Adv Comput Math (2011) 34:279–293
DOI 10.1007/s10444-010-9153-4

Adaptive Fourier series—a variation of greedy algorithm

Tao Qian · Yan-Bo Wang

Received: 5 September 2009 / Accepted: 6 April 2010 /
Published online: 8 June 2010

Abstract We study decomposition of functions in the Hardy space $H^2(\mathbb{D})$ into linear combinations of the basic functions (modified Blaschke products) in the system

$$B_n(z) = \frac{\sqrt{1-|a_n|^2}}{1-\overline{a}_n z}\prod_{k=1}^{n-1}\frac{z-a_k}{1-\overline{a}_k z}, \quad n = 1, 2, ..., \tag{1}$$

where the points a_n's in the unit disc $\mathbb{D}$ are adaptively chosen in relation to the function to be decomposed. The chosen points a_n's do not necessarily satisfy the usually assumed hyperbolic non-separability condition

$$\sum_{k=1}^{\infty}(1-|a_k|) = \infty \tag{2}$$

in the traditional studies of the system. Under the proposed procedure functions are decomposed into their intrinsic components of successively increasing non-negative analytic instantaneous frequencies, whilst fast convergence is resumed. The algorithm is considered as a variation and realization of greedy algorithm.

Communicated by Yuesheng Xu.

The work was supported by Macao FDCT 014/2008/A1 and research grant of the University of Macau No. RG-UL/07-08s/Y1/QT/FSTR.

T. Qian (✉) · Y.-B. Wang
Department of Mathematics, Faculty of Science and Technology,
University of Macau, Taipa, Macao, China
e-mail: fsttq@umac.mo

Y.-B. Wang
e-mail: ya77405@umac.mo

Keywords Rational orthonormal system · Blaschke product · Complex hardy space · Analytic signal · Instantaneous frequency · Mono-components · Adaptive decomposition of functions · Greedy algorithm

Mathematics Subject Classifications (2010) 42A50 · 32A30 · 32A35 · 46J15

1 Introduction

The rational orthonormal system (1) is often referred as Takenaka-Malmquist (TM) system. It is a generalization of the Fourier system $\{z^{n-1}\}_{n=1}^{\infty}$. Besides the trigonometric basis for the Hardy $H^2(\mathbb{D})$ space, Laguerre basis and the "two-parameter Kautz" basis are other examples of (1). TM systems have long been interested, from early last century to present, with fruitful theoretical results and ample applications in a number of areas of applied mathematics, including control theory, signal processing and system identification [1–3, 9, 11, 19]. In the large quantity of the literature all the studies are based on the condition (2). Under this condition a system (1) is dense in the Hardy spaces $H^p(\mathbb{D})$, $1 \leq p < \infty$, and in the disc algebra $\mathcal{A}(\mathbb{D})$. On the other hand, if (1) is dense in any of the above mentioned Banach spaces, then (2) holds and, therefore, it is dense in all the other mentioned Banach spaces. There is also a counterpart theory together with a counterpart condition to (2) in the upper-half-complex plane. The present study is different from all the previous ones, for we do not assume the condition (2), and the system (1) is not expected to be complete in any of the above mentioned Banach spaces. This different setting is based on choosing the points a_n's defining the system according to the function f to be decomposed. The purpose is to decompose, perhaps, only the given function f in an effective way into basic functions of which each possesses non-negative analytic instantaneous frequency. A system (1) constructed in such a way is called an *adaptive rational orthonormal system*, where the condition (2) does not necessarily hold, and hence the adaptive system may not be complete in $H^2(\mathbb{D})$. From now on when we talk about the system (1) we do not assume the condition (2). In an adaptive decomposition the modified Blaschke products are regarded as the intrinsic components of the given function f for three reasons, of which one is non-negativity of the instantaneous frequencies (See the definitions given below for mono- and pre-mono-components), the second is fast convergence, and the third is gaining successively higher and higher frequencies. As in the adaptive wavelets case this type of decomposition has useful applications in engineering practice, including in time-frequency analysis and computation of Hilbert transforms.

This study is merged from our recent study on signal decomposition into mono-components [12, 13, 15, 18]. A real-valued signal is called a *real mono-component*, if the phase function $\theta(t)$ obtained from the natural amplitude-phase representation of its *associated analytic signal*, viz. $s(t) + i\mathcal{H}s(t) = \rho(t)e^{i\theta(t)}$, satisfies the condition $\theta'(t) \geq 0$, a.e. In physics terminology a mono-component is a signal that possesses a non-negative analytic instantaneous frequency function. The associated analytic signal, in the case, is called a *complex mono-component*. Without introducing confusion, both real and complex mono-components are sometimes called briefly mono-components. It can be easily verified that the boundary values of the modified Blaschke products B_n, $n = 1, 2, ...$, are all mono-components if $a_1 = 0$. In fact, in the case, they are bounded analytic functions (as analytic signals) with the expressions

