bilinear transformation between

$$G(s) = D + C(sI - A)^{-1}B$$

and

$$G^d(z) = D^d + C^d(zI - A^d)^{-1}B^d$$

can, e.g., be described by the matrix relations [2]

$$\begin{aligned} A^d &= \left(\frac{2}{T}I + A\right)\left(\frac{2}{T}I - A\right)^{-1} \\ B^d &= \frac{2}{\sqrt{T}}\left(\frac{2}{T}I - A\right)^{-1}B \\ C^d &= \frac{2}{\sqrt{T}}C\left(\frac{2}{T}I - A\right)^{-1} \\ D^d &= D + C\left(\frac{2}{T}I - A\right)^{-1}B \end{aligned} \tag{73}$$

which imply

$$\begin{aligned} A &= \frac{2}{T}(I + A^d)^{-1}(A^d - I) \\ B &= \frac{2}{\sqrt{T}}(I + A^d)^{-1}B^d \\ C &= \frac{2}{\sqrt{T}}C^d(I + A^d)^{-1} \\ D &= D^d - C^d(I + A^d)^{-1}B^d. \end{aligned} \tag{74}$$

This particular choice of transformations has an additional advantage. The observability and controllability Gramian matrices as well as the Hankel singular values are invariant under the bilinear transformations (73) and (74) [16]. Particularly, a balanced realization remains balanced after the bilinear transformation. This implies that the transformations retain the system in a well-conditioned basis if the original system is given in a balanced realization.

VII. Identification of a Flexible Truss Structure

Although theoretical analysis is indispensable when developing new identification methods, practical experience is probably even more important. The real world is neither linear, nor are measurement errors stochastic variables drawn from some probability distribution.

This section deals with identification of linear models from measured data. The experimental data originates from a flexible mechanical structure which has a large number of lightly damped vibrational modes. We use the new algorithms as well as some classical methods. This enables a fair comparison.

A. The Data

This application considers the identification of the transfer function between a force-actuator and an accelerometer located on a flexible mechanical structure. The structure is the advanced reconfigurable control (ARC) testbed at the Jet Propulsion Laboratory (JPL), California Institute of Technology, Pasadena, California. The ARC testbed is a mechanical truss structure with several active struts and accelerometers at different locations.

The frequency data are obtained with a sampling frequency of 200 Hz using a multisine input [47] with 512 equidistant spectral lines. The frequency response data are shown in Fig. 3. As clearly seen from the phase plot in the figure, the frequency response data below 8 Hz is only noise. The response above 8 Hz appears to have a rather high signal-to-noise ratio. A quick look at the magnitude curve reveals about 16 dominant peaks.

B. Quality Measures

To assess the quality of estimated models, we will use two measures based on the fit between the data and the model. The maximum error

$$\|\hat{G}^M - G\|_{m,\infty} = \max_{\omega_k} |\hat{G}^M(e^{j\omega_k}) - G_k| \tag{75}$$

and the root-mean-square error (rms)

$$\|\hat{G}^M - G\|_{m,2} = \sqrt{\frac{1}{N}\sum_{k=1}^{N} |\hat{G}^M(e^{j\omega_k}) - G_k|^2}. \tag{76}$$

C. Model Order Determination

We start by trying to determine an appropriate model order. We do this by the technique of cross-validation [51]. The data set is divided into two disjoint sets, the estimation data and the validation data. The division is made such that every second frequency response sample is put in the estimation set and every other in the validation set. Frequency response samples at frequencies below 8 Hz are removed from the validation data since these are only noise. By only using the estimation data, models of different orders are estimated. From the validation data the model error is determined at the frequency points of the validation data. The underlying assumption is that if the model order is low, the error on validation data will decrease as the model order increases until an appropriate model order is found.

