Appl. Comput. Harmon. Anal. 67 (2023) 101568

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

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

n-Best kernel approximation in reproducing kernel Hilbert spaces

Tao Qian [1]

Macau University of Science and Technology, (Macau Center for Mathematical Sciences), Macau, China

A R T I C L E I N F O

Article history:
Received 28 March 2022
Received in revised form 22 March 2023
Accepted 11 June 2023
Available online 22 June 2023
Communicated by Ming-Jun Lai

MSC:
41A50
46E22
30H10
30H20

Keywords:
Weighted Bergman space
Weighted Hardy space
Maximum modulus principle
n-Best rational approximation
n-Best kernel approximation
Bochner space

A B S T R A C T

By making a seminal use of the maximum modulus principle of holomorphic functions we prove existence of n-best kernel approximation for a wide class of reproducing kernel Hilbert spaces of holomorphic functions in the unit disc, and for the corresponding class of Bochner type spaces of stochastic processes. This study thus generalizes the classical result of n-best rational approximation for the Hardy space and a recent result of n-best kernel approximation for the weighted Bergman spaces of the unit disc. The type of approximations has significant applications to signal and image processing and system identification, as well as to numerical solutions of the classical and the stochastic type integral and differential equations.

1. Introduction

A main form of application of mathematical analysis is approximation by basic functions of the underlying space. Various forms and topics of polynomial and rational approximations have been studied, including convergence models, capacity and rates, existence and uniqueness of best approximation, as well as algorithms, etc. See for instance the selected list of the literature [62,19,14,53,12,29,55,27,10,11,9,18,66,8,57,54] and the references therein. The present study will concentrate in approximation of reproducing kernel Hilbert spaces (RKHSs) of complex holomorphic functions, the latter being related to Z-transforms of system transfer functions. In RKHSs the most natural basic functions are the parameterized reproducing kernels. Tasks of signal and image processing are based on effective reconstruction of a given signal or image. To measure reconstruction efficiency, among the most commonly used, there are two dual models. One is, for a previ-

E-mail address: tqian@must.edu.mo.
[1] Funded by The Science and Technology Development Fund, Macau SAR (File no. 0123/2018/A3).

https://doi.org/10.1016/j.acha.2023.06.003
1063-5203/

ously given ϵ, to determine the smallest integer n such that the difference, measured in the underlying space norm, between the given function and some n-linear combination of the basic functions is already dominated by ϵ. The second model is, for a given resource limitation represented by a natural integer n, to find an n-tuple of parameterized basic functions and an n-tuple of coefficients such that the n-linear combination that they compose gives rise to the best possible approximation to the given function. The second model is abbreviated as *n-best approximation* or more briefly *n-best problem*. The present paper restricts to study the second model, and only the existence part of the n-best solutions. The corresponding algorithm part for actually finding one or all n-best solutions, as a consequence of the technical results of the existence proof, will be separately studied.

We will be based on the general concept reproducing kernel Hilbert space (RKHS). In 1930, to study the partial differential equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \alpha(x,y)\frac{\partial u}{\partial x} + \beta(x,y)\frac{\partial u}{\partial y} + \gamma(x,y)u = 0,$$

where $\alpha(x,y), \beta(x,y), \gamma(x,y) \in C^2(\Omega), \Omega$ is a bounded region, $\alpha(x,y), \beta(x,y), \gamma(x,y) \in C^2(\Omega), \Omega$ are all real-analytic functions, S. Bergman proposed the reproducing kernel concept and gave the related formulas. The theoretical study of reproducing kernel may be divided into two stages. Of which the first started from J. Mercer ([35]), at the beginning of the 20th century, who in his studies of integral equations brought up the concept positive definite kernel:

$$\sum_{i,j=1}^{n} K(y_i, y_j)\xi_i\xi_j \geq 0.$$

The second stage is the development by E. H. Moore ([36]), around 1930's, who proved that every positive Hermitian matrix induces a Hilbert space that has a kernel function $K(x,y)$ enjoying the property

$$f(y) = \langle f, K(\cdot, y)\rangle.$$

The same phenomenon was observed by S. Bochner in the convolution kernel form connected to Fourier theory ([7]). Around 1940's the most popular reproducing kernels were the Bergman type ones. Bergman developed the idea of S. Zaremba ([69]) to solve boundary value problems by using reproducing kernels, showing that reproducing kernels are effective tools to solve elliptic boundary value problems.

Combined with various science and engineering objects there have developed new theories and algorithms, including signal processing ([3]), system identification ([4,38,39,47]), scholastics processing ([25,43,50,65]), estimation theory ([22,41]), wavelet transform ([56]), reproducing kernel particle method ([24,23,26,17,60]), the moving least-square reproducing kernel method ([31,32]), multi-scale reproducing kernel particle methods ([30]), etc., with ample applications.

F. M. Larkin ([37]) and M. M. Chawla ([13]) studied the approximation aspect. Formats of interpolation with reproducing kernel have been used to numerical solutions of partial differential equations and integral equations. The latest reproducing kernel approximation methodology, called adaptive Fourier decomposition (AFD), uses the maximal energy extraction principle similar to what is in greedy algorithm or matching pursuit ([40,58]). Greedy algorithms are based on general Hilbert space theory with a dictionary ([59]). The AFD methods, originated from analytic positive frequency decomposition, validate attainability of the best suited parameters. Technically, AFDs are based on delicate mathematical analysis, and, in particular, allow repeating selection of parameters through defining multiple kernels, when necessary. The technical treatment is a blend of functional analysis, complex analysis and harmonic analytic. Recent studies on approximation in Hardy spaces, including the latest n-best and stochastic AFD approximations, may be found in a sequence of articles [44,42,1,2,15,16,46,61,43,50]. Some early studies of adaptive Fourier decomposition in Bergman and weighted Bergman spaces are given in [48,49]. Celebrating results for the n-best type approximation in weighted Bergman spaces are presented in [51].

Recent studies given by Ball et al.'s papers ([5,6]) show that approximations of Hilbert spaces of holomorphic functions have intimate connections with system identification and in particular with time-variant linear systems. In the Hardy space case there holds the Sz.-Nagy-Foias model theory for $C_{\cdot 0}$ contraction operators. The model theory combined with the Burling-Lax theorem addresses a correspondence between any two of the four kinds of objects: shift invariant subspaces, operator-valued inner functions, conservative discrete-time input/state/output linear system, and $C_{\cdot 0}$ Hilbert-space contraction operators. The studies of [5] and [6] extend such correspondence to weighted Bergman and weighted Hardy spaces. Under such frame work and via the Z-transform of the system, n-best approximation in each of the mentioned spaces determines the optimal shift invariant subspaces for effective and efficient system identification.

To complete the introduction for science and engineering motivations and involvements, we at last, but not least, mention that there have been forerunner but recent developments on the stochastic n-best model as cited in [43,50,65]. Stochastic n-best approximation of our general setting will be given in §5. Stochastic AFD offers a new approach to stochastic processes, including solutions of stochastic partial differential equations. It, in particular, stands as an alternative method to the Karhunen-Loéve decomposition together with several advantages. Further developments along this direction are to be reported in separate and forthcoming papers.

Next, we introduce the related preliminary knowledge and raise the n-best problem in our general setting. The Hardy space is defined

$$\mathbb{H}^2(\mathbf{D}) = \{f : \mathbf{D} \to \mathbf{C} \ : \ f \text{ is analytic in } \mathbf{D} \text{ and } \|f\|_{\mathbb{H}^2} = \sup_{0\le r<1} \int_0^{2\pi} |f(r\mathrm{e}^{it})|^2 < \infty\}.$$

As a fundamental result, functions in the Hardy space have non-tangential boundary limits a.e. on the unit circle $\partial\mathbf{D}$. The Hardy space is isometric to the function space $\mathbb{H}^2{}_{\partial\mathbf{D}}$ consisting of the non-tangential limiting functions. One of the alternative definitions of the space $\mathbb{H}^2{}_{\partial\mathbf{D}}$ is

$$\mathbb{H}^2{}_{\partial\mathbf{D}} = \{f : \partial\mathbf{D} \to \mathbf{C} \ : \ f \in L^2(\partial\mathbf{D}), \quad f(\mathrm{e}^{it}) = \sum_{k=0}^{\infty} c_k \mathrm{e}^{ikt}, \quad \sum_{k=0}^{\infty} |c_k|^2 < \infty\}.$$

$\mathbb{H}^2{}_{\partial\mathbf{D}}$ is a closed subspace of the Hilbert space $L^2(\partial\mathbf{D})$ equipped with the inner product

$$\langle f, g\rangle = \frac{1}{2\pi}\int_0^{2\pi} f(\mathrm{e}^{it})\overline{g(\mathrm{e}^{it})}dt. \tag{1.1}$$

For a given positive integer n, an ordered pair of polynomials (p, q) is called an *n-admissible pair* if p and q are co-prime, $q \neq 0$ in $\mathbf{D}$, and both the degrees of p and q do not exceed n. The famous *n-best rational approximation problem* in $\mathbb{H}^2(\mathbf{D})$ is as follows: For $f \in \mathbb{H}^2(\mathbf{D})$, find an n-admissible ordered pair $(\tilde{p}, \tilde{q})$ such that

$$\|f - \frac{\tilde{p}}{\tilde{q}}\|_{\mathbb{H}^2(\mathbf{D})} = \inf\{\|f - \frac{p}{q}\|_{\mathbb{H}^2(\mathbf{D})} \ : \ (p, q) \text{ is an } n\text{-}\textit{admissible pair}\}. \tag{1.2}$$

Closely related to rational approximation there exist studies on what is called Takenaka-Malmquist (TM) system, or rational orthogonal system:

$$\{B_{a_1a_2\ldots a_n}(z)\}_{n=1}^{\infty} = \left\{ \frac{\sqrt{1-|a_n|^2}}{1-\bar{a}_n z} \prod_{k=1}^{n-1} \frac{z-a_k}{1-\bar{a}_k z} \right\}_{n=1}^{\infty}, \quad a_1, \cdots, a_n, \cdots, \in \mathbf{D}, \tag{1.3}$$

where $B_{a_1 a_2 \ldots a_n}(z) = e_{a_n}(z)\phi_{a_1 a_2 \ldots a_{n-1}}(z)$, where e_{a_n} is the normalized Szegö kernel at a_n,

$$e_{a_n}(z) = \frac{\sqrt{1-|a_n|^2}}{1-\overline{a}_n z},$$

and the *canonical Blaschke product* with zeros $a_1, \cdots, a_{n-1}$,

$$\phi_{a_1 a_2 \ldots a_{n-1}}(z) = \prod_{k=1}^{n-1} \frac{z-a_k}{1-\bar{a}_k z}. \tag{1.4}$$

There further holds the relation, for $a \in \mathbf{D}$,

$$e_a(z) = \frac{\sqrt{1-|a|^2}}{1-\overline{a}z} = \frac{k_a}{\|k_a\|}, \tag{1.5}$$

where

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

are, respectively, the reproducing kernel of $\mathbb{H}^2(\mathbf{D})$ and the normalizing constant making $\|e_a\| = 1$.

