A sufficient condition for \(n\)-Best Kernel Approximation in Reproducing Kernel Hilbert Spaces

Wei Qu\(^1\), Tao Qian \(^*\)\(^2\), and Guan-Tie Deng\(^3\)

\(^1\)School of Mathematical Sciences, Beijing Normal University, Beijing, China
\(^2\)Macau Center for Mathematical Sciences, Macau University of Science and Technology, Macau, China
\(^3\)School of Mathematical Sciences, Beijing Normal University, Beijing, China

June 8, 2020

Abstract

We show that if a reproducing kernel Hilbert space \(H_K\), consisting of functions defined on \(E\), enjoys Double Boundary Vanishing Condition (DBVC) and Linear Independent Condition (LIC), then for any preset natural number \(n\), and any function \(f \in H_K\), there exists a set of \(n\) parameterized multiple kernels \(\tilde{K}_{w_1}, \ldots, \tilde{K}_{w_n}\) \(w_k \in E, k = 1, \ldots, n\), and real (or complex) constants \(c_1, \ldots, c_n\), giving rise to a solution of the optimization problem

\[
\|f - \sum_{k=1}^{n} c_k \tilde{K}_{w_k}\| = \inf \{\|f - \sum_{k=1}^{n} d_k v_k\| \mid v_k \in E, d_k \in \mathbb{R} \text{ (or } \mathbb{C}), k = 1, \ldots, n\}.
\]

By applying the theorem of this paper we show that the Hardy space and the Bergman space, as well as all the weighted Bergman spaces in the unit disc all possess \(n\)-best approximations. In the Hardy space case this gives a new proof of a classical result. Based on the obtained results we further prove existence of \(n\)-best spherical Poisson kernel approximation to functions of finite energy on the real-spheres.

MSC: 41A20; 41A65; 46E22; 30H20

keywords: Reproducing Kernel Hilbert Space, Double Boundary Vanishing Condition, \(n\)-Linearly Independent Condition, Hardy Space, Bergman space, Approximation by Rational functions of Certain Degrees

1 Introduction

Let \(H\) be a complex Hilbert space consisting of functions defined in a topological space \(E\). Assume that the point evaluation functional \(f(z)\) for any fixed \(z \in E\) is a bounded linear functional, i.e.,

\[
|f(z)| \leq C_z \|f\|,
\]
where $C_z$ is a constant depending on $z$. Then, according to the Riesz representation theorem there is a function $K_z$ with $z$ being a parameter such that

$$f(z) = \langle f, K_z \rangle,$$

for all $z \in E$. In such case we say that $H$ is a reproducing kernel Hilbert space, abbreviated as RKHS, call $K_z$ the reproducing kernel of $H$. Denote by $H_K(E)$ the Hilbert space $H$ whose corresponding reproducing kernel function is $K_z$. Indeed, any RKHS can have only one reproducing kernel. A wide class of Hilbert spaces, including the classical Hardy $H^2$-spaces, Bergman spaces, weighted Bergman spaces, and Sobolev spaces, etc., belong to the category of reproducing kernel Hilbert spaces (RKHSs).

The subject $n$-best approximation in reproducing kernel Hilbert spaces include, as a particular case, the one called best approximations to Hardy space functions by rational functions of order not exceeding $n$. The present study amounts to extending the question and solving it in a wide class of Hilbert spaces.

In below we first provide an account of the related concepts in the classical Hardy space of the unit disc. Denote by $C$ the complex plane and $D$ the open unit disc in $C$. The Hardy space in the unit disc is defined, among other equivalent definitions,

$$H^2(D) = \{ f : D \to C \mid f(z) = \sum_{k=0}^{\infty} c_k z^k, \sum_{k=0}^{\infty} |c_k|^2 < \infty \}.$$

It is a basic property of the Hardy space that for any $f \in H^2(D)$ there exists a boundary limit function, denoted $f(\bar{w}) \in L^2(\partial D)$, in both the pointwise non-tangential limit sense and in the $L^2$-convergence sense as well. It is standard knowledge that under the inner product

$$\langle f, g \rangle = \frac{1}{2\pi} \int_0^{2\pi} f(\bar{w})g(\bar{w})d\theta$$

the space $H^2(D)$ forms a Hilbert space.

Of particular importance in the Hardy space theory are the functions

$$k_w(\bar{w}) = \frac{1}{1-\bar{w}w} \quad \text{and} \quad e^{\frac{H^2}{2}}(z) = k_w z ||k_w|| = \sqrt{\frac{1-|w|^2}{1-\bar{w}z}}.$$

The function $k_w(z)$, where $w$ is considered as a parameter, is the reproducing kernel of the Hardy space $H^2(D)$; By invoking the Cauchy formula it follows that for any $f \in H^2(D)$ there holds

$$\langle f, k_w \rangle = f(w),$$

and, subsequently,

$$\langle f, \left( \frac{\partial}{\partial \bar{w}} \right)^l k_w \rangle = f^{(l)}(w), \quad l = 1, 2, \cdots.$$

**Definition 1.1** For any $n$ complex numbers $(w_1, \cdots, w_n) \in D^n$ and $n$ complex numbers $(c_1, \ldots, c_n) \in C^n$, the function

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


is called an $n$-Blaschke form, and an $n$-degenerate Blaschke form if $c_n \neq 0$, where \( \{B_k\}_{k=1}^n \) is the \( n \)-Takenaka-Malmquist (\( n \)-TM) system generated by the sequence \((w_1, \cdots, w_n)\),

\[
B_k(z) = \sqrt{\frac{1-|w_k|^2}{1-\overline{w}_k z}} \prod_{l=1}^{k-1} (z - w_l),
\]

We note that the \( n \)-TM system \( \{B_k\}_{k=1}^n \) is the orthonormalization of the \( n \)-system

\[
(\tilde{k}_w_1, \cdots, \tilde{k}_w_n),
\]

where

\[
\tilde{k}_{w_k}(z) \triangleq \left( \frac{d}{d w} \right)^{(l(w_k)-1)} k_w(z)|_{w=w_k},
\]

called the multiple reproducing kernels, where

\[
l(w_k) \triangleq \text{multiple number of } w_k \text{ in } (w_1, \cdots, w_k)
\]

\((10,17)\). Besides the multiple reproducing kernels we also use normalized multiple reproducing kernels

\[
\tilde{e}^{H_2}_{w_k}(z) = \frac{\tilde{k}_{w_k}(z)}{\|\tilde{k}_{w_k}(z)\|}.
\]

(1.2)

For fast expanding a given function into a TM system the adaptive Fourier decomposition (AFD) was proposed that is related to the Beurling-Lax decomposition of the Hardy space into the direct sum of the forward- and the backward-shift invariant subspaces \((I, II)\). AFD theory and algorithm have been generalized to matrix-valued functions defined in the disc \((I)\) and in the ball of several complex variables \((2)\).

