Adaptive Fourier Decomposition and Rational Approximation, Part I: Theory

Tao Qian *

Abstract

In this note we will give a survey on adaptive Fourier decompositions in one- and multi-dimensions. The theoretical formulations of three different types of adaptive Fourier decompositions in one-dimension, viz., Core AFD, Cyclic AFD in conjunction with best rational approximation and Unwending AFD are provided.

Key Words Möbius Transform, Blaschke Form, Mono-component, Hardy Space, Adaptive Fourier Decomposition, Rational Approximation, Rational Orthogonal System, Time-Frequency Distribution, Digital Signal Processing, Uncertainty Principle, Higher Dimensional Signal Analysis in Several Complex Variables and the Clifford Algebra Setting

1 Introduction

In the one-dimensional cases by functional decomposition of the Fourier type we refer to approximations by the rational orthogonal systems, or alternatively, the Takenaka-Malmquist (TM) systems ([21]). The real-line case and the unit circle case are analogous. In the latter, a TM system is a collection of consecutively parameterized rational functions

$$B_k(z) = \frac{\sqrt{1-|a_k|^2}}{1-\overline{a}_k z}\prod_{l=1}^{k-1}\frac{z-a_l}{1-\overline{a}_l z},\ \ k=1,2,...,$$

where $a_1,...,a_k,...$ are all in the open unit disc. Such systems have been well studied. In particular, when all the a_k's are zero, the system $\{B_k\}$ reduces to a half of the Fourier system, viz., $\{z^{k-1}\}_{k=1}^{\infty}$. For general parameters a_k's the system $\{B_k\}$ is a basis in all $H^p, 1 \le p \le \infty$, if and only if the non-separable hyperbolic condition is satisfied, viz.,

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

*Department of Mathematics, University of Macau, Macao (Via Hong Kong). Email: fsttq@umac.mo. The work was partially supported by Research Grant of University of Macau MYRG115(Y1-L4)-FST13-QT, Macao Science and Technology Fund FDCT/098/2012/A3

The case where (1) is not satisfied corresponds to a remarkable decomposition of the Hardy spaces. In the Hardy H^2 space, that is what we concentrate on in this paper, a Blaschke product $\phi(z)$ may be defined that makes use $a_1, ..., a_k, ...$ as its all zeros, including the multiples, and the Hardy space is decomposed accordingly as

$$H^2 = \overline{\text{span}\{B_k\}} \oplus \phi H^2,$$

where $\overline{\text{span}\{B_k\}}$ is a *backward-shift invariant subspace*, and ϕH^2 is a *shift invariant subspace* of the H^2 space. These invariant spaces are of the Beurling- or the Beurling-Lax type ([6]), respectively, in the unit disc or the half complex plane contexts. A TM system consists of rational functions in the Hardy space that can approximate functions in the same Hardy space. Given the relation

$$f = 2\text{Re}f^+ - c_0,$$

where $f = f^+ + f^-$ is the Hardy space decomposition of $f \in L^2(\partial \mathbf{D})$,

$$f^\pm = \frac{1}{2}(f \pm iHf),$$

where H is the Hilbert transformation of the unit circle, we learn that approximation to the Hardy space functions implies that to the L^2 functions.

All the traditional studies and applications are based on the condition (1). The approximation effects are different for different selections of the parameters a_k's. The existing studies did not show how to select the parameters to get the optimal effect. Our studies devote to parameter selection according to the give function based on an energy principle. Under such selection, however, the condition (1) is not guaranteed. But in any case, under our selection of the parameters the given function falls into the subspace $\overline{\text{span}\{B_k\}}$.

This paper will present three algorithms: Core AFD, Unwending AFD and Cyclic AFD. The story is traced back to representations of signals into basic pieces with non-negative analytic phase derivative, viz., those with well defined instantaneous frequencies (IFs). To look into this direction the first author was motivated by the work of N. Huang et al on the EMD (Imperial Mode Decomposition) algorithm of signal decomposition. Huang in their work claim that the basic pieces obtained from EMD, called IMFs (Intrinsic Mode Functions), possess such desired IFs. That is to claim that, if f is an IMF, then under the amplitude-phase representation of the associated analytic signal $f + iHf = \rho e^{i\theta}, \rho \geq 0$, a.e., there holds $\theta' \geq 0$. But, unfortunately, this is not true ([**?**]). The non-negativity of the phase derivative is indeed a strong requirement. It is in particular related to conformal mappings, as well as Hardy space properties. Realizing this we carried on a fundamental study to understand what functions or signals f would have such property. To the authors' knowledge, the related queries might first arise in the signal analysis group in Syracuse University led by Yuesheng Xu and Lixin Shen. The study was joined by Qiuhui Chen and then Luo-Qing Li, and by the group in Zhongshan University and the group in the Chinese Academy of Science represented by, respectively, Li-Hua Yang and Dun-Yan Yan. The study also motivated a new phase of the studies of Bedrosian Identity. For relevant results the reader is referred to their respective work, as well as those of others, including Hai-zhang Zhang and Li-hui Tan.

