This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2019.2913047, IEEE Transactions on Automatic Control

2-D Frequency-Domain System Identification

Xiaoyin Wang, Tao Qian, Iengtak Leong, and You Gao

***Abstract*—In this article, we propose two iterative algorithms to identify transfer functions of 2-D systems. The proposed algorithms are modifications of the two-dimensional Adaptive Fourier Decomposition (abbreviated as 2-D AFD) and Weak Pre-Orthogonal Adaptive Fourier Decomposition (abbreviated as W-POAFD). 2-D AFD and W-POAFD are newly established adaptive representation theories for multivariate functions utilizing, respectively, the product-TM system and the product-Szegö dictionary. The proposed algorithms give rise to rational approximations with real coefficients to transfer functions. Owing to the modified maximal selection principles, the algorithms achieve a fast convergence rate $O(n^{-\frac{1}{2}})$. To use 2-D AFD and W-POAFD for system identification not only the theory is revised, but also the practical algorithm codes are provided. Experimental examples show that the proposed algorithms give promising results. The theory and algorithms studied in this paper are valid for any n-D case, $n \geq 2$.**

***Index Terms*—2-D system identification, multi-dimensional rational approximation, product-TM system, product-Szegö dictionary, generalized partial backward shift operator, matching pursuit, modified maximal selection principle.**

I. Introduction

SYstem identification has had a wide range of applications in engineering (see, for instance, [1]). In recent decades, more and more complicated systems appear in various fields such as signal processing [2], [3], digital filter analysis and designing [4]–[6], model-based control [7]–[9], 2-D text synthesis and classification [10], discretization of partial differential equations (PDEs) [11]–[13] and so on. Variables in those fields tend to vary with several variables including time and space. This has led to an increased interest for identification of multi-dimensional (n-D) systems, and it has successfully been applied in, e.g., system identification for PDEs [14]–[16].

As far as we are aware of, the most popular 2-D system identification methods include those based on 2-D Hankel theory [17], subspace identification [18]–[20] and neural network [21]. They were put forward as extensions of their corresponding 1-D methods. In the 1-D case, there have been studies utilizing rational orthogonal functions to find an approximation to the transfer function [22]–[28]. Among them, we are interested in the method in [28] deriving from AFD (adaptive Fourier decomposition) given in [29]. Being different from other methods in [22]–[27] which adopt rational orthogonal bases, AFD adaptively select rational orthogonal functions according to the given function, under a maximal selection principle that does not necessarily result in any basis. We briefly introduce AFD.

For a given sequence $\mathbf{a} = \{a_1, a_2, \cdots, a_n, \cdots\}$ where each a_i belongs to the unit disk $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$, the TM system (also called the rational orthogonal system) is defined as the family of functions $\{B_n^{\mathbf{a}}(z)\}$, where

$$B_n^{\mathbf{a}}(z) = \frac{\sqrt{1-|a_n|^2}}{1-\overline{a_n}z}\prod_{l=1}^{n-1}\frac{z-a_l}{1-\overline{a_l}z}. \tag{1}$$

For $f(z) \in H_2(\mathbb{D})$ where $H_2(\mathbb{D})$ is the Hardy-2 space on $\mathbb{D}$, given previously fixed $\{a_1, a_2, \cdots, a_{n-1}\}$, AFD selects a_n at the n-th iteration such that

$$a_n = \arg\max_{\zeta\in\mathbb{D}} \left|\left\langle f, B_n^{\{a_1,\cdots,a_{n-1},\zeta\}}\right\rangle_{H_2(\mathbb{D})}\right|^2,$$

being called the maximal selection principle. $\langle\cdot,\cdot\rangle_{H_2(\mathbb{D})}$ is the inner product. Thus, by repeating the above process for each iteration, the n-partial sum $f_n(z)$ is defined by

$$f_n(z) = \sum_{l=1}^{n}\left\langle f, B_l^{\mathbf{a}}\right\rangle_{H_2(\mathbb{D})} B_l^{\mathbf{a}}(z),$$

and adopted to approximate $f(z)$. Moreover, $f_n(z)$ satisfies $\lim_{n\to\infty} f_n(z) = f(z)$ in the sense of $H_2(\mathbb{D})$ norm [29].

We expect to develop the methodology of 1-D system identification method utilizing AFD in [28] to 2-D system identification. However, in the n-D cases, it is difficult to execute an analogous rational approximation process. None of the several complex variables theory, nor Clifford algebras have as far reaching functional analysis results as one complex variable does. In fact, n-variables function theory is different from that for one complex variable. In one complex variable analytic function theory, "$f(z) = (z - z_0)g(z)$" is equivalent to "z_0 is a zero of $f(z)$". However, in several complex variables we can not always deduce "$f(z,w) = (z - z_0)(w - w_0)g(z,w)$" from the property "$(z_0, w_0)$ is a zero of $f(z,w)$". That means the pole selection in the 1-D rational approximation is difficult to carry out in the n-D cases.

In the recent mathematical paper [30], the author defines "product-TM system" and "product-Szegö dictionary" in the Hardy-2 space on poly unit disk to access the n-D rational approximation as an extension of AFD. The present paper serves as a bridge that leads the theoretical development in [30] to applications in system identification by providing practical algorithms and the necessary modifications to the theoretical formulation. The contribution of this paper is as follows. We propose two modified algorithms inspired by [30] and design a two-stage procedure for system identification using the Cauchy integral and the proposed algorithms. Moreover,

This work was supported by Macau Science and Technology Fund FD-CT079/2016/A2 and Multi-Year Research Grant of University of Macau MYRG2016-00053-FST.

X. Y. Wang and I. T. Leong are with the Department of Mathematics, Faculty of Science and Technology, University of Macau, Macau, 999078, China (e-mail: yb57448@umac.mo; itleong@umac.mo).

T. Qian is with the Macau University of Science and Technology, Macau, 999078, China (e-mail: tqian1958@gmail.com).

Y. Gao is with the School of Mathematics, SUN YAT-SEN University, ZhuHai, 519000, China (e-mail: gaoyou@mail.sysu.edu.cn).

Corresponding author is T. Qian.

we give estimates of the convergence rate of the two proposed algorithms. Utilizing the convergence rate estimations, we obtain an upper bound for the mathematical expectation of the approximation error of the two-stage procedure in the presence of stochastic noise.

The rest of the paper is organized as follows. Section II reviews some necessary preliminary and states the problem setting. Section III introduces the modified 2-D AFD and the modified W-POAFD. Section IV presents a two-stage process utilizing algorithms in Section III to identify a 2-D transfer function. With the process in the second stage, we complete the proofs of two main theorems and obtain rational approximations with real coefficients. Section V analyzes the approximation error of the identification process in Section IV under the consideration of noise. Section VI exhibits the numerical experiments in both of the two cases, viz., being with noise and without. We obtain good results for both cases. Some conclusions are drawn in Section VII and some necessary technical details are collected in the Appendices.

II. Preliminary and problem setting

In this paper, we are interested in 2-D discrete, linear time-invariant (LTI), single input single output (SISO), quadrant causal systems, whose transfer functions are proper rational functions having the following form

$$G(z_1,z_2)=\frac{Y(z_1,z_2)}{U(z_1,z_2)}=\frac{\sum\limits_{p,q=0}^{n}a_{p,q}z_1^pz_2^q}{\sum\limits_{p,q=0}^{n}b_{p,q}z_1^pz_2^q}=\sum_{p,q=0}^{\infty}g_{p,q}z_1^{-p}z_2^{-q},$$

where $b_{n,n}=1$, $Y(z_1,z_2),U(z_1,z_2)$ are the 2-D Z transformations of output $y(i,j)$ and input $u(i,j)$ in the spatial-domain, $\{g_{p,q}\}$ is the impulse response. Popular state space models, such as Attasi model, Roesser's model, the first Fornasini-Marchesini model and the second Fornasini-Marchesini model, are quadrant causal systems [31]. In addition, we assume $G(z_1,z_2)$ has no poles in the region $|z_1|\geq 1,|z_2|\geq 1$.

Through the mappings $z_1\to\frac{1}{z},z_2\to\frac{1}{w}$, we transform $G(z_1,z_2)$ into $f(z,w)$ which is holomorphic in $|z|\leq 1,|w|\leq 1$ and $f(\mathrm{e}^{\mathrm{j}t},\mathrm{e}^{\mathrm{j}u})=G(\mathrm{e}^{-\mathrm{j}t},\mathrm{e}^{-\mathrm{j}u}),t,u\in[0,2\pi)$. In the sequel, we discuss $f(z,w)$. It is noted $f(z,w)\in\mathcal{H}=H_2(\mathbb{D}^2)$ (refer to Appendix A for the definition) and $\|f\|=\|G\|_{L_2}$, where $\|G\|_{L_2}=\frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi}|G(\mathrm{e}^{\mathrm{j}t},\mathrm{e}^{\mathrm{j}u})|^2\,\mathrm{d}t\mathrm{d}u$. Moreover, $f(z,w)$ is continuous on the poly unit circle $\partial\mathbb{D}^2$ ($\partial\mathbb{D}\times\partial\mathbb{D}$).

In general, we adopt the additive noise assumption for the input and output measurements in the spatial-domain:

$$u_{measured}(i,j)=u_{real}(i,j)+\nu(i,j),$$
$$y_{measured}(i,j)=y_{real}(i,j)+\mu(i,j),$$

where ν and μ are stochastic noise sequences. Hence, we achieve the noisy transfer function model

$$G_{measured}=\frac{Y_{measured}}{U_{measured}}=G_{real}\cdot\frac{1+\frac{Z_\nu}{Y_{real}}}{1+\frac{Z_\mu}{U_{real}}},\tag{2}$$

where variable Z represents the 2-D Z transformation of the corresponding signal according to the subscript. It is noted that $Y_{measured}$ and $U_{measured}$ are random variables and each of them is a sum of a large number of random variables. In [32], [33], it is shown that the additive noise on the Fourier coefficients under discrete Fourier transformation (DFT) are asymptotically circular complex normally distributed random variables with respect to the number of samples in DFT provided that the time-domain series are stationary and weakly correlated. Fortunately, this result still holds in the 2-D case as long as we add some restrictions to the noise sequences similar with the stationary and weakly correlated condition in the 1-D case which can be easily met.

Considering the above assumptions, the formulation of **2-D frequency-domain system identification** is as follows:

Given:

A finite number MN of noisy frequency response measurements $\{E_{p,q}^{MN}\}_{p=1\dots M,q=1\dots N}$ from a 2-D system

$$E_{p,q}^{MN}=f(\mathrm{e}^{\mathrm{j}\omega_p},\mathrm{e}^{\mathrm{j}\omega_q'})(\frac{1+\Upsilon_{p,q}^Y}{1+\Upsilon_{p,q}^U}),\tag{3}$$

where $f(\mathrm{e}^{\mathrm{j}\omega_p},\mathrm{e}^{\mathrm{j}\omega_q'})=G(\mathrm{e}^{-\mathrm{j}\omega_p},\mathrm{e}^{-\mathrm{j}\omega_q'})$, $\omega_p=\frac{2\pi(p-1)}{M},\omega_q'=\frac{2\pi(q-1)}{N}$, $\Upsilon_{p,q}$ is the stochastic frequency-domain noise divided by the spectra of corresponding spatial-domain data as described in (2). Without confusion we still name them frequency-domain noise. Here we let M and N be even.

Assumption:

(i) Υ^U, as well as Υ^Y, are uncorrelated (respect to the two frequency lines t and u) circular complex normally distributed stochastic variables with zero mean.

(ii) Υ^U and Υ^Y are independent.

Find:

Approximations $f_n(z,w)$ (n is a positive integer) to $f(z,w)$, thereby approximations $G_n(\mathrm{e}^{\mathrm{j}t},\mathrm{e}^{\mathrm{j}u})=f_n(\mathrm{e}^{-\mathrm{j}t},\mathrm{e}^{-\mathrm{j}u})$ to $G(\mathrm{e}^{\mathrm{j}t},\mathrm{e}^{\mathrm{j}u})$, such that $\|G-G_n\|_{L_2}=\|f-f_n\|$ satisfies

$$\lim_{\substack{n\to\infty,M,N\to\infty\\ \Upsilon^Y,\Upsilon^U\to 0}}\mathbb{E}\|G-G_n\|_{L_2}=0\tag{4}$$

under certain conditions, which will be stated in Section V. Here, $\mathbb{E}$ represents the mathematical expectation.

III. Modified 2-D AFD and modified W-POAFD

In this section, we propose two iterative algorithms referring to the methodology and framework of 2-D AFD and W-POAFD in [30] but make some modifications. The key modification, called ***conjugate parameters selection***, is that at each iteration we select a parameters pair $(a,b),(\overline{a},\overline{b})$ instead of only (a,b). Some mathematical foundations related to the algorithms are introduced. Algorithm 1 and Algorithm 2 illustrate how we implement the proposed algorithms.

A. Modified 2-D AFD based on product-TM system

Definition 1 (product-TM system): Denote the finite TM systems of variables z and w as $\mathcal{B}_N^{\mathbf{a}}(z)=\{B_k^{\mathbf{a}}(z)\}_{k=1}^N$ and $\mathcal{B}_M^{\mathbf{b}}(w)=\{B_l^{\mathbf{b}}(w)\}_{l=1}^M$. The product-TM system is defined by

$$\mathcal{B}_N^{\mathbf{a}}\otimes\mathcal{B}_M^{\mathbf{b}}(z,w)=\{B_k^{\mathbf{a}}(z)B_l^{\mathbf{b}}(w)\}_{k=1,\cdots,N,\ l=1,\cdots,M}.$$

Clearly, $\mathcal{B}_N^{\mathbf{a}}\otimes\mathcal{B}_M^{\mathbf{b}}(z,w)$ is an orthonormal system in $\mathcal{H}$.

Here the different variables z and w are the 2 coordinates of $\mathbb{D}^2$. For brevity, we omit the variables z, w and distinguish them by parameters $\mathbf{a}, \mathbf{b}$ in the sequel.

We give a description for the n-th iteration of the **modified 2-D AFD** as follows. The complete algorithm is by repeating the same process step by step.

