Automatica 133 (2021) 109882

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

MIMO frequency domain system identification using matrix-valued orthonormal functions☆

Tao Qian [a], Xiaoyin Wang [b,*], Liming Zhang [c]

[a] Faculty of Information Technology, Macau University of Science and Technology, Macau, China
[b] Department of Mathematics, Faculty of Science and Technology, University of Macau, Macau, China
[c] Department of Computer and Information Science, Faculty of Science and Technology, University of Macau, Macau, China

ARTICLE INFO

Article history:
Received 3 March 2020
Received in revised form 17 March 2021
Accepted 8 July 2021
Available online 31 August 2021

Keywords:
MIMO systems
Blaschke–Potapov factor
Adaptive orthonormal functions
Maximal selection principle
$\mathbb{C}^{q\times q}$ right Hilbert module

ABSTRACT

In this paper, we propose a two-stage algorithm utilizing the Cauchy integral and the matrix-valued adaptive Fourier decomposition (abbreviated as matrix AFD) to identify transfer functions of linear time-invariant (LTI) multi-input multi-output (MIMO) systems in the continuous time case. In recent work of Alpay et al. (2017), a theory of adaptive rational approximation to matrix-valued Hardy space functions on the unit disk was established. The matrix-valued function theory has great potential in applications, in views of the practice of its scalar-valued counterparts. The algorithm and application aspects of the mentioned theory of Alpay et al. (2017) have not been developed. The theory was only written for the unit disk case corresponding to the discrete time systems. The contributions of the present paper are 3-fold. First, we construct an analogous adaptive approximation theory for complex matrix-valued Hardy space functions defined on a half of the complex plane, corresponding to the Laplace transforms of signals of finite energy whose Fourier transforms are supported on a half of the frequency domain. The half plane model corresponds to signals defined in the whole axis range which is an alternative case to signals defined in a compact interval. The second fold contribution lays on maximal selection of the pair (a, P) where a is a point of the right-half plane and P is an orthogonal projection. We show that the optimal selection of P is dependent on a when a is first fixed, that is $P = P(a)$, where $P : a \to P(a)$ has an explicit corresponding relation. Due to this relation we reduce the maximal selection of the pair (a, P) to only that of the parameter a. This result can be extended to the compact intervals case as well. The third fold is, with the precise rule from a to $P(a)$, we develop a practical algorithm for the adaptive approximation to the transfer function. Through an example we show that the proposed algorithm is effective in both the noise-free and noisy cases.

1. Introduction

System identification has been having a long term development motivated by the evolution of the neighbouring fields of science (Gevers, 2006) and has had a wide range of engineering applications (see, for instance, Natke and Cottin (1988)). It constructs mathematical models for dynamic systems from observed input–output data. Despite of the progresses, there are still challenges in this subject (see, for instance (Ninness, 2009)).

The two mainstreams of identification techniques are prediction error identification and subspace identification (Katayama, 2005; Ljung, 1999; Overschee & Moor, 1996; Pintelon & Schoukens, 2012). Prediction error identification methods (PEM) are based on the minimization of a prediction error criterion to estimate the system parameters. It usually involves non-linear optimization using iterative gradient search methods and an initial model is required, which may lead to a local minima. To reduce the computational burden and increase the estimation accuracy when we apply PEM for MIMO systems, some novel estimation algorithms, for instance Ding and Chen (2005), Liu, Xiao, and Zhao (2009), Ninness (2009), are studied. Subspace identification methods are based on the projection theory in a Hilbert space. The subspace methods are non-iterative, which require to determine only one structural parameter, and do not cause extra difficulty to process MIMO systems. Moreover, they are numerically efficient due to use of robust tools such as QR factorization and singular value decomposition (SVD) (Stoica & Jansson, 2000). The subspace identification methods can be used to provide a good initial state space parameterized estimation for PEM to overcome the challenge in initialization (Mckelvey, 1995).

☆ This work was supported by Macao Science and Technology Development Fund FDCT079/2016/A2 and FDCT0123/2018/A3. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Wei Xing Zheng under the direction of Editor Torsten Söderström.

* Corresponding author.
E-mail addresses: tqian1958@gmail.com (T. Qian), yb57448@connect.um.edu.mo (X. Wang), lmzhang@um.edu.mo (L. Zhang).

https://doi.org/10.1016/j.automatica.2021.109882
0005-1098/

In addition to the above mentioned two mainstreams the recent rapid developments of machine learning techniques and kernel methods shed their lights on applications to system identification.

Among the various methods in the study of system identification, identification with rational orthogonal basis functions is effective from the perspective of approximation theory, which extends the system theory of Laguerre functions and Kautz functions. It searches the "best approximating model" to the system from a priori chosen model set. The book (Heuberger, Hof, & Wahlberg, 2005) gathers the important contributions in this field during about 15 years around 2000. It is noted that these studies focus on the case of single-input single-output (SISO) systems. We aim to give an identification method for MIMO systems based on the newly established theory on matrix-valued rational orthogonal functions.

Potapov originally defined matrix-valued Blaschke–Potapov factors up to a right multiplicative constant in the work (Potapov, 1955). Inspired by this work, some researchers have developed thorough theories for matrix-valued functions (abbreviated as mvf's) as extension of the classical Hardy space theory, including rational matrix-valued Blaschke products, inner-outer factorization for mvf's, Beurling–Lax theorem for mvf's, and other related topics (see, for instance, Alpay and Gohberg (1988a, 1988b), Arov and Dym (2008)). Based on those studies, authors of Alpay, Colombo, Qian, and Sabadini (2017) propose an adaptive orthonormal decomposition method for mvf's defined on the unit disk $\mathbb{D}$, whose scalar-valued case is the adaptive Fourier decomposition (AFD) method given in Qian and Wang (2011). AFD has been receiving attentions in the last 10 years and has been successfully applied to SISO system identification (Chen, Mai, Zhang, & Mi, 2015; Fei, Mi, & Pan, 2016; Mi & Qian, 2011, 2014; Mi, Qian, & Li, 2016; Mi, Qian, & Wan, 2012). Being different from the classical identification methods (Akcay & Ninness, 1998; Makila, 1991; Ninness & Gustafsson, 1997; Wahlberg, 1994) using limited types of pre-determined orthogonal basis functions, AFD iteratively and adaptively constructs rational orthogonal functions according to the given function under the so called "maximal selection principle". Practically, AFD gives optimal approximations to the given transfer functions by achieving fast convergence.

In this paper, we develop the adaptive orthonormal decomposition method in Alpay et al. (2017) and apply it to MIMO system identification for continuous systems. For discrete systems, a parallel identification procedure can be derived by utilizing the original decomposition method in Alpay et al. (2017) with the same methodology as given in this paper. The contributions of this paper are three folds:

(1) Referring to the existing theory of adaptive orthonormal decomposition for mvf's defined on the unit disk $\mathbb{D}$, we develop the counterpart theory for mvf's defined on the right-half complex plane Π. In case there is no ambiguity, we call both the adaptive orthonormal decomposition methods for functions belonging to the Hardy spaces $H_2^{p\times q}(\mathbb{D})$ and $H_2^{p\times q}(\Pi)$ as matrix AFD, and will not specify which context. With this development a set of corresponding objects in the half-plane context are specified and further used in the mvf's system identification.
(2) The maximal selection principle used in the matrix AFD involves the selection of a parameter a in the right half-plane Π and an orthogonal projection P. Utilizing a corollary of the Poincaré Separation Theorem in the matrix inequalities theory, we prove that the selection of P depends entirely on the pre-selected a. Therefore, we find a convenient way to solve the nonlinear optimization problem under the maximal selection principle on the two independent objects.
(3) Adapted to the system identification problem, we propose a two-stage algorithm utilizing the Cauchy integral and the matrix AFD defined on Π. The proposed algorithm, in particular, gives rise to real coefficients in the rational approximations to the transfer functions of the continuous time systems. Last but not least, through some numerical experiments, we show that the proposed matrix AFD and the related algorithm are effective and converge fast.

The rest of the paper is organized as follows. Section 2 recalls preliminary knowledge and gives the problem setting. Section 3 introduces the matrix AFD for the mvf's space $H_2^{p\times q}(\Pi)$ and proves a theorem, related to the maximal selection principle in the matrix AFD, to show the precise formula of the optimal orthogonal projection P with respect to a previously determined $a \in \Pi$. $H_2^{p\times q}(\Pi)$ is the playground of the transfer functions of the continuous systems. Section 4 presents a two-stage algorithm utilizing the matrix AFD in Section 3 to identify a matrix-valued transfer function. In Section 5, we give two theorems, one states the convergence rate of the matrix AFD and the other gives an upper bound for the identification process in the presence of noise. Section 6 exhibits the numerical experiments. The proposed algorithm gives promising approximation results. Some conclusions are drawn in Section 7.

2. Preliminary and problem setting

We restrict our discussion to the following space.

Definition 1 ($H_2^{p\times q}(\Pi)$)**.** Denote the open right-half complex plane by $\Pi = \{s \in \mathbb{C} : \mathrm{Re}(s) > 0\}$, where Re means the real part. For positive integers p and q, define $H_2^{p\times q}(\Pi)$ as the set containing all $\mathbb{C}^{p\times q}$-valued functions whose entries belong to the Hardy 2-space defined on Π, denoted by $H_2(\Pi)$. It is known that all $f(s) \in H_2(\Pi)$ have non-tangential boundary limits $f(\mathrm{j}y)$ where $y \in \mathbb{R}$. The norm of $f(s)$ can be computed through its boundary limit, as $\|f\|_{H_2(\Pi)}^2 = \frac{1}{2\pi}\int_{\mathbb{R}} |f(\mathrm{j}y)|^2 \mathrm{d}y$. The induced inner product is defined by $\langle f, g\rangle = \frac{1}{2\pi}\int_{\mathbb{R}} f(\mathrm{j}y)\overline{g(\mathrm{j}y)}\mathrm{d}y$ for $f(s), g(s) \in H_2(\Pi)$. Denote $*$ as the complex conjugate transpose. For $F(s), G(s) \in H_2^{p\times q}(\Pi)$, the $\mathbb{C}^{q\times q}$-valued inner product is defined by

$$[F, G] = \frac{1}{2\pi}\int_{\mathbb{R}} G(\mathrm{j}y)^* F(\mathrm{j}y)dy, \tag{1}$$

where each entry of $F(\mathrm{j}y), G(\mathrm{j}y)$ is the non-tangential boundary limit of the corresponding entry of $F(s), G(s)$ from the inside of Π to the imaginary axis. The norm is defined by

$$\|F\| = \sqrt{\mathrm{trace}[F, F]}.$$

It is noted that if $F(s) \in H_2^{p\times q}(\Pi)$, then $F(\mathrm{j}y)$ belongs to the mvf's space $L_2^{p\times q}(\mathrm{j}\mathbb{R})$ whose entries belong to the square integrable function space $L_2(\mathrm{j}\mathbb{R})$. Obviously, the $L_2^{p\times q}(\mathrm{j}\mathbb{R})$ norm of $F(\mathrm{j}\mathbb{R})$ satisfies $\|F\|_{L_2^{p\times q}} = \|F\|$. $H_2^{p\times q}(\Pi)$ can be regarded as $\mathbb{C}^{q\times q}$ right Hilbert module (see in Lance (1994)), which is an extension of Hilbert space for mvf's space.

In this paper, we consider continuous LTI, MIMO and causal systems. The transfer function $G(s)$ satisfies $Y(s) = G(s)U(s)$ where $Y(s)$ and $U(s)$ are the Laplace transformations of the output y and input u in the time domain. Moreover, we assume that, entries of $G(s)$ have no poles in the closed region $\overline{\Pi} = \{s \in \mathbb{C} : \mathrm{Re}(s) \geq 0\}$, which means that the systems are asymptotically stable. Thus, $G(s) \in H_2^{p\times q}(\Pi)$.

In order to formulate the frequency domain system identification problem, assume that some frequency response measurements are given and corrupted by additive noise,

$$G_{\mathrm{measure}}(\mathrm{j}y_k) = G(\mathrm{j}y_k) + v_k, \tag{2}$$

where the frequency y_k belongs to a frequency nodes set Ω_M and ν_k is the matrix-valued noise. In general, we assume that entries of ν_k are deterministic and bounded or obey certain probability distribution. About this, we will give a detailed discussion later. Here, we consider a set of nodes $\Omega_M = \{\pm \mathrm{j}y_k, k = 1, 2, \ldots, \frac{M}{2}$ with $0 < y_1 < y_2 < \cdots < y_{\frac{M}{2}} = T\}$ where M is a positive even integer and $T > 0$.

