SPARSE REPRESENTATIONS OF RANDOM SIGNALS

TAO QIAN*

Abstract. Sparse (fast) representations of deterministic signals have been well studied. Among other types there exists one called adaptive Fourier decomposition (AFD) for functions in analytic Hardy spaces. Through the Hardy space decomposition of the L^2 space the AFD algorithm also gives rise to sparse representations of signals of finite energy. To deal with multivariate signals the general Hilbert space context comes into play. The multivariate counterpart of AFD in general Hilbert spaces with a dictionary has been named pre-orthogonal AFD (POAFD). In the present study we generalize AFD and POAFD to random analytic signals through formulating stochastic analytic Hardy spaces and stochastic Hilbert 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) those expressible as the sum of a deterministic signal and an error term (SAFDI); and for (ii) those from different sources obeying certain distributive law (SAFDII). In the later part of the paper we drop off 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 AFDs and POAFDs for deterministic signals, the developed stochastic POAFD algorithms offer powerful tools to approximate and thus to analyze random signals.

AMS Classification: 42A16; 42A20; 41A30; 30B99; 34K28; 35A35

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

Contents

*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 $\|f\|^2 = \sum_{-\infty}^{\infty} |c_k|^2$, where the L^2-norm is defined from 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

$$f^+(z) = \sum_{k=0}^{\infty} c_k z^k \quad \text{and} \quad f^-(z) = \sum_{k=-1}^{-\infty} c_k z^k$$

are analytic in $\mathbf{D}$ and in $\mathbf{C}\setminus\overline{\mathbf{D}}$, respectively, where $\mathbf{D}$ stands for the open unit disc in the complex plane $\mathbf{C}$. 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 transformation on the circle:

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

where $\mathrm{sgn}(k) = k/|k|$ when $k \neq 0$ and $\mathrm{sgn}(0) = 0$. 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,

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

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 $\mathbf{D}$, complex analytic methods are available for f^+. The totality of such analytic functions $f^+(z)$ in the disc is identical with the function space

$$\begin{aligned} H^2(\mathbf{D}) &\triangleq \{f : \mathbf{D} \to \mathbf{C} \mid f \text{ is analytic and } f(z) = \sum_{k=0}^{\infty} c_k z^k \text{ with } \sum_{k=0}^{\infty} |c_k|^2 < \infty\} \\ &= \{f : \mathbf{D} \to \mathbf{C} \mid f \text{ is analytic and } \sup_{0<r<1} \int_0^{2\pi} |f(re^{it})|^2 dt < \infty\}, \end{aligned} \tag{1.1}$$

called the (*complex analytic*) *Hardy H^2-space* in the unit disc. There exist other complex analytic Hardy spaces with more or less parallel theories as the one defined in the disc. In other words, the Hardy space idea to study functions may be extended to signals defined on the real line $\mathbf{R}$, or to those defined on manifolds in the higher dimensional complex spaces $\mathbf{C}^d$ in the several complex variables setting (e.g., the Hardy spaces on the n-torus [34], or the DruryCArveson space or the Hardy space of the solid ball in several complex variables [2]), or to those in the real-Euclidean spaces $\mathbf{R}^d$ in the Clifford algebra setting (the conjugate harmonic systems, [34, 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 $H^2(\mathbf{D})$ as an example to explain the adaptive Fourier decomposition (AFD) theory. In below we often abbreviate $H^2(\mathbf{D})$ as H^2. The Hardy space $H^2(\mathbf{D})$ has several equivalent characterizations that are not of interest of this paper. The disc case deals with signals defined in a compact interval on the line. That is also the model for 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 earlier established ([27]). 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) then provides a replacement of AFD in the Hilbert spaces that do not have easy-usable Blaschke product-like functions, and nor explicit and constructive orthogonal function systems like the TM system. The latter happens mostly for multivariate signals. We leave the POAFD method to be studied in the later 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 almost surely (a.s.) for a fixed probabilistic sample point $w\in\Omega$ the function $F(\cdot,w)$ is a deterministic signal of finite energy; meanwhile for almost everywhere a 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 $F(t,w)=f(e^{it},w)=f_w(t)$, and, since it a.s. has finite energy, we have the trigonometric expansion

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

The Plancherel Theorem gives

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

In this study we impose the condition

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

where $E_w|c_k(w)|^2$ stands for the mathematical expectation of the function $|c_k(w)|^2$ of the random variable w 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 measure. The quantity in (1.2) is called the *energy expectation norm* (EE-norm) of f, denoted as $\|f\|_{\mathcal{N}}$. Set,

$$L^2_w(\partial\mathbf{D},\Omega) = \{f : \partial\mathbf{D}\times\Omega \to \mathbf{C} \mid f \text{ is a RS, and } \|f\|_{\mathcal{N}} < \infty\}, \tag{1.3}$$

called *the space of random signals of finite energy.* $L^2_w(\partial\mathbf{D},\Omega)$ is written briefly as $\mathcal{N}$. The RSs in $L^2_w(\partial\mathbf{D},\Omega)$ are called *normal random signals*, or normal RSs. The space $\mathcal{N}$ is a Hilbert space under the inner product, $\langle\cdot,\cdot\rangle_{\mathcal{N}}$, induced from the EE-norm. A normal RS is almost surely a signal of finite energy in t. In below we will preserve the inner product notation $\langle\cdot,\cdot\rangle$ only for the inner product of the time-domain-space $L^2(\partial\mathbf{D})$.
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 $\mathbf{D}$. The boundary limits exist a.e. in the pointwise way, and in the L^2-convergence sense as $r = |z| \to 1$. Since f is assumed to be of real-valued, we have $c_{-k} = c_k$, that implies

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

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

$$f^+(e^{it}, w) = \frac{1}{2}(f(e^{it}, w) + iHf(e^{it}, w) + c_0), \tag{1.4}$$

where

$$\begin{aligned} Hf(e^{it}, w) &\triangleq \sum_{k=-\infty}^{\infty} (-i)\mathrm{sgn}(k)c_k(w)e^{ikt} \\ &= \frac{1}{\pi}\mathrm{v.p.}\int_{-\infty}^{\infty} \cot\left(\frac{s}{2}\right) f(e^{i(t-s)}, w)ds. \end{aligned}$$

By using the Hilbert transformation, analysis of the normal RSs is reduced to analysis of their half series. We define the *stochastic Hardy space* to be the collection of the functions f^+ in the above argument (with the superscript "+" being dropped off), denoted

$$\begin{aligned} H^2_w(\mathbf{D}) \;=\; &\{f : \mathbf{D}\times\Omega \to \mathbf{C} \mid f(z, w) \text{ is a.s. analytic in } z \text{ and} \\ &f(z, w) = \sum_{k=0}^{\infty} c_k(w)z^k \text{ with } \|f\|^2_{\mathcal{N}} = \sum_{k=0}^{\infty} E_w|c_k(w)|^2 < \infty\}. \end{aligned} \tag{1.5}$$

The space $H^2_w(\mathbf{D})$ induces a space, being the totality of the boundary limits of the RSs in $H^2_w(\mathbf{D})$, denoted as $H^2_w(\partial\mathbf{D})$. The latter is a proper closed subspace of the Hilbert space

