International Journal of Wavelets, Multiresolution
and Information Processing
Vol. 11, No. 1 (2013) 1350007 (14 pages)

DOI: 10.1142/S0219691313500070

World Scientific
www.worldscientific.com

REMARKS ON ADAPTIVE FOURIER DECOMPOSITION*

TAO QIAN*

Department of Mathematics
Faculty of Science and Technology
University of Macau Taipa, Macau,China
fsttq@umac.mo

YANBO WANG†

Department of Mathematics
Faculty of Mathematics and Computer Science
Wuhan Textile University
Wuhan City, Hubei Province,China
wybgx@yahoo.com.cn

Received 12 March 2012
Revised 2 December 2012
Published

This is ~~a~~ continuation of the study of adaptive Fourier decomposition (AFD).[15] ~~It can be treated as a variation of greedy algorithm.~~ Under a mild condition not in terms of smoothness, a convergence rate is provided. We prove that the selection of the parameters corresponding to Fourier series in the average sense is optimal. We also present the transformation matrices between the adaptive rational orthogonal system and the related sequence of the shifted Cauchy kernels and their derivatives.

Keywords: Adaptive Fourier series; orthogonal greedy algorithm; dictionary; maximal selection principle; Cauchy kernel.

AMS Subject Classification: 42A50, 32A30, 32A35, 46J15

1. Introduction

The rational orthonormal system

$$B_n(z) = B_{\{a_1,a_2,\ldots,a_n\}}(z) = \frac{\sqrt{1-|a_n|^2}}{1-\overline{a}_n z}\prod_{k=1}^{n-1}\frac{z-a_k}{1-\overline{a}_k z}, \quad n = 1,2,\ldots, \tag{1.1}$$

is known as Takenaka–Malmquist system[17] that depends on a given sequence of complex parameters $\{a_k\}$ in unit disc $\mathbb{D}$. It is a generalization of Fourier system

*Supported by University of Macau under research Grant RG056/01-02S/QT/FST
†Corresponding author.

$\{z^n\}_{n=0}^{\infty}$, the latter corresponding to $a_k = 0$ for all k. Laguerre basis and two-parameter Kautz basis[6,7] are also special cases of (1.1). The inner product that we use for $L^2(\partial\mathbb{D})$ and the boundary values of functions in $H^2(\mathbb{D})$ is

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

Under the isometric isomorphism relation between them, we identity $H^2(\mathbb{D})$ with the space of the boundary values of the functions in $H^2(\mathbb{D})$. As is well known, the condition

$$\sum_{k=1}^{\infty}(1-|a_k|) = \infty \tag{1.2}$$

is sufficient and necessary for $\{B_n\}$ to be a complete basis in the Hardy space $H^2(\mathbb{D})$. All the traditional studies of the orthonormal system are based on the condition (1.2). In Ref. 15, we introduce an approach to functional decomposition that is different from all those traditionally using the TM system. Instead of using a previously known parameter sequence $\{a_k\}$ satisfying the condition (1.1), we choose $\{a_k\}$ according to the given signal f ~~to be decomposed~~. There are two main reasons of doing such decomposition. First, such decomposition is adaptive. Intuitively, as well as supported by experiments, approximation to a given f with fast convergence in energy is achieved. Secondly, under such decomposition any physically realizable signal may be decomposed into a series of mono-components of which each possesses non-negative and thus physically meaningful instantaneous frequencies.[1,2,8,12,13] In particular, if we set $a_1 = 0$, then all $B_{\{a_1,a_2,\ldots,a_k\}}(z)$ become multi-starlike functions, and therefore their phase derivatives are non-negative on the boun

~~Subsequent~~ to the previously established convergence result ~~under the~~ greedy algorithm ~~principle~~, the present paper ~~further~~ proves a convergence rate that demonstrates the fastness of the convergence of AFD. The writing plan is as follows. In Sec. 2 we describe the AFD algorithm referred to Ref. 15. In Sec. 3 we prove the convergence rate. In Sec. 4 we show that in the average sense Fourier series is the optimal. In Sec. 5 we provide the transformation matrices between the adaptive rational orthogonal system and the related sequenc the shifted Cauchy kernels and their ~~variations with multiples larger than one.~~

2. Adaptive Fourier Decomposition

Let $f \in H^2(\mathbb{D})$. To expand $f = f_1 = g_1$ into its Fourier series we use the following process. The remainder $f_2(z) = f_1(z) - f_1(0)$ has zero at $z = 0$. Therefore, the reduced remainder $g_2(z) = f_2(z)/z \in H^2(\mathbb{D})$. Since $g_2(z) - g_2(0)$ has zero at $z = 0$, the reduced remainder $g_3 = (f_2 - f_2(0))/z \in H^2(\mathbb{D})$, and so on. We subsequently have

$$\begin{aligned} f(z) = f_1(z) &= f_2(z) + g_1(0) \\ &= zg_2(z) + g_1(0) \\ &= z^2 g_3(z) + zg_2(0) + g_1(0) \\ &\ \ \vdots \\ &= z^{n+1} g_n(z) + z^n g_{n-1}(0) + \cdots + zg_2(0) + g_1(0). \end{aligned} \tag{2.1}$$