Given $f(z,w) \in \mathcal{H}$, for $\mathbf{a} = \{a_1, \overline{a_1}, \cdots, a_n, \overline{a_n}\}, \mathbf{b} = \{b_1, \overline{b_1}, \cdots, b_n, \overline{b_n}\}$, denote

$$\begin{aligned} &S_0(f) = 0, \quad S_n(f) = \sum_{1 \le k,l \le 2n} \left\langle f, B_k^{\mathbf{a}} B_l^{\mathbf{b}} \right\rangle B_k^{\mathbf{a}} B_l^{\mathbf{b}}, \\ &D_n(f) = S_n(f) - S_{n-1}(f) = \sum_{\max\{k,l\}=2n-1}^{2n} \left\langle f, B_k^{\mathbf{a}} B_l^{\mathbf{b}} \right\rangle B_k^{\mathbf{a}} B_l^{\mathbf{b}}. \end{aligned} \tag{5}$$

$S_n(f)$ is called the n-partial sum and $D_n(f)$ is called the n-partial sum difference having $8n-4$ entries. Since $\mathcal{B}_{2n}^{\mathbf{a}} \otimes \mathcal{B}_{2n}^{\mathbf{b}}$ is an orthonormal system as mentioned in Definition 1,

$$\|S_n(f)\|^2 = \|S_{n-1}(f)\|^2 + \|D_n(f)\|^2.$$

For previously fixed $(a_1, \overline{a_1}, \cdots, \overline{a_{n-1}}) \subset \mathbb{D}$ and $(b_1, \overline{b_1}, \cdots, \overline{b_{n-1}}) \subset \mathbb{D}$, to pursue the maximal energy gain of $S_n(f)$, we apply the ***modified Maximal Selection Principle***: select $a_n, b_n \in \mathbb{D}$ such that $\|D_n(f)\|^2$ attains its maximum value among all possible selections, namely

$$\|D_n(f)\|^2 = \max_{a,b \in \mathbb{D}} \|D_n^{\tilde{\mathbf{a}},\tilde{\mathbf{b}}}(f)\|^2 \tag{6}$$

where $\tilde{\mathbf{a}} = (a_1, \overline{a_1}, \cdots, \overline{a_{n-1}}, a, \overline{a})$, $\tilde{\mathbf{b}} = (b_1, \overline{b_1}, \cdots, \overline{b_{n-1}}, b, \overline{b})$, $D_n^{\tilde{\mathbf{a}},\tilde{\mathbf{b}}}(f) = \sum\limits_{\max\{k,l\}=2n-1}^{2n} \langle f, B_k^{\tilde{\mathbf{a}}} B_l^{\tilde{\mathbf{b}}} \rangle B_k^{\tilde{\mathbf{a}}} B_l^{\tilde{\mathbf{b}}}$. $S_n(f)$ is obtained as the n-th approximation to f.

It is worth mentioning that the maximum in (6) is attainable owing to the "boundary vanish condition" for $f(z,w) \in \mathcal{H}$:

$$\lim_{|a_n| \to 1, |b_n| \to 1} \left| \left\langle f, B_k^{\mathbf{a}} B_l^{\mathbf{b}} \right\rangle \right| = 0,$$

where $\max\{k,l\} = 2n-1, 2n$.

To sum up, the modified 2-D AFD gives rise to approximations, consisting of orthogonal rational functions, to the given function using the adaptively determined product-TM system.

Algorithm 1 Modified 2-D AFD.

Input: Signal $f(\mathrm{e}^{\mathrm{j}t}, \mathrm{e}^{\mathrm{j}u}), t, u \in [0, 2\pi)$; Positive integer K.
Output: Parameters $\mathbf{a} = \{a_1, \overline{a_1}, \cdots, \overline{a_K}\}$, $\mathbf{b} = \{b_1, \overline{b_1}, \cdots, \overline{b_K}\}$; Product-TM system $\{B_k^{\mathbf{a}} B_l^{\mathbf{b}}\}_{k,l=1}^{2K}$; Coefficients $\{c_{k,l}\}_{k,l=1}^{2K}$; K-partial sum $S_K(f)$.

Initialize $S_0(f) = 0$.
for $n = 1$ to K **do**
  Get the optimal solution (a_n, b_n) by solving the maximum problem in (6).
  Compute $B_k^{\mathbf{a}} B_l^{\mathbf{b}}, c_{k,l} = \langle f, B_k^{\mathbf{a}} B_l^{\mathbf{b}} \rangle$ where $\max\{k,l\} = 2n-1$ or $2n$ using the selected parameters a_n, b_n.
  Set $S_n(f) = S_{n-1}(f) + \sum\limits_{\max\{k,l\}=2n-1}^{2n} c_{k,l} B_k^{\mathbf{a}} B_l^{\mathbf{b}}$.
end for
return $\mathbf{a} = \{a_1, \overline{a_1}, \cdots, \overline{a_K}\}$, $\mathbf{b} = \{b_1, \overline{b_1}, \cdots, \overline{b_K}\}$, $\{B_k^{\mathbf{a}} B_l^{\mathbf{b}}\}_{k,l=1}^{2K}$, $\{c_{k,l}\}_{k,l=1}^{2K}$, $S_K(f)$.

B. Modified W-POAFD based on product-Szegö dictionary

Definition 2 (product-Szegö dictionary): Denote the Szegö kernel as $e_a(z) = \frac{\sqrt{1-|a|^2}}{1-\overline{a}z}, a \in \mathbb{D}$. The set of the product-Szegö kernels

$$\mathcal{D}^2 = \{e_a(z) e_b(w) : a, b \in \mathbb{D}\}, \tag{7}$$

is a dictionary of $\mathcal{H}$. We call it the product-Szegö dictionary. For $f \in \mathcal{H}$, the reproducing property holds:

$$\langle f, e_a e_b \rangle = \sqrt{1-|a|^2}\sqrt{1-|b|^2} f(a,b).$$

Similar to the modified 2-D AFD, we give a description for the n-th iteration of the **modified W-POAFD**.

Given $f(z,w) \in \mathcal{H}$, for $\mathbf{a} = \{a_1, \overline{a_1}, \cdots, a_n, \overline{a_n}\}, \mathbf{b} = \{b_1, \overline{b_1}, \cdots, b_n, \overline{b_n}\}$, denote

$$\begin{aligned} &\tilde{S}_0(f) = 0, \\ &E_n = \{v_k\}_{k=1}^{2n} = \{e_{a_1} e_{b_1}, e_{\overline{a_1}} e_{\overline{b_1}}, \ldots, e_{\overline{a_n}} e_{\overline{b_n}}\}, \\ &\tilde{S}_n(f) = \sum_{k=1}^{n} \left[\langle f, B_k \rangle B_k + \langle f, \tilde{B}_k \rangle \tilde{B}_k \right], \\ &\tilde{D}_n(f) = \tilde{S}_n(f) - \tilde{S}_{n-1}(f), \quad g_{n+1} = f - \tilde{S}_n(f), \end{aligned} \tag{8}$$

where $\{B_1, \tilde{B}_1, \cdots, B_n, \tilde{B}_n\}$ is G-S orthonormalization of E_n as follows:

$$\begin{aligned} &u_k = v_k - \sum_{p=1}^{k-1} \frac{\langle v_k, u_p \rangle}{\|u_p\|^2} u_p, k = 1, 2, \cdots, 2n; \\ &B_k = \frac{u_{2k-1}}{\|u_{2k-1}\|}, \tilde{B}_k = \frac{u_{2k}}{\|u_{2k}\|}, k = 1, 2, \cdots, n, \end{aligned} \tag{9}$$

Note that $\tilde{S}_n(f)$ in (8) is the projection of f onto the subspace $\overline{\mathrm{span} E_n}$ of $\mathcal{H}$. $\tilde{S}_n(f)$ is called the n-partial sum and $\tilde{D}_n(f)$ is called the n-partial sum difference having 2 terms. g_{n+1} is called the orthogonal standard remainder. Clearly,

$$\|\tilde{S}_n(f)\|^2 = \|\tilde{S}_{n-1}(f)\|^2 + \|\tilde{D}_n(f)\|^2.$$

For previously fixed $(a_1, \overline{a_1}, \cdots, \overline{a_{n-1}}) \subset \mathbb{D}$ and $(b_1, \overline{b_1}, \cdots, \overline{b_{n-1}}) \subset \mathbb{D}$, to make the energy of $\tilde{S}_n(f)$ gain with a fast rate, we apply the ***modified Pre-Orthogonal ρ-Maximal Selection Principle***: select $a_n, b_n \in \mathbb{D}$ such that $\|\tilde{D}_n(f)\|^2$ is as large as possible according to the following criterion:

$$\|\tilde{D}_n(f)\|^2 \ge \rho \sup_{\alpha \in \mathcal{D}^2} \left\{ \left| \langle g_n, B_n^{\alpha} \rangle \right|^2 + \left| \langle g_n, \tilde{B}_n^{\alpha} \rangle \right|^2 \right\}, \tag{10}$$

where $0 < \rho < 1$, $\alpha = e_a e_b$, $\{B_1, \tilde{B}_1, \cdots, B_n^{\alpha}, \tilde{B}_n^{\alpha}\}$ is the G-S orthonormalization of $\{e_{a_1} e_{b_1}, e_{\overline{a_1}} e_{\overline{b_1}}, \cdots, e_{\overline{a_{n-1}}} e_{\overline{b_{n-1}}}, e_a e_b, e_{\overline{a}} e_{\overline{b}}\}$. $\tilde{S}_n(f)$ is obtained as the n-th approximation to f.

To sum up, the modified W-POAFD gives rise to approximations, consisting of orthogonal rational functions, to the given function using the projection onto an adaptively determined subspace of $\mathcal{H}$. The subspace is spanned by some items selected from the dictionary $\mathcal{D}^2$ of $\mathcal{H}$.

Remark 3: The nonlinear optimization problems in (6) and (10) are difficult to solve. One practical method to solve them is through discretization of the domain $\mathbb{D}$ of the parameters a_n, b_n with rectangular grids. Namely, search the optimal parameters pair (a_n, b_n) in a discretization set $P = \{(a^{(s)}, b^{(s)})\}_{s=1,2,\cdots,S}$ where $a^{(s)}, b^{(s)} \in \mathbb{D}$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2019.2913047, IEEE Transactions on Automatic Control

Algorithm 2 Modified W-POAFD.

Input: Signal $f(\mathrm{e}^{\mathrm{j}t},\mathrm{e}^{\mathrm{j}u}), t,u\in[0,2\pi)$; Positive integer K.
Output: Parameters $\mathbf{a}=\{a_1,\overline{a_1},\cdots,\overline{a_K}\}$, $\mathbf{b}=\{b_1,\overline{b_1},\cdots,\overline{b_K}\}$; G-S orthonormalization $\{B_1,\tilde{B}_1,\cdots,B_K,\tilde{B}_K\}$; Coefficients $\{c_l\}_{l=1}^{2K}$; K-partial sum $\tilde{S}_K(f)$.

Initialize $\tilde{S}_0(f)=0$.
for $n=1$ to K **do**
  Set $g_n=f-\tilde{S}_{n-1}(f)$.
  Get one parameters pair (a_n,b_n) satisfying (10).
  Compute the G-S orthonormalization terms $B_n,\tilde{B}_n$ of $e_{a_n}e_{b_n}, e_{\overline{a_n}}e_{\overline{b_n}}$ using the selected a_n,b_n.
  Compute $c_{2n-1}=\langle f,B_n\rangle$, $c_{2n}=\langle f,\tilde{B}_n\rangle$.
  Set $\tilde{S}_n(f)=\tilde{S}_{n-1}(f)+c_{2n-1}B_n+c_{2n}\tilde{B}_n$.
end for
return $\mathbf{a}=\{a_1,\overline{a_1},\cdots,\overline{a_K}\}$, $\mathbf{b}=\{b_1,\overline{b_1},\cdots,\overline{b_K}\}$, $\{B_1,\tilde{B}_1,\cdots,B_K,\tilde{B}_K\}$, $\{c_l\}_{l=1}^{2K}$, $\tilde{S}_K(f)$.

C. Convergence of two modified algorithms

Despite of the modifications, i.e., the conjugate parameters selection, the proposed algorithms are convergent due to the following Theorem 4 and 5. The proofs of the two theorems are similar to the ones of the convergence theorems in [30] but with some minor modifications. We omit proofs in this paper.

Theorem 4 (convergence of modified 2-D AFD): Let $f\in\mathcal{H}$. For previously fixed $\{a_1,\overline{a_1},\cdots,\overline{a_{n-1}}\}$, $\{b_1,\overline{b_1},\cdots,\overline{b_{n-1}}\}$, by selecting $(a_n,b_n),\cdots$ according to the modified Maximal Selection Principle, we have $\lim_{n\to\infty}\|f-S_n(f)\|=0$.

Theorem 5 (convergence of modified W-POAFD): For $f\in\mathcal{H}$, with a sequence of consecutively selected $\{e_{a_1}e_{b_1}, e_{\overline{a_1}}e_{\overline{b_1}},\cdots,e_{a_n}e_{b_n},e_{\overline{a_n}}e_{\overline{b_n}},\cdots\}$ from $\mathcal{D}^2$ under the modified Pre-Orthogonal ρ-Maximal Selection Principle, we have $\lim_{n\to\infty}\|f-\tilde{S}_n(f)\|=0$.

IV. Adaptive approximation for 2-D system identification

In this section, given measurements $\{E_{p,q}^{MN}\}$ of $f(z,w)$ on $\partial\mathbb{D}^2$, we design a two-stage procedure to identify $f(z,w)$ like certain nonlinear algorithms [28], [34] for 1-D system identification. The first stage involves constructing a holomorphic approximation $\tilde{f}$ on $\mathbb{D}^2$ to f making use of the measurements. The second stage involves constructing an approximation $\hat{f}$ to $\tilde{f}$, consisting of rational orthogonal entries, through utilizing the proposed algorithms in Section III.

