

Frequency analysis with multiple kernels and complete dictionary

Cuiyun Lin, Tao Qian *

Faculty of Innovation Engineering School of Computer Science and Engineering, Macao University of Science and Technology, Taipa, Macao, 999078, China

ARTICLE INFO

Keywords:

Complete Szegő kernel dictionary
Multiple kernels
Pre-orthogonal adaptive Fourier decomposition
 n -Best approximation
Unwinding Blaschke expansion
Laguerre system

ABSTRACT

In signal analysis, among the effort of seeking for efficient representations of a signal into the basic ones of meaningful frequencies, to extract principal frequency components, consecutively one after another or n at one time, is a fundamental strategy. For this goal, we define the concept of mean-frequency and develop the related frequency decomposition with the complete Szegő kernel dictionary, the latter consisting of the multiple kernels, being defined as the parameter-derivatives of the Szegő kernels. Several major energy matching pursuit type sparse representations, including greedy algorithm (GA), orthogonal greedy algorithm (OGA), adaptive Fourier decomposition (AFD), pre-orthogonal adaptive Fourier decomposition (POAFD), n -Best approximation, and unwinding Blaschke expansion, are analyzed and compared. Of which an order in reconstruction efficiency between the mentioned algorithms is given based on the detailed study of their respective remainders. The study spells out the natural connections between the multiple kernels and the related Laguerre system, and in particular, shows that both, like the Fourier series, extract out the $O(n^{-\sigma})$ order convergence rate from the functions in the Hardy-Sobolev space of order $\sigma > 0$. The existence of the n -Best approximation with the complete Szegő dictionary is proved and the related algorithm aspects are discussed. The included experiments form a significant integration part of the study, for they not only illustrate the theoretical results but also provide cross comparison between various ways of combination between the matching pursuit algorithms and the dictionaries in use. Experiments show that the complete dictionary remarkably improves approximation efficiency.

1. Introduction

Matching pursuit, as a methodology to generate sparse representations of signals, is usually based on a dictionary of the underlying Hilbert space. The most basic matching pursuit algorithm would be one called *greedy algorithm* in the context of Hilbert space with a dictionary [1,2]. In 2011, Qian and Wang proposed *AFD* or *Core AFD* [3–5], which crucially uses energy matching pursuit, as well as the complex Hardy space techniques, to develop a sparse representation in the form of a Takenaka-Malmquist system. Ever since then, there have been generalizations and variations of AFD, including *pre-orthogonal AFD* [6], the *n -Best AFD* [7], and *Unwinding AFD* [8]. A closely related method, called unwinding Blaschke expansion was proposed early by R. Coifman et al. in 2000 [9,10]. AFD was further extended to multivariate and matrix-valued cases [6,11,12]. In this paper, we restrict ourselves only to the one dimensional cases with scalar function values. AFD and its one dimensional variations are all based on the Szegő kernel dictionary. We, however, adopt a general formulation as follows.

* Corresponding author.

E-mail address: tqian@must.edu.mo (T. Qian).

Let \mathcal{H} be a Hilbert space with a dictionary \mathcal{D} . That means that \mathcal{D} consists of norm-one elements and $\text{span}\{\mathcal{D}\}$ is a dense subset of \mathcal{H} . Each of the elements of the dictionary \mathcal{D} is labeled by a parameter $q \in \mathcal{E}$, and is denoted as E_q . We will call a set \mathcal{K} of elements in \mathcal{H} a *pre-dictionary* if the unimodular normalizations of the elements in \mathcal{K} form a dictionary. In a reproducing kernel Hilbert space (RKHS), for instance, $K_q(p) \triangleq K(q, p)$ form a pre-dictionary. The parameter set \mathcal{E} is usually an open set of Euclidean space. We usually assume that the concerned E_q and K_q are smooth in q , and in particular, have as many orders of derivatives in \mathcal{E} as we use. The boundary of \mathcal{E} in the Euclidean topology is denoted $\partial\mathcal{E}$. The concerned general theory will be applied to RKHS and many non-RKHS cases as well.

In the present paper, we will study the case where \mathcal{H} is the complex Hardy space

$$\begin{aligned} H^2(\mathbf{D}) &= \{f : \mathbf{D} \rightarrow \mathbf{C} : f \text{ is analytic in } \mathbf{D}, \|f\|_{H^2(\mathbf{D})}^2 \\ &= \sup_{0 < r < 1} \frac{1}{2\pi} \int_0^{2\pi} |f(re^{it})|^2 dt < \infty\} \end{aligned} \quad (1)$$

together with the Szegő dictionary,

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

It is noted that for every $f \in H^2(\mathbf{D})$

$$\lim_{\mathbf{D} \ni z \rightarrow e^{it}} f(z) \text{ exist for a.e. } t \in [0, 2\pi],$$

where the limit $z \rightarrow e^{it}$ takes the non-tangential manner. The mapping that sends $f(z) \in H^2(\mathbf{D})$ to its non-tangential boundary limit $f(e^{it})$ on the unit circle is an isometric isomorphism between $H^2(\mathbf{D})$ and a closed subspace $H^2(\partial\mathbf{D})$ of $L^2(\partial\mathbf{D})$. In terms of the non-tangential boundary limit functions the inner product of $H^2(\mathbf{D})$ may be defined through that of the $L^2(\partial\mathbf{D})$ space, namely,

$$\langle f, g \rangle_{H^2} = \frac{1}{2\pi} \int_0^{2\pi} f(e^{it}) \overline{g(e^{it})} dt. \quad (3)$$

Using this inner product, as a consequence of the Cauchy formula, $H^2(\mathbf{D})$ is a RKHS having $k_a(z)$ as its reproducing kernel, where

$$k_a(z) = \frac{1}{1 - \bar{a}z}.$$

The Szegő dictionary element e_a is the norm-1 normalization of the Szegő kernel k_a . The collection $\mathcal{K} = \{k_a\}_{a \in \mathbf{D}}$ is a pre-dictionary.

The space of the boundary limit functions can be alternatively defined as

$$\begin{aligned} H^2(\partial\mathbf{D}) &= \{f : \partial\mathbf{D} \rightarrow \mathbf{C} : f(e^{it}) = \sum_{k=0}^{\infty} c_k e^{ikt}, \\ \|f\|_{H^2(\partial\mathbf{D})}^2 &\triangleq \sum_{k=0}^{\infty} |c_k|^2 < \infty\}. \end{aligned} \quad (4)$$

Remark 1. For any real-valued function of finite energy $g \in L^2(\partial\mathbf{D})$,

$$\begin{aligned} g^+ &\triangleq \frac{1}{2}(g + iHg + c_0) \in H^2(\mathbf{D}), \\ g^- &\triangleq \frac{1}{2}(g - iHg - c_0) \in H^2(\overline{\mathbf{D}}^c), \end{aligned}$$

the latter being the complex Hardy space outside the closure of the unit disc, and

$$g = g^+ + g^- = 2\text{Re}g^+ - c_0, \quad (5)$$

where c_0 is the average of g over the unit circle:

$$c_0 = \frac{1}{2\pi} \int_0^{2\pi} g(e^{it}) dt,$$

H is the circular Hilbert transform. For proofs of these fundamental relations, see, for instance, [13] or [14]. The relation (5), in particular, reduces the analysis of the real-valued functions on compact intervals to that of the functions in the Hardy space. The latter is, in fact, the space of the Z -transforms.

There exists a parallel theory for functions g defined on the real line with finite energy in which $g^\pm \triangleq \frac{1}{2}(g + iHg)$, where H is the Hilbert transform on the line and $g = 2\text{Reg}^+$. g^+ is identical to the Laplace transform of g . In both cases, g^+ is interpreted as the Gabor analytic signal associated with g . In the real line case, the associated Hardy space is $H^2(\mathbb{C}^+)$ that is isometric with $H^2(\partial\mathbb{C}^+) = H^2(\mathbb{R})$.

The same philosophy is obeyed by signals of several complex and several real variables (with the Clifford algebra setting), and of scalar- or vector-, and even matrix-values. Some of the cases have been developed. See [6,11,12,15] and the references therein.

In the rest of the paper we concentrate on developing the theory and practice in the unit disc case, and when we use the notation H^2 and the terminology “the Hardy space” we refer to just the unit disc case.

With the matching pursuit idea, the following n -Best question is natural.

The Ill-Posed n -Best Question:

Let $f \in \mathcal{H}$ and n be a fixed positive integer. Can one find n distinguished parameters q_1, \dots, q_n and n complex numbers c_1, \dots, c_n such that for the corresponding pre-dictionary elements K_{q_1}, \dots, K_{q_n} , there holds

$$\|f - \sum_{k=1}^n c_k K_{q_k}\|_{\mathcal{H}} = \inf \left\{ \|f - \sum_{k=1}^n d_k K_{p_k}\|_{\mathcal{H}} : \right. \\ \left. \text{all distinguished } p_k \in \mathcal{E} \text{ and all } d_k \in \mathbb{C} \right\}. \quad (6)$$

The answer to this question is dependent on the underlying Hilbert space \mathcal{H} and the pre-dictionary in use. Even for $n = 1$ the answer is not necessarily “Yes”. See, for instance, [16], for some counterexamples as weighted Hardy type spaces.

Definition 1. In the study of matching pursuit, the concept boundary vanishing condition (BVC) is established: A pair $(\mathcal{H}, \mathcal{D})$ is said to satisfy BVC if for every $f \in \mathcal{H}$ there holds

$$\lim_{q \rightarrow \partial\mathcal{E}} \langle f, E_q \rangle = 0. \quad (7)$$

If $(\mathcal{H}, \mathcal{D})$ satisfies BVC, then through a Bolzano-Weierstrass type compact argument the above n -Best problem has a solution for $n = 1$. It is proved in [3] that the Hardy space H^2 and the Szegő dictionary, as a pair, satisfies BVC, and hence there exist 1-Best solutions in the case. Verification of BVC usually involves detailed analysis. In general, if $(\mathcal{H}, \mathcal{D})$ satisfies BVC, then the powerful POAFD matching pursuit is available. See §3 below for details.

When $n > 1$ the answer to the n -Best question is “No” even for the Hardy space and the Szegő dictionary case. The following is an example of the ill-posed-ness for $n > 1$.

Example 1. We take $f(z) = \frac{1}{(2-z)^2}$, which is a function in the Hardy space $H^2(\mathbb{D})$, and take $n = 2$. The function can be infinitely approximated by linear combinations of two distinguished Szegő kernels. Or, the infimum error is zero. However, any two-term approximation cannot get a zero error.

However, if in the n -Best question the phrase “all distinguished $p_k \in \mathcal{E}$ ” is modified to be “all multiple kernels $\tilde{K}_{p_k}, p_k \in \mathcal{E}$ ”, then the problem in many cases becomes well-posed. In the Hardy space and Szegő dictionary case, for instance, multiple kernels are defined to be derivatives of K_a with respect to \bar{a} .

We now introduce multiple Szegő kernels and the complete Szegő kernel dictionary as follows [17]. Denote the set of non-negative integers by \mathbb{N} . Generating the notations in (2), we have

Definition 2. The set of multiple Szegő kernels

$$\tilde{\mathcal{K}} = \{k_{n,a}\}_{n \in \mathbb{N}, a \in \mathbb{D}},$$

is a pre-dictionary consisting of

$$k_{n,a}(z) = \left(\frac{\partial}{\partial \bar{a}} \right)^n k_a(z) = C_{n,a} \frac{z^n}{(1 - \bar{a}z)^{n+1}}, \quad (8)$$

where $C_{n,a}$ are constants depending on n and a . The totality of their unimodular normalizations

$$\tilde{\mathcal{D}} = \{e_{n,a}(z) = \frac{k_{n,a}}{\|k_{n,a}\|}\}_{n \in \mathbb{N}, a \in \mathbb{D}} \quad (9)$$

is called the *complete Szegő dictionary* with the parameter set

$$\mathcal{E} = \{(n, a)\}_{(n,a) \in \mathbb{N} \times \mathbb{D}}.$$

Returning to Example 1, $f(z) = 1/(z-2)^2$ may be expressed as $C \frac{\partial}{\partial a} k_a$, where $a = 1/2$ and C is a complex number. The 2-Best approximation (with infimum error 0) is then reached.