This process is to first project f_1 onto the unit vector one and then find the remainder. Then project the reduced remainder, g_2, that is the standard remainder f_2 being divided by z. The process from g_1 to get the reduced remainder g_2 may be called a *Fourier sifting*. Then project g_2 onto the same unit vector one, and subsequently find the next standard remainder f_3 and the reduced remainder g_3, and so on. Every time it projects the reduced remainder to the unit vector one. Projecting onto the unit vector one amounts to take average on all function values, that is equivalent to evaluate the function value at zero. In AFD, instead of projecting $f_1 = g_1$ onto the unit vector one, we project it onto the *evaluator*

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

Note that it is a generalization of one, as $e_{\{0\}} = 1$. By Cauchy's integral formula, we have

$$\langle f_1, e_{\{a\}} \rangle = \sqrt{1-|a|^2} f_1(a).$$

Below we denote, for any $f \in H^2(\mathbb{D})$,

$$A_a(f) = (1-|a|^2)|f(a)|^2. \tag{2.2}$$

We adaptively select $a = a_1 \in \mathbb{D}$ so that

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

In Ref. 15, we prove that for any $g_1 \in H^2(\mathbb{D})$ such a_1 is attainable at a point in $\mathbb{D}$. This result is called the *Maximal Selection Principle*. The standard remainder $f_2(z) = f_1(z) - \sqrt{1-|a_1|^2} f_1(a_1) e_{\{a_1\}}(z)$ is accordingly the minimized one in norm sense. We subsequently find the reduced remainder g_2 by

$$g_2(z) = f_2(z) \frac{1-\overline{a}_1 z}{z-a_1}. \tag{2.3}$$

We call the process getting g_{k+1} from g_k through such optimal selection of a_k based on the Maximal Selection Principle a *maximal sifting process*, or a *maximal sifting process through* a_k. If we algebraically deduce g_{k+1} from q_k not through an optimal selection of $e_{\{a_k\}}$ based on the Maximal Selection Principle, but through some evaluator $e_{\{b\}}$, then the corresponding process is called the *sifting process through* b. The sifting process through $a = 0$ is the so-called Fourier sifting.

A dictionary $\mathcal{D}$ in a Hilbert space H is a set of functions of unit norm with $\overline{\text{Span}\,\mathcal{D}} = H$. A dictionary is, in general, redundant. Redundancy offers greater efficiency in approximation. ~~Closely related to a dictionary in nonlinear approximation is greedy algorithm[5,3] and its variations.[16] The above-mentioned adaptive decomposition programme can be regarded as a modified orthogonal greedy algorithm as follows.~~ We begin with a dictionary $\mathcal{D}$ in the Hardy Space $H^2(D)$. Here, the special dictionary $\mathcal{D}$ is given by

$$\mathcal{D} := \left\{ e_{\{a\}}(z) = \frac{\sqrt{1-|a|^2}}{1-\bar{a}z}, a \in D \right\}. \tag{2.4}$$

1 Note that every $e_{\{a\}}(z)$ is a normalized Cauchy kernel function.

Adaptive Fourier Decomposition. Associated with AFD the following notations and properties will be used. We set $g_1 := f_1 := f$. Then, for each $m \geq 1$, we inductively define

$$H_m(f) := \text{span}\{B_{\{a_1\}}, B_{\{a_1,a_2\}}, \ldots, B_{\{a_1,\ldots,a_m\}}\}. \tag{2.5}$$

The standard remainder

$$f_{m+1} := f - P_{H_m}(f), \tag{2.6}$$

where $P_{H_m}(f)$ is the orthogonal projection of f to $H_m(f)$. f_m are standard reminders:

$$f_m = f - \sum_{k=1}^{m-1} \langle f, B_m \rangle B_m.$$

In particular,

$$\|f_m\|^2 = \|f_{m-1}\|^2 - |\langle f_{m-1}, B_{m-1}\rangle|^2. \tag{2.7}$$

~~The reduced remainders~~

$$g_m := f_m \prod_{k=1}^{m-1} \frac{1-\bar{a}_k z}{z - a_k}. \tag{2.8}$$

We have where

$$\langle f, B_m \rangle = \langle f_m, B_m \rangle = \langle g_m, e_{\{a_m\}} \rangle.$$

In AFD we employ maximal sifting processes, that is, when the proceeding $e_{\{a_l\}}, l = 1, \ldots, m-1$, have been selected, the next $e_{\{a_m\}}$ is selected according to Maximal Selection Principle, that is

$$|\langle g_m, e_{\{a_m\}} \rangle| = \max\{|\langle g_m, e_{\{a\}} \rangle| : a \in \mathbb{D}\}, \tag{2.9}$$

where

$$A_a(g_m) = |\langle g_m, e_{\{a\}} \rangle|^2 = (1 - |a|^2)|g_m(a)|^2.$$

2 After all, we have

AQ: Sentence seems incomplete. Please check.