A. First stage

Based on the Cauchy integral in complex analysis, for the counterclockwise closed curve $\partial\mathbb{D}$ whose parameterized equation is $\zeta=\mathrm{e}^{\mathrm{j}t}$ $(0\le t<2\pi)$, and $g(\zeta_1,\zeta_2)$ defined on $\partial\mathbb{D}^2$, we define the tensor product type Cauchy integral of $g(\zeta_1,\zeta_2)$ along the tensor of two curves $\partial\mathbb{D}^2$ as

$$\begin{aligned}\mathcal{C}(g)(z,w)&=\left(\frac{1}{2\pi\mathrm{j}}\right)^2\int_{\partial\mathbb{D}}\int_{\partial\mathbb{D}}\frac{g(\zeta_1,\zeta_2)}{(\zeta_1-z)(\zeta_2-w)}\,\mathrm{d}\zeta_1\mathrm{d}\zeta_2\\&=-\frac{1}{4\pi^2}\int_0^{2\pi}\!\!\int_0^{2\pi}\frac{g(\mathrm{e}^{\mathrm{j}t},\mathrm{e}^{\mathrm{j}u})}{(\mathrm{e}^{\mathrm{j}t}-z)(\mathrm{e}^{\mathrm{j}u}-w)}\,\mathrm{d}\mathrm{e}^{\mathrm{j}t}\mathrm{d}\mathrm{e}^{\mathrm{j}u}\\&=-\frac{1}{4\pi^2}\int_0^{2\pi}\!\!\int_0^{2\pi}\frac{g(\mathrm{e}^{\mathrm{j}t},\mathrm{e}^{\mathrm{j}u})}{(\mathrm{e}^{\mathrm{j}t}-z)(\mathrm{e}^{\mathrm{j}u}-w)}\,\mathrm{j}\mathrm{e}^{\mathrm{j}t}\mathrm{j}\mathrm{e}^{\mathrm{j}u}\mathrm{d}t\mathrm{d}u.\end{aligned}$$

Utilizing the tensor product type Cauchy integral above, $\tilde{f}$ is formulated as

$$\tilde{f}(z,w)=-\frac{1}{4\pi^2}\int_0^{2\pi}\!\!\int_0^{2\pi}\frac{\sum_{p,q}E_{p,q}^{MN}\chi_{p,q}(t,u)}{(\mathrm{e}^{\mathrm{j}t}-z)(\mathrm{e}^{\mathrm{j}u}-w)}\,\mathrm{d}\mathrm{e}^{\mathrm{j}t}\mathrm{d}\mathrm{e}^{\mathrm{j}u},\quad(11)$$

where

$$\chi_{p,q}(t,u)=\begin{cases}\chi_{[\omega_p,\omega_{p+1})\times[\omega'_q,\omega'_{q+1})}(t,u), & \text{case 1}\\ \chi_{[\omega_p,\omega_{p+1})\times(\omega'_{q-1},\omega'_q]}(t,u), & \text{case 2}\\ \chi_{(\omega_{p-1},\omega_p]\times[\omega'_q,\omega'_{q+1})}(t,u), & \text{case 3}\\ \chi_{(\omega_{p-1},\omega_p]\times(\omega'_{q-1},\omega'_q]}(t,u), & \text{case 4}\end{cases}$$

case 1 $=\{p=1,\cdots,\frac{M}{2},q=1,\cdots,\frac{N}{2}\}$, case 2 $=\{p=1,\cdots,\frac{M}{2},q=\frac{N}{2}+2,\cdots,N+1\}$, case 3 $=\{p=\frac{M}{2}+2,\cdots,M+1,q=1,\cdots,\frac{N}{2}\}$, case 4 $=\{p=\frac{M}{2}+2,\cdots,M+1,q=\frac{N}{2}+2,\cdots,N+1\}$,

$$\chi_{[t_1,t_2)\times[u_1,u_2)}(t,u)=\begin{cases}1, & (t,u)\in[t_1,t_2)\times[u_1,u_2)\\ 0, & \text{otherwise,}\end{cases}$$

is the characteristic function and $E_{p,q}^{MN}$ is defined in (3).

Here, when M,N is large enough and the stochastic noise is small enough, the step function $\sum_{p,q}E_{p,q}^{MN}\chi_{p,q}(t,u)$ is almost equal to $f(\mathrm{e}^{\mathrm{j}t},\mathrm{e}^{\mathrm{j}u})$. Therefore, by the Hardy space theory (see, for instance, [35]), $\tilde{f}(z,w)$ is holomorphic on $\mathbb{D}^2$ and

$$\|\tilde{f}-f\|_{L_2}=\|\tilde{f}-f\|\le C\|\sum_{p,q}E_{p,q}^{MN}\chi_{p,q}-f\|_{L_2},\quad(12)$$

where C is a constant.

Moreover, if $\{E_{p,q}^{MN}\}$ is noise-free, then $\tilde{f}(z,w)$ enjoys the conjugate symmetry property as the frequency response data $G(\mathrm{e}^{\mathrm{j}t},\mathrm{e}^{\mathrm{j}u})$ does: $G(\mathrm{e}^{\mathrm{j}t},\mathrm{e}^{\mathrm{j}u})=\overline{G(\mathrm{e}^{-\mathrm{j}t},\mathrm{e}^{-\mathrm{j}u})}$. In fact,

$$\begin{aligned}\tilde{f}(z,w)&=\overline{\tilde{f}(\overline{z},\overline{w})}\\&=-\frac{1}{4\pi^2}\Bigg[\sum_{p=1}^{\frac{M}{2}}\sum_{q=1}^{\frac{N}{2}}\ln\Big(\frac{\mathrm{e}^{\mathrm{j}\omega_{p+1}}-z}{\mathrm{e}^{\mathrm{j}\omega_p}-z}\Big)\ln\Big(\frac{\mathrm{e}^{\mathrm{j}\omega'_{q+1}}-w}{\mathrm{e}^{\mathrm{j}\omega'_q}-w}\Big)E_{p,q}^{MN}\\&+\sum_{p=1}^{\frac{M}{2}}\sum_{q=1}^{\frac{N}{2}}\ln\Big(\frac{\mathrm{e}^{-\mathrm{j}\omega_{p+1}}-z}{\mathrm{e}^{-\mathrm{j}\omega_p}-z}\Big)\ln\Big(\frac{\mathrm{e}^{-\mathrm{j}\omega'_{q+1}}-w}{\mathrm{e}^{-\mathrm{j}\omega'_q}-w}\Big)\overline{E_{p,q}^{MN}}\\&+\sum_{p=1}^{\frac{M}{2}}\sum_{q=\frac{N}{2}+2}^{N+1}\ln\Big(\frac{\mathrm{e}^{\mathrm{j}\omega_{p+1}}-z}{\mathrm{e}^{\mathrm{j}\omega_p}-z}\Big)\ln\Big(\frac{\mathrm{e}^{\mathrm{j}\omega'_q}-w}{\mathrm{e}^{\mathrm{j}\omega'_{q-1}}-w}\Big)E_{p,q}^{MN}\\&+\sum_{p=1}^{\frac{M}{2}}\sum_{q=\frac{N}{2}+2}^{N+1}\ln\Big(\frac{\mathrm{e}^{-\mathrm{j}\omega_{p+1}}-z}{\mathrm{e}^{-\mathrm{j}\omega_p}-z}\Big)\ln\Big(\frac{\mathrm{e}^{-\mathrm{j}\omega'_q}-w}{\mathrm{e}^{-\mathrm{j}\omega'_{q-1}}-w}\Big)\overline{E_{p,q}^{MN}}\Bigg].\end{aligned}$$

B. Second stage

In the second stage we formulate the partial sums that at the same time addresses the conjugate parameters selection issue. The following two theorems, Theorem 6 and 10, state that through such selection the partial sums obtained have real coefficients.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2019.2913047, IEEE Transactions on Automatic Control

1) Modified 2-D AFD:

Theorem 6: Let $f(z,w)\in\mathcal{H}$ with real coefficients, namely $f(z,w)=\overline{f(\overline{z},\overline{w})}$. $S_n(f)$ is the n-partial sum in (5) obtained by the modified 2-D AFD. Then $S_n(f)$ has real coefficients.

Before we come to the proof, we give a definition and some related results.

Definition 7: Let $h(z,w)\in\mathcal{H}$. ***The generalized partial backward shift operator with respect to variable*** z ***via*** $a\in\mathbb{D}$ is defined by

$$S_a^{(1)}(h)(z,w)=\frac{h(z,w)-\langle h,e_a\cdot 1\rangle^{(1)}(w)e_a(z)\cdot 1}{\frac{z-a}{1-\overline{a}z}},$$

where $\langle f,g\rangle^{(1)}(w)=\frac{1}{2\pi}\int_0^{2\pi}f(\mathrm{e}^{\mathrm{j}t},w)\overline{g(\mathrm{e}^{\mathrm{j}t},w)}\mathrm{d}t$ is the inner product of $H_2(\mathbb{D})$ space with respect to the variable z for $f,g\in\mathcal{H}$. For the variable w and $b\in\mathbb{D}$, we define the similar operator $S_b^{(2)}(h)$.

The operator is proposed in [30] to prove the existence of the optimal solution to the maximum problem involved in the maximal selection principle for 2-D AFD. It plays the same role in the modified 2-D AFD. The following two results regarding the operator are needed in the proof of Theorem 6. They follow from Proposition 17, which is given in Appendix B.

Corollary 8: Let $f(z,w)\in\mathcal{H}$. For positive integers m,n and parameters $\mathbf{a}=\{a_1,a_2,\cdots,a_n\}$, $\mathbf{b}=\{b_1,b_2,\cdots,b_m\}$ from $\mathbb{D}$, the following equality holds,

$$\Big\langle\prod_{l=1}^{m-1}S_{b_l}^{(2)}\prod_{k=1}^{n-1}S_{a_k}^{(1)}(f),e_{a_n}e_{b_m}\Big\rangle=\big\langle f,B_n^{\mathbf{a}}B_m^{\mathbf{b}}\big\rangle.$$

Corollary 9: Let $f(z,w)\in\mathcal{H}$ be a rational function with real coefficients, namely $f(z,w)=\overline{f(\overline{z},\overline{w})}$, and $a\in\mathbb{D}$, then $S_{\overline{a}}^{(1)}S_a^{(1)}(f)$ is a rational function with real coefficients. Furthermore, if $a\in\mathbb{R}$ then $S_a^{(1)}(f)$ is a rational function with real coefficients. Operator $S_b^{(2)}(f)$ has the same properties.

Now we go back to prove Theorem 6 utilizing the generalized partial backward shift operator.

Proof of Theorem 6:

$S_n(f)$ is a sum of $4n^2$ terms. We claim that if we divide all the terms into n^2 groups by collecting 4 terms with adjacent summation indices to one group, then each group has real coefficients. Denote $\mathbf{a}=\{\mathbf{a}_k\}_{k=1}^{2n}=\{a_1,\overline{a_1},\cdots,a_n,\overline{a_n}\}$, $\mathbf{b}=\{\mathbf{b}_k\}_{k=1}^{2n}=\{b_1,\overline{b_1},\cdots,b_n,\overline{b_n}\}$. We define $PS_n(f)$ by $PS_n(f)=\sum_{k=2n-1}^{2n}\sum_{l=1}^{2}\langle f,B_k^{\mathbf{a}}B_l^{\mathbf{b}}\rangle B_k^{\mathbf{a}}B_l^{\mathbf{b}}$ and prove, as an example, that the group $PS_n(f)$ has real coefficients. The same result holds for the other groups.

Denote $F=\prod_{k=1}^{2(n-1)}S_{\mathbf{a}_k}^{(1)}(f)$. Since f has real coefficients, by Corollary 9 and the conjugate parameters selection of $\mathbf{a}$, F is a rational function with real coefficients. Together with Corollary 8 it holds that $PS_n(f)=I_1I_2$, where

$$\begin{aligned}
I_1&=(1-|a_n|^2)(1-|b_1|^2)\tilde{B}^{\mathbf{a}}_{2(n-1)},\\
\tilde{B}^{\mathbf{a}}_{2(n-1)}&=\prod_{k=1}^{2(n-1)}\frac{z-\mathbf{a}_k}{1-\overline{\mathbf{a}_k}z};\\
I_2&=\big[c_1(1-a_nz)(1-b_1w)+c_2(z-a_n)(1-b_1w)\\
&\quad+c_3(1-a_nz)(w-b_1)+c_4(z-a_n)(w-b_1)\big]/\\
&\quad\big[(1-\overline{a_n}z)(1-a_nz)(1-\overline{b_1}w)(1-b_1w)\big],\\
c_1&=F(a_n,b_1),\\
c_2&=\frac{(1-\overline{a_n}^2)F(\overline{a_n},b_1)-(1-|a_n|^2)F(a_n,b_1)}{\overline{a_n}-a_n},\\
c_3&=\frac{(1-\overline{b_1}^2)F(a_n,\overline{b_1})-(1-|b_1|^2)F(a_n,b_1)}{\overline{b_1}-b_1},\\
c_4&=\big[(1-\overline{a_n}^2)(1-\overline{b_1}^2)F(\overline{a_n},\overline{b_1})-(1-|a_n|^2)(1-\overline{b_1}^2)\\
&\quad F(a_n,\overline{b_1})-(1-\overline{a_n}^2)(1-|b_1|^2)F(\overline{a_n},b_1)+(1-\\
&\quad |a_n|^2)(1-|b_1|^2)F(a_n,b_1)\big]/\big[(\overline{a_n}-a_n)(\overline{b_1}-b_1)\big].
\end{aligned}$$

We observe that I_1 and the denominator of I_2 have real coefficients due to the conjugate parameters selection of $\mathbf{a},\mathbf{b}$. In addition, the numerator of I_2, denoted by F_1, is a polynomial having real coefficients since

$$F_1(z,w)=C_1+C_1z+C_3w+C_4zw,$$

where

$$\begin{aligned}
C_1&=2\mathrm{Re}\big[\overline{a_n}\overline{b_1}(1-a_n^2)(1-b_1^2)F(a_n,b_1)\\
&\quad-a_n\overline{b_1}(1-\overline{a_n}^2)(1-b_1^2)F(\overline{a_n},b_1)\big],\\
C_2&=2\mathrm{Re}\big[-\overline{b_1}(1-a_n^2)(1-b_2^2)F(a_n,b_1)\\
&\quad+\overline{b_1}(1-\overline{a_n}^2)(1-b_1^2)F(\overline{a_n},b_1)\big],\\
C_3&=2\mathrm{Re}\big[-\overline{a_n}(1-a_{n+1}^2)(1-b_1^2)F(a_n,b_1)\\
&\quad+a_n(1-\overline{a_n}^2)(1-b_1^2)F(\overline{a_n},b_1)\big],\\
C_4&=2\mathrm{Re}\big[(1-a_n^2)(1-b_1^2)F(a_n,b_1)\\
&\quad-(1-\overline{a_n}^2)(1-b_1^2)F(\overline{a_n},b_1)\big].
\end{aligned}$$

Here, Re represents taking the real part of a complex number.

As a result, $PS_n(f)$ is a rational function with real coefficients. ■

2) Modified W-POAFD:

Theorem 10: Let $f(z,w)\in\mathcal{H}$ with real coefficients and $\tilde{S}_n(f)$ be the n-partial sum in (8) obtained by the modified W-POAFD. Then $\tilde{S}_n(f)$ has real coefficients.