The multiple kernel notion may be extended to general Hilbert spaces \mathcal{H} with a pre-dictionary $D = \{K_q\}, q \in \mathcal{E}$. To simplify the terminology we will call all the subjects K_q, E_q, e_a, k_a , etc., by *kernel*, although some are normalized and some are not, and we will normally denote

$$\{E_q\} = \left\{ \frac{K_q}{\|K_q\|} \right\}, \quad q \in \mathcal{E}. \quad (10)$$

Note that for a kernel $K_q(x)$ the domain \mathcal{E} for the parameter q and that for the spatial or the time variable x may not be the same. Let l be a positive integer, and (q_1, \dots, q_l) an l -tuple of parameters in \mathcal{E} , allowing multiplicity. Define $l(q_1, \dots, q_l)$ be the repeating number of q_l in $q_1, \dots, q_l, 1 \leq l \leq n$. Without ambiguity, we write $l(q_1, \dots, q_l)$ as $l(q_l)$ in short. As examples, $l(q_1) = 1$, and $l(q_l) = 1$ if q_l is different from the proceeding q_1, \dots, q_{l-1} . For \mathcal{E} being an open set of the complex number field, we define

$$\tilde{K}_{q_l} = \left[\left(\frac{\partial}{\partial q} \right)^{(l(q_l)-1)} K_q \right] (q_l) \quad (11)$$

to be the l -th multiple kernels with respect to (q_1, \dots, q_l) , and \tilde{E}_{q_l} the norm-1 normalization of \tilde{K}_{q_l} . The differential operation $\frac{\partial}{\partial q}$ in below will also be denoted as $\bar{\partial}$. If q_l are several or hyper-complex variables, then the derivatives are replaced by directional derivatives. In the RKHS case there exists the following useful relation:

$$\bar{\partial}^{l(q_l)-1} f_{q_k} = \bar{\partial}^{l(q_l)-1} \langle f, K_{q_k} \rangle = \langle f, \tilde{K}_{q_k} \rangle. \quad (12)$$

Under the multiple kernel concept, the question (1) may be re-formulated to become well-posed:

The n -Best Question (reformulation): Let $f \in \mathcal{H}, \mathcal{K}$ a pre-dictionary, and n a fixed positive integer. Can one find n parameters q_1, \dots, q_n , with multiplicity when necessary, and n complex numbers c_1, \dots, c_n , such that for the multiple kernels $\tilde{K}_{q_1}, \dots, \tilde{K}_{q_n}$ there holds

$$\|f - \sum_{k=1}^n c_k \tilde{K}_{q_k}\|_{\mathcal{H}} = \inf \left\{ \|f - \sum_{k=1}^n d_k K_{p_k}\|_{\mathcal{H}} : \right. \\ \left. \text{all distinguished } p_k \in \mathcal{E} \text{ and all } d_k \in \mathbb{C} \right\}, \quad (13)$$

where $\tilde{K}_{q_k}, k = 1, \dots, n$, are, consecutively, the multiple kernels associated with (q_1, \dots, q_n) .

Under the new formulation, the existence of a solution to the n -Best problem has been proved for a number of most commonly studied Hilbert spaces of a BVC dictionary, including the Hardy space, the Bergman space, and the weighted Bergman spaces together with the dictionaries naturally induced by their respective reproducing kernels. See [7,18] and the references therein. The n -Best approximation problem in the Hardy space is, in fact, equivalent to the best approximation problem by rational functions in the space of degrees not exceeding n [19]. Several practical algorithms for the Hardy n -Best have been proposed that, however, cannot prevent from sinking into the local minima [20–23]. Through generalizing the techniques in relation to the backward shift operator, the existence of the n -Best approximation was lately extended to a class of RKHSs, including the weighted Bergman and weighted Hardy spaces as particular cases [24]. These existence proofs also play a definitive role in seeking for a theoretical algorithm to obtain all the n -tuple minimizers. In the present paper, we prove the existence of the n -Best approximation for the Hardy space under the complete Szegő kernel dictionary. See §6.

Besides the one for the n -Best approximation, there is another reason that motivates the study of multiple kernels: We are to decompose a signal into its principal frequency components in terms of the energy, not in the degree of the frequency. In §2 we define the notion mean-frequency for functions in the Hardy space. Mean-frequency is a measurement of the total amount of the frequencies in an analytic signal by which Szegő kernels possess zero mean-frequency. All the concerned matching pursuit algorithms in the context are to select one after another dictionary elements, but unfortunately restricted to only the zero mean-frequency ones. In such a way, the high frequency terms are generated by the GS orthogonalization process, in which the order of applying GS process is a matter. What is desirable in frequency decomposition, however, would be the principal components directly related to the signal. They should be selected in terms of the greatest energy matching to the complete dictionary elements possessing any frequency, or, in other words, not be restricted to prescribed frequency levels. This question was also raised by [25] who pointed out that the existing analytic frequency decomposition, i.e., adaptive Fourier decomposition is not according to principal frequency components, but constructed from representatives of the lowest frequency. The present study points out how by using the complete dictionary and performing POAFD the goal of the principal frequency component may be achieved.

The writing plan for this paper is as follows. In §2 we establish the notion of *mean-frequency* and prove some basic results. Mean-frequency is used to measure the total amount, or degree level, of frequencies that an analytic signal contains. In §3 we give a concise summary but detailed analysis of the most commonly concerned matching pursuit algorithms. The analyzed algorithms include AFD, GA, OGA, POAFD, and n -Best. We establish an order between them in accordance with their re-constructing efficiencies. §4 is devoted to a detailed study of POAFD over the complete Szegő dictionary. In §5 we study convergence rates in relation to the multiple Szegő kernels and the Laguerre systems. The main results include that, as a generalized form of the Riemann-Lebesgue Lemma, $\langle f, e_{k,a} \rangle$ tends to zero with the order $O(n^{-\sigma})$ for functions f in the Hardy-Sobolev space of order $\sigma > 0$; and as for the Fourier series, the Fourier-Laguerre series in the σ -Hardy-Sobolev space is of the same convergence rate $O(n^{-\sigma})$. In §6 we prove

the existence of the n -Best approximation with the complete Szegő kernel dictionary. The existence cannot be deduced from the existing results for RKHSs. We also discuss theoretical and practical algorithms of the n -Best solutions. §7 contains a great number of experiments for comparison between the reconstruction efficiencies of AFD, GA, OGA, POAFD, and n -Best over, respectively, the Szegő and the complete Szegő dictionaries, as well as with the Unwinding Blaschke expansion. They stand as a significant integration part of the paper. The experiments may be divided into two types, of which one is to verify the theoretical ordering of strongness of the concerned matching pursuit algorithms; and the other is to show the strongness of the complete dictionary itself: The complete Szegő dictionary with weaker algorithms may be stronger than the Szegő dictionary used with stronger type algorithms. In §8 conclusions and comments are drawn.

2. Mean-frequency of signals in the Hardy space

It is basic knowledge that any Hardy space function $f \in H^2(\mathbf{D})$ has a factorization $f(z) = \phi(z)s(z)o(z)$, where $\phi(z)$, $s(z)$ and $o(z)$ are, respectively, the Blaschke product, the singular inner function, and the outer function parts of f [13]. The factorization is unique up to unimodular constants. The non-tangential boundary limits of the three functions have, respectively, the forms (in almost everywhere sense on the boundary)

$$\phi(e^{it}) = e^{i\theta_\phi(t)}, \quad s(e^{it}) = e^{i\theta_s(t)}, \quad o(e^{it}) = \rho_o(t)e^{i\theta_o(t)}, \quad (14)$$

where $\rho_o(t) \geq 0$, and $\theta_\phi(t)$, $\theta_s(t)$ and $\theta_o(t)$ are real-valued. As proved in [26],

$$\theta'_\phi \geq 0, \quad \theta'_s \geq 0, \quad \text{a.e.}, \quad (15)$$

and

$$\int_0^{2\pi} \theta'_\phi(t) dt = N \geq 0, \quad (16)$$

where N is the number of zeros (can be zero or ∞) of the Blaschke product ϕ . Moreover, with mild conditions on f to guarantee absolute continuity of $\theta_o(t)$, there holds

$$\int_0^{2\pi} \theta'_o(t) dt = 0. \quad (17)$$

The last equation shows that the frequency function of an outer function is negative on a set of positive Lebesgue measures if the outer function itself is not identical to the zero function. Signals that have the property

$$\theta'_\phi(t) + \theta'_s(t) + \theta'_o(t) \geq 0, \quad \text{a.e.}, \quad (18)$$

are called *mono-components* [26]. A Hardy space function may not be a mono-component. However, any Hardy space function has a none-negative *mean-frequency*, as defined in

Definition 3. For a Hardy space function f the quantity

$$\int_0^{2\pi} \theta'_\phi(t) dt + \int_0^{2\pi} \theta'_s(t) dy \quad (19)$$

is called the *mean-frequency* of f , denoted as $M_F(f)$.

We note that M_F is additive in the sense $M_F(fg) = M_F(f) + M_F(g)$.

The phase derivative of a Möbius transform

$$\tau_a(z) = \frac{z-a}{1-\bar{a}z}$$

is easily computed. Let $\tau(e^{it}) = e^{i\psi_a(t)}$. Then $\psi'_a(t)$ is the Poisson kernel [13]

$$P_{|a|}(e^{i(s-t)}), \quad a = |a|e^{is}.$$

Therefore,

$$M_F(\tau_a(z)) = \frac{1}{2\pi} \int_0^{2\pi} P_{|a|}(e^{s-t}) = 1.$$

Due to the additivity, for $n, m \geq 0$, $M_F(z^n) = n$, and

$$M_F\left(\prod_{k=1}^m \frac{z - a_k}{1 - \bar{a}_k z}\right) = m.$$

Since the Szegő kernel k_a and Szegő dictionary elements e_a , are outer functions,

$$M_F(k_a(z)) = M_F(e_a(z)) = 0.$$

There exist standard models of functions in the Hardy space that form systems with increasing mean-frequencies. Let $\mathbf{a} = (a_1, \dots, a_n)$ be an n -tuple of elements in the unit disc \mathbf{D} . Associated with \mathbf{a} there exists an order- n Takenaka-Malmquist (TM) system,

$$\{B_k\}_{k=1}^n, \quad B_k(z) = e_k(z) \prod_{l=1}^{k-1} \frac{z - a_l}{1 - \bar{a}_l z}.$$

B_k is also denoted as B_{a_1, \dots, a_k} to specify the dependence on a_1, \dots, a_k . The n -tuple \mathbf{a} can be extended to become an infinite sequence in \mathbf{D} , and thus define an order- ∞ TM system $\{B_k\}_{k=1}^\infty$. We use the notation $\{B_k\}$ to denote either a finite or an infinite TM system. In this paper, unless otherwise specified, in an n -tuple or an infinite sequence of $\{a_k\}$, multiplicities of the number a_k is allowed. A system $\{B_k\}$ is, although not necessarily complete, orthonormal in $H^2(\mathbf{D}) = H^2(\partial\mathbf{D})$ with the inner product given by (3). It is known that an infinite TM system is a basis of $H^2(\mathbf{D})$ if and only if the sequence \mathbf{a} satisfies the hyperbolic non-separability condition

$$\sum_{k=1}^\infty (1 - |a_k|) = \infty.$$

The Fourier basis $\{z^k\}_{k=0}^\infty$ is a particular case corresponding to all $a_k \equiv 0, k = 1, \dots$

We have the following result.

Theorem 1. $M_F(k_{n,a}) = M_F(e_{n,a}) = n$, and $M_F(B_n) = n - 1$.

Proof. We note that $\frac{1}{(1 - \bar{a}z)^{n+1}}$ is an outer function. As a consequence of the additivity, there holds

$$M_F(k_{n,a}) = M_F(z^n) + M_F\left(\frac{1}{(1 - \bar{a}z)^{n+1}}\right) = n.$$

As a consequence of $M_F(\tau_{a_1}) = 1$ and $M_F(e_{a_n}) = 0$, we have

$$M_F(B_n) = M_F\left(\prod_{l=1}^{n-1} \frac{z - a_l}{1 - \bar{a}_l z}\right) = n - 1.$$

Definition 4. A consecutive multiple Szegő kernel \tilde{k}_a with respect to the m -tuple (a_1, \dots, a_{m-1}, a) is defined to be

$$\tilde{k}_a(z) \triangleq k_{l(a)-1,a}(z),$$

where $l(a)$, with a little abuse of notation, denotes the repeating number of a in (a_1, \dots, a_{m-1}, a) . In particular, if there is no repeating, that is $l(a) = 1$, then $\tilde{k}_a(z) = k_a(z)$.

