SPARSE REPRESENTATIONS OF STOCHASTIC SIGNALS

TAO QIAN*

ABSTRACT. Studies of sparse representations of deterministic signals have been well developed. Among other types there exists one called the adaptive Fourier decomposition (AFD) type for the analytic Hardy spaces. This type is recently further extended to the context of Hilbert spaces with a dictionary. Through the Hardy space decomposition of the space of $L^2$-signals the AFD type algorithm gives rise to sparse representations of signals of finite energy. To deal with multivariate signals the Hilbert space context comes into play. The multivariate AFD counterpart in Hilbert spaces with a dictionary is called pre-orthogonal AFD (POAFD). In the present study we generalize AFD and POAFD to random analytic signals through formulating stochastic analytic Hardy spaces. To analyze random analytic signals we work on two models, both being called stochastic AFD, or SAFD in brief. The two models are respectively made for (i) expressible as the sum of a deterministic signal and an error term such as a white noise (SAFDI); and for (ii) being random analytic signals divided into different classes of signals obeying certain distributional law (SAFDII). In the second half of the paper we drop the analyticity assumption and generalize the SAFDI and SAFDII to what we call stochastic Hilbert spaces with a dictionary. The generalized methods are named as stochastic pre-orthogonal adaptive Fourier decompositions, SPOAFDI and SPOAFDII. Like the deterministic AFDs and POAFDs, the developed stochastic POAFD algorithms offer powerful tools to analyze random signals.

AMS Classification: 42A50; 41A30; 30B99; 60G35; 60G07; 60G10

Key words: Sparse Representation, Analytic Signals, Hardy Spaces, Dictionary, Adaptive Fourier Decomposition, Pre-Orthogonal Adaptive Fourier Decomposition, random signal, Stochastic Analytic Hardy Spaces, Stochastic Hilbert Spaces, weak stability, weak ergodicity

Contents

1. Introduction 2
2. Stochastic AFDs 6
2.1. Properties of Hardy Space Projection of Random Signals 6
2.2. The Type SAFDI: Taking Expectation First 11
2.3. The SAFDII: Taking Expectation Secondly 16
3. Stochastic SPOAFDs in Hilbert Spaces 19
3.1. POAFD in a Hilbert Space With a Dictionary Satisfying BVC 19
3.2. Stochastic POAFDs 21
4. Acknowledgement 25
5. Appendix 25
References 27

*Corresponding author.
Funded by The Science and Technology Development Fund, Macau SAR (File no. 0123/2018/A3).
1. Introduction

If $F$ is a complex-valued signal in $[0, 2\pi)$ with finite energy, then it can be expanded into its $L^2([0, 2\pi))$-convergent Fourier series:

$$F(t) = \sum_{k=-\infty}^{\infty} c_k e^{ikt}.$$  

To make convenient use of complex analysis we alter the notation and denote it as $f(e^{it}) = F(t)$. Then the Plancherel Theorem asserts the relation $\|f\|_2^2 = \sum_{k=-\infty}^{\infty} |c_k|^2$, where the $L^2$-norm is one with respect to the inner product

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

The Plancherel relation infers that $c_k$ tends to zero and therefore the complex-valued functions $f^+(z) = \sum_{k=0}^{\infty} c_k z^k$ and $f^-(z) = \sum_{k=-\infty}^{-1} c_k z^k$

are analytic in $D$ and in $C \setminus \overline{D}$, respectively, where $D$ stands for the open unit disc in the complex plane $C$. Restricted to the unit circle, in the $L^2$-convergence sense, we define

$$f^+(e^{it}) \triangleq \sum_{k=0}^{\infty} c_k e^{ikt}$$

as the analytic signal associated with $f$. Denote by $H$ the Hilbert operator on the circle defined by

$$Hf(e^{it}) = \sum_{k=-\infty}^{\infty} (-i)\text{sgn}(k)c_k e^{ikt},$$

where $\text{sgn}(k) = k/|k|$ when $k \neq 0$ and $\text{sgn}(0) = 0$. We have $f^\pm = \frac{1}{2}(f + iHf \pm c_0)$. The non-tangential boundary limit of $f^+(z)$ as $z \to e^{it}$ coincides with the above defined $L^2$-limit $f^+(e^{it})$. To be practical we assume that the test functions $f$ are real-valued. Then $c_{-n} = \overline{c_n}$, and, as a consequence,

$$f(e^{it}) = 2\text{Re}\{f^+(e^{it})\} - c_0.$$  

Due to the above relation, the analysis of a real-valued signal of finite energy can be reduced to the analysis of the associated analytic signal $f^+$. Since $f^+$ is the boundary limit of the analytic function $f^+(z)$ in $D$, complex analytic methods are available for $f^+$. The totality of all such analytic functions $f^+(z)$ in the disc constitute the function space

$$H^2(D) \triangleq \{f : D \to C \mid f \text{ is analytic and } f(z) = \sum_{k=0}^{\infty} c_k z^k \text{ with } \sum_{k=0}^{\infty} |c_k|^2 < \infty\}$$

(1.1)  

called the (complex analytic) Hardy $H^2$-space in the unit disc. There exist other complex analytic Hardy spaces having more or less parallel theories as the one defined in the disc. For instance, the Hardy space idea to study functions may be extended to signals defined on the whole real line $R$, to those defined on manifolds in the higher dimensional complex
spaces $\mathbb{C}^d$ in the several complex variables setting (e.g., the Hardy spaces on tubes [33]), or to those in the real-Euclidean spaces $\mathbb{R}^d$ in the Clifford algebra setting (the conjugate harmonic systems, [33, 8]), and with scalar, or complex, or vector values, or even matrix-values ([1, 2]), etc., all obeying the same philosophy. We will only take the context $H^2(D)$ as an example to explain the adaptive Fourier decomposition (AFD) theory. In below we often abbreviate $H^2(D)$ as $H^2$. The Hardy space $H^2(D)$ has several equivalent characterizations that are not of interest of this paper. The disc case corresponds to signals defined in a compact interval on the line. That is the model adopted by periodic signals. In the first half of this paper we mainly concentrate in stochastic lization of the Hardy space in which the adaptive Fourier decomposition, AFD or Core-AFD, was formulated ([26]). We note that AFD on the disc heavily depends on two intimately related concepts, Blaschke product and Takenaka-Malmquist system, the latter being abbreviated as TM system. AFD is, in fact, in terms of TM system. In many analytic function spaces Blaschke product-like functions are not available. Pre-orthogonal AFD (POAFD) facilitates a replacement of AFD in the Hilbert spaces that do not have easy-usable Blaschke product-like functions or T-M systems, nor explicit and constructive orthogonal function systems. The latter are in particular for multivariate signals. We leave the POAFD method to be studied in the second half of this paper in which we formulate stochastic POAFD in the general setting of stochastic Hilbert space with a dictionary.

In contrast with the deterministic signals setting, in practice, one encounters random signals: Signals are mostly corrupted with noise or together with measurement errors, or, as an alternative type, consisting of several classes of signals under certain distribution law. A practical formulation then should be a real-valued function $F(t, w)$, where for a fixed probabilistic sample point $w \in \Omega$ the function $F(\cdot, w)$ is a deterministic signal of finite energy; meanwhile for each point $t$ in the time domain or the space domain the function $F(t, \cdot)$ is a random variable. We call such signals random signals (RSs). To formulate the corresponding stochastic Hardy space theory in the case $t \in [0, 2\pi)$ we rewrite $F(t, w)$ as $F(t, w) = f(e^{it}, w)$, and we have the trigonometric expansion

$$f(e^{it}, w) = \sum_{k=-\infty}^{\infty} c_k(w)e^{ikt} = \sum_{k=-\infty}^{\infty} c_k(w)z^k, \text{ where } c_k(w) = \frac{1}{2\pi} \int_0^{2\pi} f(e^{is}, w)e^{-iks} ds.$$

The Plancherel Theorem gives

$$\|f(\cdot, w)\|^2 = \sum_{k=-\infty}^{\infty} |c_k(w)|^2.$$

In our study it is reasonable to impose the condition

$$[E_w\|f(\cdot, w)\|^2]^{\frac{1}{2}} = \left(\sum_{k=-\infty}^{\infty} E_w|c_k(w)|^2\right)^{\frac{1}{2}} < \infty, \quad (1.2)$$

where $E_w$ stands for the mathematical expectation in the underlying probability space. In the whole paper the underlying probability space, $(\Omega, \mu), w \in \Omega$, is not specified, as the theory is valid for any but fixed probability space. The quantity in (1.2) is called the energy expectation norm (EE-Norm) of $f$, denoted as $\|f\|_N$. Set,

$$L^2_w(\partial D, \Omega) = \{f : \partial D \times \Omega \to \mathbb{C} \mid f \text{ is a RS, and } \|f\|_N < \infty\}, \quad (1.3)$$
called the space of random signals of finite energy. \( L_w^2(\partial D, \Omega) \) is written briefly as \( \mathcal{N} \). The RSs in \( L_w^2(\partial D, \Omega) \) are called normal random signals, or normal RSs. The space \( \mathcal{N} \) is a Hilbert space under the inner product induced from the EE-norm. A normal RS is almost surely a signal of finite energy in \( t \). In below we will keep the inner product notation \( \langle \cdot, \cdot \rangle \) only for the inner product of the time-domain-space \( L^2(\partial D) \).

Similarly to the deterministic case we will concentrate in studying “a half” of the space \( \mathcal{N} \), consisting of the RSs with expansions in the spectrum range \( k = 0, 1, \cdots \),

\[
f^+(e^{it}, w) = \sum_{k=0}^{\infty} c_k(w) e^{ikt}, \quad \text{satisfying} \quad \sum_{k=0}^{\infty} E_w(|c_k(w)|^2) < \infty.
\]

As a consequence, almost surely

\[
\sum_{k=0}^{\infty} |c_k(w)|^2 < \infty,
\]

and thus almost surely

\[
f^+(z, w) = \sum_{k=0}^{\infty} c_k(w) z^k
\]

is an analytic function in \( D \). The boundary limits exist in the a.e. pointwise, and in the \( L^2 \)-convergence sense as \( r = |z| \to 1 \), and

\[
f(e^{it}, w) = 2\text{Re}\{f^+(e^{it}, w)\} - c_0(w).
\]

On the boundary \( \partial D \) the projection \( f^+ \), apart being obtained through the Taylor expansion, can also be obtained through the singular integral operator, the (circular) Hilbert transform, \( H \):

\[
f^+(e^{it}, w) = \frac{1}{2} (f(e^{it}, w) + i H f(e^{it}, w) + c_0),
\]

where for any \( f(e^{it}, w) = \sum_{k=-\infty}^{\infty} c_k(w) e^{ikt} \), denoting \( \text{sgn}(k) = k/|k|, k \neq 0 \) and \( \text{sgn}(0) = 0 \), the signum function,

\[
H f(e^{it}, w) \triangleq \sum_{k=-\infty}^{\infty} (-i)\text{sgn}(k)c_k(w)e^{ikt}
= \frac{1}{\pi} \text{v.p.} \int_{-\infty}^{\infty} \cot \left( \frac{s}{2} \right) f(e^{i(t-s)}, w) ds.
\]

From the second equal relation we see that the Hilbert transform maps real-valued functions to real-valued functions.