Theorem 2.1.[15] *For any function $f \in H^2(\mathbb{D})$, if $a_1, \ldots, a_k, \ldots$ are consecutively selected according to the Maximal Selection Principle, then we have*

$$f = \sum_{k=1}^{\infty} \langle f, B_{\{a_1,\ldots,a_k\}} \rangle B_{\{a_1,\ldots,a_k\}}$$

in the $H^2(\mathbb{D})$ sense.

3. Convergence Rate on $H^2(\mathcal{D}, M)$

For the dictionary $\mathcal{D}$, we define the subclasses of functions

$$H^2(\mathcal{D}, M) := \left\{ f \in H^2(D) : f = \sum_{k=1}^{\infty} c_k e_k, e_k \in \mathcal{D},\ \sum_{k=1}^{\infty} |c_k| \le M \right\}. \tag{3.1}$$

Note that the convergence in the definition takes the H^2-norm sense.

Lemma 3.1. *If f in $H^2(\mathcal{D}, M)$, then $\|f\| \le M$.*

Proof. For $f \in H^2(\mathcal{D}, M)$, there exist a sequence of complex numbers $\{c_k\}$ and a sequence $\{e_k\} \in \mathcal{D}$ such that $f = \sum_{k=1}^{\infty} c_k e_k$ with $\sum_{k=1}^{\infty} |c_k| \le M$,

$$\begin{aligned} \|f\|^2 &= |\langle f, f \rangle| \\ &= \left| \langle f, \sum_{k=1}^{\infty} c_k e_k \rangle \right| \\ &\le \sum_{k=1}^{\infty} |c_k| |\langle f, e_k \rangle|. \end{aligned} \tag{3.2}$$

From the Schwarz inequality and the characterized expansion of f in $\{e_k\}$,

$$\|f\|^2 \le M \|f\|, \tag{3.3}$$

which gives $\|f\| \le M$. □

We have:

Lemma 3.2. *Let $f \in H^2(\mathcal{D}, M)$ and $f = \sum_{k=1}^{\infty} c_k e_{\{a_k\}}$. If there exists a series of positive numbers such that $\sum_{n=1}^{\infty} \rho_n < \infty$ and*

$$\left| \sum_{k=1}^{\infty} c_k \sqrt{1 - |a_k|^2}\, \overline{a}_k^n \right| \le \rho_n,$$

then f belongs to the positive Wiener algebra W_+. In particular, if for every k, $|a_k| < r < 1$, then $f \in W_+$.

Proof. Writing each $e_{\{a_k\}}$ into its Taylor series expansion, we have

$$\begin{aligned} f(z) &= \sum_{k=1}^{\infty} c_k e_{\{a_k\}}(z) \\ &= \sum_{k=1}^{\infty} c_k \sqrt{1-|a_k|^2}\left(1+\sum_{n=1}^{\infty} \overline{a}_k^n z^n\right) \\ &= \sum_{k=1}^{\infty} c_k \sqrt{1-|a_k|^2} + \sum_{n=1}^{\infty} z^n \sum_{k=1}^{\infty} c_k \sqrt{1-|a_k|^2}\,\overline{a}_k^n. \end{aligned}$$

In the closed unit disc the series is uniformly dominated by

$$CM + \sum_{n=1}^{\infty} \rho_n,$$

and therefore is in the positive Wiener algebra. $\square$

We now turn to analysis of approximation rate of AFD. We need the following lemma.

Lemma 3.3.[4] *Let $\{d_m\}_{m=1}^{\infty}$ be a sequence of nonnegative numbers satisfying*

$$d_1 \le A,\ d_{m+1} \le d_m \left(1 - \frac{d_m}{A}\right). \tag{3.4}$$

Then there holds

$$d_m \le \frac{A}{m}.$$

Theorem 3.1. *Let $\mathcal{D}$ be the dictionary of normalized Cauchy kernels in $H^2(D)$. Then for each $f \in H^2(\mathcal{D}, M)$, decomposed by Adaptive Fourier Decomposition, we have*