The \( n \)-best rational approximation problem in the Hardy space is formulated as follows. A pair of polynomials \((p, q)\) is said to be \( n \)-admissible if it satisfies the following conditions: (i) \( p \) and \( q \) are co-prime; (ii) \( q \) does not have zeros in \( D \); and (iii) the both degrees of \( p \) and \( q \) are at most \( n \) \((10,18)\).

The \( n \)-best Rational Approximation Problem: For \( f \in H^2(D) \), find an \( n \)-admissible pair of polynomials \((p_1, q_1)\) such that

\[
\|f - \frac{p_1}{q_1}\|_{H^2(D)} = \inf \{\|f - \frac{p}{q}\|_{H^2(D)} \mid (p, q) \text{ is } n-\text{admissible}\}. \tag{1.3}
\]

The above optimization problem may be re-formulated as finding a non-degenerate Blaschke form

\[
\sum_{k=1}^{n} \langle f, B_{w_k}^w \rangle B_{w_k}^w(z),
\]

where the \( B_{w_k}^w \)'s correspond to \( w = (w_1, \cdots, w_n) \), such that

\[
\|f - \sum_{k=1}^{n} \langle f, B_{w_k}^w \rangle B_{w_k}^w \| = \inf \{\|f - \sum_{k=1}^{n} \langle f, B_{v_k}^v \rangle B_{v_k}^v\|_{H^2(D)} \mid v = (v_1, \cdots, v_n) \in D^n\}. \tag{1.4}
\]
Or, alternatively, we can ask the following question: Denotes by \( f / \text{span}\{\tilde{e}^{\mathbb{H}^2}_v, \cdots, \tilde{e}^{\mathbb{H}^2}_{vn}\} \) the projection of \( f \) into the span of \( \tilde{e}^{\mathbb{H}^2}_{v1}, \cdots, \tilde{e}^{\mathbb{H}^2}_{vn}, (v_1, \cdots, v_n) \in \mathbb{D}^n \). Find \((w_1, \cdots, w_n) \in \mathbb{D}^n\) such that

\[
\| f / \text{span}\{\tilde{e}^{\mathbb{H}^2}_{w1}, \cdots, \tilde{e}^{\mathbb{H}^2}_{wn}\} \|
\]

is maximized over all \((v_1, \cdots, v_n) \in \mathbb{D}^n\).

There have been several proofs in the literature for existence of the above specified \( n \)-best rational approximation problem in the classical Hardy spaces, see [15] (J. L. Walsh, 1962), [14] (G. Buckebusch, 1978), [3] (L. Baratchart). [10], [6]. In the last two articles the problem is reformulated in terms of \( n \)-Blaschke form. In the Hardy space case practical algorithms, including the INRIA method [19], cyclic AFD [18], and lately the gradient descent method in [11], can only claim to converge, in fact, to a local minimum. A mathematical algorithm to find the global minimum, is now still being sought.

The present paper works with the RKHS context. In a general RKHS one has a set of analogous formulation is equivalent with the following minimization problem: Find \((a_1, \cdots, a_n) \in \mathbb{E}^n\) that make the objective function

\[
A(f; a) = \| f - \sum_{k=1}^{n} \langle f, B_k^a \rangle_{H_K} B_k^a \|_{H_K}
\]

minimized, where

\[
\sum_{k=1}^{n} \langle f, B_k^a \rangle_{H_K} B_k^a
\]

is called the \( n \)-kernel orthonormal form of \( f \) corresponding to the \( n \)-tuple \((a_1, \cdots, a_n)\), where \((B_1^a, \cdots, B_n^a)\) is the G-S orthonormalization of the multiple kernels \((\tilde{E}_a, \cdots, \tilde{E}_a)\), \( k \leq n \). Note that the above formulation is equivalent with the following minimization problem: Find \((a_1, \cdots, a_n) \in \mathbb{E}^n\), \((c_1, \cdots, c_n) \in \mathbb{C}^n\), such that

\[
\| f - \sum_{k=1}^{n} c_k \tilde{K}_{ak} \| = \inf \{ \| f - \sum_{k=1}^{n} d_k \tilde{K}_{bk} \| \mid b_k \in \mathbb{E}, d_k \in \mathbb{C}, k = 1, \cdots, n \}.
\]

If for some \( d_k \) and \( b_k \)'s \( f = \sum_{k=1}^{m} d_k \tilde{K}_{bk} \), then \( f \) is said to be an \( m \)-kernel expansion.

We note that in the cases where the RKHS under study is the Hardy space inside the unit disc or the Hardy space in the upper-half complex plane, if \((B_1, \cdots, B_n)\) is the G-S orthonormalization of the \( n \)-tuple of the multiple reproducing kernels \((\tilde{e}^{\mathbb{H}^2}_{a1}, \cdots, \tilde{e}^{\mathbb{H}^2}_{an})\), then, by adding one more multiple reproducing kernel \( \tilde{e}^{\mathbb{H}^2}_{a_{n+1}} \) to the \( n \)-sequence, the corresponding \((n + 1)\)-orthonormalization system, \((B_1, \cdots, B_n, B_{n+1})\), is with the \((n + 1)\)-th term of the form \( B_{n+1} = \phi_n \tilde{e}^{\mathbb{H}^2}_{a_{n+1}} \), where \( \phi_n \) is the Blaschke product, unique up to a uni-modular constant, defined by the first \( n \) parameters \( a_1, \cdots, a_n \) as its
zeros, including the multiples. Indeed, TM systems are constructed in such way. In AFD, through a 
generalized backward shift operation, the TM systems are automatically generated ([9]). It is a 
question whether there exist other types RKHSs that possess such or similar constructive property. 
From our observation it seems that only the Hardy spaces of the classical domains possess such 
property (see [1, 2]). In the weighted Bergman spaces of the classical domains this property does not 
hold ([12, 13]).

As technical preparation we need to recall the so called \( \rho \)-weak pre-orthogonal adaptive Fourier 
decomposition (\( \rho \)-Weak-POAFD) developed in the general RKHS context. Assume that \( \mathcal{H} \) is a 
general Hilbert space with a dictionary parameterized by elements in \( \mathbb{E} \), denoted \( E_a, a \in \mathbb{E} \). Let \( \rho \in (0,1) \).

Suppose that we have obtained an \( n \)-term orthogonal expansion 
\[
f = \sum_{k=1}^{n} \langle f, B_k \rangle B_k + g_{n+1},
\]
where \( (B_1, \cdots, B_n) \) is the G-S orthonormalization of a selected \( n \)-sequence \( (a_1, \cdots, a_n) \), where the \( a_k \)'s are mutually different. Select \( a_{n+1} \), different from all the already selected \( a_k \)'s, \( k = 1, \cdots, n \), such that 
\[
|\langle f, B_{a_{n+1}} \rangle| \geq \rho \sup\{|\langle f, B_{b_{n+1}} \rangle| \mid b \in \mathbb{E}\},
\]
where for any \( b \in \mathbb{E}, (B_1, \cdots, B_n, B_{b_{n+1}}) \) is the G-S orthonormalization of \( (E_{a_1}, \cdots, E_{a_n}, E_b) \). Make such selections from \( n = 1 \) and for all consecutive \( n > 1 \), we obtain \( \rho \)-Weak-POAFD ([8, 12]).