$\mathcal{N}$ on the boundary $\partial\mathbf{D}$. The mapping that maps functions in $H^2_w(\mathbf{D})$ to their boundary limits in $H^2_w(\partial\mathbf{D})$ is, in fact, an isometry between the two spaces.
The purpose of this study is two-fold. The first fold is to develop two types of stochastic adaptive Fourier decompositions, named as SAFDI and SAFDII, for analyzing random signals. In order to make use of complex analysis methods we employ the Fourier expansion, or Hilbert transform, or the Cauchy integral to obtain the analytic functions f^+ from their boundary data. The second fold is to generalize the results obtained in the first fold to the context of Hilbert space with a dictionary. A general context in which some complex analysis methodology can be adopted is reproducing kernel Hilbert space in which the reproducing kernel plays the role of the Cauchy kernel. In the later part of the paper we establish a counterpart theory in what we call *stochastic Hilbert space*. 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 different types of analytic RSs: One is for noised signals (the ramdomization is from the noise), and the other is for a collection of RSs obeying certain probability distribution. The second type is more general than the first. In §3 we extend the theory to the context of stochastic Hilbert space with a dictionary treating also two types of RSs, and develop, accordingly two types of sparse approximations, named as SPOAFDI and SPOAFDII. The necessity of developing a theory in the general Hilbert space context lays in the demanding of applications, especially in the multivariate random signal cases, in which there do not exist analyticity properties as being used in the classical Hardy space cases. As example, by using the developed sparse representation algorithm one may analyse heart ECG signals from a group of people at one time [35, 36, 9], or numerically or with explicit formulas solve random stochastic partial differential equations [42].
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 (1.3)

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

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

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

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 $\mathcal{N}$-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. We use the terminology "sparse approximation" or "sparse representation" in the sense that a given signal is expanded by a system that is not necessarily a basis but with fast convergence. The expression "fast convergence" has the specified meaning that it is either in the classical sense, or in the best n partial sum approximation sense under n-parameters selections, where n can be any prescribed positive integer. In the classical Hardy space formulation AFD 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 ample applications in signal and image analysis as well as in system identification (see, for instance, [10, 41, 11, 12]). With the stochastic Hardy space case, as defined in §1, we generate two types of AFD-like expansions, called *stochastic AFD1* (SAFD1) and *stochastic AFD2* (SAFD2), of which each has its own merits in applications. Before studying the 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

$$f(e^{it}, w) = \tilde{f}(e^{it}) + r(e^{it}, w), \tag{2.6}$$

where $\tilde{f} = E_w f$ and $r(e^{it}, w) = f(e^{it}, w) - \tilde{f}(e^{it})$. The difference r is also called the *remainder RS*. Like in the deterministic signals case we are to reduce analysis of normal RSs to that of the associated analytic normal RSs. Given by the next two theorems, the Hardy space projections $f^+, [\tilde{f}]^+$ and r^+ preserve many properties possessed by the function $f \in \mathcal{N}$.

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

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

there hold

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

and,

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

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

$$\{E_w f\}^+ = E_w\{f^+\} \quad \text{and} \quad \|r^+\|_{\mathcal{N}} = \frac{\|r + d_0\|_{\mathcal{N}}}{\sqrt{2}}.$$

Proof We note that

$$\begin{aligned}
\left(\sum_{k=-\infty}^{\infty} |E_w(c_k(w))|^2\right)^{1/2} &\le E_w\left[\left(\sum_{k=-\infty}^{\infty} |c_k(w)|^2\right)^{1/2}\right] \quad \text{(Minkovski's inequality)} \\
&\le \left[E_w(\sum_{k=-\infty}^{\infty} |c_k(w)|^2)\right]^{1/2} [E_w(1)]^{1/2} \quad \text{(Hölder's inequality)} \\
&= \left[\sum_{k=-\infty}^{\infty} E_w(|c_k(w)|^2)\right]^{1/2} [E_w(1)]^{1/2} \\
&= \|f\|_{\mathcal{N}} < \infty. \qquad (2.7)
\end{aligned}$$

Then the Riesz-Fisher Theorem asserts that

$$g(e^{it}) = \sum_{k=-\infty}^{\infty} E_w(c_k(w))e^{ikt} \in L^2(\partial\mathbf{D}).$$

Now we show $\tilde{f} = g$. Denote $f_n(e^{it}, w) = \sum_{|k|\le n} c_k(w)e^{ikt}$. Then $E_w f_n(e^{it}, w) = \sum_{|k|\le n} E_w(c_k)e^{ikt}$. Similarly to the reasoning of (2.7), there follows

$$\begin{aligned}
\|E_w f - E_w f_n\| &= \|E_w(f - f_n)\| \\
&\le E_w\|f - f_n\| \\
&\le \left(E_w\|f - f_n\|^2\right)^{1/2} \\
&= \|f - f_n\|_{\mathcal{N}} \\
&= \left(\sum_{|k|>n} E_w(|c_k(w)^2)\right)^{1/2} \\
&\to 0, \quad \text{as } n \to \infty.
\end{aligned}$$

Since the linear functional of the m-th Fourier coefficient, C_m, is continuous, there follows

$$C_m(E_w f) = \lim_{n\to\infty} C_m(E_w f_n) = E_w(c_m(w)).$$

This shows that $E_w f = g \in L^2(\partial\mathbf{D})$ and is with the Fourier expansion

$$\tilde{f} = \sum_{k=-\infty}^{\infty} E_w(c_k(w))e^{ikt} \in L^2(\partial\mathbf{D}).$$

It then follows

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

As a consequence of (2.8), we have the orthogonality

$$E_w(|\tilde{f}(e^{it}) + r(e^{it},w)|^2) = |\tilde{f}(e^{it})|^2 + E_w|r(e^{it},w)|^2, \tag{2.9}$$

and thus the finiteness of the $\mathcal{N}$-norm of r :

$$E_w(|r(e^{it},w)|^2) = \|f\|^2_{\mathcal{N}} - \|\tilde{f}\|^2_{L^2(\partial \mathbf{D})} < \infty \quad \text{for a.e. } t \in [0,2\pi). \tag{2.10}$$

To compute the $\mathcal{N}$-norm of r^+, by taking into account $d_k = \overline{d}_{-k}$, we have

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

The proof of the theorem is complete.

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

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

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

$$\begin{aligned} & \left(\int_0^{2\pi} |E_w f(re^{it},w)|^2 dt\right)^{1/2} \\ \leq \ & E_w\left[\left(\int_0^{2\pi} |f(re^{it},w)|^2 dt\right)^{1/2}\right] \quad \text{(Minkovski's inequality)} \\ \leq \ & \left(E_w \int_0^{2\pi} |f(re^{it},w)|^2 dt\right)^{1/2} E_w(1)^{1/2} \quad \text{(Holder's inequality)} \\ \leq \ & \|f\|_{\mathcal{N}} < \infty. \end{aligned} \tag{2.11}$$

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

$$f(z,w) = \sum_0^{\infty} c_k(w) z^k = \sum_0^{\infty} c_k(w) r^k e^{ikt}, \qquad r < 1.$$

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 Szegö kernel on the circle:

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

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

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

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

$$\tilde{f}(z) = \langle f_1, e_a\rangle e_a(z) + f_2(z)\frac{z-a}{1-\overline{a}z}, \tag{2.12}$$

where f_2 is call the *reduced remainder*, given by

$$f_2(z) = \frac{f_1(z) - \langle f_1, e_a\rangle e_a(z)}{\frac{z-a}{1-\overline{a}z}} \in H^2(\mathbf{D}). \tag{2.13}$$

