Pergamon

Automatica, Vol. 32, No. 6, pp. 885–902, 1996

0005-1098/96 $15.00+0.00

PII:S0005-1098(96)00022-2

Subspace-based Identification of Infinite-dimensional Multivariable Systems from Frequency-response Data

TOMAS MCKELVEY,† HÜSEYIN AKÇAY ‡ and LENNART LJUNG †

A subspace-based identification algorithm, which takes samples of an infinite-dimensional transfer function, is shown to produce estimates which converge to a balanced truncation of the system and is applied to real and simulated data with promising results.

Key Words—System identification; frequency-response data; infinite-dimensional systems; state-space models.

Abstract—A new identification algorithm which identifies low complexity models of infinite-dimensional systems from equidistant frequency-response data is presented. The new algorithm is a combination of the Fourier transform technique with the recent subspace techniques. Given noise-free data, finite-dimensional systems are exactly retrieved by the algorithm. When noise is present, it is shown that identified models strongly converge to the balanced truncation of the identified system if the measurement errors are covariance bounded. Several conditions are derived on consistency, illustrating the trade-offs in the selection of certain parameters of the algorithm. Two examples are presented which clearly illustrate the good performance of the algorithm. Copyright © 1996 Elsevier Science Ltd.

NOTATION

j	$\sqrt{-1}$
A^T	transpose of A
A^*	complex conjugate of A
A^H	complex conjugate transpose of A
$A^\dagger$	$(A^HA)^{-1}A^H$, the Moore–Penrose pseudo inverse of full column-rank matrix A.
I_m	identity matrix of size $m \times m$
$0_{m\times p}$	zero matrix of size $m \times p$
$\mathrm{tr}(A)$	$\sum_i a_{ii}$, trace of A
$\|A\|_F$	$\sqrt{\mathrm{tr}(A^HA)}$, the Frobenius norm of A
$\sigma_i(A)$	ordered singular values of A, $\sigma_1 \geq \sigma_2 \geq \dots$
$\mathcal{H}_\infty$	Hardy space of matrix-valued bounded analytic functions in the complement of the closed unit disc of the complex plane
$\|G\|_\infty$	sup norm of G, equals $\sup_\theta \sigma_1(G(e^{j\theta}))$
$\ell_1^{p\times m}$	set of sequences in $\mathbb{R}^{p\times m}$ such that $\sum_{k=0}^\infty \|g_k\| < \infty$
ℓ_2^m	set of sequences in $\mathbb{R}^m$ such that $\sum_{k=0}^\infty \|u_k\|_2^2 < \infty$
$\beta_k = O(\alpha_k)$	given two sequences of numbers α_k and β_k, there exists an integer M and a constant K such that $\|\beta_k\| \leq K\|\alpha_k\|$ for all $k \geq M$
$O(1)$	asymptotically bounded
$\beta_k = o(\alpha_k)$	given two sequences of numbers α_k and β_k, $\lim_{k\to\infty} \|\beta_k\|/\|\alpha_k\| = 0$
$o(1)$	asymptotically vanishing
$\Gamma(G)$	Hankel operator of a linear system G
$\Gamma_i(G)$	ordered Hankel singular values $\Gamma_1(G) \geq \Gamma_2(G) \geq \dots$ of a system G
E	mathematical expectation operator
w.p. 1	with probability one
ω_G	modulus of continuity of transfer function G

Received 13 January 1995; received in final form 4 January 1996. This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor J. Bokor under the direction of Editor Törsten Söderström.
Corresponding author Dr Tomas McKelvey. Tel. +46 13 282461; Fax +46 13 282622; E-mail tomas@isy.liu.se.
† Department of Electrical Engineering, Linköping University, S-581 83 Linköping, Sweden.
‡ TÜBITAK, Marmara Research Centre, Division of Mathematics, P.0. Box 21, Gebze-KOCAELI, Turkey.

1. INTRODUCTION

Identification of infinite-dimensional systems has been much studied recently both in the time domain (Ljung and Yuan, 1985; Huang and Guo, 1990; Guo *et al.*, 1990; Mäkilä, 1991; Jacobson *et al.*, 1992; Hjalmarsson, 1993) and in the frequency domain (Helmicki *et al.*, 1991; Mäkilä and Partington, 1991; Gu and Khargonekar, 1992). Despite that low-order nominal models are preferred in most practical applications as in the design of model-based controllers, the true systems are usually of high or infinite order with unmodeled dynamics and random/deterministic noise. Thus, the basic task of system identification is to construct a simple nominal model based on the measured data generated from a complex system.

Based on how the disturbances are characterized, problem formulations in both domains can be divided into two categories. In the traditional stochastic formulations, the disturbances have been assumed to be random variables which lead to instrumental variable and prediction error methods. See, for example, the books Ljung (1987) and Söderström and Stoica (1989). The least-squares method is the archetype for such methods. Then, under suitable conditions on the unknown system and exogenous noise, letting model orders increase as the size of data grows one hopes to approximate an infinite-dimensional system well. Such a procedure, called the *black-box* identification algorithm, having desired convergence properties is described by Ljung and Yuan (1985). In the deterministic problem formulations on the other hand, (Helmicki *et al.*, 1991; Mäkilä and Partington, 1991; Jacobson *et al.*, 1992; Gu and Khargonekar, 1992), the disturbances are treated as deterministic signals and a *robust convergence* notion requiring nonlinear algorithms is introduced. The performance of the algorithm is measured by the worst-case identification error. The robust convergence simply refers to the property that worst-case errors vanish with increasing model order as noise amplitude is decreased and data size grows.

In both approaches, a prejudice-free model set of high complexity is the underlying model structure. In most practical applications on the other hand, the model is required to be of restricted complexity despite the fact that the true system might have infinite order. Thus, model reduction is a complementary step to the black-box identification. Besides the computational complexity, this step induces large approximation errors unless the system that has generated the data has a special structure, which has been overlooked in most identification studies. The robust algorithms in the $\mathcal{H}_\infty$ identification framework (Helmicki *et al.*, 1991; Mäkilä and Partington, 1991; Gu and Khargonekar, 1992) deliver bounded errors as model complexity increases unboundedly. However, the total error becomes large after model reduction.

An alternative method is to directly realize low complexity models from the experimental data. In the traditional way, a system is modeled by a parametric transfer function which is the fraction of two polynomials with real coefficients and a nonlinear least-squares fit to the data is sought (Ljung, 1993; Pintelon *et al.*, 1994b). The solution to this nonlinear parametric optimization problem is obtained by iterations. During the last few years, some noniterative subspace-based algorithms which deliver state-space models without any parametric optimization have appeared in the literature (Verhaegen and Dewilde, 1992; Van Overschee and De Moor, 1994). It is well known that models in *canonical* minimal parametrizations are numerically sensitive, particularly for high-order models, in comparison with state-space models in a balanced realization. Subspace-based algorithms are more robust to numerical inaccuracies than the canonically parametrized models since the model obtained is normally close to being balanced.

The present paper deals with a frequency-domain identification problem. In this formulation, the experimental data are taken to be noisy values of the frequency-response of a system at a given set of frequencies. In a number of applications, as in the modal analysis area of mechanical engineering, lightly damped large structures with several inputs and outputs are frequently encountered and high-order models are needed to capture the dynamics of such systems. Sophisticated data analyzers and data acquisition equipment allow large amounts of time-domain data to be compressed into a small amount of frequency-response data. The step from time-domain measurements to frequency-response data provides noise reduction if the experimental conditions are carefully chosen, e.g. the use of periodic excitation (Schoukens and Pintelon, 1991). The identification data can also be compiled from several different time-domain experiments which facilitates the determination of models which are accurate over a wide frequency band.

Frequency-domain subspace algorithms (Juang and Suziki, 1988; Liu *et al.*, 1994; McKelvey and Akçay, 1994) are based on the famous realization algorithm by Ho and Kalman (1966) or the version by Kung (1978). The realization algorithms in Ho and Kalman (1966) and Kung (1978) find a minimal state-space realization given a finite sequence of the Markov parameters. The Markov parameters or impulse-response coefficients of the system can be estimated from the inverse discrete Fourier transform (DFT) of the frequency-response data. The approach described by Juang and Suziki (1988) is exact only if the system has a finite impulse-response and therefore for lightly damped systems yields very poor estimates. This stems from the fact that the estimated impulse response, using a finite number of frequency data, is subject to aliasing effects if the system has an impulse response of infinite length. In McKelvey and Akçay (1994), the inverse DFT technique is combined with a subspace identification step yielding the true finite-dimensional system in spite of this aliasing effect of the estimated impulse response. The current paper reports extensions of the results by McKelvey and Akçay (1994) for the case of infinite-dimensional systems.

We will now outline the contents of this paper. In Section 2, we formulate the problem. In Section 3, we present a new identification algorithm. Convergence properties of the new algorithm for noise-free data are studied in Section 4. In Section 5, the main result of the paper is presented. Section 6 continues with a brief discussion on the identification of continuous-time systems and some practical aspects on the implementation of the algorithm are discussed in Section 7. In Section 8, the properties of the new and several other algorithms are studied by means of two examples. In the first example, five algorithms are tested on real data originating from a frequency-response experiment on a flexible structure testbed at the Jet Propulsion Laboratory (JPL), Pasadena, California. The JPL-data are also used in the identification studies (Gu and Khargonekar, 1993; Bayard, 1994; McKelvey and Akçay, 1994; Friedman and Khargonekar, 1995). In the second example, we simulate a system described by Gu *et al.* (1989). Section 9 contains the conclusions. A preliminary version of this paper appeared as McKelvey *et al.* (1995).

2. PROBLEM FORMULATION

In this section, we describe the low complexity identification problem of infinite-dimensional systems from equally spaced frequency-response measurements. This problem formulation is a complement to our previous finite-dimensional formulation (McKelvey and Akçay, 1994). We first focus on the discrete-time case and briefly discuss the continuous-time case in Section 6.

Let $G(z)$ denote the transfer function of a linear time-invariant (LTI), discrete-time, multi-input/multi-output (MIMO), ℓ_2-BIBO (bounded-input/bounded-output) stable real system. Then, $G \in \mathcal{H}_\infty$.

The input/output behavior of the system can be described by the impulse response coefficients g_k through the equation:

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

where $u(t) \in \mathbb{R}^m$ and $y(t) \in \mathbb{R}^p$ are inputs and outputs, respectively, and $g_k \in \mathbb{R}^{p\times m}$. The frequency response of the system is calculated as

$$G(e^{j\theta}) = \sum_{k=0}^{\infty} g_k e^{-j\theta k}, \quad 0 \le \theta \le \pi. \qquad (2)$$

Since our systems are real ($g_k \in \mathbb{R}^{p\times m}$) the frequency response satisfies the usual complex conjugate symmetry property

$$G(e^{-j\theta}) = G^*(e^{j\theta}), \quad 0 \le \theta \le \pi, \qquad (3)$$

which will be used to get the frequency response on $[\pi, 2\pi]$.

For practical purposes this type of infinite-dimensional model is rather useless since it is not possible to calculate $y(t)$ knowing a finite amount of the past inputs and outputs, usually called the state of the system. For engineering purposes, a much more practical model is a state-space model:

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k), \\ y(k) &= Cx(k) + Du(k), \end{aligned} \qquad (4)$$

where $x(k) \in \mathbb{R}^n$. In this model $y(k)$ can be calculated using only the state vector $x(k)$ of length n and the current input $u(k)$ which both are finite objects. The state-space model (4) is a special case of (1) with

$$g_k = \begin{cases} D, & k = 0, \\ CA^{k-1}B, & k > 0. \end{cases} \qquad (5)$$

It is thus of practical interest to identify a finite-dimensional model (4) which is a good approximation of the infinite-dimensional system (1).

Some further assumptions must be imposed on the system to obtain good approximations. A set of conditions can conveniently be stated in terms of the Hankel singular values of the system. Recall that the Hankel operator of the system G with symbol Γ defined on ℓ_2^m by

$$(\Gamma u)(t) \triangleq \sum_{i=0}^{\infty} g_{t+i+1} u(i), \quad t \ge 0 \qquad (6)$$

is a mapping into ℓ_2^p. Let Γ^* be the adjoint of Γ. The Hankel singular values $\Gamma_i(G)$ are defined to be the square roots of the eigenvalues of $\Gamma\Gamma^*$. Let u_i and v_i be the corresponding normalized eigenvectors of $\Gamma\Gamma^*$ and $\Gamma^*\Gamma$, respectively. The pair (v_i, u_i) is called the Schmidt pair and satisfies

$$\begin{aligned} \Gamma v_i &= \Gamma_i(G) u_i, \\ \Gamma^* u_i &= \Gamma_i(G) v_i. \end{aligned}$$

A system G is said to be *Hilbert–Schmidt* if its Hankel singular values satisfy

$$\sum_{k=1}^{\infty} \Gamma_k^2(G) < \infty \qquad (7)$$

and *nuclear* if

$$\sum_{k=1}^{\infty} \Gamma_k(G) < \infty. \tag{8}$$

From (8) we see that all finite-dimensional linear systems form a subset of nuclear systems and nuclear systems themselves are contained in the set of Hilbert–Schmidt systems.