Computation gives

$$k_{l(a)-1,a} = (-1)^{l(a)-1} (l(a) - 1)! \frac{z^{l(a)-1}}{(1 - \bar{a}z)^{l(a)}}.$$

In [27] the following result is proved.

Theorem 2. Let $\{B_k\}$ be the TM system corresponding to any given finite or infinite sequence (a_1, \dots) . Then, up to unimodular multiplicative constants, $\{B_k\}$ is the consecutive Gram-Schmidt orthonormalization of the consecutive multiple Szegő kernel \tilde{k}_{a_k} .

3. Algorithms adopting matching pursuit methodology: analysis and comparison

Apart from n -Best approximation in Hilbert spaces with a BVC dictionary, there exist a number of iterative type algorithms adopting the matching pursuit methodology. These algorithms have common, as well as different features and individual effectiveness in terms of the reconstruction of signals. In this section, we review and analyze AFD, GA, OGA, and POAFD. With a criterion, we compare them and give them an order in terms of their reconstruction efficiencies. Among the four methods, AFD and POAFD belong to the same class (the AFD type), and GA and OGA belong to another (the greedy type). The AFD type is based on delicate complex and harmonic analysis aiming at frequency decomposition and attainability of the supreme energy matching pursuit, while

the greedy type is mostly in the general functional analysis context applicable to a Hilbert space with any dictionary. The AFD type is a generalization of Fourier theory, applicable for the reproducing kernel Hilbert space setting, having generalizations to multivariate functions (several complex variables and Clifford algebra variables) with vector- and matrix-values [6,11,12,15]. We below concentrate on the basic unit disc context.

3.1. Adaptive Fourier decomposition: AFD [28]

Let $(\mathcal{H}, \mathcal{D}) = (H^2(\mathcal{D}), \{e_a\}_{a \in \mathcal{D}})$. It is proved in [3] that \mathcal{D} is a dictionary in $H^2(\mathcal{D})$ satisfying BVC. For $f \in H^2(\mathcal{D})$, $f_1 = f$, as a consequence of BVC, one can find

$$a_1 = \arg \max \{ |\langle f_1, e_b \rangle| : b \in \mathcal{D} \},$$

and

$$\|f - \langle f_1, e_{a_1} \rangle e_{a_1}\|^2$$

is hence minimized over all one-dimensional linear spaces generated by $e_b, b \in \mathcal{D}$. Define, for $k \geq 2$,

$$f_k(z) = \frac{f_{k-1}(z) - \langle f_{k-1}, e_{a_{k-1}} \rangle e_{a_{k-1}}(z)}{\frac{z - a_{k-1}}{1 - \bar{a}_{k-1}z}},$$

and select

$$a_k = \arg \max \{ |\langle f_k, e_b \rangle| : b \in \mathcal{D} \}. \quad (20)$$

We call f_k the *reduced remainder*, being the image of the a_k -generalized backward shift operator to f_{k-1} . It can be shown, inductively, $f_k \in H^2(\mathcal{D})$, and

$$f(z) = \sum_{l=1}^k \langle f_l, e_{a_l} \rangle B_l(z) + f_{k+1}(z) \prod_{l=1}^k \frac{z - a_l}{1 - \bar{a}_l z}, \quad (21)$$

where $\{B_l\}_{l=1}^k$ is the k -TM system defined by (a_1, \dots, a_k) . We note that the above procedure allows multiplicity of the parameter a_k selection. We note that the generalized backward shifts automatically generate the orthonormal k -TM system $\{B_l\}_{l=1}^k$. The orthogonality relations imply the useful relations

$$\langle f_l, e_{a_l} \rangle = \langle f_l^\dagger, B_l \rangle = \langle f, B_l \rangle, \quad (22)$$

where

$$f_l^\dagger \triangleq f - \sum_{j=1}^{l-1} \langle f_j, e_{a_j} \rangle B_j(z) = f_l(z) \prod_{j=1}^{l-1} \frac{z - a_j}{1 - \bar{a}_j z} \quad (23)$$

is the l -th-(AFD) *orthogonal remainder*. We note that from the first identical relation in (23) the l -th-(AFD) orthogonal remainder $f_l^\dagger(z)$ is orthogonal with all the terms $B_j, j < l$; and, from the second identical relation in (23),

$$\sum_{j=1}^{l-1} \langle f_j, e_{a_j} \rangle B_j(z)$$

is an interpolation rational function of f at the points a_1, \dots, a_{l-1} .

It turns out that the partial sums converge to the original function f [3].

$$f(z) = \sum_{l=1}^{\infty} \langle f_l, e_{a_l} \rangle B_l(z).$$

We will the convergence rates in the latter part of this section.

3.2. Greedy algorithm: GA [1,2,29]

Greedy algorithm is applicable to a general Hilbert space \mathcal{H} with a dictionary \mathcal{D} . To compare it with other matching pursuit algorithms we assume that \mathcal{D} satisfies BVC.

Let $f \in \mathcal{H}$, and $g_1 = f$. The first matching pursuit step is the same as AFD: Owing to BVC we can select

$$q_1 = \arg \max \{ |\langle g_1, E_p \rangle| : p \in \mathcal{E} \}. \quad (24)$$

There holds

$$f = \langle f_1, E_{q_1} \rangle E_{q_1} + g_2.$$

Since g_2 is orthogonal with E_{q_1} , we have

$$\|f\|^2 = |\langle f_1, E_{q_1} \rangle|^2 + \|g_2\|^2.$$

Due to the maximal selection of q_1 in (24) the remaining energy in g_2 is minimized. Iteratively, define g_k to be the k -th *iterative remainder* given by

$$f = \sum_{l=1}^{k-1} \langle g_l, E_{q_l} \rangle E_{q_l} + g_k, \quad (25)$$

where

$$q_l = \arg \max \{ |\langle g_l, E_p \rangle| : p \in \mathcal{E}, l = 1, \dots, k-1. \} \quad (26)$$

We note that for each $l = 1, \dots, k$, the remainder g_l is orthogonal with the last $E_{q_{l-1}}$. There follows

$$\begin{aligned} \|f\|^2 &= |\langle g_1, E_{q_1} \rangle|^2 + |\langle g_2, E_{q_2} \rangle|^2 + \|g_3\|^2 \\ &= \sum_{l=1}^{k-1} |\langle g_l, E_{q_l} \rangle|^2 + \|g_k\|^2. \end{aligned}$$

The energy of g_k decays to zero, and, as a consequence,

$$f = \sum_{l=1}^{\infty} \langle g_k, E_{q_k} \rangle E_{q_k}. \quad (27)$$

3.3. Orthogonal greedy algorithm: OGA [1,2,29]

We are under the same assumptions as for GA: We have a pair $(\mathcal{H}, \mathcal{D})$, where $\mathcal{D} = \{E_q\}, q \in \mathcal{E}$, is a dictionary of \mathcal{H} satisfying BVC. Let $f \in \mathcal{H}$ and $h_1 = f$. The first parameter selection is the same as that for AFD and GA: BVC implies that we are able to select

$$q_1 = \arg \max \{ |\langle h_1, E_p \rangle| : p \in \mathcal{E} \}.$$

There follows then

$$h_2 = f - \text{Proj}_{\{E_{q_1}\}}(f) = g_2 \perp E_{q_1},$$

where we adopt the notation $\text{Proj}_{\mathcal{A}}(f)$ for the orthogonal projection of f into the linear span of the functions in \mathcal{A} . We also use the notation $Q_{\mathcal{A}} \triangleq I - \text{Proj}_{\mathcal{A}} = \text{Proj}_{\mathcal{A}^\perp}$. $Q_{\mathcal{A}}$ is also called the *Gram-Schmidt orthogonalization operator* (GS operator) with respect to the function set \mathcal{A} .

We select

$$q_2 = \arg \max \{ |\langle h_2, E_p \rangle| : p \in \mathcal{E} \}. \quad (28)$$

Since $h_2 = g_2$, this maximal selection principle is the same as for GA. Note that E_{q_1} and E_{q_2} are not necessarily orthogonal. Define the 3rd orthogonal remainder

$$h_3 = f - \text{Proj}_{\{E_{q_1}, E_{q_2}\}}(f).$$

The Hilbert space property implies

$$\|h_3\| \leq \|g_3\|. \quad (29)$$

We will refer to this fact as “OGA is superior to GA”, or say that as a matching pursuit algorithm “OGA is stronger than GA”. Proceeding like this, we obtain the selections $\{E_{q_1}, \dots, E_{q_{k-1}}\}$ and formulate the k -th *remainder* h_k with the relation

$$f = \text{Proj}_{\{E_{q_1}, \dots, E_{q_{k-1}}\}}(f) + h_k, \quad (30)$$

where

$$q_l = \arg \max \{ |\langle h_l, E_p \rangle| : p \in \mathcal{E}, l = 1, \dots, k-1. \} \quad (31)$$

Consecutively, we select

$$q_k = \arg \max \{ |\langle h_k, E_p \rangle| : p \in \mathcal{E} \}.$$

It may be proved that $\|h_k\| \rightarrow 0$ as $k \rightarrow \infty$. As a consequence, there holds

$$f = \lim_{k \rightarrow \infty} \text{Proj}_{\{E_{q_1}, \dots, E_{q_k}\}}(f) = \sum_{l=1}^{\infty} \langle f, E_{q_1, \dots, q_l} \rangle E_{q_1, \dots, q_l},$$

where $(E_{q_1}, E_{q_1, q_2}, \dots, E_{q_1, \dots, q_l})$ is the consecutive GS orthogonalization of $(E_{q_1}, E_{q_2}, \dots, E_{q_l})$. We note that like GA under OGA the selected parameters have no multiplicity.

3.4. Pre-orthogonal adaptive Fourier decomposition: POAFD

As for GA and OGA we assume that \mathcal{H} is a general Hilbert space with a BVC dictionary $D = \{E_q\}, q \in \mathcal{E}$. Since we will involve multiple kernels, we assume E_q to be differentiable with respect to q up to the needed orders. When q is a vector, the existence of directional derivatives is assumed.

The first step matching pursuit is again the same as that for AFD, GA, and OGA:

$$q_1 = \arg \max \{ |\langle f, E_p \rangle| : p \in \mathcal{E} \}.$$

The selection of the second parameter q_2 now is different from that for GA and OGA (the latter two being the same at the q_2 selection):

$$q_2 = \arg \sup \{ |\langle f, E_{q_1, q} \rangle| : q \in \mathcal{E} \}, \quad (32)$$

where $E_{q_1, q}$ is defined by the relation that $(E_{q_1}, E_{q_1, q})$ is the GS orthonormalization of (E_{q_1}, \tilde{E}_q) . The function $E_{q_1, q}$ is unique up to an unimodular complex multiplicative constant. Owing to this step the algorithm is called “pre-orthogonal”: The orthogonalization is done prior to the maximal selection. In general, the POAFD maximal selection principle is

$$q_n = \arg \sup \{ |\langle f, E_{q_1, \dots, q_{n-1}, q} \rangle| : q \in \mathcal{E} \}, \quad (33)$$

where $(E_{q_1}, E_{q_1, q_2}, \dots, E_{q_1, \dots, q_{n-1}}, E_{q_1, \dots, q_{n-1}, q})$, regarded as TM system generated by D in \mathcal{H} , is the consecutive GS orthogonalization of

$$(E_{q_1}, \tilde{E}_{q_2}, \dots, \tilde{E}_{q_{n-1}}, \tilde{E}_q).$$

The k -th POAFD remainder denoted as h_k^\dagger , is defined through

$$f = \sum_{l=1}^{k-1} \langle f, E_{q_1, \dots, q_l} \rangle E_{q_1, \dots, q_l} + h_k^\dagger, \quad (34)$$

where the parameters are selected according to the POAFD maximal selection principle (33). Both being orthogonal remainders, the h_k^\dagger 's are different from the h_k 's for the latter are orthogonal remainders for OGA defined by (30) depending on the parameters selected according to OGA. POAFD was, in fact, suggested by the relation (22) in AFD in the Hardy space setting. In fact, writing $Q_{\tilde{E}_{q_1}, \dots, \tilde{E}_{q_l}}$ briefly as Q_{q_1, \dots, q_l} , due to its properties as a projection, there holds

$$\langle f, E_{q_1, \dots, q_l} \rangle = \langle f, Q_{q_1, \dots, q_{l-1}} E_{q_1, \dots, q_l} \rangle = \langle h_l^\dagger, E_{q_1, \dots, q_l} \rangle,$$

which is (22) in which h_l^\dagger was written as f_l^\dagger to indicate its connection with the reduced AFD remainder f_l . The coefficients in (34) can also be written in the following forms

$$\begin{aligned} & \left\langle f, \frac{Q_{q_1, \dots, q_{k-1}}(\tilde{E}_{q_k})}{\|Q_{q_1, \dots, q_{k-1}}(\tilde{E}_{q_k})\|} \right\rangle \\ &= \left\langle Q_{q_1, \dots, q_{k-1}}(f), \frac{Q_{q_1, \dots, q_{k-1}}(\tilde{E}_{q_k})}{\|Q_{q_1, \dots, q_{k-1}}(\tilde{E}_{q_k})\|} \right\rangle \\ &= \left\langle \frac{Q_{q_1, \dots, q_{k-1}}(f)}{\|Q_{q_1, \dots, q_{k-1}}(\tilde{E}_{q_k})\|}, E_{q_k} \right\rangle. \end{aligned} \quad (35)$$