Remark 1.2 \( \rho \)-Weak-POAFD is available for all RKHSs. The selection criterion (1.8) shows that it is a more optimal selection principle than the other types weak greedy algorithms in the classical 
literature ([5, 4]). When a dictionary satisfies BVC (see below), the selection corresponding to \( \rho = 1 \) is available, called POAFD. POAFD has the optimal maximal selection at each algorithm step ([8]). It reduces to AFD in the classical Hardy space ([9]).

In this paper we introduce what we call by Double Boundary Vanishing Condition (DBVC) that will play an important role in the \( n \)-best optimization problem. Assume that the parameters set \( \mathbb{E} \) is equipped with a topology. We used to work with the cases in which \( \mathbb{E} \) is a region (open and connected) of the complex plane \( \mathbb{C} \) or a region of the space of several complex variables \( \mathbb{C}^n \) under its natural 
topology. We now have the convention that, together with the finite boundary points, we add the infinite point, being included in the set of the boundary points if \( \mathbb{E} \) is unbounded, that corresponds to the one-point-compactification of the original topological space. The added point is denoted \( \infty \). Taking \( \mathbb{E} = \mathbb{C}^+ = \{z \in \mathbb{C} \mid \mathrm{Im}(z) > 0\} \) as an example, \( \mathbb{E} \) is equipped with the topology of \( \mathbb{C} \). By adding the \( \infty \) point, the sequence of open sets \( \{z \in \mathbb{C}^+ \mid \mathrm{Im}(z) < \frac{1}{n}, \mathrm{or} \mid z \mid > n\} \), were \( m, n \) are positive integers, forms a basis of open neighborhoods of \( \partial \mathbb{E} \). A RKHS is said to satisfy DBVC if for any sequence \( z_n \to \tilde{z} \in \mathbb{E} \) and \( w_n \to \tilde{w} \in \partial \mathbb{E} \), there holds 
\[
\lim_{n \to \infty} \langle E_{z_n}, E_{w_n} \rangle = 0.
\]
If DBVC holds, then we can show BVC (boundary vanishing condition) holds. That is, for any \( f \in H_K \) and \( w_n \to \tilde{w} \in \partial \mathbb{E} \), there holds 
\[
\lim_{n \to \infty} \langle f, E_{w_n} \rangle = 0.
\]
We have the following
Lemma 1.3 If $H_K$ is a RKHS satisfying DBVC, then it satisfies BVC.

Proof. Let $f \in H_K$. Since $H_K$ is a RKHS, by any type of the matching pursuit algorithm, including POAFD and Weak-POAFD, one can find $(a_1, \cdots, a_n, \cdots)$, consisting of mutually different terms in $E$, such that

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

where for any $n$, $(B_1, \cdots, B_n)$ is the G-S orthonormalization of some selected $(E_{a_1}, \cdots, E_{a_n}), n = 1, 2, \cdots$. Then, for any $\epsilon > 0$, one can find a natural number $N$ such that

$$\| f - \sum_{k=1}^{N} \langle f, B_k \rangle B_k \| \leq \epsilon/2.$$

By invoking the Cauchy-Schwarz inequality we have

$$| \langle f, E_{w_n} \rangle | = | \langle f - \sum_{k=1}^{N} \langle f, B_k \rangle B_k, E_{w_n} \rangle + \sum_{k=1}^{N} \langle f, B_k \rangle B_k, E_{w_n} \rangle |$$

$$\leq \| f - \sum_{k=1}^{N} \langle f, B_k \rangle B_k \| \| \sum_{k=1}^{N} \langle f, B_k \rangle B_k, E_{w_n} \rangle |$$

$$\leq \epsilon/2 + | \sum_{k=1}^{N} \langle f, B_k \rangle B_k, E_{w_n} \rangle |.$$  

We note that in the last summation the functions $B_1, \cdots, B_N$ can be expressed as linear combinations of $E_{a_1}, \cdots, E_{a_N}$. The inner products involving $B_1, \cdots, B_N$ then can be passed on to those with $E_{a_1}, \cdots, E_{a_N}$, and thus DBVC can be used. As a result, the last term of the above inequality chain is less than $\epsilon$ if $n$ is large enough. The proof is complete.

We need a condition on RKHS called $n$-Linearity Independent Condition ($n$-LIC): If for a fixed $n$ and any mutually distinguish $w_1, \cdots, w_n$ the corresponding function set $\{E_{w_1}, \cdots, E_{w_n}\}$ is linearly independent, then the RKHS is said to satisfy $n$-Linearity Independent Condition. This condition is rather mild, for, if it is not true, then a parameterized reproducing kernel is a linear expansion of some others. The latter implies that there exist $w_1, \cdots, w_n$, such that for all functions $f$ in the space there holds $f(w_n) = c_1 f(w_1) + \cdots + c_{n-1} f(w_{n-1})$, where $c_k$’s are fixed complex constants. A consequence of $n$-LIC, that is also the form that we use in the proof of our main Theorem 2.1, is that if $a_1, \cdots, a_k, b$ are mutually distinguish points in $E$, then the projection of $E_b$ into the span of $E_{a_1}, \cdots, E_{a_k}$, or $\| E_b - \sum_{k=1}^{k} \langle E_b, B_k \rangle B_k \| = \sqrt{1 - \sum_{l=1}^{k} | \langle E_b, B_l \rangle |^2 }$ is nonzero, where $(B_1, \cdots, B_k)$ is the G-S orthonormalization of $(E_{a_1}, \cdots, E_{a_k}), k \leq n - 1$.

The main result of this paper is

**Theorem 1.4** A RKHS $H_K$ has a solution for the $n$-best optimization problem (1.5) in the open set $E^n$ if the RKHS satisfies DBVC and $n$-LIC.

The precise statement of the theorem will be given in next section. The main effort of the proof is to show that under the conditions DBVC and $n$-LIC a solution exists and must situate in the open
set $E^n$ (interior solution). In both the theory (sifting process) and applications (model reduction) a solution being inside the open set is crucial, as having been seen in the complex Hardy space rational approximation theory (see, for instance, the enclosed references by Walsh, Baratchart, Qian, and Qu et al.). The main mechanism for such interior solutions is DBVC. In general RKHSs, a solution of the $n$-best may also happen at the boundary. Hence DBVC is not a necessary condition of existence of a general solution.

After proving the main theorem we verify that the weighted Bergman spaces in the disc satisfy DBVC and $n$-LIC, and thus conclude that the weighted Bergman spaces have $n$-best kernel approximations in the corresponding Hilbert space norms. Based on the obtained results we further prove existence of $n$-best spherical Poisson kernel approximation to functions of finite energy on the real-spheres. Except the classical Hardy spaces case, the other $n$-best existence results proved in this paper, including the version on RKHSs with a DBVC and $n$-LIC dictionary and the concrete examples with complex holomorphic function spaces and the spaces of functions of finite energy on the real-spheres, are all new results and proved for the first time.