It is possible to identify these classes with impulse-response decay rates. Examples of systems with Hilbert–Schmidt Hankel operators are systems with impulse responses which decay as

$$\|g_k\| = O(k^{-\alpha}), \quad \alpha > 1. \tag{9}$$

This is a result of the following identity:

$$\sum_{k=1}^{\infty} \Gamma_k^2(G) = \sum_{k=1}^{\infty} k\|g_k\|^2. \tag{10}$$

Other examples are systems with $\|g_k\| = O(1/k(\log k))$. Bonnet (1993) showed that a sufficient condition for the nuclearity is given in terms of the decay rate for the impulse response as

$$\|g_k\| = O(k^{-\alpha}), \quad \alpha > 3/2. \tag{11}$$

Conversely, sufficient conditions for a system to have a Hilbert–Schmidt or nuclear Hankel operator can be stated in terms of boundary behavior of the system transfer function and its derivatives. We refer the interested reader to the paper by Curtain (1985) for a discussion on the sufficient conditions for nuclearity. A full discussion of Hankel operators is beyond the scope of this paper. We summarize the requirement on G as a standing assumption.

Assumption 1. The system $G \in \mathcal{H}_\infty$ has a continuous transfer function and a Hilbert–Schmidt Hankel operator Γ

$$\sum_{k=1}^{\infty} \Gamma_k^2(G) < \infty.$$

For a fixed given n, the Hankel singular values satisfy

$$\Gamma_n(G) > \Gamma_{n+1}(G).$$

■

Next we introduce a group of smoothness classes for periodic complex-valued functions. The modulus of continuity for a complex-valued periodic function f on the unit circle is the function

$$\omega_f(t) \triangleq \sup_{|x-y|\le t} \|f(e^{jx}) - f(e^{jy})\|. \tag{12}$$

We say that f is of class Λ_α, $(0 < \alpha \le 1)$ if $\omega_f(t) = O(t^\alpha)$ as $t \to 0$.

Optimal Hankel norm and balanced truncations are two popular model reduction techniques for nuclear systems and they are known to produce the same upper bound on the approximation error by

$$\|G_n - G\|_\infty \le 2 \sum_{k=n+1}^{\infty} \Gamma_k(G), \tag{13}$$

where repeated singular values are omitted in the sum and G_n is nth-order balanced truncation of G (Hinrichsen and Pritchard, 1990).

In this paper, we will discuss methods to obtain low complexity models of the infinite-dimensional systems described above, given uniformly spaced experimental frequency-response data of the system

$$G_k \triangleq G(e^{jk\pi/M}) + e_k; \quad k = 0, \ldots, M, \tag{14}$$

where the frequency-response measurement noise e_k is assumed to satisfy some conditions.

Assumption 2. The noise e_k, $k = 0, \ldots, M$ are independent zero-mean complex random variables with uniformly-bounded second moments

$$R_k \triangleq E\{e_k e_k^H\} \le \bar{R}, \quad \forall k. \tag{15}$$

■

Since (13) is the best available bound on the approximation error, our objective is to achieve the same bound on the identification error asymptotically (with probability one), i.e.

$$\lim_{M\to\infty} \|\hat{G}_{n,M} - G\|_\infty \le 2 \sum_{k=n+1}^{\infty} \Gamma_k(G) \quad \text{w.p. } 1, \tag{16}$$

where $\hat{G}_{n,M}$ is the nth-order identified model using $M+1$ frequency data.

The above objective is achieved by many algorithms. Examples are the *so-called* two-stage algorithms. The two-stage algorithms are black-box type algorithms. In the first stage of a two-stage algorithm, a linearly parametrized model structure, is used to arrive at a *pre-identified* model and in the second stage, an nth order rational approximation to the pre-identified model gives $\hat{G}_{n,M}$. We refer the reader to Heuberger *et al.* (1995) and references therein for some interesting parametrizations. Unless the model set is suitably parametrized, a lightly damped system yields high-order pre-identified models and hence the number of data and computations needed for the accuracy increase dramatically. Therefore, a potential identification algorithm must have good performance for finite data sets in addition to satisfying (16). This is the case if the algorithm exactly retrieves the system when restricted to finite-dimensional systems and noise-free data of finite length. Such algorithms are called *correct* algorithms.

Given the problem formulation, there exist many algorithms with the aforementioned properties. In the next section, we present one such algorithm. Our algorithm is not necessarily optimal. We have not introduced an optimality criterion in this paper. Indeed, optimality depends on more restrictive assumptions than we have made on the system and noise. Our objective in this paper is to study consistency properties of a new algorithm and analyze the trade-offs in choosing the parameters to achieve consistency.

3. STATE-SPACE MODEL IDENTIFICATION IN FREQUENCY DOMAIN

In this section, we will introduce a new identification algorithm:

Algorithm 1.

(1) Expand the given frequency data (14) according to (3) as

$$G_{M+k} \triangleq G^*_{M-k}, \quad k = 1, \dots, M-1 \qquad (17)$$

and perform the $2M$-point inverse DFT on the expanded data

$$\hat{g}_i \triangleq \frac{1}{2M} \sum_{k=0}^{2M-1} G_k \, e^{j2\pi ik/2M}, \quad i = 0, \dots, q+r-1 \qquad (18)$$

to obtain the estimates of the impulse-response coefficients g_i.

(2) Construct the $q \times r$-block Hankel matrix

$$\hat{H}_{qr} \triangleq \begin{bmatrix} \hat{g}_1 & \cdots & \hat{g}_r \\ \vdots & \ddots & \vdots \\ \hat{g}_q & \cdots & \hat{g}_{q+r-1} \end{bmatrix} \qquad (19)$$

and perform a singular value decomposition for $\hat{H}_{qr}$ as follows

$$\hat{H}_{qr} \triangleq \begin{bmatrix} \hat{U}_1 & \hat{U}_2 \end{bmatrix} \begin{bmatrix} \hat{\Sigma}_1 & 0 \\ 0 & \hat{\Sigma}_2 \end{bmatrix} \begin{bmatrix} \hat{V}_1^{\mathrm{T}} \\ \hat{V}_2^{\mathrm{T}} \end{bmatrix}, \qquad (20)$$

where $\hat{\Sigma}_1$ contains the n dominant singular values on the diagonal.

(3) The system matrices are estimated as

$$\hat{A} \triangleq (J_1^q \hat{U}_1)^\dagger J_2^q \hat{U}_1, \qquad (21)$$

$$\hat{C} \triangleq J_3^q \hat{U}_1, \qquad (22)$$

$$\hat{B} \triangleq (I - \hat{A}^{2M}) \hat{\Sigma}_1 \hat{V}_1^{\mathrm{T}} J_4^r, \qquad (23)$$

$$\hat{D} \triangleq \hat{g}_0 - \hat{C}\hat{A}^{2M-1}(I - \hat{A}^{2M})^{-1}\hat{B}, \qquad (24)$$

where

$$J_1^q \triangleq \begin{bmatrix} I_{(q-1)p} & 0_{(q-1)p \times p} \end{bmatrix}, \qquad (25)$$

$$J_2^q \triangleq \begin{bmatrix} 0_{(q-1)p \times p} & I_{(q-1)p} \end{bmatrix}, \qquad (26)$$

$$J_3^q \triangleq \begin{bmatrix} I_p & 0_{p \times (q-1)p} \end{bmatrix}, \qquad (27)$$

$$J_4^r \triangleq \begin{bmatrix} I_m \\ 0_{(r-1)m \times m} \end{bmatrix}. \qquad (28)$$

(4) The resulting transfer function is

$$\hat{G}_{q,r,n,M}(z) \triangleq \hat{D} + \hat{C}(zI - \hat{A})^{-1}\hat{B}. \qquad (29)$$

We have the following result when Algorithm 1 is applied to finite-dimensional systems and data is noiseless, i.e. $e_k = 0$ in (14).

Theorem 1. Let G be a stable system of order n. Assume $q > n$, $r \geq n$ and $2M \geq q + r$. Suppose that $M + 1$ equidistant noise-free frequency-response measurements of G on $[0, \pi]$ are available and let $\hat{G}_{q,r,n,M}$ be given by Algorithm 1. Then

$$\|\hat{G}_{q,r,n,M} - G\|_\infty = 0.$$

Proof. Since G is stable, it can be represented by the following Taylor series

$$G(z) = D + C(zI - A)^{-1}B = D + \sum_{k=1}^{\infty} CA^{k-1}Bz^{-k} \qquad (30)$$

in the complement of the closed unit disk. From (14), (17), and (30) notice that $\hat{g}_k$ can be written as

$$\hat{g}_k = \sum_{i=0}^{\infty} g_{k+2iM} = \begin{cases} CA^{k-1}(I - A^{2M})^{-1}B, & k > 1 \\ D + CA^{2M-1}(I - A^{2M})^{-1}B, & k = 0 \end{cases}, \qquad (31)$$

where we used the identity

$$\sum_{i=0}^{\infty} A^{2iM} = \left(I - A^{2M}\right)^{-1}. \qquad (32)$$

The expression above for $\hat{g}_0$ shows that if $\hat{A}$, $\hat{B}$, and $\hat{C}$ are obtained from A, B, and C by a similarity transformation, then $\hat{D}$ given by (24) equals D.

Next, by introducing the extended observability and controllability matrices

$$\mathcal{O}_q = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{q-1} \end{bmatrix}, \qquad (33)$$

$$\mathcal{C}_r = \begin{bmatrix} B & AB & \cdots & A^{r-1}B \end{bmatrix} \qquad (34)$$

observe that $\hat{H}_{qr}$ can be factored as

$$\hat{H}_{qr} = \mathcal{O}_q (I - A^{2M})^{-1} \mathcal{C}_r \qquad (35)$$

for any realization of G. Minimality of the system implies that both $\mathcal{C}_r$ and $\mathcal{O}_q$ are of rank n, and hence also $\hat{H}_{qr}$ if $r \geq n$ and $q > n$. Then, $\hat{\Sigma}_2 = 0$ in (20)) and the column range spaces of $\hat{H}_{qr}$, $\mathcal{O}_q$ and $\hat{U}_1$ will be equal. Since (35) is valid for any realization, we take the realization which makes $\hat{\mathcal{O}}_q = \hat{U}_1$. Then, in this realization, by utilizing the shift structure of $\hat{\mathcal{O}}_q$, the matrices $\hat{A}$ and $\hat{C}$ are calculated by the formulae (21) and (22), respectively. Furthermore, from (20) and (35), we get

$$\hat{\Sigma}_1 \hat{V}_1^{\mathrm{T}} = (I - \hat{A}^{2M})^{-1} \hat{\mathcal{C}}_r, \tag{36}$$

which gives the formula (23) for $\hat{B}$. ∎

Notice that it suffices to let $q = n+1$ and $r = n$ to meet the requirements on r and q which imply that $M = n+1$, and consequently $n+2$ equidistant samples of the frequency-response function on $[0, \pi]$ are required. Algorithm 1 is in the class of correct algorithms when applied to data from systems of finite dimension and uses a minimum amount of data among all such algorithms. This is a remarkable advantage with respect to black-box identification algorithms using linearly parametrized model sets.

4. CONVERGENCE ANALYSIS FOR NOISE-FREE DATA

In this section, we demonstrate that the transfer function computed from the system matrices of Algorithm 1 will converge to the nth-order balanced truncation of the identified system when the system is Hilbert–Schmidt, its transfer function is continuous, and data is noise-free. This is accomplished in two steps. In the first step of our analysis, the system is approximated by a finite-impulse response model and in the second step, matrix perturbation results are applied.

Let G_n denote the nth-order balanced truncation of G. A state-space realization of G_n is given by the formulae

$$\bar{A} \triangleq U_1^{\mathrm{T}} J_2 U_1, \tag{37}$$

$$\bar{B} \triangleq U_1^{\mathrm{T}} H J_4, \tag{38}$$

$$\bar{C} \triangleq J_3 U_1, \tag{39}$$

$$\bar{D} \triangleq g_0, \tag{40}$$

$$G_n(z) \triangleq \bar{D} + \bar{C}(zI - \bar{A})^{-1}\bar{B}, \tag{41}$$

where

$$U_1 \triangleq [u_1 \ \cdots \ u_n] \tag{42}$$

contains n normalized eigenvectors of $\Gamma\Gamma^*$ corresponding to $\Gamma_1(G), \ldots, \Gamma_n(G)$ assuming $\Gamma_n(G) > \Gamma_{n+1}(G)$, where Γ is the Hankel operator of G and $\Gamma_i(G)$ are the Hankel singular values, H is the Hankel matrix formed from the impulse-response coefficients

$$H \triangleq \begin{bmatrix} g_1 & g_2 & \cdots \\ g_2 & g_3 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}, \tag{43}$$

and J_2, J_3, J_4 are defined as follows:

$$J_2 U_1 \triangleq \begin{bmatrix} u_1(2) & \cdots & u_n(2) \\ u_1(3) & \cdots & u_n(3) \\ \vdots & \ddots & \vdots \end{bmatrix}, \tag{44}$$

$$H J_4 \triangleq \begin{bmatrix} g_1 \\ g_2 \\ \vdots \end{bmatrix}, \tag{45}$$

$$J_3 U_1 \triangleq \begin{bmatrix} u_1(1) & u_2(1) & \cdots \end{bmatrix}. \tag{46}$$

By Hartman's Theorem (Partington, 1988, Theorem 3.20), Γ is compact if and only if G is the projection of a complex-valued function that is continuous on the unit circle into $\mathcal{H}_\infty$. Hence, $g \in \ell_2^{p\times m}$ and since $u_k \in \ell_2^p$ for all k in the Schmidt expansion of a compact operator, the infinite products above converge absolutely by the Cauchy–Schwarz inequality. Thus, $\bar{A}$ and $\bar{B}$ are well defined if Γ is compact. This particular realization differs from the balanced realizations described by Young (1986) or Bonnet (1993) only by a diagonal similarity transformation, i.e. scaling of the state variables.