As a reformulation of (22) this exhibits that, in a general Hilbert space with a BVC dictionary, POAFD is an analogous algorithm to Core AFD for the Hardy space. In such formation, in particular,

$$\frac{Q_{q_1, \dots, q_{k-1}}(f)}{\|Q_{q_1, \dots, q_{k-1}}(\tilde{E}_{q_k})\|}$$

has the same role as the k -th *reduced remainder*, being obtained in the Hardy space case through the generalized backward shift operators. Adopting the notation for the classical TM system,

$$B_k^{q_k} \triangleq \frac{Q_{q_1, \dots, q_{k-1}}(\tilde{E}_{q_k})}{\|Q_{q_1, \dots, q_{k-1}}(\tilde{E}_{q_k})\|}$$

can be said from the TM system in the context. And

$$Q_{q_1, \dots, q_{k-1}}(f)$$

is the k -th *orthogonal remainder* with respect to the POAFD selected parameters q_1, \dots, q_{k-1} . The relation (35) in particular, shows that the POAFD maximal selection principle corresponding to (33) is performable.

For all four types of matching pursuit algorithms, we have the following fundamental results [1–3,6,29].

Theorem 3. Corresponding to the four different types (or contexts) of parameter maximal selection principles, namely (20), (26), (31), and (33), the four remainders f_k^\dagger, g_k, h_k and h_k^\dagger defined respectively through (23), (25), (30), and (34), all tend to zero in their respective Hilbert norms. Hence the corresponding partial sums all converge to the originally given signal f . Moreover, if f belongs to

$$H_M = \{f \in \mathcal{H} : \exists \{q_k\} \subset \mathcal{E}, f = \sum_{k=1}^{\infty} c_k E_{q_k}, \sum_{k=1}^{\infty} |c_k| \leq M\},$$

then the norm of each of the above four types of k -remainders is dominated by $\frac{M}{\sqrt{k}}$.

Remark 2. The rate $O(\frac{1}{\sqrt{n}})$ of convergence is valid for all matching pursuit algorithms [29]. The reference [29] constructs concrete examples to show that the convergence rate cannot be improved. The existence of such examples may also be asserted from the Karhunen-Loeve (KL) expansions. In fact, the KL expansion of the Brownian bridge, as an example, can be precisely estimated by

$$\mathbb{E} \left[\|B - S_n\|_{L^2[0,1]}^2 \right] = \sum_{j=n+1}^{\infty} \frac{1}{\pi^2 j^2} \sim \frac{1}{\pi^2 n}$$

(page 206 of [30]). This estimation shows that there exist sample paths of Brownian bridge whose eigenfunction expansions have convergence rates as worse as $O(\frac{1}{\sqrt{n}})$. On the other hand, due to the optimality of the KL expansion over all orthonormal expansions, there must exist matching pursuit expansions whose convergence rates are as worse as $O(\frac{1}{\sqrt{n}})$ as well.

Efficiencies of the individual matching pursuit algorithms cannot be well compared in general. The parameters that give rise to the best matching pursuit may not be unique. The step by step optimality does not accumulate, and finally may not result in the overall optimality. Nevertheless, we can still draw a comparison under an intuitive criterion. We will first analyze the optimality of POAFD.

By using the GS operator, there holds, for $q \neq q_1$,

$$E_{q_1,q} = \frac{E_q - \langle E_q, E_{q_1} \rangle E_{q_1}}{\|E_q - \langle E_q, E_{q_1} \rangle E_{q_1}\|} = \frac{Q_{E_{q_1}}(E_q)}{\sqrt{1 - |\langle E_q, E_{q_1} \rangle|^2}}.$$

Since $Q_{E_{q_1}}$ is self-adjoint, there follows

$$\langle f, E_{q_1,q} \rangle = \frac{\langle f, Q_{E_{q_1}}(E_q) \rangle}{\sqrt{1 - |\langle E_q, E_{q_1} \rangle|^2}} = \frac{\langle Q_{E_{q_1}} f, E_q \rangle}{\sqrt{1 - |\langle E_q, E_{q_1} \rangle|^2}}.$$

Replacing $h_2 = Q_{E_{q_1}} f$, we have

$$\begin{aligned} \sup_q |\langle f, E_{q_1,q} \rangle| &= \sup_q |\langle h_2, E_{q_1,q} \rangle| \\ &= \sup_q \left| \frac{\langle h_2, E_q \rangle}{\sqrt{1 - |\langle E_q, E_{q_1} \rangle|^2}} \right| \\ &\geq \sup_q |\langle h_2, E_q \rangle|. \end{aligned} \quad (36)$$

Due to the continuity, this argument is also valid for the limiting case $q = q_1$. Recalling (32), the last inequality shows that the POAFD energy matching is superior to that for OGA given in (28). As a consequence we have

$$\|h_3^\dagger\| \leq \|h_3\|. \quad (37)$$

In the analyzed cases with the convention $f = f_1^\dagger = g_1 = h_1 = h_1^\dagger$, we have $g_2 = h_2, f_k^\dagger = h_k^\dagger, k \geq 3$. It is observed that examples for which the strict inequality signs in (29) or (37) hold may be constructed. The above argument may be generalized: For h_k being in the orthogonal complement of $\text{span}\{B_1, \dots, B_{k-1}\}$,

$$\begin{aligned} \langle f, B_k^q \rangle &= \langle h_k, \frac{Q_{q_1, \dots, q_{k-1}} E_q}{\sqrt{1 - \sum_{l=1}^{k-1} |\langle E_q, B_l \rangle|^2}} \rangle \\ &= \langle h_k, \frac{E_q}{\sqrt{1 - \sum_{l=1}^{k-1} |\langle E_q, B_l \rangle|^2}} \rangle \\ &= \frac{\langle h_k, E_q \rangle}{\sqrt{1 - \sum_{l=1}^{k-1} |\langle E_q, B_l \rangle|^2}}. \end{aligned}$$

In the next inequality-equality chain the left-end is the maximal selection principle of OGA, and the right-end is for POAFD, showing that POAFD is superior to OGA.

$$\begin{aligned} &\sup\{|\langle h_k, E_q \rangle| \mid q \in \mathcal{E}\} \\ &\leq \sup\left\{ \frac{|\langle h_k, E_q \rangle|}{\sqrt{1 - \sum_{l=1}^{k-1} |\langle E_q, B_l \rangle|^2}} \mid q \in \mathcal{E} \right\} \\ &= \sup\{|\langle f, B_k^q \rangle| \mid q \in \mathcal{E}\}. \end{aligned}$$

Let, in general, Algorithm 1 and Algorithm 2 be among the concerned algorithms AFD, GA, OGA, and POAFD, etc. If there exists a positive integer k_0 such that

- (i): for any signal f the energies of the l -remainders for $l < k_0$ of the two algorithms can be made to be the same through their respective optimal matching pursuit selections; and
- (ii): the energies of the k_0 th remainder of Algorithm 1 are not larger than those of Algorithm 2, and for some particular signals f strictly less than those of Algorithm 2,

then we say that Algorithm 1 is *superior to* (or *stronger than*) Algorithm 2. In this case we write Algorithm 1 \geq Algorithm 2.

Theorem 4. For any Hilbert space with a dictionary satisfying BVC the associated algorithms satisfy

$$POAFD \geq OGA \geq GA.$$

When the n -Best approximation exists, there holds

$$n - \text{Best} \geq \text{any matching pursuit algorithm}.$$

4. Pre-orthogonal adaptive Fourier decomposition with the complete Szegő dictionary

This section will be devoted to a detailed study of POAFD on the complete Szegő dictionary. Recall that in (8) and (9) we defined the multiple kernel $k_{n,a}$ and their normalizations, the dictionary elements $e_{n,a}$. The Szegő complete dictionary is denoted \tilde{D} . The quantity of the norm of $k_{n,a}$ is computed as follows.

Lemma 1.

$$\begin{aligned} &\left\| \left(\frac{\partial}{\partial \bar{a}} \right)^n k_a(z) \right\|^2 \\ &= \sum_{m=0}^n C_n^m n! \frac{(2n-m)!}{(n-m)!} (a\bar{a})^{n-m} (1-\bar{a}a)^{m-2n-1}. \end{aligned} \quad (38)$$

Proof.

$$\begin{aligned} &\left\| \left(\frac{\partial}{\partial \bar{a}} \right)^n k_a(z) \right\|^2 \\ &= \left\langle \left(\frac{\partial}{\partial \bar{a}} \right)^n k_a(z), \left(\frac{\partial}{\partial \bar{a}} \right)^n k_a(z) \right\rangle \\ &= \left(\frac{\partial}{\partial a} \right)^n \left\langle \left(\frac{\partial}{\partial \bar{a}} \right)^n k_a(z), k_a(z) \right\rangle \end{aligned}$$

$$= \left(\frac{\partial}{\partial a} \right)^n \left(\frac{\partial}{\partial \bar{a}} \right)^n k_a(a).$$

Through induction, we can show

$$\left(\frac{\partial}{\partial \bar{a}} \right)^n k_a(a) = \frac{n! a^n}{(1 - \bar{a}a)^{n+1}}.$$

Therefore,

$$\begin{aligned} & \left(\frac{\partial}{\partial a} \right)^n \left(\frac{\partial}{\partial \bar{a}} \right)^n k_a(a) \\ &= \left(\frac{\partial}{\partial a} \right)^n \left(\frac{n! a^n}{(1 - \bar{a}a)^{n+1}} \right) \\ &= \sum_{m=0}^n C_n^m (n! a^n)^{(m)} [(1 - \bar{a}a)^{-n-1}]^{(n-m)} \\ &= \sum_{m=0}^n C_n^m n! \frac{(2n-m)!}{(n-m)!} (a\bar{a})^{n-m} (1 - \bar{a}a)^{m-2n-1} \\ &= (2n)! \frac{|a|^{2n}}{(1 - |a|^2)^{2n+1}} + \dots \end{aligned} \quad (39)$$

Lemma 2. Let $f \in H^2(\mathbb{D})$. There holds for all $|a| < 1$ uniformly

$$\lim_{n \rightarrow \infty} |\langle f, e_{n,a} \rangle|(z) = 0. \quad (40)$$

Proof. For any $\varepsilon > 0$, due to L^2 -convergence of Fourier series, there exists a polynomial $h(z)$ such that

$$\|f - h\| < \varepsilon.$$

Denote by N the degree of the polynomial h . When $n > N$, in view of the Cauchy-Schwarz inequality and (12), there holds for any $a \in \mathbb{D}$

$$\begin{aligned} |\langle f, e_{n,a} \rangle| &\leq |\langle f - h, e_{n,a} \rangle| + |\langle h, e_{n,a} \rangle| \\ &< \varepsilon + |\langle h(z), e_{n,a} \rangle| \\ &= \varepsilon + \left\langle h(z), \frac{\left(\frac{\partial}{\partial \bar{a}}\right)^n k_a(z)}{\left\| \left(\frac{\partial}{\partial \bar{a}}\right)^n k_a(z) \right\|} \right\rangle \\ &= \varepsilon + \frac{1}{\left\| \left(\frac{\partial}{\partial \bar{a}}\right)^n k_a(z) \right\|} \left\langle h(z)^{(n)}, k_a(z) \right\rangle \\ &= \varepsilon. \end{aligned}$$