The above formulated n-best rational approximation problem (1.2) is, in essence, equivalent to the following *n-best Blaschke form approximation* problem ([45,46]): Let n be a given positive integer. If f itself is not an m-Blaschke form for some $m < n$, find a set of n parameters $a_1, \cdots, a_n$, all in $\mathbf{D}$, such that

$$\|f - \sum_{k=1}^n \langle f, B_{a_1\cdots a_k}\rangle B_{a_1\cdots a_k}\| = \inf\{\|f - \sum_{k=1}^n \langle f, B_{b_1\cdots b_k}\rangle B_{b_1\cdots b_k}\| \ : \ b_1, \cdots, b_n \ \in \mathbf{D}\}. \tag{1.6}$$

The formulation of the problem allows multiplicity of the zeros a_k's. A solution of (1.6) will be referred as an "n-best Blaschke form approximation"([46,45]). Regardless unimodular constants (see Lemma 3.2 below), an n-TM system is the Gram-Schmidt orthogonalization of a set of n Szegö kernels, or multiple kernels (formulated in (2.12) below) when the parameters are with multiplicities. See explanations in §2). The n-best Blaschke form approximation is, again, equivalent with the *n-best kernel approximation* formulated as: Find $(a_1, \cdots, a_n) \in \mathbf{D}^n$ and $(c_1, \cdots, c_n) \in \mathbf{C}^n$ such that

$$\begin{aligned}\|f - \sum_{j=1}^n c_j \tilde{k}_{a_j}\| = \inf\{\|f - \sum_{j=1}^n c'_j k_{a'_j}\| : a'_1, \cdots, a'_n \ \in \mathbf{D} \\ \text{are distinct, and} c'_1, \cdots, c'_n \ \in \mathbf{C}\},\end{aligned} \tag{1.7}$$

where $\tilde{k}_{a_j}$ are multiple kernels (see (2.12)).

Considerable amount of studies have been devoted to the above problem with the three equivalent forms. See [62,63,11,9,46,45,38,51]. Amongst, researchers have obtained several new proofs for existence of a solution to (1.2). The motivation of exploring new proofs of the existence, including that of the author himself's, would be at least two-folder: (i) The known existence proofs for the Hardy space case do not seem to be adaptable to prove existence of an n-best approximation in any non-Hardy spaces, including weighted Bergman spaces and weighted Hardy spaces. In fact, before the work [51] whether there is a solution to the problem in any non-Hardy space was unknown; and (ii) On top of the existence, an ultimate algorithm of finding even one n-best solution of (1.2) has yet to be sought: The commonly adopted empirical algorithms are all local that cannot theoretically avoid the possibility of sinking into a local minimum. See

[9,11,44,42,48,52] and the references therein. As an extension of the traditional Fourier method, both the n-best and the repeated one-by-one types have been found to have effective applications in signal and image processing and system identification ([47,34,38]).

The question for n-best kernel approximation can be raised in general RKHSs, or even in Hilbert spaces with a dictionary. The recently published new proof of existence of n-best approximation in the Hardy space case ([64]) uses the maximum modulus principle of holomorphic functions as a crucial technical trick. Inspired by this complex analysis method, through proving some necessary new crucial pointwise estimations of the kernels of the involved zero- and Blaschke-weighted spaces, the study [51] was able to prove existence of the n-best kernel approximation of all the weighted Bergman spaces $\mathbb{A}^2_\alpha(\mathbf{D}), -1 < \alpha < \infty$.

The study in the present paper is directly motivated by the recent new proofs of existence of n-best approximation on the Hardy space ([61]) and one on the weighted Bergman spaces ([51]). We achieve a clever and concise proof without requiring pointwise estimation of the involved reproducing kernels, contrasting with [51] in which pointwise estimations have to be used, and thus prove existence of the n-best approximation for a larger class of reproducing kernel Hilbert spaces including all the weighted Hardy spaces, having the weighted Bergman spaces as particular cases.

Based on further analysis of the orthogonalization projection operator $Q_{a_1a_2\dots a_k}$ and factorization of higher order generalized backward shift operators $Q_{a_1a_2\dots a_k}/\phi_{a_1\cdots a_{m-1}}$ (see §3), the present paper is able to avoid use of the pointwise estimations of the reproducing kernels of the involved zero spaces and the Blaschke weighted spaces. For RKHSs more general than the Bergman ones such kernel estimations may be impossible. As a result, we are able to assert existence of the n-best problem for a class of RKHSs more general than the weighted Hardy spaces. Precisely, we can declare existence of solutions of the n-best kernel approximation for all RKHSs of holomorphic functions in $\mathbf{D}$ that satisfy the following three conditions (see Theorem 2.1):

(i) The reproducing kernel $K(z,w)$ enjoys *the analyticity condition*: When $w \in \mathbf{D}$ is fixed, $K(z,w)$ is analytic for z in a neighborhood of the closed unit disc $\overline{\mathbf{D}}$, and, when $z \in \mathbf{D}$ is fixed, $K(z,w)$ is anti-analytic for w in a neighborhood of $\overline{\mathbf{D}}$;
(ii) The kernel $K(z,w)$ satisfies *the infinite-norm-property at the boundary*, that is,

$$\lim_{w\to\partial\mathbf{D}} \|K_w\| = \infty; \tag{1.8}$$

and,
(iii) $K(z,w)$ satisfies *the uniformly boundedness condition*

$$\frac{|K_w(z)|}{\|K_w\|^2} \le C_{\mathcal{H}}, \quad w, z \in \mathbf{D}, \tag{1.9}$$

where $C_{\mathcal{H}}$ is a constant depending on the space. Below, we will call a RKHS of holomorphic functions in $\mathbf{D}$ satisfying the conditions (i), (ii), (iii) as our "*general $\mathcal{H}$ space setting*".

There is an ordered sub-family, $\mathbb{H}_{W_\beta}(\mathbf{D}), -\infty < \beta < \infty$, called the Hardy-Sobolev spaces, within the family of weighted Hardy spaces ([5,6], also see §4). The index range $\beta < 0$ corresponds to the weighted Bergman spaces including the standard Bergman space case for $\beta = -1$ whose n-best existence results are proved in [51]. The $\beta = 0$ case corresponds to the Hardy space [64]. The n-best existence results for the Hardy-Sobolev spaces for $0 < \beta \le 1$ ($\beta = 1$ corresponds to the Dirichlet space) are obtained as a consequence of the main result of this paper (see §4). The Hardy-Sobolev spaces for $\beta > 1$ do not fall into the category of the RKHSs considered in the main theorem of this paper, but we show that they are governed by the Sobolev Embedding Theorem (also see [49]).

The writing plan of the paper is as follows. In §2 we discuss in detail the Gram-Schmidt orthogonalization of reproducing kernels that induces the concept multiple kernels. It is in terms of the multiple kernel concept that the n-best problem is precisely formulated. In §3 our main result, Theorem 2.1, on existence of the n-best approximation of a wide class of RKHSs is proved through a number of technical lemmas in relation to orthogonal projections and analysis of the involved reproducing kernels. In §4, as examples of using the main result Theorem 2.1, we give re-proofs of the Hardy and the weighted Bergman space results of [64] and [51], and to prove, using the unified method, the n-best existence result for the range $-\infty < \beta \le 1$ of the Hardy-Sobolev spaces $\mathbb{H}_{W_\beta}(\mathbf{D})$, in which the results for the range $(-\infty, 1]$ are known to be equivalent to the weighted Bergman and the Hardy space cases. The results for the range $\beta \in (0,1]$ are new as applications of Theorem 2.1. We include a remark in §4 concerning the range $\beta \in (1,\infty)$ through invoking the Sobolev Embedding theorem that provides complete understanding to the n-best issue for the whole range $-\infty < \beta < \infty$. In §5 we extend Theorem 2.1 to the stochastic signal case based on the Bochner type Hilbert space setting. For the existing related studies in the stochastic signal direction we refer to [43] and [50]. To the end of §5 we include a remark for the impact of this study on obtaining algorithms in finding the n-best solutions.

2. Main theorem

Let $\mathcal{H}$ be a RKHS of holomorphic functions in $\mathbf{D}$ with reproducing kernel $K_w, w \in \mathbf{D}$, satisfying the conditions (i), (ii), (iii) set in §1. Throughout the paper n is a fixed positive integer. Let $Z_k = (a_1, \cdots, a_k), 1 \le k \le n$, be an ordered k-tuple of complex numbers in $\mathbf{D}$ allowing multiplicity.

Denote by $l(a_k)$ the multiple of a_k in the k-tuple $(a_1, \cdots, a_k), k \le n$. Denote

$$\tilde{K}_{a_k}(z) = \left[\left(\frac{d}{d\overline{w}}\right)^{l(a_k)-1} K_w(z)\right]_{w=a_k}. \tag{2.10}$$

We will call $\tilde{K}_{a_k}$ the *multiple reproducing kernel corresponding to* $(a_1, \cdots, a_k)$. It is easy to show, for f being in the holomorphic function space, there holds

$$\langle f, \tilde{K}_{a_k}\rangle = f^{(l(a_k)-1)}(a_k). \tag{2.11}$$

The consecutive derivatives of the kernel function correspond to repeating use of kernel parameters.