$$\| f_m \| \le \frac{M}{\sqrt{m}}.$$

Proof. In the process of Adaptive Fourier Decomposition, we have, due to (2.7),

$$\| f_{m+1} \|^2 = \| f_m \|^2 - |\langle f_m, B_m\rangle|^2.$$

Since $f \in H^2(\mathcal{D}, M)$, there exists a sequence $\{b_k\} \in D$ such that $f = \sum_{k=1}^{\infty} c_k e_{\{b_k\}}$. Therefore,

$$\begin{aligned} \| f_m \|^2 &= |\langle f_m, f\rangle| \\ &= \left| \langle f_m, \sum_{k=1}^{\infty} c_k e_{\{b_k\}}\rangle \right| \\ &\le M \sup_{b_k} |\langle f_m, e_{\{b_k\}}\rangle| \\ &= M \sup_{b_k} \sqrt{1-|b_k|^2}\,|f_m(b_k)|. \end{aligned} \tag{3.5}$$

From Maximal Selection Principle and computation of the inner product,

$$\begin{aligned}
|\langle f_m, B_m\rangle| &= \sup_{a\in D} |\langle f_m, B_{\{a_1,\ldots,a_{m-1},a\}}\rangle| \\
&= \sup_{a\in D} |\langle g_m, e_{\{a\}}\rangle| \\
&= \sup_{a\in D} \sqrt{1-|a|^2}|f_m(a)|\left|\prod_{k=1}^{m-1}\frac{1-\bar{a}_k a}{a-a_k}\right| \\
&\geq \sup_{b_k} \sqrt{1-|b_k|^2}|f_m(b_k)|\left|\prod_{k=1}^{m-1}\frac{1-\bar{a}_k b_k}{b_k-a_k}\right| \\
&\geq \sup_{b_k} \sqrt{1-|b_k|^2}|f_m(b_k)| \\
&\geq \frac{1}{M}\, \| f_m \|^2,
\end{aligned} \tag{3.6}$$

we therefore have

$$\| f_{m+1} \|^2 \leq \| f_m \|^2 \left(1 - \frac{\| f_m \|^2}{M^2}\right). \tag{3.7}$$

By setting $A = M^2$ and using Lemma 3.3, we obtain the desired estimate. $\square$

Remark 3.1. The proved convergence rate is not a sharp estimate. It addresses the worst case, that, apart from being in $H^2(\mathcal{D}, M)$, does not assume other properties for the signal. It is, in particular, regardless degree of smoothness of the signal. The results on convergence rates of Fourier decomposition heavily rely on smoothness of functions under consideration. Effectiveness (fastness) of greedy algorithm is supported by intuition and experiments. In the concrete experimental examples one often gets small errors after a few maximal sifting processes.

4. Justification of Fourier Series

Below we give a justification on the norm convergence of the traditional Fourier expansion from the adaptive approximation point of view. Fourier expansion of a given function in $H^2(\mathbb{D})$, as described at the beginning of Sec. 2, corresponds to the selection $a_n = 0$ for all n. At every selection it takes $e_{\{0\}} = 1$ in the dictionary, and projects the function and all its reduced remainders onto this fixed elements. We call this Fourier shifting process. We show that for general signals in the Hardy space, in the average sense, the Fourier shifting process gives rise to the best result. We will introduce a probability measure $P(dg)$ of reasonably symmetric properties on the unit sphere $S(H^2(\mathbb{D}))$ of the Hardy $H^2(\mathbb{D})$ space. The first symmetric property to be required is the rotational symmetry. We require, for any $a = re^{it}$,

$$\int_{S(H^2(\mathbb{D}))} |g(a)|^2 P(dg) = L(r). \tag{4.1}$$

The average of the projected energies over all functions on the sphere then is identical with

$$\begin{aligned}\int_{S(H^2(\mathbb{D}))} |\langle g, e_{\{a\}}\rangle|^2 P(dg) &= \int_{S(H^2(\mathbb{D}))} (1-|a|^2)|g(a)|^2 P(dg)\\ &= (1-r^2)L(r), \end{aligned} \tag{4.2}$$

being independent of the orientation e^{it}. We now proceed to showing that the selection $a = 0$, among all $a \in \mathbb{D}$, gives rise to the largest average of the projected energies.