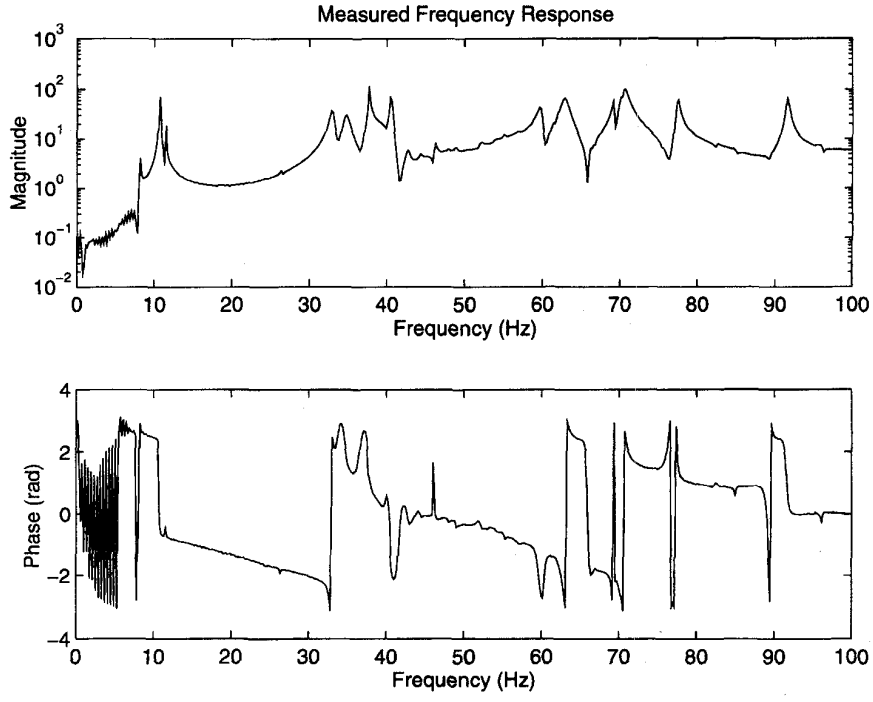

Fig. 3. Measured frequency response of the ARC testbed at JPL. The frequency response is given at 512 equidistant frequency points between 0 and 100 Hz. The sampling frequency is 200 Hz.

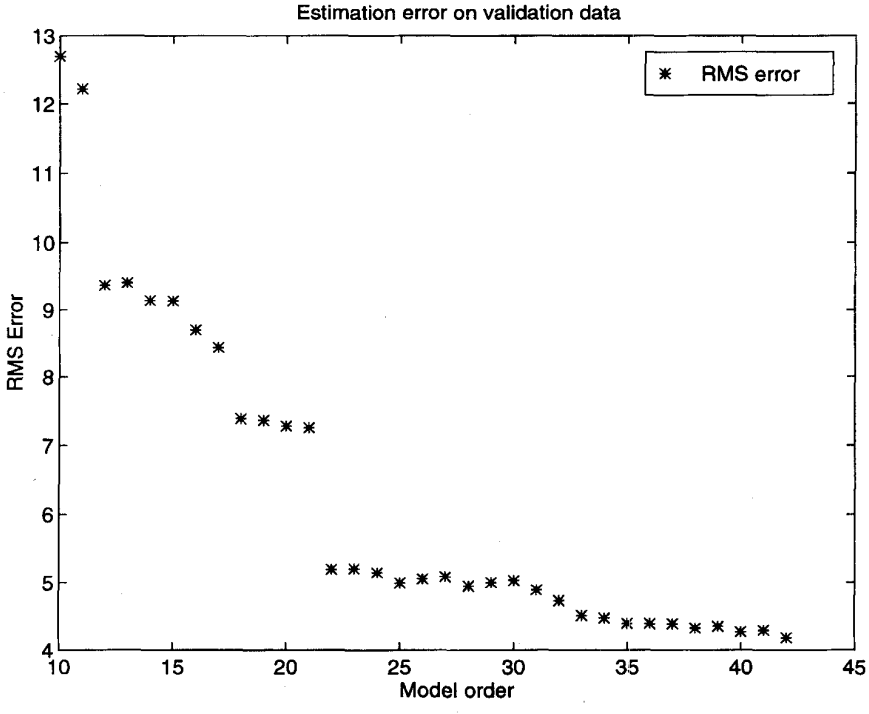

Fig. 4. Model errors (76) calculated on independent validation data plotted versus order of the estimated models using Algorithm 1.

Using Algorithm 1, a sequence of models of order 10–42 are estimated. The frequency response of each estimated model is calculated at the frequencies of the validation data and the rms error (76) is determined using the validation data set. The results are shown in Fig. 4. From the graph we can see no "knee" in the error curves, which, if present, would indicate an appropriate model order [50]. Instead the model error (on an average) decreases slowly with increasing model orders. The result indicates that the frequency data has a high signal-to-noise ratio and the best linear model has a high dimension.

