Appl. Comput. Harmon. Anal. 55 (2021) 185–198

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Letter to the Editor

A stochastic sparse representation: n-best approximation to random signals and computation [☆]

Wei Qu [a], Tao Qian [b,*], Guan-Tie Deng [a]

[a] *Laboratory of Mathematics and Complex Systems (Ministry of Education), School of Mathematical Sciences, Beijing Normal University, China*
[b] *Macao Center for Mathematical Sciences, Macau University of Science and Technology, Macau*

ARTICLE INFO

Article history:
Received 23 September 2020
Received in revised form 7 May 2021
Accepted 12 May 2021
Available online 17 May 2021
Communicated by Radu Balan

MSC:
42A50
41A30
30B99
60G35
60G07
60G10

Keywords:
Random signal
Stochastic Hilbert space
Stochastic Hardy space
Reproducing kernel
Signal analysis
Image processing

ABSTRACT

In this paper we first prove the existence of the n-best approximation in terms of the parameterized Szegö kernels in the stochastic complex Hardy space of the unit disc. It is a generalization to random signals of the corresponding result for the Hardy space of the disc, and has applications in signal analysis. The result may be generalized to the randomized polydisc case with potential applications in image processing. A practical algorithm for the approximation is proposed.

1. Introduction

In this note we first establish the existence of the n-best Szegö kernel approximation for random periodic signals. Then we extend the result to random signals on torus. The n-best approximation problem was motivated by the classical one of optimal approximation to complex Hardy space functions by rational functions of orders not exceeding n, the latter having been undergoing a long period of studies ([14,2,12]). We first give a quick review on the n-best rational approximation problem. Let n be any positive integer.

☆ Funded by The Science and Technology Development Fund, Macau SAR (File no. 0123/2018/A3).
* Corresponding author.
E-mail addresses: quwei2019@bnu.edu.cn (W. Qu), tqian@must.edu.mo (T. Qian), 96022@bnu.edu.cn (G.-T. Deng).

https://doi.org/10.1016/j.acha.2021.05.003
1063-5203/

A rational function p/q is said to be n-admissible, if p, q are co-prime polynomials, with their degrees both not exceeding n, and q does not have zero inside the unit disc. The following result has been proved by several authors (see [14,1,7]). If $f \in H^2(\mathbf{D})$, the complex Hardy H^2-space of the open unit disc $\mathbf{D}$, then there exists an n-admissible rational function p/q such that

$$\|f - p/q\|_{H^2} \tag{1.1}$$

attains the infimum value over all n-admissible rational functions.

Partial fraction decomposition for rational functions suggests that the n-best rational approximation problem has an alternative formulation, being regarded as the *kernel form*. For any $a \in \mathbf{D}$, set

$$k_a(z) = \frac{1}{1 - \overline{a}z} \quad \text{and} \quad e_a(z) = \frac{\sqrt{1 - |a|^2}}{1 - \overline{a}z}$$

being, respectively, the Szegö kernel and its $H^2(\mathbf{D})$-norm-one normalization of the disc. The inner product of the Hardy space is expressible in terms of the non-tangential boundary limit functions:

$$\langle f, g\rangle = \frac{1}{2\pi}\int_0^{2\pi} f(e^{it})\overline{g}(e^{it})dt.$$

We note that k_a is the reproducing kernel of $H^2(\mathbf{D})$. For any n-tuple $(a_1, \cdots, a_n)$ in $\mathbf{D}^n$, denote by $l(k)$ the multiplicity of a_k in the k-tuple $(a_1, \cdots, a_k), 1 \le k \le n$. With a little abuse of notation, define the $(a_1, \cdots, a_n)$-related multiple kernels to be

$$\tilde{k}_{a_k}(z) = \left[\left(\frac{\partial}{\partial \overline{a}}\right)^{l(k)-1} k_a\right]_{a=a_k}(z),$$

where for $a = s + it$,

$$\frac{\partial}{\partial \overline{a}} = \frac{1}{2}\left(\frac{\partial}{\partial s} + i\frac{\partial}{\partial t}\right).$$

Then the kernel form of the n-best rational approximation is formulated as: For any $f \in H^2(\mathbf{D})$, find an n-tuple $(a_1, \cdots, a_n)$ in $\mathbf{D}^n$, and correspondingly an n-tuple $(c_1, \cdots, c_n)$ in $\mathbf{C}^n$, where $\mathbf{C}$ stands for the complex number field, such that

$$\|f - \sum_{k=1}^n c_k \tilde{k}_{a_k}\|_{H^2} = \inf\{\|f - \sum_{k=1}^n c'_k \tilde{k}_{a'_k}\|_{H^2} \mid (a'_1, \cdots, a'_n) \in \mathbf{D}^n, (c'_1, \cdots, c'_n) \in \mathbf{C}^n\}.$$

We note that once an n-tuple $(a_1, \cdots, a_n)$ is adopted, then the corresponding best coefficients $(c_1, \cdots, c_n)$ may be found through the orthogonal projection of f to the span of $(\tilde{k}_{a_1}, \cdots, \tilde{k}_{a_n})$:

$$\sum_{k=1}^n c_k \tilde{k}_{a_k} = P_{\mathrm{span}\{\tilde{k}_{a_k}:k=1,\cdots,n\}} f.$$

To obtain the projection we use the Gram-Schmidt orthogonalization of the multiple kernels $\tilde{k}_{a_k}, k = 1, \cdots, n$. One may show that the corresponding orthonormal system, apart from modulus-one constants, coincides with the so called n-Takenaka-Malmquist system, or n-TM system, denoted by $\{B_k\}_{k=1}^n$, where

$$\{B_k(z)\}_{k=1}^n = \left\{ \frac{\sqrt{1-|a_k|^2}}{1-\overline{a}_k z} \prod_{j=1}^{k-1} \frac{z-a_j}{1-\overline{a}_j z} \right\}_{k=1}^n . \tag{1.2}$$

In terms of the TM system, the projection can be explicitly written out, as

$$P_{\mathrm{span}\{\tilde{k}_{a_k}:k=1,\cdots,n\}} f = \sum_{k=1}^n \langle f, B_k \rangle B_k.$$

For any n-sequence $(a_1, \cdots, a_n)$ the two collections of functions $\mathcal{K}_n = (\tilde{k}_{a_1}, \cdots, \tilde{k}_{a_n})$ and $\mathcal{B}_n = (B_1, \cdots, B_n)$ span the same n-dimensional function space. Between the two bases there holds

$$\mathcal{K}_n^t = \mathcal{A}_n B_n^t,$$

where $\mathcal{A}_n$ is the transform matrix between the bases, and in particular,

$$\mathcal{A}_n = (\langle \tilde{k}_{a_i}, B_j \rangle)_{i,j=1,\cdots,n},$$

and $\mathcal{K}_n^t, \mathcal{B}_n^t$ are column vectors, being respectively the transposed matrices of $\mathcal{K}_n$ and $\mathcal{B}_n$ ([11]).