To prove the theorem, we need the following proposition. In order to improve readability, the proof of the latter is deferred to Appendix C.

Proposition 11: Assume $\{v_1,v_2,\cdots v_m,\cdots\}$ is selected from $\mathcal{D}^2$ according to conjugate parameters selection. Namely, for any even number m, if $v_{m-1}=e_ae_b$, then $v_m=e_{\overline{a}}e_{\overline{b}}$. Use the same notations in the G-S orthonormalization (9),

$$u_k=v_k-\sum_{p=1}^{k-1}\frac{\langle v_k,u_p\rangle}{\|u_p\|^2}u_p,\quad k=1,2\cdots,\tag{13}$$

then for any even number m, there are reformulations

$$u_{m-1}=v_{m-1}+h_1,\quad u_m=v_m+tu_{m-1}+h_2,\tag{14}$$

where $h_1=\alpha_{m-2}v_{m-2}+\alpha_{m-3}v_{m-3}+\cdots+\alpha_2v_2+\alpha_1v_1$, $h_2=\overline{\alpha_{m-3}}v_{m-2}+\overline{\alpha_{m-2}}v_{m-3}+\cdots+\overline{\alpha_1}v_2+\overline{\alpha_2}v_1$, $\alpha_1,\cdots,\alpha_{m-2}$ are complex numbers related to u_{m-1} and $t=-\frac{\langle v_m,u_{m-1}\rangle}{\|u_{m-1}\|^2}$. In this case, if we replace the coefficients of h_1 with their conjugates, we obtain h_2.

Proof of Theorem 10:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2019.2913047, IEEE Transactions on Automatic Control

$\tilde{S}_n(f)$ is a sum of $2n$ terms. We claim that if we divide $\tilde{S}_n(f)$ into n groups by collecting 2 terms with adjacent summation indices to one group, then each group has real coefficients. We define $P\tilde{S}_n(f)$ by $P\tilde{S}_n(f) = \langle f, B_n\rangle B_n + \langle f, \tilde{B}_n\rangle \tilde{B}_n$ and prove, as an example, that the group $P\tilde{S}_n(f)$ has real coefficients.

According to Proposition 11, we reformulate u_{2n-1}, u_{2n} as

$$u_{2n-1} = v_{2n-1} + h_1, \quad u_{2n} = v_{2n} + tu_{2n-1} + h_2,$$

where $t = -\frac{\langle v_{2n}, u_{2n-1}\rangle}{\|u_{2n-1}\|^2}$. Since h_2 is the formula obtained by replacing the coefficients of h_1 with their conjugates and $\overline{f(z,w)} = f(\overline{z}, \overline{w})$, $\overline{\langle f, h_1\rangle} = \langle f, h_2\rangle$. By a direct calculation, we note

$$\begin{aligned} P\tilde{S}_n(f) =& \frac{1}{\|u_{2n-1}\|^2}\langle f, u_{2n-1}\rangle u_{2n-1} + \frac{1}{\|u_{2n}\|^2}\langle f, u_{2n}\rangle u_{2n} \\ =& d_2 v_{2n} + (d_1 + td_2)v_{2n-1} + (d_1 + td_2)h_1 + d_2 h_2, \end{aligned}$$

where $d_1 = \frac{\langle f, u_{2n-1}\rangle}{\|u_{2n-1}\|^2},\ d_2 = \frac{\langle f, u_{2n}\rangle}{\|u_{2n}\|^2}$.

It is easy to verify that $\overline{d_2} = d_1 + td_2$. In fact,

$$\begin{aligned} \overline{d_2} =& \overline{\frac{\langle f, v_{2n}\rangle + \bar{t}\langle f, v_{2n-1}\rangle + \bar{t}\langle f, h_1\rangle + \langle f, h_2\rangle}{\|u_{2n}\|^2}} \\ =& \frac{\langle f, v_{2n-1}\rangle + t\langle f, v_{2n}\rangle + t\langle f, h_2\rangle + \langle f, h_1\rangle}{\|u_{2n}\|^2}, \\ d_1 + td_2 =& \frac{(1-|t|^2)\langle f, u_{2n-1}\rangle + t\langle f, u_{2n}\rangle}{\|u_{2n}\|^2} \\ =& \big[((\langle f, v_{2n-1}\rangle + \langle f, h_1\rangle)(1-|t|^2) + t(\langle f, v_{2n}\rangle \\ &+ \bar{t}\langle f, v_{2n-1}\rangle + \bar{t}\langle f, h_1\rangle + \langle f, h_2\rangle)\big]\big/\|u_{2n}\|^2 \\ =& \frac{t\langle f, v_{2n}\rangle + \langle f, v_{2n-1}\rangle + t\langle f, h_2\rangle + \langle f, h_1\rangle}{\|u_{2n}\|^2}. \end{aligned}$$

Recall that the coefficients of v_{2n}, v_{2n-1} are conjugate to each other and the coefficients of h_1, h_2 are conjugate to each other. Taking these into account, in the formula of $P\tilde{S}_n(f)$, $d_2 v_{2n} + (d_1+td_2)v_{2n-1}$ and $(d_1+td_2)h_1 + d_2 h_2$ have real coefficients. As a result, $P\tilde{S}_n(f)$ has real coefficients. ■

Remark 12: The modified 2-D AFD and the modified W-POAFD can be generalized to the n-variables cases as long as we use the tensor form n-variables product-TM system

$$\begin{aligned} &\mathcal{B}_{N_1}^{\mathbf{a}_1} \otimes \mathcal{B}_{N_2}^{\mathbf{a}_2} \cdots \otimes \mathcal{B}_{N_n}^{\mathbf{a}_n}(z_1, z_2, \cdots, z_n) \\ =& \{B_{k_1}^{\mathbf{a}_1}(z_1) B_{k_2}^{\mathbf{a}_2}(z_2) \cdots B_{k_n}^{\mathbf{a}_n}(z_n)\}_{k_1=1,\cdots,N_1;\cdots;k_n=1,\cdots,N_n} \end{aligned}$$

and the tensor form n-variables product-Szegö dictionary

$$\mathcal{D}^n = \{e_{a_1}(z_1)e_{a_2}(z_2)\cdots e_{a_n}(z_n) : a_1, \cdots, a_n \in \mathbb{D}\}.$$

Here, $z_1, z_2, \cdots, z_n$ are the n coordinates of the polydisk $\mathbb{D}^n$. Meanwhile, the conjugate parameters selection and the modified maximal selection principles need to be presented in the n-variables structure similarly. Therefore, by utilizing the n-polydisk tensor product type Cauchy integral of function $g(\mathrm{e}^{\mathrm{j}t_1}, \mathrm{e}^{\mathrm{j}t_2}, \cdots, \mathrm{e}^{\mathrm{j}t_n})$ defined on n-torus $\partial\mathbb{D}^n$

$$\begin{aligned} &\mathcal{C}(g)(z_1, z_2, \cdots, z_n) \\ =& (\frac{1}{2\pi\mathrm{j}})^n \int_0^{2\pi} \cdots \int_0^{2\pi} \frac{g(\mathrm{e}^{\mathrm{j}t_1}, \mathrm{e}^{\mathrm{j}t_2}, \cdots, \mathrm{e}^{\mathrm{j}t_n})}{(\mathrm{e}^{\mathrm{j}t_1} - z_1)\cdots(\mathrm{e}^{\mathrm{j}t_n} - z_n)}\, \mathrm{d}\mathrm{e}^{\mathrm{j}t_1}\cdots \mathrm{d}\mathrm{e}^{\mathrm{j}t_n}, \end{aligned}$$

in the first stage, the two-stage procedure stated in Section IV can be generalized to solve the n-D system identification as well.

V. Error Analysis

To achieve (4) in Section II, which estimates the approximation error of the two-stage procedure in Section IV, we take two steps. First, we give estimates for the convergence rate of the proposed algorithms. Second, under the assumptions for the stochastic noise, we obtain an upper bound in term of mathematical expectation of the approximation error in the $\mathcal{H}$ norm sense.

A. Convergence Rate

To begin with, for the Hilbert space $\mathcal{H}$ and the dictionary $\mathcal{D}^2$ we define the class of functions

$$\begin{aligned} \mathcal{A}(\mathcal{D}^2, R) = \big\{ f \in \mathcal{H} : f = \sum_{s=1}^{\infty}(c_s w_s + \overline{c_s}\tilde{w}_s) \text{ where } w_s = \\ \mathrm{e}_{a_s}\mathrm{e}_{b_s} \in \mathcal{D}^2,\ \tilde{w}_s = \mathrm{e}_{\overline{a_s}}\mathrm{e}_{\overline{b_s}},\ \sum_{s=1}^{\infty}|c_s| \le \frac{R}{2} \text{ and } R > 0\big\}. \end{aligned}$$

Theorem 13: For $f \in \mathcal{A}(\mathcal{D}^2, R)$, let $S_n(f)$ be the n-partial sum obtained by the modified 2-D AFD, we have

$$\|f - S_{n-1}(f)\| \le \frac{R}{\sqrt{n}}.$$

Proof: Let $g_n = f - S_{n-1}(f)$. According to the definition of $\mathcal{A}(\mathcal{D}^2, R)$, on one hand we have

$$\begin{aligned} &\|g_n\|^2 \\ =& |\langle g_n, f\rangle| = |\langle g_n, \sum_{s=1}^{\infty}(c_s w_s + \overline{c_s}\tilde{w}_s)\rangle| \\ \le& \sum_{s=1}^{\infty}|c_s| \sup_s \big(|\langle g_n, w_s\rangle| + |\langle g_n, \tilde{w}_s\rangle|\big) \\ \le& \frac{R}{2}\sup_s \big(|\langle g_n, e_{a_s}e_{b_s}\rangle| + |\langle g_n, e_{\overline{a_s}}e_{\overline{b_s}}\rangle|\big) \\ =& \frac{R}{2}\sup_s \sqrt{(1-|a_s|^2)(1-|b_s|^2)}\big(|g_n(a_s, b_s)| + |g_n(\overline{a_s}, \overline{b_s})|\big), \end{aligned}$$

on the other hand, from the modified Maximal Selection Principle we have

$$\begin{aligned} &\|D_n(f)\|^2 \\ =& \sup_{a,b\in\mathbb{D}} \sum_{\max\{k,l\}=2n-1}^{2n} \big|\langle g_n, B_k^{\tilde{\mathbf{a}}} B_l^{\tilde{\mathbf{b}}}\rangle\big|^2 \\ \ge& \sup_{a,b\in\mathbb{D}} \big(\big|\langle g_n, B_{2n-1}^{\tilde{\mathbf{a}}} B_{2n-1}^{\tilde{\mathbf{b}}}\rangle\big|^2 + \big|\langle g_n, B_{2n}^{\tilde{\mathbf{a}}} B_{2n}^{\tilde{\mathbf{b}}}\rangle\big|^2\big) \\ =& \sup_{a,b\in\mathbb{D}} \big(\big|\langle h_n, e_a e_b\rangle\big|^2 + \big|\langle \tilde{h}_n, e_{\overline{a}}e_{\overline{b}}\rangle\big|^2\big) \\ \ge& \sup_s \big(\big|\langle h_n, w_s\rangle\big|^2 + \big|\langle \tilde{h}_n, \tilde{w}_s\rangle\big|^2\big) \\ \ge& \sup_s (1-|a_s|^2)(1-|b_s|^2)\big(|g_n(a_s, b_s)|^2 + |g_n(\overline{a_s}, \overline{b_s})|^2\big) \\ \ge& \sup_s (1-|a_s|^2)(1-|b_s|^2)\big(\frac{|g_n(a_s, b_s)| + |g_n(\overline{a_s}, \overline{b_s})|}{2}\big)^2, \end{aligned}$$

where $h_n = g_n \prod_{k=1}^{2n-2} \frac{1-\overline{\mathbf{a}_k}z}{z-\mathbf{a}_k}\frac{1-\overline{\mathbf{b}_k}w}{w-\mathbf{b}_k}$, $\tilde{h}_n = g_n \prod_{k=1}^{2n-1} \frac{1-\overline{\mathbf{a}_k}z}{z-\mathbf{a}_k}\frac{1-\overline{\mathbf{b}_k}w}{w-\mathbf{b}_k}$.

In the fifth equality-inequality relation of the above derivation we used $|h_n(a,b)| \geq |g_n(a,b)|$ that is because $|\frac{1-\bar{a}z}{z-a}| > 1$ for $|z| < 1,\ a \in \mathbb{D}$.

Comparing the above two inequalities, we have

$$\begin{aligned}\|g_{n+1}\|^2 &= \|g_n\|^2 - \|D_n(f)\|^2 \leq \|g_n\|^2 - \frac{1}{R^2}\|g_n\|^4 \\ &= \|g_n\|^2(1 - \frac{\|g_n\|^2}{R^2}).\end{aligned}$$

Thus, by Lemma 3.4 in [30], we obtain $\|g_n\| \leq \frac{R}{\sqrt{n}}$. ■

Theorem 14: For $f \in \mathcal{A}(\mathcal{D}^2, R)$, let $\tilde{S}_n(f)$ be the n-partial sum obtained by the modified W-POAFD, we have

$$\|f - \tilde{S}_{n-1}(f)\| \leq \frac{R_n R}{\rho}\frac{1}{\sqrt{n}},$$

where $R_n = \max\{r_1, \cdots, r_n\}$, $r_n = \sup_s\{\max\{r_n(w_s), \tilde{r}_n(\tilde{w}_s)\}\}$, $r_n(w) = \|Q_{\{v_1,\cdots,v_{n-1},\tilde{v}_{n-1}\}}(w)\|$, $\tilde{r}_n(w) = \|Q_{\{v_1,\cdots,v_{n-1},\tilde{v}_{n-1},v_n\}}(w)\|$, $Q_{\{\alpha_1,\alpha_2,\cdots,\alpha_n\}}(w) = w - \sum_{k=1}^{n-1}\langle w, X_k\rangle X_k$ and $\{X_1, \cdots, X_n\}$ is the G-S orthonormalization of $\{\alpha_1, \cdots, \alpha_n\}$.
The proof is similar with Theorem 3.3 in [30] but needs some modifications that we omit in this paper.

B. An Upper Bound for the Approximation Error

We adopt the noise model in (3) and assume $\Upsilon^U_{pq} \sim N_1^{\mathbb{C}}(0, \sigma_U^2)$, $\Upsilon^Y_{pq} \sim N_1^{\mathbb{C}}(0, \sigma_Y^2)$.