$$B_1(z) = 1, \ B_n(z) = \frac{\sqrt{1-|a_n|^2}z}{1-\overline{a}_n z}\prod_{k=2}^{n-1}\frac{z-a_k}{1-\overline{a}_k z}, \quad n = 2, 3, ...,$$

where the Blaschke product parts are of positive phase derivatives on the boundary (as Möbius transforms are, see, for instance, [5]) and the fractional linear function

$$\frac{\sqrt{1-|a_n|^2}z}{1-\overline{a}_n z}$$

is a starlike function, and in particular, a convex function, which is also of positive phase derivative on the boundary [12]. In general the members of (1) belong to the category of *pre-mono-components* which, by definition, are those becoming mono-components after being multiplied by e^{iMt}, or, in other words, becoming mono-components riding on a carrier frequency e^{iMt}, for some $M > 0$. The members in (1) are also regarded as *weighted (finite) Blaschke products* in relation to Bedrosian identity [15, 16, 18, 21] that is to find weighted unimodular mono-components from unimodular mono-components. This Bedrosian identity procedure, to the authors' knowledge, starts from Y.S. Xu. A large pool of mono-components have been found so far, including boundary values of inner functions [13], weighted forms of some unimodular mono-components [15, 18] and p-starlike functions [12]. The study of this paper belongs to the trend of finding adaptive decompositions of functions into their intrinsic mono-components (or pre-mono-components) of various types. Temptations of adaptive decomposition using mono-components are noted in literature including [14] and [20].

To the authors, the motivation of the study is the engineering algorithm called EMD (Empirical Mode Decomposition) and the related signal decomposition. The algorithm produces certain basic signals called IMFs (Intrinsic Mode Functions) that are experimentally dependent. EMD at each phase of its multiple sifting processes for producing a single IMF throws away, under a given threshold, part of the signal of unknown structure, and it is highly local. It, however, unrealistically expects the resulted IMFs being of a highly global property, viz. being mono-components. The proposed adaptive decomposition is a certain replacement of EMD: We take what is desired as our starting point, then the algorithm turns to be a completely different one.

The proposed algorithm may be treated as a variation of greedy algorithm [4, 8]. In a greedy algorithm a *dictionary*, $\mathcal{D}$, is given, that, by definition, is a linearly dense subset consisting of certain unit elements of the underlying

Hilbert space $\mathcal{H}$. Besides $\mathcal{D}$ and $\mathcal{H}$, an element $f \in \mathcal{H}$ is given, and the purpose is to select $u_1, ..., u_n, ... \in \mathcal{D}$ so that

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

converges in a fast way. To guarantee the fast convergence, at every selection step it requires, for a fixed $\alpha \in (0, 1]$,

$$|\langle f_k, u_k \rangle| \geq \alpha \sup\{|\langle f_k, u \rangle| \ : \ u \in \mathcal{D}\}, \tag{3}$$

where

$$f_k = f - \sum_{l=1}^{k-1} \langle f_l, u_l \rangle u_l.$$

In our algorithm, the Hilbert space is the complex Hardy space $H^2(\mathbb{D})$, the dictionary,

$$\mathcal{D} = \{e_{\{a\}} := \frac{\sqrt{1-|a|^2}}{1-\bar{a}z} \ : \ a \in \mathbb{D}\},$$

consists of essentially the Cauchy kernels with poles outside the compact unit disc, and we have the relations

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

where e_k's are from the dictionary, B_k's are from the adaptive rational orthonormal system (1), f_k's are the usual remainders defined by

$$f_k =: f - \sum_{l=1}^{k-1} \langle g_l, e_l \rangle B_l = f - \sum_{l=1}^{k-1} \langle f_l, B_l \rangle B_l,$$

and the functions g_k's satisfy $\langle g_k, e_k \rangle = \langle f_k, B_k \rangle$. The algorithm can not be regarded as a typical greedy algorithm for the following two reasons. The first is that the functions B_k's to give rise to a linear expansion of f are not independent to each other: the formers are constructive factors in a certain manner of the latter ones. The second is that the function that at the k-th selection step dealing with the dictionary is g_k but not the usual remainder f_k. Our algorithm, however, can be regarded as a realizable variation of greedy algorithm in the sense that we can prove the existence of $e_k \in \mathcal{D}$ such that

$$|\langle g_k, e_k \rangle| = \sup\{|\langle g_k, e \rangle| \ : \ e \in \mathcal{D}\}.$$

This is to be compared with the weaker principle given in (3).

2 The adaptive rational orthonormal system

In this section we describe the adaptive program based on the maximal selection principle given in Lemma 2.1, and prove the convergence. Throughout the paper we assume that the given function $\tilde{f} \in \mathcal{L}^2(\partial\mathbb{D})$ to be decomposed is real-valued. It can be written into its Fourier series expansion

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

where the limit takes the $\mathcal{L}^2(\partial\mathbb{D})$ sense and $\sum_{k=-\infty}^{\infty} |c_k|^2 = \|\tilde{f}\|^2$. The square-norm is defined through the inner product

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

Write

$$f^+(e^{it}) = \sum_{k=0}^{\infty} c_k e^{ikt}, \quad f^-(e^{it}) = \sum_{k=-\infty}^{-1} c_k e^{ikt},$$

which are non-tangential boundary values of, respectively,

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

The last two functions are, respectively, in $H^2(\mathbb{D})$ and $H^2(\mathbb{C}\setminus\overline{\mathbb{D}})$. Since $\tilde{f}$ is real-valued, we have $c_{-k} = \overline{c}_k$, and hence

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

The approximation of $\tilde{f}$ is thus reduced to approximation of f^+.