## 2 Existence of $n$-Best Approximation for RKHS with DBVC and $n$-LIC

**Theorem 2.1** Let $H_K$ be a RKHS that satisfies DBVC and $n$-LIC. Let $n$ be any but fixed positive integer. Then for any $f \in H_K$, if $f$ by itself is not an $m$-kernel expansion form for $0 \leq m < n$, then there exists an $n$-tuple of parameters $(a_1, \cdots , a_n) \in E^n$, with $(B_1, \cdots , B_n)$ being the associated orthonormal systems such that

$$A(f; a) = \|f - \sum_{l=1}^{n} \langle f , B_l \rangle B_l\|$$

attains the minimum value over all possible values arising from all the $n$-tuples in place of $(a_1, \cdots , a_n)$ in $E^n$.

**Proof of Theorem 2.1** Denote $b = (b_1, \cdots , b_n)$. It is obvious that $A(f; b)$ has a non-negative global infimum value for all $b$ in $E^n$, call it $d$. We show that this global infimum value is attainable at an interior point of $E^n$. Let $a^{(k)} = (a_1^{(k)}, \cdots , a_n^{(k)})$ be an $n$-tuple at which $A(f; a^{(k)}) < d + 1/k$. There then exists a subsequence tending to an $n$-tuple $a$ in $E^n$. Without loss of generality we can assume that the sequence $a^{(k)}$ itself tends to $a$. We are to show $a \in E^n$. Assume the opposite, which means that some coordinates of $a$ are on $\partial E$, and we will, in such case, introduce a contradiction. We divide the $n$ coordinates into two groups, $I$ and $B$, where for $l \in I$ there holds $\lim_{k \to \infty} a_l^{(k)} = a_l \in E$; and for $l \in B$ there holds $\lim_{k \to \infty} a_l^{(k)} = a_l \in \partial E$. We are assuming $B \neq \emptyset$. Since $A(f; a^{(k)})$ is the energy of $f$ onto the orthogonal complement of the span of the multiple reproducing kernels in the $n$-tuple $a^{(k)}$, the energy quantity being irrelevant with the order of the elements in $a^{(k)}$, we can assume, without loss of the generality, that the coordinates in $I$ are all in front of those in $B$. The point is to show that, because $\lim_{k \to \infty} a_l^{(k)} \in \partial E$ for $l \in B$, the components $a_l^{(k)}$ of $a^{(k)}$, if $l \in B$, will have no contributions to the optimization of $A(f; a)$. To simplify the argument we may assume without loss of generality that for every $k$ the $n$-tuple $a^{(k)}$ does not have multiple components, although the limiting $n$-tuple $a$ may have. Let $l_0$ be the largest index for the indices in $I$, then the indices $l_0 + t, 0 < t \leq n - l_0$ will be in the index range $B$. Since $B \neq \emptyset$, we have $l_0 < n$. 

7
Let \( R^{(k)} = \text{span}\{E_{a_1^{(k)}}, \ldots, E_{a_n^{(k)}}\} \) and \( P^{(k)} \) the orthogonal projection to \( R^{(k)} \); and likewise, \( R_1^{(k)} = \text{span}\{E_{a_1^{(k)}}, \ldots, E_{a_0^{(k)}}\} \) and \( P_1^{(k)} \) the orthogonal projection mapping into \( R_1^{(k)} \). It is easy to show that for the given function \( f \), the projections \( P_1^{(k)} f \) have a limit as \( k \to \infty \), denoted \( P_I f \), as the projection of \( f \) into span\{\(E_{a_1}, \ldots, E_{a_0}\)\}. Denote \( g = f - P_I f \).

The general form of the elements in the Gram-Schmidt orthonormalization of the system \( \{E_{a_i^{(k)}}\}_{i=1}^n \) is

\[
B_t^{(k)} = \frac{E_{a_t^{(k)}} - \sum_{j=1}^{t-1} \langle E_{a_j^{(k)}}, B_j^{(k)} \rangle B_j^{(k)}}{\|E_{a_t^{(k)}} - \sum_{j=1}^{t-1} \langle E_{a_j^{(k)}}, B_j^{(k)} \rangle B_j^{(k)}\|}, \tag{2.10}
\]

where \( \{B_1^{(k)}, \ldots, B_j^{(k)}\} \) is the Gram-Schmidt orthonormalization of \( \{E_{a_1^{(k)}}, \ldots, E_{a_j^{(k)}}\}, 1 \leq j \leq n \).

We show that for any function \( h \) in the reproducing kernel Hilbert space there holds

\[
\lim_{k \to \infty} \langle h, B_t^{(k)} \rangle = 0, \quad l_0 < t \leq n. \tag{2.11}
\]

Temporarily accepting (2.11), and using it for \( h = f \), we have

\[
d^2 = \lim_{k \to \infty} (\|f\|^2 - \sum_{t=1}^{l_0} |\langle f, B_t^{(k)} \rangle|^2) = \|f - P_I f\|^2 = \|g\|^2.
\]

We note that \( g \neq 0 \), for otherwise \( f \) is an \( m \)-kernel form with \( m = l_0 < n \), contrary with the assumption. \( g \neq 0 \) then implies \( d > 0 \). Let \( g_j^{(k)} := f - \sum_{t=1}^{j} \langle f, B_t^{(k)} \rangle B_t^{(k)}, l_0 \leq j \leq n \). We have \( \lim_{k \to \infty} g_j^{(k)} = g_0 = g, l_0 \leq j \leq n \). Find \( a \in E \) such that \( |\langle g, E_a \rangle| = \delta > 0 \). Let the new parameter matrix be

\[
b_t^{(k)} = a_t^{(k)}, 1 \leq t < n; \quad b_n^{(k)} = a,
\]

where only the last column is different from the old. Then in the new system, using \( \tilde{B}_t^{(k)} \) in place of \( B_t^{(k)}, 1 \leq t \leq n \), and \( \tilde{P}^{(k)} \) in place of \( P^{(k)} \), we have

\[
\lim_{k \to \infty} \|\tilde{P}^{(k)} f\|^2 = \lim_{k \to \infty} (\sum_{t=1}^{n-1} |\langle f, \tilde{B}_t^{(k)} \rangle|^2 + |\langle f, \tilde{B}_n^{(k)} \rangle|^2)
\]

\[
= \lim_{k \to \infty} (\| \sum_{t=1}^{l_0} |\langle f, B_t^{(k)} \rangle B_t^{(k)}\|^2 + |\langle f, B_n^{(k)} \rangle|^2)
\]

\[
= \|P_I f\|^2 + \lim_{k \to \infty} |\langle f, B_n^{(k)} \rangle|^2
\]

\[
= \|P_I f\|^2 + \lim_{k \to \infty} |\langle f, \tilde{B}_n^{(k)} \rangle|^2 \tag{2.12}
\]

where \( \langle f, \tilde{B}_n^{(k)} \rangle = \frac{\langle f, (1 - P_{n-1}^{(k)}) E_a \rangle}{\rho_k} \), where, as a consequence of LIC, \( \rho_k \in (0, 1] \). We further have \( \langle f, (I - P_{n-1}^{(k)}) E_a \rangle = \langle (I - P_{n-1}^{(k)}) f, E_a \rangle \), also \( \lim_{k \to \infty} P_{n-1}^{(k)} f = P_{l_0} f \). Taking into account \( (I - P_{l_0}) f = g \), and
\[
\lim_{k \to \infty} \rho_k = \rho = \sqrt{1 - \sum_{t=1}^{l_0} |\langle E_a, B_j \rangle|^2} \in (0, 1), \text{ as a consequence of LIC again. The last equality chain (2.12) finally equals}
\]
\[
\|Pf\|^2 = \|Pf\|^2 + \left| \frac{g, E}\rho \right|^2 = \|f\|^2 - \|g\|^2 + (\delta/\rho)^2 > \|f\|^2 - d^2.
\]