With the above assumption, we aim to find approximations $G_n(s)$ (n is a positive integer) of $G(s)$ such that

$$\lim_{\substack{n,M\to\infty \\ \nu\to 0}} \mathbb{E}\|G - G_n\|_{L_2^{p\times q}} = 0. \tag{3}$$

Here, $\mathbb{E}$ stands for the mathematical expectation and $\|\cdot\|_{L_2^{p\times q}}$ is mentioned in Definition 1.

3. Matrix AFD for functions in $H_2^{p\times q}(\Pi)$

In this section, we introduce the matrix AFD method for functions in $H_2^{p\times q}(\Pi)$. It is an adaptation of the decomposition theory on the unit disk developed in Alpay et al. (2017). The method involves, at each step, selections of a parameter $a \in \Pi$ and an orthogonal projection P of a prescribed rank to meet the maximal selection principle. We will give an explicit formula for the theoretically optimal P in terms of a prescribed $a \in \Pi$. The section includes three parts: the mathematical foundations, the matrix AFD and its convergence theorem, and the formulation of the practical and optimal P.

3.1. Mathematical foundations

This part is the base of the iterative algorithm of the matrix AFD introduced in Section 3.2. After supplying the basic and necessary ingredients, we prove a key result, Theorem 5, which plays an important role in the optimal parameter selection at each iteration step and the convergence proof of the matrix AFD.

Definition 2 (*Szegö Kernel on Π*)**.** Denote the Szegö kernel on Π by $e_a(s) = \frac{\sqrt{2\,\mathrm{Re}(a)}}{s+\bar{a}}$, $a \in \Pi$. The following properties hold:
(a) $e_a(s) \in H_2(\Pi)$,
(b) $\langle e_a, e_a\rangle = 1$,
(c) $\forall f(s) \in H_2(\Pi)$, $\langle f, e_a\rangle = \sqrt{2\,\mathrm{Re}(a)} f(a)$.

Furthermore, $H_2(\Pi)$ is a reproducing kernel Hilbert space, admitting the reproducing kernel $k_a(s) = \frac{e_a(s)}{\sqrt{2\,\mathrm{Re}(a)}}$, $a \in \Pi$, and the set $\mathcal{D} = \{e_a(s), a \in \Pi\}$ is a dictionary of $H_2(\Pi)$.

Definition 3 (*Blaschke–Potapov Factor (Arov & Dym, 2008)*)**.** An elementary Blaschke–Potapov factor on Π is defined as

$$B_{a,P}(s) = I_p - P + P\frac{s-a}{s+\bar{a}},$$

where $a \in \Pi$, p is a positive integer, P is a $\mathbb{C}^{p\times p}$-valued orthogonal projection satisfying $P = P^* = P^2$ and I_p is the $p \times p$ identity matrix. If $\mathrm{rank}(P) = k$, there exists a $k \times p$ matrix $V = [v_1, v_2, \ldots, v_k]^*$ whose row vectors are orthonormal, i.e. $VV^* = I_k$, such that $P = V^*V$. It is noted that $B_{a,P}^{-1}(s) = I_p - P + P\frac{s+\bar{a}}{s-a}$.

With the above definitions, we give the following lemma and theorem, being crucial for developing the matrix AFD, whose proofs are given in Appendix A.

Lemma 4. *For $F(s) \in H_2^{p\times q}(\Pi)$, let $a_0 \in \Pi$, P_0 be a $p\times p$ orthogonal projection and*

$$H_0(s) = P_0F(a_0)e_{a_0}(s)\sqrt{2\,\mathrm{Re}(a_0)},$$
$$H(s) = F(s) - H_0(s),$$

we have $P_0H(a_0) = 0$, $[F, F] = [H_0, H_0] + [H, H]$ and $[H_0, H_0] = 2\,\mathrm{Re}(a_0)F(a_0)^*P_0F(a_0)$. Moreover, $G(s) = B_{a_0,P_0}^{-1}(s)H(s)$ belongs to $H_2^{p\times q}(\Pi)$ and $[G, G] = [H, H]$.

Theorem 5 (*Maximal Selection Principle for mvf's in $H_2^{p\times q}(\Pi)$*)**.** *Let $F(s) \in H_2^{p\times q}(\Pi)$ and $k_0 \in \{1, 2, \ldots, p\}$. Then for all $s \in \Pi$ and orthogonal projections P of rank k_0,*
(a) trace$(2\,\mathrm{Re}(s)F(s)^*PF(s))$ has a finite supremum.
(b) there exist $a_0 \in \Pi$ and orthogonal projection P_0 with rank k_0 such that trace$(2\,\mathrm{Re}(a_0)F(a_0)^*P_0F(a_0))$ is the maximum of trace$(2\,\mathrm{Re}(s)F(s)^*PF(s))$.

3.2. Matrix AFD and its convergence theorem

In the previous part, we have made a full theoretical preparation. Based on those, we are ready to give the matrix AFD for functions in $H_2^{p\times q}(\Pi)$.

Let $F(s) \in H_2^{p\times q}(\Pi)$, $k_0 \in \{1, 2, \ldots, p\}$, and $F_1(s) = F(s)$. According to the maximal selection principle proved in Theorem 5, we can select $a_1 \in \Pi$ and an orthogonal projection $P_1 \in \mathbb{C}^{p\times p}$ of rank k_0 such that $\mathrm{trace}(2\,\mathrm{Re}(a_1)F_1(a_1)^*P_1F_1(a_1))$ attains its global maximum among all possible selections. Denote $H_1(s) = F_1(s) - P_1F_1(a_1)e_{a_1}(s)\sqrt{2\,\mathrm{Re}(a_1)}$, then from Lemma 4, $F(s)$ is decomposed as

$$F(s) = P_1F_1(a_1)e_{a_1}(s)\sqrt{2\,\mathrm{Re}(a_1)} + H_1(s), \tag{4}$$

and there exists $F_2(s) \in H_2^{p\times q}(\Pi)$ such that

$$H_1(s) = B_{a_1,P_1}(s)F_2(s), \quad [F_2, F_2] = [H_1, H_1].$$

Thus the decomposition (4) of $F(s)$ is rewritten as

$$F(s) = P_1F_1(a_1)e_{a_1}(s)\sqrt{2\,\mathrm{Re}(a_1)} + B_{a_1,P_1}(s)F_2(s).$$

Furthermore, we can decompose $F_2(s)$ using the same strategy, namely, select $a_2 \in \Pi$ and an orthogonal projection $P_2 \in \mathbb{C}^{p\times p}$ of rank k_0 according to the maximal selection principle. This gives,

$$\begin{aligned} F_2(s) &= P_2F_2(a_2)e_{a_2}(s)\sqrt{2\,\mathrm{Re}(a_2)} + H_2(s) \\ &= P_2F_2(a_2)e_{a_2}(s)\sqrt{2\,\mathrm{Re}(a_2)} + B_{a_2,P_2}(s)F_3(s), \end{aligned}$$

where $H_2(s) = F_2(s) - P_2F_2(a_2)e_{a_2}(s)\sqrt{2\,\mathrm{Re}(a_2)}$, $F_3(s) \in H_2^{p\times q}(\Pi)$ and $[F_3, F_3] = [H_2, H_2]$. This leads to the following decomposition of $F(s)$

$$\begin{aligned} F(s) =& P_1F_1(a_1)e_{a_1}(s)\sqrt{2\,\mathrm{Re}(a_1)} \\ &+ B_{a_1,P_1}(s)P_2F_2(a_2)e_{a_2}(s)\sqrt{2\,\mathrm{Re}(a_2)} \\ &+ B_{a_1,P_1}(s)B_{a_2,P_2}(s)F_3(s). \end{aligned}$$

Iterating step by step, we get two function sequences $\{F_1(s), F_2(s), \ldots\} \subset H_2^{p\times q}(\Pi)$ and $\{H_1(s), H_2(s), \ldots\} \subset H_2^{p\times q}(\Pi)$ satisfying the following recurrence relations

$$H_k(s) = F_k(s) - P_kF_k(a_k)e_{a_k}(s)\sqrt{2\,\mathrm{Re}(a_k)},$$
$$F_{k+1}(s) = B_{a_k,P_k}^{-1}(s)H_k(s).$$

Here, $H_k(s)$ is called the kth standard remainder. $F_k(s)$ is called the kth reduced remainder. Denote

$$\begin{aligned} M_k =& P_kF_k(a_k)\sqrt{2\,\mathrm{Re}(a_k)}, \\ B_k(s) =& \Big(\prod_{u=1}^{k-1} B_{a_u,P_u}(s)\Big)P_ke_{a_k}(s), \quad k = 1, 2, \ldots, \end{aligned}$$

we obtain a decomposition of $F(s)$ after N iterations as

$$F(s) = \sum_{k=1}^{N} B_k(s)M_k + \Big(\prod_{u=1}^{N} B_{a_u,P_u}(s)\Big)F_{N+1}(s),$$

where $F_{N+1}(s) \in H_2^{p\times q}(\Pi)$.

We denote the N-partial sum as $S_N(F)(s) = \sum_{k=1}^{N} B_k(s)M_k$.

Theorem 6 (*Convergence of Matrix AFD*). *Let $F(s) \in H_2^{p\times q}(\Pi)$. Then the above matrix AFD converges in the $H_2^{p\times q}(\Pi)$ norm sense, namely $\lim_{N\to\infty} \|F - S_N(F)\| = 0$. Moreover,*

$$\begin{aligned}&[S_N(F), S_N(F)]\\ &= \sum_{k=1}^{N} M_k^* M_k = \sum_{k=1}^{N} 2\,\mathrm{Re}(a_k) F_k(a_k)^* P_k F_k(a_k).\end{aligned} \tag{5}$$

The proof of this theorem is similar to the one for matrix AFD for functions in $H_2^{p\times q}(\mathbb{D})$ developed in Alpay et al. (2017). We omit it here.

3.3. Formulation of the practical and optimal P

In Theorem 5 we proved the existence of an optimal pair (a, P). It is critical to propose a practical procedure to determine an optimal pair (a, P) considering that the construction involves all orthogonal projections with the prescribed rank. In the following theorem, we give the precise corresponding rule between an optimal P_0 in relation to a pre-selected a_0.

Theorem 7. *Let $F(s) \in H_2^{p\times q}(\Pi)$ and $k_0 \in \{1, 2, \ldots, p\}$. Then an optimal pair (a_0, P_0) mentioned in Theorem 5 satisfies*

$$\begin{aligned}&\mathit{trace}(2\,\mathrm{Re}(a_0) F(a_0)^* P_0 F(a_0))\\ &= 2\,\mathrm{Re}(a_0) \sum_{l=1}^{k_0} \lambda_l(F(a_0)F(a_0)^*),\end{aligned}$$

where $\lambda_1(\cdot), \lambda_2(\cdot), \ldots, \lambda_p(\cdot)$ are the eigenvalues sorted in the descending order. In addition, $P_0 = U_0^* U_0$ where $U_0 = [\varphi_1, \varphi_2, \ldots, \varphi_{k_0}]^*$ and $\varphi_1, \varphi_2, \ldots, \varphi_{k_0}$ are the eigenvectors corresponding to the first k_0 largest eigenvalues of $F(a_0)F(a_0)^*$ satisfying $U_0 U_0^* = I_{k_0}$.

Proof. We just need to show that the maximum of $\mathrm{trace}(2\,\mathrm{Re}(s)F(s)^*PF(s))$ can be obtained by only associating it with the variable s.

From Definition 3, there exists a $k_0 \times p$ matrix V satisfying $VV^* = I_{k_0}$ such that $P = V^*V$. Denote $Q_1 = PF(s)F(s)^*$ and $Q_2 = VF(s)F(s)^*V^*$, then

$$\mathrm{trace}(Q_1) = \mathrm{trace}(Q_2) = \sum_{l=1}^{k_0} \lambda_l(Q_2) \overset{(a,b,c)}{=} \sum_{l=1}^{k_0} \lambda_l(Q_1),$$

because of the following basic results in linear algebra (see for instance (Horn & Johnson, 1990)),
(a) let $A \in \mathbb{C}^{p\times q}, B \in \mathbb{C}^{q\times p}$, AB and BA have the same nonzero eigenvalues. Moreover, $\mathrm{rank}(AB) \le \min\{\mathrm{rank}(A), \mathrm{rank}(B)\}$.
(b) the number of the nonzero eigenvalues of a square matrix A (multiplicity included) $\le$ rank (A).
(c) Q_2 is a positive semidefinite matrix and all of its eigenvalues ≥ 0.