Lemma 3. For $f \in H^2(\mathbb{D})$ there holds uniformly for $n \in \mathbb{N}$

$$\lim_{|a| \rightarrow 1} |\langle f, e_{n,a} \rangle| = 0. \quad (41)$$

Proof. Given $\varepsilon > 0$. Due to Lemma 2, we can restrict ourselves to verifying the convergence for $n \leq N$. Since the span of the parametrized Szegő kernels is dense in the whole space, the verification is further reduced to a Szegő kernel. For any but fixed $b \in \mathbb{D}$,

$$\begin{aligned} & \left| \left\langle k_b, \frac{\left(\frac{\partial}{\partial \bar{a}}\right)^n k_a}{\left\| \left(\frac{\partial}{\partial \bar{a}}\right)^n k_a \right\|} \right\rangle \right| \\ &= \frac{1}{\left\| \left(\frac{\partial}{\partial \bar{a}}\right)^n k_a \right\|} \left\langle K_b, \left(\frac{\partial}{\partial \bar{a}}\right)^n k_a \right\rangle \\ &= \frac{1}{\left\| \left(\frac{\partial}{\partial \bar{a}}\right)^n k_a \right\|} \left| \left(\frac{\partial}{\partial \bar{a}}\right)^n k_a(b) \right| \\ &\leq \frac{(1 - |a|^2)^{n+\frac{1}{2}}}{\sqrt{(2n)!} |a|^n} \frac{n! |b|^n}{|(1 - \bar{a}b)|^{n+1}} \end{aligned}$$

(using the $m = 0$ term in (39))

$$\leq \frac{(1 - |a|^2)^{n+\frac{1}{2}}}{\sqrt{(2n)!}|a|^n} \frac{n!|b|^n}{(1 - |b|)^{n+1}} \rightarrow 0, \quad \text{as } |a| \rightarrow 1.$$

Lemma 2 and Lemma 3 together show that the complete dictionary satisfies BVC:

Theorem 5. For $f \in H^2(\mathbb{D})$ there holds

$$\lim_{|a| \rightarrow 1 \text{ or } n \rightarrow \infty} |\langle f, e_{n,a} \rangle| = 0. \quad (42)$$

In the last section, we deduced that in any Hilbert space with a BVC differentiate dictionary GA, OGA and POAFD can be performed. Owing to Theorem 5, taking \mathcal{H} to be the Hardy space in the disc, and $\tilde{D} = \{E_q\}, q \in \mathcal{E}$, the complete Szegő dictionary, where

$$\mathcal{E} = \{q \in \mathbf{N} \times \mathbf{D} = \{(n, a) : n \in \mathbf{N}, a \in \mathbf{D}\}$$

and

$$E_q = e_{n,a},$$

we conclude that GA, OGA, and POAFD can be performed with respect to the Szegő complete dictionary.

The general theory of POAFD especially implies

Theorem 6. For any $m - 1$ previously given distinguished 2-tuples

$$(n_1, a_1), \dots, (n_{m-1}, a_{m-1}),$$

each in $\mathbf{N} \times \mathbf{D}$, there holds

$$\lim_{|a| \rightarrow 1 \text{ or } n \rightarrow \infty} |\langle f, B_m^{k_{n,a}} \rangle| = 0, \quad (43)$$

where $B_m^{k_{n,a}}$ is the Hardy space function, unique up to unimodular multiple constants, characterized by the condition that

$$(B_1^{k_{n_1,a_1}}, \dots, B_{m-1}^{k_{n_{m-1},a_{m-1}}}, B_m^{k_{n,a}})$$

is the GS orthonormalization of $(B_1^{k_{n_1,a_1}}, \dots, B_{m-1}^{k_{n_{m-1},a_{m-1}}}, \tilde{k}_{n_m,a_m})$. We note that POAFD through the multiple kernel notion induces the completion of the dictionary in use. In our case, the dictionary in use by itself is a complete dictionary, whose completion, therefore, remains as just the same dictionary. As a consequence, there exists $(n_m, a_m) \in \mathbf{N} \times \mathbf{D}$ such that

$$(n_m, a_m) = \arg \max \{ |\langle f, B_m^{k_{n,a}} \rangle| \mid (n, a) \in \mathbf{N} \times \mathbf{D} \}. \quad (44)$$

The system $\{B_m^{k_{n,a}}\}$ has the role as the TM system made by the multiple Szegő kernels, where they can be further specified by

$$\begin{aligned} B_m^{k_{n,a}} &= \frac{Q_{(n_1,a_1), \dots, (n_{m-1}, a_{m-1})} k_{n,a}}{\|Q_{(n_1,a_1), \dots, (n_{m-1}, a_{m-1})} k_{n,a}\|} \\ &= \frac{k_{n,a} - \sum_{l=1}^{m-1} \langle k_{n,a}, B_l \rangle B_l}{\|k_{n,a} - \sum_{l=1}^{m-1} \langle k_{n,a}, B_l \rangle B_l\|} \\ &= \frac{e_{n,a} - \sum_{l=1}^{m-1} \langle e_{n,a}, B_l \rangle B_l}{\|e_{n,a} - \sum_{l=1}^{m-1} \langle e_{n,a}, B_l \rangle B_l\|}. \end{aligned}$$

5. Behavior of multiple Szegő kernel and Laguerre system in Hardy-Sobolev space

For $\sigma > 0$, denoted by $H_\sigma^2(\mathbf{D})$ the order- σ Hardy-Sobolev space:

$$H_\sigma^2(\mathbf{D}) = \{f(z) = \sum_{n=0}^{\infty} c_n z^n : \sum_{n=0}^{\infty} |(1+n^\sigma)c_n|^2 < \infty\}.$$

The space is briefly denoted H_σ^2 . For $f \in H_\sigma^2$, define

$$\|f\|_{H_\sigma^2} \triangleq \left(\sum_{n=0}^{\infty} |(1+n^\sigma)c_n|^2 \right)^{\frac{1}{2}}.$$

H_σ^2 is a subset of H^2 , being, by itself, a reproducing kernel Hilbert space. We will consider in this section expansions of $f \in H_\sigma^2$ by multiple Szegő kernels in the norm of H^2 . By definition, $f \in H_\sigma^2$ means that f has up to the σ -th derivatives (σ can be non-integer) of which all are of finite energy. Below we restrict ourselves to the integer σ cases.

Theorem 7. If $f \in H_\sigma^2$, where σ is a non-negative integer, then

$$|\langle f, \frac{k_{n,a}}{\|k_{n,a}\|} \rangle| \leq C(\sigma, a) \frac{1}{n^\sigma} \|f\|_{H_\sigma^2}, \quad (45)$$

where $C(\sigma, a)$ is a constant depending only on σ and a .

If $a = 0$, then $\frac{k_{n,a}}{\|k_{n,a}\|} = z^n$, and the theorem reduces to a known classical result. In the general multiple kernel case, the proof uses estimates of the kernel function.

Proof. For $f \in H_\sigma^2$,

$$\begin{aligned} & |\langle f, \frac{k_{n,a}}{\|k_{n,a}\|} \rangle| \\ &= \frac{1}{\|k_{n,a}\|} |\langle f, k_{n,a} \rangle| \\ &= \frac{1}{\|k_{n,a}\|} |\langle f^{(\sigma)}, k_{n-\sigma,a} \rangle| \\ &\leq \frac{\|k_{n-\sigma,a}\|}{\|k_{n,a}\|} \|f\|_{H_\sigma^2}. \end{aligned}$$

By $\left(\frac{\partial}{\partial \bar{a}}\right)^n k_a(a) = \frac{n!a^n}{(1-\bar{a}a)^{n+1}}$, we get

$$\begin{aligned} & \frac{\|k_{n-\sigma,a}\|^2}{\|k_{n,a}\|^2} \\ &= \frac{k_{n-\sigma,a}(a)}{k_{n,a}(a)} \\ &= \frac{\left(\frac{\partial}{\partial \bar{a}}\right)^{2(n-\sigma)} k_a(a)}{\left(\frac{\partial}{\partial \bar{a}}\right)^{2n} k_a(a)} \\ &= \frac{\frac{[2(n-\sigma)]!a^{2(n-\sigma)}}{(1-\bar{a}a)^{2(n-\sigma)+1}}}{\frac{(2n)!a^{2n}}{(1-\bar{a}a)^{2n+1}}} \\ &= \frac{(2n-2\sigma)!}{(2n)!} \left(\frac{1-\bar{a}a}{a}\right)^{2\sigma}. \end{aligned}$$

By Stirling's formula,

$$\begin{aligned} & \frac{(2n-2\sigma)!}{(2n)!} \left(\frac{1-\bar{a}a}{a}\right)^{2\sigma} \\ &= \frac{\sqrt{2\pi(2n-2\sigma)} \left(\frac{2n-2\sigma}{e}\right)^{2n-2\sigma}}{\sqrt{2\pi(2n)} \left(\frac{2n}{e}\right)^{2n}} \left(\frac{1-\bar{a}a}{a}\right)^{2\sigma} \\ &= \frac{1}{2^{2\sigma}} \frac{(n-\sigma)^{2n-2\sigma+\frac{1}{2}}}{n^{2n+\frac{1}{2}}} e^{2\sigma} \left(\frac{1-\bar{a}a}{a}\right)^{2\sigma} \\ &\leq \frac{1}{2^{2\sigma}} e^{2\sigma} \left(\frac{1-\bar{a}a}{a}\right)^{2\sigma} 3^{-\frac{\sigma}{n}(2n+\frac{1}{2})} \frac{1}{n^{2\sigma}}. \end{aligned}$$

Hence,

$$\begin{aligned}
& |\langle f, \frac{k_{n,a}}{\|k_{n,a}\|} \rangle| \\
& \leq \frac{1}{2^\sigma} e^\sigma \left(\frac{1-\bar{a}a}{a} \right)^\sigma 3^{-\frac{\sigma}{n}(n+\frac{1}{4})} \frac{1}{n^\sigma} \|f\|_{H_\sigma^2}
\end{aligned}$$

tending to zero with the rate $\frac{1}{n^\sigma}$.

When all the parameters a_n are identical to 0 the system of the normalized multiple kernels is a half of the Fourier orthonormal system. When all a_n are identical to some $a \neq 0$, then the multiple kernels $k_{n,a}$ form a complete but unorthogonal system. The GS orthonormalization of such multiple kernels forms the complete orthonormal Laguerre System

$$B_{n,a}(z) = \frac{\sqrt{1-|a|^2}}{1-\bar{a}z} \left(\frac{z-a}{1-\bar{a}z} \right)^{n-1}, \quad n=1, \dots$$

(see Appendix of [27]). A Laguerre System is a particular TM system in which all a_n are equal to a non-zero $a \in \mathbf{D}$.

It is a classical result that Fourier series of functions in the H_σ^2 have convergence rate $O(n^{-\sigma})$. This result is available for Laguerre Systems as well.