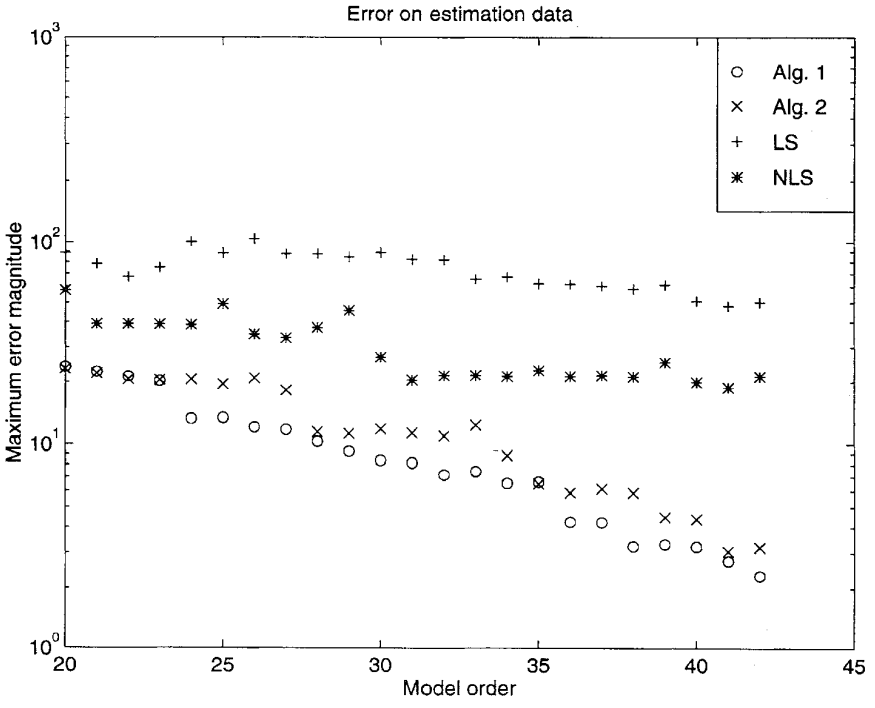

Fig. 5. Estimation errors for four different methods. The error is calculated using the estimation data.

D. A Comparison Study

Let us study the performance of the two new algorithms in comparison with some established estimation procedures. As estimation data we use all 512 frequency samples. The procedures we compare are the simple linear least-squares estimate (LS) introduced by Levy [26] and the nonlinear least squares (NLS) estimate. The LS estimate is calculated by minimizing

$$V_{LS}(\theta) = \sum_{k=1}^{M} |G_k a(e^{j\omega_k}, \theta) - b(e^{j\omega_k}, \theta)|^2$$

using two polynomials $a(z, \theta)$ and $b(z, \theta)$ with a monic and θ being the polynomial coefficients. The estimated model is $\hat{G}(z) = b(z, \hat{\theta})/a(z, \hat{\theta})$. The NLS estimate uses the LS estimate as an initial model before proceeding with Gauss-Newton iterative search to find the minimum of

$$V_{NLS}(\theta) = \sum_{k=1}^{M} \left| G_k - \frac{b(e^{j\omega_k}, \theta)}{a(e^{j\omega_k}, \theta)} \right|^2.$$

The LS and NLS estimation algorithms are implemented in the command ***invfreqz*** in MATLAB's Signal Processing Toolbox [27].

Models of order 20 to 42 are estimated using all four approaches. The dimension of the Hankel matrix $\hat{H}$ (13) in Algorithm 1 is chosen to be 512×512 and the number of rows in $\boldsymbol{G}$ for Algorithm 2 is chosen to be 250. These choices give the best accuracy. The resulting maximum errors calculated on the estimation data are shown in Fig. 5. The performance of Algorithm 1 and Algorithm 2 are significantly better than the two other least-squares methods. We notice an erratic behavior of the NLS estimate which probably is due to local minima. The rms error for all four methods behave similar as the shown maximum error.