Or,
\[
\lim_{k \to \infty} \|f - \overline{P}(k)\|^2 < d^2,
\]
being contrary with \(d\) being the global infimum value of \(A(f; b), b \in \mathbb{E}^n\). The proof of the theorem is complete.

Now we proceed to prove the relation (2.11) for \(t = l_0 + 1, \ldots, n\). First let \(t = l_0 + 1\). We have
\[
\langle f, B_{l_0+1}^{(k)} \rangle = \langle f, \frac{E_{a_{l_0+1}}^{(k)} - \sum_{j=1}^{l_0} \langle E_{a_{l_0+1}}^{(k)}, B_j^{(k)} \rangle B_j^{(k)}}{\sqrt{1 - \sum_{j=1}^{l_0} |\langle E_{a_{l_0+1}}^{(k)}, B_j^{(k)} \rangle|^2}} \rangle.
\]

Since \(H_K\) satisfies DBVC, from Lemma [1.3] \(H_K\) also satisfies BVC. As a consequence,
\[
\lim_{k \to \infty} \langle f, E_{a_{l_0+1}}^{(k)} \rangle = 0. \tag{2.13}
\]

Since \(\lim_{k \to \infty} (a_{l_0}^{(k)}, \ldots, a_{l_0}^{(k)}) = (a_1, \ldots, a_{l_0}) \in \mathbb{E}^{l_0}\), there exist the limits \(\lim_{k \to \infty} B_j^{(k)} = B_j\), being functions in \(H_K\), for \(j = 1, \ldots, l_0\). Then BVC and the Cauchy-Schwarz inequality imply
\[
\langle E_{a_{l_0+1}}^{(k)}, B_j^{(k)} \rangle = |\langle E_{a_{l_0+1}}^{(k)}, B_j \rangle| + |\langle E_{a_{l_0+1}}^{(k)}, B_j^{(k)} - B_j \rangle| \\
\leq |\langle E_{a_{l_0+1}}^{(k)}, B_j \rangle| + \|B_j^{(k)} - B_j\| \\
\to 0, \text{ as } k \to \infty, \quad j \leq l_0. \tag{2.14}
\]

In accordance with the relations (2.13) and (2.14), we have (2.11) for \(t = l_0 + 1\).

Now we prove (2.11) for \(t > l_0 + 1\). The induction hypotheses include that each term \(B_j^{(k)} , 1 \leq j \leq t - 1\), is a linear combination of \(E_{a_{s}}^{(k)}, 1 \leq s \leq j\), while the coefficients of the linear combination are all constituted by sums and products between \(\langle E_{a_{s}}^{(k)}, E_{a_{s'}}^{(k)} \rangle, 1 \leq s', s \leq j\), and divisions by \(\sqrt{1 - \sum_{t=1}^{s-1} |\langle E_{a_{s}}^{(k)}, B_l^{(k)} \rangle|^2}, 1 \leq s \leq j\), without involving universal constants; and that the
\[
\lim_{k \to \infty} \langle f, B_l^{(k)} \rangle = 0, \quad l_0 < l \leq t - 1.
\]

Write, in accordance with (2.10),
\[
\langle f, B_l^{(k)} \rangle = \frac{1}{\|E_{a_{l}}^{(k)} - \sum_{j=1}^{t-1} \langle E_{a_{l}}^{(k)}, B_j^{(k)} \rangle B_j^{(k)}\|} \left( \langle f, E_{a_{l}}^{(k)} \rangle - \sum_{j=1}^{t-1} \langle E_{a_{l}}^{(k)}, B_j^{(k)} \rangle \langle f, B_j^{(k)} \rangle \right).
\]

The assumed DBVC, its consequence BVC, and the induction hypotheses together, establish
\[
\lim_{k \to \infty} \langle f, E_{a_{l}}^{(k)} \rangle = 0, \quad \lim_{k \to \infty} \langle E_{a_{l}}^{(k)}, B_j^{(k)} \rangle = 0, \quad \text{and } \lim_{k \to \infty} \langle f, B_j^{(k)} \rangle = 0, \quad j \leq t - 1.
\]
Therefore,
\[
\lim_{k \to \infty} \langle f, B_i^{(k)} \rangle = 0.
\]

Based on the mathematical induction principle the proof is complete.

\[\textbf{Remark 2.2} \quad \text{A large amount commonly used Hilbert spaces are RKHSs in which DBVC and LIC are satisfied. The above theorem guarantees that such RKHSs have n-best kernel approximations. The recently developed cyclic and gradient descent algorithms ([18, 11]) for Hardy spaces are adaptable to abstract RKHSs with DBVC and LIC. The proof of the existence result guarantees convergence of the adapted algorithms in abstract spaces. It, in particular, serves as a useful reference in learning theory for simultaneously selecting n-parameters to optimize an energy-based objective function.}\]

\section{Hardy, Bergman and Weighted Bergman Spaces}

We briefly introduce the Bergman, and the weighted Bergman, and the weighted Hardy spaces.

The classical Bergman space together with its norm is defined as

\[
\mathbb{A}^2(D) = \{ f : D \to \mathbb{C} \mid f \text{ is holomorphic in } D, \text{ and } \|f\|^2_{\mathbb{A}^2(D)} = \int_D |f(z)|^2dA < \infty \},
\]

where \(dA\) is the normalized area measure on the unit disc: \(dA = \frac{dx dy}{\pi}, z = x + iy\).

It is a RKHS for its point evaluation functional is bounded. Denote by \(e^w_{\mathbb{A}^2}\) the normalized reproducing kernel of \(\mathbb{A}^2\), that is

\[
E_w(z) = e^w_{\mathbb{A}^2}(z) = \frac{1 - |w|^2}{(1 - wz)^2}.
\]

The weighted Bergman spaces are defined as

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

where \(dA_\alpha(z) = (1+\alpha)(1-|z|^2)^\alpha dA(z)\). With the norm \(\|\cdot\|_{\mathbb{A}^2_\alpha(D)}\) the space \(\mathbb{A}^2_\alpha\) is a reproducing kernel Hilbert space ([23]). The reproducing kernel is

\[
k^\alpha_w(z) = \frac{1}{(1 - wz)^{2 + \alpha}}, \quad \text{with} \quad \|k^\alpha_w\|^2 = k^\alpha_w(w) = \frac{1}{(1 - |w|^2)^{2 + \alpha}}. \quad (3.15)
\]

When \(\alpha = 0\), the corresponding weighted Bergman space \(\mathbb{A}^2_0 = \mathbb{A}^2\), the Bergman space. The spaces \(\mathbb{A}^2_\alpha, \alpha \to -1\), are increasing along with \(\alpha\) increasing. The relation between \(\mathbb{A}^2_\alpha, \alpha \to -1\), and the classical Hardy \(\mathbb{H}^2\) space is \(\mathbb{H}^2 \subseteq \bigcap_{\alpha \to -1} \mathbb{A}^2_\alpha\) ([13]).