In the Hardy space case, for instance, $k_a(z) = \frac{1}{1-\overline{a}z}$, and

$$\tilde{k}_{a_k}(z) = \left[\left(\frac{d}{d\overline{w}}\right)^{l(a_k)-1} k_w(z)\right]_{w=a_k} = \frac{l!\overline{a}^l}{(1-\overline{a}z)^{l+1}}. \tag{2.12}$$

In general cases, let $(a_1, \cdots, a_n)$ be any n-tuple of complex numbers in $\mathbf{D}$. Denote by $(E_{a_1}, \cdots, E_{a_1\cdots a_m})$ the Grand-Schmidt orthonormalization of $(\tilde{K}_{a_1}, \cdots, \tilde{K}_{a_m}),\ m = 1, \cdots, n$, given by

$$E_{a_1\cdots a_m}(z) = \frac{\tilde{K}_{a_m}(z) - \sum_{l=1}^{m-1}\langle \tilde{K}_{a_m}, E_{a_1\cdots a_l}\rangle E_{a_1\cdots a_l}(z)}{\sqrt{\|\tilde{K}_{a_m}\|^2 - \sum_{l=1}^{m-1}|\langle \tilde{K}_{a_m}, E_{a_1\cdots a_l}\rangle|^2}}. \tag{2.13}$$

We will denote the orthogonal projection of f into the linear subspace X by $P_X(f)$. The projection into the orthogonal complement of X is denoted $Q_X = I - P_X$. In particular, denote by $P_{a_1\cdots a_m}$ the orthogonal projection from $\mathcal{H}$ to $\mathrm{span}\{\tilde{K}_{a_1}, \cdots, \tilde{K}_{a_m}\}$, and by $Q_{a_1\cdots a_m} = I - P_{a_1\cdots a_m}$, the projection into the orthogonal complement subspace of $\mathrm{span}\{\tilde{K}_{a_1}, \cdots, \tilde{K}_{a_m}\}$. It is recognized that $Q_{a_1 a_2 \ldots a_k}$ corresponds to the Gram-Schmidt process, precisely,

$$E_{a_1 a_2 \ldots a_k} = \frac{Q_{a_1 a_2 \ldots a_{k-1}}(\tilde{K}_{a_k})}{\|Q_{a_1 a_2 \ldots a_{k-1}}(\tilde{K}_{a_k})\|}. \tag{2.14}$$

We have been using the notation $\{B_{a_1a_2,\ldots a_k}\}_{k=1}^{\infty}$ for the TM system in the Hardy space case. Now we use $\{E_{a_1a_2\ldots a_k}\}_{k=1}^{\infty}$ for the Gram-Schmidt orthonormalization of the multiple reproducing kernels $\{\tilde{K}_{a_k}\}_{k=1}^{\infty}$ in $\mathcal{H}$ given by (2.13) and (2.14). In the classical Hardy space case they are essentially the same, as, in fact, $E_{a_1a_2\ldots a_k} = c_k B_{a_1a_2\ldots a_k}$, where c_k are unimodular constants, $k = 1, \cdots$, (see Lemma 3.2 and the relevant references). None of the $E_{a_1a_2\ldots a_k}$ of any holomorphic Hilbert spaces other than the Hardy space seem to have such nice construction: the orthonormalization of the Szegö kernel k_{a_k} with respect to the span of $\{B_{a_1a_2\ldots a_{j-1}}\}_{j=1}^{k-1}$, which is $B_{a_1a_2\ldots a_k}$, is just the product of the added normalized Szegö kernel e_{a_k} with the canonical Blaschke $\phi_{a_1a_2\ldots a_{k-1}}$. This extraordinary property, together with the complex unimodular property of Blaschke products on the circle, as well as the equivalent norm property restricted to the circle, offer decisive conveniences in developing the Hardy space theory in contrast with that of the non-Hardy space cases.

The n-best kernel approximation questions in the general context are formulated as follows. Let $f \in \mathcal{H}$. Whether there exist $a_1, \cdots, a_n$, all in $\mathbf{D}$, such that

$$\|f - P_{a_1\cdots a_n} f\| = \inf\{\|f - P_{b_1\cdots b_n} f\| \ : \ (b_1 \cdots b_n) \in \mathbf{D}^n\}? \tag{2.15}$$

We note that a finite and non-negative infimum defined by the right-hand-side of (2.15) always exists. The infimum value will be denoted as $d_f(n)$ in the rest part of this paper. Since n is fixed, $d_f(n)$ is often abbreviated as d_f. The n-best problem addresses existence of an n-tuple $(a_1, \cdots, a_n)$ that gives rise to the infimum value. Further more, if there exist such $(a_1, \cdots, a_n)$, how to find all of them computationally? In the present paper we will prove the following existence result.

Theorem 2.1. *Suppose that $\mathcal{H}$ is a RKHS of holomorphic functions satisfying (i), (ii), and (iii), and n is a positive integer. Then there must hold one of the following two cases: (1) f is a linear combination of $\tilde{K}_{b_1}, \cdots, \tilde{K}_{b_{m_1}}$ for some m_1-tuple $(b_1, \cdots, b_{m_1}) \in \mathbf{D}^{m_1}, m_1 \le n$; or (2) there exists an n-tuple $(a_1, \cdots, a_n) \in \mathbf{D}^n$ such that (2.15) holds for a positive infimum $d_f > 0$. Moreover, in the case,*

$$d_f = \|f - \sum_{l=1}^{n} \langle f, E_{a_1\cdots a_l}\rangle E_{a_1\cdots a_l}\|. \tag{2.16}$$

3. Some technical lemmas

We start with giving some descriptions of the idea of the proof. We first admit that the infimum (2.15) is attainable through a sequence of n-tuples $(a_1^{(l)}, \cdots, a_n^{(l)}), l = 1, 2, \cdots$, when n parameters are truly necessary. Through a compact argument, and without loss of generality, we may assume that $\lim_{l\to\infty} a_k^{(l)} = a_k \in \overline{\mathbf{D}}, k = 1, \cdots, n$. We are to prove in such case all the limiting points a_k are, in fact, right inside the open unit disc but not on the boundary. To quantitatively show this we would have to compute the projections, through using the Gram-Schmidt (GS) orthogonalization, of f onto the spans of the involved reproducing kernels. The GS process then establishes the mutual relations between each l-level parameters. The technical difficulty arises in proving the corresponding k-projection components tending to zero along with $l \to \infty$ when the limiting point a_k being, as assumed, on the boundary, and it has to be proved irrelevantly or uniformly with locations of the other parameters $a_j^{(l)}, l \to \infty, j \ne k$. The uniform convergence is proved eventually through use of the maximum modulus principle in one complex variable. To proceed the proof one needs to analyze the relations between zeros, orthogonality with reproducing kernels, orthogonal projections, zero-reproducing kernel Hilbert spaces and Blaschke-weighted reproducing kernel Hilbert spaces, and boundedness behavior of an analytic function after applying GS operation. The details are treated through a series of technical lemmas.

We will use the notation $f_{a_1a_2\ldots a_k} = Q_{a_1a_2\ldots a_k} f$, where $a_j, j = 1, \cdots, k$, are allowed to repeat.

Lemma 3.1. *If a_j is among $a_1, \cdots, a_k$, then $f_{a_1a_2\ldots a_k}(a_j) = 0$, including the multiplicity.*

Proof. The proof is straightforward if $a_1, \cdots, a_k$ are all different. In that case $l(a_j) = 1, j = 1, \cdots, k$. Due to the self-adjoint property of the projection operators and the orthogonality gained from G-S process, for a_j being among $a_1, \cdots, a_k$,

$$\begin{aligned} f_{a_1a_2\ldots a_k}(a_j) &= \langle Q_{a_1a_2\ldots a_k}f, K_{a_j}\rangle \\ &= \langle f, Q_{a_1a_2\ldots a_k}K_{a_j}\rangle \\ &= \langle f, (I - P_{a_1a_2\ldots a_k})K_{a_j}\rangle \\ &= \langle f, 0\rangle \\ &= 0. \end{aligned}$$

Let a_j have multiplicity $l(a_j) > 1$, and $a_{s_1} = a_{s_2} = \cdots = a_{s_{l(a_j)}} = a_j$. For $m = 1, \cdots, l(a_j)$, in view of (2.11),

$$\begin{aligned} \left(\frac{d}{dz}\right)^{m-1}[f_{a_1a_2\ldots a_k}](a_j) &= \langle \left(\frac{d}{dz}\right)^{m-1} Q_{a_1a_2\ldots a_k}f, K_{a_j}\rangle \\ &= \langle f, Q_{a_1a_2\ldots a_k}\left(\frac{d}{d\overline{w}}\right)^{m-1} K_w(z)|_{w=a_j}\rangle \\ &= \langle f, (I - P_{a_1a_2\ldots a_k})\tilde{K}_{a_j}\rangle \\ &= 0. \end{aligned}$$

So, $f_{a_1a_2\ldots a_k}$ has $l(a_j)$-multiple zero at a_j. $\square$

The new methodology of this paper involves *higher order generalized backward shift operator*

$$\frac{Q_{a_1\cdots a_{m-1}}(f)(z)}{\phi_{a_1\cdots a_{m-1}}(z)}$$

and its factorization (see Lemma 3.2 below). The *order-1 generalized backward shift operator* $Q_a/[\frac{z-a}{1-\overline{a}z}]$ gives rise to what we call reduced remainder ([44]) that plays a central role in the formulation of AFD.

Lemma 3.2. *Let $a_1, \cdots, a_m$ be complex numbers in $\mathbf{D}$ allowing multiplicity. Then*

(1) In the general $\mathcal{H}$ space setting the Gram-Schmidt orthonormalization has the explicit representations

$$E_{a_1\cdots a_m}(z) = \frac{Q_{a_1\cdots a_{m-1}}(K_{a_m})(z)}{\|Q_{a_1\cdots a_{m-1}}(K_{a_m})\|} = \frac{\tilde{K}_{a_m}(z) - \sum_{l=1}^{m-1}\langle \tilde{K}_{a_m}, E_{a_1\cdots a_l}\rangle E_{a_1\cdots a_l}(z)}{\|\tilde{K}_{a_m} - \sum_{l=1}^{m-1}\langle \tilde{K}_{a_m}, E_{a_1\cdots a_l}\rangle E_{a_1\cdots a_l}\|}. \tag{3.17}$$

(2) In the Hardy space case, with the notations defined in (1.5), (1.3) and (1.4), the function given by (3.17) is identical with $\mathrm{e}^{ic}e_{a_m}(z)\phi_{a_1\cdots a_{m-1}}(z)$, where

$$\mathrm{e}^{ic} = \frac{\overline{\phi}_{a_1\cdots a_{m-1}}(a_m)}{|\overline{\phi}_{a_1\cdots a_{m-1}}(a_m)|}.$$

There further hold the relations

$$\begin{aligned} \mathrm{e}^{ic}e_{a_m}(z)\phi_{a_1\cdots a_{m-1}}(z) &= \mathrm{e}^{ic}B_{a_1\cdots a_m}(z), \\ \langle f, E_{a_1\cdots a_m}\rangle E_{a_1\cdots a_m} &= \langle f, B_{a_1\cdots a_m}\rangle B_{a_1\cdots a_m}, \end{aligned} \tag{3.18}$$

and

$$\langle f, E_{a_1 \cdots a_m} \rangle = \frac{Q_{a_1 \cdots a_{m-1}}(f)(a_m)}{\phi_{a_1 \cdots a_{m-1}}(a_m)} \sqrt{1 - |a_m|^2}. \tag{3.19}$$