We proceeded our study as follows. We named the sort of functions in the Hardy spaces H^2 as *mono-component*. Precisely, they are the functions whose analytic phase derivatives as measurable functions are non-negative. The basic examples are the monomials z^k, but there are many more. A rigorous definition of analytic phase derivative involves non-tangential boundary limits of holomorphic functions in the relevant domain [12], [5], [4]. Through a combined effort a large pool of mono-components is found. Next, one wishes to seek for apropriate decompositions of a signal into mono-components. The Fourier decomposition of Hardy space functions is a particular example. One faces two facts. One is that there are infinitely many ways to decompose a Hardy space function into mono-components. This is together with the observation that the faster is the convergence, the more stable the decomposition is. The second is that there exist mono-component decompositions that converge as fast as one can wish. These facts suggest that one restricts to certain types of mono-components proceeding the decomposition, as only in that case comparison between different signals in terms of composing mono-components make sense. The type that we choose for mono-component decomposition started from the weighted Blaschke products, or B_k, constituting TM systems.

Indeed, each B_k is a so called *pre-mono-component.* This is due to the fact that zB_k is a mono-component. Moreover, if letting $a_1 = 0$, then every B_k is a mono-component. A TM system has the following property: The phase derivatives of B_{k+1} is strictly larger than the phase derivatives of B_k. This implies that if B_k is a mono-component, then all $B_{k+m}, m \geq 0$, are mono-components.

Adaptive Fourier decomposition (AFD) offers fast decompositions into TM systems in which the parameters are selected consecutively according to the given signal. The fast decomposition is based on a maximal selection principle (see §1) together with a generalization of backward shift operator. In [9] the author propose an improvement of AFD, or Core AFD, called Unwending AFD. It incorporates at each recursive step a factorization process based on Nevanlinna's Factorization Theorem. Not only Unwending is faster, but also it treats the signals that are essentially of high frequencies. Unwending AFD, however, encounters computation of Hilbert transforms. There is another variation of AFD that uses higher order Szegö kernels, also treating signals of high frequencies. In the present paper, however, we do not pursue this particular direction. The next variation is Cyclic AFD that gives rise to a conditional solution of the open algorithm problem in finding a rational function of a given degree best approximating a Hardy space function.

Below we mention about higher dimensional generalizations. A similar theory, but not in full, is available in the quaternionic context. In the quaternionic space there is Möbius transformation. The Möbius transformation itself, however, is not a quaternionic regular (monogenic) function. As result, there are no Blaschke products. But there exists a Szegö kernel in either the unit ball or the upper half space context. In the complex number context there exist the useful relations

$$\langle f, B_k \rangle = \langle f_k, e_{a_k} \rangle = \langle \tilde{f}_k, B_k \rangle, \tag{2}$$

where for any $a \in \mathbf{D}$, e_a is the L^2-normalized Szegö kernel

$$e_a(z) = \frac{\sqrt{1-|a|^2}}{1-\overline{a}z} = n(a)z(a), \tag{3}$$

where $z(a)$ itself denotes the Szegö kernel or the reproducing kernel, and $\|z(a)\| = 1/n(a) = \frac{1}{\sqrt{1-|a|^2}}$, $\tilde{f}_k$ is the $(k-1)$-*th standard remainder*, realized by the *remainder operator*:

$$\mathcal{R}_{k-1}(f_1) = \tilde{f}_k = f - \sum_{l=1}^{k-1} \langle f, B_l \rangle B_l, \quad \tilde{f}_1 = f = f_1,$$

and f_k, called the $(k-1)$-*th reduced remainder*, is the result of applying the order $(k-1)$ generalized backward shift operator induced by $a_1, ..., a_{k-1}$, denoted as $\mathcal{B}_{k-1} = \mathcal{B}_{a_1,...,a_{k-1}}$. There hold the relations

$$\mathcal{B}_m \mathcal{B}_l = \mathcal{B}_{k+l} \quad \text{and} \quad \mathcal{B}_{k-1}(f_1) = f_k.$$