By using Hilbert transformation study of the normal RSs can be reduced to that of their half series. We define the stochastic Hardy space as follows (with the superscript “+” dropped off), denoted

\[
H_w^2(D) = \{ f : D \times \Omega \to \mathbb{C} \mid f(z, w) \text{ is a.s. analytic in } z \text{ and } \}
\]

\[
f(z, w) = \sum_{k=0}^{\infty} c_k(w) z^k \text{ with } \| f \|_N^2 = \sum_{k=0}^{\infty} E_w|c_k(w)|^2 < \infty \}.
\]

There then is an induced space, being the totality of the boundary limits of the RSs in \( H_w^2(D) \), denoted as \( H_w^2(\partial D) \). The latter is a proper closed subspace of the space \( \mathcal{N} \) on the boundary \( \partial D \).
The purpose of this study is to develop stochastic adaptive Fourier decompositions (SAFDs) for analyzing random signals of two types. We will develop two models of stochastic AFD (SAFD), namely SAFDI and SAFDII. What makes the complex analysis methods a great power is that there is a Cauchy kernel and a Cauchy formula, the latter reproduces the function values of an analytic function using its boundary data. A direct generalization of the analytic function theory would be one for reproducing kernel Hilbert spaces. In the later half of this paper we extend the theory for analytic RSs further: We establish a counterpart theory in what we call *stochastic Hilbert space* with a dictionary. A Hilbert space with a dictionary is a more general concept than a reproducing kernel Hilbert space. The writing plan is as follows. In §2 with the stochastic Hardy space context we establish two types of sparse approximations, SAFDI and SAFDII, for treating two categories of analytic RSs: One is for noised deterministic signals, and the other is a collection of several classes of signals obeying certain probability distribution. In §3 we extend the theory to the context of stochastic Hilbert space with a dictionary treating also two categories of RSs, and develop, two types of sparse approximations, that we name, respectively, as SPOAFDI and SPOAFDII. The necessity of developing a theory in the general Hilbert space context rests in the tendency of studying multivariate random signals in which there does not exist good analyticity properties as may be used in the classical Hardy space cases.

For the reader’s convenience we give the following abbreviations list:

**AFD:** adaptive Fourier decomposition (for deterministic signals in the classical Hardy spaces consisting of analytic signals of finite energy on the boundary, associated with a Blaschke product structure)

**BVC:** boundary vanishing condition

**MSP:** maximal selection principle

**POAFD:** pre-orthogonal adaptive Fourier decomposition (Applicable for Hilbert spaces with a dictionary satisfying BVC)

**SBVC:** stochastic boundary vanishing condition

**RS:** random signal

**Normal RS:** normal random signal, or a signal in the space $H^2_w(D)$

**$\mathcal{N}$:** the Hilbert space consisting of normal RSs.

$H^2_w(D)$: the stochastic Hardy space on the disc, corresponding to $c_k(w) = 0$ for $k < 0$

$H^2_w(\partial D)$: the space of the functions as boundary limits of those in $H^2_w(D)$ defined on $\partial D$

**SHS:** a stochastic Hilbert space, or a Hilbert space of RSs possessing finite variation
SAFD, SAFDI, SAFDII: stochastic AFDs (SAFDs) are divided into two types: the type I, SAFDI, is for the RSs that are expressible as a deterministic signal corrupted with a noise of small $N$-norm; the type II, SAFDII, is for a general stochastic Hardy space.

SPOAFD, SPOAFDI, SPOAFDII: stochastic POAFDs (SPOAFDs) in SHS consist of two types; the type I, SPOAFDI, is for the RSs being expressible as noised signals; the type II, SPOAFDII, is for any general SHS.

2. Stochastic AFDs

In the deterministic signal analysis AFD is a sparse approximation methodology using a suitably adapted Takenaka-Malmquist (TM) system. In the classical Hardy space formulation it well fits with the Beurling-Lax Theorem, where any specific function belongs to a backward-shift-invariant subspace in which the function is the limit of a fast converging TM series. The AFD type expansions have found many applications in signal and image analysis as well as in system identification (see, for instance, [9, 37, 10, 11]). With the stochastic Hardy space defined in §1 we present two types of AFD-like expansions, called stochastic AFDs (SAFDs), of which each has its own merits in application. Before studying SAFDs we develop some aspects in relation to Hardy space projections of normal RSs.

2.1. Properties of Hardy Space Projection of Random Signals. Normal RSs $f(e^{it}, w)$ can all be represented into the form

$$f(e^{it}, w) = \tilde{f}(t) + \tilde{r}(e^{it}, w),$$

where $\tilde{f} = E_w f$. The difference $\tilde{r}$ is sometimes called the remainder RS. In this section we reduce the analysis of ordinary normal RSs to that of the analytic normal RSs. The philosophy support of this methodology is the relation (1.4). Given by the next two theorems, the Hardy space projections $f^+, \tilde{f}^+ \text{ and } \tilde{r}^+$ enjoy many good properties of those from which they are projected.

**Theorem 2.1.** If $f \in \mathcal{N}$, then $\tilde{f} \in L^2(\partial \mathbb{D}), \tilde{r} \in \mathcal{N}, E\tilde{r} = 0$. In writing

$$f(e^{it}, w) = \sum_{k=-\infty}^{\infty} c_k(w)e^{ikt} \text{ and } \tilde{r}(e^{it}, w) = \sum_{k=-\infty}^{\infty} d_k(w)e^{ikt},$$

there hold

$$\tilde{f}(e^{it}) = \sum_{k=-\infty}^{\infty} (E_w c_k)e^{ikt},$$

where $E_w c_k = E_w (c_k(w))$, and,

$$d_k(w) = c_k(w) - E_w c_k, \quad E_w d_k = 0, \quad k = 0, \pm 1, \pm 2 \ldots$$

The Hardy space projections $f^+, \tilde{f}^+, \tilde{r}^+$, respectively, belong to $H^2_w(\partial \mathbb{D}), H^2(\partial \mathbb{D})$, and in $H^2_w(\partial \mathbb{D})$. There hold

$$\{E_w f\}^+ = E_w \{f^+\} \quad \text{and} \quad \|\tilde{r}^+\|_{\mathcal{N}} = \frac{\|\tilde{r} + d_0\|_{\mathcal{N}}}{\sqrt{2}}.$$
Proof We note that
\[
\left( \sum_{k=-\infty}^{\infty} |E_w(c_k(w))|^2 \right)^{1/2} \leq E_w \left[ \left( \sum_{k=-\infty}^{\infty} |c_k(w)|^2 \right)^{1/2} \right] \quad \text{(Minkowski's inequality)}
\]
\[
\leq \left[ E_w \left( \sum_{k=-\infty}^{\infty} |c_k(w)|^2 \right) \right]^{1/2} [E_w(1)]^{1/2} \quad \text{(Hölder's inequality)}
\]
\[
= \left[ \sum_{k=-\infty}^{\infty} E_w(|c_k(w)|^2) \right]^{1/2} [E_w(1)]^{1/2}
\]
\[
(2.7) \quad = \|f\|_N < \infty.
\]

Then the Riesz-Fisher Theorem asserts that
\[
g(e^{it}) = \sum_{k=-\infty}^{\infty} E_w(c_k(w))e^{ikt} \in L^2(\partial D).
\]

Now we show \(\tilde{f} = g\). Denote \(f_n(e^{it}, w) = \sum_{|k| \leq n} c_k(w)e^{ikt}\). Then \(E_w f_n(e^{it}, w) = \sum_{|k| \leq n} E_w(c_k) e^{ikt}\).

Similarly to the reasoning of (2.7), there follows
\[
\|E_w f - E_w f_n\| = \|E_w(f - f_n)\|
\]
\[
\leq E_w\|f - f_n\|
\]
\[
\leq (E_w\|f - f_n\|^2)^{1/2}
\]
\[
= \|f - f_n\|_N
\]
\[
= \left( \sum_{|k| > n} E_w(|c_k(w)|^2) \right)^{1/2}
\]
\[
\to 0, \quad \text{as} \quad n \to \infty.
\]

Since the linear functional of the \(m\)-th Fourier coefficient, \(C_m\), is continuous, there follows
\[
C_m(E_w f) = \lim_{n \to \infty} C_m(E_w f_n) = E_w(c_m).
\]

This shows that \(E_w f = g \in L^2(\partial D)\) and is with the Fourier expansion
\[
\tilde{f} = \sum_{k=-\infty}^{\infty} E_w(c_k(w))e^{ikt} \in L^2(\partial D).
\]

It then follows
\[
(2.8) \quad E_w(\tilde{r}(e^{it}, w)) = E_w d_k = 0, \quad \forall t \in [0, 2\pi) \quad \text{and} \quad k = 0, \pm 1 \cdots
\]

As a consequence of (2.8), we have the orthogonality
\[
(2.9) \quad E_w(\tilde{r}(e^{it}, w)^2) = |\tilde{f}(e^{it})|^2 + E_w(|\tilde{r}(e^{it}, w)|^2),
\]
and thus the finiteness of the \(N\)-norm of \(\tilde{r}\):
\[
(2.10) \quad E_w(|\tilde{r}(e^{it}, w)|^2) = \|f\|_N^2 - \|\tilde{f}\|_{L^2(\partial D)}^2 < \infty \quad \text{for a.e.} \ t \in [0, 2\pi).
\]
To compute the $N$-norm of $\tilde{r}^+$, by taking into account $d_k = \overline{d}_{-k}$, we have

$$\|r^+\|_N^2 = E_w \int_0^{2\pi} |r^+(e^{it}, w)|^2 dt = \sum_{k=0}^{\infty} E_w |d_k(w)|^2 = \frac{\|\tilde{r} + d_0\|_N^2}{2}.$$ 

The proof of the theorem is complete. \qed

A particular example is that the remainder $\tilde{r}$ in question is the white Gaussian noise $N(0, \sigma^2)$, when the relation (2.10) becomes

$$E_w(\tilde{r}^2(e^{it}, w)) = \sigma^2, \quad \forall t \in [0, 2\pi).$$

We would be interested in properties imposed to the remainder $RS \tilde{r}$ not as special as white noise. What have in mind are weakly stationary, or little more further, ergodic $RS \tilde{r}$. Since we already have $E_w \tilde{r} = 0$, recall that if the autocorrelation function of $\tilde{r}$ (autocovariance function of $f$ itself) depends only on the time difference, that is, if there holds for some deterministic signal $\tilde{r}_1$,

$$\tilde{\gamma}(t, s) = E_w(\tilde{r}(e^{it}, w)\tilde{r}(e^{is}, w)) \triangleq \tilde{r}_1(s - t), \quad (2.11)$$

then $\tilde{r}$ is called a weakly stationary $RS$.