4.1. *Finite-impulse response approximation*

We will now establish the convergence of the Hankel singular values and Schmidt pairs for compact Hankel operators $\Gamma^{(k)}$ converging to Γ.

Lemma 1. Let $\Gamma^{(k)}$ be a sequence of compact Hankel operators such that $\|\Gamma^{(k)} - \Gamma\| \to 0$, where Γ is the Hankel operator of a system G. Let $\Gamma_i^{(k)}$ and $(v_i^{(k)}, u_i^{(k)})$ denote, respectively, singular values and the Schmidt pairs of $\Gamma^{(k)}$ and $\Gamma_i(G)$ and (v_i, u_i) those of Γ. Suppose that $\Gamma_n(G) > \Gamma_{n+1}(G)$. Let U_1 be as in (42) and $U_2 \triangleq [u_{n+1} \ u_{n+2} \ \cdots \ ]$. Let $U_1^{(k)} \triangleq \left[u_1^{(k)} \ \cdots \ u_n^{(k)}\right]$. Then

(1) $\lim_{k\to\infty} |\Gamma_i^{(k)} - \Gamma_i(G)| = 0$ for all i.

(2) For all sufficiently large k, there exist a sequence of nonsingular matrices $T^{(k)} \in \mathbb{R}^{n\times n}$ and a sequence of semi-infinite matrices $P^{(k)}$ such that $\|P^{(k)}\|_F \to 0$ and

$$U_1^{(k)} = (U_1 + U_2 P^{(k)}) T^{(k)}. \tag{47}$$

Proof. See Appendix A.

Using Lemma 1, we can establish that the sequence of balanced truncations of systems $G^{(k)}$ converges to the balanced truncation of G if the sequence of the associated Hankel operators converges to Γ.

Lemma 2. Let $\Gamma^{(k)}$ be a sequence of compact Hankel operators of systems $G^{(k)}$ such that $\|\Gamma^{(k)}-\Gamma\| \to 0$, where Γ is the Hankel operator of the system G. Let g and $g^{(k)}$ denote the impulse responses of G and $G^{(k)}$, respectively. Assume that $g_0 = g_0^{(k)}$. Let $G_n^{(k)}$ and G_n denote, respectively, nth-order balanced truncations of $G^{(k)}$ and G. Then

$$\lim_{k\to\infty} \|G_n^{(k)} - G_n\|_\infty = 0.$$

Proof. Let $\bar{A}^{(k)}, \bar{B}^{(k)}, \bar{C}^{(k)}, \bar{D}^{(k)}$ be the realization of $G_n^{(k)}$ computed by the formulae (37)–(40).

Given three matrices $A, C \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$ where $n \ge p$, the following inequalities deriving from them will frequently be used.

$$\|A\|_F\ \sigma_p(B) \le \|AB\|_F \le \|A\|_F\ \sigma_1(B), \quad (48)$$

$$\sigma_1(A)\ \sigma_p(B) \le \sigma_1(AB) \le \sigma_1(A)\ \sigma_1(B), \quad (49)$$

$$\max_{1\le i\le n} |\sigma_i(A) - \sigma_i(C)| \le \sigma_1(A-C). \quad (50)$$

See Theorem 3.9 for a proof of (48)–(49) and Theorem 4.11 for (50), both in Stewart and Sun (1990).

From Lemma 1

$$U_1^{(k)} = (U_1 + U_2P^{(k)})T^{(k)} \triangleq \tilde{U}_1^{(k)}T^{(k)},$$

where $\|P^{(k)}\|_F \to 0$. Note that $\|U_i\|_2 = \|U_i^{(k)}\|_2 = 1$, $i = 1,2$; $\|J_i\|_2 = 1$, $i = 1,\cdots,4$; and $\|U_1\|_F \le \sqrt{n}$. Thus

$$\begin{aligned}
&\|((T^{(k)})^{\mathrm T})^{-1}\bar{A}^{(k)}(T^{(k)})^{-1} - \bar{A}\|_F \\
&= \|(\tilde{U}_1^{(k)})^{\mathrm T}J_2\tilde{U}_1^{(k)} - U_1^{\mathrm T}J_2U_1\|_F \\
&= \|(\tilde{U}_1^{(k)})^{\mathrm T}J_2U_2P^{(k)} + (U_2P^{(k)})^{\mathrm T}J_2U_1\|_F \\
&\le 2\ \|P^{(k)}\|_F \to 0, \\
&\|((T^{(k)})^{\mathrm T})^{-1}\bar{B}^{(k)} - \bar{B}\|_F \\
&= \|U_1^{\mathrm T}(H^{(k)} - H)J_4 + (U_2P^{(k)})^{\mathrm T}H^{(k)}J_4\|_F \\
&\le \sqrt{n}\ \|\Gamma^{(k)} - \Gamma\| + \|P\|_F\|\Gamma^{(k)}\| \to 0, \\
&\|\bar{C} - \bar{C}^{(k)}(T^{(k)})^{-1}\|_F = \|J_3U_2P^{(k)}\|_F \\
&\le \|P\|_F \to 0.
\end{aligned}$$

Factor $G^{(k)}$ as

$$\begin{aligned}
G^{(k)} = \bar{C}^{(k)}(T^{(k)})^{-1}\Big(&z(T^{(k)}(T^{(k)})^{\mathrm T})^{-1} \\
&-((T^{(k)})^{\mathrm T})^{-1}\bar{A}^{(k)}(T^{(k)})^{-1}\Big)^{-1} \\
&\times((T^{(k)})^{\mathrm T})^{-1}\bar{B}^{(k)} + g_0^{(k)}.
\end{aligned}$$

Hence to finish the proof, it suffices to show that $T^{(k)}(T^{(k)})^{\mathrm T} \to I_n$. From $(U_1^{(k)})^{\mathrm T}U_1^{(k)} = I_n$

$$(T^{(k)})^{\mathrm T}T^{(k)} = I_n - (P^{(k)}T^{(k)})^{\mathrm T}P^{(k)}T^{(k)} \to I_n \quad \text{as } k\to\infty \quad (51)$$

and thus $T^{(k)}(T^{(k)})^{\mathrm T} \to I_n$.

■

Lemma 1 and Lemma 2 imply only that Γ is a compact Hankel operator, which is fulfilled by the Hilbert–Schmidt condition in Assumption 1 since Hilbert–Schmidt Hankel operators are compact.

We will now use a particular sequence of systems $G^{(k)}$ and relate their truncated balanced realizations to Algorithm 1. Consider the following truncation of g by finite impulse responses

$$G^{(k)} \triangleq \sum_{i=0}^{k-1} g_i z^{-i}. \quad (52)$$

Since

$$\|\Gamma - \Gamma^{(k)}\|^2 \le \sum_{i=1}^{\infty}\Gamma_i^2(G - G^{(k)}) = \sum_{i=k}^{\infty} i\|g_i\|^2 \to 0, \quad (53)$$

it follows from Lemma 2 that

$$\lim_{k\to\infty} \|G_n^{(k)} - G_n\|_\infty = 0. \quad (54)$$

The Hankel matrices of $G^{(k)}$ have only finitely many nonzero elements contained in the following matrix

$$H^{(k)} \triangleq \begin{bmatrix} g_1 & \cdots & g_{k-1} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ g_{k-1} & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 0 \end{bmatrix}. \quad (55)$$

Thus, Hankel singular values of $G^{(k)}$ coincide with the singular values of $H^{(k)}$ and the normalized Schmidt pairs are normalized right and left singular vectors of $H^{(k)}$ extended by zero padding. An nth-order balanced truncation of $G^{(k)}$ can be obtained from the singular value decomposition of $H^{(k)}$

$$H^{(k)} \triangleq \begin{bmatrix} U_1^{(k)} & U_2^{(k)} \end{bmatrix} \begin{bmatrix} \Sigma_1^{(k)} & 0 \\ 0 & \Sigma_2^{(k)} \end{bmatrix} \begin{bmatrix} V_1^{(k)\mathrm T} \\ V_2^{(k)\mathrm T} \end{bmatrix}, \quad (56)$$

where $\Sigma_1^{(k)}$ contains n dominant singular values, as follows:

$$\bar{A}^{(k)} \triangleq (J_1^kU_1^{(k)})^{\mathrm T}J_2^kU_1^{(k)}, \quad (57)$$

$$\bar{C}^{(k)} \triangleq J_3^kU_1^{(k)}, \quad (58)$$

$$\bar{B}^{(k)} \triangleq (U_1^{(k)})^{\mathrm T}H^{(k)}J_4^k, \quad (59)$$

$$\bar{D}^{(k)} \triangleq g_0. \quad (60)$$

Notice that (37) is a special case of (57) for $(J_1^kU_1^{(k)})^{\mathrm T} \to U_1^{\mathrm T}$ as $k \to \infty$.

The system matrices $\hat{A}$ and $\hat{C}$ of Algorithm 1 are calculated exactly by the same formulae as (57)–(59) except the factor $((J_1^q\hat{U}_1)^{\mathrm T}J_1^q\hat{U}_1)^{-1}$ in $\hat{A}$ which tends to I_n. Furthermore, if the magnitudes of the eigenvalues of $\hat{A}$ are bounded away from one for all large M, then $\hat{B}$ will approximately be calculated by (59). Therefore, it can be claimed that when data are noise free, Algorithm 1 will converge to a transfer function described by the realization (57)–(60) if $\hat{H}_{qr}$ tends to $H^{(k)}$ in the Frobenius norm. A couple of conditions on the system and the parameters q, r, M suffice to make $\|\hat{H}_{qr} - H^{(k)}\|_F \to 0$. The proof of this claim will be based on standard matrix perturbation results and is the topic of the next section.

4.2. *Perturbation analysis*

Recall that $\hat{G}_{q,r,n,M}$ denotes the identified transfer function computed from Algorithm 1. We will complete our convergence analysis by showing

$$\lim_{q,r,M\to\infty} \|\hat{G}_{q,r,n,M} - G_n^{(q+r)}\|_\infty = 0$$

when data are noise free and q and r are suitably chosen increasing functions of M. First, we need to derive some bounds on the Frobenius norm of the difference between the Hankel matrix $\hat{H}_{q,r}$ of Algorithm 1 and the Hankel matrix composed of the impulse response of G.

Let $k = q + r$ in (52) and partition (55) as

$$H^{(q+r)} \triangleq \begin{pmatrix} H_{qr} & \Delta_1 \\ \Delta_2 & \Delta_3 \end{pmatrix}, \tag{61}$$

where H_{qr} is q by r-block Hankel matrix formed from impulse-response coefficients

$$H_{qr} \triangleq \begin{pmatrix} g_1 & \cdots & g_r \\ \vdots & \ddots & \vdots \\ g_q & \cdots & g_{q+r-1} \end{pmatrix}. \tag{62}$$

By Assumption 1, we have

$$\sum_{i=1}^{3} \|\Delta_i\|_F^2 < \sum_{i=\min\{q,r\}}^{\infty} i\,\|g_i\|^2,$$

which tends to zero as q and r grow to infinity.

Next observe that $\hat{g}_i$ is a Riemann sum approximation of g_i

$$g_i = \frac{1}{2\pi} \int_0^{2\pi} G(e^{j\theta})\, e^{ji\theta}\, d\theta. \tag{63}$$

Thus, $g_i - \hat{g}_i$ can be bounded as

$$\begin{aligned} \|g_i - \hat{g}_i\| &\le \frac{1}{2M} \sum_{l=0}^{2M-1} \sup_{\frac{\pi l}{M} \le \theta \le \frac{\pi(l+1)}{M}} \|G(e^{j\theta})e^{ji\theta} \\ &\quad - G(e^{j\pi l/M})e^{ji\pi l/M}\| \\ &\le \frac{1}{2M} \sum_{l=0}^{2M-1} \sup_{\frac{\pi l}{M} \le \theta \le \frac{\pi(l+1)}{M}} \|G(e^{j\theta}) \\ &\quad - G(e^{j\pi l/M})\| + \|G\|_\infty \sup_{0\le\theta\le\frac{\pi}{M}} |e^{ji\theta} - 1| \\ &\le \omega_G(\frac{\pi}{M}) + \frac{\pi(q+r)}{M}\, \|G\|_\infty, \end{aligned} \tag{64}$$

where ω_G is the modulus of continuity of $G(e^{j\theta})$. Hence

$$\begin{aligned} \|\hat{H}_{qr} - H_{qr}\|_F^2 &= \sum_{s=1}^{q} \sum_{t=1}^{r} \|\hat{g}_{s+t-1} - g_{s+t-1}\|^2 \\ &\le 2qr\, \omega_G^2(\frac{\pi}{M}) \\ &\quad + 2\pi\, \frac{qr(q+r)^2}{M^2}\, \|G\|_\infty. \end{aligned} \tag{65}$$

If q and r are chosen to satisfy $\sqrt{qr}\,\omega_G(\pi/M) \to 0$ and $\sqrt{qr}\,(q+r)/M \to 0$, then it follows that $\sum_{i=1}^{3} \|\Delta_i\|_F^2 + \|\hat{H}_{qr} - H_{qr}\|_F^2$ will tend to zero as q, r and M jointly tend to infinity.