Due to the orthogonalization we have

$$\|\tilde{f}\|_{H^2(\mathbf{D})} = |\langle f_1, e_a\rangle|^2 + \|f_2\|_{H^2(\mathbf{D})}. \tag{2.14}$$

Thus, the larger the quantity $|\langle f_1, e_a\rangle|^2$ is, the smaller the energy of the reduced remainder f_2 is. Although $\mathbf{D}$ is an open set it can be proved (see [27], for instance) that

$$\sup\{|\langle f_1, e_a\rangle|^2 \mid a \in \mathbf{D}\}$$

is attainable at a point of $\mathbf{D}$. Hence, one theoretically selects

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

Such maximal selection is phrased as *Maximal Selection Principle* (MSP) of the Hardy space ([27]). The MSP is evidenced by the boundary vanishing condition (BVC) of the Szegö kernel dictionary in the Hardy space (see §3 for a more general formulation). Using this a_1 in place of a in (2.12), (2.13) and (2.14), we have that the corresponding reduced remainder f_2 has its least possible norm. In our terminology this is n-best approximation with $n = 1$.
To f_2 we carry on the same decomposition procedure, and so on, after n-iterations, we have

$$\tilde{f}(z) = \sum_{k=1}^{n}\langle f_k, e_{a_k}\rangle B_k(z) + f_{n+1}(z)\prod_{k=1}^{n}\frac{z-a_k}{1-\overline{a}_k z}, \tag{2.15}$$

where $\{B_k\}_{k=1}^{\infty}$ is the automatically generated orthonormal Takenaka-Malmquist (TM) system corresponding to the maximally selected $a_1, \cdots, a_k, \cdots$, all in $\mathbf{D}$, where, explicitly,

$$B_k(z) = e_{a_k}(z)\prod_{l=1}^{k-1}\frac{z-a_l}{1-\overline{a}_l z}, \tag{2.16}$$

$$a_k = \max\{|\langle f_k, e_a\rangle|^2 \mid a \in \mathbf{D}\}, \tag{2.17}$$

$$f_{k+1}(z) = \frac{f_k(z) - \langle f_k, e_{a_k}\rangle e_{a_k}(z)}{\frac{z-a_k}{1-\overline{a}_k z}} \in H^2(\mathbf{D}). \tag{2.18}$$

We note that $\{B_k\}$ is automatically an orthonormal system, although not necessarily a basis. It turns out that under the maximal selections of $a_k, k = 1, 2, \cdots$, there holds the convergence:

$$\tilde{f}(z) = \sum_{k=1}^{\infty}\langle f_k, e_{a_k}\rangle B_k(z). \tag{2.19}$$

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 ([27]).

Remark 2.2. Any sequence $(a_1, \cdots, a_n, \cdots)$ in $\mathbf{D}$ can define a TM system $\{B_k\}_{k=1}^{\infty}$ by (2.16). 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 ([37]). 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, $\{z^{k-1}\}_{k=1}^{\infty}$, corresponding to all a_n being identical with zero, is a particular example of the basis case. The study [27] opens a new era of adaptive use of TM systems through maximal selections of the parameters according to the data of the given signal. The MSP of AFD declares the best selection principle at the one-step selection. This is due to attainability of the global maximum at each step, that rests in the availability of repeating selections 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 and feasibility of repeating selections of the parameters. AFD found close connections to the Beurling Theorem for $H^2(\mathbf{D})$ asserting directional-sum decomposition of the space into shift- and backward shift-invariant subspaces:

$$H^2(\mathbf{D}) = \overline{\text{span}}\{B_k\}_{k=1}^{\infty} \oplus bH^2(\mathbf{D}), \tag{2.20}$$

where $\{B_k\}_{k=1}^{\infty}$ is the TM system and b is the Blaschke product, when can be defined, with the sole zeros $a_1, a_2, \cdots$, including the multiplicities. The Blaschke product is well defined if and only if $\sum_{k=1}^{\infty}(1-|a_k|) < \infty$. If the sequence cannot define a Blaschke product, then

$$H^2(\mathbf{D}) = \overline{\text{span}}\{B_k\}_{k=1}^{\infty}. \tag{2.21}$$

With the AFD formulation we know that $\tilde{f} \in \overline{\text{span}}\{B_k\}_{k=1}^{\infty}$, the backward shift-invariant subspace in (2.20) or (2.21).

Remark 2.3. AFD was motivated by intrinsic positive frequency 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, in the sparse approximation aspect, to multi-dimensions with the Clifford and several complex variables setting with scalar- to matrix-valued signals ([18, 28, 38, 39, 31, 1, 2]). In the one dimensional case a variation called unwinding Blaschke expansion was first studied by Coifman and Nahon in 2000, and then joined by Steinerberger and Peyriére making further connections with Blaschke products and outer functions ([6, 7]). Unwinding method was also separately developed in [19], and further developed in a recent paper on maximal unwinding AFD [29]. AFD has also been generalized to Hilbert spaces with a dictionary satisfying BVC ([21, 23]). The AFD generalization in Hilbert spaces is called pre-orthogonal adaptive Fourier decomposition (POAFD). Amongst, POAFD algorithms in weighted Bergman spaces and weighted Hardy spaces (Non-Hardy type Hilbert spaces of holomorphic functions) were developed in ([25, 26]). AFD and its one-dimensional variations, as well as its generalizations to the non-Hardy type and general Hilbert spaces, have become powerful tools in signal and image analysis, and in system identification ([11, 12, 5, 10, 41]).

Remark 2.4. In the AFD algorithm, as a consequence of the orthogonality, there hold the relations:

$$\langle f_k, e_{a_k}\rangle = \langle g_k, B_k\rangle = \langle \tilde{f}, B_k\rangle, \qquad k \geq 2, \tag{2.22}$$

where

$$g_k(z) = \tilde{f}(z,w) - \sum_{l=1}^{k-1} \langle f_l, e_{a_l}\rangle B_l(z), \quad k \geq 2, \tag{2.23}$$

is the *k-th standard remainder*. It is the relation (2.22) 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 in the underlying Hilbert space as Gram-Schmidt orthogonalization of the Szegö kernels as in the unit disc or the half complex plane case.