Recall that a weakly stationary $RS$, say $x(t, w)$, is weakly ergodic if and only if

$$E_w x = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x(t, w) dt, \quad a.s., \quad (2.12)$$

and

$$E_w (x(t, w) \overline{x}(t - \tau, w)) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x(t, w) \overline{x}(t - \tau, w) dt, \quad a.s. \quad (2.13)$$

The condition (2.12) implies that the common quantity of the LHS and the RHS of the equality (2.12) is a.s. a constant. The condition (2.13) implies that the common quantity of the LHS and the RHS of the equality (2.13) is a.s. a function of the time difference $\tau$. Since $E_w \tilde{r} = 0$, the relation (2.6) implies that, under the condition $\tilde{r}$ being weakly stationary, $f$ is weakly stationary if and only if $\tilde{f}$ is a.s. a constant function; and, $f$ is weakly ergodic if and only if $\tilde{f}$ is a.s. the zero function. A pure random variable $f(w)$ is stationary. If it is further ergodic, then it has to be a constant almost surely. This observation together with the above one hints that it would be necessary to assume that $d_0 = 0$ when discuss stationarity and ergodicity of RSs.

**Theorem 2.2.** Under the assumptions as in Theorem 2.1, if further $d_0 = 0$, a.s., then weak stationarity of $\tilde{r}$ implies weak stationarity of $\tilde{r}^+$; and, weak ergodicity of $\tilde{r}$ implies weak ergodicity of $\tilde{r}^+$.

We need first prove the following lemma.

**Lemma 2.3.** The Hilbert transform $H$ and the expectation operator $E_w$ are commutative.

**Proof** As proved in the beginning of the proof of Theorem 2.1, the series

$$\sum_{k=-\infty}^{\infty} |E_w c_k|^2$$
is convergent. It implies that, for each \( t \in [0, 2\pi) \),

\[
\sum_{k=\infty}^{\infty} (-i) \text{sgn}(k)(E_w c_k)e^{ikt}
\]

is absolutely convergent. This implies

\[
E_w \sum_{k=\infty}^{\infty} (-i) \text{sgn}(k)c_k(w)e^{ikt} = \sum_{k=\infty}^{\infty} (-i) \text{sgn}(k)(E_w c_k)e^{ikt}.
\]

Hence,

\[
(E_w H) f(e^{it}) = E_w \sum_{k=\infty}^{\infty} (-i) \text{sgn}(k)c_k(w)e^{ikt} = \sum_{k=\infty}^{\infty} (-i) \text{sgn}(k)(E_w c_k)e^{ikt} = H(E_w f)(e^{it}).
\]

The proof is complete. \( \square \)

**Proof of Theorem 2.2** We first show that if the autocorrelation function of \( \tilde{r} \) is a function of, merely, the time difference, then that of \( \tilde{r}^+ \) is the same. For this goal we first note that weakly stationarity of \( \tilde{r} \) implies the orthogonality under the expectation operation: \( E_w(\tilde{d}_k \tilde{d}_l) = \delta_k(l) \), where \( \delta_k(l) \) is the Dirac Delta function. As a consequence of it, we have, by invoking the respective Fourier expansions of \( \tilde{r}^+(e^{it}, w) \) and \( \tilde{r}^+(e^{is}, w) \),

\[
E_w(\tilde{r}^+(e^{it}, w)\tilde{r}^+(e^{is}, w)) = \sum_{k=0}^{\infty} E_w(|c_k(w)|^2)e^{ikt-s}.
\]

Next, we assume \( \tilde{r} \) is weakly stationary and weakly ergodic. We first show that the expectation is ergodic. By invoking the commutativity between \( H \) and \( E_w \) proved in Lemma 2.3 and the property that the Hilbert transform \( H \) annihilates constant functions, we have, almost surely,

\[
E_w(\tilde{r}^+) = \frac{1}{2} E_w(\tilde{r} + iH\tilde{r}) = \frac{1}{2} (E_w \tilde{r} + iE_w H\tilde{r}) = \frac{1}{2} E_w \tilde{r} + i\frac{1}{2} HE_w \tilde{r} = \frac{1}{2} E_w \tilde{r}.
\]
On the other hand,
\[
\lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} \tilde{r}^+(e^{it}, w) dt = \frac{1}{2} \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} \tilde{r}(e^{it}, w) dt + \frac{i}{2} \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} H\tilde{r}(e^{it}, w) dt
\]
\[
= \frac{1}{2} \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} \tilde{r}(e^{it}, w) dt + \frac{i}{2} \lim_{T \to \infty} H(\frac{1}{2T} \int_{-T}^{T} \tilde{r}(e^{it}, w) dt)
\]
\[
= \frac{1}{2} E_w\tilde{r} + \frac{i}{2} H(E_w\tilde{r})
\]
\[
= \frac{1}{2} E_w\tilde{r}.
\]

Therefore the expectation is ergodic. To show that the autocorrelation is also ergodic we proceed similarly. We first write
\[
E_w(\tilde{r}^+(e^{it}, w)\tilde{r}^-(e^{t-s}, w)) = \frac{1}{4} E_w([\tilde{r}(e^{it}, w) + iH\tilde{r}(e^{it}, w)][\tilde{r}(e^{i(t-s)}, w) - iH\tilde{r}(e^{i(t-s)}, w)]).
\]
The RHS of the last identity can be expressed as a complex linear combination of the following four terms:
\[
E_w(\tilde{r}(e^{it}, w)\tilde{r}(e^{i(t-s)}, w)), \quad E_w(H\tilde{r}(e^{it}, w)\tilde{r}(e^{i(t-s)}, w)),
\]
\[
E_w(\tilde{r}(e^{it}, w)H\tilde{r}(e^{i(t-s)}, w)) \quad \text{and} \quad E_w(H\tilde{r}(e^{it}, w)H\tilde{r}(e^{i(t-s)}, w)).
\]

For the first term, due to the ergodicity, we have
\[
E_w(\tilde{r}(e^{it}, w)\tilde{r}(e^{i(t-s)}, w)) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} \tilde{r}(e^{it}, w)\tilde{r}(e^{i(t-s)}, w) dt.
\]

We show that with each of the rest three terms the expectation operator may commute with the partial circular Hilbert transforms. The commutativity then leads to the respective ergodicity. With a little abuse of the notation, temporarily denoting
\[
H_u g(e^{i(t-u)}) = \frac{1}{\pi} \text{v.p.} \int_{-\pi}^{\pi} \cot \frac{u}{2} g(e^{i(t-u)}) du,
\]
we have
\[
E_w(H\tilde{r}(e^{it}, w)\tilde{r}(e^{i(t-s)}, w)) = E_w(H_u(\tilde{r}(e^{i(t-u)}, w)\tilde{r}(e^{i(t-s)})))
\]
\[
= H_u(E_w(\tilde{r}(e^{i(t-u)}, w)\tilde{r}(e^{i(t-s)})))
\]
\[
= H_u(\lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} \tilde{r}(e^{i(t-u)}, w)\tilde{r}(e^{i(t-s)}) dt)
\]
\[
= \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} H_u(\tilde{r}(e^{i(t-u)}, w)\tilde{r}(e^{i(t-s)}) dt)
\]
\[
= \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} H\tilde{r}(e^{it}, w)\tilde{r}(e^{i(t-s)}, w) dt.
\]

Similarly, we have
\[
E_w(\tilde{r}(e^{it}, w)H\tilde{r}(e^{i(t-s)}, w)) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} \tilde{r}(e^{it}, w)H\tilde{r}(e^{i(t-s)}, w) dt.
\]
For the last term we have

\[
E_w(H \tilde{f}(e^{it}, w)H \tilde{f}(e^{i(t-s)}, w)) = E_w(H_u H_v(\tilde{f}(e^{i(t-u)}, w)\tilde{f}(e^{i(t-s-v)}))) \\
= H_u H_v(E_w(\tilde{f}(e^{i(t-u)}, w)\tilde{f}(e^{i(t-s-v)}))) \\
= H_u H_v(\lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} \tilde{f}(e^{i(t-u)}, w)\tilde{f}(e^{i(t-s-v)})dt) \\
= \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} H_u H_v(\tilde{f}(e^{i(t-u)}, w)\tilde{f}(e^{i(t-s-v)})dt) \\
= \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} H \tilde{f}(e^{it}, w)H \tilde{f}(e^{i(t-s)}, w)dt.
\]

For functions in the underlying function space through a density argument based on functions in nice subspaces of functions the above exchange of taking limits may be justified. Then the same complex linear combination of the four just obtained ergodic identities leads to

\[
E_w(\tilde{r}^+(e^{it}, w)\overline{\tilde{r}^+(e^{i(t-s)}, w)}) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} \tilde{r}^+(e^{it}, w)\overline{\tilde{r}^+(e^{i(t-s)}, w)}dt,
\]

as desired. Thus, ergodicity of the autocorrelation is proved. The proof of Theorem 2.2 is complete. \( \square \)

2.2. The Type SAFDI: Taking Expectation First. In this section we assume that \( f(e^{it}, w) \) is \( H^2_u(D) \). Letting \( \tilde{f} = E_w(f(e^{it}, w)) \), we, as in the last section, have

\[
f(e^{it}, w) = \tilde{f}(e^{it}) + \tilde{r}(e^{it}, w).
\]

The function \( \tilde{f} \) is, in fact, in \( H^2(D) \). This is a consequence of Theorem 2.1 or can be proved by the similar but integral inequalities as, for \( r < 1 \),

\[
\left( \int_{0}^{2\pi} |E_wf(re^{it}, w)|^2dt \right)^{1/2} \\
\leq E_w \left[ \left( \int_{0}^{2\pi} |f(re^{it}, w)|^2dt \right)^{1/2} \right] \quad \text{(Minkovski’s inequality)} \\
\leq \left( E_w \int_{0}^{2\pi} |f(re^{it}, w)|^2dt \right)^{1/2} E_w(1)^{1/2} \quad \text{(Holder’s inequality)} \\
\leq \|f\|_{L^2} < \infty.
\]

(2.14)

We also note that, as a consequence of the last inequality, for a.s. \( w \in \Omega \), \( f(re^{it}, w) \) is a function in the classical analytic Hardy space with the power series expansion

\[
f(re^{it}, w) = \sum_{k=0}^{\infty} c_k(w)r^k e^{ikt}, \quad r < 1.
\]