A function of the form

$$\sum_{k=1}^n \langle f, B_k \rangle B_k$$

is said to be an *n-Blaschke form* of f. The n-best rational approximation problem may be, in essence, reformulated as follows. Let f be any function in the Hardy space and n be any positive integer. Show that either there exist an integer $m_1, 0 \le m_1 < n$, and correspondingly an m_1-tuple of parameters, $(a_1, \cdots, a_{m_1})$, such that f coincides with its m_1-Blaschke form induced by $(a_1, \cdots, a_{m_1})$; or otherwise, there exists an n-tuple $(a_1, \cdots, a_n)$ such that

$$\|f - \sum_{k=1}^n \langle f, B_k \rangle B_k\|_{H^2}$$

attains the infimum over all possible n-Blaschke form of f. There is an associated algorithm problem to actually find the optimal m_1-, or n-tuple for the attainability of the infimum. This algorithm problem is so far still open. The existing literatures can only claim finding local minimum values ([1,2,12,7,13]).

The above is the formulation for one-dimensional deterministic signals. The present study is to raise and solve a similar problem for random signals. The motivation of studying random signals is from two sides. One is that a practical signal data obtained through various methods is usually corrupted with noises or errors. The other is that a signal under study is maybe just one of some several types of signals that altogether obey certain distribution law. Both of these two types are random signals that come from practice. A random signal theory in relation to sparse representation is developed in the recent paper [9]. In the present paper the stochastic n-best approximations are developed. The one and several complex variables theory are, respectively, given in §2 and §3. In §4 we introduce an algorithm called by random cyclic AFD that does not theoretically guarantee the global minimum but practical in applications. For applications of n-best approximation in system identification we refer the reader to [5,6].

In this note we will concentrate in the disc and the polydisc contexts. The disc context corresponds to one-dimensional periodic signals. The polydisc context corresponds to images. We will introduce the *randomized disc* and *randomized polydisc* in the following sections. For random signals in the whole time or whole space range there exist essentially the same theories with the one or several complex variables, or the Clifford number variables settings.

2. n-best approximation in the randomized unit disc

In the beginning of this section we introduce the concept of random signal in the unit disc case.

Let $F(t,\omega)$ be a random signal, $t \in T$, where T is a set in the time or the space range, and $\omega \in \Omega$ is an arbitrary but fixed probability space. Here we assume that for each fixed $t \in T$, the function $F(t,\cdot)$ is a random variable in Ω with the probability measure $d\mu$; and for each fixed ω in Ω, $F(\cdot,\omega)$ is a real-valued function in $L^2(T)$. We will be considering the case $T = [0, 2\pi)$. In order to use complex analytic methods we rewrite the above function notation: $F(t,\omega) = f(e^{it},\omega), e^{it} \in \partial\mathbf{D}$. A natural space for such random signals is

$$L^2_\omega(\partial\mathbf{D}) \triangleq \{f(e^{it},\omega) \mid \forall\omega \in \Omega : f_\omega \triangleq f(\cdot,\omega) \in L^2(\partial\mathbf{D});$$
$$\forall t \in [0,2\pi),\ f(e^{it},\cdot) \text{ is a random variable and } E_\omega\|f_\omega\|^2_{L^2(\partial\mathbf{D})} < \infty\},$$

where E_ω denotes the probability expectation. The quantity $E_\omega\|f_\omega\|^2_{L^2(\partial\mathbf{D})}$ is also denoted by $\|f\|^2_{\mathcal{N}}$. The space $L^2_\omega(\partial\mathbf{D})$ is then denoted as $\mathcal{N}(\partial\mathbf{D})$, or briefly $\mathcal{N}$. The random signals in $\mathcal{N}$ are regarded as *random signals with finite energy* or *normal random signals*. It is standard knowledge of measure theory and Hilbert space theory that such formulated space $\mathcal{N}$ is a Hilbert space with the inner product $\langle f, g\rangle_{\mathcal{N}} = E_\omega\langle f_\omega, g_\omega\rangle_{L^2(\partial\mathbf{D})}$. A normal random signal has a Fourier series expansion:

$$f(e^{it},\omega) = \sum_{-\infty}^{\infty} c_k(\omega)e^{ikt}, \quad c_k(\omega) = \frac{1}{2\pi}\int_0^{2\pi} f(e^{it},\omega)dt.$$

The Plancherel Theorem implies that $E_\omega\|f_\omega\|^2 = \sum_{k=-\infty}^{\infty} E_\omega|c_k(\omega)|^2 < \infty$.

To make use of analytic function methodology we consider the stochastic Hardy space as a closed subspace of the space $\mathcal{N}$ on the boundary $\partial\mathbf{D}$:

$$H^2_\omega(\partial\mathbf{D}) = \{f(e^{it},\omega) = \sum_{k=0}^{\infty} c_k(\omega)e^{ikt} \mid \sum_{k=0}^{\infty} E_\omega|c_k(\omega)|^2 < \infty\}.$$

The closedness of $H^2_\omega(\partial\mathbf{D})$ in $L^2_\omega(\partial\mathbf{D})$ may be proved through the Fourier series expansions of the functions in these spaces and by invoking the Plancherel Theorem. This space coincides with one consisting of the non-tangential boundary limits of the analytic functions in the space

$$H^2_\omega(\mathbf{D}) = \{f(z,\omega) = \sum_{k=0}^{\infty} c_k(\omega)z^k \mid \sum_{k=0}^{\infty} E_\omega|c_k(\omega)|^2 < \infty\}$$

(see [3]). The last two spaces are isometric. We accordingly define *random n-Blaschke forms* to be the functions in $H^2_\omega(\mathbf{D})$ of the form

$$\sum_{k=1}^{n} c_k(\omega)B_k(z),$$

which is associated with an n-tuple of parameters $(a_1,\cdots,a_n)$ determining the n-TM system $\{B_k\}_{k=1}^n$. For any $f \in \mathcal{N}$, the forms

$$\sum_{k=1}^{n}\langle f_\omega, B_k\rangle B_k(z)$$

are called the *random n-Blaschke forms* of f. If for some positive integer m there holds

$$f = \sum_{k=1}^{m} \langle f_\omega, B_k \rangle B_k(z),$$

where $\{B_k\}_{k=1}^m$ is the TM system corresponding to an m-sequence of the parameters $\{a_1, \cdots, a_m\}$ in $\mathbf{D}$, then we say that f *is by itself a random m-Blaschke form*. We note that in the above definitions of random Blaschke forms the involved parameters $a_k \in \mathbf{D}$ defining the TM systems are all constants and not random variables.

We note that For real-valued random signals $f \in \mathcal{N}$ there holds $c_{-n}(\omega) = \overline{c}_n(\omega)$. As a consequence, between a real-valued random signal $f \in \mathcal{N}$ and its stochastic Hardy space projection $\{f_\omega\}^+ \in H^2_\omega(\mathbf{D})$, there holds the relation

$$f_\omega = 2\mathrm{Re}\{f_\omega\}^+ - c_0(\omega).$$

In such a way, harmonic analysis of a function f in $\mathcal{N}(\partial\mathbf{D})$ is reduced to complex analysis of the associated f^+ in $H^2_\omega(\mathbf{D})$. This paper is to study the theme of best approximation to f^+ by n-Blaschke form. After obtaining a solution for f^+, by taking the real part, we obtain the related n-best approximation for the original real-valued function f.