We now continue our sparse representation theme for random signals. For an analytic random signal f in $H^2_w(\mathbf{D})$, we obtain a sequence of parameters $a_1, a_2, \cdots$, and an associated TM system $\{B_k\}_{k=1}^\infty$ that gives rise to an AFD sparse representation of the deterministic $\tilde{f} = E_w f_w$. The question is that when we use the system $\{B_k\}_{k=1}^\infty$ to expand the original random signal $f(e^{it}, w) = f_w(e^{it})$ for fixed w, then in what extent the related TM series expansion can represent f as a RS? Or namely, what is the difference

$$d_f(e^{it}, w) = f_w(e^{it}) - \sum_{k=1}^\infty \langle f_w, B_k\rangle B_k(e^{it})? \tag{2.24}$$

In view of the Beurling Theorem, when there holds $\sum_{k=1}^\infty (1-|a_k|) < \infty$, it would well happen that for some w, $f_w \in H^2(\mathbf{D}) \ominus \overline{\mathrm{span}}\{B_k\}_{k=1}^\infty = bH^2(\mathbf{D})$, and in the case the difference $d_f(e^{it}, w)$ is a non-zero function. We have the following

Theorem 2.5. *Let $f \in H^2_w(\mathbf{D}), \tilde{f} = E_w f$, 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 $\{B_k\}$,

$$E_w d_f(e^{it}, w) = 0, \quad \forall t \in [0, 2\pi). \tag{2.25}$$

There holds the relation

$$E_w \| f_w - \sum_{k=1}^n \langle f_w, B_k\rangle B_k\|^2_{H^2_w} = \|d_f\|^2_{\mathcal{N}} + \sum_{k=n+1}^\infty E_w |\langle f_w, B_k\rangle|^2, \tag{2.26}$$

with

$$\lim_{n\to\infty} \sum_{k=n+1}^\infty E_w |\langle f_w, B_k\rangle|^2 = 0. \tag{2.27}$$

And, in terms of the error $r = f - \tilde{f}$ the difference, d_f is estimated

$$\|d_f\|^2_{\mathcal{N}} = \|r\|^2_{\mathcal{N}} - \sum_{k=1}^\infty E_w |\langle r_w, B_k\rangle|^2 = \|P_{bH^2(\mathbf{D})} r\|^2_{\mathcal{N}} = \|P_{bH^2(\mathbf{D})} f\|^2_{\mathcal{N}}, \tag{2.28}$$

where P_X is, in general, denoted as the projection operator into the closed subspace X.

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

$$\sum_{k=1}^{\infty} E_w|\langle f_w, B_k\rangle|^2 \le \|f\|_{\mathcal{N}}^2,$$

that implies the desired relation (2.27). 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 $H_w^2(\mathbf{D})$. Hence the difference $d_f(w,\cdot)$ belongs to $H_w^2(\mathbf{D})$. All these functions are in $\mathcal{N}$.
Since the underlying product measure space of $\mathcal{N}$ 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

$$\begin{aligned} E_w(f_w - \sum_{k=1}^{\infty}\langle f_w, B_k\rangle B_k) &= \tilde{f} - E_w(\sum_{k=1}^{\infty}\langle f_w, B_k\rangle B_k) \\ &= \tilde{f} - \sum_{k=1}^{\infty} E_w\langle f_w, B_k\rangle B_k \\ &= \tilde{f} - \sum_{k=1}^{\infty}\langle \tilde{f}, B_k\rangle B_k \\ &= 0, \end{aligned}$$

as desired by (2.25).
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 (2.26).
Since

$$d_f = (f_w - \tilde{f}) - \sum_{k=1}^{\infty}\langle f_w - \tilde{f}, B_k\rangle B_k = r_w - \sum_{k=1}^{\infty}\langle r_w, B_k\rangle B_k,$$

there follows

$$\begin{aligned} \|d_f\|_{\mathcal{N}}^2 &= E_w \int_0^{2\pi} |r_w(e^{it}) - \sum_{k=1}^{\infty}\langle r_w, B_k\rangle B_k(e^{it})|^2 dt \\ &= E_w\left(\|r_w\|_{L^2}^2 - \sum_{k=1}^{\infty}|\langle r_w, B_k\rangle|^2\right) \\ &= \|r\|_{\mathcal{N}}^2 - \sum_{k=1}^{\infty} E_w|\langle r_w, B_k\rangle|^2. \end{aligned}$$

The proof of the theorem is complete.

Remark 2.6. The sparse random approximation established in Theorem 2.5 would mostly concern a deterministic signal corrupted with non-significant errors. The effectiveness of the averaging TM system represented by the $\mathcal{N}$-norm of the *general difference* $d_f(e^{it}, w)$ defined through 2.24 is identical with the $\mathcal{N}$-energy of f on $bH^2(\mathbf{D})$, dominated by the $\mathcal{N}$-variation of the error term r. In the following section we develop a sparse representation for analytic random signal that enjoys $d_f = 0$ almost surely in Ω.

2.3. **The SAFDII: Taking Mean Secondly.**

Theorem 2.7. *Let $f \in H^2_w(\mathbf{D})$. Then there exists $a_1 \in \mathbf{D}$ such that*

$$a_1 = \arg\max\{E_w|\langle f_w, e_a\rangle|^2 \mid a \in \mathbf{D}\}.$$

Proof Since $E_w|\langle f_w, e_a\rangle|^2$ is a continuous function in $a \in \mathbf{D}$, it suffices to show that the quantity satisfies a *stochastic boundary vanishing condition* (SBVC), that is

$$\lim_{|a|\to 1} E_w|\langle f_w, e_a\rangle|^2 = 0. \tag{2.29}$$

Then a continuity argument based on (2.29) concludes the theorem.
The condition $f \in \mathcal{N}$ implies

$$E_w \sum_{k=0}^{\infty} |c_k(w)|^2 < \infty \tag{2.30}$$

As a consequence of the integrability, almost surely

$$\sum_{k=0}^{\infty} |c_k(w)|^2 < \infty.$$

This implies, almost surely $f_w(z) = \sum_{k=0}^{\infty} c_k(w) z^k \in H^2(\mathbf{D})$. Thanks to the BVC of the classical Hardy space ([27]), we have almost surely

$$\lim_{|a|\to 1} |\langle f_w, e_a\rangle|^2 = 0. \tag{2.31}$$

On the other hand, when $|a| \to 1$, the function $|\langle f_w, e_a\rangle|^2$ uniformly in $a \in \mathbf{D}$ has a uniform positive dominating function. In fact, using the Cauchy-Schwarz inequality,

$$|\langle f_w, e_a\rangle|^2 \leq \|f_w\|^2 = \sum_{k=0}^{\infty} |c_k(w)|^2.$$

Given by (2.30), the dominating function is E_w-integrable. The Lebesgue domination convergence theorem then can be used to conclude the desired SBVC (2.29). The proof is complete.
The SAFDII proceeds as follows: Guaranteed by the Theorem 2.7, in the same iterative steps as for the classical AFD, one can select, at the k- step, an optimal a_k :

$$a_k = \arg\max\{E_w|\langle (f_k)_w, e_a\rangle|^2 \mid a \in \mathbf{D}\}, \tag{2.32}$$

where $f = f_1$, and

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

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 (2.16), corresponding to the selections $a_1, a_2, \cdots$, 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

$$\langle (f_k)_w, e_{a_k}\rangle = \langle (g_k)_w, B_k\rangle = \langle f_w, B_k\rangle, \tag{2.33}$$

where

$$(g_k)_w(z) = g_k(z,w) = f(z,w) - \sum_{l=1}^{k-1}\langle f_w, B_l\rangle B_l(z), \quad k \geq 2, \tag{2.34}$$

is the k-th standard remainder. The relations (2.33) imply

$$E_w|\langle (f_k)_w, e_{a_k}\rangle|^2 = E_w|\langle (g_k)_w, B_k\rangle|^2 = E_w|\langle f_w, B_k\rangle|^2. \tag{2.35}$$