The next lemma provides perturbation bounds for invariant subspaces of a matrix when matrix dimensions as well as matrix elements are perturbed.

Lemma 3. Let $X^1 \in \mathbb{R}^{m\times p}$ and $X^2 \in \mathbb{R}^{q\times r}$, where $q \ge r$, $q \ge m > n$, and $r \ge p > n$. Partition X^2 as

$$X^2 \triangleq \begin{bmatrix} Z & \Delta_1 \\ \Delta_2 & \Delta_3 \end{bmatrix},$$

where $Z \in \mathbb{R}^{m\times p}$. In the partition, Δ_1 and Δ_3 are omitted if $p = r$ and Δ_2 and Δ_3 if $m = q$. Suppose that $\|X^1 - Z\|_F^2 + \sum_{i=1}^{3} \|\Delta_i\|_F^2 \le \epsilon$. Perform singular value decompositions for X^1 and X^2

$$X^i \triangleq \begin{bmatrix} U_1^i & U_2^i \end{bmatrix} \begin{bmatrix} \Sigma_1^i & 0 \\ 0 & \Sigma_2^i \end{bmatrix} \begin{bmatrix} V_1^{iT} \\ V_2^{iT} \end{bmatrix}, \quad i = 1, 2, \tag{66}$$

where singular values in $\Sigma_1^i \in \mathbb{R}^{n\times n}$ and $\Sigma_2^i \in \mathbb{R}^{(r-n)\times(r-n)}$ are nonincreasing along the diagonals. Assume that

$$4\epsilon < \sigma_n(X^2) - \sigma_{n+1}(X^2). \tag{67}$$

Then, there exists a nonsingular matrix $T \in \mathbb{R}^{n\times n}$ and a matrix $P \in \mathbb{R}^{(r-n)\times n}$, such that

$$U_1^1 = \begin{bmatrix} I_m & 0_{m\times(q-m)} \end{bmatrix} (U_1^2 + U_2^2 P)T \tag{68}$$

and

$$\|P\|_F \le \frac{4\epsilon}{\sigma_n(X^2) - \sigma_{n+1}(X^2)}. \tag{69}$$

Proof. See Appendix B.

4.3. *Convergence result*

Now we use the perturbation results to obtain the following key lemma.

Lemma 4. Suppose that $M + 1$ equidistant noise-free frequency-response measurements (14) of G on

$[0, \pi]$ are available. Let G satisfy Assumption 1. Suppose that

$$\lim_{q,r,M\to\infty} \sqrt{qr}\, \frac{q+r}{M} = 0, \quad (70)$$

$$\lim_{q,r,M\to\infty} \sqrt{qr}\, \omega_G(\frac{\pi}{M}) = 0, \quad (71)$$

where ω_G is the modulus of continuity of G. Let $\hat{G}_{q,r,n,M}$ be given by Algorithm 1. Then

$$\lim_{q,r,M\to\infty} \|\hat{G}_{q,r,n,M} - G_n^{(q+r)}\|_\infty = 0. \quad (72)$$

Proof. Let $k = q+r$. We have $\lim_{k\to\infty} \|\Gamma^{(k)} - \Gamma\| = 0$ by (53). Then, from Lemma 1, $\sigma_n(H^{(k)}) > \Gamma_n(G)/2$ for all large k. Partition $U_1^{(k)}$ and $U_2^{(k)}$ as follows:

$$\begin{bmatrix} U_1^{(k)} & U_2^{(k)} \end{bmatrix} \triangleq \begin{bmatrix} U_{11}^{(k)} & U_{21}^{(k)} \\ U_{12}^{(k)} & U_{22}^{(k)} \end{bmatrix}, \quad (73)$$

where $U_{11}^{(k)} \in \mathbb{R}^{qp\times nm}$. From (61), (56) and (48), we have

$$\begin{aligned} \sigma_n(H^{(k)})\ \|U_{12}^{(k)}\|_F &\le \| \begin{bmatrix} U_{12}^{(k)}\Sigma_1^{(k)} & U_{22}^{(k)}\Sigma_2^{(k)} \end{bmatrix} \\ &\quad \times \begin{bmatrix} V_1^{(k)} & V_2^{(k)} \end{bmatrix}^{\mathrm{T}} \|_F \\ &= \| [\Delta_2\ \Delta_3] \|_F \to 0, \end{aligned} \quad (74)$$

which shows that $\|U_{12}^{(k)}\|_F \to 0$. Let $x_q^{(k)}$ denote the last block row of $U_{11}^{(k)}$. Then, by the Hilbert–Schmidt assumption on the system, we have $\|x_q^{(k)}\|_F \to 0$ from

$$\begin{aligned} \sigma_n(H^{(k)})\ \|x_q^{(k)}\|_F &\le \| \begin{bmatrix} x_q^{(k)}\Sigma_1^{(k)} & y\Sigma_2^{(k)} \end{bmatrix} \\ &\quad \times \begin{bmatrix} V_1^{(k)} & V_2^{(k)} \end{bmatrix}^{\mathrm{T}} \|_F \\ &= \| \begin{bmatrix} g_q & \cdots & g_{q+r} & 0 & \cdots & 0 \end{bmatrix} \|_F \\ &\to 0, \end{aligned} \quad (75)$$

where y is the last block row of $U_{21}^{(k)}$. Next, we have from Lemma 3 and (73)

$$\begin{aligned} \hat{U}_1 &= \begin{bmatrix} I_{qp} & 0_{qp\times rp} \end{bmatrix} (U_1^{(k)} + U_2^{(k)}P^{(k)})T^{(k)} \\ &= (U_{11}^{(k)} + U_{21}^{(k)}P^{(k)})T^{(k)} \end{aligned}$$

for some nonsingular matrix $T^{(k)} \in \mathbb{R}^{n\times n}$ and $P^{(k)}$ which tends to zero in the Frobenius norm as $q, r, M \to \infty$ by (71) and the Hilbert–Schmidt assumption on the system. Notice that $(T^{(k)})^{\mathrm{T}}T^{(k)} \to I_n$ which implies $T^{(k)}(T^{(k)})^{\mathrm{T}} \to I_n$. To see this, write

$$I_n = (T^{(k)})^{\mathrm{T}}(I_n + X)T^{(k)},$$

where

$$\begin{aligned} \|X\|_F = \| &- (U_{12}^{(k)})^{\mathrm{T}}U_{12}^{(k)} + (U_{11}^{(k)})^{\mathrm{T}}U_{21}^{(k)}P^{(k)} \\ &+ (U_{21}^{(k)}P^{(k)})^{\mathrm{T}}U_{11}^{(k)} \\ &+ (U_{21}^{(k)}P^{(k)})^{\mathrm{T}}U_{21}^{(k)}P^{(k)} \|_F \to 0. \end{aligned}$$

Then

$$\begin{aligned} (J_1^q\hat{U}_1)^{\mathrm{T}}J_1^q\hat{U}_1 &= (T^{(k)})^{\mathrm{T}}(I_n - (x_q^{(k)})^{\mathrm{T}}x_q^{(k)} \\ &\quad + o(1))T^{(k)} \\ &= I_n + o(1), \\ (J_1^q\hat{U}_1)^{\mathrm{T}}J_2^q\hat{U}_1 &= (T^{(k)})^{\mathrm{T}}((J_1^kU_1^{(k)})^{\mathrm{T}}J_2^{(k)}U_1^{(k)} \\ &\quad + o(1))T^{(k)} \\ &= (T^{(k)})^{\mathrm{T}}\bar{A}^{(k)}T^{(k)} + o(1). \end{aligned}$$

It follows that

$$\hat{A} = (T^{(k)})^{\mathrm{T}}\bar{A}^{(k)}T^{(k)} + o(1),$$

where $(\bar{A}^{(k)}, \bar{B}^{(k)}, \bar{C}^{(k)}, \bar{D}^{(k)})$ is the realization of $G^{(k)}$ in (57)–(60). We derive the following expressions for the other matrices:

$$\begin{aligned} \hat{B} &= \hat{U}_1^{\mathrm{T}}\hat{H}_{qr}J_4^r - \hat{A}^{2M}\hat{u}_1^{\mathrm{T}}\hat{H}_{qr}J_4^r \\ &= (T^{(k)})^{\mathrm{T}}(U_{11}^{(k)})^{\mathrm{T}}H_{qr}J_4^r + o(1) \\ &= (T^{(k)})^{\mathrm{T}}(U_1^{(k)})^{\mathrm{T}}H^{(k)}J_4^k + o(1) \\ &= (T^{(k)})^{\mathrm{T}}\bar{B}^{(k)} + o(1), \\ \hat{C} &= J_3^q\hat{U}_1 = J_3^kU_1^{(k)}T^{(k)} + o(1) = \bar{C}^{(k)}T^{(k)} + o(1), \\ \hat{D} &= \hat{g}_0 + o(1) = g_0 + o(1) = g_0^{(k)} + o(1). \end{aligned}$$

Finally

$$\begin{aligned} &\hat{C}(zI_n - \hat{A})^{-1}\hat{B} + \hat{D} \\ &\quad = \bar{C}^{(k)}T^{(k)}\left(zI_n - (T^{(k)})^{\mathrm{T}}\bar{A}^{(k)}T^{(k)}\right)^{-1} \\ &\quad \times (T^{(k)})^{\mathrm{T}}\bar{B}^{(k)} + g_0^{(k)} + o(1) \\ &\quad = \bar{C}^{(k)}\left((T^{(k)}(T^{(k)})^{\mathrm{T}})^{-1} - \bar{A}^{(k)}\right)^{-1} \\ &\quad + g_0^{(k)} + o(1) \\ &\quad = G^{(k)} + o(1). \end{aligned}$$

■

By combining the results in Lemma 2 and Lemma 4 with the triangle inequality we obtain the main result of this section.

Theorem 2. Let G be a linear system satisfying Assumption 1. Let ω_G from (12) be the modulus of continuity of G and assume q, r satisfy the conditions (70) and (71). Let G_n be the balanced truncation of G of order n. Let $\hat{G}_{q,r,n,M}$ be given by Algorithm 1 using $M+1$ noise-free frequency-response measurements (14) of G equidistantly spaced on $[0, \pi]$. Then

$$\lim_{q,r,M\to\infty} \|\hat{G}_{q,r,n,M} - G_n\|_\infty = 0. \quad (76)$$

In the rest of this section, we will briefly discuss the class of systems considered in this paper and the convergence conditions.

The Hilbert–Schmidt assumption on the system merely implies that $\|g_k\| = o(1/\sqrt{k})$. The set of Hilbert–Schmidt systems is not contained in $\mathcal{H}_\infty$ (Duren, 1970, Exercise 6–7 in Chapter 6). On the other hand, even if $\ell_1^{p\times m}$ is not contained in the set of Hilbert–Schmidt system. For example, the system described by

$$g_k = \begin{cases} \frac{1}{\sqrt{k}}, & \text{for } k = 1, 2^4, 3^4 \cdots, \\ 0, & \text{otherwise,} \end{cases}$$

is in ℓ_1 but not in the set of Hilbert–Schmidt systems since $\sum_{k=1}^{\infty} k g_k^2 = \infty$. We can generate many interesting examples considering sequences which tend to zero extremely slowly and the gap between nonzero elements are arbitrarily large. These examples clearly illustrate that Assumption 1, imposed on the identified system is rather weak and satisfied by some systems with frequency responses characterized by a modulus of continuity tending to zero extremely slowly.

Recall the condition $\sqrt{qr}\ \omega_G(\pi/M) \to 0$. This condition implies that for the convergence result to hold the number of data must increase faster than the size of the Hankel matrix at a rate determined by the modulus of continuity.

To appreciate this condition, consider now the class of systems characterized by their impulse-response decay rates $\|g_k\| = O(k^{-\alpha})$. If $\alpha > 1$, such systems are Hilbert–Schmidt and in ℓ_1. The modulus of continuity of a system in this class is estimated by the following lemma.

Lemma 5. Assume that $\|g_k\| = O(k^{-\alpha})$ for some $\alpha > 1$. Then, $G(e^{j\theta}) \in \Lambda_{\min(2,\alpha)-1}$.

Proof. Assume $\alpha \le 2$. For some constants c_i, we have

$$\sum_{k=1}^{N} k\|g_k\| \le c_1 \sum_{k=1}^{N} k^{1-\alpha} \le c_2 \int_1^N x^{1-\alpha}\,\mathrm{d}x = O(N^{2-\alpha}).$$

Thus, $G(e^{j\theta}) \in \Lambda_{\alpha-1}$ (Duren, 1970, Exercise 1 in Chapter 5).

■

Hence, for this class, we have a convergence requirement

$$qr = o(M^{2\alpha-2}) \quad \text{for} \quad 1 < \alpha \le 2. \tag{77}$$

This requirement drops out for $\alpha > 2$ since we already have $M > q, r$. Lemma 5 is sharp. Thus, as α gets closer to one, more and more data are required for the convergence to take place.