The set of functions $S(H^2(\mathbb{D}))$, being identical with the unit sphere of the l^2 space

$$\left\{(c_0, c_1, \ldots, c_n, \ldots) | \sum_{k=0}^{\infty} |c_k|^2 = 1\right\}, \tag{4.3}$$

is viewed as the direct product of the sets

$$X_1 = \{(|c_0|, \ldots, |c_n|, \ldots) | \sum_{n=0}^{\infty} |c_n|^2 = 1\},$$

and

$$X_2 = \{(e^{i\theta_0}, \ldots, e^{i\theta_n}, \ldots) |\, \theta_n \in [0, 2\pi), n = 0, 1, \ldots\},$$

i.e.

$$S(H^2(\mathbb{D})) = X_1 \times X_2.$$

Let $P(d\rho)$ and $P(d\theta)$ denote the probability measures on X_1 and X_2, respectively, where $P(dg)$ is the product probability of $P(d\rho)$ and $P(d\theta)$, i.e. $P(dg) = P(d\rho) \times P(d\theta)$. $P(d\theta)$ is defined by the independent identical distributions (i.i.d.) of its factor spaces $\{\theta_k : \theta_k \in [0, 2\pi)\}$ of which each is the normalized Lebesgue measure in $[0, 2\pi)$. $P(d\rho)$ is defined by evenly distributed $|c_n|^2$ in $[0, 1]$ for each n. For different n they are not independent, but with the constraint condition given in the definition of the space X_1. Adopting the above defined probability over the unit sphere $S(H^2(\mathbb{D}))$, and considering the random variable

$$A_a(g) = |\langle g, e_{\{a\}}\rangle|^2 = (1-|a|^2)|g(a)|^2, \quad g \in S(H^2(\mathbb{D})), \tag{4.4}$$

we have

Theorem 4.1. *Under the probability defined on $S(H^2(\mathbb{D}))$ the mathematical expectation $E(A_a)$ takes its maximum value at $a = 0$.*

Proof. We have, for any $a = re^{i\alpha} \in \mathbb{D}$,

$$\begin{aligned}\int_{S(H^2(\mathbb{D}))} |g(a)|^2 P(dg) = L(r) &= \frac{1}{2\pi}\int_0^{2\pi} L(r)dt\\ &= \frac{1}{2\pi}\int_0^{2\pi}\int_{S(H^2(\mathbb{D}))} |g(re^{it})|^2 P(dg)dt\end{aligned}$$

$$
\begin{aligned}
&= \int_{S(H^2(\mathbb{D}))} \frac{1}{2\pi} \int_0^{2\pi} |g(re^{it})|^2 dt P(dg) \\
&= \int_{S(H^2(\mathbb{D}))} \sum_{n=0}^{\infty} |c_n|^2 r^{2n} P(dg) \\
&= \int_{X_1} \sum_{n=0}^{\infty} |c_n|^2 r^{2n} P(d\rho). \qquad (4.5)
\end{aligned}
$$

Denoting the probability event $|c_0|^2 \in [\frac{k-1}{N}, \frac{k}{N})$ by E_k, then $P(E_k) = \frac{1}{N}$, and the energy left for $\sum_{k=1}^{\infty} |c_k|^2$ is, approximately, $1 - \frac{k}{N}$. Denote by $P(d\rho/E_k)$ the conditional probability, $k = 1, \ldots, L$, then the last entry of (4.5) is equal to

$$
\begin{aligned}
&\lim_{N\to\infty} \sum_{k=1}^{N} P(E_k) \int_{X_1/E_k} \sum_{n=0}^{\infty} |c_n|^2 r^{2n} P(d\rho/E_k) \\
&= \lim_{N\to\infty} \sum_{k=1}^{N} \frac{1}{N} \int_{X_1/E_k} \left(\frac{k}{N} + r^2 \sum_{n=0}^{\infty} |c_{n+1}|^2 r^{2n} \right) P(d\rho/E_k) \\
&= \lim_{N\to\infty} \left(\sum_{k=1}^{N} \frac{k}{N^2} + r^2 \sum_{k=1}^{N} L(r) \left(1 - \frac{k}{N}\right) \frac{1}{N} \right) \\
&= \left(\int_0^1 t dt + r^2 L(r) \int_0^1 (1-t) dt \right) \\
&= \left(\frac{1}{2} + \frac{r^2}{2} L(r) \right). \qquad (4.6)
\end{aligned}
$$

Comparing (4.5) with (4.6), we obtain

$$
L(r) = \frac{1}{2 - r^2},
$$

and, by (4.4),

$$
\sup_{a\in\mathbb{D}} E(A_a) = \sup_{a\in\mathbb{D}} (1 - |a|^2) L(|a|) = \frac{1}{2}, \quad \text{at } a = 0. \qquad \square
$$

This shows that if we do not know any information about g_1, but only $g_1 = g \in H^2(\mathbb{D})$, then the wise selection of a is $a_1 = 0$. Since the reduced remainder, g_2 obtained through the a_1-sifting is also a general element in $H^2(\mathbb{D})$, the next wise selection is $a_2 = 0$, and so on. Thus Fourier series would be the wisest for decomposing a general element in $H^2(\mathbb{D})$. If, however, specific information of $g_1 \in H^2(\mathbb{D})$ is known, for instance, through the given concrete data, then the Fourier series should not be the best. In general the introduced AFD provides a fast convergence in energy, while the obtained decomposition better frequency aspects than Fourier series. One could also define the probability $P(d\rho)$ on X_1 by the condition that