The rest of this section will be devoted to approximation of $f = f^+ \in H^2(\mathbb{D})$ by using a series (1) formulated through a sequence of points $\{a_k\} \subset \mathbb{D}$ adaptively chosen according to f. The procedure is in the spirit of greedy algorithm [4, 8]. Below we also write $B_n = B_{\{a_1,\ldots,a_n\}}$. For $f \in H^2(\mathbb{D})$, by invoking the Cauchy Integral Formula, we have

$$\begin{aligned}\langle f, B_{\{a\}}\rangle &= \frac{\sqrt{1-|a|^2}}{2\pi}\int_0^{2\pi} f(e^{it})\overline{\frac{1}{1-\overline{a}e^{it}}}dt \\ &= \sqrt{1-|a|^2}\frac{1}{2\pi i}\int_{\partial\mathbb{D}} f(\zeta)\frac{1}{\zeta-a}d\zeta \\ &= \sqrt{1-|a|^2}\, f(a). \end{aligned} \tag{4}$$

To stress on the effect that the inner product $\langle f, B_{\{a\}}\rangle$ is essentially the evaluation of f at the point a, we give the particular notation $B_{\{a\}} = e_{\{a\}}$. Note that

$$\{e_{\{a\}} \ : \ a \in \mathbb{D}\}$$

is the dictionary introduced in the last section, where $e_a(z) = \frac{\sqrt{1-|a|^2}}{1-\bar{a}z}$ is essentially a shifted Cauchy kernel.

We use notations consistent with those used in Section 1 and set $f = f_1 = g_1$. The first step of the decomposition is

$$\begin{aligned} f_1(z) &= \big(f_1(z) - \langle f_1, e_{\{a_1\}}\rangle e_{\{a_1\}}(z)\big) + \langle g_1, e_{\{a_1\}}\rangle e_{\{a_1\}}(z) \\ &= f_2(z) + \langle g_1, e_{\{a_1\}}\rangle B_{\{a_1\}}(z) \\ &= g_2(z)\frac{z-a_1}{1-\bar{a}_1 z} + \langle g_1, e_{\{a_1\}}\rangle B_{\{a_1\}}(z), \end{aligned} \tag{5}$$

where

$$\begin{aligned} f_2(z) &= f_1(z) - \langle f_1, e_{\{a_1\}}\rangle e_{\{a_1\}}(z) \\ &= g_1(z) - (1-|a_1|^2)\frac{g_1(a_1)}{1-\bar{a}_1 z} \end{aligned}$$

having zero at $z = a_1$, and thus

$$g_2(z) = f_2(z)\frac{1-\bar{a}_1 z}{z-a_1}$$

is, again, in $H^2(\mathbb{D})$. The above relations hold for any $a_1 \in \mathbb{D}$. Now in (5) we wish to minimize the energy $\| f_2\|^2$ of the remainder f_2. Because of the orthogonality between $f - f_2$ and the orthogonal projection $\langle g_1, e_{\{a_1\}}\rangle B_{\{a_1\}}$ this is equivalent to maximizing the quantity (see (4))

$$|\langle g_1, e_{\{a_1\}}\rangle|^2 = (1-|a_1|^2)|g_1(a_1)|^2.$$

We can show that at an interior point a_1 of the disc $\mathbb{D}$ (a *critical point of* g_1) the maximum value is actually attainable, that is

$$(1-|a_1|^2)|g_1(a_1)|^2 = \sup\{|\langle g_1, e_{\{a\}}\rangle|^2 : a \in \mathbb{D}\}. \tag{6}$$

This fact is stated as

Lemma 2.1 (The maximal selection principle) *For any function $g \in H^2(\mathbb{D})$ there exists $a \in \mathbb{D}$ such that* (6) *holds.*

Remark 2.1 The existence of such a_1 turns to be crucial for the recursive steps that leads to success of the algorithm. It, however, is similar to greedy algorithm in the aspect that the critical points $a_1 \in \mathbb{D}$ may not be unique. In Section 3 we will provide some aspects on uniqueness and continuity of such a_1.