Together with Lemma 16 in Appendix B and the fact that $F(s)F(s)^*$ is a positive semidefinite matrix, we have

$$\begin{aligned}&\mathrm{trace}(2\,\mathrm{Re}(s)F(s)^*PF(s))\\ &= 2\,\mathrm{Re}(s)\mathrm{trace}(Q_1) = 2\,\mathrm{Re}(s) \sum_{l=1}^{k_0} \lambda_l(Q_1)\\ &\overset{(*)}{\le} 2\,\mathrm{Re}(s) \sum_{l=1}^{k_0} \lambda_l(F(s)F(s)^*).\end{aligned}$$

Here, "=" in the above inequality $(*)$ holds if and only if $P = U^*U$ following the statement of Lemma 16, namely columns of U^* are set to be the orthonormal eigenvectors corresponding to the first k_0 largest eigenvalues of $F(s)F(s)^*$.

Therefore, we note that by searching the maximum of $2\,\mathrm{Re}(s)\sum_{l=1}^{k_0} \lambda_l(F(s)F(s)^*)$, which is only associated with the variable s, we can obtain the maximum of $\mathrm{trace}(2\,\mathrm{Re}(s)F(s)^*PF(s))$. In addition, the optimal P_0 can be represented by the optimal a_0 using the eigenvectors of $F(a_0)F(a_0)^*$. □

Remark 8. Due to (5), we conclude that the maximal selection principle guarantees that at each iteration, the norm increasing of $S_N(F)$ is maximal. Here, the optimal pair (a_k, P_k) at each iteration is not unique since the optimal a_k is not unique. In addition, according to Theorem 7, we notice that at each iteration, by prescribing a larger k_0, which is the rank of the orthogonal projection, we obtain a larger norm increasing of $S_N(F)$.

4. Two-stage procedure for frequency-domain MIMO system identification

To apply the decomposition method introduced in Section 3 to MIMO system identification, we design a two-stage process referring to the work in Gu and Khargonekar (1992), Mi and Qian (2011) and Mo, Qian, and Mi (2015). In the first stage, we use the Cauchy integral to formulate an analytic mvf whose boundary value is the frequency response of the transfer function. In the second stage, we decompose the analytic mvf obtained in the previous stage utilizing the matrix AFD and use the N-partial sum to approximate the transfer function.

4.1. The first stage

In the classical Hardy space theory, the following results involving Cauchy integral representation of the Hardy space on the half plane are well known (see, for instance, Garnett (1987), Li, Deng, and Qian (2016, 2018)). Let $f(\mathrm{j}y) \in L_2(\mathrm{j}\mathbb{R})$. The Cauchy integral of $f(\mathrm{j}y)$ is defined by

$$\mathcal{C}(f)(s) = \frac{1}{2\pi\mathrm{j}} \int_{\mathbb{R}} \frac{f(\mathrm{j}y)}{y - \mathrm{j}s}\mathrm{d}y, \quad s \in \Pi, \tag{6}$$

where Π is the right-half plane. Then $\mathcal{C}(f)(s) \in H_2(\Pi)$ and there exists a constant $C > 0$ such that

$$\|\mathcal{C}(f)\|_{H_2(\Pi)} \le C\|f\|_{L_2(\mathrm{j}\mathbb{R})},$$

where the $L_2(\mathrm{j}\mathbb{R})$ norm is $\|f\|_{L_2(\mathrm{j}\mathbb{R})}^2 = \frac{1}{2\pi}\int_{\mathbb{R}} |f(\mathrm{j}y)|^2\mathrm{d}y$. Conversely, for all $f(s) \in H_2(\Pi)$, $f(s)$ is the Cauchy integral of its boundary limit on the imaginary axis $f(\mathrm{j}y)$, namely $f(s) = \mathcal{C}(f)(s)$.

Based on the above results, given the $\mathbb{C}^{p\times q}$-valued frequency response $G(\mathrm{j}y) = (g_{lm}(\mathrm{j}y)), y \in \mathbb{R}$, by taking the Cauchy integral of each entry $g_{lm}(\mathrm{j}y)$, we construct the analytic mvf $F(s)$ as

$$F(s) = \mathcal{C}(G)(s),$$

and there exists a constant $C > 0$ such that

$$\begin{aligned}\|F\|^2 &= \sum_{l=1}^{p}\sum_{m=1}^{q} \|\mathcal{C}(g_{lm})\|_{H_2(\Pi)}^2\\ &\le C \sum_{l=1}^{p}\sum_{m=1}^{q} \|g_{lm}\|_{L_2(\mathrm{j}\mathbb{R})}^2 = C\|G\|^2.\end{aligned} \tag{7}$$

Given some measurements of the frequency response, $\{G_{measure}(\pm\mathrm{j}y_k)\}_{k=1}^{\frac{M}{2}}$ as mentioned in Section 2, we can apply the classical interpolatory quadrature methods (see, for instance (Dahlquist & Björck, 2008; Sun & Dang, 2019)) to calculate the Cauchy integral in (6). In this paper, we use the basic rectangular interpolation to construct a step function $\hat{G}(\mathrm{j}y)$, whose value

defined outside a finite interval is equal to 0, to approximate $G(\mathrm{j}y)$ as below.

$$\hat{G}(\mathrm{j}y)=\begin{cases} G(\mathrm{j}y_k), & y_k\le y<y_{k+1} \\ & \text{and } k=-\frac{M}{2},\dots,-1, \\ G(\mathrm{j}y_{k+1}), & y_k<y\le y_{k+1} \\ & \text{and } k=0,\dots,\frac{M}{2}-1, \\ 0, & y<y_{-\frac{M}{2}} \text{ or } y>y_{\frac{M}{2}}, \end{cases} \tag{8}$$

where we set $y_0=0$. It is easy to verify that $\overline{\mathcal{C}(\hat{G})(\bar{s})}=\mathcal{C}(\hat{G})(s)$. Denote $\hat{G}_{measure}(\mathrm{j}y)$ as the function obtained by (8) when we use the noisy frequency response $G_{measure}(\mathrm{j}y_k)$ in place of $G(\mathrm{j}y_k)$, then in the noise-free case, the analytic mvf $F(s)=\mathcal{C}(\hat{G}_{measure})(s)$ satisfies the symmetric property as the transfer function does, i.e., $\overline{F(\bar{s})}=F(s)$.

The Cauchy integral method to compute the first stage analytic approximation is referred to Qian, Zhang, and Li (2011).

Remark 9. In practice, frequency response functions are often measured in frequency bands that are away from the zero frequency. The Carleman Theorem asserts that if boundary limits of a Hardy H_p ($p\in[1,\infty)$) function are known on any set of positive Lebesgue measure, then the Hardy space function can be fully recovered from the related Carleman formula (Aizenberg, 1993). According to this result, on any frequency band of positive Lebesgue measure, practically being away from the zero frequency, if one takes a set of dense enough frequency response measurements, then one can recover the corresponding transfer function.

4.2. The second stage

Applying the matrix AFD to $F(s)=\mathcal{C}(\hat{G}_{measure})(s)$ constructed in the first stage, we obtain the N-partial sum $S_N(F)(s)$ of $F(s)$. We use $S_N(F)(s)$ to construct an approximation sequence to $G(s)$ with real coefficients. Denote by $\widetilde{S_N}(F)(s)$ the function obtained by taking the conjugates of the coefficients of $S_N(F)(s)$, and set

$$RS_N(F)(s)=\frac{S_N(F)(s)+\widetilde{S_N}(F)(s)}{2}.$$

Then the following theorem holds.

Theorem 10. *Let $F(s)\in H_2^{p\times q}(\Pi)$. Applying the matrix AFD to $F(s)$, then $RS_N(F)(s)\in H_2^{p\times q}(\Pi)$ and $RS_N(F)(s)$ is a mvf whose elements are rational functions having real coefficients and with the same force of approximation to $F(s)$.*

Proof. Given $F(s)\in H_2^{p\times q}(\Pi)$, we note that $S_N(F)(s)\in H_2^{p\times q}(\Pi)$ and

$$\begin{aligned}\|\widetilde{S_N}(F)\|^2&=\mathrm{trace}\Big(\frac{1}{2\pi}\int_{\mathbb{R}}\overline{S_N(F)(-\mathrm{j}y)}^{*}\overline{S_N(F)(-\mathrm{j}y)}\mathrm{d}y\Big)\\&=\|S_N(F)\|^2.\end{aligned}$$

Therefore, $RS_N(F)(s)\in H_2^{p\times q}(\Pi)$.

$RS_N(F)(s)$ is a mvf whose elements are rational functions of real coefficients due to the relations

$$\begin{aligned}\overline{RS_N(F)(\bar{s})}&=\frac{\overline{S_N(F)(\bar{s})+\widetilde{S_N}(F)(\bar{s})}}{2}\\&=\frac{\widetilde{S_N}(F)(s)+S_N(F)(s)}{2}=RS_N(F)(s).\end{aligned}$$

Furthermore, since $\|\widetilde{S_N}(F)-F\|=\|S_N(F)-F\|$,

$$\begin{aligned}\|RS_N(F)-F\|&\le\frac{1}{2}\|\widetilde{S_N}(F)-F\|+\frac{1}{2}\|S_N(F)-F\|\\&=\|S_N(F)-F\|.\end{aligned}$$

Thus, $RS_N(F)(s)$ is an approximation of $F(s)$ of the same force. □

5. Convergence rate and error analysis

In this section, we give an estimation for the convergence rate of the matrix AFD and use it to analyse the approximation error of the two-stage algorithm proposed in Section 4.

5.1. Convergence rate of matrix AFD

Define a subset of $H_2^{p\times q}(\Pi)$ as

$$\begin{aligned}H(\mathcal{D},R)=\{&F(s)\in H_2^{p\times q}(\Pi):F(s)=\sum_{k=1}^{\infty}C_k e_{w_k}(s),\\&\text{where } C_k\in\mathbb{C}^{p\times q},\ w_k\in\Pi,\ \text{and}\\&\sum_{k=1}^{\infty}\sqrt{\mathrm{trace}(C_k^*C_k)}<R\},\quad R>0.\end{aligned}$$

We analyse the convergence rate of matrix AFD within $H(\mathcal{D},R)$.

Proposition 11. *Let $F(s)\in H(\mathcal{D},R)$. Then $\|F\|<R$.*

Proof. Let $F(s)=(f_{lm}(s))\in H(\mathcal{D},R)$. By directly computing according to the triangle inequality and the Cauchy–Schwartz inequality, we obtain

$$\begin{aligned}&\|F\|^2\\&=\Big|\mathrm{trace}[F,\sum_{k=1}^{\infty}C_k e_{w_k}]\Big|\\&=\Big|\mathrm{trace}(\sum_{k=1}^{\infty}C_k^*\begin{bmatrix}\langle f_{11},e_{w_k}\rangle & \cdots & \langle f_{1q},e_{w_k}\rangle\\ \vdots & & \vdots\\ \langle f_{p1},e_{w_k}\rangle & \cdots & \langle f_{pq},e_{w_k}\rangle\end{bmatrix})\Big|\\&\le\sum_{k=1}^{\infty}\Big|\mathrm{trace}(C_k^*(\langle f_{lm},e_{w_k}\rangle))\Big|\\&\le\sum_{k=1}^{\infty}\sqrt{\mathrm{trace}(C_k^*C_k)}\sqrt{\mathrm{trace}((\langle f_{lm},e_{w_k}\rangle)^*(\langle f_{lm},e_{w_k}\rangle))}\\&\le\sum_{k=1}^{\infty}\sqrt{\mathrm{trace}(C_k^*C_k)}\sqrt{\sum_{l=1}^{p}\sum_{m=1}^{q}\|f_{lm}\|^2_{H_2(\Pi)}}\\&=\sum_{k=1}^{\infty}\sqrt{\mathrm{trace}(C_k^*C_k)}\|F\|.\end{aligned}$$

Therefore, $\|F\|\le\sum_{k=1}^{\infty}\sqrt{\mathrm{trace}(C_k^*C_k)}<R$. □

The following theorem and its proof are suitable adaptations of the scalar-valued case.