$|c_n|^p, p \neq 2$, is evenly distributed in $[0,1]$. Theorem 4.1 may be elaborated as follows. Assume that the probability on X_2 satisfies the same orientation symmetric property and the probability on X_1 complies with the following probability law:

(i) For any k the probability density of the distribution of $|c_k|^2$ in $[0,1]$ is $\alpha(t)$:

$$\int_0^1 \alpha(t)dt = 1;$$

and

(ii) When $|c_0|^2$ is fixed belonging to the event $\{|c_0|^2 \in E\}$, the conditional probability distribution $P(d\rho/E)$ for $\left\{\frac{|c_1|^2}{1-|c_0|^2}, \ldots, \frac{|c_n|^2}{1-|c_0|^2}, \ldots\right\}$ on the sphere $S(H^2(\mathbb{D}))$ is the same as that for $\{|c_0|^2, \ldots, |c_n|^2, \ldots\}$ on the unit sphere. Then there holds

Theorem 4.2. *Under the probability distribution α for X_1 and the i.i.d. symmetric distribution for X_2, the mathematical expectation $E(A_a)$ satisfies the relation*

$$E(A_a) = \frac{A(1-|a|^2)}{1-|a|^2(1-A)}$$

that takes the maximum value A at $a = 0$, and

$$A = \int_0^1 t\alpha(t)dt.$$

Proof. Define $L(r)$ by (4.1) and denote E_k^α the event $|c_0|^2 \in \left[\frac{k-1}{N}, \frac{k}{N}\right]$ that has the probability

$$\int_{\frac{k-1}{N}}^{\frac{k}{N}} \alpha(t)dt \approx \alpha\left(\frac{k}{N}\right)\frac{1}{N}.$$

Similarly to the proof of Theorem 4.1, we have

$$\begin{aligned}
L(r) &= \lim_{N\to\infty} \sum_{k=1}^N P(E_k^\alpha) \int_{X_1/E_k^\alpha} \sum_{n=0}^\infty |c_n|^2 r^{2n} P(d\rho/E_k^\alpha) \\
&= \lim_{N\to\infty} \sum_{k=1}^N \alpha\left(\frac{k}{N}\right)\frac{1}{N} \int_{X_1/E_k^\alpha} \left(\frac{k}{N} + r^2 \sum_{n=0}^\infty |c_{n+1}|^2 r^{2n}\right) P(d\rho/E_k) \\
&= \lim_{N\to\infty} \left(\sum_{k=1}^N \alpha\left(\frac{k}{N}\right)\frac{1}{N}\frac{k}{N} + r^2 \sum_{k=1}^N L(r)\left(1-\frac{k}{N}\right)\alpha\left(\frac{k}{N}\right)\frac{1}{N}\right) \\
&= \int_0^1 t\alpha(t)dt + r^2 L(r)\int_0^1 (1-t)\alpha(t)dt \\
&= A + r^2 L(r)(1-A).
\end{aligned} \tag{4.7}$$

Solving the equation for $L(r)$, we have

$$L(r) = \frac{A}{1-r^2(1-A)}.$$

AQ: Please check needs to rephrase?

Therefore,

$$\begin{aligned}\sup_{a\in\mathbb{D}} E(A_a) &= \sup_{a\in\mathbb{D}}(1-|a|^2)L(|a|)\\ &= A \sup_{r\in[0,1)} \frac{1-r^2}{1-r^2(1-A)}\\ &= A, \quad \text{at } r=0. \qquad \square\end{aligned}$$

Remark 4.1. The probability distribution of X_1 in Theorem 4.1 corresponds to $\alpha(t)\equiv 1$, and that for "$|c_n|^p, p\neq 2$, being evenly distributed in $[0,1]$" corresponds to $\alpha(t)=\frac{p}{2}t^{\frac{p-2}{2}}$. In the two cases, respectively, $A=\frac{1}{2}$ and $A=\frac{p}{p+2}$.

Remark 4.2. It is well known that better smoothness gives ~~implies~~ faster convergence of Fourier series. In the probability language this may be interpreted as $\alpha(t)$ having greater values nearby one. In the case A is close to one, and, by Theorem 4.2, the Fourier series has a faster convergence rate in the average sense.