In particular, when $k = 2$,

$$\mathcal{B}_1(f)(z) = \frac{f(z) - \langle f, e_a \rangle e_a(z)}{\frac{z-a}{1-\overline{a}z}},$$

Due to the density property of the dictionary $\{e_a\}_{a\in\mathbf{D}}$ and the construction of the weighted Blaschke products the second equal relation in (2) is equivalent with

$$\mathcal{B}_{k-1} f_1 = f_k = \mathcal{R}_{k-1}(f_1) \prod_{l=1}^{k-1} \frac{z-a_l}{1-\overline{a}_l z}.$$

We note that $\{B_l\}_{l=1}^k$ is produced by the Gram-Schmidt orthogonalization process applied to $\{e_{a_l}\}_{l=1}^k$ (at least for the distinguished a_l's case, while the non-distinguished case is achieved by taking a limit approach):

$$B_k = \frac{\mathcal{R}_{k-1}(e_{a_k})}{\|\mathcal{R}_{k-1}(e_{a_k})\|}.$$

In a more general setting other than for complex numbers, due to the reproducing property of $z(a_k)$, the first equal relation in (2) implies

$$\langle f, \frac{\mathcal{R}_{k-1}(e_{a_k})}{\|\mathcal{R}_{k-1}(e_{a_k})\|} \rangle = \langle \frac{\mathcal{R}_{k-1}(f)}{\|\mathcal{R}_{k-1}(e_{a_k})\|}, e_{a_k} \rangle = n(a_k) \frac{\mathcal{R}_{k-1}(f)(a_k)}{\|\mathcal{R}_{k-1}(e_{a_k})\|}, \tag{4}$$

where

$$\|\mathcal{R}_{k-1}(e_{a_k})\|^2 = 1 - \sum_{l=1}^{k-1} |\langle e_{a_k}, B_l\rangle|^2.$$

The generalization to quaternions of the theory for complex numbers is based on (4) [15]. There is an obstacle for the theory being generalized to Clifford algebra due to the fact that for Clifford algebra-valued functions f the usually defined inner product $\langle f, f\rangle$ is not necessarily scaler-valued. The inner product being scalar-valued, however, is crucial in the Gram-Schmidt process. Nevertheless, an analogous theory in the Clifford algebra setting can be established under the idea of matching pursuit [19].

The traditional spherical harmonics expansion, viz., the Fourier-Laplace series expansion on the sphere falls into the same theoretical frame under the Clifford algebra setting. It is a realization of reproducing kernel representation of the Clifford Hardy H^2 space. On the unit sphere $\mathbf{S}^{n-1}$ of the Euclidean space $\mathbf{R}^n, n > 2$, for a scalar-valued or Clifford algebra-valued function of finite energy $f \in L^2(\mathbf{S}^{n-1})$, there holds, in the L^2 convergence sense,

$$f(x) = \sum_{k=0}^{\infty} f_k(x), \tag{5}$$

where for each k, f_k is a k-spherical harmonics. One can show that for $k > 0$ there holds the decomposition

$$f_k = P_k(f) + P_{-k-(n-2)}(f), \tag{6}$$

where for each l among $..., -n-1, -n, -n+1, 0, 1, 2, ...$, the Clifford algebra-valued function $P_l(f)$ is the boundary limit of a monogenic function in the ball that possesses homogeneity of degree l. The complex unit circle case corresponds to $n = 2$ the k-spherical harmonics decomposition is

$$\cos kt = \frac{1}{2}(e^{ikt} + e^{-ikt}).$$

The projection functions $P_l(f)$ are given by integral operators against the l-multiple Szegö kernels at the zero (see [17]). In the Clifford algebra setting of the Euclidean space things follow the same philosophy. The decomposition (5) under (6) becomes

$$f(x) = \sum_{k \neq -1, \ldots, -n+2} P_k(f)(x)$$

that is the corresponding Fourier decomposition whose adaptive forms are given in [15] and [19].

In the more traditional setting for multivariate functions, viz., the several complex variables setting, a similar theory can be established via tensor type products of the TM systems. Since there is no Laurent series in the context in view of the Hartogs Theorem, again, it is not dividable, and the analogue is not in full but only a half. We adopt the square partial sums setting with the multiple Fourier series on the cube, or, equivalently, on the torus We only explain our results for the 2-torus.