Proof It suffices to show

$$\lim_{|a|\to 1-} \|g - \langle g, B_{\{a\}}\rangle B_{\{a\}}\| = \|g\|. \tag{7}$$

Let P_r denote the Poisson kernel for the unit circle at the point $r \in (0, 1)$. For $\epsilon > 0$, we can choose r sufficiently close to 1 so that owing to the $\mathcal{L}^2$ approximation property of the Poisson integral, there holds

$$\begin{aligned} \|g\| &\geq \|g - \langle g, B_{\{a\}}\rangle B_{\{a\}}\| \\ &\geq \|P_r * (g - \langle g, B_{\{a\}}\rangle B_{\{a\}})\| \\ &\geq \|P_r * g\| - |\langle g, B_{\{a\}}\rangle| \|P_r * B_{\{a\}}\| \\ &\geq (1-\epsilon)\|g\| - \|g\| \|P_r * B_{\{a\}}\|. \end{aligned} \tag{8}$$

Now with the fixed r, since $B_{\{a\}} \in H^\infty(\mathbb{D})$, there holds ([5], Corollary 3.2, p. 58), for $z = re^{it}$,

$$P_r * B_{\{a\}}(e^{it}) = B_{\{a\}}(z).$$

Therefore, we have

$$\begin{aligned} \|P_r * B_{\{a\}}\|^2 &= \frac{1}{2\pi}\int_0^{2\pi} \frac{1-|a|^2}{|1-\overline{a}re^{it}|^2}dt \\ &= \frac{1}{2\pi}\frac{1-|a|^2}{1-r^2|a|^2}\int_0^{2\pi} \frac{1-r^2|a|^2}{|1-|ra|e^{it}|^2}dt \\ &= \frac{1-|a|^2}{1-r^2|a|^2}\int_0^{2\pi} P_{r|a|}(e^{it})dt \\ &= \frac{1-|a|^2}{1-r^2|a|^2}. \end{aligned}$$

When $|a|$ is close to 1, the inequality (8) gives

$$\|g\| \geq \|g - \langle g, B_{\{a\}}\rangle B_{\{a\}}\| \geq (1-2\epsilon)\|g\|.$$

This shows the desired limit (7). The proof is complete. □

Remark 2.2 The proof of Lemma 2.1 is based on the density of the Poisson integrals, and the availability of the desired relation for the Poisson integrals (also see [17]). What is proved is actually a particular case of a known result: For $g \in H^p(\mathbb{D}),\ 0 < p < \infty$, there holds ([6], p. 123)

$$g(z) = o\left(\frac{1}{(1-|z|^2)^{1/p}}\right), \qquad |z| \to 1. \tag{9}$$

Similar estimates also hold for functions in Bergman spaces ([7], p. 54). These more general results are proved by using the density of analytic polynomials and the availability of the desired relations for those polynomials. In spite of existence of the results, in order to make the article self-containing and easy for the future referring especially in the contexts where analytic approximation is not available (multiply-connected domains, several real variables etc.), we choose to include the proof using Poisson kernel argument.

The adopted notation implies $\langle f_1, B_{\{a_1\}}\rangle = \langle f, B_{\{a_1\}}\rangle = \langle g_1, e_{\{a_1\}}\rangle$. Now to g_2 repeating the same procedure we have

$$\begin{aligned} f(z) &= g_2(z)\frac{z-a_1}{1-\bar{a}_1 z} + \langle g_1, e_{\{a_1\}}\rangle e_{\{a_1\}}(z) \\ &= \left(g_3(z)\frac{z-a_2}{1-\bar{a}_2 z} + \langle g_2, e_{\{a_2\}}\rangle e_{\{a_2\}}(z)\right)\frac{z-a_1}{1-\bar{a}_1 z} + \langle g_1, e_{\{a_1\}}\rangle e_{\{a_1\}}(z) \\ &= g_3(z)\frac{z-a_2}{1-\bar{a}_2 z}\frac{z-a_1}{1-\bar{a}_1 z} + \langle g_2, e_{\{a_2\}}\rangle B_{\{a_1,a_2\}} + \langle g_1, e_{\{a_1\}}\rangle B_{\{a_1\}}(z), \end{aligned}$$

where a_2 is chosen in $\mathbb{D}$ so that

$$|\langle g_2, e_{\{a_2\}}\rangle|^2 = (1-|a_2|^2)|g_2(a_2)|^2 = \sup\{|\langle g_2, e_{\{a\}}\rangle|^2 : a \in \mathbb{D}\}. \tag{10}$$

According to Lemma 2.1 this is always possible. We also have

$$\langle g_2, e_{\{a_2\}}\rangle = \langle f_2, B_{\{a_1,a_2\}}\rangle = \langle f, B_{\{a_1,a_2\}}\rangle.$$