5. Transformation Matrices Between T-M and Shifted Cauchy Kernel Systems

In Ref. 14, we show, for any given m-tuple $\{a_1,\ldots,a_n\}$,

$$\mathrm{Span}\{B_1,B_2,\ldots,B_n\}=\mathrm{Span}\{E_{\{a_1\}},E_{\{a_1,a_2\}},\ldots,E_{\{a_1,\ldots,a_n\}}\}, \tag{5.1}$$

where if $a_k\neq 0$ having multiplicity l in $\{a_1,\ldots,a_k\}$, then

$$E_{\{a_1,\ldots,a_k\}}=\frac{1}{(1-\overline{a}_k z)^l}, \quad l\geq 1$$

and if $a_k=0$ having multiplicity l in $\{a_1,\ldots,a_k\}$, then

$$E_{\{a_1,\ldots,a_k\}}=z^{l-1}, \quad l\geq 1.$$

The system

$$\{E_k\}_{k=1}^n=\{E_{\{a_1\}},E_{\{a_1,a_2\}},\ldots,E_{\{a_1,\ldots,a_n\}}\}$$

is called the *shifted Cauchy kernel system*, or the *Cauchy wavelet system* by some authors. Although it is not orthogonal, it has some advantage over the TM system $\{B_k\}_{k=1}^n$. For instance, if a real-valued signal s can be expressed by

$$s(e^{it})=\mathrm{Re}\sum_{k=1}^n c_k E_k(e^{it}),$$

which is easy to compute, then the Hilbert transform of $s(t)$ is

$$Hs(e^{it})=\mathrm{Im}\sum_{k=1}^n c_k E_k(e^{it}),$$

which is also easy to compute.

Proposition 5.1. *For arbitrary n, given a sequence $\{a_k\}_{k=1}^n$, denote $\mathbb{B}_n=\{B_k\}_{k=1}^{n\ T}$, $\mathbb{E}_n=\{E_k\}_{k=1}^{n\ T}$. Then the invertible transformation matrix T_n such* *that* $\mathbb{E}_n = T_n\mathbb{B}_n$ *is given by* $T_n = \{c_{kj}\}_{n\times n}$ *where*

$$c_{kj} = \frac{\sqrt{1-|a_j|^2}}{1-\bar{a}_k a_j}\prod_{i=1}^{j-1}\frac{\bar{a}_k - \bar{a}_i}{1-\bar{a}_k a_i},$$

when all $\{a_k\}$ *are distinct*; *or*

$$c_{kj} = \begin{cases} \overline{D^{q-1}[z^{q-1}B_j(z)](a_m)}, & a_m \neq 0, \\ \overline{\mathcal{D}^{(q-1)}[B_j(z)](0)}, & a_m = 0, \end{cases}$$

where m *and* q *are uniquely determined by* k. $\mathcal{D}^{q-1}$ *denoting the* $(q-1)$*th derivative, when* $\{a_k\}$ *has the multiplicity.*

Proof. There are two cases to consider.

Case (i). Let $\{a_k\}$ be a sequence of distinct points in $\mathbb{D}$. Since $\mathbb{B}_n$ is obtained from $\mathbb{E}_n$ through Gram–Schmidt procedure, for finite n, Span $\mathbb{B}_n$ = Span $\mathbb{E}_n$, and elements in $\mathbb{B}_n$ are orthogonal, so $E_k = \sum_{j=1}^{k} c_{kj}B_j$, where

$$\begin{aligned} c_{kj} &= \langle E_k, B_j\rangle \\ &= \overline{\langle B_j, E_k\rangle} \\ &= \overline{B_j(a_k)} \\ &= \frac{\sqrt{1-|a_j|^2}}{1-\bar{a}_k a_j}\prod_{i=1}^{j-1}\frac{\bar{a}_k - \bar{a}_i}{1-\bar{a}_k a_i}, \quad k = 1, 2, \ldots, n. \end{aligned} \tag{5.2}$$

Case(ii). When some a_k has multiplicity larger than one, the corresponding E_k changes. Suppose, for the given n, there are totally N different points $\{a_1, a_2, \cdots, a_N\}$, with l_m being the corresponding multiplicity of a_m, $l_1 + l_2 + \cdots + l_N = n$. In this case, Span $\mathbb{B}_n$ = Span $\mathbb{E}_n$ is irrelevant to the order of the points. We may set the order to be $\{a_1, \ldots, a_1, a_2, \ldots, a_2, \ldots, a_N, \ldots, a_N\}$, and, accordingly, $E_k = \sum_{j=1}^{k} c_{kj}B_j$, and

$$\begin{aligned} c_{kj} &= \langle E_k, B_j\rangle \\ &= \overline{\langle B_j, E_k\rangle}, \quad j \leq k. \end{aligned} \tag{5.3}$$