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

$$\lim_{k\to\infty} E_w|\langle f_w, B_k\rangle|^2 = 0. \tag{2.36}$$

In view of (2.35), the SMSP (2.32) is reduced to the form

$$a_k = \arg\max\{E_w|\langle f_w, B_k^a\rangle|^2 \mid a \in \mathbf{D}\}, \tag{2.37}$$

where

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

We now prove

Theorem 2.8. *Let $f(w, e^{it}) \in H^2_w(\mathbf{D})$ and $(a_1, \cdots, a_n, \cdots)$ be a sequence selected according to the SMSP given in (2.32). Then there holds, in the $\mathcal{N}$-norm sense,*

$$f(z,w) = \sum_{k=1}^\infty \langle f_w, B_k\rangle B_k(z). \tag{2.38}$$

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

$$f(z,w) = \sum_{k=1}^\infty \langle f_w, B_k\rangle B_k(z) + g(z,w), \quad \|g\|_{\mathcal{N}} > 0. \tag{2.39}$$

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

$$\|g\|_N^2 = \|f\|_{\mathcal{N}}^2 - \sum_{k=1}^\infty E_w|\langle f_w, B_k\rangle|^2. \tag{2.40}$$

In particular,

$$\lim_{k\to\infty} E_w|\langle f_w, B_k\rangle|^2 = 0. \tag{2.41}$$

We show that there exists $b \in \mathbf{D}$ such that

$$E_w|\langle g_w, e_b\rangle|^2 = \delta^2 > 0$$

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

$$\langle g_w, e_b \rangle = 0.$$

Due to the density of e_b in $H^2(\mathbf{D})$ we would have, for a.s. $w \in \Omega$, $g_w = 0$ 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 $a_k, k = 1, 2, \cdots$. In below we will fix this $b \in \mathbf{D}$ 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 (2.34), 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,

$$(1 - |b|^2)^2 E_w|(g_k)(b)|^2 \geq \delta^2/4. \tag{2.42}$$

Since pointwise there holds

$$f_k = g_k/b_k \quad \text{and} \quad |b_k(b)| < 1, \tag{2.43}$$

, where b_k is the Blaschke product generated by $a_1, \cdots, a_k$ as sole zeros, there follows $|f_k| \geq |g_k|$. Hence,

$$(1 - |b|^2)^2 E_w|(f_k)(b)|^2 \geq \delta^2/4.$$

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

$$E_w|\langle f_k, e_b \rangle|^2 = E_w|\langle f_w, B_k \rangle|^2 \geq \delta^2/4$$

for large enough k. This is contradictory to (2.41). The proof is thus complete.

Remark 2.9. The proof is an adaptation of one used in [27] to the stochastic case, in which, as in the classical case, the relation (2.43) is crucial.

3. Stochastic SPOAFDs in Hilbert Spaces

Our discussions on stochastic Hilbert spaces will be based on the related 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 ([23], also see [21, 22, 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 collection of certain elements of unit norm whose span is dense in the Hilbert space. The unit norm requirement for a dictionary is not essential. We usually assume that in the underlying Hilbert space $\mathcal{H}$ there a subclass of elements $K_q, q \in \mathbf{E}$, whose linear span is dense. We call such a set a *pre-dictionary*. The parameter set $\mathbf{E}$ is an open set of the complex plane, or more generally an open set of $\mathbf{R}^d$ or $\mathbf{C}^{d'}$, or a product of two such open sets, etc. We denote the normalizations of K_q by E_q, that is, $E_q = K_q/\|K_q\|, q \in \mathbf{E}$. Below we often call the K_q's in a pre-dictionary by kernels. We borrow this terminology from reproducing kernel Hilbert space. Indeed, the parameterized reproducing kernels constitute a subclass that induces through normalization a dictionary of the space.

Now, we introduce what is called and assumed *Boundary Vanishing Condition* (*BVC*) in our Hilbert space context: For any but fixed $G \in \mathcal{H}$, if $p_n \in \mathbf{E}$ and $p_n \to \partial\mathbf{E}$ (including ∞ if $\mathbf{E}$ is unbounded while in the case we use the compactification topology for the added infinity point), then

$$\lim_{n\to\infty} |\langle G, E_{p_n}\rangle| = 0.$$

We next define what we call by "multiple kernels". Let $(q_1, \cdots, q_n)$ be any n-tuple of parameters in $\mathbf{E}$. We denote by $l(k)$ the multiplicity of q_k in the k-tuple $(q_1, \cdots, q_k)$. *Multiple kernels* are defined as follows. For any $k \le n$, denote

$$\tilde{K}_k = \left[\left(\frac{\partial}{\partial \overline{q}}\right)^{(l(k)-1)} K_q\right]_{q=q_k}.$$

With a little abuse of notation, we will also denote $\tilde{K}_k$ by $\tilde{K}_{q_k}, k = 1, 2, \cdots, n$, 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 $\{q_1, \cdots, q_{n-1}\}$, allowing multiplicities, corresponding to the $(n-1)$-tuple of kernels, $\{\tilde{K}_{q_1}, \cdots, \tilde{K}_{q_{n-1}}\}$. By performing the G-S orthonormalization process consecutively we obtain an equivalent $(n-1)$-orthonormal system, $\{B_1, \cdots, B_{n-1}\}$. For any given G in the Hilbert space we wish to investigate whether there exists a q_n that gives rise to the supreme value

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

where the finiteness of the supreme is guaranteed by the Cauchy-Schwarz inequality, and B_n^q be such that $\{B_1, \cdots, B_{n-1}, B_n^q\}$ is the G-S orthonormalization of $\{\tilde{K}_{q_1}, \cdots, \tilde{K}_{q_{n-1}}, K_q\}$. Since q is distinct from the proceeding $q_1, \cdots, q_{n-1}$, B_n^q is given by

$$B_n^q = \frac{K_q - \sum_{k=1}^{n-1}\langle K_q, B_k\rangle_{\mathcal{H}} B_k}{\|K_q - \sum_{k=1}^{n-1}\langle K_q, B_k\rangle_{\mathcal{H}} B_k\|}. \tag{3.44}$$

Under BVC a compact argument leads that there exists a point $q_n \in \mathbf{E}$ 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

$$\lim_{l\to\infty} |\langle G, B_n^{q^{(l)}}\rangle| = \sup\{|\langle G, B_n^q\rangle| \ : \ q \in \mathbf{E}, q \neq q_1, \cdots, q_{n-1}\} = |\langle G, B_n^{q_n}\rangle|, \tag{3.45}$$

where

$$B_n^{q_n} = \frac{\tilde{K}_{q_n} - \sum_{k=1}^{n-1}\langle \tilde{K}_{q_n}, B_k\rangle_{\mathcal{H}} B_k}{\sqrt{\|\tilde{K}_{q_n}\|^2 - \sum_{k=1}^{n-1}|\langle \tilde{K}_{q_n}, B_k\rangle_{\mathcal{H}}|^2}}, \tag{3.46}$$

proved through an argument involving Taylor series expansion (see [22, 5]). The BVC together with multiple kernels are theoretical guarantee 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 $\{q_k\}_{k=1}^{\infty}$ one eventually obtains, with a fast convergent pace,

$$F = \sum_{k=1}^{\infty}\langle F, B_k\rangle_{\mathcal{H}} B_k \tag{3.47}$$

([21, 22, 5]).