The same data has been used by Gu and Khargonekar [19] and Friedman and Khargonekar [15]. In [19] they estimate discrete-time stable models with an algorithm inspired by the recent theory of identification in $\mathcal{H}_\infty$. They use SK iterations as a first step and fail to use the second step (which is needed for the algorithm to be robustly convergent). For model orders 24 and 42, they obtain maximum errors 13 and 6.1, respectively. In [15] the Fourier coefficients are used as an FIR model which is reduced, by FIR balanced truncation, to a 24th-order rational model with an approximate maximum error of 22. The maximum errors obtained by Algorithm 1 are 13.2 and 2.3, for models of order 24 and 42, respectively. This clearly indicates that the use of an FIR model as an intermediate step in the identification leads to less accurate models as compared with a direct approximation of a rational model to the given data using a correct algorithm.

Bayard has also successfully estimated models from these data using polynomial models [8]. However, in [8] the estimated transfer functions are shown without any explicit quantitative results.

Finally we conclude by showing the excellent fit obtained for the 42nd-order model estimated with Algorithm 1. The estimated stable transfer function is shown in Fig. 6.

E. Summary

High-order models have successfully been estimated from real measured data, using the new algorithms. The application shows that high-order state-space models of high quality easily are derived with the two algorithms. By using these subspace algorithms we obtain high-quality models without the

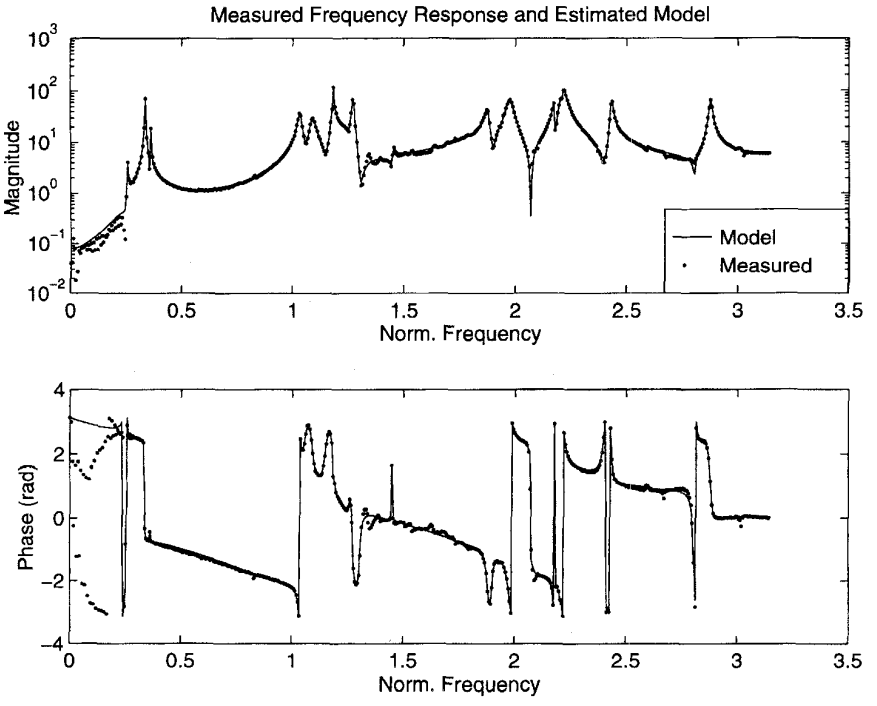

Fig. 6. Measured frequency response and estimated model using Algorithm 1. The model is stable and of order 42.

need for an explicit parameterization or an iterative nonlinear optimization.

VIII. Conclusions

In this paper, we have developed two simple state-space identification algorithms to identify linear MIMO systems from samples of the frequency response function. We have shown that both algorithms are correct. The first algorithm which uses data sampled at equidistantly frequencies is shown to be strongly consistent if the noise is zero mean and has a covariance function which is uniformly bounded. The consistency of the second algorithm is dependent on the *a priori* knowledge of the noise covariance function. However, this algorithm can use an arbitrary frequency spacing. The algorithms were used to identify a high-order flexible structure and a comparison with a nonlinear least-squares iterative algorithm was made. The results show that the new subspace algorithms outperform the nonlinear least-squares algorithm, for these data, and are therefore a viable alternative to classical iterative methods.

Acknowledgment

The authors are indebted to Dr. D. S. Bayard at the Jet Propulsion Laboratory, Pasadena, CA, who provided the experimental data used in this paper and motivated this work. The second author would also like to thank Prof. P. P. Khargonekar of the University of Michigan for introducing him to this problem during his doctoral studies at the University of Michigan.

References

[1] J. L. Adcock, "Curve fitter for pole-zero analysis," *Hewlett-Packard J.*, pp. 33–36, Jan. 1987.