Theorem 12. *Let $F(s)\in H(\mathcal{D},R)$. Use matrix AFD for $H_2^{p\times q}(\Pi)$ to decompose $F(s)$. Denote the kth remainder by $R_k(s)=F(s)-S_{k-1}(F)(s)$. If at each iteration of applying the maximal selection principle, we select $a_u\in\Pi$ and the orthogonal projection P_u of full rank, i.e., $\mathrm{rank}(P_u)=p$, then the kth remainder satisfies*

$$\|R_k\|\le\frac{R}{\sqrt{k}}.$$

Proof. We adopt the notations of Section 3. Since $\{B_u(s)\}$ is an orthonormal system, we have

$$[F(\cdot),B_k(\cdot)]=[R_k(\cdot),B_k(\cdot)]=M_k.$$

Then

$$\begin{aligned}\|R_{k+1}\|^2&=\mathrm{trace}[R_k(\cdot)-B_k(\cdot)M_k,R_k(\cdot)-B_k(\cdot)M_k]\\&=\|R_k\|^2-\mathrm{trace}(M_k^*M_k).\end{aligned}$$

If $F(s) \in H(\mathcal{D}, R)$, there exist sequences $\{w_k\} \subset \Pi$ and $\{C_k\} \subset \mathbb{C}^{p\times q}$ such that $F(s) = \sum_{k=1}^{\infty} C_k e_{w_k}(s)$. Hence, on one hand,

$$\begin{aligned}
&\|R_k\|^2\\
=&\text{trace}[R_k(\cdot), R_k(\cdot)] = \text{trace}[R_k(\cdot), F(\cdot)]\\
=&\text{trace}[R_k(\cdot), \sum_{k=1}^{\infty} C_k e_{w_k}(\cdot)]\\
=&\text{trace}(\sum_{k=1}^{\infty} C_k^* R_k(w_k)\sqrt{2\,\text{Re}(w_k)})\\
\le&\sum_{k=1}^{\infty}\sqrt{2\,\text{Re}(w_k)}\sqrt{\text{trace}(C_k^*C_k)}\sqrt{\text{trace}(R_k(w_k)^*R_k(w_k))}\\
\le&\sum_{k=1}^{\infty}\sqrt{\text{trace}(C_k^*C_k)}\sup_k\sqrt{2\,\text{Re}(w_k)\text{trace}(R_k(w_k)^*R_k(w_k))}\\
\le&R\sup_k\sqrt{2\,\text{Re}(w_k)\text{trace}(R_k(w_k)^*R_k(w_k))}.
\end{aligned}$$

On the other hand, since $B_{a,P}^{-1}(s)^*B_{a,P}^{-1}(s) \ge I_p$ for $s \in \Pi$, we have

$$\begin{aligned}
&\text{trace}(M_k^*M_k)\\
=&\sup_{\substack{s\in\Pi\\ P \text{ of rank } p}}\text{trace}(2\,\text{Re}(s)F_k(s)^*PF_k(s))\\
=&\sup_{s\in\Pi}2\,\text{Re}(s)\ \text{trace}(F_k(s)^*F_k(s))\\
=&\sup_{s\in\Pi}2\,\text{Re}(s)\ \text{trace}((\prod_{u=1}^{k-1}B_{a_k,P_k}(s)^{-1}R_k(s))^*\\
&\prod_{u=1}^{k-1}B_{a_k,P_k}(s)^{-1}R_k(s))\\
\ge&\sup_{s\in\Pi}2\,\text{Re}(s)\text{trace}(R_k(s)^*R_k(s))\\
\ge&\sup_{w_k}2\,\text{Re}(w_k)\text{trace}(R_k(w_k)^*R_k(w_k)).
\end{aligned}$$

Combining the above two aspects, we have $\text{trace}(M_k^*M_k) \ge \frac{\|R_k\|^4}{R^2}$. Thus,

$$\|R_{k+1}\|^2 \le \|R_k\|^2 - \frac{\|R_k\|^4}{R^2} = \|R_k\|^2(1-\frac{\|R_k\|^2}{R^2}).$$

From Proposition 11, we know $\|R_1\| = \|F\| < R$. According to the Lemma 3.3 in Qian and Wang (2013), we obtain that $\|R_k\| \le \frac{R}{\sqrt{k}}$. The proof is complete. □

5.2. Error analysis

In this part, we give an upper bound for the approximation error of the two-stage process designed in Section 4 for system identification in the presence of noise. Given measurements $\{G_{measure}(\mathrm{j}y_k)\}$, in the first stage, we construct an analytic mvf $F(s) = \mathcal{C}(\hat{G}_{measure})(s)$ using the measurements. In the second stage, we apply the matrix-valued AFD to $F(s)$ and use $RS_N(F)(s) = \frac{R_N(F)(s)+\widetilde{S_N}(F)(s)}{2}$ to approximate the transfer function $G(s)$. Due to the two-stage process, we divide the whole approximation error into two parts, one is caused by the measurements and computations, the other is caused by the truncated partial sum $S_N(F)(s)$. In the following theorem, we analyse the above two parts separately, leading to an upper bound for the whole approximation error.

Theorem 13. *Assume that each element of $G(\mathrm{j}y_k)$ is corrupted by additive noise ν_k, i.e., $G_{measure}(\mathrm{j}y_k) = G(\mathrm{j}y_k) + \nu_k$, $\pm k = 1, 2, \ldots, \frac{M}{2}$ with $0 < y_1 < y_2 < \cdots < y_{\frac{M}{2}} = T$, $\nu_k = \overline{\nu_{-k}}$, and ν_k satisfies one of the following two conditions:*

(1) All entries are bounded. We denote by $|\nu_k| < \epsilon$ for a small value $\epsilon > 0$.

(2) All entries are subjected to a complex Gaussian distribution $N^{\mathbb{C}}(0, \sigma^2)$ and the real and imaginary parts are independent and have equal variance $\frac{\sigma^2}{2}$. We denote this by $\nu_k \sim N^C(0, \sigma^2)$. This condition is based on the fact that after a discrete Fourier transform, filtered white noise is asymptotically circular complex normally distributed (Picinbono, 1993; Pintelon & Schoukens, 2012).

If $F(s) \in H(\mathcal{D}, R)$ and the frequency y belongs to a finite interval $[-T, T]$ where $T > 0$, then $\mathbb{E}\|RS_N(F) - G\|^2$ has an upper bound E as below:

$$E = \begin{cases} \frac{2R^2}{N} + 8CpqT(\Delta^2+\epsilon^2), & \text{when } |\nu| < \epsilon,\\ \frac{2R^2}{N} + 8CpqT(\Delta^2+\sigma^2), & \text{when } \nu \sim N^C(0,\sigma^2).\end{cases}$$

where Δ is the maximum of all elements in the matrix $\max_k \max_{y\in[y_k,y_{k+1})}|G(\mathrm{j}y) - G(\mathrm{j}y_k)|$. Furthermore,

$$\lim_{\substack{N,M\to\infty\\ \nu\to 0}}\mathbb{E}\|RS_N(F) - G\| = 0. \tag{9}$$

Proof. Due to $F(s) = \mathcal{C}(\hat{G}_{measure})(s)$ and $G(s) = \mathcal{C}(G)(s)$, by applying the triangle inequality and the norm inequality (7), we note that

$$\begin{aligned}
&\mathbb{E}\|S_N(F) - G\|^2\\
\le&2\mathbb{E}\|S_N(F) - F\|^2 + 2\mathbb{E}\|\mathcal{C}(\hat{G}_{measure}) - \mathcal{C}(G)\|^2\\
\le&2\mathbb{E}\|S_N(F) - F\|^2 + 4C\mathbb{E}\|\hat{G} - G\|^2 + 4C\mathbb{E}\|\nu\|^2.
\end{aligned}$$

Here, $\|\mathcal{C}(\hat{G}_{measure}) - \mathcal{C}(G)\|^2$ causes the first-stage error; and $\|S_N(F) - F\|^2$ causes the second-stage error.

Denote Δ as the maximum of all elements in the matrix $\max_k \max_{y\in[y_k,y_{k+1})}|G(\mathrm{j}y) - G(\mathrm{j}y_k)|$. Together with the convergence rate in Theorem 12, we have

(1) when $|\nu_k| < \epsilon$,

$$\begin{aligned}
\mathbb{E}\|S_N(F) - G\|^2 \le&\frac{2R^2}{N} + 4Cpq\int_{-T}^{T}(\Delta^2+\epsilon^2)\mathrm{d}y\\
\le&\frac{2R^2}{N} + 8CpqT(\Delta^2+\epsilon^2).
\end{aligned}$$

(2) when $\nu_k \sim N^{\mathbb{C}}(0, \sigma^2)$,

$$\begin{aligned}
\mathbb{E}\|S_N(F) - G\|^2 \le&\frac{2R^2}{N} + 4Cpq\int_{-T}^{T}(\Delta^2+\mathbb{E}|\nu|^2)\mathrm{d}y\\
\le&\frac{2R^2}{N} + 8CpqT(\Delta^2+\sigma^2).
\end{aligned}$$

Since $\|\widetilde{S_N}(F) - G\| = \|S_N(F) - G\|$, we obtain

$$\begin{aligned}
\mathbb{E}\|RS_N(F) - G\|^2 =&\mathbb{E}\|\frac{S_N(F)+\widetilde{S_N}(F)}{2} - G\|^2\\
\le&\frac{1}{2}\mathbb{E}(\|S_N(F) - G\|^2 + \|\widetilde{S_N}(F) - G\|^2).
\end{aligned}$$

The proof is complete. □

Remark 14. Theorem 13 declares the convergence of the approximation error when the noise level ν tends to zero. It may be concluded from Pintelon and Schoukens (2012) that $\nu \to 0$ holds, where a periodic excitation consisting of a fixed number of harmonically related frequencies is used, and the number of the measured time domain samples Q tends to infinity. It is noted that in this case the standard deviation of the additive frequency noise ν at the excited DFT frequencies decreases as $\mathcal{O}(Q^{-\frac{1}{2}})$. It is worth emphasizing that for broadband random excitations we cannot conclude the convergence formula (9). That is because the standard derivation of the additive frequency noise ν at the excited DFT frequencies is of the level $\mathcal{O}(Q^0)$, amounting that the noise level ν remains steady. However, the upper bound obtained in Theorem 13 is still valid.

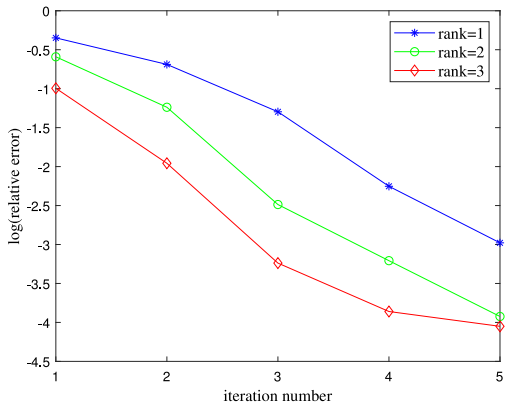


Fig. 1. The REs obtained by the matrix AFD in the previous 5 iterations when we choose the orthogonal projection with different rank values using the noise-free measurements with the number of measurements $M = 4000$. The frequency $\{y_k\}$ of the measurements belongs to the frequency band $[-200, 200]$. The REs are shown using a base-10 logarithmic scale.