Repeating this process up to n-times, we obtain

$$\begin{aligned} f(z) \;=\; f_1(z) &= g_{n+1}(z)\prod_{l=1}^{n}\frac{z-a_l}{1-\bar{a}_l z} + \sum_{l=1}^{n}\langle g_l, e_{\{a_l\}}\rangle B_{\{a_1,\ldots,a_l\}}(z) \\ &= g_{n+1}(z)\prod_{l=1}^{n}\frac{z-a_l}{1-\bar{a}_l z} + \sum_{l=1}^{n}\langle f, B_{\{a_1,\ldots,a_l\}}\rangle B_{\{a_1,\ldots,a_l\}}(z), \end{aligned}$$

where, for $l = 2, \ldots, n+1$, as recursive formula,

$$\begin{aligned} g_l(z) &= \left(g_{l-1}(z) - \langle g_{l-1}, e_{\{a_{l-1}\}}\rangle e_{\{a_{l-1}\}}(z)\right)\frac{1-\bar{a}_{l-1}z}{z-a_{l-1}} \\ &= \left(g_{l-1}(z) - (1-|a_{l-1}|^2)\frac{g_{l-1}(a_{l-1})}{1-\bar{a}_{l-1}z}\right)\frac{1-\bar{a}_{l-1}z}{z-a_{l-1}}, \end{aligned} \tag{11}$$

and

$$|\langle g_l, e_{\{a_l\}}\rangle|^2 = (1-|a_l|^2)|g_l(a_l)|^2 = \max\{|\langle g_l, e_{\{a\}}\rangle|^2 : a \in \mathbb{D}\}, \quad l = 1, \ldots, n. \tag{12}$$

Denote by f_l the usual remainder,

$$f_l(z) = g_l(z) \prod_{k=1}^{l-1} \frac{z - a_k}{1 - \bar{a}_k z}, \tag{13}$$

we have the relations

$$\langle g_l, e_{\{a_l\}} \rangle = \langle f_l, B_{\{a_1,\ldots,a_l\}} \rangle = \langle f, B_{\{a_1,\ldots,a_l\}} \rangle. \tag{14}$$

The inductive procedure results in an infinite sequence $\{a_k\}$ in $\mathbb{D}$ such that at each step we have chosen a_k that gives rise to the best approximation. Getting the best at each step does not automatically guarantee the convergence. In our case, however, we have

Theorem 2.2 *For a given function $f \in H^2(\mathbb{D})$ under the maximal selection principle in Lemma 2.1 starting from $g_1 = f$, we have*

$$f = \sum_{k=1}^{\infty} < f, B_k > B_k, \tag{15}$$

where $B_k = B_{\{a_1,\ldots,a_k\}}$.

Remark 2.3 We note that $\{B_k\}_{k=1}^{\infty}$ may not form a basis for $H^2(\mathbb{D})$ (also see Section 3). What we are interested is only a fast convergence in the energy sense to the given function. This program results a decomposition of the given function into its intrinsic components in the rational orthonormal system.

To prove the theorem we first prove a lemma.

Lemma 2.3 *Let $0 \neq g \in H^2(\mathbb{D})$. Denote by Z_g the set of the complex numbers $a \in \mathbb{D}$ such that*

$$\langle g, e_{\{a\}} \rangle = 0.$$

Then the accumulation points of Z_g are contained in $\partial\mathbb{D}$.

Proof Apart from a constant multiple, $\langle g, e_{\{a\}} \rangle$ is the evaluation of g at the point a. Then the theorem of isolated zeros of analytic function concludes the lemma. □

Proof of Theorem 2.2 We prove the assertion by introducing a contradiction. If (15) does not hold, then

$$0 \neq g = f - \sum_{k=1}^{\infty} \langle f, B_k \rangle B_k, \tag{16}$$

and

$$0 < \|g\|^2 = \|f\|^2 - \sum_{k=1}^{\infty} |\langle f, B_k \rangle|^2, \tag{17}$$

and, owing to (14) and (17),

$$\lim_{k\to\infty}\langle f, B_k\rangle = \lim_{k\to\infty}\langle f_k, B_k\rangle = \lim_{k\to\infty}\langle g_k, e_{\{a_k\}}\rangle = 0, \tag{18}$$

where g_k and f_k are defined through (11) and (13). They satisfy

$$f_k = f - \sum_{l=1}^{k-1}\langle f, B_l\rangle B_l.$$

By Lemma 2.3, there exists $b \in \mathbb{D}$ such that

$$|\langle g, e_{\{b\}}\rangle| = \delta > 0.$$

We may, in particular, select $b \in \mathbb{D}$ that distinguishes from all the chosen a_k in the sequence. Set

$$h_k = -\sum_{l=k}^{\infty}\langle f, B_l\rangle B_l.$$

Thus, $g = f_k + h_k$. Due to (17), when k is large,

$$|\langle h_k, e_{\{b\}}\rangle| \le \|h_k\|\|e_{\{b\}}\| < \delta/2,$$

and hence

$$|\langle f_k, e_{\{b\}}\rangle| + \delta/2 \; > \; |\langle f_k, e_{\{b\}}\rangle + \langle h_k, e_{\{b\}}\rangle| = \delta.$$