The type SAFDI is based on AFD of the deterministic signal \( \tilde{f} \). For the self-containing purpose we now go through a full AFD expansion of \( \tilde{f} \). We will be using the \( L^2 \)-normalized
Szegő kernel on the circle:

\[ e_a(z) = \frac{\sqrt{1 - |a|^2}}{1 - \overline{a}z}, \quad a \in \mathbb{D}. \]

In \( H^2(\mathbb{D}) \) it has the reproducing kernel property: For any \( g \in H^2(\mathbb{D}) \),

\[ \langle g, e_a \rangle = \sqrt{1 - |a|^2} g(a). \]

Let \( f_1 = \tilde{f} \). For any \( a \in \mathbb{D} \) we have the following identity as an orthogonal decomposition

\[ (2.15) \quad \tilde{f}(z) = \langle f_1, e_a \rangle e_a(z) + f_2(z) \frac{z - a}{1 - \overline{a}z}, \]

where \( f_2 \) is called the reduced remainder, given by

\[ (2.16) \quad f_2(z) = \frac{f_1(z) - \langle f_1, e_a \rangle e_a(z)}{\frac{z - a}{1 - \overline{a}z}} \in H^2(\mathbb{D}). \]

Due to the orthogonalization we have

\[ (2.17) \quad \| \tilde{f} \|_{H^2(\mathbb{D})} = |\langle f_1, e_a \rangle|^2 + \| f_2 \|_{H^2(\mathbb{D})}. \]

Thus, the larger is the quantity \( |\langle f_1, e_a \rangle|^2 \), the smaller the energy of the reduced remainder \( f_2 \) is. Although \( \mathbb{D} \) is an open set it can be proved (see [26], for instance) that

\[ \sup \{ |\langle f_1, e_a \rangle|^2 \mid a \in \mathbb{D} \} \]

is attainable at a point of \( \mathbb{D} \). Hence, one practically selects

\[ a_1 = \arg \max \{ |\langle f_1, e_a \rangle|^2 \mid a \in \mathbb{D} \}. \]

Such maximal selection is phrased as \textit{Maximal Selection Principle} (MSP) of the Hardy space ([26]). The MSP is evidenced by the boundary vanishing condition (BVC) of the Szegő kernel dictionary in the Hardy space (see §3 for a more general formulation). Using this \( a_1 \) in place of \( a \) in (2.15), (2.16) and (2.17), we have that the corresponding reduced remainder \( f_2 \) has its least possible norm. To \( f_2 \) perform the same decomposition process, and so on, after \( n \)-iterations, we have

\[ (2.18) \quad \tilde{f}(z) = \sum_{k=1}^{n} \langle f_k, e_{a_k} \rangle B_k(z) + f_{n+1}(z) \prod_{k=1}^{n} \frac{z - a_k}{1 - \overline{a_k}z}, \]

where \( \{B_k\}_{k=1}^\infty \) is the Takenaka-Malmquist system determined by \( a_1, \ldots, a_k, \ldots \), all in \( \mathbb{D} \), where

\[ (2.19) \quad B_k(z) = e_{a_k}(z) \prod_{l=1}^{k-1} \frac{z - a_l}{1 - \overline{a_l}z}, \]

\[ (2.20) \quad a_k = \max \{ |\langle f_k, e_a \rangle|^2 \mid a \in \mathbb{D} \}, \]

\[ (2.21) \quad f_{k+1}(z) = \frac{f_k(z) - \langle f_k, e_{a_k} \rangle e_{a_k}(z)}{\frac{z - a_k}{1 - \overline{a_k}z}} \in H^2(\mathbb{D}). \]
We note that \( \{B_k\} \) is automatically an orthonormal system, although not necessarily a basis. It turns out that under the maximal selections of \( a_k, k = 1, 2, \ldots \), there holds the convergence:

\[
\tilde{f}(z) = \sum_{k=1}^{\infty} (f_k, e_{a_k}) B_k(z).
\]

Due to the consecutive optimal selections of the parameters \( a_k \) the convergence is in a fast pace. Although on the unit circle the Hardy space functions may not be smooth, it admits a promising convergence rate \([26]\).

**Remark 2.4.** Any sequence \((a_1, \ldots, a_n, \ldots)\) in \( \mathbb{D} \) can define a TM system \( \{B_k\}_{k=1}^{\infty} \) by \((2.19)\). A TM system is alternatively called a rational orthonormal system. In the area of rational approximation, the study of TM systems together with their applications has a long history \([34]\). A TM system is an \( H^p \)-basis, \( 1 < p < \infty \), if and only if \( \sum_{k=1}^{\infty}(1-|a_k|) = \infty \). A half of the Fourier basis, \( \{z^{k-1}\}_{k=1}^{\infty} \), is a particular example of the basis cases. The study \([26]\) opens a new era of use of TM systems through adaptive selections of the parameters according to the signals in practice. The MSP of AFD declares the best selection principle at the one-step selection strategy. This is due to its attainability of the global maximum at each step, that rests, in particular, in the availability of repeating selection of the parameters when needed. AFD shares the same idea as greedy algorithm for the one-step-optimal selection strategy, the latter, however, does not address the issue concerning attainability of the global maximal in the parameters, nor address necessity of repeating selections of the parameters. Through addressing those points missed by greedy algorithm AFD stands as a mathematical theory. AFD found close connections with the Beurling Theorem of the Hardy \( H^2(\mathbb{D}) \) asserting the directional-sum decomposition of the space into shift- and backward shift-invariant subspace:

\[
H^2(\mathbb{D}) = \text{span}\{B_k\}_{k=1}^{\infty} \oplus \phi H^2(\mathbb{D}),
\]

where \( \{B_k\}_{k=1}^{\infty} \) is the TM system and \( \phi \) is the Blaschke product, when can be defined. A sequence of complex numbers \( a_1, \ldots, a_k, \ldots \) in the unit disc can define a Blaschke product having those numbers as its zeros, including multiples, if and only if \( \sum_{k=1}^{\infty}(1-|a_k|) < \infty \). If the sequence cannot define a Blaschke product, then

\[
H^2(\mathbb{D}) = \text{span}\{B_k\}_{k=1}^{\infty}.
\]

With the AFD formulation we know that \( \tilde{f} \in \text{span}\{B_k\}_{k=1}^{\infty} \), the backward shift-invariant subspace in \((2.23)\) or \((2.24)\).

**Remark 2.5.** AFD was initially motivated by intrinsic positive phase derivative decomposition of analytic signals. It automatically generates a fast converging orthogonal expansion of which each entry has a meaningful instantaneous frequency. It has several variations, namely cyclic AFD, unwinding AFD, and be generalized to multi-dimensions with the Clifford and several complex variables setting with scalar- to matrix-valued signals \([17, 27, 35, 36, 30, 1, 2]\). In particular a variation called unwinding Blaschke expansion was studied by Coifman, Steinerberger and Peyrière making further connections with Blaschke products and outer functions \([6, 7]\), being also separately developed in \([18]\), and further developed in a recent paper on maximally unwinding AFD \([28]\). AFD has also been generalized to Hilbert spaces with a dictionary satisfying BVC \([20, 22]\).
The AFD generalization in Hilbert spaces is called pre-orthogonal adaptive Fourier decomposition (POAFD), that in particularly includes Hilbert spaces other than the Hardy type spaces \([23, 24]\). AFD and its variations, as well as its generalizations, have become powerful tools in signal and signal analysis \([10, 11, 5, 9, 37]\).

**Remark 2.6.** In the AFD algorithm, as a consequence of the orthogonality, there hold the relations:

\[
\langle f_k, e_{a_k} \rangle = \langle g_k, B_k \rangle = \langle \tilde{f}, B_k \rangle, \quad k \geq 2,
\]

where

\[
g_k(z) = \tilde{f}(z, w) - \sum_{l=1}^{k-1} \langle f_l, e_{a_l} \rangle B_l(z), \quad k \geq 2
\]

is the \(k\)-th standard remainder. It is the relation (2.25) that allows AFD to be generalized to Hilbert spaces with a dictionary satisfying BVC. In the latter there is no reduced remainder structure, nor explicit TM system as Gram-Schmidt orthogonalization of the Szegő kernels in the underlying Hilbert space.

Next we continue our sparse representation theme. For an analytic random signal \(f\) in \(H^2_w(\mathbb{D})\), we obtain a sequence of parameters \(a_1, a_2, \ldots\), and an associated TM system \(\{B_k\}_{k=1}^\infty\) that gives rise to an AFD sparse representation of the deterministic \(\tilde{f}\). The question is if we use the system \(\{B_k\}_{k=1}^\infty\) to expand the original random signal \(f(e^{it}, w) = f_w(e^{it})\), then in what extent the obtained series can represent the original RS \(f\)? Or namely, what is the difference

\[
d_f(e^{it}, w) = f_w(e^{it}) - \sum_{k=1}^{\infty} \langle f_w, B_k \rangle B_k(e^{it})?
\]

We note that the RS \(d_f\) has dependance on the TM system \(\{B_k\}_{k=1}^\infty\).

In view of the Beurling Theorem, it would well happen that for some \(w\) the difference \(d_f(e^{it}, w)\) is non-zero. We have the following

**Theorem 2.7.** Let \(f \in H^2_w(\mathbb{D})\), \(\tilde{f} = E_w f\), and

\[
\tilde{f} = \sum_{k=0}^{\infty} \langle \tilde{f}, B_k \rangle B_k
\]

be an AFD expansion of \(\tilde{f}\). Then, with the same \(\{B_k\}\),

\[
E_w d_f(e^{it}, w) = 0, \quad \forall t \in [0, 2\pi).
\]

There holds the relation

\[
E_w \| f_w - \sum_{k=1}^{n} \langle f_w, B_k \rangle B_k \|_{H^2_w}^2 = \| d_f \|_{N}^2 + \sum_{k=n+1}^{\infty} E_w |\langle f_w, B_k \rangle|^2,
\]

with

\[
\lim_{n \to \infty} \sum_{k=n+1}^{\infty} E_w |\langle f_w, B_k \rangle|^2 = 0.
\]
And, in terms of the error $\tilde{r} = f - \tilde{f}$ the difference, $d_f$ is estimated

\begin{equation}
(2.31) \quad \|d_f\|_{N'}^2 = \|r\|_{N'}^2 - \sum_{k=1}^{\infty} E_w|\langle r_w, B_k \rangle|^2.
\end{equation}

**Proof** Since $\{B_k\}_{k=1}^{\infty}$ is an orthonormal system in the $N$-space, the projection function $\sum_{k=1}^{\infty} \langle f_w, B_k \rangle B_k$ is in the Hilbert space $N$. The Bessel inequality gives

\[ \sum_{k=1}^{\infty} E_w|\langle f_w, B_k \rangle|^2 \leq \|f\|_{N'}^2, \]

that implies the desired relation (2.30). As a consequence of the Riesz-Fisher Theorem the infinite series

\[ \sum_{k=1}^{\infty} \langle f_w, B_k \rangle B_k \]

is well defined for a. s. $w$ as a function in $H^2_w(D)$. Hence the difference $d_f(w, \cdot)$ belongs to $H^2_w(D)$. All these functions are in $N$.