(3) In the general $\mathcal{H}$ space setting, the higher order generalized shift operators (to generalize the reduced remainders) can be factorized as

$$\frac{Q_{a_1 \cdots a_{m-1}}(f)(z)}{\phi_{a_1 \cdots a_{m-1}}(z)} = \left(\frac{Q_{a_{m-1}}}{\phi_{a_{m-1}}} \circ \cdots \circ \frac{Q_{a_1}}{\phi_{a_1}} \right) (f)(z). \tag{3.20}$$

Proof. The proofs of (1) and (2) are referred to [43]. We now prove (3). For $k > 1$, denote by g_k the *k-reduced remainder* ([44])

$$g_{k+1}(z) = \frac{g_k(z) - \langle g_k, E_{a_k} \rangle E_{a_k}(z)}{\phi_{a_k}(z)} = \left(\frac{Q_{a_k}}{\phi_{a_k}} \right) (g_k)(z),$$

where $g_1 = f$. Inductively there holds

$$g_{k+1}(z) = \left(\frac{Q_{a_k}}{\phi_{a_k}} \circ \frac{Q_{a_{k-1}}}{\phi_{a_{k-1}}} \right) (g_{k-1})(z) = \left(\frac{Q_{a_k}}{\phi_{a_k}} \circ \frac{Q_{a_{k-1}}}{\phi_{a_{k-1}}} \circ \cdots \frac{Q_{a_1}}{\phi_{a_1}} \right) (f)(z).$$

On the other hand, the AFD formulation given in [44] implies

$$Q_{a_1 \cdots a_k} f = g_{k+1} \prod_{j=1}^{k} \phi_{a_j}.$$

We thus have

$$g_{k+1} = \frac{Q_{a_1 \cdots a_k} f}{\prod_{j=1}^{k} \phi_{a_j}} = \left(\frac{Q_{a_k}}{\phi_{a_k}} \circ \frac{Q_{a_{k-1}}}{\phi_{a_{k-1}}} \circ \cdots \frac{Q_{a_1}}{\phi_{a_1}} \right) (f). \quad \square$$

Remark 3.3. It is recognized that the operator $\frac{Q_a}{\phi_a}$ is the generalized backward shift operator defined in [44]. Repeating use of the operator yields the reduced remainders $\frac{Q_{a_1 \cdots a_k} f}{\prod_{j=1}^{k} \phi_{a_j}}$. The following lemma shows that the reduced remainders of a function bounded by M are still bounded with explicit bounds in terms of M and the involved parameters $a_1, \cdots, a_k$. The result plays a crucial role in the proof of the main result of the paper.

Lemma 3.4. *Let f be an analytic function in an open neighborhood of $\overline{\mathbf{D}}$ and*

$$|f(z)| \leq M$$

for some $M > 0$ on $\overline{\mathbf{D}}$. Then for any sequence $a_1, a_2, ..., a_k \in \mathbf{D}$ the reduced remainder functions

$$\frac{f_{a_1 a_2 \ldots a_k}(z)}{\phi_{a_1 a_2 \ldots a_k}(z)}$$

are analytic in an open neighborhood of $\overline{\mathbf{D}}$ with the bounds over $\overline{\mathbf{D}}$:

$$\left| \frac{f_{a_1 a_2 \ldots a_k}(z)}{\phi_{a_1 a_2 \ldots a_k}(z)} \right| \leq M(1 + C_{\mathcal{H}})^k, \quad z \in \overline{\mathbf{D}}, \tag{3.21}$$

where $C_{\mathcal{H}}$ is the constant in (1.9).

Proof. For $a_1 \in \mathbf{D}$ and $z \in \partial\mathbf{D}$,

$$\begin{aligned}|f_{a_1}(z)| &= |f(z) - \langle f, E_{a_1}\rangle E_{a_1}(z)| \\ &\leq |f(z)| + |f(a_1)|\frac{|K_{a_1}(z)|}{\|K_{a_1}\|^2} \\ &\leq |f(z)| + |f(a_1)|C_{\mathcal{H}} \\ &\leq M(1 + C_{\mathcal{H}}).\end{aligned}$$

Since the zero of ϕ_{a_1} is a zero of $Q_{a_1}f$, $\frac{Q_{a_1}f}{\phi_{a_1}}$ is a holomorphic function in an open neighborhood of $\overline{\mathbf{D}}$. The maximum modulus principle over $\overline{\mathbf{D}}$ gives

$$\left|\frac{Q_{a_1}f(w)}{\phi_{a_1}(w)}\right| \leq \max\{|f_{a_1}(z)| \ : \ z \in \partial\mathbf{D}\} \leq M(1 + C_{\mathcal{H}})$$

for all $w \in \overline{\mathbf{D}}$. By invoking the result of (iii), Lemma 3.2, and repeating k times the above estimation procedure for $\frac{Q_{a_1}f}{\phi_{a_1}}$, we obtain the bounds claimed in the statement of the lemma. □

We will first prove the following result which is directly related to the $n = 1$ case of existence of n-best kernel approximation, and on the other hand, it is by itself important.

Lemma 3.5. *With the general $\mathcal{H}$ space setting, Boundary Vanishing Condition (BVC) holds. That is,*

$$\lim_{|a| \to 1-} |\langle f, E_a\rangle| = 0. \tag{3.22}$$

Proof. Due to the assumption of the underlying Hilbert space the given function $f \in \mathcal{H}$ may be approximated in energy within an error $\epsilon > 0$ by a bounded holomorphic function g as a linear combination of some parameterized reproducing kernels. By the Cauchy-Schwarz inequality, we have

$$|\langle f, E_a\rangle| \leq |\langle f - g, E_a\rangle| + |\langle g, E_a\rangle| \leq \|f - g\| + \frac{|g(a)|}{\|K_a\|} \leq \epsilon + \frac{|g(a)|}{\|K_a\|}.$$

As a consequence of (1.8) and boundedness of g, BVC (3.22) is concluded. □

For any zero set Z possibly with multiplicities we use the general notation $K_Z(z, a)$ for the reproducing kernel at a of the zero space $\mathcal{H}_Z$, where

$$\mathcal{H}_Z = \{f \in \mathcal{H} \ : \ f \text{ vanishes at points in } Z \text{ including multiplicities}\}.$$

The space $\mathcal{H}_Z$ uses the same inner product as $\mathcal{H}$. We denote by $\mathcal{H}_{\phi_Z}$ the Hilbert space

$$\mathcal{H}_{\phi_Z} = \{f : \mathbf{D} \to \mathbf{C} \ : \ f \text{ is analytic}, \ \|f\phi_Z\|_{\mathcal{H}} < \infty\},$$

where ϕ_Z is the canonical Blaschke product generated by the elements of Z including multiplicities. The inner product of $\mathcal{H}_{\phi_Z}$ is denoted as $\langle \cdot, \cdot\rangle_{\mathcal{H}_{\phi_Z}}$. The reproducing kernel of $\mathcal{H}_{\phi_Z}$ is denoted K_{ϕ_Z}. In this paper we only need to treat zero sets Z with finite points. Note that $\mathcal{H} \subset \mathcal{H}_{\phi_Z}$, and $\|f\|_{\mathcal{H}_{\phi_Z}} \leq \|f\|_{\mathcal{H}}$.

The next two lemmas follow similar idea in [28,67,68,20,51].

Lemma 3.6. *For any finite zero set Z, by denoting $K_Z(z, w)$ the reproducing kernel of the zero space $\mathcal{H}_Z$, there holds*

$$K_Z(z,w) = \phi_Z(z) K_{\phi_Z}(z,w) \overline{\phi_Z(w)} \tag{3.23}$$

and

$$\|K_Z(\cdot,w)\|_{\mathcal{H}} \le \|K_{\phi_Z}(\cdot,w)\|_{\mathcal{H}_{\phi_Z}}, \tag{3.24}$$

where ϕ_Z is the canonical Blaschke product defined by Z, $\mathcal{H}_{\phi_Z}$ is the $|\phi_Z|^2$-weighted $\mathcal{H}$ space, K_{ϕ_Z} is its reproducing kernel. As a consequence, the normalized reproducing kernel is

$$\frac{K_Z(z,w)}{\|K_Z(\cdot,w)\|} = \frac{\overline{\phi_Z(w)}}{|\phi_Z(w)|} \frac{\phi_Z(z) K_{\phi_Z}(z,w)}{K^{1/2}_{\phi_Z}(w,w)}. \tag{3.25}$$

Proof. We note that $K_Z(z,w)$ has zero set Z for the variable z when w is fixed in a neighborhood of $\overline{\mathbf{D}}$; and zero set Z for the variable w when z is fixed in a neighborhood of $\overline{\mathbf{D}}$. Therefore, $\phi_Z^{-1}(z)K_Z(z,w)\overline{\phi}_Z^{-1}(w)$ is analytic for z in a neighborhood of $\overline{\mathbf{D}}$, and anti-analytic for w in a neighborhood of $\overline{\mathbf{D}}$. Let $f \in \mathcal{H}_{\phi_Z}$. In the case $f\phi_Z \in \mathcal{H}_Z$. We have

$$\begin{aligned}
\langle f, \phi_Z^{-1}(\cdot)K_Z(\cdot,w)\overline{\phi}_Z^{-1}(w)\rangle_{\mathcal{H}_{\phi_Z}} &= \phi_Z^{-1}(w)\langle f\phi_Z, K_Z(\cdot,w)\rangle_{\mathcal{H}_Z} \\
&= \phi_Z^{-1}(w)\langle f\phi_Z, K_Z(\cdot,w)\rangle_{\mathcal{H}_Z} \\
&= \phi_Z^{-1}(w) f(w)\phi_Z(w) \\
&= f(w).
\end{aligned}$$

Therefore, $\mathcal{H}_{\phi_Z}$ is a RKHS. Due to uniqueness of reproducing kernel, its kernel $K_{\phi_Z}(z,w)$ satisfies the relation (3.23). To prove (3.24) we have

$$\begin{aligned}
\|K_Z(\cdot,w)\|^2_{\mathcal{H}} &= K_Z(w,w) \\
&= \phi_Z(w)K_{\phi_Z}(w,w)\overline{\phi_Z(w)} \\
&= K_{\phi_Z}(w,w)|\phi_Z(w)|^2 \\
&\le K_{\phi_Z}(w,w) \\
&= \|K_{\phi_Z}(\cdot,w)\|^2_{\mathcal{H}_{\phi_Z}}.
\end{aligned}$$