There exist some unique m and q such that $E_k = \frac{1}{(1-\bar{a}_m z)^q}$, $a_m \neq 0$ or $E_k = z^{q-1}$, $a_m = 0$, where $1 \leq q \leq l_m$. From Residue theorem, for $j \leq k$, for the first case,

$$\begin{aligned} c_{kj} &= \langle E_k, B_j\rangle \\ &= \overline{\langle B_j, E_k\rangle} \\ &= \overline{\frac{1}{2\pi}\int_0^{2\pi} B_j(e^{it})\frac{1}{(1-a_m e^{-it})^q}dt} \end{aligned}$$

$$= \overline{\frac{1}{2\pi i}\int_{z\in\mathbb{D}} B_j(z)z^{q-1}\frac{1}{(z-a_m)^q}dz}$$
$$= \frac{1}{(q-1)!}\overline{D^{(q-1)}[z^{q-1}B_j(z)](a_m)} \tag{5.4}$$

and, for the second case,

$$\begin{aligned} c_{kj} &= \langle E_k, B_j\rangle \\ &= \overline{\langle B_j, E_k\rangle} \\ &= \overline{\frac{1}{2\pi}\int_0^{2\pi} B_j(e^{it})\frac{1}{e^{i(q-1)t}}dt} \\ &= \overline{\frac{1}{2\pi i}\int_{z\in\mathbb{D}} B_j(z)\frac{1}{z^q}dz} \\ &= \frac{1}{(q-1)!}\overline{D^{(q-1)}[B_j(z)](0)}. \end{aligned} \tag{5.5}$$

In both cases, for $j > k$,

$$B_j \perp \mathrm{Span}\{B_1, \ldots, B_k\} = \mathrm{Span}\{E_1, \ldots, E_k\},$$

and thus $B_j \perp E_k, j > k$. So, $c_{kj} = 0, j > k$. Therefore, writing the n-dimensional vector $\mathbb{E}_n, \mathbb{B}_n$ in the matrix version, there exists

$$T_n = \begin{pmatrix} c_{11} & 0 & \cdots & 0 \\ c_{21} & c_{22} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix},$$

such that $\mathbb{E}_n = T_n\mathbb{B}_n$. Note that $c_{kk} \neq 0$ and T_n is invertible. $\square$

References

1. Q. Chen, L. Li and T. Qian, Stability of frames generated by nonlinear Fourier atoms, *Int. J. Wavelets, Multiresolut. Inf. Process.* **3** (2005) 465–476.
2. Q. Chen, L. Li and T. Qian, Two families of unit analytic signals with nonlinear phase, *Phys. D: Nonlinear Phenomena* **221** (2006) 1–12.
3. G. Davis, S. Mallat and M. Avellaneda, Adaptive greedy approximations, *Constructive Approx.* **13** (1997) 57–98.
4. R. A. Devore and V. N. Temlyakov, Some remarks on greedy algorithm, *Adv. Comput. Math.* **5** (1996) 173–187.
5. K. J. Lee, A simple lemma on greedy approximation in Hilbert space and convergence rates for projection pursuit regression and neural network training, *The Annals Statist.* **20** (1992) 608–613.
6. B. Ninness and F. Gustafsson, A unified construction of orthogonal bases for system identification, *IEEE Trans. Auto. Control* **42** (1997) 515–522.
7. B. Ninness, H. Hjalmarsson and F. Gustafasson, The fundamental role of general orthogonal bases in system identification, *IEEE Trans. Auto. Control* **42** (1999) 1384–1407.

8. B. Picinbono, On instantaneous amplitude and phase of signals, *IEEE Trans. Signal Process.* **45** (1997) 552–560.
9. T. Qian, Characterization of boundary values of functions in Hardy spaces with application in signal analysis, *J. Integral Equations Approx* **17** (2005) 159–198.
10. T. Qian, Analytic signals and harmonic measures, *J. Math. Anal. Appl.* **314** (2006) 526–536.
11. T. Qian, Mono-components for decomposition of signals, *Math. Methods Appl. Sci.* **29** (2006) 1187–1198.
12. T. Qian, Boundary derivatives of the phases of inner and outer functions and applications, *Math. Methods Appl. Sci.* **32** (2009) 253–263.
13. T. Qian, Q. Chen and L. Li, Analytic unit quadrature signals with nonlinear phase, *Phy. D: Nonlinear Phenomena*, **203** (2005) 80–87.
14. T. Qian and E. Wegert, Optimal approximation by Blaschke forms, *Complex Variables Elliptic Equations* DOI:10.1080/17476933.2011.557152.
15. T. Qian and Y. Wang, Adaptive Fourier series — A variation of greedy algorithm, *Adv. Comput. Math.* **34** (2011) 279–293.
16. V. N. Temlyakov, The best m-term approximation and greedy algorithm, *Adv. Comput. Math.* **8** (1998) 249–265.
17. J. L. Walsh, *Interpolation and Approximation by Rational Functions in the Complex Plain* (Amer. Math. Soc. Colloq. Publications, 1969).