[2] U. M. Al-Saggaf and G. F. Franklin, "Model reduction via balanced realizations: An extension and frequency weighting techniques," *IEEE Trans. Automat. Contr.*, vol. 33, pp. 687–692, July 1988.

[3] K. J. Åström and B. Wittenmark, *Computer Controlled Systems.* Englewood Cliffs, NJ: Prentice-Hall, 1984.

[4] D. Bayard, "Statistical plant set estimation using Schroeder-phased multisinusoidal input design," in *Proc. Amer. Contr. Conf.*, Chicago, IL, 1992, pp. 1707–1712.

[5] ——, "An algorithm for state-space frequency domain identification without windowing distortions," in *Proc. 31st IEEE Conf. Dec. Contr.*, Tucson, AZ, Dec. 1992, pp. 1707–1712.

[6] ——, "Multivariate state-space identification in the delta and shift operators: Algorithms and experimental results," in *Proc. Amer. Contr. Conf.*, San Francisco, CA, June 1993, pp. 3038–3042.

[7] ——, "High-order multivariable transfer function curve fitting: Algorithms, sparse matrix methods and experimental results," *Automatica*, vol. 30, no. 9, pp. 1439–1444, 1994.

[8] D. R. Brillinger, *Time Series: Data Analysis and Theory.* New York: McGraw-Hill, 1981.

[9] K. L. Chung, *A Course in Probability Theory.* New York: Academic, 1974.

[10] R. L. Dailey and M. S. Lukich, "MIMO transfer function curve fitting using Chebyshev polynomials," in *Proc. SIAM 35th Anniv. Meet.*, Denver, CO, 1987.

[11] B. De Moor, "The singular value decomposition and long and short spaces of noisy matrices," *IEEE Trans. Signal Processing*, vol. 41, pp. 2826–2838, 1993.

[12] B. De Moor and J. Vandewalle, "A geometrical strategy for the identification of state space models of linear multivariate systems with singular value decomposition," in *Proc. 3rd Int. Symp. Appl. Multivariable Syst. Tech.*, Plymouth, U.K., Apr. 13–15, 1987, pp. 59–69.

[13] J. C. Doyle, "Structured uncertainty in control systems design," in *Proc. 24th IEEE Conf. Dec. Contr.*, Fort Lauderdale, FL, Dec. 1985, pp. 260–265.

[14] J. H. Friedman and P. P. Khargonekar, "A comparative applications study of frequency domain identification techniques," in *Proc. Amer. Contr. Conf.*, Seattle, WA, June 1995, pp. 3055–3059.

[15] K. Glover, "All optimal Hankel-norm approximations of linear multivariable systems and their L^∞-error bounds," *Int. J. Contr.*, vol. 39, no. 6, pp. 1115–1193, 1984.

[16] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1989.

[17] G. Gu and P. P. Khargonekar, "A class of algorithms for identification in $\mathcal{H}_\infty$," *Automatica*, vol. 37, pp. 299–312, 1992.