6. Numerical experiments

We give an example to show the effectiveness of the proposed algorithm. Given measurements of any function $F(s)$ and its approximation $F_n(s)$, the relative error (RE) below is defined to evaluate the approximating performance:

$$RE = \frac{\sum_{k=1}^{M} \text{trace}[(F(\mathrm{j}y_k) - F_n(\mathrm{j}y_k))^*(F(\mathrm{j}y_k) - F_n(\mathrm{j}y_k))]}{\sum_{k=1}^{M} \text{trace}(F(\mathrm{j}y_k)^*F(\mathrm{j}y_k))}.$$

Example. We consider a continuous time system used in Garnier, Gilson, Bastogne, and Mensler (2008) and show it in the following transfer function form:

$$G(s) = \frac{1}{E(s)} \begin{bmatrix} g_{11}(s) & g_{12}(s) & g_{13}(s) \\ g_{21}(s) & g_{22}(s) & g_{23}(s) \\ g_{31}(s) & g_{32}(s) & g_{33}(s) \end{bmatrix}$$

where

$$\begin{aligned} E(s) =& s^3 + 8.6813s^2 + 26.3493s + 23.4799; \\ g_{11}(s) =& -4.7167s^2 - 31.4326s - 70.9014; \\ g_{12}(s) =& 1.4503s^2 + 9.7398s + 11.5931; \\ g_{13}(s) =& 1.4213s^2 + 4.4633s + 6.7293; \\ g_{21}(s) =& -2.6966s^2 - 10.7190s - 5.4670; \\ g_{22}(s) =& -2.3246s^2 - 6.4052s - 4.3889; \\ g_{23}(s) =& 4.8762s^2 + 26.5256s + 28.5751; \\ g_{31}(s) =& -0.5145s^2 + 3.3836s + 2.4385; \\ g_{32}(s) =& 3.2158s^2 + 22.8679s + 27.2990; \\ g_{33}(s) =& 3.1153s^2 + 6.9512s + 3.7446. \end{aligned}$$

In Fig. 1, we show the REs obtained by the matrix AFD when we set the chosen orthogonal projection by different ranks as $1, 2, 3$ respectively in the noise-free case. Being consistent with the theoretical analysis in Theorem 7 and the statement in Remark 8, the larger the rank of the orthogonal projection is, the faster the algorithm converges. In addition, the algorithm converges to a satisfactory approximation within several iterations in all the 3 cases. As supplement to Fig. 1, we give the REs data when we choose orthogonal projection with different ranks and different numbers of measurements by using noise-free and noisy measurements respectively in Table 1. From the table, we find that under interference of noise, the algorithm is still effective.

Table 1
The REs obtained by the matrix AFD with different types of noise: (1) The noise $N^{\mathbb{C}}(0, \sigma^2)$ is Gaussian, σ being set to be $\delta \max_k \{|G(\mathrm{j}y_k)|\}$ where δ is a parameter. (2) Bounded noise ν satisfies $|\nu| < \epsilon$, ϵ being set to be $\delta \max_k \{|G(\mathrm{j}y_k)|\}$ where δ is a parameter. (3) N is the number of iterations. (4) k_0 is the rank of the chosen orthogonal projection. (5) M is the number of measurements.

(a) noise-free						
M	N	1	2	3	4	5
2000	$k_0 = 1$	0.4486	0.2031	0.0498	0.0053	9.97e−04
	$k_0 = 2$	0.2566	0.0578	0.0032	6.037e−04	1.18e−04
	$k_0 = 3$	0.1012	0.0110	5.91e−04	1.57e−04	1.15e−04
4000	$k_0 = 1$	0.4485	0.2048	0.0505	0.0055	0.0011
	$k_0 = 2$	0.2566	0.0578	0.0032	6.20e−04	1.19e−04
	$k_0 = 3$	0.1012	0.0110	5.77e−04	1.39e−04	8.91e−05
(b) Gaussian noise with $\delta = 0.01$						
M	N	1	2	3	4	5
2000	$k_0 = 1$	0.4486	0.1966	0.0483	0.0050	9.97e−04
	$k_0 = 2$	0.2565	0.0575	0.0035	6.56e−04	1.97e−04
	$k_0 = 3$	0.1012	0.0111	6.33e−04	2.05e−04	1.71e−04
4000	$k_0 = 1$	0.4486	0.1961	0.0491	0.0048	9.31e−04
	$k_0 = 2$	0.2567	0.0570	0.0033	6.41e−04	1.72e−04
	$k_0 = 3$	0.1012	0.0111	6.03e−04	1.65e−04	1.18e−04
(c) Gaussian noise with $\delta = 0.02$						
M	N	1	2	3	4	5
2000	$k_0 = 1$	0.4487	0.2174	0.0521	0.0061	0.0014
	$k_0 = 2$	0.2560	0.0593	0.0033	9.33e−04	4.40e−04
	$k_0 = 3$	0.1013	0.0112	7.91e−04	4.57e−04	5.23e−04
4000	$k_0 = 1$	0.4486	0.1941	0.0480	0.0051	0.0011
	$k_0 = 2$	0.2569	0.0595	0.0029	7.72e−04	2.74e−04
	$k_0 = 3$	0.1012	0.0112	7.10e−04	3.89e−04	3.74e−04
(d) bounded noise with $\delta = 0.01$						
M	N	1	2	3	4	5
2000	$k_0 = 1$	0.4486	0.2000	0.0489	0.0052	9.77e−04
	$k_0 = 2$	0.2564	0.0571	0.0033	6.17e−04	1.38e−04
	$k_0 = 3$	0.1012	0.0110	5.83e−04	1.52e−04	1.09e−05
4000	$k_0 = 1$	0.4486	0.2054	0.0503	0.0056	0.0011
	$k_0 = 2$	0.2566	0.0577	0.0033	6.46e−04	1.16e−04
	$k_0 = 3$	0.1012	0.0110	5.77e−04	1.39e−04	9.72e−05
(e) bounded noise with $\delta = 0.02$						
M	N	1	2	3	4	5
2000	$k_0 = 1$	0.4486	0.1943	0.0484	0.0050	0.0011
	$k_0 = 2$	0.2565	0.0575	0.0033	7.09e−4	2.01e−04
	$k_0 = 3$	0.1012	0.0111	6.81e−04	2.63e−04	2.95e−04
4000	$k_0 = 1$	0.4486	0.1963	0.0484	0.0049	9.52e−04
	$k_0 = 2$	0.2566	0.0571	0.0035	6.51e−04	1.73e−04
	$k_0 = 3$	0.1012	0.0111	6.24e−04	1.98e−04	1.94e−04

Besides, the relative errors are not sensitive to the noise level and the number of measurements. As in the scalar-valued cases, this shows the robustness of the matrix AFD. In Table 2, we list the parameter of the Blaschke–Potapov factor chosen by the matrix AFD in the previous 3 iterations in both the noise-free and noisy case. We choose to show the results corresponding to the case $\delta = 0.02$ and $M = 4000$ in Table 1. As we can see from the table, the selection of the parameter is relatively stable.

As comparison, we use the output error model estimation (OE) in the system identification toolbox of matlab to estimate the transfer function as well.

In the OE method, the transfer function is assumed to be a rational function $G(s) = \frac{B(s)}{A(s)}$ where $B(s), A(s)$ are polynomials and the method gives an estimation for the coefficients of $B(s), A(s)$. Considering that it is possible for the OE method to give complex valued estimation for the coefficients, we apply the same technique mentioned in Section 4.2 to obtain a transfer function estimation with real coefficients. In Table 3, we present the relative errors obtained by the OE method with different degrees of $B(s), A(s)$, different types of noisy data and different numbers

Table 2
The parameters $\{a_n\}$ of the Blaschke–Potapov factor selected by the matrix AFD in the previous 3 iterations with different types of noise: (1) The noise $N^{\mathbb{C}}(0, \sigma^2)$ is Gaussian, σ being set to be $0.02 \max_k \{|G(\mathrm{j}y_k)|\}$. (2) Bounded noise ν satisfies $|\nu| < \epsilon$, ϵ being set to be $0.02 \max_k \{|G(\mathrm{j}y_k)|\}$. (3) The number of measurements is 4000. (4) N is the number of iterations. (5) k_0 is the rank of the orthogonal projection.

(a) $k_0 = 1$

N	noise-free	Gaussian noise	bounded noise
1	2.1277	2.0957	2.1277
2	3.5319 + 1.0638j	3.5000 + 1.2057j	3.4681 + 1.2057j
3	2.8936 − 1.0638j	2.8936 − 1.2057j	2.8936 − 1.2057j

(b) $k_0 = 2$

N	noise-free	Gaussian noise	bounded noise
1	2.5106	2.5106	2.5106
2	2.9255	2.8936 − 0.0709j	2.8936 − 0.0709j
3	3.2128	3.1489 + 0.0709j	3.3404 + 0.0709j

(c) $k_0 = 3$

N	noise-free	Gaussian noise	bounded noise
1	2.6064	2.6064	2.6064
2	2.7660	2.7021	2.7660
3	2.7340	2.7340 − 0.0709j	2.6064 − 0.0709j

Table 3
The REs obtained by the OE method with different degrees of the polynomials $B(s), A(s)$ and different kinds of noise (the additive Gaussian noise $N(0; \sigma^2)$ and the additive bounded noise with bound ϵ), d representing the degree, degree $B(s)$ = degree $A(s)$, the level of the noise as σ(or ϵ) equal to the maximum of the matrix $\delta \max_k\{|G(\mathrm{j}y_k)|\}$ with $\delta = 0.01, 0.02$ respectively, M being the number of the measurements.

(a) $M = 2000$

d	noise-free	Gaussian noise		Bounded noise	
	–	$\delta = 0.01$	$\delta = 0.02$	$\delta = 0.01$	$\delta = 0.02$
1	0.0124	0.0117	0.0137	0.0132	0.0512
2	2.17e−04	0.0177	0.0206	0.0100	0.0090
3	4.13e−08	0.0110	0.0390	0.0013	0.0445
4	1.62e−11	0.0337	0.0767	2.07e−04	0.0517
5	7.31e−10	0.0012	0.0088	2.97e−04	4.47e−04

(b) $M = 4000$

d	noise-free	Gaussian noise		Bounded noise	
	–	$\delta = 0.01$	$\delta = 0.02$	$\delta = 0.01$	$\delta = 0.02$
1	0.0124	0.0132	0.0133	0.0116	0.0116
2	2.17e−04	0.0103	0.0357	0.0087	0.0116
3	4.01e−08	0.0082	0.0107	0.0109	0.0526
4	2.77e−07	0.0190	0.1402	0.0023	0.0138
5	2.35e−07	0.0041	0.8277	3.96e−04	0.0016

of the measurements. We found that even in the noise free case, the RE when $d = 3$, which is the order of the true system, is not the smallest. It is unexpected, from the theory of OE implies that the OE method gives asymptotically unbiased estimation. This phenomenon is caused by sinking into a local minimum inherited from the gradient type methods. The crucial thing is the initial state to start with. In the noisy case, the OE method does not give results as stable as those given by the proposed method. The results of the OE method can be greatly influenced by the noise. The fact is that due to the large number of the estimated parameters, MIMO systems are likely to get stuck in local minima and unstable when applying the PEM methods. Table 4 shows the REs obtained by the OE method starting from an initialization system deduced from the subspace method (n4sid in Matlab). It is observed that in the noise free case, the RE when $d = 3$ is the smallest and almost 0, and in the noisy case the results are more stable. Although the OE method seems superior, in the OE model the initialization is important and the degree numbers of $B(s), A(s)$ are critical. It requires an appropriate model order selection, for example, using the AIC or MDL criterion.

Table 4
The REs obtained by the OE method under an initialization derived by the subspace method with notations as the same as those in Table 3.

(a) $M = 2000$

d	noise-free	Gaussian noise		Bounded noise	
	–	$\delta = 0.01$	$\delta = 0.02$	$\delta = 0.01$	$\delta = 0.02$
1	0.0387	0.0384	0.0399	0.0382	0.0387
2	0.0093	0.0081	0.0064	0.0072	0.0088
3	3.13e−32	1.32e−04	4.92e−04	3.50e−04	1.38e−04
4	4.18e−32	1.84e−04	6.43e−04	4.45e−04	2.24e−04
5	4.44e−32	2.18e−04	8.87e−04	6.21e−04	2.77e−04

(b) $M = 4000$

d	noise-free	Gaussian noise		Bounded noise	
	–	$\delta = 0.01$	$\delta = 0.02$	$\delta = 0.01$	$\delta = 0.02$
1	0.0387	0.0387	0.0370	0.0384	0.0368
2	0.0093	0.0087	0.0089	0.0073	0.0106
3	3.83e−32	5.24e−05	3.00e−04	2.69e−05	7.31e−05
4	4.17e−32	6.23e−05	3.73e−04	3.04e−05	8.96e−05
5	5.01e−32	9.48e−05	4.44e−04	4.33e−05	1.42e−04

Table 5
The REs obtained by the subspace method using the function n4sid in Matlab with notations as the same as those in Table 3.