Since the underlying product measure space $N$ is of finite total measure, both the convergence and the projection function are also in $L^1$. As a consequence of the Fubini Theorem we can first take integral with respect to the probability, and get

\[
E_w(f_w - \sum_{k=1}^{\infty} \langle f_w, B_k \rangle B_k) = \tilde{f} - E_w(\sum_{k=1}^{\infty} \langle f_w, B_k \rangle B_k)
\]

\[
= \tilde{f} - \sum_{k=1}^{\infty} E_w\langle f_w, B_k \rangle B_k
\]

\[
= \tilde{f} - \sum_{k=1}^{\infty} \langle \tilde{f}, B_k \rangle B_k
\]

\[
= 0,
\]

as desired by (2.28).

Noting that for each $w$, $d_f$ is orthogonal with all $B_k$’s, we have the orthogonal decomposition

\[
f_w - \sum_{k=1}^{n} \langle f_w, B_k \rangle B_k = d_f + \sum_{k=n+1}^{\infty} \langle f_w, B_k \rangle B_k,
\]

that implies the desired Pythagoras relation (2.29).

Since

\[
d_f = (f_w - \tilde{f}) - \sum_{k=1}^{\infty} \langle f_w - \tilde{f}, B_k \rangle B_k = r_w - \sum_{k=1}^{\infty} \langle r_w, B_k \rangle B_k,
\]
estimating of \( \|d_f\|^2_N \) proceeds as
\[
\|d_f\|^2_N = E_w \int_0^{2\pi} |r_w(e^{it}) - \sum_{k=1}^{\infty} \langle r_w, B_k \rangle B_k(e^{it})|^2 dt
= E_w \left( \|r_w\|^2_{L^2} - \sum_{k=1}^{\infty} |\langle r_w, B_k \rangle|^2 \right)
= \|r\|^2_N - \sum_{k=1}^{\infty} E_w |\langle r_w, B_k \rangle|^2.
\]

The proof of the theorem is complete. \( \square \)

**Remark 2.8.** The sparse random approximation corresponding to Theorem 2.7 is designed for a deterministic signal with noise. Examples fitting into this theorem include those \( r \) being the white noise. In the following section we develop a sparse representation for analytic random signal that enjoys \( d_f = 0 \) almost surely in \( \Omega \).

### 2.3. The SAFDII: Taking Expectation Secondly.

**Theorem 2.9.** Let \( f \in H^2_w(D) \). Then there exists \( a_1 \in D \) such that
\[
a_1 = \arg \max \{ E_w |\langle f_w, e_a \rangle|^2 \mid a \in D \}.
\]

**Proof** Our effort will be rest on showing that the quantity under study satisfies a *statistical boundary vanishing condition* (SBVC), that is
\[
\lim_{|a| \to 1} E_w |\langle f_w, e_a \rangle|^2 = 0. \tag{2.32}
\]
Then a density argument based on the SBVC concludes the theorem. Since \( f \in \mathcal{N} \), the property
\[
E_w \sum_{k=0}^{\infty} |c_k(w)|^2 < \infty \tag{2.33}
\]
implies that almost surely
\[
\sum_{k=0}^{\infty} |c_k(w)|^2 < \infty.
\]
As a consequence, almost surely \( f_w(z) = \sum_{k=0}^{\infty} c_k(w) z^k \in H^2(D) \). Thanks to the BVC of the classical Hardy space ([26]), we have almost surely
\[
\lim_{|a| \to 1} |\langle f_w, e_a \rangle|^2 = 0. \tag{2.34}
\]
Now we show that there is a positive function of finite expectation dominating \( |\langle f_w, e_a \rangle|^2 \) a.s. in the process \( |a| \to 1 \).
In fact, for any \( a \in D \) almost surely
\[
|\langle f_w, e_a \rangle|^2 \leq \|f_w\|^2 = \sum_{k=0}^{\infty} |c_k(w)|^2.
\]
The last positive random variable function, as a dominating function, has a finite expectation as shown in (2.33). By taking into account (2.34) as well, the Lebesgue domination convergence theorem can be used to conclude the desired SBVC (2.32). The proof is complete. □

The SAFDII proceeds as follows: Guaranteed by the Theorem 2.9, in the same iterative steps as for the classical AFD, one can select, at the $k$-step, an optimal $a_k$:

$$a_k = \arg \max \{E_w | \langle (f_k)_w, e_a \rangle |^2 \ | \ a \in D \},$$

where $f = f_1$, and

$$f_k(z, w) = (f_k)_w(z) = \frac{(f_{k-1})_w(z) - \langle (f_{k-1})_w, e_{a_{k-1}} \rangle e_{a_{k-1}}(z)}{1 - a_{k-1} z}, \quad k \geq 2.$$  

The above maximal selection is called *stochastic maximal selection principle*, abbreviate as SMSP. We then construct a TM system $\{B_k\}_{k=1}^\infty$, as given in (2.19), corresponding to the selected $a_1, a_2, \cdots$, and have the association

$$f(z, w) \sim \sum_{k=1}^\infty \langle f_w, B_k \rangle B_k(z).$$

On the RHS of the last relation we also have

$$\langle (f_k)_w, e_{a_k} \rangle = \langle (g_k)_w, B_k \rangle = \langle f_w, B_k \rangle,$$

where

$$g_k(z, w) = g_k(z, w) = f(z, w) - \sum_{l=1}^{k-1} \langle f_w, B_l \rangle B_l(z), \quad k \geq 2,$$

is the $k$-th standard remainder. The relations (2.36) imply

$$E_w | \langle (f_k)_w, e_{a_k} \rangle |^2 = E_w | \langle (g_k)_w, B_k \rangle |^2 = E_w | \langle f_w, B_k \rangle |^2.$$

The Bessel inequality for $f$ in $N$ with respect to the orthonormal system $\{B_k\}$ implies

$$\lim_{k \to \infty} E_w | \langle f_w, B_k \rangle |^2 = 0.$$

In view of (2.38), the SMSP (2.35) is reduced to the form

$$a_k = \arg \max \{E_w | \langle f_w, B_k^a \rangle |^2 \ | \ a \in D \},$$

where

$$B_k^a(z) = e_a(z) \prod_{l=1}^{k-1} \frac{z - a_l}{1 - a_l z}.$$  

We now prove

**Theorem 2.10.** Let $f(w, e^{iw}) \in H^2_w(D)$ and $(a_1, \cdots, a_n, \cdots)$ be a sequence selected according to the SMSP given in (2.35). Then there holds, in the $N$-norm sense,

$$f(z, w) = \sum_{k=1}^\infty \langle f_w, B_k \rangle B_k(z).$$
Proof By assuming the opposite we prove the convergence through a contradiction. If the RHS does not converge to the LHS, then there is a non-trivial normal RS, \( g \in \mathcal{N} \), such that

\[
(2.42) \quad f(z, w) = \sum_{k=1}^{\infty} \langle f_w, B_k \rangle B_k(z) + g(z, w), \quad \|g\|_\mathcal{N} > 0.
\]

We note that \( g \) is orthogonal with all \( B_1, B_2, \cdots, B_k, \cdots \), and

\[
(2.43) \quad \|g\|^2_\mathcal{N} = \|f\|^2_\mathcal{N} - \sum_{k=1}^{\infty} E_w |\langle f_w, B_k \rangle|^2.
\]

In particular,

\[
(2.44) \quad \lim_{k \to \infty} E_w |\langle f_w, B_k \rangle|^2 = 0.
\]

We show that there exists \( b \in \mathbb{D} \) such that

\[
\langle g_w, e_b \rangle = \delta^2 > 0
\]

for some \( \delta > 0 \). For, if this were not true, then almost surely for all \( b \in \mathbb{D} \)

\[
\langle g_w, e_b \rangle = 0.
\]

Due to the density of \( e_b \) in \( H^2(\mathbb{D}) \) we would have, for each \( w \in \Omega \), \( g_w = 0 \) as a function of \( t \), being contradictory to the condition \( \|g\|_\mathcal{N} > 0 \). We, in particular, can choose \( b \) being different from all the selected \( a_k, k = 1, 2, \cdots \) We in below will fix this \( b \in \mathbb{D} \) and proceed to derive a contradiction.

Set

\[
h_k = - \sum_{l=k}^{\infty} \langle f_w, B_l \rangle B_l.
\]

From the definition of \( g_k \) in (2.37), there follows the orthogonal decomposition

\[
g = g_k + h_k.
\]

The Bessel inequality implies, when \( k \) is large,

\[
E_w |\langle h_k, e_b \rangle|^2 \leq E_w \|h_k\|^2 \leq \delta^2 / 4.
\]

Hence

\[
2E_w |\langle g_k, e_b \rangle|^2 + \delta^2 / 2 \geq E_w |\langle g_k, e_b \rangle + \langle h_k, e_b \rangle|^2 = \delta^2,
\]

which implies

\[
E_w |\langle g_k, e_b \rangle|^2 \geq \delta^2 / 4.
\]

Due to the reproducing kernel property of \( e_b \), for a large \( k \),

\[
(2.45) \quad (1 - |b|^2)^2 E_w |\langle g_k \rangle(b)|^2 \geq \delta^2 / 4.
\]

Since pointwise there hold

\[
(2.46) \quad f_k = g_k / B_k \quad \text{and} \quad |B_k(b)| < 1,
\]

there follows \( |f_k| \geq |g_k| \). Therefore,

\[
(1 - |b|^2)^2 E_w |\langle f_k \rangle(b)|^2 \geq \delta^2 / 4.
\]

By using the reproducing property of \( e_b \) again, the inner product form of the last equality has the form

\[
E_w |\langle f_k, e_b \rangle|^2 = E_w |\langle f_w, B_k \rangle|^2 \geq \delta^2 / 4
\]
Remark 2.11. The proof is crucially based on the relation (2.46), which makes the one originally used in [26] adaptable to the associated stochastic case. Among the succeeded generalizations of AFD, some are based on the special Blaschke product structure in the context, the latter including the matrix-valued Hardy space over the unit disc, the Drury-Arveson Space of several complex variables ([1, 2]), and over the n-torus with the product basis formulation (20). At least for those in which (2.46) is valid the AFD can also have a stochastic version with a similar proof as above. In the following section we study the formulation in which Blaschke products are not crucial.