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 $q_n \in \mathbf{E}$ can be found, different from the previously selected $q_k, k = 1, \cdots, n-1$, to have

$$|\langle G_n, B_n^{q_n}\rangle| \geq \rho \sup\{\langle G_n, B_n^{q}\rangle \ : \ q \in \mathbf{E}, q \neq q_1, \cdots, q_{n-1}\}. \tag{3.48}$$

The corresponding algorithm is called *Weak Pre-orthogonal Adaptive Fourier Decomposition* (WPOAFD). With WPOAFD one may have all the selected $q_1, \cdots$ distinguished, and thus $l(k) \equiv 1$ and $K_k = \tilde{K}_k$ for all k. WPOAFD is still a fast converging algorithm although at each step it does not reach the optimal.

Remark 3.2. An order $O(1/\sqrt{n})$ of the convergence rate can be proved: For $M > 0$, by defining

$$\mathcal{M}_M = \{F \in \mathcal{H} : \exists \{c_n\},\ \{E_{q_n}\} \text{ s. t. } F = \sum_{n=1}^{\infty} c_n E_{q_n} \text{ with } \sum_{n=1}^{\infty}|c_n| \leq M\}, \tag{3.49}$$

for any $F \in \mathcal{M}_M$, the POAFD partial sums satisfy

$$\|F - \sum_{k=1}^{n}\langle F, B_k\rangle_{\mathcal{H}} B_k\|_{\mathcal{H}} \leq \frac{M}{\sqrt{n}}.$$

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 (2.22), the MSP (2.17) for AFD reduces to the MSP (3.45) for POAFD, and AFD reduces to POAFD. The algorithm codes of AFD and POAFD, as well as those of the related ones are available at request in 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 than the n maximal consecutive kernel expansion as given in (3.47). The n-best approximation is motivated by the classical problem, yet still open in its ultimate algorithm, called the best approximation to Hardy space functions by rational functions of degree not exceeding n ([3, 4, 30]). The gradient descending method for cyclic AFD ([30]) and cyclic AFD separately ([18]) 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 problems in in various types of concrete Hilbert spaces, including Hardy and non-Hardy types, and those with hypercomplex variables and matrix-valued functions. For related publications see [19, 6, 1, 2, 29].

3.2. **Stochastic POAFDs.** Let $\mathcal{H}$ be a Hilbert space with a pre-dictionary $\{K_q\}$ parameterized in an open set $\mathbf{E} : q \in \mathbf{E}$. We assume that the pre-dictionary satisfies BVC

$$\lim_{q\to\partial\mathbf{E}} |\langle F, E_q\rangle| = 0, \tag{3.50}$$

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. $w \in \Omega$, $F(\cdot,w) = F_w \in \mathcal{H}$; and for any $t \in T$, $F(t,\cdot)$ is a random variable. Define

$$\begin{aligned} \mathcal{N}(\mathcal{H},\Omega) \;\; &= \;\; \{F(t,w) : F(\cdot,w) \in \mathcal{H}, \text{for a.s. w; and } F(t,\cdot) \text{ being a random} \\ &\qquad \text{variable for each fixed } t;\ \text{ and } E_w\|F(\cdot,w)\|^2_{\mathcal{H}} < \infty.\} \end{aligned} \tag{3.51}$$

This formulation supports two types of stochastic POAFDs, abbreviated as SPOAFDI and SPOAFDII.

SPOAFDI is one to treat a noised deterministic signal. It corresponds to first take the mean and then do maximal energy extractions. We need to show $E_wF_w \in \mathcal{H}$. Following what is done in (2.11), by using the Minkovski inequality followed by the Cauchy-Schwarz inequality, we get

$$\|E_wF_w\|_{\mathcal{H}} \le E_w\|F_w\|_{\mathcal{H}} \le \left(E_w\|F_w\|^2_{\mathcal{H}}\right)^{1/2} = \|F\|_{\mathcal{N}(\mathcal{H},\Omega)} < \infty.$$

This shows that the mean belongs to the underlying Hilbert space $\mathcal{H}$. Since $\mathcal{H}$ has a pre-dictionary that satisfies BVC, one can perform POAFD in $\mathcal{H}$. The difference $d(t,w) = F(t,w) - E_wF(t,w)$ enjoys the zero-mean 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 mean and of a small $\mathcal{N}(\mathcal{H},\Omega)$ norm.

To perform the SPOAFDII we first need to prove the stochastic boundary vanishing condition, or SBVC,

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

The proof follows the same route as for the SAFDII. To show SBVC we again use the Lebesgue Dominated Convergence Theorem for the probability space, through showing

1. For a.s. $w \in \Omega$

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

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 $\{E_q\}_{q\in\mathbf{E}}$ in $\mathcal{H}$. To show 2, we have, by the Cauchy-Schwarz inequality,

$$E_w|\langle F_w, E_q\rangle|^2 \le E_w\|F_w\|^2 = \|F\|^2_{\mathcal{N}(\mathcal{H},\Omega)} < \infty.$$

This concludes that $\|F_w\|^2$ is a desired dominating function for $|\langle F_w, E_q\rangle|^2$ in the probability space. The SBVC is hence proved.
Based on the 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*

$$q_k = \arg\sup\{E_w|\langle (G_k)_w, B_k^q\rangle|^2 \mid q\in\mathbf{E}\},$$

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,

$$F(z,w) = \sum_{k=1}^{\infty}\langle F_w, B_k\rangle B_k(z). \tag{3.52}$$

Remark 3.5. The proof of Theorem 2.8 crucially depends on the property $|b_k(z)|\le 1$ of the classical Blaschke products. In the general Hilbert spaces case there may not exist Blaschke product-like functions. 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 [20] or [22], or [5], where [5] is English equivalent to [22]).

Proof of Theorem 3.4 We will prove the theorem by contradiction. If the RHS series of (3.52) does not converges to the LHS function, then there is a non-trivial random signal $H\in\mathcal{N}(\mathcal{H},\Omega)$ such that

$$F(t,w) = \sum_{k=1}^{\infty}\langle F_w, B_k\rangle B_k(z) + H(z,w), \quad \|H\|_{\mathcal{N}(\mathcal{H},\Omega)} > 0. \tag{3.53}$$

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

$$0 < \|H\|^2_{\mathcal{N}(\mathcal{H},\Omega)} = \|F\|^2_{\mathcal{N}(\mathcal{H},\Omega)} - \sum_{k=1}^{\infty}E_w|\langle F_w, B_k\rangle|^2. \tag{3.54}$$

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

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

For, if this were not true, then almost surely for all $q\in\mathbf{E}$

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

Due to the density of K_q in $\mathcal{H}$, we would have almost surely $H_w = 0$ as a function of t, being contradictory to $\|H\|_{\mathcal{N}(\mathcal{H},\Omega)} > 0$. We may, in particular, choose q being distinguished from all the selected $q_k, k=1,2,\cdots$. In below such $q\in \mathbf{E}$ 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 (3.53) as

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

where

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

The Bessel inequality implies

$$\lim_{M\to\infty}\|\tilde{G}_{M+1}\|_{\mathcal{N}(\mathcal{H},\Omega)} = 0. \tag{3.55}$$

On one hand, we have, from the orthogonality and (2.36), for large M,

$$E_w|\langle (G_{M+1})_w, B_{M+1}\rangle|^2 = E_w|\langle F_w, B_{M+1}\rangle|^2 = E_w|\langle F_w, B^{q_{M+1}}_{M+1}\rangle|^2 < \delta^2/16. \tag{3.56}$$