Theorem 15: Let $f(z,w) \in \mathcal{H}$ and be continuous on $\overline{\mathbb{D}}^2$. With an extra condition $|1+\Upsilon^U_{pq}| \geq C_U$ where C_U is a constant $\in (0,1)$, if $\tilde{f} \in \mathcal{A}(\mathcal{D}^2, R)$, then by using the modified 2-D AFD, we have

$$\mathbb{E}\|f - S_n(f)\|^2 \leq 2C\Delta^2_{MN} + 2CI^2_{MN}\frac{(\sigma_U^2 + \sigma_Y^2)}{C_U^2} + 2\frac{R^2}{n},$$

where $\Delta_{MN} = \max_{\text{constraint}}|f(\mathrm{e}^{\mathrm{j}t_1}, \mathrm{e}^{\mathrm{j}u_1}) - f(\mathrm{e}^{\mathrm{j}t_2}, \mathrm{e}^{\mathrm{j}u_2})|$, constraint $= \{t_1, t_2, u_1, u_2 \in [0, 2\pi),\ |t_1 - t_2| < \frac{2\pi}{M}, |u_1 - u_2| < \frac{2\pi}{N}\}$, $I_{MN} = \max_{1\leq p\leq M, 1\leq q\leq N}|f(\mathrm{e}^{\mathrm{j}\omega_p}, \mathrm{e}^{\mathrm{j}\omega'_q})|$, $C > 0$ is a constant. Similarly, by using the modified W-POAFD, we have

$$\mathbb{E}\|f - \tilde{S}_n(f)\|^2 \leq 2C\Delta^2_{MN} + 2CI^2_{MN}\frac{(\sigma_U^2 + \sigma_Y^2)}{C_U^2} + 2\frac{R_n^2R^2}{\rho^2 n},$$

where R_n is defined in Theorem 14.
Proof: We prove the theorem in the case of using the modified 2-D AFD. The proof is similar for the modified W-POAFD.

By the triangle inequality for the norm, we have

$$\begin{aligned}\mathbb{E}\|f - S_n(f)\|^2 &\leq \mathbb{E}(\|f - \tilde{f}\| + \|\tilde{f} - S_n(f)\|)^2 \\ &\leq 2\mathbb{E}\|f - \tilde{f}\|^2 + 2\mathbb{E}\|\tilde{f} - S_n(f)\|^2.\end{aligned}$$

That means, corresponding to the two stages in Section IV, the mathematical expectation of the total error $\mathbb{E}\|f - S_n(f)\|^2$ can be estimated separately by $\mathbb{E}\|f - \tilde{f}\|^2$ and $\mathbb{E}\|\tilde{f} - S_n(f)\|^2$, caused, respectively, by the noise in the measurements and the error by the proposed algorithms. It follows from Theorem 13 that $\mathbb{E}\|\tilde{f} - S_n(f)\|^2 \leq \frac{R^2}{n}$. To estimate $\mathbb{E}\|f - \tilde{f}\|^2$, we first introduce the following inequality (15).

Assume ξ and η are two independently complex normally distributed random variables, namely $\xi \sim N_1^{\mathbb{C}}(0, \sigma_\xi^2)$, $\eta \sim N_1^{\mathbb{C}}(0, \sigma_\eta^2)$. Then the expectation $\mathbb{E}|\frac{1+\xi}{1+\eta} - 1|^2 = E|\frac{\xi-\eta}{1+\eta}|^2$ is infinity for the singularity in the denominator. However if we add a condition $|1+\eta| \geq t$ where $0 < t < 1$, the expectation is computable and

$$\mathbb{E}|\frac{\xi - \eta}{1+\eta}|^2 \leq \frac{1}{t^2}\mathbb{E}|\xi - \eta|^2 = \frac{\sigma_\xi^2 + \sigma_\eta^2}{t^2}. \tag{15}$$

In fact, condition $|1+\eta| \geq t$ excludes a small neighbourhood on the complex plane around the singularity $\eta = -1$ as proceeded in [36] and thus can be estimated according to the literature.

Therefore, from (12) and (15),

$$\begin{aligned}&\mathbb{E}\|f - \tilde{f}\|^2 \\ \leq& 2C\mathbb{E}\big\|\big[\sum_{p,q} f(\mathrm{e}^{\mathrm{j}\omega_p}, \mathrm{e}^{\mathrm{j}\omega'_q}) - \sum_{p,q} E^{MN}_{p,q}\big]\chi_{p,q}(t,u)\big\|^2_{L^2} \\ &+ 2C\mathbb{E}\|f(\mathrm{e}^{\mathrm{j}t}, \mathrm{e}^{\mathrm{j}u}) - \sum_{p,q} f(\mathrm{e}^{\mathrm{j}\omega_p}, \mathrm{e}^{\mathrm{j}\omega'_q})\chi_{p,q}(t,u)\|^2_{L^2} \\ \leq& 2C\mathbb{E}\|\sum_{p,q} f(\mathrm{e}^{\mathrm{j}\omega_p}, \mathrm{e}^{\mathrm{j}\omega'_q})(\frac{1+\Upsilon^Y_{p,q}}{1+\Upsilon^U_{p,q}} - 1)\chi_{p,q}(t,u)\|^2_{L^2} \\ &+ 2C\Delta^2_{MN} \\ \leq& 2CI^2_{MN}\frac{\sigma_U^2 + \sigma_Y^2}{C_U^2} + 2C\Delta^2_{MN},\end{aligned}$$

where C is a constant. The proof is complete. ■

Consequently, if we take $G_n(\mathrm{e}^{\mathrm{j}t}, \mathrm{e}^{\mathrm{j}u}) = S_n(f)(\mathrm{e}^{-\mathrm{j}t}, \mathrm{e}^{-\mathrm{j}u})$ by using the modified 2-D AFD (or $G_n(\mathrm{e}^{\mathrm{j}t}, \mathrm{e}^{\mathrm{j}u}) = \tilde{S}_n(f)(\mathrm{e}^{-\mathrm{j}t}, \mathrm{e}^{-\mathrm{j}u})$ by using the modified W-POAFD) as the approximation to $G(\mathrm{e}^{\mathrm{j}t}, \mathrm{e}^{\mathrm{j}u})$, we conclude (4) utilizing Theorem 15 under the assumptions in Theorem 15.

VI. Numerical experiments

We apply the proposed system identification procedure to the examples in [37], and we solve the optimization problems (6) and (10) by discretization, as suggested in Remark 3. In addition, we will make the comparison with the multi-Fourier expansion, as, in fact, Fourier series of several complex variables is the generalization of FIR model in the n-D cases. In particular, if we select parameters $a_n, b_n = 0$ at every iteration, the modified 2-D AFD gives rise to Fourier series. The code is implemented in Matlab.

We give two evaluation criteria to evaluate the approximation $f_n(z,w)$ to $f(z,w)$ obtained by the proposed methods, which is $S_n(f)$ in the modified 2-D AFD and $\tilde{S}_n(f)$ in the modified W-POAFD.

(1) Given measurements $E^{MN}_{p,q} = f(\mathrm{e}^{\mathrm{j}\omega_p}, \mathrm{e}^{\mathrm{j}\omega'_q})\frac{1+\Upsilon^Y_{p,q}}{1+\Upsilon^U_{p,q}}, p = 1, \cdots, M,\ q = 1, \cdots, N$, the *relative error* between f and f_n is defined to evaluate the accuracy of energy approximation as below

$$RE = \frac{\sum_{p=1}^{M}\sum_{q=1}^{N}|f(\mathrm{e}^{\mathrm{j}\omega_p}, \mathrm{e}^{\mathrm{j}\omega'_q}) - f_n(\mathrm{e}^{\mathrm{j}\omega_p}, \mathrm{e}^{\mathrm{j}\omega'_q})|^2}{\sum_{p=1}^{M}\sum_{q=1}^{N}|f(\mathrm{e}^{\mathrm{j}\omega_p}, \mathrm{e}^{\mathrm{j}\omega'_q})|^2}.$$

(2) If the input $\{u(i,j)\}$ is a zero-mean wide-sense stationary process, the power spectral densities $S_u(\omega, \omega')$, $S_y(\omega, \omega')$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2019.2913047, IEEE Transactions on Automatic Control

of input $u(i,j)$ and output $y(i,j)$ and the transfer function $G(z_1,z_2)$ have the following relation

$$S_y(\omega,\omega') = |G(\mathrm{e}^{\mathrm{j}\omega},\mathrm{e}^{\mathrm{j}\omega'})|^2 S_u(\omega,\omega').$$

Besides the above quantitative criteria, in the comparison we also draw color graphs to show $\log_{10}|f_n(\mathrm{e}^{\mathrm{j}\omega_p},\mathrm{e}^{\mathrm{j}\omega'_q})|^2$ which reflects the details of f_n at the frequency (ω_p,ω'_q).

Example 1: The transfer function is

$$G(z_1,z_2) = \frac{(2+z_1^{-1})+(3-.5z_1^{-1})z_2^{-1}}{(1-1.6z_1^{-1}+1.4z_1^{-2}-.48z_1^{-3})(1-.6z_2^{-1}+.25z_2^{-2})}.$$

We apply the proposed methods to $f(z,w)$ obtained from $G(z_1,z_2)$ through the mappings $z_1 \to \frac{1}{z}, z_2 \to \frac{1}{w}$.

TABLE I

THE REs BY THE MODIFIED 2-D AFD WITH VARYING OUTPUT AND INPUT NOISE FOR EXAMPLE 1. THE NOISE OF OUTPUT AND INPUT ARE ADDED IN THE FOLLOWING WAY. TAKING THE OUTPUT AS AN EXAMPLE, ASSUME THE STOCHASTIC NOISE $\Upsilon_{p,q}^Y \sim N_1^{\mathbb{C}}(0,\sigma^2)$. THE POWER OF THE NOISE-FREE OUTPUT Y AND NOISE $Y\Upsilon^Y$ ARE CALCULATED AND σ IS SELECTED TO ACHIEVE THE DESIRED SNR. NOTATION SNR=X MEANS WE ONLY ADD OUTPUT NOISE WITH SNR=X. NOTATION SNR=X/Y MEANS WE ADD OUTPUT NOISE WITH SNR=X AND INPUT NOISE WITH SNR=Y. $M=N=L$ AND MN IS THE NUMBER OF MEASUREMENTS. n IS THE NUMBER OF ITERATIONS.

(a) SNR=∞

n	L=64	L=128	L=256	L=512	L=1024
1	0.0422	0.0423	0.0423	0.0423	0.0423
2	0.0155	0.0155	0.0155	0.0155	0.0155
3	7.29e-04	7.29e-04	7.29e-04	7.29e-04	7.29e-04
4	1.76e-04	1.76e-04	1.76e-04	1.76e-04	1.76e-04

(b) SNR=20

n	L=64	L=128	L=256	L=512	L=1024
1	0.0423	0.0423	0.0423	0.0423	0.0423
2	0.0157	0.0155	0.0155	0.0155	0.0155
3	0.0012	7.64e-04	7.67e-04	7.37e-04	6.90e-04
4	8.17e-04	4.18e-04	2.23e-04	1.88e-04	1.79e-04

(c) SNR=10

n	L=64	L=128	L=256	L=512	L=1024
1	0.0445	0.0424	0.0423	0.0423	0.0423
2	0.0117	0.0161	0.0156	0.0155	0.0155
3	0.0076	0.0022	9.04e-04	7.63e-04	7.50e-04
4	0.0138	0.0022	7.01e-04	2.92e-04	2.09e-04

(d) SNR=20/20

n	L=64	L=128	L=256	L=512	L=1024
1	0.0427	0.0423	0.0423	0.0423	0.0423
2	0.0191	0.0156	0.0155	0.0155	0.0155
3	0.0020	9.24e-04	7.17e-04	7.44e-04	6.46e-04
4	0.0023	5.63e-04	2.90e-04	1.97e-04	1.83e-04

From TABLE I, it is noted that given different number of measurements, we achieve relative errors $\le 10^{-2}$ within 3 iterations no matter whether it is in the presence of noise or not. With the increase of noise, it needs more measurements to achieve a similar accuracy compared with the noise-free case.

TABLE II lists the parameters pairs of the product-TM system selected by the modified 2-D AFD in the previous 3 iterations (omit the conjugate pairs) with varying noise when $L=1024$. We notice that in some iterations parameters change greatly when the measurements are disrupted by noise. This phenomenon may happen since the optimal solution of the maximum problem in (6) may not be unique although the maximum value of (6) is unique. To verify the statement, we give fix scatter plots in Fig. 1 for the parameters satisfying certain condition (refer to the caption of Fig. 1) in the previous 3 iterations in the noise-free case. To a certain extent, all those parameters can be regarded as the optimal solutions approximately since they correspond to $\|D_n^{\tilde{\mathbf{a}},\tilde{\mathbf{b}}}(f)\|^2$ which is almost equal to the maximum value $\|D_n(f)\|^2$. From Fig. 1, we observe that $S_2(f), S_3(f)$ obtained by the modified 2-D AFD are not unique.

TABLE II

THE PARAMETERS PAIRS OF THE PRODUCT-TM SYSTEM SELECTED BY THE MODIFIED 2-D AFD IN THE PREVIOUS 3 ITERATIONS WITH VARYING OUTPUT AND INPUT NOISE WHEN $L=1024$ FOR EXAMPLE 1. n IS THE NUMBER OF ITERATIONS.