[18] ——, "Frequency domain identification of lightly damped systems: The JPL example," in *Proc. Amer. Contr. Conf.,* San Francisco, CA, 1993, pp. 3052–3056.
[19] G. Gu and P. Misra, "Identification of linear time-invariant systems from frequency response data corrupted by bounded noise," *IEE Proc. D,* vol. 139, no. 2, pp. 135–140, Mar. 1992.
[20] A. J. Helmicki, C. A. Jacobson, and C. N. Nett, "Control-oriented system identification: A worst-case/deterministic approach in $\mathcal{H}_\infty$," *IEEE Trans. Automat. Contr.,* vol. 36, pp. 1163–1176, 1991.
[21] B. L. Ho and R. E. Kalman, "Effective construction of linear state-variable models from input/output functions," *Regelungstechnik,* vol. 14, no. 12, pp. 545–548, 1966.
[22] J. N. Juang and R. S. Pappa, "An eigensystem realization algorithm for modal parameter identification and model reduction," *J. Guidance, Contr. Dynamics,* vol. 8, no. 5, pp. 620–627, 1985.
[23] J. N. Juang and H. Suzuki, "An eigensystem realization algorithm in frequency domain for modal parameter identification," *J. Vibration, Acoust., Stress, Reliab. Des.,* vol. 110, pp. 24–29, Jan. 1988.
[24] S. Y. Kung, "A new identification and model reduction algorithm via singular value decomposition," in *Proc. 12th Asilomar Conf. Circuits, Syst. Comput.,* Pacific Grove, CA, 1978, pp. 705–714.
[25] E. C. Levy, "Complex curve fitting," *IRE Trans. Automat. Contr.,* vol. AC-4, pp. 37–44, May 1959.
[26] J. Little and L. Shure, *Signal Processing Toolbox.* The Mathworks, Inc., 1988.
[27] K. Liu, R. N. Jacques, and D. W. Miller, "Frequency domains structural system identification by observability range space extraction," in *Proc. Amer. Contr. Conf.,* vol. 1, Baltimore, MD, June 1994, pp. 107–111.
[28] K. Liu and R. E. Skelton, "Q-Markov covariance equivalet realization and its application to flexible structure identification," *AIAA J. Guidance, Contr. Dynamics,* vol. 16, no. 2, pp. 308–319, 1993.
[29] L. Ljung, *System Identification: Theory for the User.* Englewood Cliffs, NJ: Prentice-Hall, 1987.
[30] ——, "Some results on identifying linear systems using frequency domain data," in *Proc. 32nd IEEE Conf. Dec. Contr.,* San Antonio, TX, Dec. 1993, pp. 3534–3538.
[31] J. M. Maciejowski, "Guaranteed stability with subspace methods," *Syst. Contr. Lett.,* vol. 26, no. 2, pp. 153–156, 1995.
[32] T. McKelvey, "Identification of state-space models from time and frequency data," Ph.D. dissertation, Linköping Univ., May 1995.
[33] T. McKelvey and H. Akçay, "An efficient frequency domain state-space identification algorithm," in *Proc. 33rd IEEE Conf. Dec. Contr.,* Lake Buena Vista, FL, Dec. 1994, pp. 3359–3364.
[34] ——, "An efficient frequency domain state-space identification algorithm: Robustness and stochastic analysis," in *Proc. 33rd IEEE Conf. Dec. Contr.,* Lake Buena Vista, FL, Dec. 1994, pp. 3348–3353.
[35] T. McKelvey, L. Ljung, and H. Akçay, "Identification of infinite dimensional systems from frequency response data," in *Proc. 3rd European Contr. Conf.,* vol. 3, Rome, Italy, 1995, pp. 2106–2111.
[36] B. C. Moore, "Principal component analysis in linear systems: Controllability, observability, and model reduction," *IEEE Trans. Automat. Contr.,* vol. AC-26, pp. 17–32, 1981.
[37] B. Ottersten and M. Viberg, "A subspace based instrumental variable method for state-space system identification," in *Proc. 10th IFAC Symp. Syst. Identif.,* vol. 2, Copenhagen, Denmark, July 1994, pp. 139–144.
[38] J. R. Partington, "Robust identification and interpolation in H_∞," *Int. J. Contr.,* vol. 54, pp. 1281–1290, 1991.
[39] L. E. Pfeffer, "The RPM toolbox: A system for fitting linear models to frequency response data," in *Proc. 1993 MATLAB Conf.,* Cambridge, MA, Oct. 1993.
[40] R. Pintelon, P. Guillaume, Y. Rolain, J. Schoukens, and H. Van Hamme, "Parametric identification of transfer functions in the frequency domain—A survey," *IEEE Trans. Automat. Contr.,* vol. 39, pp. 2245–2260, Nov. 1994.
[41] R. Pintelon, J. Schoukens, and H. Chen, "On the basic assumptions in the identification of continuous time systems," in *Proc. 10th IFAC Symp. Syst. Identif.,* vol. 3, Copenhagen, Denmark, July 1994, pp. 143–152.
[42] K. Poolla, P. P. Khargonekar, A. Tikku, J. Krause, and K. M. Nagpal, "A time domain approach to model validation," *IEEE Trans. Automat. Contr.,* vol. 39, pp. 951–959, 1994.
[43] Y. Rolain, R. Pintelon, K. Q. Xu, and H. Vold, "On the use of orthogonal polynomials in high order frequency domain system identification and its application to modal parameter estimation," in *Proc. 33rd IEEE Conf. Dec. Contr.,* Lake Buena Vista, FL, Dec. 1994, pp. 3365–3373.
[44] R. Roy and T. Kailath, "ESPRIT—Estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 37, pp. 984–995, July 1989.
[45] C. K. Sanathanan and J. Koerner, "Transfer function synthesis as a ratio of two complex polynomials," *IEEE Trans. Automat. Contr.,* vol. AC-8, pp. 56–58, Jan. 1963.
[46] J. Schoukens and R. Pintelon, *Identification of Linear Systems: A Practical Guideline to Accurate Modeling.* London: Pergamon, 1991.
[47] N. K. Sinha and G. P. Rao, *Identification of Continuous-Time Systems.* Norwell, MA: Kluwer, 1994.
[48] R. S. Smith and J. C. Doyle, "Model validation: A connection between robust control design and identification," *IEEE Trans. Automat. Contr.,* vol. 37, pp. 942–952, 1992.
[49] T. Söderström and P. Stoica, *System Identification.* Hemel Hempstead, Hertfordshire: Prentice-Hall, 1989.
[50] P. Stoica, P. Eykhoff, P. Janssen, and T. Söderström, "Model structure selection by cross validation," *Int. J. Contr.,* vol. 43, pp. 1841–1878, 1986.
[51] L. Swindlehurst, R. Roy, B. Ottersten, and T. Kailath, "A subspace fitting method for identification of linear state-space models," *IEEE Trans. Automat. Contr.,* vol. 40, pp. 311–316, Feb. 1995.
[52] A. J. van der Veen, E. F. Deprettere, and A. L. Swindlehurst, "SVD-based estimation of low-rank system parameters," in *Algorithms and Parallel VLSI Architectures,* E. F. Deprettere and A. J. van der Veen, Eds. New York: Elsevier, 1991.
[53] P. Van Overschee and B. De Moor, "N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems," *Automatica,* vol. 30, no. 1, pp. 75–93, 1994.
[54] ——, "Frequency domain subspace identification algorithms," in *Proc. 14th Benelux Meet. Syst. Contr.,* Houthalen, Belgium, 1995.
[55] M. Verhaegen and P. Dewilde, "Subspace model identification. Part I: The output-error state-space model identification class of algorithms," *Int. J. Contr.,* vol. 56, pp. 1187–1210, 1992.
[56] M. Viberg, B. Ottersten, B. Wahlberg, and L. Ljung, "A statistical perspective on state-space modeling using subspace methods," in *Proc. 30th IEEE Conf. Dec. Contr.,* Brighton, England, 1991, pp. 1337–1342.
[57] A. H. Whitfield, "Asymptotic behavior of transfer function synthesis methods," *Int. J. Contr.,* vol. 45, no. 3, pp. 1083–1092, 1987.
[58] H. P. Zieger and A. J. McEwen, "Approximate linear realizations of given dimension via Ho's algorithm," *IEEE Trans. Automat. Contr.,* vol. AC-19, p. 153, Apr. 1974.