So,

$$|\langle f_k, e_{\{b\}}\rangle| > \delta/2. \tag{19}$$

This implies that for large k,

$$|\sqrt{1-b^2}\, f_k(b)| > \delta/2. \tag{20}$$

Since Blaschke products are dominated by the constant 1, the relation (13) gives

$$|g_k(z)| \ge |f_k(z)|, \tag{21}$$

and thus

$$|\sqrt{1-b^2} g_k(b)| > \delta/2, \tag{22}$$

or, equivalently,

$$|\langle g_k, e_{\{b\}}\rangle| > \delta/2. \tag{23}$$

On the other hand, when k is large, (18) gives

$$|\langle g_k, e_{\{a_k\}}\rangle| < \delta/2. \tag{24}$$

If k if fixed large enough to satisfy both (23) and (24), then we arrive a contradiction. In fact, the maximal selection principle proved in Lemma 2.1 asserts that we should not have chosen a_k but b. This proves the theorem. □

Remark 2.4 As already mentioned the obtained B_n's, in general, may only be pre-mono-components. If $a_1 = 0$, and the other a_n's are chosen according to Lemma 2.1, then we obtain an adaptive system consisting of mono-components. Apart from modifying the algorithm at the beginning step, one can first get $g(z) = \frac{1}{z}(f(z) - f(0))$, and then to $g(z)$ perform the algorithm defined by Lemma 2.1 and Theorem 2.2. The adaptive mono-component decomposition of f then will be obtained from the adaptive decomposition of g and the relation $f(z) = f(0) + zg(z)$.

Remark 2.5 As already mentioned, such chosen a_n's (as in Theorem 2.2) give rise to a complete system $\{B_n\}$ if and only if condition (2) is met. Some minor modifications may be made in order to make $\{B_k\}$ to be a complete system. The expense is to give up the total adaptivity. For instance, if, besides the maximal selection principle, we add the constraint condition $|a_n| \leq 1 - r_n$, where $\{r_n\}$ are any non-negative numbers satisfying $\sum_{r=1}^{\infty} r_n = \infty$. Then the chosen a_n's satisfy the condition (2) and thus gives rise to a complete system. Also see Theorem 3.2 below.

3 More aspects on critical points and convergence

In this section we will deal with two theoretical aspects in relation to the algorithm. The first one concerns uniqueness and continuity of the maximizers (critical points) of $A_G^2(z)$ defined by (25) for $G \in H^2(\mathbb{D})$.

For $G \in H^2(\mathbb{D})$ define m_G to be the maximal inner product modular

$$m_G := \max\{A_G(z) \ : \ z \in \mathbb{D}\},$$

where A_G is

$$A_G(z) := |\langle G, e_{\{z\}}\rangle| = \sqrt{1 - |z|^2}|G(z)|. \tag{25}$$

Denote by

$$z_G := \{z \in \mathbb{D} \ : \ A_G(z) = m_G\}, \tag{26}$$

the set of maximizers. For a given $G \in H^2(\mathbb{D})$ the set z_G may contain more than one points. We have

Theorem 3.1 *Let F, G, G_n be functions in $H^2(\mathbb{D})$, and $G_n \to G$ in $H^2(\mathbb{D})$ and $\|G\|_{H^2(\mathbb{D})} > 0$. Let $z_n \in z_{G_n}$. Then*

(i)

$$|m_F - m_G| \leq \|F - G\|_{H^2(\mathbb{D})}.$$

(ii) *There exist a subsequence $z_{n_k} \to z^*$ such that $z^* \in z_G$.*

(iii) *If z_G contains only one element z^*, then $z_n \to z^*$.*

Proof

(i) Without loss of generality we may assume $m_F \le m_G$. In the case, by the triangle inequality for complex numbers, for $z \in z_G$,

$$\begin{aligned} m_G - m_F &\le A_G(z) - A_F(z) \\ &\le \sqrt{1-|z|^2}|(G-F)(z)| \\ &= |\langle G-F, e_z\rangle| \\ &\le \|G-F\|_{H^2(\mathbb{D})}. \end{aligned}$$

(ii) Take $\epsilon \in (0, \frac{1}{4}m_G)$. Owing to Lemma 2.1, there exists $\delta > 0$ such that

$$A_G(z) < \epsilon \quad \text{whenever} \quad |z| > 1-\delta. \tag{27}$$

Chose n large enough so that

$$|m_{G_n} - m_G| \le \|G_n - G\|_{H^2(\mathbb{D})} < \epsilon \tag{28}$$

and, for any $z \in \mathbb{D}$,

$$|A_{G_n}(z) - A_G(z)| \le \|G_n - G\|_{H^2(\mathbb{D})} < \epsilon \tag{29}$$

simultaneously hold. There follows

$$|A_{G_n}(z_n) - m_G| < \epsilon < \frac{1}{4}m_G,$$

and thus

$$A_{G_n}(z_n) > \frac{1}{2}m_G \quad \text{and} \quad A_G(z_n) \ge A_{G_n}(z_n) - \epsilon > \frac{1}{4}m_G.$$

The last inequality shows that when n is large,

$$|z_n| \le 1-\delta.$$

Therefore there exists a subsequence z_{n_k} with

$$\lim_{k\to\infty} z_{n_k} = z^* \in \overline{(1-\delta)\mathbb{D}} \quad \text{and} \quad \lim_{k\to\infty} A_G(z_{n_k}) = A_G(z^*).$$