The relation (3.25) is just by dividing $K_Z(z,w)$ with $\|K_Z(\cdot,w)\| = \sqrt{K_Z(w,w)}$ and invoking (3.23). □

Lemma 3.7. *For any Hilbert space $\mathcal{H}$ with reproducing kernel K and any $a \in \mathbf{D}$ there hold*

$$K(a,a) = \sup\{|f(a)|^2 \ : \ f \in \mathcal{H}, \|f\| \le 1\} \tag{3.26}$$

and

$$K(a,a) \le K_{\phi_Z}(a,a). \tag{3.27}$$

Proof. Recall that $E_a(z) = K(z,a)/\|K_a\|$. On one hand, $\|E_a\| = 1$. On the other hand, for any f satisfying $\|f\| = 1$, using the Cauchy-Schwarz inequality,

$$|f(a)|^2 = |\langle f, K_a\rangle|^2 \le \|K_a\|^2 = K(a,a).$$

So, E_a is a solution for the extremal problem. Using this argument also to $\mathcal{H}_{\phi_Z}$ and K_{ϕ_Z}, we obtain

$$\begin{aligned} K(a,a) &= \sup\{|f(a)|^2 \ : \ f \in \mathcal{H}, \|f\|_{\mathcal{H}} \le 1\} \\ &\le \sup\{|f(a)|^2 \ : \ f \in \mathcal{H}_{\phi_Z}, \|f\|_{\mathcal{H}_{\phi_Z}} \le 1\} \\ &= K_{\phi_Z}(a,a), \end{aligned}$$

as desired. □

4. The proof of Theorem 2.1

The proof of Theorem 2.1 will not be according to mathematical induction. The case $n = 1$, therefore, does not need to be firstly addressed. Since the proof for $n = 1$ contains a basic idea for general n, we incorporate its proof for easy reference later.

To prove existence of 1-best approximation amounts to finding $a_1 \in \mathbf{D}$ such that

$$|\langle f, E_{a_1}\rangle| = \frac{|f(a_1)|}{\|K_{a_1}\|}$$

attains the global maximum value over all possible choices of the parameter in $\mathbf{D}$ (this is what we called Maximal Selection Principle (MSP) in the previous relevant studies ([44,48,49])). In fact, if f is not identical with the zero function, due to density of the span of the parameterized reproducing kernels, there exists $b \in \mathbf{D}$ such that $|\langle f, E_b\rangle| > 0$. Denote $|\langle f, E_b\rangle| = \delta$. BVC proved in Lemma 3.5 amounts that there exists $0 < r_1 < 1$ such that $|a| > r_1$ implies $|\langle f, E_a\rangle| < \delta/2$. Therefore,

$$\max\{|\langle f, E_a\rangle| \ : \ |a| \le r_1\} = \sup\{|\langle f, E_a\rangle| \ : \ a \in \mathbf{D}\}. \tag{4.28}$$

Note that $|\langle f, E_a\rangle|$ is a continuous function in a defined in the compact set $\{a \in \mathbf{C} \ : \ |a| \le r_1\}$. Thus the global maximum of $|\langle f, E_a\rangle|$ in $\mathbf{D}$ is attainable, and, in fact, in $\{a \in \mathbf{C} \ : \ |a| \le r_1\}$.

Next we turn to the general $n \ge 1$ cases. For the $n = 1$ case, all the extra cumbersome such as those in the equality and inequality chain (4.31) vanish. We first note that $(a_1, \cdots, a_n) \in \mathbf{D}^n$ gives rise to equality (2.16) if and only if

$$\sum_{l=1}^{n} |\langle f, E_{a_1\cdots a_l}\rangle|^2 = \sup\{\sum_{l=1}^{n} |\langle f, E_{b_1\cdots b_l}\rangle|^2, (b_1, \cdots, b_n) \in \mathbf{D}^n\}. \tag{4.29}$$

To begin with the proof we assume that f itself is not expressible by a linear combination of m_1 reproducing kernels for any $m_1 < n$. Based on definition of supreme, one can find a sequence of n-tuples $(a_1^{(l)}, \cdots, a_n^{(l)}), l = 1, 2, \cdots$, with mutually distinct and non-zero components, that corresponds to a sequence of n-tuples of reproducing kernels $(K_{a_1^{(l)}}, \cdots, K_{a_n^{(l)}})$, such that the square norms of the projections $P_{a_1^{(l)}\cdots a_n^{(l)}}(f)$ tend to the supreme (4.29). The distinct and non-zero requirements can be met owing to continuity of the inner product. Since $(a_1^{(l)}, \cdots, a_n^{(l)}) \in \overline{\mathbf{D}}^n$, through a Bolzano-Weierstrass compact argument, we may assume, without loss of generality, that the sequence of the n-tuples $(a_1^{(l)}, \cdots, a_n^{(l)})$ itself converges to $(a_1, \cdots, a_n) \in \overline{\mathbf{D}}^n$. If we have $a_1, \cdots, a_n$ all in $\mathbf{D}$, then we are done due to continuity of the inner product, with multiple kernels when multiplicities occur. This gives rise to the case (1) of the Theorem 2.1 for $m_1 = n$ when $d_f = 0$; and the case (2) when $d_f > 0$.

Now we deduce that if not all the limiting points $a_1, \cdots, a_n$ locate within $\mathbf{D}$, then there will hold the case (1) for some $m_1 < n$, being contrary with our assumption beneath (4.29). Assume that at least one of $a_1, \cdots, a_n$ are on the boundary $\partial\mathbf{D}$. Since the projections $P_{a_1^{(l)}\cdots a_n^{(l)}}(f)$ and $Q_{a_1^{(l)}\cdots a_n^{(l)}}(f)$ are irrelevant with the order, by re-ordering when necessary, we may assume without loss of generality that $a_1, \cdots, a_{m_1}$ are in $\mathbf{D}$, and $a_{m_1+1}, \cdots, a_n$ are on $\partial\mathbf{D}$, with $m_1 < n$, and in particular, $\lim_{l\to\infty} |a_n^{(l)}| = 1$. In the case we will show

$$\lim_{l\to\infty} |\langle f, E_{a_1^{(l)}a_2^{(l)}\dots a_n^{(l)}}\rangle| = 0, \tag{4.30}$$

regardless the locations of $a_k^{(l)}, k = 1, \cdots, n-1$ and $l = 1, 2, \cdots$. If (4.30) can be proved, by repeating the same argument $n - m_1$ times we result in that the latter $n - m_1$ rows of the l-sequence of the n-tuples all have no contribution. We show, in the case, $d_f > 0$ cannot hold. For, if $d_f > 0$ held, then d_f had rooms to be further reduced involving more reproducing kernels: Like what we did in proving the case $n = 1$, a_{m_1+1} can now be selected, due to the density of the span of the parameterized reproducing kernels, such that $a_{m_1+1}^{(l)} = a_{m_1+1}, l = 1, \cdots$, and $|\langle f - P_{a_1\cdots a_{m_1}} f, E_{a_{m_1+1}}\rangle| > 0$, and so on. This is contrary with d_f being the infimum. But, d_f cannot be zero either, for in such case we got that f is a linear combination of $m_1 < n$ multiple reproducing kernels, again contrary with our assumption. Thus, all that remain to be proved is (4.30). By using the same density argument as we prove BVC in Lemma 3.5 we may assume that f itself is an analytic function in a neighborhood of the closed unit disc $\overline{\mathbf{D}}$ with a bound M.

Now we proceed with the main technical step of the proof. Denote by $Z_{n-1}^{(l)}$ the l-level zero set $(a_1^{(l)}, a_2^{(l)}, ..., a_{n-1}^{(l)})$. With the above preparations we have

$$\begin{aligned}
&\langle f, E_{a_1^{(l)}a_2^{(l)}\dots a_n^{(l)}}\rangle_{\mathcal{H}} \\
&= \langle f_{a_1^{(l)}a_2^{(l)}\dots a_{n-1}^{(l)}}, E_{a_1^{(l)}a_2^{(l)}\dots a_n^{(l)}}\rangle_{\mathcal{H}} \qquad \left(E_{a_1^{(l)}a_2^{(l)}\dots a_n^{(l)}} = \frac{Q_{a_1^{(l)}a_2^{(l)}\dots a_{n-1}^{(l)}}(K_{a_n^{(l)}})}{\|Q_{a_1^{(l)}a_2^{(l)}\dots a_{n-1}^{(l)}}(K_{a_n^{(l)}}\|} = \frac{Q^2_{a_1^{(l)}a_2^{(l)}\dots a_{n-1}^{(l)}}(K_{a_n^{(l)}})}{\|Q_{a_1^{(l)}a_2^{(l)}\dots a_{n-1}^{(l)}}(K_{a_n^{(l)}}\|} \right) \\
&= \left\langle f_{a_1^{(l)}a_2^{(l)}\dots a_{n-1}^{(l)}}, \frac{K_{Z_{n-1}^{(l)}}(\cdot, a_n^{(l)})}{\|K_{Z_{n-1}^{(l)}}(\cdot, a_n^{(l)})\|} \right\rangle_{\mathcal{H}} \qquad \left(Q_{a_1^{(l)}a_2^{(l)}\dots a_{n-1}^{(l)}}(K_{a_n^{(l)}}) = K_{Z_{n-1}^{(l)}}(\cdot, a_n^{(l)}) \right) \\
&= \frac{\overline{\phi_{Z_{n-1}^{(l)}}(a_n^{(l)})}}{|\phi_{Z_{n-1}^{(l)}}(a_n^{(l)})|} \langle f_{a_1^{(l)}a_2^{(l)}\dots a_{n-1}^{(l)}}, K_{\phi_{Z_{n-1}^{(l)}}}(a_n^{(l)}, a_n^{(l)})^{-1/2} \phi_{Z_{n-1}^{(l)}} K_{\phi_{Z_{n-1}^{(l)}}}(\cdot, a_n^{(l)})\rangle_{\mathcal{H}} \qquad \text{(Lemma 3.6)} \\
&= \frac{\overline{\phi_{Z_{n-1}^{(l)}}(a_n^{(l)})}}{|\phi_{Z_{n-1}^{(l)}}(a_n^{(l)})|} \left\langle \frac{f_{a_1^{(l)}a_2^{(l)}\dots a_{n-1}^{(l)}}}{\phi_{Z_{n-1}^{(l)}}}, |\phi_{Z_{n-1}^{(l)}}|^2 K_{\phi_{Z_{n-1}^{(l)}}}(\cdot, a_n^{(l)}) \right\rangle_{\mathcal{H}} \frac{1}{\sqrt{K_{\phi_{Z_{n-1}^{(l)}}}(a_n^{(l)}, a_n^{(l)})}} \qquad \text{(Lemma 3.6)} \\
&= \frac{\overline{\phi_{Z_{n-1}^{(l)}}(a_n^{(l)})}}{|\phi_{Z_{n-1}^{(l)}}(a_n^{(l)})|} \left\langle \frac{f_{a_1^{(l)}a_2^{(l)}\dots a_{n-1}^{(l)}}}{\phi_{Z_{n-1}^{(l)}}}, K_{\phi_{Z_{n-1}^{(l)}}}(\cdot, a_n^{(l)}) \right\rangle_{\mathcal{H}_{\phi_{Z_{n-1}^{(l)}}}} \frac{1}{\sqrt{K_{\phi_{Z_{n-1}^{(l)}}}(a_n^{(l)}, a_n^{(l)})}} \\
&= \frac{\overline{\phi_{Z_{n-1}^{(l)}}(a_n^{(l)})}}{|\phi_{Z_{n-1}^{(l)}}(a_n^{(l)})|} \frac{f_{a_1^{(l)}a_2^{(l)}\dots a_{n-1}^{(l)}}(a_n^{(l)})}{\phi_{Z_{n-1}^{(l)}}(a_n^{(l)})} \frac{1}{\sqrt{K_{\phi_{Z_{n-1}^{(l)}}}(a_n^{(l)}, a_n^{(l)})}} \qquad \text{(Lemma 3.6 and reproducing kernel property)}.
\end{aligned} \tag{4.31}$$