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

$$E_w|\langle (G_{M+1})_w, B^q_{M+1}\rangle|^2 > 9\delta^2/16, \tag{3.57}$$

where B^q_{M+1} 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 $\mathcal{N}(\mathcal{H},\Omega)$-norm,

$$\left(E_w|\langle (G_{M+1})_w, B^q_{M+1}\rangle|^2\right)^{1/2} \geq \left(E_w|\langle H_w, B^q_{M+1}\rangle|^2\right)^{1/2} - \left(E_w|\langle (\tilde{G}_{M+1})_w, B^q_{M+1}\rangle|^2\right)^{1/2}.$$

Using the Gauchy-Schwarz inequality and then (3.55), for large enough M we have

$$E_w|\langle (\tilde{G}_{M+1})_w, B^q_{M+1}\rangle|^2 \leq \|\tilde{G}_{M+1}\|^2_{\mathcal{N}(\mathcal{H},\Omega)} \leq \delta^2/16.$$

Therefore,

$$\left(E_w|\langle (G_{M+1})_w, B^q_{M+1}\rangle|^2\right)^{1/2} \geq \left(E_w|\langle H_w, B^q_{M+1}\rangle|^2\right)^{1/2} - \delta/4. \tag{3.58}$$

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 just $E_w|\langle H_w, B^q_{M+1}\rangle|^2$, for 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

$$E_w|\langle H_w, B^q_{M+1}\rangle|^2 \geq E_w|\langle H_w, E_q\rangle|^2 = \delta^2.$$

Combining with (3.58), we have

$$\left(E_w|\langle (G_{M+1})_w, B^q_{M+1}\rangle|^2\right)^{1/2} \geq 3\delta/4.$$

Thus we proved (3.57) that is contradictory with (3.56). 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 2.8 and Theorem 3.4 have separate proofs. Theorem 2.8 is, as a matter of fact, a special case of Theorem 3.4. The question is whether validity of the former can be reduced to the latter. The answer is "Yes". In 2.8 we do not use the G-S orthogonalization, but the backward shift process to obtain the orthogonality. Whether the two methodologies result in the same orthonormal system? In Appendix we prove that the TM system, obtained in AFD through the backward shift process to the Szegö kernel, coincides with that from the G-S orthogonalization to the same kernels. This validates the above "Yes" answer. Precisely, we will prove

Theorem 3.7. *Let $\{a_1, \cdots, a_n\}$ be any n-tuple of parameters in $\mathbf{D}$ in which multiplicities are allowed. Denote by $l(m)$ the multiplicity of a_m in the m-tuple $\{a_1, \cdots, a_m\}, 1 \le m \le n$. For each m, denote by*

$$\tilde{k}_{a_m}(z) = \frac{\partial^{l(m)-1}}{(\partial \overline{a})^{l(m)-1}} k_a(z)|_{a=a_m}, \text{ where } k_a(z) = \frac{1}{1-\overline{a}z}.$$

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 $\{B_1, \cdots, B_m\}$ (2.16) defined through the ordered m-tuple $\{a_1, \cdots, a_m\}$.

There exist different proofs for this result. In Appendix we give a constructive proof. As far as the author is aware, the unit disc and a half of the complex plane are the only cases to which the equivalence of the two processes 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 $a_1, \cdots, a_m$ by

$$b_{a_1,\cdots,a_m}(z) = \prod_{l=1}^{m} \frac{z-a_l}{1-\overline{a}_l z}.$$

We first show that for any $a \in \mathbf{D}$ being different from $a_1, \cdots, a_{m-1}$ there holds

$$k_a(z) - \sum_{l=1}^{m-1} \langle k_a, B_l \rangle B_l(z) = \overline{b}_{a_1,\cdots,a_{m-1}}(a) b_{a_1,\cdots,a_{m-1}}(z) k_a(z). \tag{5.59}$$

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

$$\begin{aligned} k_a - \langle k_a, B_1\rangle B_1(z) &= \frac{1}{1-\overline{a}z} - \overline{B}_1(a)B_1(z) \\ &= \frac{1}{1-\overline{a}z} - \frac{\alpha}{1-\overline{a}_1 z}, \qquad \text{with} \quad \alpha = \frac{1-|a_1|^2}{1-a_1\overline{a}}, \\ &= \frac{\overline{a}-a_1}{1-a_1\overline{a}}\frac{z-a_1}{1-\overline{a}_1 z}\frac{1}{1-\overline{a}z} \\ &= \overline{b}_{a_1}(a)b_{a_1}(z)k_a(z). \end{aligned}$$

Assume that (5.59) holds for a general $m-1$. Under the inductive hypothesis, we have

$$\begin{aligned} k_a(z) - \sum_{l=1}^{m-1}\langle k_a, B_l\rangle B_l(z) &= [k_a(z) - \sum_{l=1}^{m-2}\langle k_a, B_l\rangle B_l(z)] - \langle k_a, B_{m-1}\rangle B_{m-1}(z) \\ &= \overline{b}_{a_1,\cdots,a_{m-2}}(a)b_{a_1,\cdots,a_{m-2}}(z)k_a(z) - \langle k_a, B_{m-1}\rangle B_{m-1}(z) \\ &= \overline{b}_{a_1,\cdots,a_{m-2}}(a)b_{a_1,\cdots,a_{m-2}}(z)k_a(z) - \overline{B}_{m-1}(a)B_{m-1}(z) \\ &= \overline{b}_{a_1,\cdots,a_{m-2}}(a)b_{a_1,\cdots,a_{m-2}}(z)\left[k_a(z) - \frac{1-|a_{m-1}|^2}{(1-a_{m-1}\overline{a})(1-\overline{a}_{m-1}z)}\right] \\ &= \overline{b}_{a_1,\cdots,a_{m-1}}(a)b_{a_1,\cdots,a_{m-1}}(z)k_a(z). \end{aligned}$$

We hence proved (5.59). Next we deal with multiplicity of parameters. Now we are with the new inductive hypothesis that the Gram-Schmidt orthonormalization of $\{\tilde{k}_{a_1}, \cdots, \tilde{k}_{a_{m-1}}\}$ is the $(m-1)$-TM system $\{B_1, \cdots, B_{m-1}\}$. First assume that a_m is different from all the preceding $a_k, k = 1, \cdots, m-1$. In (5.59) let $a = a_m$. By taking the norm on the both sides of (5.59) and invoking the orthonormality of the TM system we have

$$\|k_{a_m}(z) - \sum_{l=1}^{m-1}\langle k_{a_m}, B_l\rangle B_l(z)\| = e^{-ic_m}\overline{b}_{a_1,\cdots,a_{m-1}}(a_m)\frac{1}{\sqrt{1-|a_m|^2}},$$

where c_m is a real number depending on $a_1, \cdots, a_{m-1}, a_m$, and precisely,

$$e^{ic_m} = \frac{|b_{a_1,\cdots,a_{m-1}(a_m)}|}{b_{a_1,\cdots,a_{m-1}}(a_m)}.$$

We thus conclude that

$$\frac{k_{a_m}(z) - \sum_{l=1}^{m-1}\langle k_{a_m}, B_l\rangle B_l(z)}{\|k_{a_m}(z) - \sum_{l=1}^{m-1}\langle k_{a_m}, B_l\rangle B_l(z)\|} = e^{ic_m}b_{a_1,\cdots,a_{m-1}}(z)e_{a_m}(z). \tag{5.60}$$