The weighted Hardy spaces form a class of RKHSs larger than that of the weighted Bergman spaces ([24]). A special Chain, yet general enough, within the weighted Hardy spaces is

\[
\mathbb{H}^2_\beta(D) = \{ f : D \to \mathbb{C} \mid f \text{ is holomorphic in } D, \text{ and } \|f\|^2_{\mathbb{H}^2_\beta(D)} = \sum_{k=0}^{\infty} (k+1)^3 |c_k|^2 < \infty \},
\]
and the reproducing kernels are

\[ k^\beta(w, z) = \sum_{k=0}^{\infty} \frac{1}{(k+1)\beta} (wz)^k. \]

Moreover, \( ||k^\beta(w, z)||^2 = k^\beta(w, w) = \sum_{k=0}^{\infty} \frac{1}{(k+1)\beta} |w|^{2k}. \)

Based on the Gamma function estimation we have that \( \mathbb{H}^2_\beta \) is identical with the Bergman spaces \( \mathbb{A}^2_\alpha \), in the set theoretic sense, with equivalent norms, if \( \beta = -(\alpha + 1) \), \( \beta < 0 \), or, equivalent, \( \alpha > -1 \).

The collection of the weighted Hardy spaces corresponding to \( \beta > 0 \) would be right to be called Hardy-Sobolev spaces. We specially note that for \( \beta > 1 \) the spaces \( \mathbb{H}^2_\beta \) no longer possess BVC, that, as some Hardy-Sobolev spaces, is outside the \( \beta < 0 \) range for the weighted Bergman spaces.

In this section we will give applications of the main theorem in §2. We will first show that DBVC and LIC hold for the classical Bergman \( \mathbb{A}^2 \) space, and for all the weighted Bergman spaces of the unit disc as well. As a conclusion of Theorem [21] there exist solutions to the \( n \)-best kernel approximation problem for all the weighted Bergman spaces. Consequently, we have an \( n \)-best rational approximation in the weighted Bergman norm sense when the reproducing kernel is a rational function.

To show that the Bergman space \( \mathbb{A}^2 \) satisfies DBVC what we need is to verify that if \( \tilde{z} \in \overline{D} \) and \( w_k \to \tilde{w} \in \partial D \), then

\[
\lim_{k \to \infty} \langle e^{A^2_{\tilde{z}k}}, e^{A^2_{wk}} \rangle = 0. \tag{3.16}
\]

For the first term we use the Cauchy-Schwarz inequality, and for the second term we refer to [12] for the desired zero limit. To verify the case when both \( \tilde{w} \) and \( \tilde{z} \) belong to the boundary \( \partial D \), we are to show that

\[
\langle e^{A^2_{\tilde{z}k}}, e^{A^2_{wk}} \rangle = \frac{(1 - |z_k|^2)(1 - |w_k|^2)}{(1 - |\tilde{z}k||\tilde{w}k|^2)}
\]

tends to zero. It is easy to exam that if \( \tilde{w} \neq \tilde{z} \), then the zero limit is obvious. Now assume that \( \tilde{w} = \tilde{z} \), but \( w_k \neq z_k \). By using the polar coordinates, we write \( w_k = r_k e^{i\phi_k}, z_k = \rho_k e^{i\psi_k}, r_k, \rho_k < 1, r_k \to 1, \rho_k \to 1, \phi_k \to \theta, \psi_k \to \theta, \phi_k \neq \psi_k, \forall k \).

We will equivalently show

\[
\lim_{k \to \infty} \frac{1}{\sqrt{1 - |\langle e^{A^2_{w_k}}, e^{A^2_{z_k}} \rangle|^2}} = 1.
\]

The latter desired relation is reduced to, explicitly,

\[
\lim_{k \to \infty} \frac{(1 - \overline{w_k}z_k)(1 - w_k\overline{z_k})}{\sqrt{A^2 - B^2}} = 1,
\]

where \( A = (1 - \overline{w_k}z_k)(1 - w_k\overline{z_k}), B = (1 - |w_k|^2)(1 - |z_k|^2) \). Simple computation gives \( A - B = |w_k - z_k|^2 \).
Due to the relation $A + B \geq A - B = |w_k - z_k|^2$, we have
\[
\frac{(1 - w_k z_k)(1 - w_k \bar{z}_k)}{\sqrt{A^2 - B^2}} \leq \frac{|1 - w_k z_k|}{|w_k - z_k|} \frac{|1 - w_k z_k|}{\sqrt{A + B}} \leq \frac{|1 - w_k z_k|^2}{|w_k - z_k|^2} \leq \frac{1}{r_k} \left( \frac{|1 - r_k \rho_k e^{i(\phi_k - \psi_k)}|}{|1 - \frac{\rho_k}{r_k} e^{i(\phi_k - \psi_k)}|} \right)^2.
\]
Note that $\phi_k - \psi_k \neq 0$. Elementary geometry asserts that whenever $\theta \in (0, \pi/4)$, $\cos \theta < t < s$, there holds $|1 - te^{i\theta}| < |1 - se^{i\theta}|$. Letting $\theta = \phi_k - \psi_k$, $t = r_k \rho_k$, $s = \frac{\rho_k}{r_k}$, we hence have
\[
|1 - r_k \rho_k e^{i(\phi_k - \psi_k)}| < |1 - \frac{\rho_k}{r_k} e^{i(\phi_k - \psi_k)}|.
\]
Therefore, the upper limit
\[
\lim_{k \to \infty} \frac{(1 - w_k z_k)(1 - w_k \bar{z}_k)}{\sqrt{A^2 - B^2}} \leq 1.
\]
On the other hand, the lower limit is not smaller than 1. We thus conclude (2.11) in the classical Bergman space.

The above result can be easily extended to the classical Hardy $\mathbb{H}^2$ and the weighted Bergman spaces $\mathbb{A}_\alpha^2$, $\alpha > -1$. In fact, with a temporary abuse of notation by denoting $\mathbb{H}^2 = \mathbb{A}_0^2$ for $\alpha = -1$, we have
\[
\langle e_w^2, e_z^2 \rangle = \left( \langle e_w^2, e_z^{-2} \rangle \right)^{2+\alpha},
\]
where $\alpha = 0$ corresponds to the Bergman space.