For a general function in $\mathcal{N}$, not necessarily real-valued, one can alternatively project it, through applying the Hilbert transform for instance, into the stochastic inner- and stochastic outer-Hardy spaces and correspondingly gets the decomposition $f = f^+ + f^-$, and to each of the two components performs the approximation, and then puts together. The approximation theories of the stochastic inner- and outer-Hardy spaces are parallel. The non-stochastic $L^2(\partial\mathbf{D})$ case along the same line was studied in the early literature (see, for instance, [10,7]).

For one-dimensional random periodic signals we will prove the following stochastic n-best Blaschke form approximation result.

Theorem 2.1. *Let $f \in H^2_\omega(\mathbf{D})$ be a non-zero random signal. For any positive integer n, there must hold one of the following two cases: (i) For some $1 \le m_1 \le n$, there exists an m_1-tuple of constant parameters $(a_1, \cdots, a_{m_1}) \in \mathbf{D}^{m_1}$ such that f is precisely expressible by*

$$f(z,\omega) = \sum_{k=1}^{m_1} \langle f_\omega, B_k \rangle B_k(z), \tag{2.3}$$

where $\{B_k\}_{k=1}^{m_1}$ is the TM system generated by $(a_1, \cdots, a_{m_1})$; or (ii) There exists an n-tuple of constant parameters $(a_1, \cdots, a_n) \in \mathbf{D}^n$ such that

$$\|f - \sum_{k=1}^{n} \langle f_\omega, B_k \rangle_{L^2(\partial\mathbf{D})} B_k\|_{\mathcal{N}} \tag{2.4}$$

attains its positive infimum over all possible random n-Blaschke forms of f.

The proof of Theorem 2.1 for deterministic signals given in [7] may be adapted to give a proof for the present stochastic case. Both the proofs rely on the following result proved in the literature ([12,9]).

Theorem 2.2. *Let $(a_1, \cdots, a_n)$ be an n-tuple of complex numbers in $\mathbf{D}^n$. Apart from unimodular constants, the Gram-Schmidt orthonormalization of $(\tilde{k}_{a_1}, \cdots, \tilde{k}_{a_n})$ is the corresponding TM system $\{B_k\}_{k=1}^n$.*

Proof of Theorem 2.1. Denote by d the infimum value of (2.4). Let $\mathbf{s}^{(l)} = (a_1^{(l)}, \cdots, a_n^{(l)})$ be a sequence of n-tuples that leads to the infimum value d along with $l \to \infty$, that is

$$\lim_{l\to\infty} \|f - \sum_{k=1}^{n} \langle f_\omega, B_k^{(l)}\rangle_{L^2(\partial\mathbf{D})} B_k^{(l)}\|_{\mathcal{N}} = d \geq 0. \tag{2.5}$$

Since $\overline{\mathbf{D}}^n$ is compact in $\mathbf{C}^n$, there exists a subsequence of $\mathbf{s}^{(l)}$ having a limit $\mathbf{s} = (a_1, \cdots, a_n)$ in $\overline{\mathbf{D}}^n$. Without loss of generality we may assume that the sequence $\mathbf{s}^{(l)}$ itself has this limit. The complex numbers $a_k, k = 1, \cdots, n$, may be divided into two groups $\mathcal{I}$ and $\mathcal{B}$: $\mathcal{I}$ is the set of the limit points a_k inside the disc; and $\mathcal{B}$ is the set of the limit points $a_k, k = 1, \cdots, n$, on the boundary of the disc. If $\mathcal{B}$ is empty, then the limiting points are all in $\mathbf{D}$. Based on continuity of inner product there holds

$$d = \|f_\omega - \sum_{k=1}^{n} \langle f_\omega, B_k\rangle_{L^2(\partial\mathbf{D})} B_k\|_{\mathcal{N}},$$

where $\{B_k\}_{k=1}^n$ is associated with $(a_1, \cdots, a_n)$. If $d = 0$ we obtain $m_1 = n$ in the case (i), and if $d > 0$, we obtain the case (ii) described by (2.4). Now we assume $\mathcal{B}$ is non-empty. We show that in such case there must hold $d = 0$, and we obtain what is described in the case (i) for $1 \leq m_1 < n$, and no other possibilities. This will conclude the theorem. Since $\sum_{k=1}^{\infty} \langle f_\omega, B_k^{(l)}\rangle_{L^2(\partial\mathbf{D})} B_k^{(l)}$ is the projection of f into the span$\{\tilde{k}_{a_k^{(l)}} : k = 1, \cdots, n\}$, the latter being irrelevant with the order of the $a_k^{(l)} : k = 1, \cdots, n$, we can re-arrange the order of the indices k so that all the indices in $\mathcal{I}$ are smaller than those in $\mathcal{B}$. Let $\mathcal{I} = \{1, \cdots, m_1\}, m_1 < n$, and $\mathcal{B} = \{m_1 + 1, \cdots, n\} \neq \emptyset$. We will first show that

$$\lim_{l\to\infty} \|\sum_{k\in\mathcal{B}} \langle f_\omega, B_k^{(l)}\rangle_{L^2(\partial\mathbf{D})} B_k^{(l)}\|_{\mathcal{N}} = 0.$$

Define

$$g_j^{(l)} \triangleq f_\omega - \sum_{k=1}^{j-1} \langle f_\omega, B_k^{(l)}\rangle_{L^2(\partial\mathbf{D})} B_k^{(l)} = f_\omega - \sum_{k=1}^{j-1} \langle (g_k^{(l)})_\omega, B_k^{(l)}\rangle_{L^2(\partial\mathbf{D})} B_k^{(l)} \quad (g_1^{(l)} = f_\omega)$$

to be the standard orthogonal remainders, where $j = 1, \cdots, n$. There follows that the energy terms

$$\|g_j^{(l)}\|_{\mathcal{N}}^2 = \|f\|_{\mathcal{N}}^2 - \sum_{k=1}^{j-1} E_\omega |\langle (g_k^{(l)})_\omega, B_k^{(l)}\rangle|^2$$

form a decreasing sequence along with j increasing. Denote

$$R_{\mathcal{I}}^{(l)} = f_\omega - \sum_{k\in\mathcal{I}} \langle (g_k^{(l)})_\omega, B_k^{(l)}\rangle_{L^2(\partial\mathbf{D})} B_k^{(l)}.$$

We have, in particular, $R_{\mathcal{I}}^{(l)} = g_{m_1+1}^{(l)}$. Hence,

$$\|R_{\mathcal{I}}^{(l)}\|_{\mathcal{N}}^2 = \|f\|_{\mathcal{N}}^2 - \sum_{k\in\mathcal{I}} E_\omega |\langle (g_k^{(l)})_\omega, B_k^{(l)}\rangle|^2 \geq \|g_j^{(l)}\|_{\mathcal{N}}^2, \quad \text{for } j \geq m_1 + 1. \tag{2.6}$$