(a) $M = 2000$

d	noise-free	Gaussian noise		Bounded noise	
	–	$\delta = 0.01$	$\delta = 0.02$	$\delta = 0.01$	$\delta = 0.02$
1	0.4521	0.4520	0.4530	0.4520	0.4518
2	0.2083	0.2088	0.2077	0.2081	0.2078
3	2.71e−30	5.92e−05	0.0046	3.40e−05	2.50e−04
4	3.69e−30	1.50e−04	0.0073	8.89e−05	2.81e−05
5	2.38e−29	0.0051	0.0553	2.61e−04	0.0033

(b) $M = 4000$

d	noise-free	Gaussian noise		Bounded noise	
	–	$\delta = 0.01$	$\delta = 0.02$	$\delta = 0.01$	$\delta = 0.02$
1	0.4521	0.4526	0.4525	0.4519	0.4519
2	0.2083	0.2082	0.2073	0.2083	0.2080
3	2.35e−30	5.11e−05	9.52e−04	1.27e−05	7.31e−05
4	6.38e−30	7.16e−05	0.0012	1.59e−05	1.32e−04
5	6.02e−30	1.31e−04	0.0624	3.71e−05	4.74e−04

In Table 5, we show the REs obtained by the subspace method (n4sid in Matlab). Comparing Table 4 with Table 5, it is true that we achieve more accurate approximations by applying the OE method starting from an initialization obtained by the subspace method than by simply applying the subspace method.

We give some figures to intuitively compare the approximations obtained by four methods, which are the proposed method, the OE method, the OE method with an initialization deduced from the subspace method and the subspace method, with the true system in the noisy case.

(1) In Fig. 2, we show the Nyquist plots of the 11-th entry of the true system, the 11-th entries of the approximations obtained by the proposed method with different iteration numbers, the 11-th entries of the approximations obtained by the OE method with different degrees based on an initialization deduced from the subspace method and the approximations obtained by the subspace method with different degrees.
(2) In Figs. 3 and 4, we separately show the singular values of the matrix-valued frequency response $G(\mathrm{j}y)$ at different frequencies $\{y_k\}$ of the true system, the approximation systems obtained by the proposed method with different ranks of the selected orthogonal projections, the approximation systems obtained by the OE method, the approximation systems with an initialization and the approximation systems obtained by the subspace method.

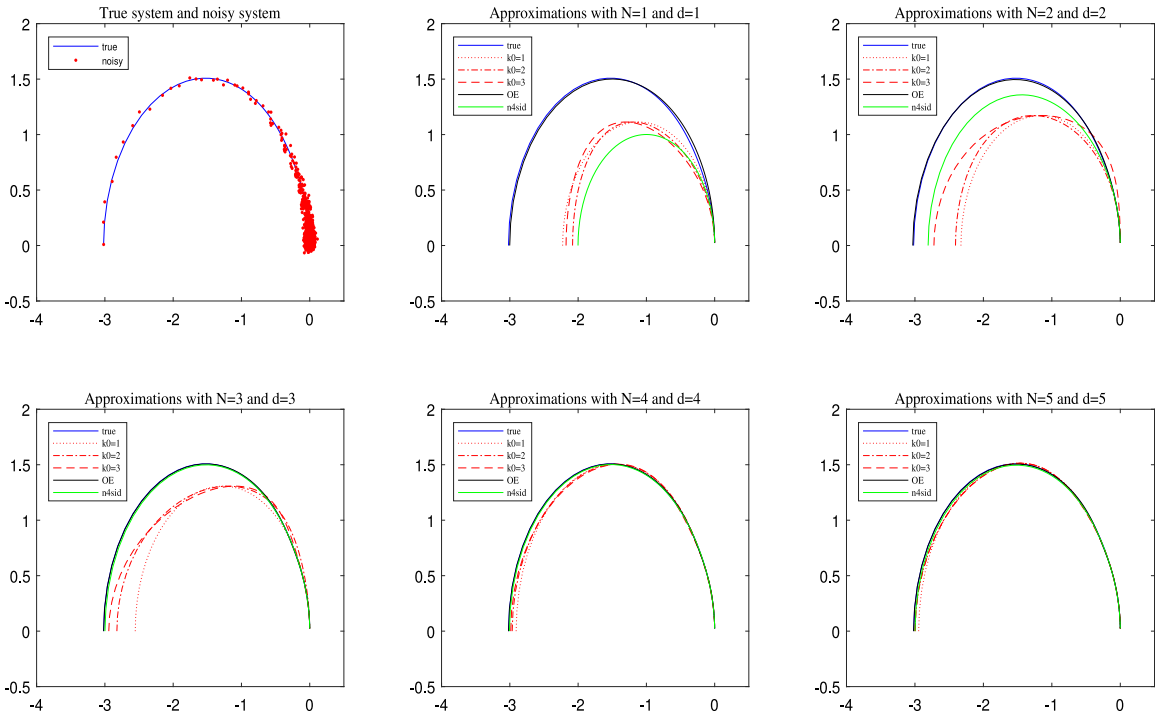


Fig. 2. The Nyquist plots of the true system, the noisy system, the approximations obtained by the proposed method, the approximations obtained by the OE method starting from an initialization deduced by the subspace method (n4sid in Matlab) and the approximations obtained by the subspace method (n4sid in matlab) for the entry at the first row and the first column in the transfer function of the example. The subfigure in the first row and the first column shows the Nyquist plots of the noise-free and noisy frequency response data. Here the noise is additive Gaussian noise and $\delta = 0.01$. The other subfigures are the Nyquist plots of the approximations obtained by using the noisy data but setting different iteration numbers N (for the proposed method) and different degree values d of $B(z), A(z)$ (for the OE method). k_0 represents the rank of the selected orthogonal projection at each iteration by the proposed method. In the title of these subfigures, we indicate the value of N and d. Here, the number of the measurements $M = 4000$.

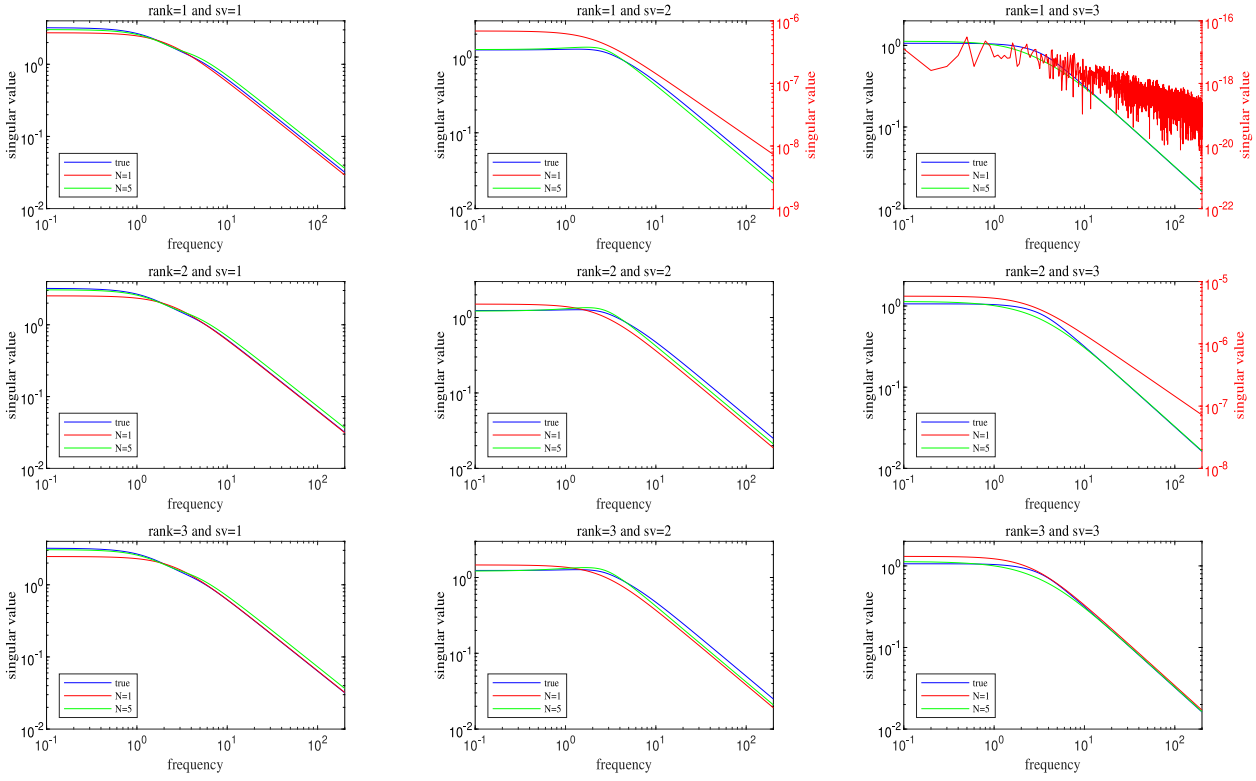


Fig. 3. The singular values of the frequency responses of the true system and the approximations obtained by the proposed method in the noisy case when the iteration number $N = 1, 5$. Here, the noise is the additive Gaussian noise and $\delta = 0.01$ and the number of the measurements $M = 4000$. In the 3 subfigures of each row, we successively show the 3 singular values of the 3×3-valued frequency responses in descending order. sv stands for the order number of the sorted singular values. In the 3 subfigures of each column, we successively show the singular values of the approximations obtained by the proposed method by choosing orthogonal projections with rank = 1,2,3. In all the subfigures, when the labels of the left y axis and the right y axis are both presented, the data on the red line is read according to the right y axis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

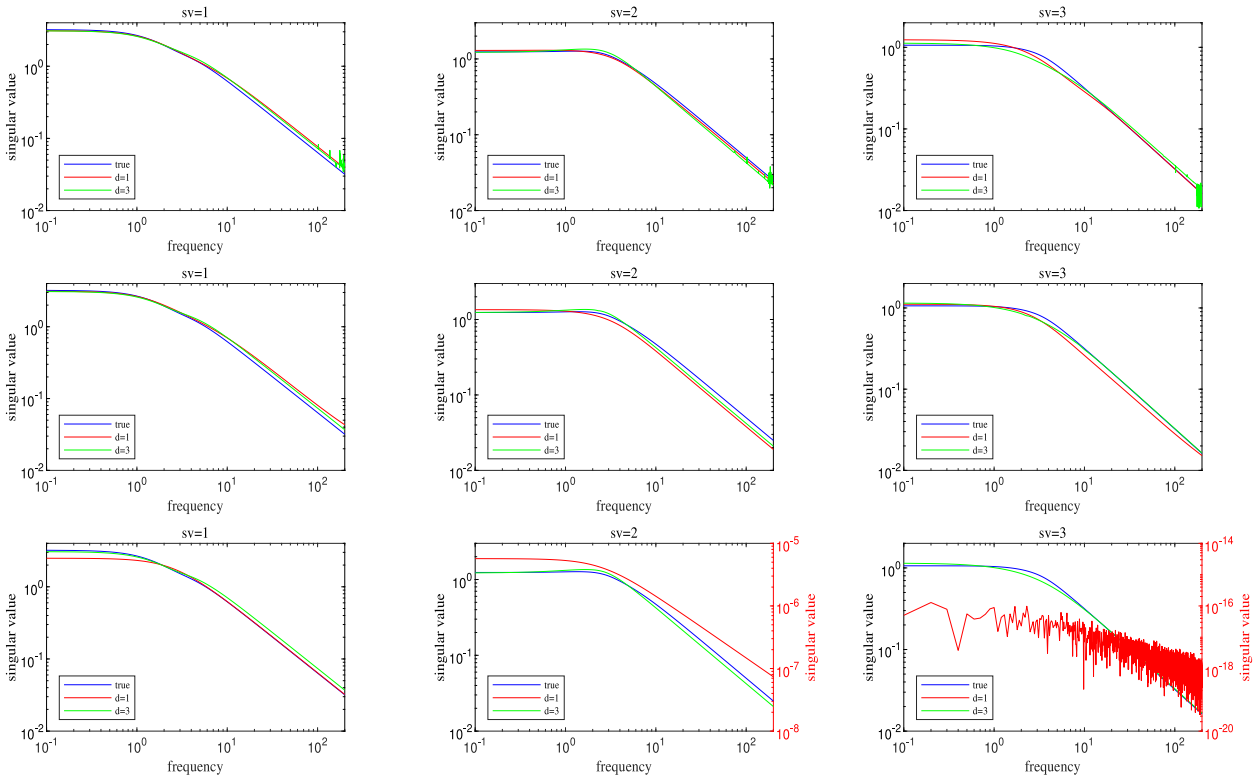


Fig. 4. The singular values of the frequency responses of the true system, the approximations obtained by the OE method, the approximations obtained by the OE method with an initialization and the approximations obtained by the subspace method in the noisy case when the degree number $d = 1, 3$. Here, the noise is the additive Gaussian noise and $\delta = 0.01$ and the number of the measurements $M = 4000$. In the 3 subfigures of each row, we successively show the 3 singular values of the 3×3-valued frequency responses in descending order. sv stands for the order number of the sorted singular values. In the first row, the singular values of the approximations obtained by directly applying the OE method are shown. In the second row, the singular values of the approximations obtained by applying the OE method starting from an initialization using the subspace method are shown. In the third row, the singular values of the approximations obtained by directly applying the subspace method (n4sid in Matlab) are shown. In all the subfigures, when the labels of the left y axis and the right y axis are both presented, the data on the red line is read according to the right y axis.