We show that $A_G(z^*) = m_G$. Due to the continuity of A_G and the inequalities in (28) and (29), for any $\epsilon_1 > 0$ and large enough k, the following three inequalities simultaneously hold:

$$|A_G(z_{n_k}) - A_G(z^*)| < \epsilon_1,$$

$$|A_G(z_{n_k}) - A_{G_{n_k}}(z_{n_k})| < \epsilon_1,$$

$$|A_{G_{n_k}}(z_{n_k}) - m_G| < \epsilon_1.$$

Therefore,

$$|A_G(z^*) - m_G| < 3\epsilon_1,$$

and thus $A_G(z^*) = m_G$.

(iii) In the proof of (ii) we show that for large enough n, there holds $z_n \in \overline{(1-\delta)\mathbb{D}}$. Therefore, if $\{z_n\}$ does not converge, then it has more than one accumulation pints in $\mathbb{D}$. The result proved in (ii) asserts that all the accumulation points of $\{z_n\}$ are in z_G, contrary to the assumption that z_G only contains a single point. The proof is complete. □

In practice the convergence result proved in Theorem 2.2 is sufficient. In fact, no application requires expansions of all functions in the space. In system identification, for instance, only the system function of the rational function type is to be sought. The following theorem is a refinement of Theorem 2.2 characterizing the subspace in which the chosen system is complete. Before we state the theorem we need to introduce some notation.

We say that b_k *has the multiplicity* l in the n-tuple $\{b_1, ..., b_n\}$, $n \geq l$, if there totally exist l entries, $b_{n_1}, ..., b_{n_l}$, $1 \leq n_1 < \cdots < n_l = k$, such that $b_{n_1} = \cdots = b_{n_l} = b_k$. In other words, up to the entry b_k the number b_k altogether appears l times.

We call

$$\tilde{e}_{\{b_1\}}, \tilde{e}_{\{b_2\}}, ..., \tilde{e}_{\{b_n\}},$$

the n-system associated with the n-tuple $\{b_1, ..., b_n\}$ if

$$\tilde{e}_{\{b_k\}} = \frac{1}{(1-\overline{b}_k z)^l}$$

when $b_k \neq 0$ and has the multiplicity l; and

$$\tilde{e}_{\{b_k\}} = z^{l-1}$$

when $b_k = 0$ and has the multiplicity l, $1 \leq l \leq n$. The n-system associated with an n-tuple is not orthogonal.

Both the definitions for the multiplicity of b_k in an n-tuple and the associated n-tuple functions $\tilde{e}_{\{b_k\}}$ may be extended to $n = \infty$.

We have

Theorem 3.2 *For a sequence* $\{b_n\}$ *in* $\mathbb{D}$, *if the condition* (2) *does not hold, then*

$$H^2(\mathbb{D}) = \overline{\mathrm{span}\{\tilde{e}_{\{b_1\}}, ..., \tilde{e}_{\{b_n\}}, ...\}} \oplus \phi H^2(\mathbb{D}), \tag{30}$$

where ϕ *is the Blaschke product that has and only has the* b_n*'s as its zeros including the multiplicity.*

Proof It is obvious that both $\overline{\mathrm{span}\{\tilde{e}_{\{b_1\}}, ..., \tilde{e}_{\{b_n\}}, ...\}}$ and $\phi H^2(\mathbb{D})$ are closed subspaces of $H^2(\mathbb{D})$. To show that they are orthogonal complements to each other it suffices to show that they are orthogonal, and each function in $H^2(\mathbb{D})$ may be decomposed in a unique way into a sum of two functions in the respective two subspaces [10]. To show the orthogonality it suffices to show each $\tilde{e}_{\{b_k\}}$ is orthogonal with any function of the form ϕg, $g \in H^2(\mathbb{D})$. By computing the corresponding inner product this turns to be a conclusion of Cauchy's Theorem. Bear in mind that the span of the n-tuple of functions $\{\tilde{e}_{\{b_1\}}, ..., \tilde{e}_{\{b_n\}}\}$

is the same as the span of the n-system $\{B_{\{b_1\}}, ..., B_{\{b_n\}}\}$, and the same is true for the whole sequence $\{\tilde{e}_{\{b_1\}}, ..., \tilde{e}_{\{b_n\}}, ...\}$ and the system $\{B_{\{b_1\}}, ..., B_{\{b_n\}}, ...\}$. Now, for any function $f \in H^2(\mathbb{D})$, the remainder $f_{n+1} = f - \sum_{k=1}^{n} \langle f, B_k \rangle B_k$, in view of (13), has zeros $b_1, ...b_n$, including the multiplicities. For arbitrary n, $f - \sum_{k=1}^{\infty} \langle f, B_k \rangle B_k = f - \sum_{k=1}^{n} \langle f, B_k \rangle B_k - \sum_{k=n+1}^{\infty} \langle f, B_k \rangle B_k$, where every B_k, $k \geq n+1$, has zeros $b_1, b_2, ..., b_n$ with the multiplicities. Therefore, $f - \sum_{k=1}^{\infty} \langle f, B_k \rangle B_k$ has zeros $b_1, ..., b_n, ...$, including the multiplicities. As a consequence,