Let P_r be the Poisson kernel. For any fixed $r < 1$ the convolution operator $P_r * f$ is contractive in $\mathcal{N}$. In fact,

$$\|P_r * f\|_{\mathcal{N}}^2 = E_\omega \|P_r * f_\omega\|_{L^2(\partial\mathbf{D})}^2 \leq E_\omega \|f_\omega\|_{L^2(\partial\mathbf{D})}^2 = \|f\|_{\mathcal{N}}.$$

As in $L^2(\partial\mathbf{D})$, the convolution operator P_r*, as $r \to 1-$, is an approximation to the identity in $\mathcal{N}$. It comes from a Lebesgue dominated convergence argument as follows. For any random signal g in $\mathcal{N}$, and any fixed $\omega \in \Omega$,

$$\lim_{r\to 1-} \|g_\omega - P_r * g_\omega\|^2_{L^2(\partial\mathbf{D})} = 0;$$

and

$$\|g_\omega - P_r * g_\omega\|^2_{L^2(\partial\mathbf{D})} \leq 4\|g_\omega\|^2_{L^2(\partial\mathbf{D})} \in L^1(\Omega, d\mu).$$

Then the Lebesgue Dominated Convergence Theorem concludes

$$\lim_{r\to 1-} \|g - P_r * g\|_{\mathcal{N}} = 0. \tag{2.7}$$

The last relation amounts that P_r* is an approximation to the identity in $\mathcal{N}$.

The following inequality chain uses orthogonality of the projections and the remainders, the contraction property of the Poisson integral operator in $\mathcal{N}$, the triangle inequality, the Cauchy-Schwarz inequality together with the norm-one property of $B_k^{(l)}$, and finally the relation (2.6):

$$\begin{aligned}\|R_{\mathcal{I}}^{(l)}\|_{\mathcal{N}} &\geq \|R_{\mathcal{I}}^{(l)} - \sum_{k\in\mathcal{B}} \langle (g_k^{(l)})_\omega, B_k^{(l)}\rangle_{L^2(\partial\mathbf{D})} B_k^{(l)}\|_{\mathcal{N}} \\ &\geq \|P_r * (R_{\mathcal{I}}^{(l)} - \sum_{k\in\mathcal{B}} \langle (g_k^{(l)})_\omega, B_k^{(l)}\rangle_{L^2(\partial\mathbf{D})} B_k^{(l)})\|_{\mathcal{N}} \\ &\geq \|P_r * R_{\mathcal{I}}^{(l)}\|_{\mathcal{N}} - \sum_{k\in\mathcal{B}} \|g_k^{(l)}\|_{\mathcal{N}} \|P_r * B_k^{(l)}\|_{\mathcal{N}} \\ &\geq \|P_r * R_{\mathcal{I}}^{(l)}\|_{\mathcal{N}} - \|R_{\mathcal{I}}^{(l)}\|_{\mathcal{N}} \sum_{k\in\mathcal{B}} \|P_r * B_k^{(l)}\|_{\mathcal{N}}. \end{aligned} \tag{2.8}$$

Continuity of the inner product justifies

$$\lim_{l\to\infty} R_{\mathcal{I}}^{(l)} = R_{\mathcal{I}}, \qquad \text{in } \mathcal{N},$$

where, by definition,

$$R_{\mathcal{I}} \triangleq f_\omega - \sum_{k\in\mathcal{I}} \langle f_\omega, B_k\rangle_{L^2(\partial\mathbf{D})} B_k,$$

where $(B_1, \cdots, B_{m_1})$ is the TM system corresponding to $(a_1, \cdots, a_{m_1})$ with the cited order.

Note that the above formulation allows $\mathcal{I} = \emptyset$. Since P_r* is an approximation to the identity, for any given $\epsilon \in (0,1)$, one can choose $r_1 < 1$ sufficiently close to 1 such that

$$\|P_{r_1} * R_{\mathcal{I}}\|_{\mathcal{N}} > (1 - \epsilon/2)\|R_{\mathcal{I}}\|_{\mathcal{N}}.$$

Now turn to the entries of summation (2.8). Since the convolution operator $P_{r_1}*$ is a continuous operator, for the fixed r_1, for large enough l there also holds

$$\|P_{r_1} * R_{\mathcal{I}}^{(l)}\|_{\mathcal{N}} > (1 - \epsilon/2)\|R_{\mathcal{I}}^{(l)}\|_{\mathcal{N}}. \tag{2.9}$$

In the following argument let r_1 be fixed. For each $k \in \mathcal{B}$, the convolution operator $P_{r_1}*$ applied to the non-tangential boundary limit of $B_k^{(l)}$ reproduces the analytic function $B_k^{(l)}(r_1 e^{it})$ ([3], Ch. 2, Section 3):

$$(P_{r_1} * B_k^{(l)})(e^{it}) = B_k^{(l)}(r_1 e^{it}).$$

We have, for each $k \in \mathcal{B}$,

$$\begin{aligned}
\|P_{r_1} * B_k^{(l)}\|^2 &= \frac{1}{2\pi}\int_0^{2\pi} |B_k^{(l)}(r_1 e^{it})|^2 dt \\
&\le \frac{1}{2\pi}\int_0^{2\pi} |e_{a_k^{(l)}}(r_1 e^{it})|^2 dt \\
&= \frac{1}{2\pi}\frac{\sqrt{1-|a_k^{(l)}|^2}}{1-|r_1 a_k^{(l)}|^2}\int_0^{2\pi} \frac{1-|r_1 a_k^{(l)}|^2}{|1-r_1\overline{a}_k^{(l)} e^{-it}|^2} dt \\
&= \frac{1}{2\pi}\frac{\sqrt{1-|a_k^{(l)}|^2}}{1-|r_1 a_k^{(l)}|^2} \\
&\to 0, \qquad \text{as} \quad |a_k^{(l)}| \to 1.
\end{aligned} \tag{2.10}$$

We note that the above second inequality relation is a crucial step, that is based on the fact that values of Blaschke products inside the unit disc are dominated by 1, so through introducing a dominating term the roles of the other parameters $a_j^{(l)}, j \neq k$, do not show in the final estimation:

$$|B_k^{(l)}(r_1 e^{it})| = \left| \frac{\sqrt{1-|a_k^{(l)}|^2}}{1-\overline{a}_k^{(l)} r_1 e^{it}} \prod_{j=1}^{k-1} \frac{r_1 e^{it} - a_l^{(l)}}{1-\overline{a}_l^{(l)} r_1 e^{it}} \right| \le \left| \frac{\sqrt{1-|a_k^{(l)}|^2}}{1-\overline{a}_k^{(l)} r_1 e^{it}} \right| = |e_{a_k^{(l)}}(r_1 e^{it})|.$$

As a consequence of (2.10), for l large enough there holds

$$\|R_{\mathcal{I}}^{(l)}\|_{\mathcal{N}} \sum_{k\in\mathcal{B}} \|P_{r_1} * B_k^{(l)}\|_{\mathcal{N}} \le \frac{\epsilon}{2}\|R_{\mathcal{I}}^{(l)}\|_{\mathcal{N}}.$$