The condition (70) becomes $q = o(M^{1/2})$ if $r = O(q)$ and for $\alpha > 3/2$, (77) reads off $q = o(M^{\alpha-1})$. Therefore, with the choice $q = O(r)$, we observe that q must satisfy $q = o(M^{1/2})$ if $\alpha > 3/2$ and $q = o(M^{\alpha-1})$ if $1 < \alpha < 3/2$. Recall that if $\alpha > 3/2$, the system is nuclear. Hence, for nuclear systems characterized by $\alpha > 3/2$, the only convergence requirement is $q = o(M^{1/2})$ if $q = O(r)$.

5. CONSISTENCY ANALYSIS

In this section, we show that Algorithm 1 is strongly consistent. The consistency proof will be performed via two lemmas. First using our convergence results in Section 4, we will show that Algorithm 1 is strongly consistent provided that the system Hankel matrix is consistently estimated. Next, we derive further size conditions on the system Hankel matrix to obtain consistency. To illustrate the trade-offs, the results are applied to the systems (9). At the end of this section, two other related consistent algorithms are briefly discussed.

The matrix $\hat{H}_{qr}$ in (19) is a linear function of the noisy data G_k. Since the noise term e_k in (14) is additive in G_k, it will be additive in $\hat{H}_{qr}$. Let E_{qr} denote the part of the Hankel matrix (19) that has its origin in e_k in (14). Then, E_{qr} is given by

$$E_{qr} \triangleq \begin{bmatrix} \hat{e}_1 & \cdots & \hat{e}_r \\ \vdots & \ddots & \vdots \\ \hat{e}_q & \cdots & \hat{e}_{q+r-1} \end{bmatrix}, \tag{78}$$

where

$$e_{M+k} \triangleq e_{M-k}^*, \quad k = 1, \cdots, M-1,$$

$$\hat{e}_i \triangleq \frac{1}{2M} \sum_{k=0}^{2M-1} e_k\, e^{j2\pi ik/2M}, \quad i = 0, \cdots, q+r-1.$$

The frequency-response noise e_k, $k = 0, \cdots, M$ is assumed to be independent, zero-mean complex random variables with a bounded covariance function. For more information on complex noise models, see Brillinger (1981) and Schoukens and Pintelon (1991).

Lemma 6. Let the assumptions in Theorem 2 hold with the addition that the data are noisy, i.e. $e_k \neq 0$. Let E_{qr} be given by (78) and let $\hat{G}_{q,r,n,M}$ be given by Algorithm 1. Then

$$\lim_{q,r,M\to\infty} \|\hat{G}_{q,r,n,M} - G_n\|_\infty = 0, \quad \text{w.p. } 1 \tag{79}$$

if

$$\lim_{q,r,M\to\infty} \|E_{qr}\|_F = 0, \quad \text{w.p. } 1. \tag{80}$$

Proof. See Appendix C.

Lemma 7. Let E_{qr} be given by (78), where e_k, satisfy Assumption 2. Let r and q increase to infinity at rates at most $O(\sqrt{M}\,(\log M)^{-\beta})$ for some $\beta > 1/2$. Then

$$\lim_{q,r,M\to\infty} \|E_{qr}\|_F = 0, \quad \text{w.p. } 1. \tag{81}$$

Proof. Without loss of generality, we may assume that the system is single-input/single-output. Let

$w = 2M$ and $\phi(l+m) = 2\pi(l+m-1)$. The lm-th element of E_{qr} is then

$$\left[E_{qr}\right]_{lm} = \frac{1}{w}\sum_{k=0}^{w-1} e_k e^{j\phi(l+m)k/w} \triangleq \frac{1}{w}S_w(l+m). \quad (82)$$

We have by (82)

$$E\,|S_w(l+m)|^2 = \sum_{k=0}^{w-1}\sum_{l=0}^{w-1} E(e_k e_l^*)\, e^{j(k-l)\phi(l+m)/w} = \sum_{k=0}^{w-1} E|e_k|^2 \le \bar{R}w, \quad (83)$$

where we used the following relation for the conjugate symmetric complex noise

$E(e_{M-k}e^*_{M+k}) = E(e^2_{M-k}) = 0,\ k = 1, \cdots, M-1.$

Hence, for each $\epsilon > 0$ by Chebyshev's inequality (Chung, 1974)

$$P\left(\|E_{qr}\|_F > \epsilon\right) \le \frac{1}{\epsilon^2}E\left\|E_{qr}\right\|_F^2 = \frac{1}{w^2\epsilon^2}\sum_{l=1}^{q}\sum_{m=1}^{r} E\,|S_w(l+m)|^2 \le \frac{qr\bar{R}}{w\epsilon^2}. \quad (84)$$

Taking a subsequence S_{w^2}, we get

$$\sum_w P\left(\|S_{w^2}\|_F > w^2\epsilon\right) = \sum_w \frac{qr\bar{R}}{w^2\epsilon^2} < \infty \quad (85)$$

since q and r are at most $O(\sqrt{M}\,(\log M)^{-\beta})$, $\beta > 1/2$. Hence, by Borel–Cantelli's Lemma (Chung, 1974), we have

$$\frac{\|S_{w^2}\|_F}{w^2} \to 0, \quad \text{w.p. 1 as } w \to \infty. \quad (86)$$

We have thus proved the desired result for a subsequence. Now it will be extended to the whole sequence. For $w^2 < k < (w+1)^2$ from

$$S_k(l+m) - S_{w^2}(l+m) = \sum_{i=0}^{w^2-1} e_i\left[e^{j\phi(l+m)i/k} - e^{j\phi(l+m)i/w^2}\right] + \sum_{i=w^2}^{k-1} e_i e^{j\phi(l+m)i/k},$$

we have

$$E\,|S_k(l+m) - S_{w^2}(l+m)|^2 \le 2E\left|\sum_{i=0}^{w^2-1} e_i\left[e^{j\phi(l+m)i/k} - e^{j\phi(l+m)i/w^2}\right]\right|^2 + 2E\left|\sum_{i=w^2}^{k-1} e_i e^{j\phi(l+m)i/k}\right|^2 \le 8\bar{R}w^2\left(\sin\frac{\phi(l+m)}{w}\right)^2 + 4\bar{R}w \le 4\bar{R}\left[8\pi^2(q+r)^2 + w\right].$$

Let

$$D_w(l+m) \triangleq \max_{w^2<k<(w+1)^2} |S_k(l+m) - S_{w^2}(l+m)|. \quad (87)$$

Then we have

$$E\left[D_w^2(l+m)\right] \le E\left[\sum_{k=w^2+1}^{(w+1)^2-1} |S_k(l+m) - S_{w^2}(l+m)|^2\right] \le 8\bar{R}\left[8\pi^2(q+r)^2 + w\right]w$$

and consequently applying Chebyshev inequality, we obtain that

$$P(\|D_w\|_F > w^2\epsilon) \le \sum_{l=1}^{q}\sum_{m=1}^{r}\frac{E\left[D_w^2(l+m)\right]}{w^4\epsilon^2} \le \frac{8\bar{R}qr\left[8\pi^2(q+r)^2 + w\right]}{w^3\epsilon^2}.$$

It follows as before that

$$\frac{\|D_w\|_F}{w^2} \to 0, \quad \text{w.p. 1 as } w \to \infty. \quad (88)$$

For $w^2 < k < (w+1)^2$ from the following inequality

$$\frac{|S_k(l+m)|}{k} \le \frac{|S_{w^2}(l+m)| + D_w(l+m)}{w^2}, \quad (89)$$

we get

$$\frac{\|S_w\|_F}{w} \le \frac{\|\sqrt{2}S_{w^2}\|_F}{w^2} + \frac{\|\sqrt{2}D_w\|_F}{w^2}. \quad (90)$$

Hence, from (86) and (88), (80) follows. ■

It is possible to get rates faster than $O(\sqrt{M})$ for q and r in the lemma if higher-order moments of the noise are bounded. In particular, if the noise is uniformly bounded, the rate $o(M)$ for q and r is attained. When $q = O(r)$, however, by (70), the growth rates of q and r are limited by $o(M^{1/2})$. Then, (70) is satisfied with $q = O(\sqrt{M}\,(\log M)^{-\beta})$ for some $\beta > 1/2$.

Combining Theorem 2, Lemma 6 and Lemma 7, we obtain the following main result of the paper.

Theorem 3. Let G be a linear system satisfying Assumption 1. Let ω_G from (12) be the modulus of continuity of G. Assume q, r satisfy the condition (71) and be at most $O(\sqrt{M}\,(\log M)^{-\beta})$ for some $\beta > 1/2$. Let G_n be the balanced truncation of G of order n. Let $\hat{G}_{q,r,n,M}$ be given by Algorithm 1 using $M+1$ frequency-response measurements of G equidistantly spaced on $[0, \pi]$. Let the measurement noise e_k in (14) satisfy Assumption 2. Then

$$\lim_{q,r,M\to\infty} \|\hat{G}_{q,r,n,M} - G_n\|_\infty = 0, \quad \text{w.p. 1.} \quad (91)$$

Furthermore if the Hankel operator of G is nuclear, i.e. satisfies (8), then

$$\lim_{q,r,M\to\infty} \|\hat{G}_{q,r,n,M} - G\|_\infty \le 2 \sum_{k=n+1}^{\infty} \Gamma_k(G) \quad \text{w.p. } 1, \tag{92}$$

where repeated singular values are omitted in the sum.

Theorem 3 shows that any Hilbert–Schmidt system with a continuous transfer function can be identified consistently if suitable rates for q and r are chosen. Note the two competing rates: almost square root $\sqrt{M}\,(\log M)^{-1/2}$ and $(\omega_G(\pi/M))^{-1}$ when $r = O(q)$. The former is related to measurement noise and the latter to approximation errors caused by aliasing.

When applied to the class of systems (9), Theorem 3 yields the following important result.

Corollary 1. Let G be a linear system satisfying Assumption 1. Assume that the impulse response of G satisfies $\|g_k\| = O(k^{-\alpha})$ for some $\alpha > 1$. Let q and r be at most $O(\sqrt{M}\,(\log M)^{-\beta})$ for some $\beta > 1/2$ and satisfy $qr = o(M^{2\min(\alpha,2)-2})$. Let $\hat{G}_{q,r,n,M}$ be given by Algorithm 1 using $M + 1$ frequency-response measurements of G equidistantly spaced on $[0, \pi]$. Let the measurement noise e_k in (14) satisfy Assumption 2. Let G_n denote the balanced truncation of G. Then

$$\lim_{q,r,M\to\infty} \|\hat{G}_{q,r,n,M} - G_n\|_\infty = 0, \quad \text{w.p. } 1. \tag{93}$$

Furthermore, if $\alpha > 3/2$ then

$$\lim_{q,r,M\to\infty} \|\hat{G}_{q,r,n,M} - G\|_\infty \le 2 \sum_{k=n+1}^{\infty} \Gamma_k(G) \quad \text{w.p. } 1, \tag{94}$$

where repeated singular values are omitted in the sum.

Assuming $q = O(r)$, then the consistency condition in Corollary 1 becomes

$$q = \begin{cases} o(M^{\alpha-1}), & \text{if } \alpha < 3/2, \\ O(\sqrt{M}\,(\log M)^{-\beta});\ \beta > 1/2, & \text{if } \alpha \ge 3/2. \end{cases}$$

Hence, if $\alpha < 3/2$, rates for q and r depend on the smoothness of the system impulse response. For the nuclear systems characterized by $\alpha > 3/2$, rates are determined by approximation errors caused by the measurement noise.

5.1. *Two related consistent algorithms*

We introduce two more algorithms as follows.

Let $\hat{A}$, $\hat{C}$ be calculated as in Algorithm 1, $\hat{B}$ from

$$\hat{B} = \hat{\Sigma}_1 \hat{V}_1^{\mathrm{T}} \begin{bmatrix} I_m \\ 0_{(r-1)m\times m} \end{bmatrix} \tag{95}$$

and $\hat{D} = \hat{g}_0$. This algorithm, which we call Algorithm 2, was studied in a modal analysis context by Juang and Suziki (1988). It is a biased algorithm. Indeed, Example 1 illustrates poor performance of Algorithm 2 on real data of finite length when it is applied to lightly damped systems. However, the bias term vanishes asymptotically and the algorithm yields truncated balanced realizations of the identified system under the system and noise assumptions of Theorem 3.

In the third algorithm, which we call Algorithm 3, the impulse-response coefficients g_i are estimated as in Algorithm 1 and a *pre-identified* model is calculated by

$$\hat{G}_{pi} \triangleq \sum_{i=0}^{k} \hat{g}_i z^{-i}. \tag{96}$$

The nth-order identified model is obtained by model reduction by balanced truncation from $\hat{G}_{pi}$. Algorithm 3 is a special case of the *two-stage* algorithms outlined by Gu and Khargonekar (1992), where the pre-identified model structure is allowed to be infinite-impulse response to counter possible worst-case noise. Owing to finite-impulse response structure, $\hat{G}_{pi}$ can be reduced by a recursively implemented balanced truncation technique. Interestingly, Algorithm 1 contains Algorithm 3 as a special case. For when applied to the Hankel matrix

$$\hat{H}' = \begin{bmatrix} \hat{g}_1 & \cdots & \hat{g}_k & 0 \\ \vdots & \ddots & \vdots & \vdots \\ \hat{g}_k & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 0 \end{bmatrix}.$$

Algorithm 1 yields the nth-order balanced truncation of $\hat{G}_{pi}$. Thus, Algorithm 3 is also consistent under the assumptions of Theorem 3 though it is biased for finite data sets.