Among all the singular values of the frequency response $G(y_k)$ at certain frequency y_k, it is observed that in terms of the singular values, the smaller the true singular value is, the more difficult to obtain accurate approximations. The numerical experiments show that the true system is well approached by the proposed method when the iteration number $N = 5$ and the rank of the selected orthogonal projection rank $= 3$, and is also well approached by the OE method starting from an initialization deduced by the subspace method when the degree number $d = 3$.

Remark 15. Robustness to the noise is an inherent advantage of the proposed matrix AFD. Due to the reproducing property of Szegö kernel $e_a(s), a \in \Pi$, we can obtain $2\,\mathrm{Re}\,(a)F(a)F(a)^*$ appearing in Theorem 7 by the inner products $\{\langle f_{lm}, e_a\rangle\}$, where $f_{lm}(s)$ is the lm-th entry of $F(s)$. The integrals smooth the noise out on the measurements $\{F(\mathrm{j}y_k)\}$. In addition, out of a considerable amount of measurements $\{G_{\mathrm{measure}}(\mathrm{j}y_k)\}$, the rational approximations are determined by a few parameters, which avoids overfitting in a certain extent. However, in the numerical experiments, accumulation of the calculation errors may lead to ineffective approximations. To avoid this, in practice, when we find that the relative error obtained does not decrease with a prescribed threshold, we stop the iteration.

7. Conclusion

The present work extends the matrix AFD result by Alpay et al. (2017), that is for the functions on the unit disc context corresponding to Z-transformations of the discrete systems, to the functions on the right-half complex plane context corresponding to Laplace transformations of the continuous systems. The key point of the AFD type methods is that they perform the most efficient matching pursuit strategy with the one-step-optimal selection model. The matrix-valued contexts concerned in Alpay et al. (2017) and the present paper are generalizations of Qian and Wang (2011). The corresponding algorithms in the two matrix-valued contexts are made practical through crucial use of a corollary of the famous Poincaré Separation Theorem (Durbin & Watson, 1950). The efficiency of the proposed algorithm is evidenced by the mathematical estimation on the convergence rate $\mathcal{O}(k^{-\frac{1}{2}})$, as well as experiments in comparison with the OE model, in both the noisy and the noise-free cases. The AFD type methods have been successfully applied to the field of frequency domain system identification including SISO system identification in Mi and Qian (2011) and SISO multi-dimensional system identification in Wang, Qian, Leong, and Gao (2020). With the present study we successfully extend the applications to the MIMO type system identification by using the matrix AFD. It is prospected that one can consider identification problems of more complex systems, including nonlinear systems and time-variant systems, and explore the possibility of AFD type algorithms in the contexts in which Blaschke products are unavailable (Qu & Dang, 2019a, 2019b).

Acknowledgements

The authors sincerely thank D. Alpay, F. Colombo and I. Sabadini for the previous collaborative studies and knowledge exchanges in relation to this work.

Appendix A. Some proofs involved in Section 3.1

Proof of Lemma 4. Using Definition 2 directly, we obtain

$$P_0H(a_0) = P_0F(a_0) - P_0^2F(a_0)\frac{2\,\mathrm{Re}(a_0)}{a_0+\overline{a_0}} = 0.$$

According to the $\mathbb{C}^{q\times q}$-valued inner product (1) and the properties (b), (c) of the Szegö kernel $e_a(s)$ in Definition 2, we obtain

$$\begin{aligned}
&[H, H_0]\\
=&\frac{1}{2\pi}\int_{\mathbb{R}}(P_0F(a_0)e_{a_0}(\mathrm{j}y)\sqrt{2\,\mathrm{Re}(a_0)})^*F(\mathrm{j}y)\mathrm{d}y\\
&-\frac{1}{2\pi}\int_{\mathbb{R}}2\,\mathrm{Re}(a_0)F(a_0)^*P_0F(a_0)\overline{e_{a_0}(\mathrm{j}y)}e_{a_0}(\mathrm{j}y)\mathrm{d}y\\
=&F(a_0)^*P_0F(a_0)2\,\mathrm{Re}(a_0) - F(a_0)^*P_0F(a_0)2\,\mathrm{Re}(a_0)\\
=&0.
\end{aligned}$$

Thus, $[F, F] = [H + H_0, H + H_0] = [H, H] + [H_0, H_0]$.

From $P_0H(a_0) = 0$, we can formulate $P_0H(s) = (s - a_0)R(s)$, where $R(s)$ is $\mathbb{C}^{p\times q}$-valued and analytic at a_0. Due to the formulation of $B^{-1}_{a,P}(s)$ in Definition 3,

$$\begin{aligned}
B^{-1}_{a_0,P_0}(s)H(s) &= (I_p - P_0 + P_0\frac{s+\overline{a_0}}{s-a_0})H(s)\\
&= H(s) + 2\,\mathrm{Re}(a_0)R(s),
\end{aligned}$$

meaning that $B^{-1}_{a_0,P_0}(s)H(s)$ is analytic at a_0. Moreover, because of $B^{-1}_{a_0,P_0}(\mathrm{j}y)^*B^{-1}_{a_0,P_0}(\mathrm{j}y) = I_p$,

$$\begin{aligned}
[G, G] &= [B^{-1}_{a_0,P_0}H, B^{-1}_{a_0,P_0}H]\\
&= \frac{1}{2\pi}\int_{\mathbb{R}}H(\mathrm{j}y)^*B^{-1}_{a_0,P_0}(\mathrm{j}y)^*B^{-1}_{a_0,P_0}(\mathrm{j}y)H(\mathrm{j}y)\mathrm{d}y\\
&= \frac{1}{2\pi}\int_{\mathbb{R}}H(\mathrm{j}y)^*H(\mathrm{j}y)\mathrm{d}y = [H, H].
\end{aligned}$$

Therefore, $G(s) \in H_2^{p\times q}(\Pi)$. The proof is complete. □

Proof of Theorem 5. (a) We show that $\mathrm{trace}(2\,\mathrm{Re}(s)F(s)^*PF(s))$ is uniformly bounded by $\|F\|^2$, then it has a finite supremum.

Write $F(s) \in H_2^{p\times q}(\Pi)$ in its entries $f_{lm}(s)$, $l = 1, \ldots, p$, $m = 1, \ldots, q$ as $F(s) = (f_{lm}(s))$. Recall that for $a \in \Pi$ and entry $f_{lm}(s)$ we have

$$|\sqrt{2\,\mathrm{Re}(a)}f_{lm}(a)| = |\langle f_{lm}, e_a\rangle| \le \|f_{lm}\|_{H_2(\Pi)}.$$

Thus, since the orthogonal projection P satisfies $P \le I_p$, we obtain

$$\begin{aligned}
&\mathrm{trace}(2\,\mathrm{Re}(s)F(s)^*PF(s))\\
\le&\ \mathrm{trace}(2\,\mathrm{Re}(s)F(s)^*F(s))\\
=&\sum_{l=1}^{p}\sum_{m=1}^{q}|\sqrt{2\,\mathrm{Re}(s)}f_{lm}(s)|^2\\
\le&\sum_{l=1}^{p}\sum_{m=1}^{q}\|f_{lm}\|^2_{H_2(\Pi)} = \|F\|^2.
\end{aligned}$$

(b) We claim that among all possible selections of $s \in \Pi$ and orthogonal projection P with rank k_0, the finite supremum of $\mathrm{trace}(2\,\mathrm{Re}(s)F(s)^*PF(s))$ is attainable at a point inside Π. It involves a density argument. Similar proofs can be found in Mi et al. (2012), Qian (2010) and Qu and Dang (2019a). Below we include the details in the present case.

We first indicate the particulars with this unbounded domain situation for $f \in H_2(\Pi)$. We show that for any $f(s) \in H_2(\Pi)$ and $\epsilon > 0$, there exist $\delta > 0$, $R > 0$, and the associated neighbourhood of the boundary of Π,

$$B(\delta, R) = \{s \in \Pi : \mathrm{Re}(s) < \delta\} \cup \{s \in \Pi : |s| > R\},$$

such that whenever $a \in B(\delta, R)$, there holds that $|\langle f, e_a\rangle| < \epsilon$. It amounts that $|\langle f, e_a\rangle| \to 0$ when a tends to the boundary of Π.

From Definition 2, we know that $\mathcal{D} = \{e_a(s), a \in \Pi\}$ is a dictionary of $H_2(\Pi)$, which implies that $\mathcal{D}$ is dense in $H_2(\Pi)$. Thus $\mathrm{span}\{\mathcal{D}\}$, which consists of all finite linear combinations of the elements in $\mathcal{D}$, is also dense in $H_2(\Pi)$. As a consequence, any function $f(s) \in H_2(\Pi)$ can be infinitely approximated by a sequence of functions of which each is a finite linear combination of parameterized Szegö kernels. Therefore, for $\forall\epsilon > 0$, there exist a positive integer N_0 and a function $g(s)$ having the following form

$$g(s) = \sum_{l=1}^{N_0}\frac{c_l}{s+\overline{a_l}},$$

where $c_l \in \mathbb{C}, a_l \in \Pi$, such that $\|f - g\|_{H_2(\Pi)} < \frac{\epsilon}{2}$. From the triangle inequality and the Cauchy–Schwartz inequality, for $\forall a \in \Pi$, we have

$$|\langle f, e_a\rangle| \le |\langle g, e_a\rangle| + \|f - g\|_{H_2(\Pi)} < |\langle g, e_a\rangle| + \frac{\epsilon}{2}.$$

Denote $C_0 = \max_{1\le l\le N_0}|c_l|$, $\delta_1 = \min_{1\le l\le N_0}\mathrm{Re}(a_l) > 0$, $R_1 = \max_{1\le l\le N_0}|a_l|$. From the reproducing property (c) in Definition 2 and $|a + \overline{a_l}| \ge \mathrm{Re}(a + \overline{a_l}) > \mathrm{Re}(\overline{a_l}) \ge \delta_1$,

$$|\langle g, e_a\rangle| = \sqrt{2\,\mathrm{Re}(a)}|\sum_{l=1}^{N_0}\frac{c_l}{a+\overline{a_l}}| \le \sqrt{2\delta}\frac{N_0C_0}{\delta_1}.$$

It implies $|\langle g, e_a\rangle| < \frac{\epsilon}{2}$ whenever $\delta < (\frac{\delta_1\epsilon}{2\sqrt{2}N_0C_0})^2$. If $a \in \Pi$ satisfies $|a| > 2R_1$, we obtain $|a + a_l| \ge |a| - |a_l| \ge \frac{|a|}{2}$. Then if we let $R = \max\{R_1, (\frac{4\sqrt{2}N_0\max_{1\le l\le N_0}\{|c_l|\}}{\epsilon})^2\}$, whenever $|a| > R$,

$$|\langle g, e_a\rangle| \le \sqrt{2|a|}\ \frac{N_0\max_{1\le l\le N_0}\{|c_l|\}}{\frac{|a|}{2}} < \frac{\epsilon}{2}.$$

To sum up, whenever $a \in B(\delta, R)$, there holds $|\langle f, e_a\rangle| < |\langle g, e_a\rangle| + \frac{\epsilon}{2} < \epsilon$.

Considering that

$$\mathrm{trace}(2\,\mathrm{Re}(a)F(a)^*F(a)) = \sum_{\substack{l=1,\ldots,p,\\ m=1,\ldots,q}}\|\sqrt{2\,\mathrm{Re}(a)}f_{lm}(a)\|^2_{H_2(\Pi)},$$

for $\forall\epsilon > 0$, denote by $B(\delta_{lm}, R_{lm})$ the set such that whenever $a \in B(\delta_{lm}, R_{lm})$, there holds $|\langle f_{lm}, e_a\rangle| < \epsilon$. Then the supremum of $\mathrm{trace}(2\,\mathrm{Re}(a)F(a)^*F(a))$ is attainable in a compact set having the form $\mathbb{S} = \Pi \setminus \cap_{\substack{l=1,\ldots,p,\\ m=1,\ldots,q}} B(\delta_{lm}, R_{lm})$.