Theorem 8. For a function f in the Sobolev space H_σ^2 there holds

$$\|f - \sum_{l \geq n} \langle f, B_{l,a} \rangle B_{l,a}\| \leq C_a \frac{1}{n^\sigma} \|f\|_{H_\sigma^2}. \quad (46)$$

Proof. Performing change of variable $e^{i\theta} = \frac{e^{it}-a}{1-\bar{a}e^{it}}$, simple computation gives $dt = \frac{|1-\bar{a}e^{it}|^2}{1-|a|^2} d\theta$ and

$$\begin{aligned}
& \langle f, B_{n,a} \rangle \\
& = \frac{1}{2\pi} \int_0^{2\pi} f(e^{it}) \frac{\sqrt{1-|a|^2}}{1-\bar{a}e^{-it}} \left(\frac{e^{-it}-\bar{a}}{1-\bar{a}e^{-it}} \right)^{n-1} dt \\
& = \frac{1}{2\pi} \frac{1}{\sqrt{1-|a|^2}} \int_0^{2\pi} f(e^{it}) (1-\bar{a}e^{it}) e^{-i(n-1)\theta} d\theta \\
& = \frac{1}{2\pi} \frac{1}{\sqrt{1-|a|^2}} \int_0^{2\pi} f\left(\frac{e^{i\theta}+a}{1+\bar{a}e^{i\theta}}\right) (1-\bar{a}\frac{e^{i\theta}+a}{1+\bar{a}e^{i\theta}}) e^{-i(n-1)\theta} d\theta \\
& = \frac{1}{2\pi} \frac{1}{\sqrt{1-|a|^2}} \int_0^{2\pi} F_a(\theta) e^{-i(n-1)\theta} d\theta \\
& = \frac{1}{\sqrt{1-|a|^2}} \langle F_a, e_{n,0} \rangle,
\end{aligned}$$

where $F_a(\theta) = f\left(\frac{e^{i\theta}+a}{1+\bar{a}e^{i\theta}}\right) (1-\bar{a}\frac{e^{i\theta}+a}{1+\bar{a}e^{i\theta}})$. Since $|a|$ has a positive distance to 1, we have $\|F_a\|_{H_\sigma^2} \simeq \|f\|_{H_\sigma^2}$. Hence, through a brutal estimation,

$$\begin{aligned}
& \left\| \sum_{l \geq n} \langle f, B_{l,a} \rangle B_{l,a} \right\|^2 \\
& = \sum_{l \geq n} |\langle f, B_{l,a} \rangle|^2 \\
& = \frac{1}{\sqrt{1-|a|^2}} \sum_{l \geq n} |\langle F_a, e_{n,0} \rangle|^2 \\
& \leq \frac{1}{\sqrt{1-|a|^2}} \frac{1}{(1+n^\sigma)^2} \sum_{l \geq n} |(1+l^\sigma) \langle F_a, e_{n,0} \rangle|^2 \\
& = \frac{1}{\sqrt{1-|a|^2}} \frac{1}{(1+n^\sigma)^2} \|F_a\|_{H_\sigma^2}^2 \\
& \lesssim C_a \frac{1}{(1+n^\sigma)^2} \|f\|_{H_\sigma^2}^2.
\end{aligned}$$

Below we provide the explicit transformation matrices between the n -Laguerre system and the corresponding first n multiple kernels.

Proposition 1. For arbitrary but fixed n and $a \in \mathbb{D}$, $a \neq 0$, denote by the row matrix $B = \{B_{l,a}\}_{l=1}^n$ the n -Laguerre system, and the row matrix $\mathcal{K} = \{k_{l,a}\}_{l=1}^n$ the corresponding n -tuple of multiple kernels. Then the invertible transformation matrix \mathcal{T} such that $\mathcal{K} = \mathcal{T}B$ is given by $\mathcal{T} = \{c_{kj}\}_{n \times n}$,

$$c_{kj} = \begin{cases} \frac{(j+1)!}{(k-j)!} \frac{a^{k-j}}{(1-|a|^2)^{k+\frac{1}{2}}}, & j \leq k \leq n, \\ 0, & n \geq j > k, \end{cases}$$

and $B = \mathcal{T}^{-1}\mathcal{K}$ with $\mathcal{T}^{-1} = \{d_{kj}\}_{n \times n}$,

$$d_{kj} = \begin{cases} \frac{(-a)^{k-j}(1-|a|^2)^{j-\frac{1}{2}}}{k!(k-j)!}, & j \leq k \leq n, \\ 0, & n \geq j > k. \end{cases}$$

Proof. Since $\text{span } B = \text{span } \mathcal{K}$ and $B_{l,a}$ are orthonormal, we have $k_{k,a} = \sum_{j=1}^k c_{kj} B_{j,a}$, where $c_{kj} = \langle k_{k,a}, B_{j,a} \rangle$. For $j \leq k \leq n$,

$$\begin{aligned} \bar{c}_{kj} &= \langle B_{j,a}, k_{k,a} \rangle \\ &= \left\langle \frac{\sqrt{1-|a|^2}}{1-\bar{a}z} \left(\frac{z-a}{1-\bar{a}z} \right)^{j-1}, \frac{z^{k-1}}{(1-\bar{a}z)^k} \right\rangle \\ &= \frac{1}{2\pi} \int_0^{2\pi} \frac{\sqrt{1-|a|^2}}{1-\bar{a}e^{it}} \left(\frac{e^{it}-a}{1-\bar{a}e^{it}} \right)^{j-1} \frac{e^{-i(k-1)t}}{(1-\bar{a}e^{-it})^k} dt \\ &= \frac{1}{2\pi} \sqrt{1-|a|^2} \int_0^{2\pi} \frac{(e^{it}-a)^{j-1}}{(1-\bar{a}e^{it})^j} \frac{1}{(e^{it}-a)^k} de^{it} \\ &= \frac{1}{2\pi} \sqrt{1-|a|^2} \int_{\partial\mathbb{D}} \frac{(\xi-a)^{j-1}}{(1-\bar{a}\xi)^j} \frac{1}{(\xi-a)^k} d\xi \\ &= \frac{1}{2\pi} \sqrt{1-|a|^2} \int_{\partial\mathbb{D}} \frac{1}{(1-\bar{a}\xi)^j} \frac{1}{(\xi-a)^{k+1-j}} d\xi \\ &= \frac{j!}{(k-j)!} \frac{\bar{a}^{k-j}}{(1-|a|^2)^{k-\frac{1}{2}}}. \end{aligned}$$

For $n \geq j > k$, $B_{j,a} \perp \text{span } B = \text{span } \mathcal{K}$, and thus $B_{j,a} \perp k_{k,a}$. We hence have $c_{kj} = 0$, $j > k$.

To compute the entries d_{kj} of the inverse matrix \mathcal{T}^{-1} , by using mathematical induction, we have $d_{kj} = \frac{(-a)^{k-j}(1-|a|^2)^{j-\frac{1}{2}}}{k!(k-j)!}$, $j \leq k$, and $d_{kj} = 0$, $j > k$.

6. n -Best approximation with the complete Szegő kernel dictionary: existence and algorithm

Since derivatives of parameterized Szegő kernels and multiple Szegő kernels are still multiple Szegő kernels, the completion of the Szegő complete dictionary $\{e_{m,a}\}$, $m = 0, 1, 2, \dots$, $a \in \mathbb{D}$, is itself. Hence the n -Best problem with (1) returns to one with (1).

Theorem 9. In the setting of the Hardy H^2 space with the complete Szegő dictionary there exist solutions to the n -Best approximation problem.

It is noted that the n -Best approximation with the complete Szegő dictionary is different from that with the Szegő kernel dictionary. For the former, the total number of the involved multiple kernels is required not exceeding n , and for the latter, the sum of all the multiples of the involved parameters is required not exceeding n . The theorem cannot be proved by directly invoking the existing n -Best results [7,18,24]. We outline a proof by referring to the one in [7].

Proof. Let f be a fixed function in the Hardy space. We assume that f is not identical to any j -Best approximation for $j < n$, and hence exactly n parameters are necessary to get an n -Best approximation. In this case, suppose that we have a sequence of n -tuples of parameter pairs

$$\left((a_1^{(l)}, m_1^{(l)}), \dots, (a_n^{(l)}, m_n^{(l)}) \right), l = 1, 2, \dots,$$

such that the corresponding n -tuples of multiple kernels $k_{a_j^{(l)}, m_j^{(l)}}, j = 1, \dots, n$, give rise to, with the limit procedure $l \rightarrow \infty$, the infimum of (1). We claim that the integers $m_j^{(l)}$ for $j = 1, \dots, n$ and $j = 1, 2, \dots$, have to be bounded, and the complex numbers $a_j^{(l)}$ s for $j = 1, \dots, n$ and $j = 1, 2, \dots$, have to be in a compact disc contained in \mathbf{D} . Once these claims are proved, by invoking the Bolzano-Weierstrass Theorem on the existence of a convergent subsequence in the compact set we conclude the existence of n -Best approximation in our case.

The boundedness of $m_j^{(l)}$ is assured through the argument employed in the proof of Lemma 2. Next, by Proposition 1, each of the involved multiple kernels $k_{a_j^{(l)}, m_j^{(l)}}$ is a linear combination of the first $m_j^{(l)}$ functions of the corresponding Laguerre system. Then $(a_1^{(l)}, m_1^{(l)}), \dots, (a_n^{(l)}, m_n^{(l)})$ altogether induce n finite Laguerre systems of, respectively, the orders $m_1^{(l)}, \dots, m_n^{(l)}$. Put the n finite Laguerre systems together and construct the equivalent orthonormal system by using the GS process. The obtained orthonormal system is a finite TM system. The argument of the proof of the main result of [7] can be adopted to obtain that any $|a_j^{(l)}|$ tending to 1 along with $l \rightarrow \infty$ will result in that the sequence $k_{a_j^{(l)}, m_j^{(l)}}$ having no contribution in the approximation, and thus may be deleted. This concludes n parameters are unnecessary. This is contradictory to the assumption at the beginning of the proof.

Remark 3. Since the Fourier system is contained in the complete Szegő dictionary, the n -Best approximation to functions in H_σ^2 by the complete dictionary has a convergence rate at least as good as $O(n^{-\sigma})$. Taking into account the convergence rate $O(n^{-1/2})$ for general matching pursuit algorithms, it is natural to guess, but we are unable to prove so far, that the n -Best approximation in H_σ^2 by the complete dictionary has the convergent rate $O(n^{-\sigma-1/2})$.

We hereby cite two types of algorithms for finding one or all the n -Best solutions.

6.1. A global and theoretical algorithm for finding all the n -Best solutions

By a *global method*, we refer to finding all the global minimum solutions, and, in particular, not being trapped in the local minimums. Such methods, therefore, have to be theoretical. Since, as asserted, the $m_j^{(l)}$ are altogether bounded, say, by N , and all $a_j^{(l)}$ are in a compact disc $\overline{\mathbf{D}}_r, 0 < r < 1$, we reduce the problem to finding $(k_{m_1, a_1}, \dots, k_{m_n, a_n})$ such that $(m_j, a_j) \in \{1, \dots, N\} \times \overline{\mathbf{D}}_r$. This is to find all the global minimizers of the Lipschitz target function defined in the compact set. There exist theoretical, as well as practical methods to solve the type of problems in the optimization specialty. See, for instance, [24,31], and the references thereby.

6.2. A practical algorithm for finding an n -Best solution

Practically, by the Gaussian gradient type or other similar methods based on local comparison of the target function values one may find an n -Best solution. By local comparison, although there is no guarantee of finding a global solution, it is practical and can often be used [20–23]. We hereby recommend a local comparison type method, *Cyclic POAFD*, with theoretical clarity and computational simplicity. It can be used with any dictionary [22].

Suppose we have an initial n -tuple of dictionary elements $(e_{q_1^{(0)}}, \dots, e_{q_n^{(0)}})$. The n -tuple of parameters $(q_1^{(0)}, \dots, q_n^{(0)})$ may be obtained randomly, or through an n -POAFD procedure. We rather use the kernels $k_{q_j^{(0)}}$ (pre-dictionary) instead of using the dictionary elements $e_{q_j^{(0)}}$ to save the computation of normalization. The idea is to improve the already obtained n -tuple of parameters in one by one cyclic manner: At each iteration step we replace a kernel already existing in the n tuple of kernels, and keep the other $n - 1$ kernels unchanged. This cyclic procedure is performed continuously until each of the n -tuple of parameters cannot be improved with respect to a tolerant error threshold.

7. Experiments

The experiments will select parameters in \mathbf{D} which are expressed in polar coordinates:

$$a = \frac{x-1}{X} e^{i2\pi \frac{y-1}{Y}}, \quad (49)$$

for $x = 1, 2, \dots, X$, and $y = 1, 2, \dots, Y$, where X, Y correspond to the radial and angular discretization of the unit disc, respectively. We let $X = 100, Y = 100$ in the following experiments. The abbreviations of the algorithms with the specifications including the type of dictionary in use, the type of matching pursuit algorithms, the iteration times, and whether it is n -Best etc. are given in Table 1. For Complete POAFD as an example, we obtain the partial sum series:

$$S_K = \sum_{k=1}^K \langle g_k, B_k \rangle B_k, \quad (50)$$

with the relative error given by

Table 1
Abbreviations of the algorithms.

p-GA	p iterations of GA algorithm with the Szegő dictionary
p-OGA	p iterations of OGA algorithm with the Szegő dictionary
p-POAFD	p iterations of POAFD algorithm with the Szegő dictionary
p-Best	p-Best approximation with the Szegő dictionary through some cycles
p-Complete GA	p iterations of GA algorithm with the complete Szegő dictionary
p-Complete OGA	p iterations of OGA algorithm with the complete Szegő dictionary
p-Complete POAFD	p iterations of POAFD algorithm with the complete Szegő dictionary
p-Best Complete	p-Best approximation with the complete Szegő dictionary through some cycles
p-unwinding	p iterations of unwinding algorithm

Table 2
Parameters.