The bias error of Algorithm 3 has two components: the first-stage error $\|G - \hat{G}_{pi}\|_\infty$ and the approximation error $\|\hat{G}_{pi} - \hat{G}\|_\infty$. The total error is bounded above by the sum of $\|G - \hat{G}\|_\infty$ and $\|G - \hat{G}_{pi}\|_\infty$. In the same example in Section 8, Algorithm 3 performs poorly on the same data due to large approximation errors. The example demonstrates that in the choice of a potential identification algorithm, the *posterior* error caused by model reduction and correctness in addition to asymptotic properties must be taken into account.

6. IDENTIFICATION OF CONTINUOUS-TIME SYSTEMS

Most processes subject to modeling are of continuous-time character. However, measured input/output signals are almost without exception sampled. We can distinguish between two different modeling goals. Either a discrete-time model is sought which accurately describes the sampled properties of the system or a continuous-time model is desired which accurately describes the true continuous-time input/output properties of the system.

For the case of discrete-time modeling the results in this paper are directly applicable if we see the zero-order hold (ZOH) equivalence of the (infinite-dimensional) continuous-time system as an infinite-dimensional discrete-time system. Frequency-response data, for this modeling philosophy, can accurately be found using periodic excitation and the discrete Fourier transform.

If a continuous-time model is required, two different approaches can be identified. If the dynamics of the system (poles and zeros) are below the Nyquist frequency, a continuous-time model can be found by using the inverse zero-order hold sampling. In this case, first, a discrete-time model is identified from data and, secondly, a continuous-time model is obtained by inverse ZOH sampling. A second approach is to excite the system with a periodic input signal which is band-limited, i.e. no signal power above the Nyquist frequency. In this case the discrete Fourier transform of the sampled input–output data $(U(\theta_k), Y(\theta_k))$ satisfies the simple relation (for normalized frequencies), see Pintelon *et al.* (1994a)

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

Hence, samples of the continuous-time transfer function are directly obtained from the Fourier transform of one period of the periodic signals. A third (but expensive) way of obtaining samples of the continuous transfer function is to use stepped sine excitation, i.e. for each frequency a sinusoid is applied and amplitude and phase are measured. If samples of the continuous transfer function are at hand, the results of this paper again hold if the data are transformed to discrete-time by the bilinear transformation $z = \frac{2+sT}{2-sT}$. A discrete-time model is estimated and converted back to continuous time by the inverse bilinear transformation. The transformation of data from continuous time to discrete time is equal to the mapping of continuous-time frequencies θ^c to discrete-time frequencies θ according to $\tan(\theta/2) = \theta^c T/2$. Equidistant spacing of discrete-time frequencies requires nonequidistant spacing of the continuous-time frequencies. However, the spacing required is approximately logarithmic which is often used when describing continuous-time transfer functions.

7. ALGORITHMIC ASPECTS

7.1. *Computational complexity*

When facing a practical identification problem, many models of different orders are estimated and compared in order to find a suitable "best" model. In the presented algorithm, most of the computational effort lies in the SVD factorization (20). Given the factorization (20), all models of order less than q are easily obtained from the rest of the algorithm by letting the model order 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 low computational cost.

In the choice of model order, the size of the singular values of $\hat{H}_{qr}$ provides some useful information. They converge to the Hankel singular values of the system, see the Proof of Lemma 6. The sum of the neglected singular values in $\hat{\Sigma}_2$ together with (13) will thus give an indication of the magnitude of the approximation error.

7.2. *Stability issues*

Stability is often a desirable feature of the estimated model. Algorithm 1 does not guarantee stability of the estimated models when using a finite number of noisy frequency data. However, stability can be ensured by adding an extra projection step after (21). In this step all unstable eigenvalues of $\hat{A}$ are projected into the unit circle. This idea can be implemented in the following way:

- Transform $\hat{A}$ to the diagonal form (or complex Schur form if $\hat{A}$ is defective), with the eigenvalues λ_i on the diagonal.
- Project any diagonal elements (eigenvalues) satisfying $1 < |\lambda_i| \leq 2$, into the unit disc by $\lambda_i' \triangleq \lambda_i(\frac{2}{|\lambda_i|} - 1)$. 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. $\lambda_i' \triangleq \lambda_i(1-\epsilon)$.
- 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 change the consistency of the algorithm 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.

A stable $\hat{A}$ (all eigenvalues inside the unit circle) can be also guaranteed by the following procedure (Maciejowski, 1995):

$$\hat{A} = \hat{U}_1^{\dagger} \begin{bmatrix} J_2^q \hat{U}_1 \\ 0_{p \times n} \end{bmatrix}.$$

Of course, this step should only be applied when the original $\hat{A}$ is unstable since it gives a biased estimate.

8. IDENTIFICATION EXAMPLES

The algorithm outlined in the previous section differs from McKelvey and Akçay (1994), which we call Algorithm 4, only in the calculation of $\hat{B}$ and $\hat{D}$:

$$\hat{B}, \hat{D} = \arg \min_{\substack{B \in \mathbf{R}^{n \times m} \\ D \in \mathbf{R}^{p \times m}}} \sum_{i=0}^{2M-1} \left\| G_i - \tilde{G}(e^{j\pi i/M}) \right\|_F^2, \quad (97)$$

where

$$\tilde{G}(z) = D + \hat{C}(zI - \hat{A})^{-1}B. \quad (98)$$

Algorithm 4 is also correct and needs only minimal data when restricted to finite-dimensional systems. In Algorithm 1, $\hat{B}$ and $\hat{D}$ were modified to obtain truncation errors in (16) while maintaining correctness of Algorithm 2 over finite-dimensional systems.

The least-squares procedure (97) to estimate $\hat{B}$ and $\hat{D}$ is a particular case of the nonlinear least-squares identification algorithm (NLS) where $\hat{A}$ and $\hat{C}$ also are estimated. The NLS is not suitable for narrow band data if model fit is measured in the $\mathcal{H}_\infty$-norm. To reduce model mismatch, model orders should be increased. As this happens, pole-zero sensitivity of the model increases. Example 1 of this section illustrates a model error fluctuation at high orders for the NLS. Since Algorithm 1 and Algorithm 4 yield identical asymptotic poles, the asymptotic performance of Algorithm 1 should be expected between the NLS and Algorithm 1.

Example 1. We will use a real data set obtained at the Jet Propulsion Laboratory, Pasadena, California. The data origin from a frequency-response experiment on a flexible structure. The JPL-data consist of a total of $M = 512$ complex frequency samples in the frequency range [1.23, 628] and have several lightly damped modes. The discrete-time models matching the given frequency response were constructed applying zero-order hold sampling equivalence and five algorithms. In the data set the static gain, G_0, is missing and was estimated by visual inspection of the transfer function to be -0.04. The size of the Hankel matrix $\hat{H}_{qr}$ is taken as $q = r =$ 512. In Fig. 1, the error between the predicted, $\hat{G}(e^{j\theta_k})$, and measured, G_k, frequency responses

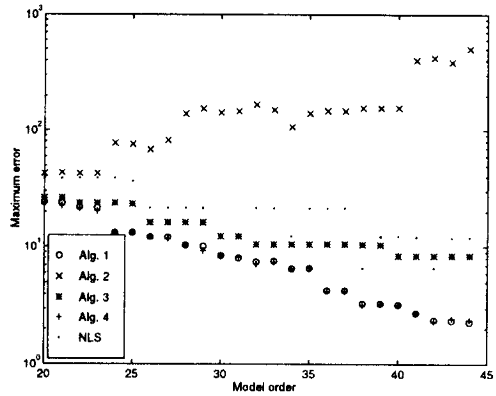

Fig. 1. Plot of $\|G - \hat{G}\|_{m,\infty}$ for different model orders and algorithms in Example 1.

$$\|G - \hat{G}\|_{m,\infty} \triangleq \max_k |G_k - \hat{G}(e^{j\theta_k})|$$

is plotted for different estimated models with orders ranging from 20 to 44. In the estimation, Algorithms 1–4 and the NLS are used. Notice the bias for Algorithm 2 and the fluctuation of the NLS estimates at high orders. This example clearly shows the relevance of using either Algorithm 1 or Algorithm 4 to estimate B and D. The three subspace algorithms delivered stable models for all estimated model orders. If the model order is further increased some eigenvalues of $\hat{A}$ matrix move outside the unit disc and the models become unstable.

A two-stage nonlinear identification algorithm outlined by Gu and Khargonekar (1992) was tested on the JPL-data by Friedman and Khargonekar (1995). In Friedman and Khargonekar (1995), the second stage of the algorithm was eliminated due to the numerical difficulty in reducing very high-order rational systems. Thus the resulting algorithm is Algorithm 3 which is a linear, black-box type. The pre-identified model had a finite impulse response represented by 1024 coefficients and was reduced by a recursively implemented model reduction procedure. With this choice of model order, the data are entirely explained by the model. For a comparison, we included the results obtained by Friedman and Khargonekar (1995) in Fig. 1 (Algorithm 3). 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.

Example 2. We will consider the problem of approximating the infinite-dimensional transfer function

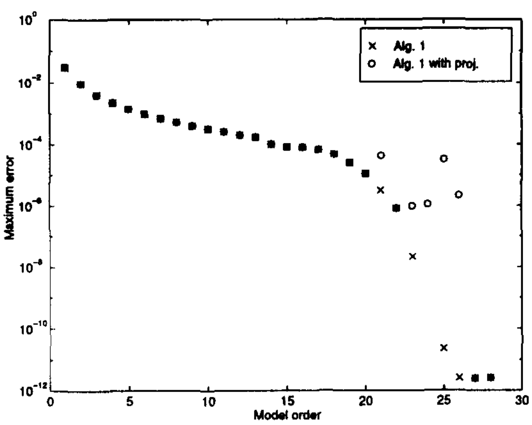

Fig. 2. Plot of $\|G - \hat{G}\|_{m,\infty}$ for different model orders in Example 2 using Algorithm 1 with "x" and without "o" projecting unstable eigenvalues of $\hat{A}$ into the unit disc.

$$G(s) = \frac{1}{s + 1 - e^{-2-s}} \tag{99}$$

with a finite-dimensional linear model. This particular problem has also been studied by Gu *et al.* (1989). As in Gu *et al.* (1989), we use 512 uniformly spaced noise-free frequency-response data on $[0, \pi]$ derived from (99) by use of the bilinear map. In Fig. 2, $\|G-\hat{G}\|_{m,\infty}$ is plotted for model orders ranging from 1 to 28 and Algorithm 1 shown by "x" on the figure and Algorithm 1 with the added projection of all unstable eigenvalues into the unit disc as discussed in Section 7 shown by "o". Here we take $q = r = 512$ which gives the maximal size Hankel matrix. From the figure we notice a deviation at model orders 21, 23, 24, 25, and 26 which are due to the unstable initial models which after the projection give an increased model error. The first-order approximation has the error 3.1×10^{-2} to be compared with the first-order model in Gu *et al.* (1989) with error 3.2×10^{-2}. The error is reduced by increasing the model order. The 24th-order stable approximation of Algorithm 1 has a quite small error 1.4×10^{-6} to be compared with the 24th-order approximant obtained by Gu *et al.* (1989) with error 7.9×10^{-3}. A further increase of the model order to $n = 27$ gives the almost negligible error 2.4×10^{-12}.

9. CONCLUSIONS

In this paper, we presented a correct, frequency-domain subspace-based identification algorithm yielding w.p. 1 a state-space model with a transfer function equal to the balanced truncation of the identified system under a range of conditions on the measurement errors and the parameters q and r of the Hankel matrix H_{qr}. For practical use, an extension of the algorithms are also outlined which guarantees stability of the estimated models. Two examples were used to illustrate the properties of different algorithms and show the practical applicability of the algorithms.

Acknowledgements— This work was supported in part by the Swedish Research Council for Engineering Sciences (TFR), which is gratefully acknowledged. The authors would like to thank Dr D. S. Bayard at the Jet Propulsion Laboratory, Pasadena, California who provided the experimental data used in Example 1.

REFERENCES

Bayard, D. S. 1994). High-order multivariable transfer function curve fitting: Algorithms, sparse matrix methods and experimental results. *Automatica*, **30**, 1439–1444.

Bonnet, C. (1993). Convergence and convergence rate of the balanced realization truncations for infinite-dimensional discrete-time systems. *Syst. & Control Lett.*, **20**, 353–359.

Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory.* McGraw-Hill Inc., New York.

Chung, K. L. (1974). *A Course in Probability Theory.* Academic Press, San Diego, CA.

Curtain, R. F. (1985). Sufficient conditions for infinite-rank Hankel operators to be nuclear. *J. Math. Control and Information*, **2**, 171–180.

Duren, P. L. (1970). *Theory of H^p Spaces.* Academic Press, New York and London.

Friedman, J. H. and P. P. Khargonekar (1995). Application of identification in $\mathcal{H}_\infty$ to lightly damped systems: Two case studies. *IEEE Trans. on Control Systems Technology*, **3**, 279–289.

Glover, K., R. F. Curtain and J. R. Partington (1988). Realizations and approximations of infinite-dimensional systems with error bounds. *SIAM J. Control and Optimization*, **26**, 863–898.

Golub, G. H. and C. F. Van Loan (1989). *Matrix Computations*, second edition. The John Hopkins University Press, Baltimore, MD.