The above estimation together with (2.9) and (2.8) gives, for the large enough l,

$$\|R_{\mathcal{I}}^{(l)} - \sum_{k\in\mathcal{B}} \langle (g_k^{(l)})_\omega, B_k^{(l)}\rangle_{L^2(\partial\mathbf{D})} B_k^{(l)}\|_{\mathcal{N}} \ge (1-\epsilon)\|R_{\mathcal{I}}^{(l)}\|.$$

Hence,

$$\begin{aligned}
\|R_{\mathcal{I}}^{(l)} - \sum_{k\in\mathcal{B}} \langle (g_k^{(l)})_\omega, B_k^{(l)}\rangle_{L^2(\partial\mathbf{D})} B_k^{(l)}\|_{\mathcal{N}}^2 &= \|R_{\mathcal{I}}^{(l)}\|^2 - \|\sum_{k\in\mathcal{B}} \langle f_\omega, B_k^{(l)}\rangle_{L^2(\partial\mathbf{D})} B_k^{(l)}\|_{\mathcal{N}}^2 \\
&> (1-\epsilon)^2\|R_{\mathcal{I}}^{(l)}\|_{\mathcal{N}}^2.
\end{aligned}$$

By taking limit $l \to \infty$ this shows

$$\overline{\lim}_{l\to\infty} \|\sum_{k\in\mathcal{B}} \langle f_\omega, B_k^{(l)}\rangle_{L^2(\partial\mathbf{D})} B_k^{(l)}\|_{\mathcal{N}} \le \sqrt{2\epsilon - \epsilon^2}\|f\|_{\mathcal{N}}. \tag{2.11}$$

Since ϵ is arbitrary small and positive, the above limit is actually 0. We hence have

$$d = \|f - \sum_{k\in\mathcal{I}} \langle f, B_k\rangle_{L^2(\partial\mathbf{D})} B_k\|_{\mathcal{N}} = \|R_{\mathcal{I}}\|_{\mathcal{N}}.$$

Next, we divide into two cases: $d = 0$ and $d > 0$. The $d = 0$ case shows that f itself is a random m_1-Blaschke form where $m_1 < n$. For the $d > 0$ case, since $R_{\mathcal{I}} \neq 0$ and $\{k_a\}_{a\in\mathbf{D}}$ is dense in $\mathcal{N}$, one can always find an $(n - m_1)$-tuple $(b_{m_1+1}, \cdots, b_n) \in \mathbf{D}^{n-m_1}$, consisting of distinguished elements, such that

$$P_{\text{span}\{\tilde{k}_{a_1},\cdots,\tilde{k}_{a_{m_1}},k_{b_{m_1+1}},\cdots,k_{b_n}\}\ominus\text{span}\{\tilde{k}_{a_1},\cdots,\tilde{k}_{a_{m_1}}\}}(R_{\mathcal{I}}) > 0.$$

This immediately gives a random n-Blaschke form

$$\sum_{k=1}^{n}\langle f_\omega, \tilde{B}_k\rangle_{L^2(\partial\mathbf{D})}\tilde{B}_k,$$

where $(\tilde{B}_1, \cdots, \tilde{B}_n)$ is the TM system corresponding to $(a_1, \cdots, a_{m_1}, \tilde{b}_{m_1+1}, \cdots, \tilde{b}_n)$, possessing the property

$$\|f - \sum_{k=1}^{n}\langle f, \tilde{B}_k\rangle_{L^2(\partial\mathbf{D})}\tilde{B}_k\|_{\mathcal{N}} = d' < d.$$

This is contrary to the previously assumed infimum property of d in (2.5). The proof is complete. □

3. n-best approximation in the randomized polydisc

We will write out the detailed theory for the 2-torus case. The general m-torus theory is similar. In image processing one deals with functions on the square $\{(t, s) \mid t, s \in [0, 2\pi)\}$. By making the correspondence $F(t, s) = f(e^{it}, e^{is})$, we, instead, work on functions defined on the torus $\mathbf{T}^2 = \{(e^{it}, e^{is}) \mid t, s \in [0, 2\pi)\}$. In the deterministic case we work with the Hardy space with two complex variables:

$$H^2(\mathbf{D}^2) = \{f : \mathbf{D}^2 \to \mathbf{C} \mid f(z, w) = \sum_{k,l\geq 0} c_{kl}z^k w^l, \sum_{k,l\geq 0}|c_{kl}|^2 < \infty\}.$$

Recall that

$$L^2(\mathbf{T}^2) = \{f : \mathbf{T}^2 \to \mathbf{C} \mid f(e^{it}, e^{is}) = \sum_{-\infty<k,l<\infty} c_{kl}e^{ikt}e^{ils}, \sum_{-\infty<k,l<\infty}|c_{kl}|^2 < \infty\}.$$

$\mathbf{T}^2$ is called the *characteristic boundary* of $\mathbf{D}^2$. Like in the unit circle case $L^2(\mathbf{T}^2)$ is divided into a " nearly direct sum" of four spaces of which each consists of the radial boundary limits of a space like the Hardy space $H^2(\mathbf{D}^2)$. More details of the formulation together with AFD algorithm and its variations, as well as some applications, can be found in [8,4,15].

The stochastic-lization of $H^2(\mathbf{D}^2)$ is defined as

$$H^2_\omega(\mathbf{D}^2) = \{f : \mathbf{D}^2 \to \mathbf{C} \mid f(\omega; z, w) = \sum_{k,l\geq 0} c_{kl}(\omega)z^k w^l, \sum_{k,l\geq 0} E_\omega|c_{kl}(\omega)|^2 < \infty\},$$

where E_ω stands for, again, the expectation with respect to a general probability space $(\Omega, d\mu)$. The Plancherel Theorem of $H^2_\omega(\mathbf{D}^2)$ asserts

$$E_\omega\|f_\omega\|_{L^2(\mathbf{T}^2)} = \sum_{k,l\geq 0} E_\omega|c_{kl}(\omega)|^2,$$

where $f_\omega(e^{it}, e^{is}) = f(\omega; e^{it}, e^{is})$, and the space $L^2(\mathbf{T}^2)$ is equipped with the usual inner product

$$\langle u, v\rangle_{L^2(\mathbf{T}^2)} = \frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^{2\pi} u(e^{it}, e^{is})\overline{v}(e^{it}, e^{is})dtds.$$

The Fourier coefficients of $f \in H^2_\omega(\mathbf{D}^2)$ are given by

$$c_{kl}(\omega) = \frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^{2\pi} f_\omega(e^{it}, e^{is})e^{-ikt}e^{-ils}dtds, \quad k,\ l \geq 0.$$