The LIC condition for the weighted Bergman spaces is easy to be verified. Their reproducing kernels, in fact, have explicit representation formulas as meromorphic functions showing where the poles are. Suppose a kernel can be expressed as a linear combination of some other kernel functions, then in terms of distribution of the poles we immediately have a contradiction. We thus obtain the following theorem.

**Theorem 3.1** For the Hardy space $\mathbb{H}^2$ and the weighted Bergman spaces $\mathbb{A}_\alpha^2$, $\alpha > -1$, the $n$-best optimization problems (1.7) all have interior solutions. When the reproducing kernel is a rational function we obtain correspondingly $n$-best rational function approximations in the RKHS norm.

We finally remark that the commonly used algorithms of finding $n$-best approximations in the Hardy space case may be adapted to finding $n$-best approximations in general RKHSs. Those algorithms, however, are mostly through comparison of nearby values of the objective functions, and finally may lead to local extremes. By adopting certain practical methods global extremal values may be attainable (see [22] [8] [6] [12] [13]).

### 4 n-Best Spherical Poisson Kernel Approximation

Next we work on $L^2(S^{d-1})$, the Hilbert space of square integrable functions on the $(d-1)$-dimensional unit sphere centered at the origin, $d \geq 2$, where the inner product of $L^2(S^{d-1})$ is
\[
\langle f, g \rangle_{L^2(S^{d-1})} = \int_{S^{d-1}} f(t)g(t)\,d\sigma(t),
\]
and \( d\sigma(t) \) is the normalized Lebesgue measure on the sphere. Let the parameter set \( E = B^d \), being the \( d \)-dimensional unit ball centered at the origin. For a point \( q \in B^d \), the function \( h_q \) in \( H - H_K \) formulation (\[10\]) of the context is the Poisson kernel of the ball: with \( q = rt, r = |q|, t \in S^{d-1} \),

\[
h_q(t) = P_q(t) = c_n \frac{1 - r^2}{|q - t|^n}. \tag{4.17}
\]

The operator \( L \) and its images \( Lf, f \in L^2(S^{d-1}) \), are given

\[
u(q) = Lf(q) = \langle f, h_q \rangle_{L^2(S^{d-1})},
\]

We also use the notation \( L(h_p) \) to denote the parameterized Poisson kernel that is the \( L \)-image of \( h_p = P_p(t) \in \mathcal{H} \). The range \( R(L) \) is the harmonic Hardy space on the unit ball \( B^d \):

\[
h^2(B^d) = \{ u : B^d \to \mathbb{R} : \triangle u = 0, \sup_{0 \leq r < 1} \int_{S^{d-1}} |u(r)|^2 d\sigma(t) < \infty \}.
\]

It is a result of harmonic analysis on the sphere that \( \{ h_q \}_{q \in B^d} \) is dense in \( L^2(S^{d-1}) \) (\[20\]). The RKHS \( H_K = h^2(B^d) \) is the range set \( R(L) \) equipped with the norm induced from their non-tangential boundary limits (NTBs): For \( u = Lf \), there holds

\[
\lim_{r \to 1} u(rt) = f(t),
\]

in both the \( L^2 \)- and in the a.e. pointwise sense, and

\[
\|u\|_{H_K} \overset{\triangle}{=} \|f\|_{L^2(S^{d-1})}.
\]

For this reason we write \( f(t) = u(t) \) under the correspondence \( u = Lf \). We also have the relations

\[
\|u\|_{H_K}^2 = \|f\|_{L^2(S^{d-1})}^2 = \|u(\cdot)\|_{L^2(S^{d-1})}^2 = \sup_{0 \leq r < 1} \int_{S^{d-1}} |u(rt)|^2 d\sigma(t).
\]

The reproducing kernel of the space \( H_K \) is computed, for \( q = rt, p = \rho s, t, s \in S^{d-1} \),

\[
K(q, p) = \langle h_q, h_p \rangle_{L^2(S^{d-1})} = P_{\rho s}(t) = P_{\rho s}(t).
\tag{4.18}
\]

The last equality is due to the relation \( |\rho t - s|^2 = |\rho s - t|^2 \) and the expressions of the corresponding Poisson kernels. The second last equality of the above refers, or is equivalent, to the semi-group property of the spherical Poisson kernel. It gives computational convenience in the approximation. To verify the reproducing property of \( K_q \), we have, for \( u \in H_K = h^2(S^{d-1}), q = rt \),

\[
\langle u, K_q \rangle_{H_K} = \langle u(\cdot), P_{rt}(\cdot) \rangle_{L^2(S^{d-1})} = u(rt) = u(q).
\]

The norm of a general reproducing kernel \( K_w, w = \rho s, s \in S^{d-1}, 0 < \rho < 1 \), is computed

\[
\|K_w\|_{H_K}^2 = \langle K_w, K_w \rangle_{H_K} = K(w, w) = P_{\rho s}(s) = \frac{c_n(1 + \rho^2)}{(1 - \rho^2)^{n-1}}.
\]
Therefore, the norm-one normalization of $K_q$, as denoted

$$E_w = \frac{K_w}{\|K_w\|} = \frac{(1 - \rho^2)^{(n-1)/2}}{\sqrt{c_n(1 + \rho^2)}} K_w.$$ 

The boundary vanishing condition (BVC) in the case stands for

$$\lim_{B^d \ni w \to S_{d-1}} |\langle u, E_w \rangle_{H_K}| = 0,$$ 

where $u$ is any function in $H_K$. Due to density of the span of the parameterized Poisson kernels verification of the BVC for general functions $u$ reduces to verifying the BVC of any but fixed parameterized Poisson kernel $K_q, q = rt, t \in S^{d-1}$. Thanks to the reproducing kernel expression (4.18) we have,

$$\langle K_q, E_w \rangle_{H_K} = \frac{(1 - \rho^2)^{(n-1)/2}}{\sqrt{c_n(1 + \rho^2)}} P_{\rho^2}(r).$$ 

When $\rho \to 1^-$, the quantities $P_{\rho^2}(r)$ are bounded uniformly in $t, s$, and the bounds depend on the fixed $r < 1$. Since the factor in front tends to zero for $n \geq 2$, the whole tends to zero, uniformly in $t, s$. We thus conclude BVC and have POAFD for this model.

Below we show that the sphere Poisson kernel context further satisfies DBVC so that in the case $n$-best approximation by the spherical Poisson kernel does exist.