To conclude the theorem it is sufficient to show that the above quantity tends to zero along with $\mathbf{D} \ni a_n^{(l)} \to a_n \in \partial\mathbf{D}$ uniformly in $a_1^{(l)}, \cdots, a_{n-1}^{(l)} \in \mathbf{D}$ for $l = 1, 2 \cdots$ It then suffices to prove

1^o.

$$\frac{f_{a_1^{(l)}a_2^{(l)}\dots a_{n-1}^{(l)}}(a_n^{(l)})}{\phi_{Z_{n-1}^{(l)}}(a_n^{(l)})}$$

is bounded uniformly in $a_1^{(l)},\cdots,a_{n-1}^{(l)}$ and $a_n^{(l)}, l=1,2,\cdots$; and

2^o.

$$\lim_{l\to\infty} K_{\phi_{Z_{n-1}^{(l)}}}(a_n^{(l)},a_n^{(l)})=\infty$$

uniformly in $a_1^{(l)},\cdots,a_{n-1}^{(l)}, l=1,2,\cdots$

Now we show assertion 1^o. First by Lemma 3.1 the function

$$g_{Z_{n-1}^{(l)}}(z)=\frac{f_{a_1^{(l)}\cdots a_{n-1}^{(l)}}(z)}{\phi_{Z_{n-1}^{(l)}}(z)}$$

is analytic in a neighborhood of $\overline{\mathbf{D}}$. By invoking the maximum modulus principle for one complex variable in $\overline{\mathbf{D}}$, Lemma 3.4, as well as the fact that all finite Blaschke products are of modulus 1 on the boundary $\partial\mathbf{D}$, we have

$$\begin{aligned}\max\{|g_{Z_{n-1}^{(l)}}(z)| \ : \ z\in\overline{\mathbf{D}}\} &= \max\{|g_{Z_{n-1}^{(l)}}(\zeta)| \ : \ \zeta\in\partial\mathbf{D}\}\\ &= \max\{\left|f_{a_1^{(l)}\cdots a_{n-1}^{(l)}}(\mathrm{e}^{it})\right| \ : \ t\in\partial\mathbf{D}\}\\ &\le M(1+C_{\mathcal{H}})^{n-1},\end{aligned}$$

concluding the uniform boundedness claimed of 1^o. The assertion 2^o is a consequence of the condition (1.8) and Lemma 3.7.

5. Applications

5.1. The classical Hardy space

By taking $\mathcal{H}=\mathbb{H}^2(\mathbf{D})$, we are with the inner product (1.1), and the reproducing kernel $K(z,w)=k_w(z)=\frac{1}{1-\overline{w}z}$. Since

$$K(a,a)=\frac{1}{1-|a|^2}\to\infty \text{ as } |a|\to 1, \text{ and } \frac{|K_a(z)|}{K(a,a)}=\frac{1-|a|^2}{|1-\overline{a}z|}\le 2,$$

the conditions (1.8) and (1.9) are satisfied. We hence have existence of the n-best approximation.

5.2. The Bergman spaces

Let $\mathcal{H}$ be the weighted Bergman spaces with the definition and notation

$$\mathbb{A}^2_\alpha(\mathbf{D})=\{f:\mathbf{D}\to\mathbf{C} \mid f \text{ is holomorphic in } \mathbf{D}, \text{and } \|f\|^2_{\mathbb{A}^2_\alpha(\mathbf{D})}=\int_{\mathbf{D}}|f(z)|^2dA_\alpha<\infty\},$$

where $\alpha\in(-1,\infty), dA_\alpha=(1+\alpha)(1-|z|^2)^\alpha dA(z)$, and $dA=\frac{dxdy}{\pi}, z=x+iy$, is the normalized area measure of the unit disc. The inner product of $\mathbb{A}^2_\alpha(\mathbf{D})$ is defined as

$$\langle f,g\rangle_{\mathbb{A}^2_\alpha(\mathbf{D})}=\int_{\mathbf{D}} f(z)\overline{g(z)}dA_\alpha.$$

In the sequel we sometimes write $\|\cdot\|_{\mathbb{A}^2_\alpha(\mathbf{D})}$ and $\langle\cdot,\cdot\rangle_{\mathbb{A}^2_\alpha(\mathbf{D})}$ briefly as $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$, and $\mathbb{A}^2_\alpha(\mathbf{D})$ as $\mathbb{A}^2_\alpha$.

$\mathbb{A}^2_\alpha$ is a RKHS with reproducing kernel

$$k_a^\alpha(z) = \frac{1}{(1-\overline{a}z)^{2+\alpha}}.$$

By invoking the reproducing kernel property we have

$$\|k_a^\alpha\|^2 = k_a^\alpha(a) = \frac{1}{(1-|a|^2)^{2+\alpha}}. \tag{5.32}$$

This shows that the condition (1.8) holds for all $\mathbb{A}^2_\alpha$. A simple computation gives

$$\frac{k_a^\alpha(z)}{k_a^\alpha(a)} = \frac{(1-|a|^2)^{2+\alpha}}{(1-\overline{a}z)^{2+\alpha}} \le 2^{2+\alpha}.$$

Hence, the reproducing kernel satisfies the condition (1.9). Therefore, an n-best approximation exists in all the weighted Bergman spaces. This is a re-proof of the main result of [51].

5.3. The weighted Hardy spaces

Let $W(k)$ be a sequence of non-negative numbers satisfying $\lim_{k\to\infty} W(k)^{\frac{1}{k}} \ge 1$ ([33]). Denote by $\mathbb{H}_W(\mathbf{D})$ the W-weighted Hardy $\mathbb{H}^2$-space defined by

$$\mathbb{H}_W(\mathbf{D}) = \{f \;:\; \mathbf{D} \to \mathbf{C} \;:\; f(z) = \sum_{k=0}^\infty c_k z^k,\ z \in \mathbf{D}, \|f\|_{\mathbb{H}_W} = \sum_{k=0}^\infty W(k)|c_k|^2 < \infty\}.$$

We will be considering an ordered sequence of W-weighted Hardy spaces defined by the weights $W_\beta(k) = (1+k)^\beta, -\infty < \beta < \infty$. This class of function spaces is a generalization of the Hardy and the weighted Bergman spaces. In fact, $\mathbb{H}_{W_0}(\mathbf{D}) = \mathbb{H}^2(\mathbf{D})$, and $\mathbb{H}_{W_\beta}(\mathbf{D}) = \mathbb{A}^2_\alpha(\mathbf{D}), \alpha = -\beta - 1, \beta < 0\ (\alpha > -1)$, $\mathbb{H}_{W_{-1}}(\mathbf{D})$ is the standard Bergman, and $\mathbb{H}_{W_1}(\mathbf{D})$ is the Dirichlet space in $\mathbf{D}$. The spaces $\mathbb{H}_{W_\beta}(\mathbf{D})$ are, as a matter of fact, equivalent with the Hardy-Sobolev spaces $W^{\frac{\beta}{2},2}$. From the last two subsections we know that the spaces $\mathbb{H}_{W_\beta}(\mathbf{D}), \beta \le 0$, have n-best approximation. We now extend the result to $0 < \beta \le 1$.

The inner product of $\mathbb{H}_{W_\beta}(\mathbf{D}), -\infty < \beta < \infty$, is

$$\langle f, g\rangle = \sum_{k=0}^\infty (k+1)^\beta c_k \overline{d}_k,$$

where c_k and d_k are, respectively, the coefficients of the Taylor expansions of f and g. From this it can be directly verified that the reproducing kernel of $\mathbb{H}_{W_\beta}(\mathbf{D})$ is

$$k_a^\beta(z) = \sum_{k=0}^\infty \frac{(z\overline{a})^k}{(k+1)^\beta}.$$

The function

$$k_a^\beta(a) = \sum_{k=0}^\infty \frac{|a|^{2k}}{(k+1)^\beta} \tag{5.33}$$

is an increasing function in $|a|$, and for any large N,

$$\underline{\lim}_{|a|\to 1-} k_a^\beta(a) \geq \lim_{|a|\to 1-} \sum_{k=0}^{N} \frac{|a|^{2k}}{(k+1)^\beta} = \sum_{k=0}^{N} \frac{1}{(k+1)^\beta}.$$

Therefore, for all $\beta \leq 1$,

$$\lim_{|a|\to 1-} k_a^\beta(a) = \infty,$$

verifying (1.8). Next we show that the weighted Hardy spaces kernels k_a^β satisfy the condition (1.9). This requires to prove that the function $\frac{|k_a^\beta(z)|}{k_a^\beta(a)}$ is uniformly bounded in $a, z \in \mathbf{D}$. The following estimation uses the well known technique for summing up series of positive decreasing entries: If f is a positive decreasing function integrable over $(0,\infty)$, then

$$\int_1^\infty f(t)dt \leq \sum_{k=1}^{\infty} f(k) \leq \int_0^\infty f(t)dt.$$