It is noted that $H^2_\omega(\mathbf{D}^2)$ is a Hilbert space, being identical with $H^2_\omega(\mathbf{T}^2)$, consisting of the radial boundary limits of those in the former. The two spaces are, in fact, isometric under the natural correspondence between an analytic function and its boundary limit. Like in the one complex variable case we are interested in the optimization problem: Let n_1, n_2 be two positive integers. Find $\mathbf{a} = (a_1, \cdots, a_{n_1}) \in \mathbf{D}^{n_1}$ and $\mathbf{b} = (b_1, \cdots, b_{n_2}) \in \mathbf{D}^{n_2}$ such that

$$\|f - \sum_{1\leq k\leq n_1, 1\leq l\leq n_2} \langle f_\omega, B^{\mathbf{a}}_k \otimes B^{\mathbf{b}}_l\rangle_{L^2(\mathbf{T}^2)} B^{\mathbf{a}}_k \otimes B^{\mathbf{b}}_l\|_{H^2_\omega(\mathbf{D}^2)} \tag{3.12}$$

is minimized, where $\{B^{\mathbf{a}}_k\}_{k=1}^{n_1}$ is the n_1-TM system generated by $a_1, \cdots, a_{n_1}$; and $\{B^{\mathbf{b}}_l\}_{l=1}^{n_2}$ is the n_2-TM system generated by $b_1, \cdots, b_{n_2}$, and $(B^{\mathbf{a}}_k \otimes B^{\mathbf{b}}_l)(z, w) = B^{\mathbf{a}}_k(z)B^{\mathbf{b}}_l(w), 1 \leq k \leq n_1, 1 \leq l \leq n_2$, are tensor products. We note that the system $\{B^{\mathbf{a}}_k \otimes B^{\mathbf{b}}_l\}_{1\leq k\leq n_1, 1\leq l\leq n_2}$ for any $a_1, \cdots, a_{n_1}$ and $b_1, \cdots, b_{n_2}$ is an orthonormal system in $H^2(\mathbf{D}^2)$, as well as an orthonormal system in $H^2_\omega(\mathbf{D}^2)$. The function

$$\sum_{1\leq k\leq n_1, 1\leq l\leq n_2} \langle f_\omega, B^{\mathbf{a}}_k \otimes B^{\mathbf{b}}_l\rangle_{L^2(\mathbf{T}^2)} B^{\mathbf{a}}_k \otimes B^{\mathbf{b}}_l$$

is said to be the *random* $(n_1 \times n_2)$*-Blaschke form of* f *generated by* $(a_1, \cdots, a_{n_1})$ and $(b_1, \cdots, b_{n_2})$. A function of the form

$$\sum_{1\leq k\leq n_1, 1\leq l\leq n_2} c_{kl}(\omega) B^{\mathbf{a}}_k \otimes B^{\mathbf{b}}_l$$

is a general *random* $(n_1 \times n_2)$*-Blaschke form.* We have the following theorem.

Theorem 3.1. *Let $f \in H^2_\omega(\mathbf{D}^2)$ be a non-zero random signal. Let n_1 and n_2 be any two positive integers. Then one of the following two cases must happen:* (i) *f itself is a random $(m_1 \times m_2)$-Blaschke form for some $m_1 \leq n_1$ and $m_2 \leq n_2$; or* (ii) *there exists an n_1-tuple $(a_1, \cdots, a_{n_1})$ and an n_2-tuple $(b_1, \cdots, b_{n_2})$ such that*

$$\|f - \sum_{1\leq k\leq n_1, 1\leq l\leq n_2} \langle f_\omega, B^{\mathbf{a}}_k \otimes B^{\mathbf{b}}_l\rangle_{L^2(\mathbf{T}^2)} B^{\mathbf{a}}_k \otimes B^{\mathbf{b}}_l\|_{H^2_\omega(\mathbf{D}^2)} \tag{3.13}$$

attains its positive infimum over all possible $(n_1 \times n_2)$-Blaschke forms.

To the best knowledge of the authors even for the deterministic case such multivariate result has not been aware in the literature. Theorem 3.1 asserts the result directly in the randomized polydisc case. Theorem 3.1 generalizes immediately to $H^2_\omega(\mathbf{D}^m), m > 2$.

Proof of Theorem 3.1. Let $f \in H^2_\omega(\mathbf{D}^2)$ be the given non-zero function. By using a similar argument as in the unit disc case we may assume without loss of generality that $\lim_{j\to\infty} a^{(j)}_k = a_k, \lim_{j\to\infty} b^{(j)}_l = b_l, k = 1, \cdots, n_1, l = 1, \cdots, n_2$, such that

$$\lim_{j\to\infty}\|f-\sum_{1\le k\le n_1,1\le l\le n_2}\langle f_\omega, B_k^{\mathbf{a}^j}\otimes B_l^{\mathbf{b}^j}\rangle_{L^2(\mathbf{T}^2)}B_k^{\mathbf{a}^j}\otimes B_l^{\mathbf{b}^j}\|_{H^2_\omega(\mathbf{D}^2)}=d,$$

where d is the targeted infimum. We can re-arrange the order of the indices k and l, and find a pair of indices k_0 and l_0 such that if and only for $k\le k_0$ the limit a_k is an interior point of $\mathbf{D}$; and if and only for $l\le l_0$ the limit b_l is an interior point of $\mathbf{D}$. Note that k_0 or l_0, or both of them, can be zero. We accordingly form

$$\mathcal{R}^{(j)}_{k_0+1,l_0+1}=f-\sum_{1\le k\le k_0;1\le l\le l_0}\langle f_\omega, B_k^{\mathbf{a}^j}\otimes B_l^{\mathbf{b}^j}\rangle_{L^2(\mathbf{T}^2)}B_k^{\mathbf{a}^j}\otimes B_l^{\mathbf{b}^j},$$

the (k_0+1,l_0+1)-standard orthogonal remainder, and

$$g^{(j)}_{k',l'}=f-\sum_{1\le k<k';1\le l<l'}\langle f_\omega, B_k^{\mathbf{a}^j}\otimes B_l^{\mathbf{b}^j}\rangle_{L^2(\mathbf{T}^2)}B_k^{\mathbf{a}^j}\otimes B_l^{\mathbf{b}^j},$$

the (k',l')-standard orthogonal remainders, in general.

Due to continuity of the inner product, we have

$$\lim_{j\to\infty}\mathcal{R}^{(j)}_{k_0+1,l_0+1}=f-\sum_{1\le k\le k_0;1\le l\le l_0}\langle f_\omega, B_k^{\mathbf{a}}\otimes B_l^{\mathbf{b}}\rangle_{L^2(\mathbf{T}^2)}B_k^{\mathbf{a}}\otimes B_l^{\mathbf{b}}\triangleq\mathcal{R}_{k_0+1,l_0+1},$$

where $\{B_k^{\mathbf{a}}\otimes B_l^{\mathbf{b}}\}_{1\le k\le k_0,1\le l\le l_0}$ is the tensor product system of the two TM systems generated by $a_1,\cdots,a_{k_0}$ and $b_1,\cdots,b_{j_0}$, respectively. At this point, if $k_0=n_1$ and $j_0=n_2$, then one can discuss two cases: $d=0$ corresponds to the $m_1=n_1$ and $m_2=n_2$ case within the case (i)) in the statement of the theorem; and $d>0$ corresponds to the stated case (ii) of the theorem. In the following we discuss the otherwise cases: $k_0<n_1$ or $j_0<n_2$, or both. We are to show that in such case for any $\epsilon\in(0,1)$ and large enough j there holds

$$\begin{aligned}&\|\mathcal{R}^{(j)}_{k_0+1,l_0+1}\|_{H^2_\omega(\mathbf{D}^2)}\\ &\ge\|\mathcal{R}^{(j)}_{k_0+1,l_0+1}-\sum_{k'>k_0\ \text{or}\ l'>l_0}\langle g^{(j)}_{k',l'}, B_{k'}^{\mathbf{a}^j}\otimes B_{l'}^{\mathbf{b}^j}\rangle_{L^2(\mathbf{T}^2)}B_{k'}^{\mathbf{a}^j}\otimes B_{l'}^{\mathbf{b}^j}\|_{H^2_\omega(\mathbf{D}^2)}\\ &\ge(1-\epsilon)\|\mathcal{R}^{(j)}_{k_0+1,l_0+1}\|_{H^2_\omega(\mathbf{D}^2)}.\end{aligned}\tag{3.14}$$