3. Stochastic SPOAFDs in Hilbert Spaces

Our discussions on stochastic Hilbert spaces will be based on one on deterministic Hilbert spaces, the latter being assumed to have a dictionary satisfying BVC. For the self-containing purpose we give a brief exposition on POAFD algorithm for deterministic signals ([22], also see [20, 21, 5]).

3.1. POAFD in a Hilbert Space With a Dictionary Satisfying BVC. The classical formulation of sparse representation of a Hilbert space is often under the assumption that the space has a dictionary that, by definition, is a dense subset of elements of the space of which each has unit norm. The unit norm requirement for a dictionary is not essential. We, therefore, release the norm-one requirement and only assume that the underlying Hilbert space $H$ has a dense subclass of elements $K_q, q \in E$, where $E$ is an open set of the complex plane, or more generally an open set of a product space between $R^d$ and $C^d$, the latter denoting the real or complex Euclidean spaces, respectively. We denote the normalizations of $K_q$ by $E_q$, where $E_q = K_q/\|K_q\|$, $q \in E$. Below we often call the $K_q$’s by kernels. We now define what we call by “multiple kernels”. Let $(q_1, \ldots, q_n)$ be any $n$-tuple of parameters in $E$. Each of the terms $q_k, k = 1, \ldots, n$, may has multiplicity in the $k$-tuple $(q_1, \ldots, q_k)$. We denote by $l(k)$ the multiplicity of $q_k$ in $(q_1, \ldots, q_k)$. We accordingly introduce what we call multiple kernels as follows. For any $k \leq n$, denote

$$\tilde{K}_k = \left( \frac{\partial^{(l(k))} K_q}{\partial q} \right) (q_k),$$

where $l(k)$ is the multiple of $q_k$ in $(q_1, \ldots, q_k)$. With a little abuse of the notation, we will also denote $\tilde{K}_k$ by $\tilde{K}_{q_k}, k = 1, 2, \ldots, n$, indicating the parameter sequence in use. The concept multiple kernel is a necessity of the pre-orthogonal maximal selection principle (POMSP): Suppose we already have an $(n-1)$-tuple $\{q_1, \ldots, q_{n-1}\}$, with repetition or without, corresponding to the $(n-1)$-tuple $\{K_{q_1}, \ldots, K_{q_{n-1}}\}$. By doing the G-S orthonormalization process consecutively we obtain an equivalent $(n-1)$-orthonormal basis $\{B_1, \ldots, B_{n-1}\}$. For any given $G$ in the Hilbert space we wish to find a $q_n$ that gives rise to supreme value

$$\sup\{ |\langle G, B_n^q \rangle | : q \in E, q \neq q_1, \ldots, q_{n-1} \},$$

where the finiteness of the supreme is guaranteed by the Cauchy-Schwartz inequality, and $B_n^q$ be such that $\{B_1, \ldots, B_{n-1}, B_n^q\}$ is the G-S orthonormalization of $\{\tilde{K}_{q_1}, \ldots, \tilde{K}_{q_{n-1}}, K_q\}$,
where $B^q_n$ is precisely given by
\begin{equation}
B^q_n = \frac{K_q - \sum_{k=1}^{n-1} \langle K_q, B_k \rangle H B_k}{\sqrt{\|K_q\|^2 - \sum_{k=1}^{n-1} |\langle K_q, B_k \rangle H|^2}}.
\end{equation}

The second crucial ingredient of POAFD that is *Boundary Vanishing Condition (BVC)* in the context: For any but fixed $G \in H$, if $p_n \in E$ and $p_n \to \partial E$ (including $\infty$ if $E$ is unbounded while in the case we use the compactification topology for the added infinity point), then
\[
\lim_{n \to \infty} |\langle G, E_{p_n} \rangle| = 0.
\]

Under BVC a compact argument leads that there exists a point $q_n \in E$ and $q^{(l)}, l = 1, 2, \cdots$, such that $q^{(l)}$ are all different from $q_1, \cdots, q_{n-1}$, $\lim_{l \to \infty} q^{(l)} = q_n$, and
\begin{equation}
\lim_{l \to \infty} |\langle G, B^{q(l)}_n \rangle| = \sup\{|\langle G, B^q_n \rangle| : q \in E, q \neq q_1, \cdots, q_{n-1}\} = |\langle G, B^{q_\infty}_n \rangle|,
\end{equation}
where
\begin{equation}
B^{q_\infty}_n = \frac{\tilde{K}_q - \sum_{k=1}^{n-1} \langle \tilde{K}_{q_n}, B_k \rangle H B_k}{\sqrt{\|\tilde{K}_q\|^2 - \sum_{k=1}^{n-1} |\langle \tilde{K}_{q_n}, B_k \rangle H|^2}}.
\end{equation}

BVC and multiple kernels both are unavoidable for existence of such $q_n$ and thus for the availability of POAFD method: We iteratively apply the above process to $G = G_n$, where $G_n$ is the standard remainder
\[
G_n = F - \sum_{k=1}^{n-1} \langle F, B_k \rangle B_k,
\]
and $(B_1, \cdots, B_n)$ is the G-S orthogonalization of $(\tilde{K}_{q_1}, \cdots, \tilde{K}_{q_n})$. Under the consecutive maximal selections of $\left\{ q_k \right\}_{k=1}^{\infty}$ one eventually obtains, with a fast convergent pace,
\begin{equation}
F = \sum_{k=1}^{\infty} \langle F, B_k \rangle H B_k
\end{equation}
(\cite{20, 21, 25}).

**Remark 3.1.** We note that repeating selections of parameters can be avoided in practice. By definition of supreme, for any $\rho \in (0, 1)$, a parameter $q_n \in E$ can be found, different from the previously selected $q_k, k = 1, \cdots, n - 1$, to have
\begin{equation}
|\langle G_n, B^{q_\infty}_n \rangle| \geq \rho \sup\{|\langle G_n, B^q_n \rangle| : q \in E, q \neq q_1, \cdots, q_{n-1}\}.
\end{equation}

The corresponding algorithm for consecutively finding such a sequence $\{q_n\}_{n=1}^{\infty}$ is called *Weak Pre-orthogonal Adaptive Fourier Decomposition (WPOAFD)*. With WPOAFD one may choose all $q_1, \cdots$ being distinguished. Under such selections we still get convergence \textbf{(3.50)} with a little less fast pace.

**Remark 3.2.** An order $O(1/\sqrt{n})$ of the convergence rate can be proved: For $M > 0$, by defining
\begin{equation}
\mathcal{M}_M = \{ F \in H : \exists \{ c_n \}, \{ E_{q_n} \} s. t. F = \sum_{n=1}^{\infty} c_n E_{q_n} \text{ with } \sum_{n=1}^{\infty} |c_n| \leq M \},
\end{equation}
for any $F \in \mathcal{M}_M$, the POAFD partial sums satisfy
\[ \| F - \sum_{k=1}^{n} \langle F, B_k \rangle H B_k \|_H \leq \frac{M}{\sqrt{n}}. \]

We note that the above convergence rate is the same as that of the Shannon expansion into the sinc functions of bandlimited entire functions. In the POAFD case the orthonormal system $\{B_1, \cdots, B_n, \cdots\}$ is not necessarily a basis but a system adapted to the given function $F$. For the Hardy space case, due to the relations in (2.25), the MSP (2.20) of AFD reduces to the MSP (3.43) of POAFD, and thus AFD reduces to POAFD. The algorithm codes of AFD and POAFD, as well as those of several related ones are available at request (http://www.fst.umac.mo/en/staff/fsttq.html).

Remark 3.3. AFD and POAFD have been seen to have two directions of developments. One is $n$-best kernel expansion. That is to determine $n$-parameters at one time, being obviously of better optimality in the sparse kernel approximation. The $n$-best approximation is motivated by the classical problem, yet still open in its ultimate global algorithm, called the best approximation to Hardy space functions by rational functions of degree not exceeding $n$ ([3, 4, 29]). The gradient descending method for cyclic AFD ([29]) and cyclic AFD separately ([17]) may be adopted to give practical (not mathematical) $n$-best algorithms in Hilbert spaces with a dictionary satisfying BVC. The second direction of development of POAFD is related to exploration of Blaschke product-like functions and interpolation type problems in general Hilbert spaces. For related publications see [18, 6, 1, 2, 28].

3.2. Stochastic POAFDs. Let $\mathcal{H}$ be a Hilbert space with a dense subset $\{K_q\}$ parameterized in an open set $E : q \in E$. We assume that the dictionary satisfies BVC
\[ (3.53) \quad \lim_{q \to \partial E} |\langle F, E_q \rangle| = 0, \]
where $E_q = K_q/\|K_q\|$. Let us consider random signals $F(t, w), t \in T, w \in \Omega$, where for a.s. $w \in \Omega, F(\cdot, w) \in \mathcal{H}$; and for any $t \in T, F(t, \cdot)$ is a random variable. Define
\[ (3.54) \quad \mathcal{N}(\mathcal{H}, \Omega) = \{ F(t, w) : F(\cdot, w) \in \mathcal{H}, \text{for a.s. } w; \text{and } F(t, \cdot) \text{ being a random variable for each fixed } t, \text{ and } E_w\|F(\cdot, w)\|_H < \infty. \} \]
This formulation governs two types of stochastic POAFDs, abbreviated as SPOAFDI and SPOAFDII.

SPOAFDI is one to treat a noised deterministic signal by first taking the expectation and then doing maximal energy extractions. We need to show $E_wF(t, w) \in \mathcal{H}$. Following what is done in (2.14), by using the Minkovski inequality followed by the H"{o}lder inequality, we get
\[ \|E_wF(\cdot, w)\|_{\mathcal{H}} \leq E_w\|F(\cdot, w)\|_{\mathcal{H}} \leq (E_w\|F(\cdot, w)\|_{\mathcal{H}}^2)^{1/2} = \|F\|_{\mathcal{N}(\mathcal{H}, \Omega)} < \infty. \]
This shows that the expectation belongs to the underlying Hilbert space $\mathcal{H}$. Since $\mathcal{H}$ has a dictionary that satisfies BVC one can perform POAFD in $\mathcal{H}$. The difference $d(t, w) = F(t, w) - E_w F(\cdot, w)$ enjoys the zero-expectation property and all the related quantities may be analyzed as in the subsection 2.2. This approach gives rise to the type SPOAFDI that is suitable for analyzing signals corrupted with noise of zero expectation and of a small $\mathcal{N}(\mathcal{H}, \Omega)$ norm.
To perform the SPOAFDII type algorithm we first need to prove the stochastic boundary vanishing condition, or SBVC,

\[ \lim_{q \to \partial E} E_w |\langle F_w, E_q \rangle|^2 = 0. \]

To show this we still use the Lebesgue Dominated Convergence Theorem in the probability space, through showing

1. For a.s. \( w \in \Omega \)

\[ \lim_{q \to \partial E} |\langle F_w, E_q \rangle|^2 = 0; \]

and,

2. For all \( q \) the function \( |\langle F_w, E_q \rangle|^2 \) is dominated by a positive integrable function in the probability space.