The estimation amounts to numerically comparing some elementary integrals. Denote by $|a| = r < 1$. Then

$$\begin{aligned}
\frac{|k_a^\beta(z)|}{k_a^\beta(a)} &= \frac{|\sum_{k=0}^\infty \frac{(\overline{a}z)^k}{(1+k)^\beta}|}{\sum_{k=0}^\infty \frac{|a|^{2k}}{(1+k)^\beta}} \\
&\leq \frac{\sum_{k=0}^\infty \frac{r^k}{(1+k)^\beta}}{\sum_{k=0}^\infty \frac{r^{2k}}{(1+k)^\beta}} \\
&\leq \frac{\int_0^\infty \frac{r^x}{(1+x)^\beta}dx}{\int_1^\infty \frac{r^{2x}}{(1+x)^\beta}dx} \\
&= \frac{\int_0^\infty \frac{r^x}{(1+x)^\beta}dx}{2^{\beta-1}\int_2^\infty \frac{r^x}{(2+x)^\beta}dx} \qquad \text{(change of variable)} \\
&= \frac{\int_0^\infty \frac{r^x}{(1+x)^\beta}dx}{2^{\beta-1}\left(\int_0^\infty \frac{r^x}{(2+x)^\beta}dx - \int_0^2 \frac{r^x}{(2+x)^\beta}dx\right)} \qquad \left(\frac{\infty}{\infty} \text{ type when } \ r \to 1-\right) \\
&\leq \frac{\int_0^\infty \frac{r^x}{(1+x)^\beta}dx}{2^{\beta-2}\int_0^\infty \frac{r^x}{(1+x)^\beta}\frac{(1+x)^\beta}{(2+x)^\beta}dx} \qquad \text{(if } r \geq \text{ some } r_0 \in (0,1)) \\
&\leq 4. \qquad \text{(mean} - \text{value theorem of integration)}
\end{aligned}$$

For $r \leq r_0$ the estimated quantity also has a uniform bound. Hence, for $\beta \leq 1$, the spaces $\mathbb{H}_{W_\beta}(\mathbf{D})$ satisfy the three conditions (i), (ii), and (iii). By invoking Theorem 2.1 the n-best approximation problems have solutions in those spaces.

Remark 5.1. (For the spaces $\mathbb{H}_{W_\beta}(\mathbf{D}), \beta > 1$) The spaces $\mathbb{H}_{W_\beta}(\mathbf{D}), \beta > 1$, do not fall into the category governed by Theorem 2.1, as, owing to (5.33), the condition (1.8) is not satisfied. Recall the Sobolev Embedding theorem asserting that $W^{k,p} \subset C^{r,\alpha}$ if $m < pk, \frac{1}{p} - \frac{k}{m} = -\frac{r+\alpha}{m}, m$ is the dimension. In our case $m = 1, p = 2, k = \frac{\beta}{2}$, and, in particular, $\beta = pk > 1$. It hence concludes that the functions in the spaces $\mathbb{H}_{W_\beta}(\mathbf{D}), \beta > 1$, are all continuously extendable to the closed unit disc, and the norm square $\|k_a^\beta\|^2 = k_a^\beta(a)$ given by (5.33) does not have singularity for $|a| = 1$. Based on these, as well as continuity of the inner product, we conclude existence of n-best approximations of the spaces for $\beta > 1$.

The results of this section are summarized as

Theorem 5.2. *For all Hardy-Sobolev spaces $\mathbb{H}_{W_\beta}(\mathbf{D}), -\infty < \beta < \infty$, there exist solutions to the n-best kernel approximation problem.*

6. Stochastic n-best approximation

Let $(\Omega, \mathcal{F}, d\mathbb{P})$ be a probability space, and, as in the previous sections, $\mathcal{H}$ be a RKHS of analytic functions in $\mathbf{D}$ with reproducing kernel $K_w, w \in \mathbf{D}$. We will be studying stochastic signals $f(z,\xi)$, where $\xi \in \Omega$ and $z \in \mathbf{D}$: We assume that for a.s. $\xi \in \Omega, f(\cdot,\xi)$ is a function in $\mathcal{H}$; and, for a.e. $z \in \mathbf{D}, f(z,\cdot)$ is a random variable. We will use the notation $f_\xi(z) = f(z,\xi)$. Associated with the probability space and the RKHS we define a Bochner type space ([21])

$$L^2(\mathcal{H},\Omega) = \{f : \mathbf{D} \times \Omega \to \mathbf{C} \ : \ \|f\|^2_{L^2(\mathcal{H},\Omega)} = E_\xi \|f_\xi\|^2_{\mathcal{H}} < \infty\}, \tag{6.34}$$

where E_ξ denotes the expectation, and precisely,

$$E_\xi \|f_\xi\|^2_{\mathcal{H}} = \int\limits_\Omega \|f_\xi\|^2_{\mathcal{H}} d\mathbb{P}(\xi).$$

We often use the simplified notation $\mathcal{N} = L^2(\mathcal{H},\Omega)$.

The following theorem generalizes the existence result for stochastic n-best approximation for the Hardy space (see [50]) to the RKHSs satisfying the conditions (i), (ii), and (iii) as assumed in Theorem 2.1.

Theorem 6.1. *Let $\mathcal{H}$ be a RKHS of analytic functions in the unit disc satisfying the conditions (i), (ii), and (iii) in §1, and Ω a probability space. Let $\mathcal{N} = L^2(\mathcal{H},\Omega)$ be the associated Bochner type space as above defined. Let f be any non-zero random signal. Then for any positive integer n, either of the following two cases holds: (1) 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 identical with the orthogonal expansion*

$$f(z,\xi) = \sum_{k=1}^{m_1} \langle f_\xi, E_k \rangle_{\mathcal{H}} E_k(z); \tag{6.35}$$

or (2) There exists an n-tuple of constant parameters $(a_1, \cdots, a_n) \in \mathbf{D}^n$ such that

$$\|f - \sum_{k=1}^{n} \langle f_\xi, E_k \rangle_{\mathcal{H}} E_k \|_{\mathcal{N}} \tag{6.36}$$

attains its positive infimum over all possible n-orthonormal systems $\{E_k\}$, where in both cases, $\{E_k\}_{k=1}^m$ is the orthonormal system generated by $(\tilde{K}_{a_1}, \cdots, \tilde{K}_{a_m}), 1 \le m \le n$.

The proof of Theorem 2.1 of [50] for the stochastic Hardy space case cannot be directly adopted, for, in the present case no density argument based on the boundary value of the given function on $\partial\mathbf{D}$ is available. In §3 we established, by using a new technical method, the pointwise convergence result (4.30) which is crucial to in proving Theorem 6.1.

Proof. In the proof of Theorem 2.1 we already show that, for each ξ outside an event in Ω of probability zero, there holds uniformly in $a_1^{(l)}, a_2^{(l)}, ..., a_{n-1}^{(l)}, l = 1, 2, \cdots$, that

$$\lim_{l\to\infty} |\langle f_\xi, E_{a_1^{(l)} a_2^{(l)} \dots a_n^{(l)}} \rangle_{\mathcal{H}}|^2 = 0.$$

By using the Cauchy-Schwarz inequality for the space $\mathcal{H}$ we have a dominating function of the function sequence on the left hand side:

$$|\langle f_\xi, E_{a_1^{(l)} a_2^{(l)} \dots a_n^{(l)}} \rangle_{\mathcal{H}}|^2 \leq \|f_\xi\|^2 \in L^1(\Omega).$$

Then the Lebesgue dominated convergence theorem be invoked to conclude

$$\lim_{l\to\infty} E_\xi |\langle f_\xi, E_{a_1^{(l)} a_2^{(l)} \dots a_n^{(l)}} \rangle_{\mathcal{H}}|^2 = 0 \tag{6.37}$$

uniformly in $a_1^{(l)}, a_2^{(l)}, ..., a_{n-1}^{(l)}, l = 1, 2, \cdots$. Based on this the contradiction argument used in proving Theorem 2.1 of [50] may be adopted to conclude the theorem. □

Remark 6.2. (Impact to Algorithm) This present paper only treats the existence aspect of the n-best problem. Based on the obtained estimates, however, a mathematical algorithm to actually get a solution is now on its way. We now cite the crucial step to reduce the problem to a global optimization one of a differential function defined in a compact set. Separate studies will be devoted to the computation aspect. To have an n-best approximation algorithm we are under the assumption that the given function f is not expressible by any m-linear combination of multiple kernels for $m \leq n-1$. This implies that $d_f(n-1) > 0$, and there exists an n-tuple $(b_1, \cdots, b_n) \in \mathbf{D}^n$ such that for some $\epsilon > 0$

$$\|f - P_{b_1 \cdots b_n} f\| = d_f(n-1) - \epsilon. \tag{6.38}$$

By using (4.30) one can find $\delta > 0$ such that if $|a_n| > 1 - \delta$, then

$$\|f - P_{a_1 \cdots a_{n-1} a_n} f\| > d_f(n-1) - \epsilon$$

for any $a_1, \cdots, a_{n-1}$ in $\mathbf{D}$. Since $P_{a_1 \cdots a_{n-1} a_n} f$ is symmetric in $a_1, \cdots, a_n$, we conclude

$$\|f - P_{a_1 \cdots a_{n-1} a_n} f\| > d_f(n-1) - \epsilon$$

whenever $|a_k| > 1 - \delta$ for some $k = 1, \cdots, n$. Owing to the observation (6.38) we have $d_f(n) \leq d_f(n-1) - \epsilon$. When the infimum attains at $(\tilde{a}_1, \cdots, \tilde{a}_{n-1}, \tilde{a}_n)$, that is,

$$\|f - P_{\tilde{a}_1 \cdots \tilde{a}_{n-1} \tilde{a}_n} f\| = d_f(n),$$

there holds $|\tilde{a}_k| \leq 1 - \delta$ for all $k = 1, \cdots, n$. This concludes that the global minimum value $d_f(n)$ is only attainable in the compact set $\overline{(1-\delta)\mathbf{D}}^n$. In view of the relation (6.37) the same conclusion holds for the stochastic case.

Declaration of competing interest

This work does not have any conflicts of interest.

Data availability

No data was used for the research described in the article.

Acknowledgment

The author wishes to express his sincere thanks to Y.B. Wang, K.H. Zhu and W. Qu for the previous works done with the author, as well as for their inspiriting discussions on this and related subjects.