Tomas McKelvey was born in Lund, Sweden, in 1966. He received the M.Sc. degree in electrical engineering from Lund Institute of Technology, Sweden, in 1991 and the Ph.D. degree in automatic control from Linköping University, Sweden, in 1995.

Since 1995 he has been a Research Associate in the Department of Electrical Engineering, Linköping University. His research interests include various aspects of system identification and signal processing.

Hüseyin Akçay was born in Antalya, Turkey, in 1958. He received the Engineer degree from the Istanbul Technical University in 1981, the M.Sc. degree from the Massachusetts Institute of Technology, Cambridge, in 1988, and the Ph.D. degree from the University of Michigan, Ann Arbor, in 1992, all in mechanical engineering, and the M.A. degree in mathematics from the University of Michigan in 1991.

During 1993 he spent one year as a Postdoctoral Researcher in the Department of Electrical Engineering, Linköping University, Sweden. Currently he is a Research Scientist at Tübitak, Marmara Research Centre, Mathematics Division. His research interests include control-oriented system identification and signal processing.

Lennart Ljung (S'74–M'75–SM'83–F'85) was born in Malmö, Sweden. He received the Ph.D. degree from Lund University, Sweden, in 1974.

Since 1976 he has been the Professor of the Chair of Automatic Control at Linköping University, Sweden. He is an Associate Editor of several journals.

Dr. Ljung is a member of the Royal Swedish Academy of Engineering Sciences, the Royal Swedish Academy of Science, and also an IFAC Advisor.