According to (4.20) and (4.17) the sphere case DBVC is interpreted as under the limiting process $r, \rho \to 1$, and on the sphere $t' \neq s', t', s' \to x'$, there holds

$$\lim \frac{(1 - \rho^2)^{(n-1)/2}(1 - r^2)^{(n-1)/2}}{|t' - r\rho s'|^n} = 0.$$ 

Since

$$\frac{1 - r^2 \rho^2}{|t' - r\rho s'|} \leq 1,$$

the quantity in (4.21) under the limit is dominated by

$$\frac{(1 - \rho^2)^{(n-1)/2}(1 - r^2)^{(n-1)/2}}{|t' - r\rho s'|^{2(n-1)/2}}.$$ 

Therefore, it suffices to prove

$$\lim \frac{(1 - \rho^2)(1 - r^2)}{|t' - r\rho s'|^2} = 0.$$ 

This turns to be of the same form as we proved in the last section for the Bergman space case in the unit disc, where with $z_k = \rho_k e^{i\theta_k}, w_k = r e^{i\psi_k}, r, \rho \to 1; \theta_k \neq \psi_k, \theta_k, \psi_k \to \alpha$, we showed

$$\lim_{k \to \infty} \frac{|z_k - w_k|^2}{|1 - \overline{w_k} z_k|^2} = 0.$$ 

To show the limit we created a circumstance where we could assert an equivalent limit through an upper limit being dominated by a lower limit. To make this part self-containing we repeat the method.
Let \( A_k = |t'_k - r_k \rho s'_k|^2 \), \( B_k = (1 - \rho^2_k)(1 - r_k^2) \). To prove \( \lim_{k \to \infty} \frac{B_k}{A_k} = 0 \) is reduced to proving the equivalent assertion

\[
\lim_{k \to \infty} \frac{1}{\sqrt{1 - \frac{B_k^2}{A_k^2}}} = 1.
\]

(4.22)

We, however, have

\[
\frac{1}{\sqrt{1 - \frac{B_k^2}{A_k^2}}} = \frac{A_k}{A_k^2 - B_k^2} \leq \frac{A_k}{A_k - B_k},
\]

where, for \( \cos \phi_k = \langle s'_k, t'_k \rangle_{\mathbb{R}^n} \),

\[
A_k - B_k = |t'_k - r_k \rho s'_k|^2 - (1 - \rho^2)(1 - r^2) = (1 - 2r_k \rho_k \cos \phi_k + r_k^2 \rho_k^2) - (1 - r_k^2 - \rho_k^2 + r_k^2 \rho_k^2) = |\rho_k t'_k - r_k s'_k|^2 = \rho_k^2 |t'_k - \frac{r_k}{\rho_k} s'_k|^2.
\]

The geometry of the sphere implies, for \( r_k \rho_k \) being close to 1,

\[
|t'_k - r_k \rho_k s'_k| < |t'_k - \frac{r_k}{\rho_k} s'_k|.
\]

There then follows

\[
\lim_{k \to \infty} \frac{1}{\sqrt{1 - \frac{B_k^2}{A_k^2}}} \leq \lim_{k \to \infty} \frac{1}{\rho_k^2} \frac{|t'_k - r_k \rho_k s'_k|^2}{|t'_k - \frac{r_k}{\rho_k} s'_k|^2} \leq 1.
\]

On the other hand,

\[
\lim_{k \to \infty} \frac{1}{\sqrt{1 - \frac{B_k^2}{A_k^2}}} \geq 1.
\]

We hence have the desired DBVC (4.21).

References

[1] D. Alpay, F. Colombo, T. Qian, I. Sabadini, *Adaptive orthonormal systems for matrix-valued functions*, Proceedings of the American Mathematical Society, 2017, 145(5): 2089-2106.

[2] D. Alpay, F. Colombo, T. Qian, and I. Sabadini, *Adaptative decomposition: The case of the Drury-Arveson space*, Journal of Fourier Analysis and Applications, 2017, 23(6): 1426-1444.

[3] L. Baratchart, *Existence and generic properties of L^2 approximations for linear systems*, Math. Control Inform., 3: 89-101.

[4] E. D. Livshitz, V. N. Temlyakov, *On convergence of weak greedy algorithms*, South Carolina University Columbia DEPT of Mathmatics, 2000.
[5] S. Mallat, Z. Zhang, *Matching pursuits with time-frequency dictionaries*, IEEE Trans. Signal Process, 1993, **41**: 3397-3415.

[6] W. Mi, T. Qian, F. Wan, *A fast adaptive model reduction method based on Takenaka-Malmquist systems*, Systems and Control Letters, 2012, **61**(1): 223-230.

[7] T. Qian, *Reproducing Kernel Sparse Representations in Relation to Operator Equations*, Complex Anal. Oper. Theory 14 (2020), no. 2, 1C15.

[8] T. Qian, Two-Dimensional Adaptive Fourier Decomposition, Mathematical Methods in the Applied Sciences, 2016, **39**(10): 2431-2448.

[9] T. Qian, Y.B. Wang, Adaptive Fourier series—a variation of greedy algorithm, Advances in Computational Mathematics, 2011, **34**(3): 279–293.

[10] T. Qian, E. Wegert, *Optimal approximation by Blaschke forms*, Complex Variables and Elliptic Equations, 2013, **58**(1): 123-133.

[11] T. Qian, J. Z. Wang, W. X. Mai, *An Enhancement Algorithm for Cyclic Adaptive Fourier Decomposition*, Applied and Computational Harmonic Analysis, available, 2019.

[12] W. Qu, P. Dang, *Rational approximation in a class of weighted Hardy spaces*, Complex Analysis and Operator Theory, 2019, **13**(4): 1827-1852.

[13] W. Qu, P. Dang, *Reproducing kernel approximation in weighted Bergman spaces: Algorithm and applications*, Mathematical Methods in the Applied Sciences, 2019, **42**(12): 4292-4304.

[14] G. Ruckebusch, *Sur l’approximation rationnelle des filtres*, Report No 35 CMA Ecole Polytechnique, 1978.

[15] J. L. Walsh, *Interpolation and approximation by rational functions in the complex domain*, American Mathematical Soc. Publication, 1962, **10**.

[16] T. Qian, Reproducing Kernel Sparse Representations in Relation to Operator Equations. Complex Anal. Oper. Theory 14 (2020), no. 2, 1C15.

[17] T. Qian, Y.B. Wang, Remarks on adaptive Fourier decomposition, International Journal of Wavelets, Multiresolution and Information Processing, 2013, **11**(01).

[18] T. Qian, Cyclic AFD Algorithm for best approximation by rational functions of given order, Mathematical Methods in the Applied Sciences, 2014, **37**(6): 846-859.

[19] Baratchart, L., Cardelli, M., Olivi, M., Identification and rational $L^2$ approximation a gradient algorithm, Automatica, 1991, **27**(2): 413-417.

[20] E. Stein, G. Weiss, *Introduction to Fourier Analysis in Euclidean Spaces*, Princeton University Press, 1970.

[21] L. H. Tan, T. Qian, Q. H. Chen, *New aspects of Beurling-Lax shift invariant subspaces*, Applied Mathematics and Computation, 2015, **256**: 257-266.
[22] X. Y. Wang, T. Qian, I. T. Leong, Y. Gao, *Two-Dimensional Frequency-Domain System Identification*, IEEE Transactions on Automatic Control, 2019, DOI: 10.1109/TAC.2019.2913047.

[23] B. Korenblum, H. Hedenmalm, K. Zhu, D. Békollé, Theory of Bergman spaces, The Mathematical Intelligencer, 2005, 27 (1): 85–86.

[24] B. MacCluer, Elementary functional analysis. Springer Science & Business Media, 2008, 253.