Note that here we are with the case $l(m) = 1$ and $k_{a_m} = \tilde{k}_{a_m}$. Next we consider the case $l(m) > 1$, and we are to show

$$\frac{\tilde{k}_{a_m}(z) - \sum_{l=1}^{m-1}\langle \tilde{k}_{a_m}, B_l\rangle B_l(z)}{\|\tilde{k}_{a_m}(z) - \sum_{l=1}^{m-1}\langle \tilde{k}_{a_m}, B_l\rangle B_l(z)\|} = e^{ic_m}\phi_{a_1,\cdots,a_{m-1}}(z)e_{a_m}(z). \tag{5.61}$$

For b being sufficiently close to a_m in $\mathbf{D}$ we have up to the $(l(m)-1)$-order power series expansion in the variable b :

$$\begin{aligned} k_b(z) &= \sum_{l=0}^{l(m)-1} \frac{1}{l!}\left[\left(\frac{\partial}{\partial \overline{a}}\right)^l k_a(z)\right]_{a=a_m} (b-a_m)^l + o((b-a_m)^{(l(m)-1)}) \\ &= T(z) + \frac{1}{(l(m)-1)!}\tilde{k}_{a_m}(z)(b-a_m)^{l(m)-1} + o((b-a_m)^{(l(m)-1)}), \end{aligned}$$

where

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

Now, according to the inductive hypothesis, $B_1, \cdots, B_{m-1}$ involve the derivatives of the reproducing kernel up to the $(l(m)-2)$-order, and hence T is in the linear span of $B_1, \cdots, B_{m-1}$. As a consequence,

$$T(z) - \sum_{k=1}^{m-1} \langle T, B_k \rangle B_k = 0. \tag{5.62}$$

Inserting the left-hand-side of (5.62) into (5.60), where a_m is replaced by b with $b \to a_m$ horizontally (meaning that $\mathrm{Im}(b) = \mathrm{Im}(a_m)$), and from the right hand side, while dividing both the denominator and numerator part of the left hand side quotient by $(b-a_m)^{l(m)-1} > 0$, we have

$$\frac{\frac{k_b(z)-T(z)}{(b-a_m)^{l(m)-1}} - \sum_{l=1}^{m-1}\langle \frac{k_b-T}{(b-a_m)^{l(m)-1}}, B_l\rangle B_l(z)}{\|\frac{k_b-T}{(b-a_m)^{l(m)-1}} - \sum_{l=1}^{m-1}\langle \frac{k_b-T}{(b-a_m)^{l(m)-1}}, B_l\rangle B_l(z)\|} = e^{ic_m}\phi_{a_1,\cdots,a_{m-1}}(z)k_b(z).$$

Letting $b - a_m \downarrow 0$ and noticing that the Taylor series remainder is

$$k_b - T = \tilde{k}_{a_m}(z)(b-a_m)^{l(m)-1} + o(b-a_m)^{l(m)-1},$$

we obtain the desired relation (5.61). The proof is complete.

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, *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] 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,* Advancements in Complex Analysis. Springer, Cham, 2020: 1-37.

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

[7] R. Coifman, J. Peyriére, *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 $\mathbf{R}_+^{n+1}$ for* $1p$, J. Math. Anal. Appl. 483 (2020), no. 1, 123598, 19 pp.

[9] Lei Dai, Liming Zhang, Tao Qian, *A universal signal and image compression method based on Stochastic Adaptive Fourier Decomposition*, preprint.
[10] 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.
[11] W. Mi, T. Qian, *Frequency-domain identification: An algorithm based on an adaptive rational orthogonal system, Automatica*, 2012, 48(6), 1154-1162.
[12] W. Mi, T. Qian, "On backward shift algorithm for estimating poles of systems," *Automatica*, 2014, 50(6), 1603-1610.
[13] *Rational Approximation in Hardy Spaces on Strips,* Complex Variables and Elliptic Equations, **63** (2018), 1721-1738.
[14] *Aveiro Method in Reproducing Kernel Hilbert Spaces under Complete Dictionary*, Mathematical Methods in the Applied Sciences, **40** (2017), 7240-7254.
[15] *The Fourier Type Expansions on Tubes*, submitted for publication, 2018.
[16] E. D. Livshitz, V. N. Temlyakov, *On convergence of weak greedy algorithms*, SOUTH CAROLINA UNIV COLUMBIA DEPT OF MATHEMATICS, 2000.
[17] S. Mallat, Z. Zhang, *Matching pursuits with time-frequency dictionaries,* IEEE Trans. Signal Process. 41, 3397C3415 (1993)
[18] T. Qian, *Cyclic AFD Algorithm for Best Rational Approximation,* Mathematical Methods in the Applied Sciences, 2014, 37(6), 846-859.
[19] T. Qian, Intrinsic mono-component decomposition of functions: An advance of Fourier theory, Mathematical Methods in Applied Sciences, 2010, 33(7): 880-891.
[20] T. Qian, *Adaptive Fourier Decomposition,* Beijing: Science Press, 2015.
[21] T. Qian, Two-Dimensional Adaptive Fourier Decomposition, Mathematical Methods in the Applied Sciences, 2016, 39(10) : 2431-2448.
[22] T. Qian, *A novel Fourier theory on non-linear phase and applications,* Advances in Mathematics (China), 2018 **47**(3), pp 321-347 (in Chinese).
[23] T. Qian, *Reproducing Kernel Sparse Representations in Relation to Operator Equations,* Complex Anal. Oper. Theory 14 (2020), no. 2, 1C15.
[24] T. Qian, L.-H. Tan, J.-C. Chen, *A class of iterative greedy algorithmsrelated to Blaschke products,* accepted by SCIENCE CHINA, Mathematics.
[25] W. Qu, P. Dang, *Rational approximation in a class of weighted Hardy spaces,* Complex Analysis and Operator Theory, 13(4)(2019), 1827-1852.
[26] 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.
[27] T. Qian, Y.B. Wang, Adaptive Fourier series-a variation of greedy algorithm, Advances in Computational Mathematics 34 (3) (2011) 279–293.
[28] 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.
[29] 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.
[30] 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.
[31] 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.
[32] G. Ruckebusch, *Sur l'approximation rationnelle des filtres,* Report No 35 CMA Ecole Polytechnique, 1978.
[33] S. Saitoh, Y. Sawano, *Theory of Reproducing Kernels and Applications,* Singapore: Springer, 2016.
[34] E. M. Stein, G. Weiss, *Introduction to fourier analysis on Euclidean spaces,* Princeton University Press, 1971.
[35] C.Y. Tan, L.M. Zhang, T. Qian, *A New Supervised Learning Approach: Statistical Adaptive Fourier Decomposition (SAFD)*, preprint.
[36] C.Y. Tan, L.M. Zhang, T. Qian, S. Bras, and A. J. Pinho, *Statistical n-best AFD based Sparse Representation for ECG Biometric Identification*, preprint.

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

[38] 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.

[39] 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.

[40] Z. Wang et al., *Adaptive Fourier decomposition for multi-channel signal analysis: Finding common adaptive basis across channels*,

[41] 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.

[42] F. Yang, W. Qu, T. Qian, J.C. Chen, *Sparse representation solution for stochastic boundary value problems*, preprint.

Tao QIAN, Macao Centre for Mathematical Sciences, Macau University of Science and Technology, Macao

E-mail address: tqian@must.edu.mo