Gu, G. and P. P. Khargonekar (1993). Frequency domain identification of lightly damped systems. In *Proc. Amer. Control Conf.*, San Francisco, CA, pp. 3052–3056.

Gu, G. and P. P. Khargonekar (1992). A class of algorithms for identification in $\mathcal{H}_\infty$. *Automatica*, **37**, 299–312.

Gu, G., P. P. Khargonekar and E. B. Lee (1989). Approximation of infinite-dimensional systems. *IEEE Trans. Automat. Control*, **34**, 610–618.

Guo, L., D. W. Huang and E. J. Hannan (1990). On ARX(∞) approximation. *J. Multivariate Analysis*, **32**, 17–47.

Helmicki, A. J., C. A. Jacobson and C. N. Nett (1991). Control-oriented system identification: A worst-case/deterministic approach in $\mathcal{H}_\infty$. *IEEE Transactions on Automatic Control*, **36**, 1163–1176.

Heuberger, P. S. C., P. M. J. Van den Hof and O. H. Bosgra (1995). A generalized orthonormal basis for linear dynamical systems. *IEEE Trans. Automatic Control*, **40**, 451–465.

Hinrichsen, D. and A. J. Pritchard (1990). An improved error estimate for reduced order models of discrete-time systems. *IEEE Trans. Automat. Control*, **35**, 317–320.

Hjalmarsson, H. (1993). Aspects on incomplete modeling in system identification. PhD thesis, Linköping University.

Ho, B. L. and R. E. Kalman (1966). Effective construction of linear state-variable models from input/output functions. *Regelungstechnik*, **14**, 545–548.

Huang, D. W. and L. Guo (1990). Estimation of nonstationary ARMAX models based on Hannan–Rissanen method. *The Annals of Statistics*, **18**, 1729–1756.

Jacobson, C. A., C. N. Nett and J. R. Partington (1992). Worst case system identification in ℓ_1: optimal algorithms and error bounds. *Syst. & Control Lett.*, **19**, 419–424.

Juang, J. N. and H. Suzuki (1988). An eigensystem realization algorithm in frequency domain for modal parameter identification. *J. Vibration, Acoustics, Stress, and Reliability in Design*, **110**, 24–29.

Kung, S. Y. (1978). A new identification and model reduction algorithm via singular value decomposition. In *Proc. 12th Asilomar Conf. on Circuits, Systems and Computers*, Pacific Grove, CA, pp. 705–714.

Liu, K., R. N. Jacques and D. W. Miller (1994). Frequency domain structural system identification by observability range space extraction. In *Proc. Amer. Control Conf.*, Baltimore, MD, pp. 107–111,

Ljung, L. (1993). Some results on identifying linear systems using frequency domain data. In *Proc. 32nd IEEE Conf. on Decision and Control*, San Antonio, TX, pp. 3534–3538.

Ljung, L. (1987). *System Identification: Theory for the User*. Prentice-Hall, Englewood Cliffs, NJ.

Ljung, L. and Z. D. Yuan (1985). Asymptotic properties of black-box identification of transfer functions. *IEEE Trans. Automat. Control*, **30**, 514–530.

Maciejowski, J. M. (1994). Guaranteed stability with subspace methods. *Syst. & Control Lett.*, **26**, 153–156.

Mäkilä, P. M. (1991). Robust identification and Galois sequences. *Int. J. Control*, **54**, 1189–1200.

Mäkilä, P. M. and J. R. Partington (1991). Robust approximation and identification in $\mathcal{H}_\infty$. In *Proc.* 1991 *Amer. Control Conf.*, Boston, MA, pp. 70–76.

McKelvey, T., L. Ljung and H. Akçay (1995). Identification of Infinite Dimensional Systems from Frequency Response Data. In *Proc. 3rd European Control Conf.*, Rome, Italy, pp. 2106–2111.

McKelvey, T. and H. Akçay (1994). An efficient frequency domain state-space identification algorithm. In *Proc. 33rd IEEE Conf. on Decision and Control*, Lake Buena Vista, FL, pp. 3359–3364.

Partington, J. R. (1988). *An Introduction to Hankel Operators*. Cambridge University Press, Cambridge, U.K.

Pintelon, R., J. Schoukens and H. Chen (1994a). On the basic assumptions in the identification of continuous time systems. In *Proc. 10th IFAC Symp. on System Identification*, **3**, Copenhagen, Denmark, pp. 143–152.

Pintelon, R., P. Guillaume, Y. Rolain, J. Schoukens and H. Van Hamme (1994b). Parametric identification of transfer functions in the frequency domain — a survey. *IEEE Trans. Automat. Control*, **39**, 2245–2260.

Schoukens, J. and R. Pintelon (1991). *Identification of Linear Systems: A Practical Guideline to Accurate Modeling*. Pergamon Press, London.

Söderström, T. and P. Stoica (1990). *System Identification*. Prentice Hall International, Hemel Hempstead, U.K.

Stewart, G. W. and J. G. Sun (1990). *Matrix Perturbation Theory*. Academic Press, Inc., New York.

Van Overschee, P. and B. De Moor (1994). N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*, **30**, 75–93.

Verhaegen, M. and P. Dewilde (1992). Subspace model identification. Part I: The output-error state-space model identification class of algorithms. *Int. J. Control*, **56**, 1187–1210.

Young, N. J. (1986). Balanced realizations in infinite dimensions. In *Operator Theory: Advances and Applications*. Birkhäuser, Basel.

APPENDIX A. PROOF OF LEMMA 1

This lemma was partly proven by Glover *et al.* (1988) (Appendix 2) for distinct singular values case given an arbitrary sequence of compact operators on Hilbert space converging in norm to a compact operator. The proof here is extended for the repeated singular values case.

Given any two compact operators S and T with singular values $\sigma_i(S)$ and $\sigma_i(T)$, we have from Corollary 1.5 in Partington (1988)

$$\sigma_i(S+T) \le \sigma_i(S) + \sigma_1(T)$$

and

$$\sigma_i(S) \le \sigma_i(S+T) + \sigma_1(T).$$

Being a limit of compact operators, Γ is compact. Substitute $\Gamma^{(k)}$ for S and $\Gamma - \Gamma^{(k)}$ for T. Thus

$$|\Gamma_i^{(k)} - \Gamma_i(G)| \le \|\Gamma^{(k)} - \Gamma\| \to 0.$$

For the second part, suppose that $\Gamma_1(G) = \ldots = \Gamma_m(G) > \Gamma_{m+1}(G)$. Fix $i \le m$ and write $u_i^{(k)} \triangleq \sum_{l=1}^m r_{li}^{(k)} u_l + x_i^{(k)}$, where $(x_i^{(k)}, u_l) = 0$ so that $\|x_i^{(k)}\| = \sqrt{1 - \sum |r_{li}^{(k)}|^2}$. Then

$$\begin{aligned}
\Gamma_i^{(k)} &= \|(\Gamma^{(k)})^* u_i^{(k)}\| \le \|\Gamma^* u_i^{(k)}\| + \|(\Gamma^{(k)})^* - \Gamma^*\| \\
&= \|\sum_{l=1}^m \Gamma_l(G) r_{li}^{(k)} v_l + \Gamma^* x_i^{(k)}\| + \|\Gamma^{(k)} - \Gamma\| \\
&= (\Gamma_1^2(G) \sum_{l=1}^m |r_{li}^{(k)}|^2 + \|\Gamma^* x_i^{(k)}\|^2)^{1/2} + \|\Gamma^{(k)} - \Gamma\| \\
&\le (\Gamma_1^2(G) \sum_{l=1}^m |r_{li}^{(k)}|^2 + \Gamma_{m+1}^2(G)(1 - \sum_{l=1}^m |r_{li}^{(k)}|^2))^{1/2} \\
&\quad + \|\Gamma^{(k)} - \Gamma\|,
\end{aligned}$$

since $(\Gamma^* x_i^{(k)}, v_l) = (x_i^{(k)}, \Gamma v_l) = \Gamma_l(G)(x_i^{(k)}, u_l) = 0$, $l = 1, \cdots, m$. Hence

$$\begin{aligned}
&\Gamma_1^2(G) \sum_{l=1}^m |r_{li}^{(k)}|^2 + \Gamma_{m+1}^2(G)(1 - \sum_{l=1}^m |r_{li}^{(k)}|^2) \\
&\ge (\Gamma_1^{(k)} - \|\Gamma^{(k)} - \Gamma\|)^2 \\
&\ge (\Gamma_1(G) - 2\|\Gamma^{(k)} - \Gamma\|)^2.
\end{aligned}$$

Thus

$$\sum_{l=1}^m |r_{li}^{(k)}|^2 \ge \frac{(\Gamma_1(G) - 2\|\Gamma^{(k)} - \Gamma\|)^2 - \Gamma_{m+1}^2(G)}{\Gamma_1^2(G) - \Gamma_{m+1}^2(G)} \to 1$$
$$\text{as } k \to \infty.$$

It follows that $\|x_i^{(k)}\| \to 0$ for each $i \le m$ as $k \to \infty$.

Let $U_{11} \triangleq [u_1 \cdots u_m]$, $U_{11}^{(k)} \triangleq [u_1^{(k)} \cdots u_m^{(k)}]$, and $X_1 \triangleq [x_1^{(k)} \cdots x_m^{(k)}]$. Let R^1 denote the matrix with elements $[R^1]_{li} \triangleq r_{li}^{(k)}$. Then, we have shown that $U_{11}^{(k)}$ can be decomposed as $U_{11}^{(k)} = U_{11}R^1 + X_1$ for some X_1 that is orthogonal to the columns of U_{11} and $\|X_1\|_F = o(1)$. From $(R^1)^T R^1 = I_m - X_1^T X_1$, we have $R^1(R^1)^T = I_m + o(1)$.

Consider now the following Hilbert–Schmidt expansions of Γ and $\Gamma^{(k)}$:

$$\Gamma = \Gamma_1(G) \sum_{l=1}^m (\cdot, v_l) u_l + \sum_{l=m+1}^\infty \Gamma_l(G)(\cdot, v_l) u_l,$$

$$\Gamma^{(k)} = \sum_{l=1}^m \Gamma_l^{(k)}(\cdot, v_l^{(k)}) u_l^{(k)} + \sum_{l=m+1}^\infty \Gamma_l^{(k)}(\cdot, v_l^{(k)}) u_l^{(k)}$$

and let

$$\tilde{\Gamma} \triangleq \sum_{l=m+1}^{\infty} \Gamma_l(G)(\cdot, v_l)u_l,$$

$$\tilde{\Gamma}^{(k)} \triangleq \sum_{l=m+1}^{\infty} \Gamma_l^{(k)}(\cdot, v_l^{(k)})u_l^{(k)}.$$

Let p be the multiplicity of $\Gamma_{m+1}(G)$. If $\|\tilde{\Gamma} - \tilde{\Gamma}^{(k)}\| \to 0$, the previous procedure can be applied to the leading singular values and the Hilbert–Schmidt pairs of $\tilde{\Gamma}$ and $\tilde{\Gamma}^{(k)}$ to obtain $U_{12}^{(k)} \triangleq U_{12}R^2 + X_2$ for some $R^2 \in \mathbb{R}^{p\times p}$ and X_2 that is orthogonal to the columns of U_{12} and $\|X_2\|_F = o(1)$, where $U_{12} \triangleq \left[u_{m+1} \ \cdots \ u_{m+p}\right]$ and $U_{12}^{(k)} \triangleq \left[u_{m+1}^{(k)} \ \cdots \ u_{m+p}^{(k)}\right]$. Furthermore, $R^2(R^2)^{\mathrm{T}} = I_p + o(1)$. Continue this procedure until $\Gamma_n(G)(\cdot, v_n)u_n$ is not contained in $\tilde{\Gamma}$. Let

$$R \triangleq \begin{bmatrix} R^1 & 0 & \cdots \\ 0 & R^2 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix},$$

$$X \triangleq [X_1 \ X_2 \ \cdots \ ].$$

Thus, we obtain by this procedure

$$U_1^{(k)} = U_1R + X$$

for some $R \in \mathbb{R}^{n\times n}$ and X such that $RR^{\mathrm{T}} = I_n + o(1)$ and $\|X\|_F = o(1)$. Then, orthogonal decomposition of X into U_1 and U_2 yields

$$U_1^{(k)} = U_1(R + U_1^{\mathrm{T}}X) + U_2(U_2^{\mathrm{T}}X) \triangleq U_1T^{(k)} + U_2S.$$

From our construction, $T^{(k)}$ is invertible since it is a perturbed block diagonal matrix of invertible matrices and the amount of perturbation is bounded by $\|X\|_F$. Next, S has the property $\|S\|_F = o(1)$. Since $T^{(k)}$ is invertible, U_2S can be further factorized as $U_2S = U_2(S(T^{(k)})^{-1})T^{(k)}$. Set $P^{(k)} \triangleq S(T^{(k)})^{-1}$. Furthermore, $\|P^{(k)}\|_F \to 0$ since $\sigma_n(T^{(k)}) \to 1$. This completes the proof provided that $\|\tilde{\Gamma} - \tilde{\Gamma}^{(k)}\| \to 0$ which is equivalent to

$$\sum_{l=1}^{m} \Gamma_l^{(k)}(x, v_l^{(k)})u_l^{(k)} \to \Gamma_1(G)\sum_{l=1}^{m}(x, v_l)u_l, \ \forall x \in \ell_2$$

for $\|\Gamma - \Gamma^{(k)}\| \to 0$. Because $\Gamma^{(k)} \to \Gamma$, we have for each $l \le m$

$$(x, v_l^{(k)}) = (x, (\Gamma^{(k)})^* u_l^{(k)}) = (\Gamma^{(k)}x, \sum_{i=1}^{m} r_{il}^{(k)}u_i + x_i^{(k)})$$
$$= \sum_{i=1}^{m} r_{il}^{(k)}(\Gamma x, u_i) + o(1) = \sum_{i=1}^{m} r_{il}^{(k)}(x, v_i) + o(1).$$