References

[1] D. Alpay, F. Colombo, T. Qian, I. Sabadini, Adaptive orthonormal systems for matrix-valued functions, Proc. Am. Math. Soc. 145 (5) (2017) 2089–2106.
[2] D. Alpay, F. Colombo, T. Qian, I. Sabadini, Adaptative decomposition: the case of the Drury-Arveson space, J. Fourier Anal. Appl. 23 (6) (2017) 1426–1444.
[3] D. Alpay, H. Dym, On a new class of realization formulas and their application, Linear Algebra Appl. 241 (1996) 3–84.
[4] H. Akcay, B. Ninness, Orthonormal basis functions for modelling continuous-time system, Signal Process. 77 (3) (1999) 261–274.
[5] J.A. Ball, V. Bplotnikov, Weighted Begman spaces: shift invariant subspaces and input/state/output linear systems, Integral Equ. Oper. Theory 76 (3) (2013) 301–356.
[6] J.A. Ball, V. Bplotnikov, Weighted Hardy spaces: shift invariant and coinvariant subspaces, linear system and operator model theory, arXiv:1405.2974 [math.CA].
[7] S. Bochner, Vorlesungen Ueber Fouriersche Integrale, Leipzig, 1932.
[8] H.H. Bauschke, P.L. Combettes, D.R. Luke, Finding best approximation pairs relative to two closed convex sets in Hilbert spaces, J. Approx. Theory 127 (2) (2004) 178–192.
[9] L. Baratchart, M. Cardelli, M. Olivi, Identification and rational L^2 approximation-a gradient algorithm, Automatica 27 (2) (1991) 413–417.
[10] D.W. Byun, S. Saitoh, Best approximation in reproducing kernel Hilbert spaces, in: Second International Colloquium on Numerical Analysis, 1994, p. 55.
[11] L. Baratchart, E.B. Staff, F. Wielonsky, A criterion for uniqueness of a critical point in H^2 rational approximation, J. Anal. Math. 70 (1) (1996) 225–266.
[12] N.I. Chernykh, Best approximation of periodic functions by trigonometric polynomials in L^2, Math. Notes Acad. Sci. USSR 2 (5) (1967) 803–808.
[13] M.M. Chawla, V. Kaul, Optimal rules with polynomial precision for Hilbert spaces possessing reproducing kernel functions, Numer. Math. 22 (3) (1974) 207–218.
[14] E.W. Cheney, H.L. Loeb, Generalized rational approximation, J. Soc. Ind. Appl. Math. 1 (1) (1964) 11–25.
[15] R. Coifman, S. Steinerberger, Nonlinear phase unwinding of functions, J. Fourier Anal. Appl. 23 (4) (2017) 778–809.
[16] R. Coifman, J. Peyriére, Phase unwinding, or invariant subspace decompositions of Hardy spaces, J. Fourier Anal. Appl. 25 (2019) 684–695.
[17] J.S. Chen, S. Yoon, H.P. Wang, W.K. Liu, An improved reproducing kernel particle method for nearly incompressible finite elasticity, Comput. Methods Appl. Mech. Eng. 181 (1–3) (2000) 117–145.
[18] F. Deutsch, Best Approximation in Inner Product Spaces, Springer, 2001.
[19] J. Descloux, Approximations in L^p and Chebyshev approximations, J. Soc. Ind. Appl. Math. 11 (4) (1963) 1017–1026.
[20] P. Duren, A. Schuster, S. Alexander, Bergman Spaces, American Mathematical Society, 2004.
[21] L.C. Evans, Partial Differ. Equ. 19 (4) (1998) 7.
[22] C. Gu, Penalized likelihood estimation: convergence under incorrect model, Stat. Probab. Lett. 36 (4) (1998) 359–364.
[23] F. Gunther, W.K. Liu, Implementation of boundary conditions for meshless methods, Comput. Methods Appl. Mech. Eng. 163 (1–4) (1998) 205–230.
[24] F. Gunther, W.K. Liu, D. Diachin, M.A. Christon, Multi-scale meshfree parallel computations for viscous, compressible flows, Comput. Methods Appl. Mech. Eng. 190 (3–4) (2000) 279–303.
[25] H. Hult, Approximating some Volterra type stochastic integrals with applications to parameter estimation, Stoch. Process. Appl. 105 (1) (2003) 1–32.
[26] W. Han, X. Meng, Error analysis of the reproducing kernel particle method, Comput. Methods Appl. Mech. Eng. 190 (46–47) (2001) 6157–6181.
[27] R. Khalil, W. Deeb, Best approximation in $L^p(\mu, X)$, II, J. Approx. Theory 59 (3) (1989) 296–299.
[28] B. Korenblum, H. Hedenmalm, K.H. Zhu, D. Békollé, Theory of Bergman spaces, Math. Intell. 27 (1) (2005) 85–86.
[29] F.M. Larkin, Optimal approximation in Hilbert spaces with reproducing kernel functions, Math. Comput. 24 (112) (1970) 911–921.
[30] W.K. Liu, Y. Chen, R.A. Uras, C.T. Chang, Generalized multiple scale reproducing kernel particle methods, Comput. Methods Appl. Mech. Eng. 139 (1–4) (1996) 91–157.
[31] S. Li, W.K. Liu, Moving least-square reproducing kernel method part II: Fourier analysis, Comput. Methods Appl. Mech. Eng. 139 (1–4) (1996) 159–193.
[32] S. Li, W.K. Liu, T. Belytschko, Moving least-square reproducing kernel methods (I) methodology and convergence, Comput. Methods Appl. Mech. Eng. 143 (1–2) (1997) 113–154.
[33] B.D. MacCluer, Elementary Functional Analysis, Springer, New York, 2009.
[34] 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.
[35] J. Mercer, Function of positive and negative and their connection with the theory of integral equation, Philos. Trans. R. Soc. Lond. Ser. A 209 (1909) 415–446.

[36] E.H. Moore, R.W. Barnard, General Analysis I, The American Philosophical Society, 1935.
[37] F.M. Larkin, Optimal approximation in Hilbert spaces with reproducing kernel functions, Math. Comput. 24 (112) (1970) 911–921.
[38] W. Mi, T. Qian, Frequency domain identification: an algorithm based on adaptive rational orthogonal system, Automatica 48 (6) (2012) 1154–1162.
[39] W. Mi, T. Qian, On backward shift algorithm for estimating poles of systems, Automatica 50 (4) (2014) 1603–1610.
[40] S. Mallat, Z. Zhang, Matching pursuit with time-frequency dictionaries, IEEE Trans. Signal Process. 41 (1993) 3397–3415.
[41] B.U. Park, W.C. Kim, M.C. Jones, On identity reproducing nonparametric regression estimators, Stat. Probab. Lett. 32 (3) (1997) 279–290.
[42] T. Qian, Two-dimensional adaptive Fourier decomposition, Math. Methods Appl. Sci. 39 (10) (2016) 2431–2448.
[43] T. Qian, Sparse representations of random signals, Math. Methods Appl. Sci. 45 (8) (2022) 4210–4230.
[44] T. Qian, Y.B. Wang, Adaptive Fourier series-a variation of greedy algorithm, Adv. Comput. Math. 34 (3) (2011) 279–293.
[45] T. Qian, E. Wegert, Optimal approximation by Blaschke forms, Complex Var. Elliptic Equ. 58 (1) (2013) 123–133.
[46] 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.
[47] T. Qian, X.Y. Wang, L.M. Zhang, MIMO frequency domain system identification using matrix-valued orthonormal functions, Automatica 133 (2021) 109882.
[48] W. Qu, P. Dang, Rational approximation in a class of weighted Hardy spaces, Complex Anal. Oper. Theory 13 (4) (2019) 1827–1852.
[49] W. Qu, P. Dang, Reproducing kernel approximation in weighted Bergman spaces: algorithm and applications, Math. Methods Appl. Sci. 42 (12) (2019) 4292–4304.
[50] W. Qu, T. Qian, G.T. Deng, A stochase sparse representation: n-best approximation to random signals and computation, Appl. Comput. Harmon. Anal. 55 (1) (2021) 185–198.
[51] W. Qu, T. Qian, H.C. Li, K.H. Zhu, Best kernel approximation in Bergman spaces, Appl. Math. Comput. (2021).
[52] W. Qu, C.K. Chui, G.T. Deng, T. Qian, Sparse approximation to the Dirac-δ distribution, Anal. Appl. (2021).
[53] I. Singer, The Theory of Best Approximation and Functional Analysis, Society for Industrial and Applied Mathematics, 1974, pp. 259–269.
[54] K. Seip, Interpolation and Sampling in Spaces of Analytic Functions, American Mathematical Soc., 2004.
[55] S.A. Sahab, M.S. Khan, S. Sessa, A result in best approximation theory, J. Approx. Theory 55 (3) (1988) 349–351.
[56] D.J. Strauss, G. Steidl, Hybrid wavelet-support vector classification of waveforms, J. Comput. Appl. Math. 148 (2) (2002) 375–400.
[57] L.N. Trefethen, Approximation Theory and Approximation Practice, extended edition, Society for Industrial and Applied Mathematics, 2019.
[58] V.N. Temlyakov, Greedy algorithm and m-term trigonometric approximation, Constr. Approx. 107 (2009) 971–980.
[59] V.N. Temlyakov, Greedy Approximation, Cambridge University Press, 2011.
[60] T.E. Voth, M.A. Christon, Discretization errors associated with reproducing kernel methods: one-dimensional domains, Comput. Methods Appl. Mech. Eng. 190 (18–19) (2001) 2429–2446.
[61] Y.B. Wang, T. Qian, Pseudohyperbolic distance and n-best rational approximation, Math. Methods Appl. Sci., https://doi.org/10.22541/au.160630203.39582717/v1.
[62] J.L. Walsh, The existence of rational functions of best approximation, Trans. Am. Math. Soc. 33 (3) (1931) 668–689.
[63] J.L. Walsh, Interpolation and Approximation by Rational Functions in the Complex Plane, AMS, 1965.
[64] Y.B. Wang, T. Qian, Pseudo-hyperbolic distance and n-best rational approximation in H^2 space, Math. Methods Appl. Sci. 44 (11) (2021) 8497–8504.
[65] F. Yang, M. Chen, J.C. Chen, Y. Zhang, T. Qian, Sparse representations of solutions to a class of random boundary value problems, arXiv:submit/4064143 [math.GM].
[66] D.X. Zhou, Capacity of reproducing kernel spaces in learning theory, IEEE Trans. Inf. Theory 49 (7) (2003) 1743–1752.
[67] K.H. Zhu, Interpolating and recapturing in reproducing Hilbert spaces, Bull. Hong Kong Math. Soc. 1 (1997) 21–33.
[68] K.H. Zhu, Operator Theory in Function Spaces, American Mathematical Society, 2007, p. 138.
[69] S. Zaremba, Sur le calcul numerique des fonctions demandees dans le probleme de Dirichlet et le probleme hydrodynamique, Bull. Int. I Acad. Sci. Cracovie (1908) 125–195.