(a) parameter a

n	1	2	3
SNR=∞	$0.5037+0.6986$j	$0.7409+0.1768$j	$0.4815+0.7586$j
SNR=20	$0.5037+0.6986$j	$0.7409-0.1723$j	$0.4815+0.7586$j
SNR=10	$0.5037+0.6986$j	$0.7409-0.1768$j	$0.4815-0.7586$j
SNR=20/20	$0.5037-0.6986$j	$0.7364-0.1723$j	$0.4815-0.7586$j

(b) parameter b

n	1	2	3
SNR=∞	$0.2999-0.3984$j	$0.2999-0.3984$j	$0.2999-0.3984$j
SNR=20	$0.2999+0.4006$j	$0.5800-0.6875$j	$0.5867-0.6809$j
SNR=10	$0.2999+0.3984$j	$0.7839+0.4406$j	$0.5778+0.6898$j
SNR=20/20	$0.2999-0.3984$j	$0.4437+0.6000$j	$0.7706+0.4628$j

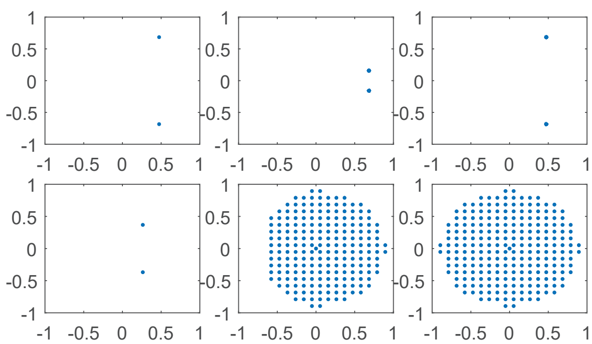


Fig. 1. The six figures show the parameters a,b in $\tilde{a}=\{a_1,\cdots,\overline{a_{n-1}}, a,\overline{a}\}, \tilde{b}=\{b_1,\cdots,\overline{b_{n-1}}, b,\overline{b}\}$ corresponding to $\|D_n^{\tilde{a},\tilde{b}}(f)\|^2 \ge (1-10^{-4})\times\|D_n(f)\|^2$ at the n-th iteration when $L=1024$ and the measurements are noise-free by the modified 2-D AFD for example 1. The three columns are in turn for the 1-th, 2-th, 3-th iteration. The first row is for parameter a and the second row is for parameter b.

In TABLE III, we show relative errors when the modified W-POAFD is applied to example 1 with varying output and input noise. In TABLE IV, we show the parameters pairs selected by the modified W-POAFD in the previous 6 iterations when $L=1024$. Compared with the modified 2-D AFD, it needs more iterations to achieve a similar accuracy. Besides, the parameter pairs chosen by two proposed algorithms are different because of the usage of different modified maximal selection principles. On the other hand, corresponding to the same iteration number n, the modified 2-D AFD has $4n^2$ orthogonal terms while the modified W-POAFD only has $2n$ orthogonal terms.

In TABLE V, we list the relative errors obtained by Fouri-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2019.2913047, IEEE Transactions on Automatic Control

TABLE III
THE REs BY THE MODIFIED W-POAFD WITH VARYING OUTPUT AND INPUT NOISE FOR EXAMPLE 1. THE NOTATIONS USED HERE ARE THE SAME AS THOSE IN TABLE I.

(a) SNR=∞

n	L=64	L=128	L=256	L=512	L=1024
1	0.2611	0.2611	0.2611	0.2611	0.2611
2	0.1140	0.1140	0.1140	0.1140	0.1140
3	0.0587	0.0587	0.0587	0.0587	0.0587
6	0.0069	0.0069	0.0069	0.0069	0.0069
9	8.20e-04	8.21e-04	8.21e-04	8.21e-04	8.21e-04
10	5.28e-04	5.28e-04	5.28e-04	5.28e-04	5.28e-04

(b) SNR=20

n	L=64	L=128	L=256	L=512	L=1024
1	0.2611	0.2611	0.2611	0.2611	0.2611
2	0.1140	0.1140	0.1140	0.1140	0.1140
3	0.0573	0.0588	0.0588	0.0587	0.0587
6	0.0061	0.0071	0.0061	0.0069	0.0069
9	0.0022	0.0011	0.0015	7.17e-04	8.22e-04
10	0.0015	6.93e-04	8.93e-04	4.35e-04	5.29e-04

(c) SNR=10

n	L=64	L=128	L=256	L=512	L=1024
1	0.2614	0.2611	0.2611	0.2611	0.2611
2	0.1144	0.1140	0.1140	0.1140	0.1140
3	0.0595	0.0580	0.0573	0.0587	0.0587
6	0.0080	0.0061	0.0057	0.0060	0.0069
9	0.0061	0.0017	0.0014	0.0014	8.34e-04
10	0.0062	0.0012	8.73e-04	8.10e-04	5.42e-04

(d) SNR=20/20

n	L=64	L=128	L=256	L=512	L=1024
1	0.2611	0.2611	0.2611	0.2611	0.2611
2	0.1140	0.1140	0.1140	0.1140	0.1140
3	0.0582	0.0588	0.0587	0.0587	0.0587
6	0.0093	0.0067	0.0082	0.0069	0.0069
9	0.0022	0.0022	0.0015	8.12e-04	8.22e-04
10	0.0017	0.0015	9.22e-04	5.31e-04	5.30e-04

TABLE IV
THE SZEGÖ KERNEL PARAMETERS PAIRS SELECTED BY THE MODIFIED W-POAFD WITH VARYING OUTPUT AND INPUT NOISE SNR=∞, 20, 10, 20/20 WHEN $L = 1024$ FOR EXAMPLE 1. THE SELECTED PARAMETERS ARE THE SAME IN THE 4 CASES.

n	parameter a	parameter b
1	$0.5111 + 0.6889\mathrm{i}$	$0.5611 + 0.2389\mathrm{i}$
2	$0.5111 + 0.6889\mathrm{i}$	$0.2167 - 0.6167\mathrm{i}$
3	$0.5111 + 0.7333\mathrm{i}$	$0.3944 + 0.5833\mathrm{i}$
4	$0.7278 + 0.0276\mathrm{i}$	$0.5167 - 0.4278\mathrm{i}$
5	$0.5333 + 0.7111\mathrm{i}$	$0.3167 - 0.5389\mathrm{i}$
6	$0.4889 + 0.7111\mathrm{i}$	$0.4611 + 0.2056\mathrm{i}$

er series and relative errors corresponding to the identified transfer functions given in [37] in the presence of noise or not. The method in [37] identify the transfer functions by calculating the coefficients of polynomials of two complex variables in the numerator and denominator. For example 1, the authors of [37] assume that the numerator is already known and only calculate the denominator. From TABLE I, TABLE III, TABLE V, we find out that to obtain a similar accuracy with the method in [37], the modified 2-D AFD, the modified W-POAFD and Fourier series need 3, 9 and 14 iterations respectively in the noise-free case. When the measurements are disrupted by noise, the modified 2-D AFD and modified W-POAFD still give comparable results within almost equal number of iterations, but Fourier series does not.

In Fig. 2, color graphs of $\log_{10}|f_n(\mathrm{e}^{\mathrm{j}\omega_p}, \mathrm{e}^{\mathrm{j}\omega'_q})|^2$ for the

TABLE V
THE REs OBTAINED BY TWO OTHER METHODS WITH VARYING OUTPUT AND INPUT NOISE WHEN $L = 1024$ FOR EXAMPLE 1.

(a) relative errors by the Fourier series

n	1	2	4	8	14	20
SNR=∞	0.5941	0.3086	0.1362	0.0187	0.0013	1.08e-04
SNR=20	0.6035	0.3180	0.1458	0.0282	0.0109	0.0097
SNR=10	0.6909	0.4051	0.2330	0.1152	0.0978	0.0965
SNR=20/20	0.6135	0.3281	0.1555	0.0381	0.0206	0.0194

(b) relative errors by the method in [37]

	SNR=∞	SNR=20	SNR=10
method in [37]	0.0027	0.0014	3.30e-04

approximations obtained by two proposed algorithms at some iterations when $L = 256$ and SNR=10 are shown. We can see the approaching details at given frequencies.

Example 2: The transfer function is

$$G(z_1, z_2) = \frac{(1 + z_1^{-1}) + (3 + z_1^{-1})z_2^{-1}}{(1 + .6z_1^{-1} + .36z_1^{-2} + .48z_1^{-3})(1 + .7z_2^{-1})}.$$

We apply the proposed methods to $f(z, w)$ obtained from $G(z_1, z_2)$ through the mappings $z_1 \to \frac{1}{z}, z_2 \to \frac{1}{w}$.

TABLE VI
THE REs OBTAINED BY THE MODIFIED 2-D AFD, MODIFIED W-POAFD AND FOURIER SERIES WITH SNR=20/20 WHEN L=256 FOR EXAMPLE 2.

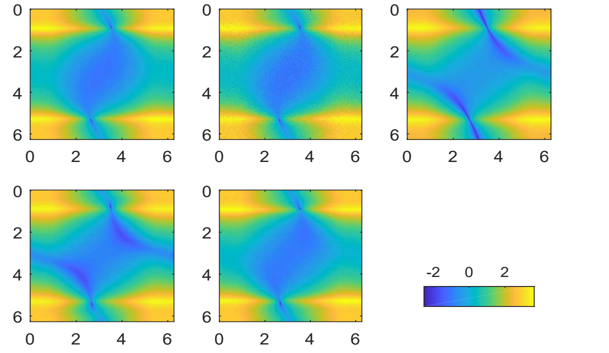
	modified 2-D AFD		modified W-POAFD		Fourier series	
n	SNR=∞	SNR=20/20	SNR=∞	SNR=20/20	SNR=∞	SNR=20/20
1	0.0280	0.0280	0.1014	0.1014	0.4792	0.4991
2	8.02e-04	8.12e-04	0.0686	0.0686	0.1088	0.1280
3	2.10e-05	5.27e-05	0.0149	0.0150	0.0258	0.0449
4	5.83e-07	9.96e-05	4.43e-03	4.45e-03	6.10e-03	0.0253

[37] only gives estimation in the noise-free case as

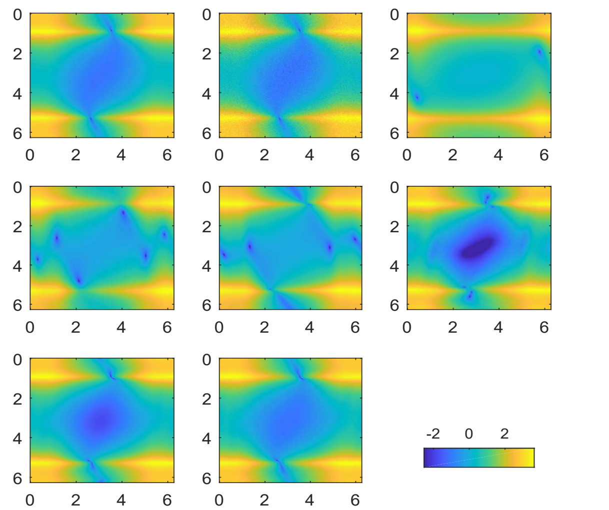
$$\hat{G}(z_1, z_2) = \frac{1 + 1.0364z_1^{-1} + 3.2713z_2^{-1} + 1.0485z_1^{-1}z_2^{-1}}{(1 + .6027z_1^{-1} + .3897z_1^{-2} + .0551z_1^{-3})(1 + .6978z_2^{-1})}.$$

Relative error between $\hat{G}$ and G is 0.0154. From TABLE VI, we find that the modified 2-D AFD and the modified W-POAFD give models having comparable relative errors with $\hat{G}$ after 2 and 4 iterations even when we add noise. Here, for example 2, we do not include the parameters pairs and figures for the identified models obtained by the proposed methods.

Remark 16: What [37] addresses is reconstruction of sparse signals, and thus, compared with the proposed algorithms of the same system order, the results of [37] are better. The AFD type decompositions, however, are designed for reconstruction of general signals that are not necessarily sparse. The latter uses the energy matching principle that, as a result, gives rise to a sparse representation in the energy sense. By this principle, if, for instance, one is given a signal composed of two reproducing kernels of almost balanced energy, then the first step of the energy matching would not give any one of the two reproducing kernels, and due to the result of the first step, the second step would not give any one of them, either. Although the proposed algorithms do not give rise to the correct decomposition of a sparse signal, the algorithms



(a) The five color graphs from left to right and from the first row to the second are, respectively, for f, noisy f, $S_1(f)$, $S_2(f)$, $S_3(f)$ by the modified 2-D AFD when $L = 256$ and SNR=10. The sixth color graph is the colorbar of the previous 5 graphs.



(b) The eight color graphs from left to right and from the first row to the second and then the third are, respectively, for f, noisy f, $\tilde{S}_1(f)$, $\tilde{S}_2(f)$, $\tilde{S}_3(f)$, $\tilde{S}_6(f)$,$\tilde{S}_9(f)$, $\tilde{S}_{10}(f)$ by the modified W-POAFD when $L = 256$ and SNR=10. The ninth graph is the colorbar of the previous 8 graphs.

Fig. 2. The color in the graphs represents the value of $\log_{10}|g(\mathrm{e}^{\mathrm{j}\omega_p}, \mathrm{e}^{\mathrm{j}\omega'_q})|^2$ for function g and the x-axis and y-axis represent the frequency ω and ω' respectively.

do lead to a fast converging decomposition of the signal and the results in a sparse representation.

A variation of AFD can get the correct decomposition of a sparse signal: A cyclic AFD method is established for this purpose. 1-D Cyclic AFD is developed in [38] to treat the n-best rational approximation problem and gets promising applications in model reduction [39]. This article does not develop this aspect. For a given and fixed n, cyclically use the algorithms proposed in the present paper one can indeed get the correct decomposition of a sparse signal in the 2-D case as well.

VII. Conclusion

In this paper, we propose two practical methods to solve 2-D system identification. The key advantage of the proposed algorithms is that the system identifying partial sum sequence, based on the adaptively chosen parameters, converges to the original transfer function at a fast rate, and thus offers a sparse representation in the energy sense. Moreover, the methodology of the modified W-POAFD can be generalized to any reproducing kernel Hilbert space with the so called boundary vanishing condition [30], [40].

Appendix A
Definition of Hardy-2 space on poly unit disk

Denote by $\mathcal{H} = H_2(\mathbb{D}^2)$ the Hardy space defined on $\mathbb{D}^2$ consisting of all complex valued holomorphic functions satisfying

$$\sup_{0\le r_1,r_2<1}\int_0^{2\pi}\int_0^{2\pi}|f(r_1\mathrm{e}^{\mathrm{j}t}, r_2\mathrm{e}^{\mathrm{j}u})|^2\ \mathrm{d}t\mathrm{d}u < \infty,$$

where $\mathbb{D}^2 = \{(z,w) : |z| < 1, |w| < 1\}$. $\mathcal{H}$ is a reproducing kernel Hilbert space with inner product $\langle\cdot,\cdot\rangle$ defined by

$$\langle f, g\rangle = \frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi} f(\mathrm{e}^{\mathrm{j}t}, \mathrm{e}^{\mathrm{j}u})\overline{g(\mathrm{e}^{\mathrm{j}t}, \mathrm{e}^{\mathrm{j}u})}\mathrm{d}t\mathrm{d}u,$$

where $\overline{g}$ is the conjugate of g. We call $\|f\|^2$ the "energy" of f. The norm $\|f\| = \sqrt{\langle f, f\rangle}$ is induced by the inner product.