Thus

$$\sum_{l=1}^{m} \Gamma_l^{(k)}(x, v_l^{(k)})u_l^{(k)} = \Gamma_1(G)\sum_{l=1}^{m}\sum_{i=1}^{m}\sum_{s=1}^{m} r_{il}^{(k)}r_{sl}^{(k)}(x, v_i)u_s + o(1)$$
$$= \Gamma_1(G)\sum_{i=1}^{m}\sum_{s=1}^{m}(R^1(R^1)^{\mathrm{T}}))_{is}(x, v_i)u_s + o(1)$$
$$= \Gamma_1(G)\sum_{i=1}^{m}(x, v_i)u_i + o(1).$$

APPENDIX B. PROOF OF LEMMA 3

We will prove for $q > m$ and $r > p$ since equality cases are included. Let

$$\tilde{X}^1 \triangleq \begin{bmatrix} X^1 & 0_{m\times(r-p)} \\ 0_{(q-m)\times r} & \end{bmatrix}$$

and $\tilde{U}_1^1$ denote the matrix containing n left singular vectors of $\tilde{X}^1$ as in (66). Nonzero singular values of X_1 and $\tilde{X}_1$ are equal. Thus, $\sigma_n(\tilde{X}^1) > 0$ and $\tilde{U}_1^1$ must be in the form

$$\tilde{U}_1^1 = \begin{bmatrix} U_1^1 \\ 0_{(q-m)p\times n} \end{bmatrix} \tag{B.1}$$

for some U_1^1 in (66). Let $\gamma \triangleq \sigma_n(X^2) - \sigma_{n+1}(X^2)$ and

$$\begin{bmatrix} U_1^{2\mathrm{T}} \\ U_2^{2\mathrm{T}} \end{bmatrix}(\tilde{X}^1 - X^2)\begin{bmatrix} V_1^2 & V_2^2 \end{bmatrix} = \begin{bmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{bmatrix} = E. \tag{B.2}$$

From (B.2), we notice that $\|E\|_F = \|\tilde{X}^1 - X^2\|_F \le \epsilon$ and also $\|E_{ij}\|_2 \le \|E_{ij}\|_F$ since the spectral norm is upper bounded by the Frobenius norm. Therefore, it is clear that

$$\delta \triangleq \gamma - \|E_{11}\|_2 - \|E_{22}\|_2 \ge \gamma - \|E_{11}\|_F - \|E_{22}\|_F$$
$$> \gamma - 2\epsilon > \gamma/2$$

by assumption (67). We also have

$$\frac{\|E_{21} \ E_{12}^{\mathrm{T}}\|_F}{\delta} < \frac{2\epsilon}{\gamma} < \frac{1}{2}.$$

Then, by Theorem 8.3.5 in Golub and Van Loan (1989) there exists a matrix P satisfying

$$\|P\|_F \le \frac{2\epsilon}{\delta} \le \frac{4\epsilon}{\gamma},$$

such that range spaces of $\tilde{U}_1^1$ and $U_1^2 + U_2^2P$ are equal. Since the range spaces are equal and U_1^2 is of full rank there exists a unique nonsingular matrix T such that

$$\tilde{U}_1^1 = (U_1^2 + U_2^2P)T. \tag{B.3}$$

The equalities (B.1) and (B.3) give (68).

APPENDIX C. PROOF OF LEMMA 6

First, observe that a lower bound $M^* \le M$ yields lower bounds $q(M^*) \le q$ and $r(M^*) \le r$ since q and r are given two increasing functions. Therefore, we will consider fixed values only for M in what follows. Let $H_{qr}^s \triangleq \hat{H}_{qr} - E_{qr}$ denote the noise-free part of $\hat{H}_{qr}$ in (19) and A_s, B_s, C_s, D_s, system matrices estimated from H_{qr}^s and $\hat{g}_0$ by Algorithm 1.

As $M \to \infty$, we have $\|\Gamma^{(q+r)} - \Gamma\| \to 0$, where $G^{(q+r)}$ denotes truncated impulse response in (52). Recall that the Hankel singular values of $G^{(q+r)}$ coincide with the singular values of $H^{(q+r)}$ in (55). Then, by Lemma 1, it follows that $\sigma_i(H^{(q+r)}) \to \Gamma_i(G)$ for all i as $M \to \infty$, where $\sigma_i(H^{(q+r)})$ are the singular values of $H^{(q+r)}$ and $\Gamma_i(G)$ the Hankel singular values of G. Set

$$\tilde{H}^s \triangleq \begin{pmatrix} H_{qr}^s & 0_{qp\times qm} \\ 0_{rp\times rm} & 0_{rp\times qm} \end{pmatrix}.$$

Under the hypothesis of the lemma, $\|\tilde{H}^s - H^{(q+r)}\|_F \to 0$ as $M \to \infty$, which implies $\sigma_n(H_{qr}^s) \to \sigma_n(H^{(q+r)})$ and $\sigma_{n+1}(H_{qr}^s) \to \sigma_{n+1}(H^{(q+r)})$ as $M \to \infty$ since the singular values of H_{qr}^s and $\tilde{H}^s$ are the same. Therefore, there exists a number M_1 such that for all $M \ge M_1$

$$\sigma_n(H_{qr}^s) - \sigma_{n+1}(H_{qr}^s) > \frac{1}{2}\left(\Gamma_n(G) - \Gamma_{n+1}(G)\right).$$

By Theorem 2, $G_s \triangleq C_s(zI_n - A_s)^{-1}B_s + D_s$ converges to $G_n = \bar{C}(zI_n - \bar{A})^{-1}\bar{B} + \bar{D}$ as $M \to \infty$ where $\bar{A}, \bar{B}, \bar{C}, \bar{D}$ are the system matrices in (37)–(40). In particular, the spectral radius of A_s ($\triangleq \rho(A_s)$), radius of the smallest disk at the origin containing all eigenvalues of A_s) is bounded away from one for all sufficiently large M since balanced truncations of G are stable ,i.e. $\rho(\bar{A}) < 1$ and eigenvalues of A_s converge to those of $\bar{A}$. Hence, choose $M_2 \geq M_1$ such that for all $M \geq M_2$

$$\rho(A_s) < \frac{1 + \rho(\bar{A})}{2}.$$

Next we apply Lemma 3 with $X^1 = \hat{H}_{qr}$ and $X^2 = H^s_{qr}$. Suppose $\|H^s_{qr} - \hat{H}_{qr}\|_F \leq \epsilon$ for all $M \geq M_2$ and $\epsilon \leq (\Gamma_n(G) - \Gamma_{n+1}(G))/8$. Then there exists a $P \in \mathbb{R}^{(qp-n)\times n}$ and a $T \in \mathbb{R}^{n\times n}$ satisfying

$$\hat{U}_1 = (U_1^s + U_2^s P)T \tag{C.1}$$

and

$$\|P\|_F < \frac{8}{\Gamma_n(G) - \Gamma_{n+1}(G)}\,\epsilon \overset{\Delta}{=} c\epsilon, \tag{C.2}$$

where $U_1^s \in \mathbb{R}^{qp\times n}$ and $U_2^s \in \mathbb{R}^{qp\times(rm-n)}$ are formed from normalized left singular vectors of H^s_{qr} as in (20) for $\hat{U}_1$ and $\hat{U}_2$. Note that (C.1) and (C.2) hold for all M and ϵ provided $M \geq M_2$ and $\epsilon \leq (\Gamma_n(G) - \Gamma_{n+1}(G))/8$ although P and T may also depend on M and ϵ.

Now we proceed as in the Proof of Lemma 4. First, from (C.1) we get $T^{\mathrm{T}}T = I_n + O(\epsilon)$ or $TT^{\mathrm{T}} = I_n + O(\epsilon)$. From

$$(J_1^q\hat{U}_1)^{\mathrm{T}}J_1^q\hat{U}_1 = T^{\mathrm{T}}(J_1^qU_1^s)^{\mathrm{T}}J_1^qU_1^sT + O(\epsilon),$$
$$(J_1^q\hat{U}_1)^{\mathrm{T}}J_2^q\hat{U}_1 = T^{\mathrm{T}}(J_1^qU_1^s)^{\mathrm{T}}J_2^qU_1^sT + O(\epsilon),$$

it follows that

$$\hat{A} = T^{-1}(J_1^qU_1^s)^{\dagger}J_2U_1^sT + O(\epsilon) = T^{-1}A_sT + O(\epsilon).$$

Hence, $\rho(\hat{A}) \leq \rho(A_s) + O(\epsilon) < (1 + \rho(\bar{A}))/2 + O(\epsilon)$. Thus, choose $\bar{\epsilon} \leq (\Gamma_n(G) - \Gamma_{n+1}(G))/8$ to satisfy $\rho(\hat{A}) < \bar{\rho} < 1$ for all $\epsilon \leq \bar{\epsilon}$. Then, $\hat{A}^{2M} = O(\bar{\rho}^{2M}) = o(1)$ for all $\epsilon \leq \bar{\epsilon}$ and $M \geq M_2$. We derive the following expressions for the other matrices:

$$\hat{B} = \hat{U}_1^{\mathrm{T}}\hat{H}_{qr}J_4^r - \hat{A}^{2M}\hat{U}_1^{\mathrm{T}}\hat{H}_{qr}J_4^r = T^{\mathrm{T}}(U_1^s)^{\mathrm{T}}H^s_{qr}J_4^r + O(\epsilon) + o(1)$$
$$= T^{\mathrm{T}}B_s + O(\epsilon) + o(1),$$
$$\hat{C} = J_3^q\hat{U}_1 = J_3^qU_1^sT + O(\epsilon) = C_sT + O(\epsilon),$$
$$\hat{D} = \hat{g}_0 + o(1).$$

Finally

$$\hat{C}(zI_n - \hat{A})^{-1}\hat{B} + \hat{D}$$
$$= C_sT\left(zI_n - T^{-1}A_sT\right)^{-1}T^{\mathrm{T}}B_s$$
$$+\hat{g}_0 + O(\epsilon) + o(1)$$
$$= C_s(zI_n - A_s)^{-1}TT^{\mathrm{T}}B_s + \hat{g}_0 + O(\epsilon) + o(1)$$
$$= G_s + (\hat{g}_0 - D_s) + O(\epsilon) + o(1).$$

To sum up, there exist two numbers M_2 and $\bar{\epsilon}$ and an absolute constant c_1 such that for all $M \geq M_2$ and $\epsilon \leq \bar{\epsilon}$

$$\|E_{qr}\|_F \leq \epsilon \Longrightarrow \|\hat{G} - G_s\|_\infty \leq \|\hat{g}_0 - D_s\| + c_1\epsilon + \alpha_M$$

for some sequence α_M which tends to zero as $M \to \infty$. Since G_s converges to G_n, we have

$$\|G_s - G_n\|_\infty \leq \beta_M$$

for some sequence β_M tending to zero as $M \to \infty$. Given $\epsilon \leq \bar{\epsilon}$, choose a number $M^* \geq M_2$ such that $\alpha_M + \beta_M \leq \epsilon$. Then

$$\|E_{qr}\|_F \leq \epsilon \text{ and } \|\hat{g}_0 - D_s\| \leq \epsilon;$$
$$M \geq M^* \Longrightarrow \|\hat{G} - G_n\|_\infty \leq (2 + c_1)\epsilon;\ M \geq M^*.$$

Thus, for the events above, we get

$$\{\|\hat{G} - G_n\|_\infty > (2 + c_1)\epsilon,\ M \geq M^*\}$$
$$\subseteq \{\|E_{qr}\|_F > \epsilon \text{ or } \|\hat{g}_0 - D_s\| > \epsilon,\ M \geq M^*\},$$

which implies

$$\text{Prob}\,(\|E_{qr}\|_F > \epsilon \text{ i. o.}) + \text{Prob}\,(\|\hat{g}_0 - D_s\| > \epsilon \text{ i. o.})$$
$$\geq \text{Prob}\,(\|\hat{G} - G_n\|_\infty > (2 + c_1)\epsilon \text{ i. o.}),$$

where i. o. stands for "infinitely often". It is known that a sequence of random variables X_M converges to zero w.p. 1 as $M \to \infty$ if and only if for each $\epsilon > 0$, Prob $(|X_M| > \epsilon$ i. o.$) = 0$ (Chung, 1974). Hence, the desired result will follow from this inequality and the hypothesis if it holds that $\hat{g}_0 - D_s \to 0$ w.p. 1. Note that $\hat{g}_0 - D_s$ is the sample mean of extended frequency-response noise which is an independent sequence of covariance-bounded random variables. Thus, $\hat{g} - D_s \to 0$ w.p. 1 as $M \to \infty$.