The property 1 is a consequence of BVC of the dictionary \( \{ E_q \}_{q \in E} \) in \( \mathcal{H} \). To show 2, we have, by the Hölder inequality,

\[ E_w|\langle F_w, E_q \rangle|^2 \leq E_w \| F_w \|^2 = \| F \|_{\mathcal{N}(\mathcal{H}, \Omega)}^2 < \infty, \]

where \( \| F_w \|^2 \) is the dominating function in the probability space. The SBVC is hence proved.

Based on the just proved SBVC we have the following theorem.

**Theorem 3.4.** Let \( F(t, w) \in \mathcal{N}(\mathcal{H}, \Omega) \) and \( (q_1, \cdots, q_n, \cdots) \) be a consecutively selected kernel sequence under SMSP

\[ q_k = \arg \sup \{ E_w |\langle (G_k)_w, B^q_k \rangle|^2 | q \in E \}, \]

where

\[ (G_k)_w = F_w - \sum_{l=1}^{k-1} \langle F_w, B_l \rangle B_l, \]

and \( (B_1, \cdots, B_{k-1}, B_k) \) is the G-S orthonormalization of \( (B_1, \cdots, B_{k-1}, \tilde{K}_{q_k}) \). Then there holds, in the \( \mathcal{N}(\mathcal{H}, \Omega) \)-norm sense,

\[ F(z, w) = \sum_{k=1}^{\infty} \langle F_w, B_k \rangle B_k(z). \quad (3.55) \]

**Remark 3.5.** The proof of Theorem 2.10 crucially depends on the property \( |B(z)| \leq 1 \) of the classical Blaschke products. In the general Hilbert spaces case there may not exist Blaschke product-like functions, and, when there exist such functions, say \( B \), they may not enjoy the property \( |B(z)| \leq 1 \). Below we give a proof of Theorem 3.4 that does not depend on Blaschke product-like functions. The proof is an adaptation of one for the deterministic signal case (see [19] or [21], or [5], the last being essentially the English equivalence of the former).

**Proof of Theorem 3.4** We will prove the theorem by contradiction. If the RHS series of (3.55) does not converge to the LHS function, then there is a non-trivial random signal
$H \in \mathcal{N}(\mathcal{H}, \Omega)$ such that

\begin{equation}
F(t, w) = \sum_{k=1}^{\infty} \langle F_w, B_k \rangle B_k(z) + H(z, w), \quad \|H\|_{\mathcal{N}(\mathcal{H}, \Omega)} > 0.
\end{equation}

We note that $H$ is orthogonal with all $B_1, B_2, \cdots, B_k, \cdots$, and

\begin{equation}
0 < \|H\|_{\mathcal{N}(\mathcal{H}, \Omega)}^2 = \|F\|_{\mathcal{N}(\mathcal{H}, \Omega)}^2 - \sum_{k=1}^{\infty} E_w|\langle F_w, B_k \rangle|^2.
\end{equation}

We claim that the fact $\|H\|_{\mathcal{N}(\mathcal{H}, \Omega)} > 0$ implies that there exists $q \in \mathbb{E}$ such that

$E_w|\langle H_w, E_q \rangle|^2 = \delta^2 > 0,$

for some $\delta > 0$. For, if this were not true, then almost surely for all $q \in \mathbb{E}$

$\langle H_w, E_q \rangle = 0.$

Due to the density of $K_q$ in $\mathcal{N}(\mathcal{H}, \Omega)$ we would have almost surely $H_w = 0$ as a function of $t$, being contradictory to the condition $\|H\|_{\mathcal{N}(\mathcal{H}, \Omega)} > 0$. We, in particular, can choose $q$ being distinguished from all the selected $q_k, k = 1, 2, \cdots$. In below such $q \in \mathbb{E}$ will be fixed. The following argument will lead to a contradiction with the selections of $q_M$ for large enough $M$.

Based on the notation $G_k$ for standard remainders defined in the theorem we rewrite the relation (3.56) as

\begin{align*}
F_w &= \left( \sum_{k=1}^{M} + \sum_{k=M+1}^{\infty} \right) \langle (G_k)w, B_k \rangle B_k + H \\
&= \sum_{k=1}^{M} \langle (G_k)w, B_k \rangle B_k + \tilde{G}_{M+1} + H \\
&= \sum_{k=1}^{M} \langle (G_k)w, B_k \rangle B_k + G_{M+1},
\end{align*}

where

$\tilde{G}_{M+1} = \sum_{k=M+1}^{\infty} \langle (G_k)w, B_k \rangle B_k$ and $G_{M+1} = \tilde{G}_{M+1} + H.$

The Bessel inequality implies

\begin{equation}
\lim_{M \to \infty} \|\tilde{G}_{M+1}\|_{\mathcal{N}(\mathcal{H}, \Omega)} = 0.
\end{equation}

On one hand, we have, from (2.39), for a large $M$,

\begin{equation}
E_w|\langle (G_{M+1})w, B_{M+1} \rangle|^2 = E_w|\langle F_w, B_{M+1} \rangle|^2 = E_w|\langle F_w, B_{M+1}^q \rangle|^2 < \delta^2/16.
\end{equation}

On the other hand, we can show, for large $M$, there holds

\begin{equation}
E_w|\langle (G_{M+1})w, B_{M+1}^q \rangle|^2 > 9\delta^2/16,
\end{equation}

where $B_{M+1}^q$ is the last function of the Gram-Schmidt orthonormalization of the $(M + 1)$-system $(B_1, B_2, \cdots, B_M, K_q)$ in the given order. From the triangle inequality of the
The energy is then
\[ E_w|\langle (G_{M+1})_w, B_{M+1}^q \rangle|^2 \leq \|G_{M+1}\|_{\mathcal{N}(H,\Omega)}^2 \leq \delta^2/16. \]
Using the Cauchy-Schwarz inequality and then (3.58), for large enough \( M \) we have
\[ E_w|\langle (G_{M+1})_w, B_{M+1}^q \rangle|^2 \geq \left( E_w|\langle H_w, B_{M+1}^q \rangle|^2 \right)^{1/2} - \left( E_w|\langle (G_{M+1})_w, B_{M+1}^q \rangle|^2 \right)^{1/2} \]

Therefore,
\[ (3.61) \quad E_w|\langle (G_{M+1})_w, B_{M+1}^q \rangle|^2 \geq \left( E_w|\langle H_w, B_{M+1}^q \rangle|^2 \right)^{1/2} - \delta/4. \]
Next we compute the energy of the projection of \( H_w \) into the span of \( \{B_1, \cdots, B_M, E_q\} \). The energy is then
\[ E_w|\langle H_w, B_{M+1}^q \rangle|^2, \]
as \( H_w \) is orthogonal with \( B_1, \cdots, B_M. \) However, the span is just the same if we alter the order \( \{B_1, \cdots, B_M, E_q\} \) to \( \{E_q, B_1, \cdots, B_M\} \). As a consequence, the energy of the projection into the span is surely not less than the energy of \( H_w \) projected onto the first function \( E_q \). This gives rise to the relation
\[ E_w|\langle H_w, B_{M+1}^q \rangle|^2 \geq E_w|\langle H_w, E_q \rangle|^2 = \delta^2. \]
Combining with (3.61), we have
\[ \left( E_w|\langle (G_{M+1})_w, B_{M+1}^q \rangle|^2 \right)^{1/2} \geq 3\delta/4. \]
Thus we proved (3.60) that is contradictory with (3.59). This shows that the selection of \( q_{M+1} \) did not obey SMSP, for we would better select \( q \) instead of \( q_{M+1} \) at the \( (M + 1) \)-th step. The proof of the theorem is hence complete.

**Remark 3.6.** Theorem 2.10 and Theorem 3.4 have separate proofs. Theorem 2.10 is, as a matter of fact, a special case of Theorem 3.4. The question is whether the former can refer to the latter for its validity. The answer is “Yes” but one has to do some work before refer to Theorem 3.4. In 2.10 we do not use G-S orthogonalization, but backward shift process for the orthogonality. Whether the two methodologies result in the same orthonormal system? In Appendix we prove that the TM system, obtained in AFD through a backward shift process on the Szegő kernel, coincides with the result of the G-S orthogonalization on the same kernels. This facilitates the above “Yes” answer. Precisely, we will prove

**Theorem 3.7.** Let \( \{a_1, \cdots, a_n\} \) be any \( n \)-tuple of parameters in \( \mathbb{D} \) in which multiplicities are allowed. Denote by \( l(m) \) the multiplicity of \( a_m \) in the \( m \)-tuple \( \{a_1, \cdots, a_m\}, 1 \leq m \leq n. \) For each \( m, \) denote by
\[ \tilde{k}_{a_m}(z) = \frac{\frac{\partial^{l(m)-1}}{(\log a)^{l(m)-1}} k_a(z)}{\big| a=a_m, \ \text{where} \ k_a(z) = \frac{1}{1-a^z}. \]

Then the Gram-Schmidt orthonormalization of \( \{\tilde{k}_{a_1}, \cdots, \tilde{k}_{a_m}\} \) in the given order coincides with the \( m \)-TM system \( \{B_1, \cdots, B_m\} \) (2.19) defined through the ordered \( m \)-tuple \( \{a_1, \cdots, a_m\} \).

There surely exist different proofs for this result. In Appendix we give a constructive proof. As far as the author is aware of, the unit disc and a half of the complex plane are the only cases to which the equivalence of the two processes, i.e., the Blaschke product formulation and the G-S orthogonalization, has been proved.
4. Acknowledgement

The author wishes to express his sincere thankfulness to Dr Chen Wei-Guo, Dr Wang Shi-Lin, Dr Cheng Han-Sheng, Prof Chen Qiu-Hui, Prof Leong Ieng Tak for their interest and encouragement to study this topic, and useful comments on draft material of this article.