Experiment 1		
k	b_k	c_k
1	-0.4750 + 0.3050i	-0.5861 - 0.04445i
2	-0.1800 + 0.7150i	0.2428 - 0.6878i
3	0.2600 - 0.7300i	0.4423 - 0.3309i
4	0.5400 + 0.3600i	-0.2703 - 0.8217i
5	-0.4850 - 0.2150i	-0.8085 + 0.3774i

Table 3
Relative error.

Experiment 1	Fig. 1.			
3 iterations	GA	OGA	POAFD	3-Best
relative error	0.4581	0.4314	0.4246	0.2613

$$\frac{\|f - S_K\|}{\|f\|}. \quad (51)$$

We take a similar definition to get the relative energy and relative error of other methods. Below we use the following convention and notation: If for a particular signal f the relative approximation error (51) of Algorithm 1 is less than that of Algorithm 2, then we denote this fact by

$$\text{Algorithm1}(f) \geq \text{Algorithm2}(f).$$

The simulations are performed in a MATLAB environment using the data collected from the formula of the signal $f(z)$, $z = e^{it}$. The samples of f are from $t_j = \frac{2\pi(j-1)}{100}$, $j = 1, 2, \dots, 100$. Six experiments are included.

Experiment 1. This experiment contains two lots of comparisons of which one is 3 iterations of, respectively, GA, OGA, POAFD, and n-Best algorithm, with respect to the Szegő kernel dictionary; and the other is the same but with respect to the complete Szegő dictionary. The tested toy function f_1 is given by the finite Blaschke form

$$f_1(z) = \sum_{k=1}^5 c_k B_{\{b_1, \dots, b_k\}}(e^{it}), \quad (52)$$

where $B_{\{b_1, \dots, b_k\}}$ is the TM system

$$B_{\{b_1, \dots, b_k\}} = \frac{\sqrt{1 - |a_k|^2}}{1 - \bar{a}_k z} \prod_{l=1}^{k-1} \frac{z - a_l}{1 - \bar{a}_l z}, \quad a \in \mathbb{D}, \quad k = 1, 2, \dots,$$

and $t \in (0, 2\pi)$. The parameters (b_k, c_k) , $k = 1, 2, \dots, 5$ are given in Table 2.

We note that 3-Best by using POAFD with the Szegő dictionary to create the initial 3-tuple, and 3-Best Complete by using POAFD with the complete Szegő dictionary to create the initial 3-tuple.

According to the relative error given in Table 3, and the results in Fig. 1, we conclude that $n\text{-Best}(f_1) \geq \text{POAFD}(f_1) \geq \text{OGA}(f_1) \geq \text{GA}(f_1)$.

From Table 4, and the results in Fig. 2 we see that $n\text{-Best Complete}(f_1) \geq \text{Complete POAFD}(f_1) \geq \text{Complete OGA}(f_1) \geq \text{Complete GA}(f_1)$. The order of superiority of the algorithms remains unchanged when using the complete dictionary, but by using the complete dictionary the approximation is more accurate.

Comparing POAFD and Complete GA algorithm, the results in Fig. 3, and the relative error given in Table 3, 4, we conclude that $\text{Complete GA}(f_1) \geq \text{POAFD}(f_1)$.

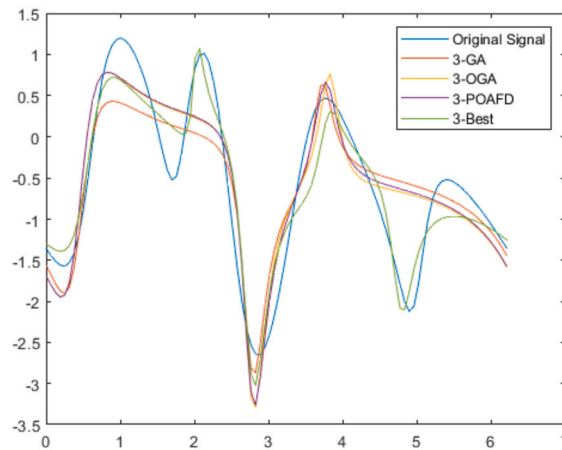


Fig. 1. 3 iterations with the Szegő dictionary.

Table 4
Relative error.

Experiment 1	Fig. 2.			
3 iterations relative error	Complete GA	Complete OGA	Complete POAFD	3-Best Complete
	0.2941	0.2733	0.2590	0.0671

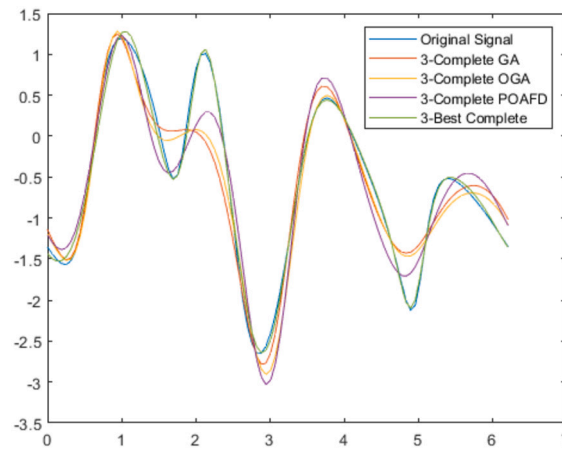


Fig. 2. 3 iterations with the complete Szegő dictionary.

Experiment 2. This experiment is to evaluate the performance of POAFD with the two dictionaries: the Szegő one and the complete Szegő one. The toy function is still (52) but with the parameters (b_k, c_k) , $k = 1, 2, \dots, 5$ given in Table 5 that we note it f_2 .

From Table 6, and the results in Fig. 4, 5 we know that Complete POAFD can give a good approximation to f through 9 iterations, while POAFD gives an approximation at a similar level through 18 iterations. Hence, $\text{Complete POAFD}(f_2) \geq \text{POAFD}(f_2)$.

Experiment 3. This experiment is to evaluate the performance of n -Best with the two dictionaries: the Szegő one and the complete Szegő one. The toy function is still (52) but with the parameters (b_k, c_k) , $k = 1, 2, \dots, 4$ given in Table 7 that we note it f_3 .

We note that 8-Best by using POAFD with the Szegő dictionary to create the initial 8-tuple, and 4-Best Complete by using POAFD with the complete Szegő dictionary to create the initial 4-tuple.

From Table 8, and the results in Fig. 6, 7 we know that 4-Best Complete POAFD can give a good approximation to f by 3 cycles, while 8-Best POAFD gives an approximation at a similar level by 5 cycles. Hence, $n\text{-Best Complete}(f_3) \geq n\text{-Best}(f_3)$.

Experiment 4. We note that unwinding Blaschke expansion is not a matching pursuit type algorithm, it, however, is of the same nature. As a very effective signal analysis method we add it to the pool of comparison [32]. This experiment is to compare unwinding, Complete POAFD, and n -Best algorithm, the signal is given by the samples of the following function

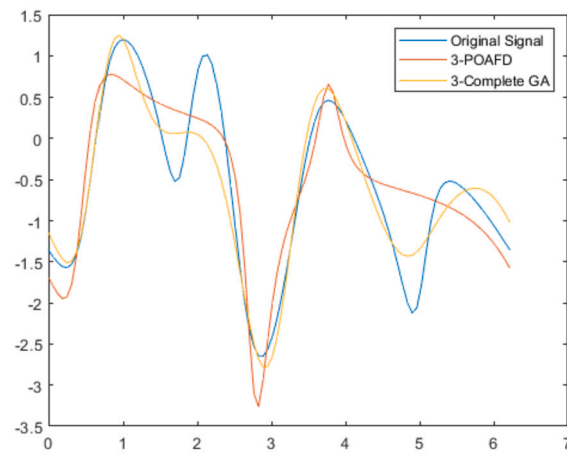


Fig. 3. 3 iterations of POAFD and Complete GA.

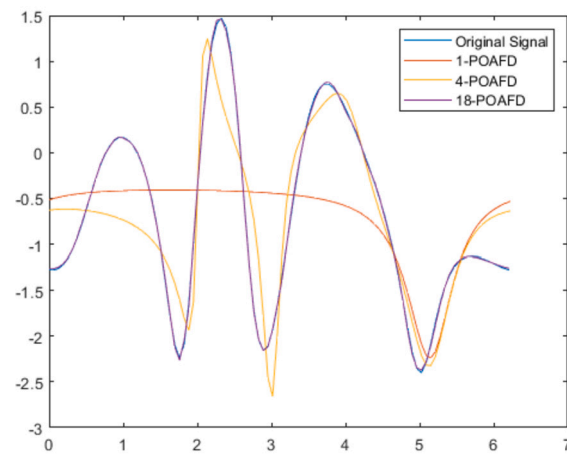


Fig. 4. POAFD iterations with the Szegő dictionary.

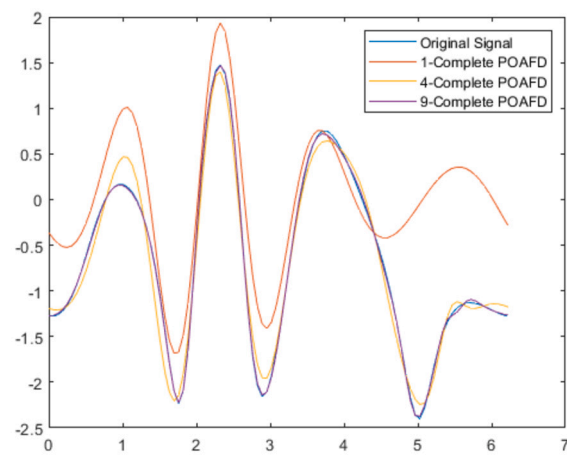


Fig. 5. POAFD iterations with the complete Szegő dictionary.

Table 5
Parameters.

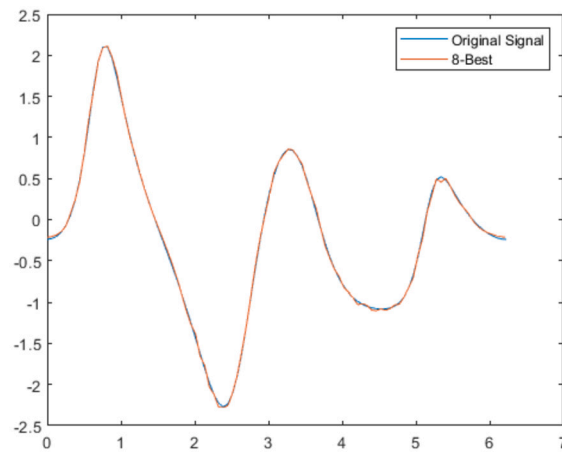
Experiment 2		
k	b_k	c_k
1	-0.5850+0.2930i	-0.3861-0.0515i
2	0.4806+0.2513i	-0.2802-0.7235i
3	0.2505-0.6823i	0.4505-0.4325i
4	-0.2005+0.6950i	0.2539-0.7136i
5	-0.4512-0.1825i	-0.7562+0.4265i

Table 6
Relative error.

Experiment 2						
Fig. 4. / Fig. 5.	POAFD			Complete POAFD		
K iterations	$K=1$	$K=4$	$K=18$	$K=1$	$K=4$	$K=9$
relative error	0.7906	0.4259	0.0173	0.7306	0.1334	0.0173

Table 7
Parameters.