Appendix B
Some results related to the generalized partial backward shift operator

Here, we omit the factor 1 in the tensor, for instance, we write $e_a \cdot 1$ as e_a in brief.

Proposition 17: Let $f(z,w) \in \mathcal{H}$, then

$$\prod_{l=1}^{m} S_{b_l}^{(2)} \prod_{k=1}^{n} S_{a_k}^{(1)}(f)(z,w)$$
$$= \frac{1}{\widetilde{B}_n^{\mathbf{a}}(z)\widetilde{B}_m^{\mathbf{b}}(w)}\Big[f(z,w) - \sum_{k=1}^{n}\langle f, B_k^{\mathbf{a}}\rangle^{(1)}(w)B_k^{\mathbf{a}}(z) -$$
$$\sum_{l=1}^{m}\left\langle f, B_l^{\mathbf{b}}\right\rangle^{(2)}(z)B_l^{\mathbf{b}}(w) + \sum_{l=1}^{m}\sum_{k=1}^{n}\left\langle f, B_k^{\mathbf{a}}B_l^{\mathbf{b}}\right\rangle B_k^{\mathbf{a}}(z)B_l^{\mathbf{b}}(w)\Big]$$

holds for all nonnegative integer m,n where $\widetilde{B}_n^{\mathbf{a}}(z) = \prod_{k=1}^{n}\frac{z-a_k}{1-\overline{a}_k z}$, $\widetilde{B}_m^{\mathbf{b}}(w) = \prod_{l=1}^{m}\frac{w-b_l}{1-\overline{b}_l w}$, $\mathbf{a} = \{a_1,\cdots,a_n\}$, $\mathbf{b} = \{b_1,\cdots,b_m\}$.

Proof: We use mathematical induction.

Firstly, the cases $m,n = 0$ and $m,n = 1$ hold trivially.

Secondly, we assume the proposition holds for all integers $m \le p-1$ and $n \le q-1$ for fixed p,q.

Finally, we prove the cases $m \le p$ and $n \le q$. We divide them into two subcases: (m,q) where $m \le p-1$ and (p,n) where $n \le q$. We discuss the first subcase and the other is similar.

It is noticed that if $f \in \mathcal{H}$, then $S_a^{(1)}(f)$ and $S_b^{(2)}(f)$ are still in $\mathcal{H}$ for $a,b \in \mathbb{D}$. By computation we know

$$\frac{1}{\frac{z-a}{1-\overline{a}z}}\Big|_{z=\mathrm{e}^{\mathrm{j}t}} = \frac{1-\overline{a}\mathrm{e}^{\mathrm{j}t}}{\mathrm{e}^{\mathrm{j}t}-a} = \overline{\frac{z-a}{1-\overline{a}z}\Big|_{z=\mathrm{e}^{\mathrm{j}t}}}.$$

It is easy to deduce that

$$\begin{aligned}&\Big\langle \prod_{l=1}^{m} S_{b_l}^{(2)} \prod_{k=1}^{q-1} S_{a_k}^{(1)}(f), e_{a_q}\Big\rangle^{(1)}\\
=&\Big\langle \frac{1}{\widetilde{B}_m^{\mathbf{b}}}\Big[f - \sum_{k=1}^{q-1}\big\langle f, B_k^{\mathbf{a}}\big\rangle^{(1)} B_k^{\mathbf{a}} - \sum_{l=1}^{m}\big\langle f, B_l^{\mathbf{b}}\big\rangle^{(2)} B_l^{\mathbf{b}}\\
&+\sum_{l=1}^{m}\sum_{k=1}^{q-1}\big\langle f, B_k^{\mathbf{a}}B_l^{\mathbf{b}}\big\rangle B_k^{\mathbf{a}}B_l^{\mathbf{b}}\Big], B_q^{\mathbf{a}}\Big\rangle^{(1)}\\
=&\frac{\big\langle f, B_q^{\mathbf{a}}\big\rangle^{(1)} - \sum\limits_{l=1}^{m}\big\langle f, B_q^{\mathbf{a}}B_l^{\mathbf{b}}\big\rangle}{\widetilde{B}_m^{\mathbf{b}}},\end{aligned}$$

holds for $m \le p-1$. By directly plugging in the last equation, the proposition holds for the subcase (m, q) where $m \le p-1$ immediately since

$$\begin{aligned}&\prod_{l=1}^{m} S_{b_l}^{(2)} \prod_{k=1}^{q} S_{a_k}^{(1)}(f)(z,w)\\
=&S_{a_q}^{(1)}\Big(\prod_{l=1}^{m} S_{b_l}^{(2)} \prod_{k=1}^{q-1} S_{a_k}^{(1)}(f)\Big)(z,w)\\
=&\frac{\prod\limits_{l=1}^{m} S_{b_l}^{(2)} \prod\limits_{k=1}^{q-1} S_{a_k}^{(1)}(f) - \Big\langle \prod\limits_{l=1}^{m} S_{b_l}^{(2)} \prod\limits_{k=1}^{q-1} S_{a_k}^{(1)}(f), e_{a_q}\Big\rangle^{(1)} e_{a_q}}{\frac{z-a_q}{1-\overline{a_q}z}}.\end{aligned}$$

The proof is complete. ■

Corollary 8 is directly derived from Proposition 17. ■

Proof of Corollary 9: By Proposition 17, Corollary 8 and the reproducing property of the Szegö kernel, we directly calculate

$$\begin{aligned}S_{\overline{a}}^{(1)} S_a^{(1)}(f)(z,w) =&\frac{|a|^2z^2 - 2\mathrm{Re}(a)z + 1}{z^2 - 2\mathrm{Re}(a)z + |a|^2} f(z,w)\\
&- \frac{1-|a|^2}{z^2 - 2\mathrm{Re}(a)z + |a|^2}\frac{F(z,w)}{2\mathrm{Im}(a)\mathrm{j}},\end{aligned}$$

where $F(z,w) = (1-a^2)(z-\overline{a})f(a,w) - (1-\overline{a}^2)(z-a)f(\overline{a},w)$, Re and Im represent taking the real part and the imaginary part of a complex number. Since $f(z,w) = \overline{f(\overline{z},\overline{w})}$, the first term in the last equality has real coefficients. As for the second term, it is noted that $F(z,w) = -\overline{F(\overline{z},\overline{w})}$. Then together with the imaginary part $2\mathrm{Im}(a)\mathrm{j}$ in the denominator, it has real coefficients. Consequently, $S_{\overline{a}}^{(1)} S_a^{(1)}(f)$ is a rational function with real coefficients.

If $a \in \mathbb{R}$, it is obvious that $S_a^{(1)}(f) = [(1-az)f(z,w) - (1-|a|^2)f(a,w)]/(z-a)$ is a real rational function. ■

Appendix C
Some results related to G-S orthonormalization

Proof of Proposition 11: We use mathematical induction.

Firstly, the proposition holds trivially when $m = 2$.

Secondly, let n be an even number and assume the proposition holds for all even numbers $m \le n$.

Finally, we aim to prove the proposition holds when $m = n+2$ using the above assumption. Actually, with the above assumption, we can conclude the following results (a) and (b). For all even numbers $m \le n$,

(a) $\|u_m\|^2 = \|u_{m-1}\|^2 - \frac{|\langle v_m, u_{m-1}\rangle|^2}{\|u_{m-1}\|^2}$,

(b) $\frac{\langle v_{m+1}, u_m\rangle\langle v_m, u_{m-1}\rangle}{\|u_m\|^2\|u_{m-1}\|^2} - \frac{\langle v_{m+1}, u_{m-1}\rangle}{\|u_{m-1}\|^2} = -\frac{\overline{\langle v_{m+2}, u_m\rangle}}{\|u_m\|^2}$.

We prove (a) using mathematical induction.

Trivially, note that $\|u_2\|^2 = 1 - \frac{|\langle v_2, u_1\rangle|^2}{\|u_1\|^2} = \|u_1\|^2 - \frac{|\langle v_2, u_1\rangle|^2}{\|u_1\|^2}$. Assume that (a) holds for all even numbers $k < n$, next we derive the n case. We claim

$$\sum_{p=k-1}^{k}\frac{|\langle v_n, u_p\rangle|^2}{\|u_p\|^2} = \sum_{p=k-1}^{k}\frac{|\langle v_{n-1}, u_p\rangle|^2}{\|u_p\|^2}, \tag{16}$$

holds for all even numbers $k < n$.

In fact, from the assumption, u_{k-1}, u_k are reformulated as

$$u_{k-1} = v_{k-1} + h_1, u_k = v_k + su_{k-1} + h_2,$$

where $s = -\frac{\langle v_k, u_{k-1}\rangle}{\|u_{k-1}\|^2}, h_1 = \beta_{k-2}v_{k-2} + \beta_{k-3}v_{k-3} + \cdots + \beta_2 v_2 + \beta_1 v_1, h_2 = \overline{\beta_{k-3}}v_{k-2} + \overline{\beta_{k-2}}v_{k-3} + \cdots + \overline{\beta_1}v_2 + \overline{\beta_2}v_1$, and $\beta_1, \cdots, \beta_{k-2}$ are complex numbers. Here, if we replace the coefficients of function h_1 with their conjugates we will obtain function h_2. This relation is a key point to deduce the formulas below:

$$\begin{aligned}&\frac{|\langle v_n, u_{k-1}\rangle|^2}{\|u_{k-1}\|^2} + \frac{|\langle v_n, u_k\rangle|^2}{\|u_k\|^2}\\
=&\frac{|\langle v_n, u_{k-1}\rangle|^2(1-|s|^2)\|u_{k-1}\|^2 + |\langle v_n, u_k\rangle|^2\|u_{k-1}\|^2}{\|u_{k-1}\|^2\|u_k\|^2}\\
=&\Big\{\big[|\langle v_n, v_k\rangle|^2 + |\langle v_n, v_{k-1}\rangle|^2\big] + \big[|\langle v_n, h_1\rangle|^2 +\\
&|\langle v_n, h_2\rangle|^2\big] + \big[\langle v_n, h_1\rangle\overline{\langle v_n, v_{k-1}\rangle} + \overline{\langle v_n, h_2\rangle}\langle v_n, v_k\rangle\big]\\
&+\big[\overline{\langle v_n, h_1\rangle}\langle v_n, v_{k-1}\rangle + \langle v_n, h_2\rangle\overline{\langle v_n, v_k\rangle}\big] + \big[\overline{s}\langle v_n, h_1\rangle\\
&\overline{\langle v_n, v_k\rangle} + \overline{s}\overline{\langle v_n, h_2\rangle}\langle v_n, v_{k-1}\rangle\big] + \big[s\overline{\langle v_n, h_1\rangle}\langle v_n, v_k\rangle +\\
&s\langle v_n, h_2\rangle\overline{\langle v_n, v_{k-1}\rangle}\big] + \big[s\langle v_n, v_k\rangle\overline{\langle v_n, v_{k-1}\rangle} +\\
&\overline{s}\langle v_n, v_{k-1}\rangle\overline{\langle v_n, v_k\rangle} + \overline{s}\langle v_n, h_1\rangle\overline{\langle v_n, h_2\rangle}\\
&+ s\overline{\langle v_n, h_1\rangle}\langle v_n, h_2\rangle\big]\Big\}/\|u_k\|^2\\
=&\Big\{\big[|\langle v_{n-1}, v_k\rangle|^2 + |\langle v_{n-1}, v_{k-1}\rangle|^2\big] + \big[|\langle v_{n-1}, h_1\rangle|^2 +\\
&|\langle v_{n-1}, h_2\rangle|^2\big] + \big[\langle v_{n-1}, h_1\rangle\overline{\langle v_{n-1}, v_{k-1}\rangle} + \overline{\langle v_{n-1}, h_2\rangle}\\
&\langle v_{n-1}, v_k\rangle\big] + \big[\overline{\langle v_{n-1}, h_1\rangle}\langle v_{n-1}, v_{k-1}\rangle + \langle v_{n-1}, h_2\rangle\\
&\overline{\langle v_{n-1}, v_k\rangle}\big] + \big[\overline{s}\langle v_{n-1}, h_1\rangle\overline{\langle v_{n-1}, v_k\rangle} + \overline{s}\overline{\langle v_{n-1}, h_2\rangle}\\
&\langle v_{n-1}, v_{k-1}\rangle\big] + \big[s\overline{\langle v_{n-1}, h_1\rangle}\langle v_{n-1}, v_k\rangle + s\langle v_{n-1}, h_2\rangle\\
&\overline{\langle v_{n-1}, v_{k-1}\rangle}\big] + \big[s\langle v_{n-1}, v_k\rangle\overline{\langle v_{n-1}, v_{k-1}\rangle} +\\
&\overline{s}\langle v_{n-1}, v_{k-1}\rangle\overline{\langle v_{n-1}, v_k\rangle} + \overline{s}\langle v_{n-1}, h_1\rangle\overline{\langle v_{n-1}, h_2\rangle} +\\
&s\overline{\langle v_{n-1}, h_1\rangle}\langle v_{n-1}, h_2\rangle\big]\Big\}/\|u_k\|^2\\
=&\frac{|\langle v_{n-1}, u_{k-1}\rangle|^2}{\|u_{k-1}\|^2} + \frac{|\langle v_{n-1}, u_k\rangle|^2}{\|u_k\|^2}.\end{aligned}$$

In order to obtain the third equality in last formulas, we adjust the order of terms in each bracket in the second equality.

Hence, we attain equation (a) easily because $\|u_p\|^2 = 1 - \sum\limits_{q=1}^{p-1}\frac{|\langle v_p, u_q\rangle|^2}{\|u_q\|^2}$ holds for any integer p.