5. Appendix

Proof of Theorem 3.7 Denote the canonical Blaschke product determined by \( a_1, \ldots, a_m \) by

\[
\phi_{a_1, \ldots, a_m}(z) = \prod_{l=1}^{m} \frac{z - a_l}{1 - \overline{a_l}z}.
\]

We first show that for any \( a \in D \) being different from \( a_1, \ldots, a_{m-1} \) there holds

\[
k_a(z) - \sum_{l=1}^{m-1} \langle k_a, B_l \rangle B_l(z) = \overline{\phi_{a_1, \ldots, a_{m-1}}(a)} \phi_{a_1, \ldots, a_{m-1}}(z) k_a(z).
\]

For this aim we use mathematical induction. First we verify the case \( m = 2 \). Using the reproducing kernel property of \( k_a \), there follows

\[
k_a - \langle k_a, B_1 \rangle B_1(z) = \frac{1}{1 - \overline{a}z} - \overline{B_1(a)} B_1(z)
\]

\[
= \frac{1}{1 - \overline{a}z} - \frac{\alpha}{1 - \overline{a_1}z}, \quad \alpha = \frac{1 - |a|²}{1 - a \overline{a}},
\]

\[
= \frac{1 - \alpha a}{1 - \overline{a_1}z} = \frac{1}{1 - a \overline{a} - a_1 \overline{a}} = \phi_{a_1}(a) \phi_{a_1}(z) k_a(z).
\]

Assume that (5.62) holds for \( m \) being replaced by \( m - 1 \). Under this inductive hypothesis, we have

\[
k_a(z) - \sum_{l=1}^{m-1} \langle k_a, B_l \rangle B_l(z) = [k_a(z) - \sum_{l=1}^{m-2} \langle k_a, B_l \rangle B_l(z)] - \langle k_a, B_{m-1} \rangle B_{m-1}(z)
\]

\[
= \overline{\phi_{a_1, \ldots, a_{m-2}}(a)} \phi_{a_1, \ldots, a_{m-2}}(z) k_a(z) - \langle k_a, B_{m-1} \rangle B_{m-1}(z)
\]

\[
= \overline{\phi_{a_1, \ldots, a_{m-2}}(a)} \phi_{a_1, \ldots, a_{m-2}}(z) k_a(z) - \overline{B_{m-1}(a)} B_{m-1}(z)
\]

\[
= \overline{\phi_{a_1, \ldots, a_{m-2}}(a)} \phi_{a_1, \ldots, a_{m-2}}(z) k_a(z) \left[ k_a(z) - \frac{1 - |a_{m-1}|²}{(1 - a_{m-1} \overline{a})(1 - \overline{a_{m-1}}z)} \right]
\]

\[
= \overline{\phi_{a_1, \ldots, a_{m-1}}(a)} \phi_{a_1, \ldots, a_{m-1}}(z) k_a(z).
\]

We hence proved (5.62). Next we deal with the orthonormalization allowing repetition of the parameters. Now we are with the new inductive hypothesis that the Gram-Schmidt orthonormalization of \( \{ k_{a_1}, \ldots, k_{a_{m-1}} \} \) is the \( (m-1) \)-TM system \( \{ B_1, \ldots, B_{m-1} \} \). First assume \( a_m \) is different from all the preceding \( a_k, k = 1, \ldots, m - 1 \). In (5.62) let \( a = a_m \). By taking the norm on the both sides of (5.62) and invoking the orthonormality of the
where \( c \) is a real number depending on \( a_m \) and \( a_1, \ldots, a_{m-1} \). We thus conclude that

\[
\frac{k_{am}(z) - \sum_{l=1}^{m-1} \langle k_{am}, B_l \rangle B_l(z)}{\|k_{am}(z) - \sum_{l=1}^{m-1} \langle k_{am}, B_l \rangle B_l(z)\|} = e^{ie_\phi_{a_1, \ldots, a_{m-1}}(z)} e_{am}(z).
\]

Note that here we have the case \( k_{am} = \vec{k}_{am} \) and \( l(m) = 1 \). Next we extend the above relation to the cases that \( a = a_m \) coincides with some of the preceding \( a_1, \ldots, a_{m-1} \). In that case we have \( l(m) > 1 \), and we are to show

\[
\frac{\vec{k}_{am}(z) - \sum_{l=1}^{m-1} (\vec{k}_{am}, B_l) B_l(z)}{\|\vec{k}_{am}(z) - \sum_{l=1}^{m-1} (\vec{k}_{am}, B_l) B_l(z)\|} = e^{ie_\phi_{a_1, \ldots, a_{m-1}}(z)} e_{am}(z),
\]

where \( c \) depends on \( a_1, \ldots, a_m \). For \( b \) being sufficiently close to \( a_m \) in \( D \) we have up to the \((l(m) - 1)\)-order power series expansion in the variable \( b \):

\[
k_b(z) = \sum_{l=0}^{l(m)-1} \frac{1}{l!} \left[ \frac{\partial}{\partial a} \right]^l k_a(z)|_{a=a_m} (b - a_m)^l + o((b - a_m)^{(l(m)-1)}
\]

\[
= T(z) + \frac{1}{(l(m) - 1)!} \vec{k}_{am}(z)(b - a_m)^{(l(m)-1)} + o((b - a_m)^{(l(m)-1)}),
\]

where

\[
T(z) = \sum_{l=0}^{l(m)-2} \frac{1}{l!} \left[ \frac{\partial}{\partial a} \right]^l k_a(z)|_{a=a_m} (b - a_m)^l.
\]

Now, according to the inductive hypothesis, \( B_1, \ldots, B_{m-1} \) involve the derivatives of the reproducing kernel up to the \((l(m) - 2)\)-order, and hence

\[
T(z) - \sum_{k=1}^{m-1} \langle T, B_k \rangle B_k = 0.
\]

Inserting the left-hand-side of (5.65) into (5.63), where \( a_m \) is replaced by \( b \) with \( b \rightarrow a_m \) horizontally (meaning that \( \text{Im}(b) = \text{Im}(a_m) \)), while dividing by \((b - a_m)^{(l(m)-1)} > 0 \), we have

\[
\frac{k_b(z) - T(z)}{(b - a_m)^{(l(m)-1)}} - \sum_{l=1}^{m-1} \left( \frac{k_b - T}{(b - a_m)^{(l(m)-1)}}, B_l \right) B_l(z) = e^{ie_\phi_{a_1, \ldots, a_{m-1}}(z)} k_b(z).
\]

Letting \( b - a_m \downarrow 0 \) and noticing that the Taylor series remainder is an infinitesimal of an order higher than \((b - a_m)^{(l(m)-1)} \), we obtain the desired relation (5.64). The proof is complete. \( \square \)
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):2089C2106.

[2] D. Alpay, F. Colombo, T. Qian, and I. Sabadini, *Adaptive 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] Baratchart, L., Cardelli, M., Olivi, M., Identification and rational L2 approximation a gradient algorithm. Automatica, 1991, 27(2), 413-417.

[5] Q.-H. Chen, T. Qian, L.-H. Tan, *A Theory on Non-Constant Frequency Decompositions and Applications*, In: Advancements in Complex Analysis: From Theory to Practice, D. Breaz and M. Th. Rassias (Eds.), Springer, (to appear).

[6] R. Coifman, S. Steinerberger, *Nonlinear phase unwinding of functions*, J Fourier Anal Appl. 2017;23(4):778-809.

[7] R. Coifman, J. Peyrière, *Phase unwinding, or invariant subspace decompositions of Hardy spaces*, J Fourier Anal Appl 25, 684-695 (2019).

[8] P. Dang, W. X. Mai, T. Qian, *Fourier spectrum of Clifford $H_p$ spaces on $\mathbb{R}^{n+1}_+$ for 1p*, J. Math. Anal. Appl. 483 (2020), no. 1, 123598, 19 pp.

[9] Y. T. Li, L. M. Zhang, T. Qian, *2D Partial Unwinding - A Novel Non-Linear Phase Decomposition of Images*, IEEE Transactions on Image Processing, 2019, DOI: 10.1109/TIP.2019.2914000.

[10] W. Mi, T. Qian, *Frequency-domain identification: An algorithm based on an adaptive rational orthogonal system*, Automatica, 2012, 48(6), 1154-1162.

[11] W. Mi, T. Qian, “On backward shift algorithm for estimating poles of systems,” Automatica, 2014, 50(6), 1603-1610.

[12] *Rational Approximation in Hardy Spaces on Strips*, Complex Variables and Elliptic Equations, 63 (2018), 1721-1738.

[13] *Aveiro Method in Reproducing Kernel Hilbert Spaces under Complete Dictionary*, Mathematical Methods in the Applied Sciences, 40 (2017), 7240-7254.

[14] *The Fourier Type Expansions on Tubes*, submitted for publication, 2018.

[15] E. D. Livshitz, V. N. Temlyakov, *On convergence of weak greedy algorithms*, SOUTH CAROLINA UNIV COLUMBIA DEPT OF MATHEMATICS, 2000.

[16] S. Mallat, Z. Zhang, *Matching pursuits with time-frequency dictionaries*, IEEE Trans. Signal Process. 41, 3397C3415 (1993)

[17] T. Qian, *Cyclic AF D Algorithm for Best Rational Approximation*, Mathematical Methods in the Applied Sciences, 2014, 37(6), 846-859.

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

[19] T. Qian, *Adaptive Fourier Decomposition*, Beijing: Science Press, 2015.

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

[21] T. Qian, *A novel Fourier theory on non-linear phase and applications*, Advances in Mathematics (China), 2018 47(3), pp 321-347 (in Chinese).

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

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

[24] 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.

[25] T. Qian, L.-H. Tan, J.-C. Chen, *A class of iterative greedy algorithmsrelated to Blaschke products*, accepted by SCIENCE CHINA, Mathematics.

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

[27] T. Qian, W. Sproessig, J. X. Wang, *Adaptive Fourier decomposition of functions in quaternionic Hardy spaces*, Mathematical Methods in the Applied Sciences, 2012, 35(1): 43C64.
[28] T. Qian, L.-H. Tan, J.-C. Chen, A class of iterative greedy algorithms related to Blaschke product, accepted to appear in Science China, Mathematics.

[29] T. Qian, J. Z. Wang, W. X. Mai, An Enhancement Algorithm for Cyclic Adaptive Fourier Decomposition, Applied and Computational Harmonic Analysis, available online 19 January 2019.

[30] T. Qian, J. X. Wang, Y. Yang, Matching Pursuits among Shifted Cauchy Kernels in Higher-Dimensional Spaces, Acta Mathematica Scientia, 2014, 34(3): 660-672.

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

[32] S. Saitoh, Y. Sawano, Theory of Reproducing Kernels and Applications, Singapore: Springer, 2016.

[33] E. M. Stein, G. Weiss, Introduction to fourier analysis on Euclidean spaces, Princeton University Press, 1971.

[34] J. L. Walsh, Interpolation and approximation by rational functions in the complex domain, American Mathematical Soc. Publication, vol 10 (1962).

[35] J. X. Wang, T. Qian, Approximation of Functions by Higher Order Szegö Kernels I. Complex Variable Cases, Complex Variables and Elliptic Equations, 2015, 60(6), 733-747.

[36] J. X. Wang, T. Qian, Approximation of monogenic functions by higher order Szegö kernels on the unit ball and the upper half space, Sciences in China: Mathematics, 2014, 57(9), 1785-1797.

[37] 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.

Tao QIAN, Macao Institute of Systems Engineering, Macau University of Science and Technology, Macau
E-mail address: tqian@must.edu.mo