Experiment 3		
k	b_k	c_k
1	-0.4750+0.3050i	-0.5861-0.4444i
2	0.3600-0.6300i	0.4423-0.3308i
3	0.5400+0.4600i	-0.2702-0.8217i
4	-0.4850-0.2150i	-0.7085+0.3773i

**Fig. 6.** 8-Best approximation with the Szegő dictionary through 5 cycles.**Table 8**
Relative error.

Experiment 3		
Fig. 6. / Fig. 7.	8-Best POAFD	4-Best Complete POAFD
K iterations	$K=8$	$K=4$
l cycles	$l=5$	$l=3$
relative error	0.0224	0.0200

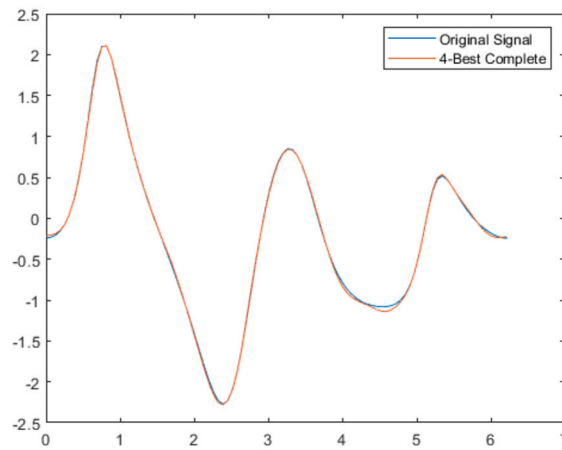


Fig. 7. 4-Best approximation with the complete Szegő dictionary through 3 cycles.

Table 9
Parameters.

Experiment 4	
k	d_k
1	3.1017-2.5305i
2	-6.1205+2.3674i
3	-5.4678-2.2502i
4	-4.4217+7.6913i

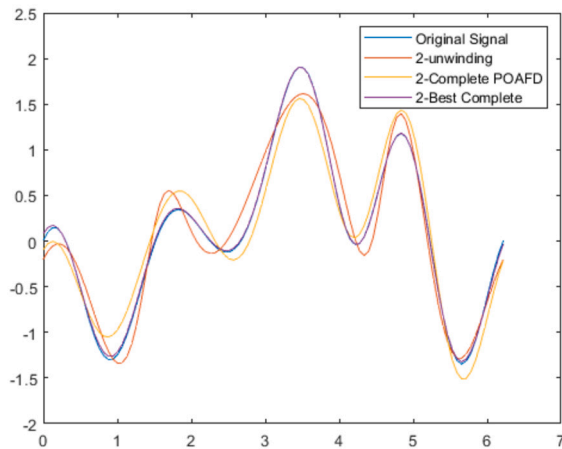


Fig. 8. 2 iterations of unwinding, Complete POAFD and 2-Best Complete.

Table 10
Relative error.

Experiment 4	Fig. 8.		
2 iterations	unwinding	Complete POAFD	2-Best Complete
relative error	0.2113	0.0863	0.0082

$$f_4(z) = \frac{(z^4 - d_1)(d_2 - z)^5}{(d_3 - z)^3(d_4 - z)^2},$$

where $t \in (0, 2\pi)$. The parameters d_k , $k = 1, 2, \dots, 4$ are given in Table 9.

We note that 2-Best Complete by using POAFD with the complete Szegő dictionary to create the initial 2-tuple.

From Table 10, and the results in Fig. 8, it is seen that $n\text{-BestComplete}(f_4) \geq \text{Complete POAFD}(f_4) \geq \text{unwinding}(f_4)$.

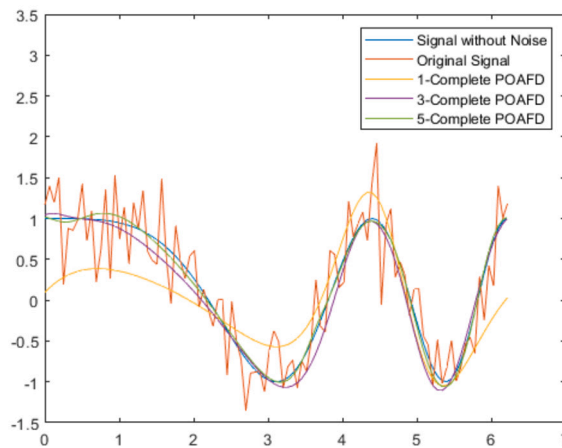


Fig. 9. Denoising the chirp signal by Complete POAFD.

Experiment 5. Using Complete POAFD algorithm to denoise a noisy signal, the data are collected from the chirp signal without noise

$$f(z) = e^{i\frac{t^2}{\pi}},$$

where $t \in (0, 2\pi)$. And the original signal is a noisy signal with additive Gaussian white noise of f .

From the denoising effect given in Fig. 9, we see Complete POAFD may denoise noisy signals.

To summarize the experiments:

1. GA, OGA, POAFD (=AFD in HARDY), n -BEST with either the Szegő or the complete Szegő dictionary, and unwinding have the signal reconstruction efficiency following this order: n -Best Complete(f_1) \geq Complete POAFD(f_1) \geq Complete OGA(f_1) \geq Complete GA(f_1) \geq POAFD(f_1) \geq OGA(f_1) \geq GA(f_1), n -Best Complete(f_3) \geq n -Best(f_3), and n -Best Complete(f_4) \geq Complete POAFD(f_4) \geq unwinding(f_4).
2. In spite of the ordering theoretically or experimentally proved, referring to 1., experiments show that the differences between their efficiencies are not much.
3. Experiments show that Complete GA(f_1) \geq POAFD(f_1). A weaker algorithm with the complete dictionary is usually stronger than a stronger algorithm with a weaker dictionary, say the original Szegő dictionary. This shows that it is the dictionary that is important.
4. Complete POAFD algorithm shows promising denoise effect.

8. Conclusion

The necessity of the notion multiple kernels is justified by the existence of the natural question on n -Best kernel approximation, and by the desire for principal frequency decomposition of signals. For the latter, the concept mean-frequency is introduced. As the Fourier basis functions do, multiple kernels can extract the same rates of decaying to zero of signals in the Sobolev spaces. The work spells out the natural connections between the multiple Szegő kernels and the Laguerre system. Well-posed-ness of n -Best approximation with the complete Szegő dictionary is proved. With any dictionary, n -Best is the strongest, and POAFD is superior to all the other concerned matching pursuit methods in the one by one manner. Precisely, it is proved that the concerned matching pursuit methods from strong to weak are in the order n -Best, POAFD, OGA, and GA. Through concrete examples, we show that the complete Szegő dictionary has great potential in sparse representation, for, especially, the weakest matching pursuit GA combined with the complete dictionary can inspire greater efficiency than what POAFD with the Szegő dictionary does.

Future works include reducing the computational complexity and increasing the speed of the algorithm process.

Data availability

The data that has been used is confidential.

Acknowledgements

This work was supported by the research fund of Macao Government Research Funding (grant no. FDCT0128/2022/A). The author wishes to express her sincere thankfulness to professor Tao Qian for his interest and encouragement to study this topic and useful comments on the draft material of this article. The authors would like to thank the referees for their helpful suggestions and comments.

References

- [1] S. Mallat, Z. Zhang, Matching pursuits with time-frequency dictionaries, *IEEE Signal Proces.* 41 (12) (1993) 3397–3415.
- [2] V. Temlyakov, *Greedy Approximation*, Cambridge University Press, 2011, p. 20.
- [3] T. Qian, Y.B. Wang, Adaptive Fourier series-a variation of greedy algorithm, *Adv. Comput. Math.* 34 (3) (2011) 279–293.
- [4] Q.-H. Chen, T. Qian, L.-H. Tan, A theory on non-constant frequency decompositions and applications, in: *Advancements in Complex Analysis*, Springer, 2020, pp. 1–37.
- [5] T. Qian, Adaptive Fourier decompositions and rational approximations, Part I: Theory, *Int. J. Wavelets Multiresolut. Inf. Process.* 12 (5) (2014) 1461008.
- [6] T. Qian, Two-dimensional adaptive Fourier decomposition, *Math. Methods Appl. Sci.* 39 (10) (2016) 2431–2448.
- [7] Y.B. Wang, T.T. Qian, Pseudohyperbolic distance and n -Best rational approximation in H^2 space, *Math. Methods Appl. Sci.* 44 (11) (2021) 8497–8504.
- [8] T. Qian, Intrinsic mode-component decomposition of functions: an advance of Fourier theory, *Math. Methods Appl. Sci.* 33 (7) (2010) 880–891.
- [9] R. Coifman, S. Steinerberger, Nonlinear phase unwinding of functions, *J. Fourier Anal. Appl.* 23 (4) (2017) 778–809.
- [10] R. Coifman, J. Peyrière, Phase unwinding, or invariant subspace decompositions of Hardy spaces, *J. Fourier Anal. Appl.* 25 (2019) 684–695.
- [11] D. Alpay, F. Colombo, T. Qian, I. Sabadini, Adaptive orthonormal system for matrix-valued functions, *Proc. Am. Math. Soc.* 145 (5) (2017) 2089–2106.
- [12] D. Alpay, F. Colombo, T. Qian, I. Sabadini, Adaptive decomposition: the case Drury-Arveson space, *J. Fourier Anal. Appl.* 23 (6) (2017) 1426–1444.
- [13] J.B. Garnett, *Bounded Analytic Functions*, Academic Press, New York, 1987.
- [14] W. Rudin, *Real and Complex Analysis*, McGraw Hill International Book Company, 1974.
- [15] T. Qian, W. Sproessig, J.X. Wang, Adaptive Fourier decomposition of functions in quaternionic Hardy spaces, *Math. Methods Appl. Sci.* 35 (1) (2012) 43–C64.
- [16] W. Qu, P. Dang, Reproducing kernel approximation in weighted Bergman spaces: algorithm and applications, *Math. Methods Appl. Sci.* 42 (12) (2019) 4292–4304.
- [17] T. Qian, *Adaptive Fourier Transform* (in Chinese), The Chinese Science Press, 2015.
- [18] W. Qu, T. Qian, H.C. Li, K. Zhu, Best kernel approximation in Bergman spaces, *Appl. Math. Comput.* 416 (2022) 126749.
- [19] J.L. Walsh, *Interpolation and Approximation by Rational Functions in the Complex Domain*, American Mathematical Society Publication, vol. 10, 1962.
- [20] L. Baratchart, Existence and generic properties of L^2 approximation for linear systems, *Math. Control Inf.* 3 (1986) 89–101.
- [21] L. Baratchart, M. Cardelli, M. Olivi, Identification and rational L^2 approximation: a gradient algorithm, *Automatica* 27 (2) (1991) 413–417.
- [22] T. Qian, Cyclic AFD algorithm for the best rational approximation, *Math. Methods Appl. Sci.* 37 (6) (2014) 846–859.
- [23] T. Qian, J. Wang, W. Mai, An enhancement algorithm for cyclic adaptive Fourier decomposition, *Appl. Comput. Harmon. Anal.* 47 (2) (2019) 516–525.
- [24] T. Qian, *N-Best Kernel Approximation in Reproducing Kernel Hilbert Spaces*, Available at SSRN 4090819.
- [25] A. Borowicz, Improving energy compaction of adaptive Fourier decomposition, in: *2020 28th European Signal Processing Conference (EUSIPCO)*, IEEE, 2021, pp. 2348–2352.
- [26] T. Qian, Boundary derivatives of the phases of inner and outer functions and applications, *Math. Methods Appl. Sci.* 32 (2009) 253–263.
- [27] T. Qian, Sparse representations of random signals, *Math. Methods Appl. Sci.* 45 (8) (2022) 4210–4230.
- [28] W. Mai, T. Qian, The Fourier type expansions on tubes, *Complex Var. Elliptic Equ.* 67 (2) (2022) 433–461.
- [29] R.A. DeVore, V.N. Temlyakov, Some remarks on greedy algorithms, *Adv. Comput. Math.* 5 (1996) 173–187.
- [30] G.J. Lord, C.E. Powell, T. Shardlow, *An Introduction to Computational Stochastic PDEs*, Cambridge University Press, 2014.
- [31] T. Qian, L. Dai, L.-M. Zhang, Z.-H. Chen, Granular sieving algorithm for selecting best n parameters, *Math. Methods Appl. Sci.* 613 (October 2022) 288–308.
- [32] T. Qian, H. Li, M. Stessins, Comparison of adaptive mono-component decompositions, *Nonlinear Anal., Real World Appl.* 14 (2) (2013) 1055–C1074.