Temporarily accepting (3.14), we have that, due to the orthogonality, the sum corresponding to " $k'>k_0$ or $l'>l_0$" has no contribution, that is,

$$\lim_{j\to\infty}\|\sum_{k'>k_0\ \text{or}\ l'>l_0}\langle g^{(j)}_{k',l'}, B_{k'}^{\mathbf{a}^j}\otimes B_{l'}^{\mathbf{b}^j}\rangle_{L^2(\mathbf{T}^2)}B_{k'}^{\mathbf{a}^j}\otimes B_{l'}^{\mathbf{b}^j}\|_{H^2_\omega(\mathbf{D}^2)}=0,\tag{3.15}$$

and, as a consequence,

$$d=\lim_{j\to\infty}\|f-\sum_{1\le k\le k_0;1\le l\le l_0}\langle f_\omega, B_k^{\mathbf{a}^j}\otimes B_l^{\mathbf{b}^j}\rangle_{L^2(\mathbf{T}^2)}B_k^{\mathbf{a}^j}\otimes B_l^{\mathbf{b}^j}\|_{H^2_\omega(\mathbf{D}^2)}.\tag{3.16}$$

As in the unit disc context if $d=0$ then we have case in (i); and if $d>0$ we can accordingly derive a contradiction. All the task left for proving the theorem is estimation (3.14).

The strategy to prove (3.14) is to employ the tensor product type Poisson kernels, namely, $P_r^{(1)}\otimes P_\rho^{(2)}$, where $P_r^{(1)}$ and $P_\rho^{(2)}$ denote the Poisson kernels for, respectively, the first and the second circular variable. We will use $*_1$ and $*_2$ to denote the convolutions with respect to the first and the second circular variable, respectively. We will show that in $H^2_\omega(\mathbf{D}^2)$ the operator $(P_r \otimes P_\rho)*$ is a contraction, as well as an approximation to the identity.

To show that the tensor product Poisson convolution is a contraction we have

$$\begin{aligned}
\|(P_r^{(1)} \otimes P_\rho^{(2)}) * f_\omega\|^2_{H^2_\omega(\mathbf{D}^2)} &= E_\omega \|(P_r^{(1)} \otimes P_\rho^{(2)}) * f_\omega\|^2_{L^2(\partial\mathbf{D}\times\partial\mathbf{D})} \\
&= E_\omega \|P_r^{(1)} *_1 (P_\rho^{(2)} *_2 f_\omega)\|^2_{L^2(\partial\mathbf{D}\times\partial\mathbf{D})} \\
&\leq E_\omega \|P_\rho^{(2)} *_2 f_\omega\|^2_{L^2(\partial\mathbf{D}\times\partial\mathbf{D})} \\
&\leq E_\omega \|f_\omega\|^2_{L^2(\partial\mathbf{D}\times\partial\mathbf{D})} \\
&= \|f\|_{H^2_\omega(\mathbf{D}^2)}.
\end{aligned}$$

Now we show that the tensor product Poisson convolution is an approximation to the identity in $H^2_\omega(\mathbf{D}^2)$. Denote the identity operators for the first and the second variable by I_1 and I_2, respectively. Since the partial Poisson operators are contractions, as well as approximation to the identity in their respective spaces, we have

$$\begin{aligned}
\|(P_r^{(1)} \otimes P_\rho^{(2)}) * f_\omega - f_\omega\|_{H^2_\omega(\mathbf{D}^2)} &\leq \|P_\rho^{(2)} *_2 (P_r^{(1)} *_1 f_\omega) - P_\rho^{(2)} *_2 f_\omega\|_{H^2_\omega(\mathbf{D}^2)} + \|P_\rho^{(2)} *_2 f_\omega - f_\omega\|_{H^2_\omega(\mathbf{D}^2)} \\
&= \|P_\rho^{(2)} *_2 (P_r^{(1)} *_1 - I_1) f_\omega\|^2_{H^2_\omega(\mathbf{D}^2)} + \|(P_\rho^{(2)} *_2 - I_2) f_\omega\|_{H^2_\omega(\mathbf{D}^2)} \\
&\leq \|(P_r^{(1)} *_1 - I_1) f_\omega\|^2_{H^2_\omega(\mathbf{D}^2)} + \|(P_\rho^{(2)} *_2 - I_2) f_\omega\|_{H^2_\omega(\mathbf{D}^2)} \\
&\to 0, \qquad \text{as } r \to 1- \quad \text{and } \rho \to 1-,
\end{aligned}$$

where, as in the proof of (2.7), the last step invokes the Lebesgue Dominated Convergence Theorem. In the way of (2.8) and (2.9) we also have

$$\|\mathcal{R}^{(j)}_{k_0,l_0}\|_{H^2_\omega(\mathbf{D}^2)} \geq (1-\epsilon/2)\|\mathcal{R}^{(j)}_{k_0,l_0}\|_{H^2_\omega(\mathbf{D}^2)} - \|\mathcal{R}^{(j)}_{k_0,l_0}\|_{H^2_\omega(\mathbf{D}^2)} \sum_{k>k_0 \text{ or } l>l_0} \|(P_{r_1}^{(1)} \otimes P_{\rho_1}^{(2)}) * (B_k^{\mathbf{a}^j} \otimes B_l^{\mathbf{b}^j})\|_{H^2(\mathbf{T}^2)}. \tag{3.17}$$

To prove the relations in the polydisc context counterpart to (2.10), and that to (2.11), we have, for $k > k_0$ or $l > l_0$,

$$\begin{aligned}
\|(P_{r_1}^{(1)} \otimes P_{\rho_1}^{(2)}) * (B_k^{\mathbf{a}^j} \otimes B_l^{\mathbf{b}^j})\|^2_{H^2(\mathbf{T}^2)} &= \|P_{r_1}^{(1)} *_1 B_k^{\mathbf{a}^j}\|^2 \|P_{\rho_1}^{(2)} *_2 B_l^{\mathbf{b}^j}\|^2_{H^2(\mathbf{T}^2)} \\
&\leq \frac{1}{4\pi^2} \int_0^{2\pi} |e_{a_k^{(j)}}(r_1 e^{it})|^2 dt \int_0^{2\pi} |e_{b_l^{(j)}}(\rho_1 e^{is})|^2 ds \\
&\to 0,
\end{aligned}$$

where r_1 and ρ_1 are previously chosen to satisfy (3.17) and fixed. The proof is thus complete. □