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

for some $g \in H^2(\mathbb{D})$. The uniqueness of the decomposition is obvious. The proof is complete. □

Corollary 3.3 *If the chosen $\{a_n\}$ according to the given function $f \in H^2(\mathbb{D})$ as in Theorem* 2.2 *does not satisfy the condition* (2)*, then $\{B_n\}$ is complete in the closed subspace* $\overline{\text{span}\{\tilde{e}_{\{a_1\}}, ..., \tilde{e}_{\{a_n\}}, ...\}}$ *that contains f.*

Acknowledgements The authors would like to thank L.X. Yan for his kind help in getting useful references. The first named author wishes to sincerely thank C. Micchelli, K. H. Zhu, M. I. Stessin, Y. S. Xu and L. X. Shen for their kind invitations to, and instructive and inspiring discussions in University at Albany and Syracuse University in November 2009. Special thanks are due to Micchelli and Xu whose questions motivated the author working out the results on uniqueness and continuity in relation to the selections of the a_n's (Theorem 3.1) and the convergence result in relation to the shift-invariance space (Theorem 3.2). The authors also wish to thank the referees whose comments and suggestions helped to improve the quality of the paper.

References

1. Akcay, H., Niness, B.: Orthonormal basis functions for modelling continuous-time systems. Signal Process. **77**, 261–274 (1999)
2. Bultheel, A., Carrette, P.: Takenaka-Malmquist basis and general Toeplitz matrices. SIAM J. Opt. **41**, 1413–1439 (2003)
3. Bultheel, A., Gonzalez-Vera, P., Hendriksen, E., Njastad, O.: Orthogonal rational functions. In: Cambridge Monographs on Applied and Computational Mathematics, vol. 5. Cambridge University Press (1999)
4. Davis, G., Mallat, S., Avellaneda, M.: Adaptive greedy approximations. Constr. Approx. **13**, 57–98 (1997)
5. Garnett, J.B.: Bounded Analytic Functions. Academic, New York (1987)
6. Zhu, K.: Spaces of Holomorphic Functions in the Unit Ball. GTM, Springer, New York (2005)
7. Hedenmalm, H., Korenblum, B., Zhu, K.: Theory of Bergman Spaces. GTM, Springer, New York (2000)
8. Mallat, S., Zhang, Z.: Matching pursuits with time-frequency dictionaries. IEEE Trans. Signal Process. **41**, 3397–3415 (1993)
9. Ninness, B., Hjalmarsson, H., Gustafsson, F.: Generalized Fourier and Toeplitz results for rational orthonormal bases. SIAM J. Control Optim. **37**(2), 429–460 (1999)
10. Partington, J.R.: Interpolation, Identification and Sampling, pp. 44–47. Clarendon Press, Oxford (1997)
11. Heuberger, P.S.C., Van den Hof, P.M.J., Wahlberg, B.: Modelling and Identification with Rational Orthogonal Basis Functions. Springer, London (2005)

12. Qian, T.: Mono-components for decomposition of signals. Math. Methods Appl. Sci. **29**, 1187–1198 (2006)
13. Qian, T.: Boundary derivative of the phases of inner and outer functions and applications. Math. Methods Appl. Sci. **32**, 253–263 (2009)
14. Qian, T., Ho, I.T., Leong, I.T., Wang, Y.B.: Adaptive decomposition of functions into pieces of non-negative instantaneous frequencies. Accepted by International Journal of Wavelets, Multiresolution and Information Processing, vol. 8(5) (2010)
15. Qian, T., Wang, R., Xu, Y.S., Zhang, H.Z.: Orthonormal bases with nonlinear phases. Adv. Comput. Math. **33**, 75–95 (2010)
16. Qian, T., Xu, Y.S., Yan, D.Y., Yan, L.X., Yu, B.: Fourier spectrum charaterization of hardy spaces and applications. Proc. Am. Math. Soc. **137**, 971–980 (2009)
17. Temlyakov, V.N.: Greedy algorithm and m-term trigonometric approximation. Constr. Approx. **107**, 569–587 (1998)
18. Tan, L.H., Shen, L.X., Yang, L.H.: Rational orthogonal bases satisfying the Bedrosian identity. Adv. Comput. Math. doi:10.1007/s10444-009-9133-8
19. Walsh, J.L.: Interpolation and Approximation by Rational Functions in the Complex Plane. AMS (1969)
20. Wang, R., Xu, Y.S., Zhang, H.Z.: Fast non-linear Fourier expansions. AADA **1**(3), 373–405 (2009)
21. Xu, Y.S., Yan, D.Y.: The Bedrosian identity fot the Hilbert transform of product functions. Proc. Am. Math. Soc. **134**, 2719–2728 (2006)