It is noted that the orthogonal projections of rank k_0 constitute a compact set, denoted by $\mathbb{P}$. For the orthogonal projection $P = V^*V$, as mentioned in Definition 3,

$$\begin{aligned}
&\mathrm{trace}(2\,\mathrm{Re}(s)F(s)^*PF(s))\\
=&\mathrm{trace}(2\,\mathrm{Re}(s)(VF(s))^*VF(s)).
\end{aligned}$$

Since the entries of $VF(s)$ are linear combinations of the entries of $F(s)$, we note that by adding the orthogonal projection P, $\mathrm{trace}(2\,\mathrm{Re}(s)F(s)^*PF(s))$ also tends to 0 when s tends to the boundary of Π. Thus the supremum of $\mathrm{trace}(2\,\mathrm{Re}(s)F(s)^*PF(s))$ is attainable in a compact set having the form $\mathbb{S}\times\mathbb{P}$ as product of two compact sets. Therefore, a pair (a_0, P_0) giving rise to the supremum of $\mathrm{trace}(2\,\mathrm{Re}(s)F(s)^*PF(s))$ exists. □

Appendix B. A corollary of the famous Poincaré separation theorem

Lemma 16 (*Durbin & Watson, 1950*). *Let A be a $p\times p$ positive semidefinite matrix and P be a $p\times p$ orthogonal projection such that $\mathrm{rank}(P) = k_0$, where $k_0 \in \{1, 2, \ldots, p\}$. Denote*

$\Phi = [\varphi_1, \varphi_2, \ldots, \varphi_p]^*$, where the column vectors $\varphi_1, \varphi_2, \ldots, \varphi_p$ are eigenvectors of A corresponding to $\lambda_1(A), \lambda_2(A), \ldots, \lambda_p(A)$ respectively and satisfy $\Phi\Phi^* = I_p$. Here, $\lambda_1(A), \lambda_2(A), \ldots, \lambda_p(A)$ are sorted in the descending order satisfying $\lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_p(A)$. Then,

$$\lambda_{p-k_0+l}(A) \leq \lambda_l(PA) \leq \lambda_l(A), \quad l = 1, 2, \ldots, k_0.$$

Here, "=" in the right inequality holds for all $l = 1, 2, \ldots, k_0$ if and only if $P = \Phi^*_{(k_0)}\Phi_{(k_0)}$ where $\Phi_{(k_0)} = [\varphi_1, \varphi_2, \ldots, \varphi_{k_0}]^*$, and "=" in the left inequality holds for all $l = 1, 2, \ldots, k_0$ if and only if $P = \Phi^*_{[k_0]}\Phi_{[k_0]}$ where $\Phi_{[k_0]} = [\varphi_{p-k_0+1}, \ldots, \varphi_{p-1}, \varphi_p]^*$.

References

Aizenberg, L. (1993). *Carleman's formulas in complex analysis: theory and applications*. Dordrecht: Springer Science+Buisness Media.

Akcay, H., & Ninness, B. (1998). Rational basis functions for robust identification from frequency and time domain measurements. *Automatica*, *34*, 1101–1117.

Alpay, D., Colombo, F., Qian, T., & Sabadini, I. (2017). Adative orthonormal systems for matrix-valued functions. *Proceedings of the Americal Mathematical Society*, *145*(5), 2089–2106.

Alpay, D., & Gohberg, I. (1988a). On orthogonal matrix polynomials. In *Orthogonal matrix valued polynomials and applications, Vol. 34* (pp. 25–46). Basel: Birkhäuser.

Alpay, D., & Gohberg, I. (1988b). Unitary rational matrix functions. In *Topics in interpolation theory of rational matrix-valued functions, Vol. 33* (pp. 175–222). Basel: Birkhäuser.

Arov, D. Z., & Dym, H. (2008). *J-contractive matrix valued functions and related topics*. Cambridge University Press.

Chen, Q. H., Mai, W. X., Zhang, L. M., & Mi, W. (2015). System identification by discrete rational atoms. *Automatica*, *56*, 53–59.

Dahlquist, G., & Björck, r. (2008). *Numerical methods in scientific computing, Vol. 1*. Philadelphia, PA: Society for Industrial Mathematics.

Ding, F., & Chen, T. W. (2005). Hierarchical gradient-based identification of multivariable discrete-time systems. *Automatica*, *41*(2), 315–325.

Durbin, J., & Watson, G. S. (1950). Testing for serial correlation in least squares regression. *Biometrika*, *37*, 409–428, Corrections, 38:177-178, 1951.

Fei, M. G., Mi, W., & Pan, Y. B. (2016). On model reduction with consecutively selected rational basis. *International Journal of Wavelets, Multiresolution and Information Processing*, *14*(1), 1650002, 13 pp.

Garnett, J. B. (1987). *Bounded analytic functions*. Cambridge: Academic Press.

Garnier, H., Gilson, M., Bastogne, T., & Mensler, M. (2008). The CONTSID Toolbox: A software support for data-based continuous-time modelling. In *Identification of continuous-time models from sampled data*. Springer.

Gevers, M. (2006). A personal view of the development of system identification. *IEEE Control System Magazine*, *26*(6), 93–105.

Gu, G., & Khargonekar, P. P. (1992). A class of algorithms for identification in H_∞. *Automatica*, *28*, 299–312.

Heuberger, P. S. C., Hof, P. M. J. V. D., & Wahlberg, B. (2005). *Modelling and identification with rational orthogonal basis functions*. London: Springer-Verlag.

Horn, R. A., & Johnson, C. R. (1990). *Matrix analysis*. Cambridge University Press.

Katayama, T. (2005). *Subspace methods for system identification*. Springer.

Lance, C. (1994). *London mathematical society lecture note series book: vol. 210, Hilbert C*-modules: A toolkit for operator algebraists*. London: Cambridge University Press.

Li, H. C., Deng, G. T., & Qian, T. (2016). Hardy space decomposition of L_p on the unit circle: $0 < p < 1$. *Complex Variables and Elliptic Equations: An International Journal*, *61*(4), 510–523.

Li, H. C., Deng, G. T., & Qian, T. (2018). Fourier spectrum characterizations of H^p spaces on tubes over cones for $1 \leq p \leq \infty$. *Complex Analysis and Operator Theory*, *12*(5), 1193–1218.

Liu, Y. J., Xiao, Y. S., & Zhao, X. L. (2009). Multi-innovation stochastic gradient algorithm for multiple-input single-output systems using the auxiliary model. *Applied Mathematics and Computation*, *215*(4), 1477–1483.

Ljung, L. (1999). *System identification: Theory for the user (2nd edition)*. New Jersey: Prentice-Hall.

Makila, P. (1991). Laguerre methods and H^∞ identification of continuous-time systems. *International Journal of Control*, *53*, 689–707.

Mckelvey, T. (1995). *Identification of state space models from time and frequency data* (Ph.D. thesis), Linköping University.

Mi, W., & Qian, T. (2011). Frequency domain identification: An algorithm based on adaptive rational orthogonal system. *Automatica*, *48*, 1154–1162.

Mi, W., & Qian, T. (2014). On backward shift algorithm for estimating poles of systems. *Automatica*, *50*(6), 1603–1610.

Mi, W., Qian, T., & Li, S. (2016). Basis pursuit for frequency-domain identification. *Mathematical Methods in the Applied Science*, *39*(3), 498–507.

Mi, W., Qian, T., & Wan, F. (2012). A fast adaptive model reduction method based on takenaka-malmquist systems. *Systems & Control Letters*, *61*(1), 223–230.

Mo, Y., Qian, T., & Mi, W. (2015). Sparse representation in Szegö Kernels through reproducing Kernel Hilbert space theory with applications. *International Journal of Wavelet, Multiresolution and Information Processing*, *13*(4), 1550030, 20pp.

Natke, H. G., & Cottin, N. (1988). *Application of system identification in engineering*. Vienna: Springer.

Ninness, B. (2009). Some system identification challenges and approaches. In *Proc. 15th IFAC symp. syst. ident*, Saint-Malo, France (pp. 1–20).

Ninness, B., & Gustafsson, F. (1997). A unifying construction of orthonormal bases for system identification. *IEEE Transactions on Automatic Control*, *42*, 515–521.

Overschee, P. V., & Moor, B. D. (1996). *Subspace identification for linear systems*. Kluwer Academic Pub..

Picinbono, B. (1993). *Random signals and systems*. Englewood Cliffs, NJ: Prentice-Hall.

Pintelon, R., & Schoukens, J. (2012). *System identification: A frequency domain approach*. Hoboken, New Jersey: Wiley-IEEE Press.

Potapov, V. P. (1955). The multiplicative structure of J-contractive matrix functions (Russian). *Trudy Moskovskogo Matematicheskogo Obshchestva*, *4*, 125–236.

Qian, T. (2010). Intrinsic mono-component decomposition of functions: An advance of Fourier theory. *Mathematical Methods in Applied Sciences*, *33*(7), 880–891.

Qian, T., & Wang, Y. B. (2011). Adaptive Fourier series - a variation of greedy algorithm. *Advances in Computational Mathematics*, *34*, 279–293.

Qian, T., & Wang, Y. B. (2013). Remarks on adaptive Fourier decomposition. *International Journal of Wavelets, Multiresolution and Information Processing*, *11*(1), 1–14.

Qian, T., Zhang, L. M., & Li, Z. X. (2011). Algorithm of adaptive Fourier decomposition. *IEEE Transactions on Signal Processing*, *59*(2), 5899–5906.

Qu, W., & Dang, P. (2019a). Rational approximation in a class of weighted Hardy spaces. *Complex Analysis and Operator Theory*, *13*, 1827–1852.

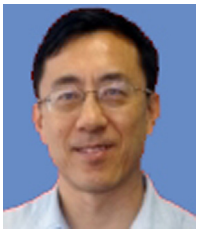
Qu, W., & Dang, P. (2019b). Reproducing kernel approximation in weighted Bergman space algorithm and application. *Mathematical Methods in the Applied Science*, *42*(12), 4292–4304.

Stoica, P., & Jansson, M. (2000). MIMO system identification : state space and subspace approximations versus transfer function and instrumental variables. *IEEE Transactions on Singnal Processing*, *48*, 3087–3099.

Sun, X.-Y., & Dang, P. (2019). Numerical stability of circular Hilbert transform and its application to signal decomposition. *Applied Mathematics and Computation*, *359*, 357–373.

Wahlberg, B. (1994). System identification using Kautz models. *IEEE Transactions on Automatic Control*, *AC-39*, 1276–1282.

Wang, X. Y., Qian, T., Leong, I. T., & Gao, Y. (2020). Two dimensional frequency-domain system identification. *IEEE Transactions on Automatic Control*, *65*(2), 577–590.



Tao Qian received the MA.Sc. and Ph.D. degrees, both in harmonic analysis, from Peking University, Beijing, China, in 1981 and 1984, respectively. From 1984 to 1986, he worked in Institute of Systems Science, the Chinese Academy of Sciences. Then he worked as Research Associate and Research Fellow in Australia till 1992 (Macquarie University, Flinders University of South Australia). He worked as Lecturer (Level B and Level C, English System) at New England University, Australia, from 1992 to 2000. He started working at University of Macau, Macau, China, from 2000 as Associate Professor. He got Full Professorship in 2003, and has been Head of Department of Mathematics from 2005 to 2011. He has been Distinguished Professor from 2013 to 2018. He started working at Macau University of Science and Technology as Professor from 2019. His research interests include harmonic analysis in Euclidean spaces, complex and Clifford analysis and signal analysis. Up to now he has published more than 180 papers.

Xiaoyin Wang received the B.S. in Mathematics from China University of Mining and Technology, Jiangsu, China in 2011 and the MA.Sc. in Mathematics from Wuhan University, Hubei, China in 2013. She received her Ph.D. degree in Mathematics from University of Macau, Macau in 2020. Her main interests are Fourier analysis and complex analysis.

Liming Zhang received the B.S. degree in Computer Software at Nankai University, China and M.S. degree in Signal Processing at Nanjing University of Science and Technology, China. She received her Ph.D. degree in image processing at University of New England, Australia. She is currently an assistant professor in Faculty of Science and Technology, University of Macau. Her research interests include signal processing, Image processing, Computer vision and IT in Education.