4. Random cyclic AFD algorithm

AFD, as abbreviation of *adaptive Fourier decomposition*, is established in [10]. Although its energy pursuing idea is the same as greedy algorithm, the advantage of AFD lies on the fact that the AFD types look into attainability of the optimal energy gain at each of the iteration step. The attainability of the optimal energy is, in particular, based on admitting the unavoidble multiplicities of the parameters. The general greedy algorithm principle, on the other hand, does not allow repetition of selection of parameters. In this study we go two steps further by concerning attainability of simultaneous selection of a set of n parameters, as well as the randomization. AFD methods, as a generalization of Fourier theory, involve complex analysis, especially Blaschke products. The motivation of AFD is sparse representation of signals into those with positive analytic instantaneous frequency. In [7] and [13] we propose a practical cyclic AFD algorithm to solve the n-best problem for the deterministic signals case. In view of the methodology, cyclic AFD is not a theoretical solution for the n-best problem. It is only a practical method in the sense that it cannot prevent itself from falling into local minimums, and the way to treat this is to start the algorithm with multiple and maybe various initial values. This stands as a common computational problem. With this understanding in mind cyclic AFD can be extended to the random signals case as a practical method, called *random cyclic AFD*. Cyclic AFD for deterministic signals is summarized as follows.

Let $f \in H^2(\mathbf{D})$. To get a start, let $\mathbf{a}^0 = (a_1^{(0)}, \cdots, a_n^{(0)})$ be any n-tuple of complex numbers in $\mathbf{D}^n$. Note that in the following process multiplicities of parameters are always allowed. This initial n-tuple generates a TM system $(B_1^{(0)}, \cdots, B_n^{(0)})$. Denote the span of $B_1^{(0)}, \cdots, B_n^{(0)}$ by $\mathcal{B}^0$. Now we throw away the first entry in the n-tuple, namely, $a_1^{(0)}$, and replace it with a better $a \in \mathbf{D}$, if exists, and denote this better a by $a = a_1^{(1)}$. Here " better" takes the sense that the energy of the projection, $\|P_{\mathcal{B}^1} f\|^2_{H^2(\mathbf{D})}$, being larger than $\|P_{\mathcal{B}^0} f\|^2_{H^2(\mathbf{D})}$, is maximized by $a_1^{(1)}$, where $\mathcal{B}^1$ is the TM system corresponding to $\mathbf{a}^1 \triangleq (a_1^{(1)}, a_2^{(0)}, \cdots, a_n^{(0)}) \triangleq (a_1^{(1)}, a_2^{(1)}, \cdots, a_n^{(1)})$. This is the first re-selection of the parameters. Next, we re-select the entry $a_2^{(1)}$ according to the maximal projection principle, while the other parameters remain unchanged, and denote the newly selected better entry by $a_2^{(2)}$, and denote $\mathbf{a}^2 \triangleq (a_1^{(1)}, a_2^{(2)}, a_3^{(1)}, \cdots, a_n^{(1)}) \triangleq (a_1^{(2)}, a_2^{(2)}, a_3^{(2)}, \cdots, a_n^{(2)})$. Next, we re-select the entry in the third position, $a_3^{(2)}$, to get $a_3^{(3)}$, and so on. After the first round of n re-selections we can proceed the second one, and cyclically, until the energy is seen to have little gain with any new re-selection. As is proved in [7] the n-tuple sequence $\mathbf{a}^l$ converges to an n-tuple $\mathbf{a} \in \mathbf{D}^n$, being a practical solution of the n-best problem for f. For more details of the algorithm and theoretical remarks the author is referred to [7] and [13]. While the structure is the same, the difference of stochastic cyclic AFD lies on the random maximal selection principle: As is proved in [9] for every $f \in H^2_\omega(\mathbf{D})$ there exists $a \in \mathbf{D}$ such that

$$E_\omega |\langle f_\omega, e_a \rangle_{L^2(\partial \mathbf{D})}|^2 = \max\{E_\omega |\langle f_\omega, e_b \rangle_{L^2(\partial \mathbf{D})}|^2 \mid b \in \mathbf{D}\}.$$

In practice the expectation can be computed through taking a sample of sufficiently large capacity, being based on the Law of Large Numbers, namely,

$$E_\omega |\langle f_\omega, e_a \rangle_{L^2(\partial \mathbf{D})}|^2 = \lim_{N\to\infty} \frac{\sum_{m=1}^N |\langle (f_m)_\omega, e_a \rangle_{L^2(\partial \mathbf{D})}|^2}{N},$$

where $f_m, m = 1, \cdots, N$, are samples.

Acknowledgments

The authors would like to sincerely thank the reviewer who pointed out the intrinsic connections between this study and the Karhunen-Loéve Theorem.

References

[1] L. Baratchart, Existence and generic properties of L^2 approximations for linear systems, IMA J. Math. Control Inf. 2 (3) (1986) 89–101.
[2] L. Baratchart, M. Cardelli, M. Olivi, Identification and rational L^2 approximation a gradient algorithm, Automatica 27 (2) (1991) 413–417.
[3] J. Garnett, Bounded Analytic Functions, Springer, New York, 1981.
[4] Y.T. Li, L.M. Zhang, T. Qian, 2D partial unwinding - a novel non-linear phase decomposition of images, IEEE Trans. Image Process. 28 (10) (2019) 4762–4773.

[5] W. Mi, T. Qian, Frequency-domain identification: an algorithm based on an adaptive rational orthogonal system, Automatica 48 (6) (2012) 1154–1162.
[6] W. Mi, T. Qian, On backward shift algorithm for estimating poles of systems, Automatica 50 (6) (2014) 1603–1610.
[7] T. Qian, Cyclic AFD algorithm for best rational approximation, Math. Methods Appl. Sci. 37 (6) (2014) 846–859.
[8] T. Qian, Two-dimensional adaptive Fourier decomposition, Math. Methods Appl. Sci. 39 (10) (2016) 2431–2448.
[9] T. Qian, Sparse representations of random signals, arXiv:2008.10473 [math.PR], 2020.
[10] T. Qian, Y.B. Wang, Adaptive Fourier series — a variation of greedy algorithm, Adv. Comput. Math. 34 (3) (2011) 279–293.
[11] T. Qian, Y.B. Wang, Remarks on adaptive Fourier decomposition, Int. J. Wavelets Multiresolut. Inf. Process. 11 (1) (2013) 1–14.
[12] T. Qian, E. Wegert, Optimal approximation by Blaschke forms, Complex Var. Elliptic Equ. 58 (1) (2013) 123–133.
[13] T. Qian, J.Z. Wang, W.X. Mai, An enhancement algorithm for cyclic adaptive Fourier decomposition, Appl. Comput. Harmon. Anal. 47 (2) (2019) 516–525.
[14] J.L. Walsh, Interpolation and Approximation by Rational Functions in the Complex Domain, American Mathematical Society, 1895.
[15] X.Y. Wang, T. Qian, I.T. Leong, Y. Gao, 2-D frequency domain system identification, IEEE Trans. Autom. Control 65 (2) (2020) 577–590.