(b) can be inferred from (a) and the assumption that the proposition, namely (14), holds for all even number $m \le n$. Actually if denote $t = -\frac{\langle v_m, u_{m-1}\rangle}{\|u_{m-1}\|^2}$, there is

$$\frac{\langle v_{m+1}, u_m\rangle\langle v_m, u_{m-1}\rangle}{\|u_m\|^2\|u_{m-1}\|^2} - \frac{\langle v_{m+1}, u_{m-1}\rangle}{\|u_{m-1}\|^2}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2019.2913047, IEEE Transactions on Automatic Control

$$
\begin{aligned}
&= -t\frac{\langle v_{m+1}, u_m\rangle}{\|u_m\|^2} - \frac{\langle v_{m+1}, u_{m-1}\rangle}{\|u_{m-1}\|^2}\\
&\overset{(a)}{=} \frac{[-t\langle v_{m+1}, u_m\rangle - \langle v_{m+1}, u_{m-1}\rangle(1-|t|^2)]\|u_{m-1}\|^2}{\|u_m\|^2\|u_{m-1}\|^2}\\
&= -\frac{\overline{\langle v_{m+2}, u_m\rangle}}{\|u_m\|^2}.
\end{aligned}
$$

Now we are prepared to prove the proposition when $m = n+2$. We just verify that the coefficient of the term v_1 in u_{n+2} can be determined by the coefficients of the terms v_1, v_2 in u_{n+1}. The remain formulas can be proved similarly. Denote the coefficient of v_1 in u_k as x_k and the coefficient of v_2 in u_k as y_k. Thus

$$
\begin{aligned}
&x_{n+2}\\
&\overset{(13)}{=} -\frac{\langle v_{n+2}, u_{n+1}\rangle}{\|u_{n+1}\|^2}x_{n+1} - \frac{\langle v_{n+2}, u_n\rangle}{\|u_n\|^2}x_n - \cdots\\
&\quad -\frac{\langle v_{n+2}, u_2\rangle}{\|u_2\|^2}x_2 - \frac{\langle v_{n+2}, u_1\rangle}{\|u_1\|^2}x_1\\
&= -\frac{\langle v_{n+2}, u_{n+1}\rangle}{\|u_{n+1}\|^2}x_{n+1} - \frac{\langle v_{n+2}, u_n\rangle}{\|u_n\|^2}\Big[-\frac{\langle v_n, u_{n-1}\rangle}{\|u_{n-1}\|^2}\\
&\quad x_{n-1} + \overline{y_{n-1}}\Big] - \frac{\langle v_{n+2}, u_{n-1}\rangle}{\|u_{n-1}\|^2}x_{n-1} - \cdots\\
&\quad -\frac{\langle v_{n+2}, u_2\rangle}{\|u_2\|^2}\Big[-\frac{\langle v_2, u_1\rangle}{\|u_1\|^2}x_1 + \overline{y_1}\Big] - \frac{\langle v_{n+2}, u_1\rangle}{\|u_1\|^2}x_1\\
&\overset{(b)}{=} -\frac{\langle v_{n+2}, u_{n+1}\rangle}{\|u_{n+1}\|^2}x_{n+1} - \frac{\overline{\langle v_{n+1}, u_n\rangle}}{\|u_n\|^2}x_{n-1} +\\
&\quad \Big[\frac{\overline{\langle v_{n+1}, u_n\rangle}}{\|u_n\|^2}\frac{\overline{\langle v_n, u_{n-1}\rangle}}{\|u_{n-1}\|^2} - \frac{\overline{\langle v_{n+1}, u_{n-1}\rangle}}{\|u_{n-1}\|^2}\Big]\overline{y_{n-1}} + \cdots\\
&\quad -\frac{\overline{\langle v_{n+1}, u_2\rangle}}{\|u_2\|^2}x_1 + \Big[\frac{\overline{\langle v_{n+1}, u_2\rangle}}{\|u_2\|^2}\frac{\overline{\langle v_2, u_1\rangle}}{\|u_1\|^2} -\\
&\quad \frac{\overline{\langle v_{n+1}, u_1\rangle}}{\|u_1\|^2}\Big]\overline{y_1}\\
&= -\frac{\langle v_{n+2}, u_{n+1}\rangle}{\|u_{n+1}\|^2}x_{n+1} - \frac{\overline{\langle v_{n+1}, u_n\rangle}}{\|u_n\|^2}\Big[x_{n-1} -\\
&\quad \frac{\overline{\langle v_n, u_{n-1}\rangle}}{\|u_{n-1}\|^2}\overline{y_{n-1}}\Big] - \frac{\overline{\langle v_{n+1}, u_{n-1}\rangle}}{\|u_{n-1}\|^2}\overline{y_{n-1}}\cdots\\
&\quad -\frac{\overline{\langle v_{n+1}, u_2\rangle}}{\|u_2\|^2}\Big[x_1 - \frac{\overline{\langle v_2, u_1\rangle}}{\|u_1\|^2}\overline{y_1}\Big] - \frac{\overline{\langle v_{n+1}, u_1\rangle}}{\|u_1\|^2}\overline{y_1}\\
&= -\frac{\langle v_{n+2}, u_{n+1}\rangle}{\|u_{n+1}\|^2}x_{n+1} - \frac{\overline{\langle v_{n+1}, u_n\rangle}}{\|u_n\|^2}\overline{y_n} - \frac{\overline{\langle v_{n+1}, u_{n-1}\rangle}}{\|u_{n-1}\|^2}\\
&\quad \overline{y_{n-1}}\cdots - \frac{\overline{\langle v_{n+1}, u_2\rangle}}{\|u_2\|^2}\overline{y_2} - \frac{\overline{\langle v_{n+1}, u_1\rangle}}{\|u_1\|^2}\overline{y_1}\\
&\overset{(13)}{=} -\frac{\langle v_{n+2}, u_{n+1}\rangle}{\|u_{n+1}\|^2}x_{n+1} + \overline{y_{n+1}}.
\end{aligned}
$$

This is exactly the expression for the coefficient of the term v_1 in u_{n+2} by this proposition. The proof is complete. ∎

REFERENCES

[1] H.G. Natke and N. Cottin, *Application of System Identification in Engineering*. Vienna: Springer, 1988.

[2] R.P. Roesser, "A discrete state-space model for image processing," *IEEE Trans. Automat. Contr.*, vol. AC-20, pp. 1-10, 1975.

[3] Y. Zhao, H. Liebgott, and C. Cachard, "Comparison of the existing tool localisation methods on two-dimensional ultrasound images and their tracking results," *IET Control Theory & Applications*, vol. 9, pp. 1124-1134, 2015.

[4] L. Wu, P. Shi, H. Gao, and C. Wang, "H_∞ filtering for 2D Markovian jump systems," *Automatica*, vol. 44, pp. 1849-1858, 2008.

[5] J.J. Shyu, S.C. Pei, Y.D. Huang, and Y.S. Chen, "A new structure and design method for variable fractional-delay 2-D FIR digital filters," *Multidimensional Systems and Signal Processing*, vol. 25, pp. 511-529, 2014.

[6] X. Hong, X. Lai, and R. Zhao, "A fast design algorithm for elliptic-error and phase-error constrained LS 2-D FIR filters," *Multidimensional Systems and Signal Processing*, vol. 27, pp. 477-491, 2016.

[7] R. Yang, L. Xie, and C. Zhang, "H-2 and mixed H-2/H-infinity control of two-dimensional systems in Roesser model," *Automatica*, vol. 42, pp. 1507-1514, 2006.

[8] J. Shi, F. Gao, and T.J. Wu, "Robust design of integrated feedback and iterative learning control of a batch process based on a 2D Roesser system," *Journal of Process Control*, vol. 15, pp. 907-924, 2005.

[9] L. Wang, S. Mo, H. Qu, D. Zhou, and F. Gao, "H_∞ design of 2D controller for batch processes with uncertainties and interval time-varying delays," *Control Engineering Practice*, vol. 20, pp. 1321-1333, 2013.

[10] T. Ding, G. M. Sznaier, and O. Camps, "Robust identification of 2-D periodic systems with applications to texture synthesis and classification," *Proceedings of the IEEE Conference on Decision and Control*, pp. 3678-3683, San Diego, CA, 2006.

[11] C. Chen, J. Tsai and L. Shieh, "Modeling of variable coefficient Roesser's model for systems described by second order partial differnetial equation," *Circuits, Systems, and Signal Processing,* vol. 22, pp. 423-463, 2003.

[12] M. Dillabough, "Discrete state-space models for distributed parameter systems," M.S. thesis, School of Graduate Studies Laurential University, Sudbury, 2007.

[13] K. Galkowski, "Higher roder discretization of 2-D systems," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications,* vol. 47, pp. 713-722, 2000.

[14] Z. Hidayat, A. Nunez, R. Babuska, and B. D. Schutter, "Identification of distributed-parameter systems with missing data," *Proceedings of the IEEE International Conference on Control Applications*, pp. 1014-1019, Dubrovnik, Croatia, 2012.

[15] J. K. Rice and M. Verhaegen, "Efficient system identification of heterogeneous distributed systems via a structure exploiting extended Kalman filter," *IEEE Transactioins on Automatic Control*, vol. 56, pp. 1713-1718, 2011.

[16] J. Schorsch, H. Garnier, M. Gilson and P. C. Young, "Instrumental variable methods for identifying partial differential equation models," *International Journal of Control,* vol. 86, pp. 2325-2335, 2013.

[17] C. Chen and Y. Kao, "Identification of two-dimensional transfer function from finite input-output data," *IEEE Transactions on Automatic Control*, vol. 24, pp. 748-752, 1979.

[18] B. Lashgari, L.M. Silverman, and J. Abramatic, "Approximation of 2-d separable in denominator filters," *IEEE Trans Circuits and Systems*, vol. 30, pp. 107-121, 1983.

[19] J. Ramos, "A subspace algorithm for identifying 2-d separable in denominator filters," *IEEE Transaction on Circuits and Systems*, vol. 41, pp. 63-67, 1994.

[20] J.A. Ramos, A. Alenany, H. Shang, and P.J.L.dos Santos, "Subspace algorithms for identifying separable-in-denominator two-dimensional systems with deterministic inputs," *IET Control Theory & Applications*, vol. 5, pp. 1748-1765, 2011.

[21] D. Wang and A. Zilouchian, "Identification of 2-D Discrete Systems Using Netural Network," *Intelligent Automation and Soft Computing*, vol. 8, pp. 315-324, 2002.

[22] D.K.de Vries and P.M.J. Van den Hof, "Frequency domain identification with generalized orthonomal basis functions," *IEEE Trans. Automat. Control*, vol. 43, pp. 656-669, 1998.

[23] H. Akcay and B. Ninness, "Rational basis functions for robust identification from frequency and time domain measurements," *Automatica*, vol. 34, pp. 1101-1117, 1998.

[24] B. Ninness, "Frequency domain estimation using orhtonormal bases," IFAC Proceedings Volumns, vol. 29(1), pp. 4309-4314, 1996.

[25] B. Ninness and F. Gustafsson, "A unifying construction of orthonormal bases for system identification," *IEEE Trans. Automat. Control*, vol. 42, pp. 515-512, 1997.

[26] P. Mäkilä, "Laguerre methods and H^∞ identification of continuous-time systems," *Internat. J. Control*, vol. 53, pp. 689-707, 1991.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2019.2913047, IEEE Transactions on Automatic Control

[27] B. Wahlberg, "System identification using Kautz models," *IEEE Trans. Automat. Control*, vol. AC-39, pp. 1276-1282, 1994.

[28] W. Mi and T. Qian, "Frequency Domain Identification: An Algorithm Based On Adaptive Rational Orthogonal System," *Automatica*, vol. 48, pp. 1154-1162, 2011.

[29] T. Qian and Y.B. Wang, "Adaptive Fourier series–a variation of greedy algorithm," *Adv. Comput. Math.*, vol. 34, pp. 279–293, 2011.

[30] T. Qian, "Two-dimensional adaptive Fourier decomposition," *Math. Methods Appl. Sci.*, vol. 39, pp. 2431-2448, 2016.

[31] C. Cariou, O. Alata, and J.-M. le Caillec, "Basic elements of 2-D signal processing," in *Two-Dimensional Signal Analysis*, R.Garello, Ed., ISTE & Wiley, 2008, pp. 17-64.

[32] D.R. Brillinger, *Time Series Data Analysis and Theory*. New York: Holt, Reinhart and Winter, 1975, pp. 88-116.

[33] J. Schoukens and J. Renneboog, "Modeling the noise influence on the Fourier coefficients after discrete Fourier transform," *IEEE Trans. Instrum. Meas.*, vol. IM-35, pp. 278-286, 1986.

[34] G. Gu and P.P. Khargonekar, "A class of algorithms for identification in $\mathcal{H}_\infty$," *Automatica*, vol. 28, pp. 299-312, 1992.

[35] H. C. Li, G. T. Deng, and T. Qian, "Hardy space decomposition of on the unit circle: $0 < p < 1$," *Complex Variables and Elliptic Equations: An International Journal*, vol. 61(4), pp. 510-523, 2016.

[36] P. Guillaume, L. Kollár, and R. Pintelon, "Statistical analysis of nonparametric transfer function estimates," *IEEE Transactions on Intrumentation and Measurement*, vol. IM-45, pp. 594-600, 1996.

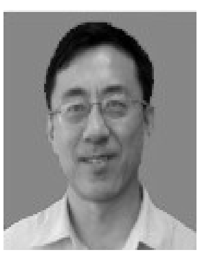
[37] H.M. Valenzuela and A.D. Salvia, "Modeling of Two-Dimensional Systems Using Cumulants," *Proc.ICASSP*, pp. 2913-2916, 1991.

[38] T. Qian, "Cyclic AFD Algorithm for Best Rational approximation," *Mathematical Methods in the Applied Sciences*, vol. 37(6), pp. 846-859, 2014.

[39] W. Mi and T. Qian, "A fast adaptive model reduction method based on Takenaka-Malmquist systems," *Systems & Control Letters*, vol. 61(1), pp. 223-230, 2012.

[40] T. Qian, *Adaptive Fourier Decomposition: A Mathematical Method Through Complex Analysis, Harmonic Analysis and Signal Analysis.* The Chinese Science Press (in Chinese), 2015.

Xiaoyin Wang received the B.S. in Mathematics from China University of Mining and Technology, Jiangsu, China in 2011 and received the MA.Sc. in Mathematics from Wuhan University, Hubei, China in 2013. She is currently pursuing a doctor's degree from University of Macau, Macau. Her main interests are Fourier analysis and complex analysis.



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. His research interests include harmonic analysis in Euclidean spaces, complex and Clifford analysis and signal analysis. Up to now he has published more than 170 papers.

Iengtak Leong received the B.A. from University of Hong Kong in 1987 and Ph.D. from Purdue University in 1994. He has been working as an Associate Professor at University of Macau, Macau, China from 2017. His research interests include wavelets, algebraic geometry and image processing.

You Gao received the B.S. degree from Yantai University in 2010. She received the MA.Sc. degree from South China Normal University in 2013, and the Ph.D. degree from University of Macau in 2017. She started working at Sun Yat-sen University, Zhuhai campus, from 2017 as an associate researcher. Her recent research interests include signal analysis and algorithm design.