Let $\mathbf{a}$ denote a finite or infinite sequence $\{a_n\}$ of complex numbers $a_1, a_2, ...$ in the unit disc $\mathbf{D}$, and $\mathcal{B}^{\mathbf{a}}$ the finite or infinite TM system defined by the sequence $\mathbf{a}$, i.e.

$$\mathcal{B}^{\mathbf{a}} = \{B_{\{a_1,...,a_n\}}\} = \{B_n^{\mathbf{a}}\}.$$

If $\mathcal{B}_N^{\mathbf{a}}$ and $\mathcal{B}_M^{\mathbf{b}}$ are two finite TM systems, then $\mathcal{B}_N^{\mathbf{a}} \bigotimes \mathcal{B}_M^{\mathbf{b}}$ is an orthonormal system in $L^2(\mathbf{T}^2)$. When $\mathcal{B}^{\mathbf{a}}$ and $\mathcal{B}^{\mathbf{b}}$ are two bases of $H^2(\mathbf{T})$, then $\mathcal{B}^{\mathbf{a}} \bigotimes \mathcal{B}^{\mathbf{b}}$ is a basis of $H^2(\mathbf{T}^2)$.

Denote, for $f \in H^2(\mathbf{T}^2)$,

$$S_n(f) = \sum_{1\le k,l\le n} \langle f, B_k^{\mathbf{a}} \otimes B_l^{\mathbf{b}} \rangle B_k^{\mathbf{a}} \otimes B_l^{\mathbf{b}}, \quad D_n(f) = S_n(f) - S_{n-1}(f).$$

Note that $D_n(f)$ has $2n-1$ entries.

Based on such setting a maximal selection theorem is available, and a related adaptive decomposition can be established ([11]).

We have been explaining the theory of adaptive Fourier decomposition on the circles, the unit spheres and the n-torus. An analogous theory is available in the real-line and the half spaces that replace the integral Fourier transformations and offer fast series expansions in terms of the special functions in the respective contexts. Adaptive Fourier decompositions have strong backgrounds in both the theoretical and application aspects. Apart from the "shift analysis" aspect it also has close relation to rational function approximation (see [10] and §3). The AFD formulation is also related to compressed sensing and learning theory. In the application aspect we accomplished some studies in relation to system identification ([8], [7]), time-frequency distribution in signal analysis, speech analysis and distortion reversing in image processing in relation to harmonic mappings, etc.

In approximation one concerns convergence rates. In the classical settings convergence rates are in terms of degrees and types of smoothness. In AFD we treat general functions as boundary limits of Hardy space functions that can be non-smooth functions. We did obtain the convergence rate $O(1/n)$ under conditions that do not directly address smoothness ([14]) which, therefore, does not look nice. Some further study on this topic has been carrying out under the frame work of statistical learning theory.

The last point to make in the introduction part is rational approximation. Even in the one complex variable case the best rational approximation problem has not been solved, it is let alone the similar problems in the several complex variables and quaternionic and Clifford algebra variables contexts. To the authors knowledge, questions on rational function approximation to multi-variable functions in appropriate Hardy spaces have not been properly addressed. The proposed study based on reproducing kernels in various spaces would un-doubtably open new research and application directions.

In the following sections we will concentrate in introducing three types of AFD in the one complex variable.

2 Core AFD

AFD adaptively uses the TM system: the parameters a_k are selected according to the signals to be decomposed. By AFD a signal is decomposed in a fast way into a sum of mono-components or *pre-mono-components.* By a pre-mono-component we mean a signal that becomes mono-component after we multiple it by an exponential function of the form $\exp(iMt)$ with $M > 0$.

Suppose we are given a signal f in the Hardy H^2 space, that means

$$f(z) = f(z) = \sum_{l=0}^{\infty} c_l z^l, \quad \sum_{l=0}^{\infty} |c_l|^2 < \infty.$$

We seek for a decomposition into a TM system with selected parameters. Set $f = f_1$. For any complex a_1 in the unit disc we have the identity

$$f(z) =< f_1, e_{a_1} > e_{a_1}(z) + f_2(z)\frac{z - a_1}{1 - \overline{a}_1 z}, \tag{7}$$

with

$$f_2(z) = \frac{f_1(z) - < f_1, e_{a_1} > e_{a_1}(z)}{\frac{z-a_1}{1-\overline{a}_1 z}}.$$

We call the transformation from f_1 to f_2 the *generalized backward shift via* a_1, and, accordingly, f_2 *the generalized backward shift transform of* f_1 *via* a_1 (also called *reduced reminder*). The terminology was motivated by the classical backward shift operator

$$S(f)(z) = c_1 + c_2 z + \cdots + c_{k+1} z^k + \cdots = \frac{f(z) - f(0)}{z}.$$

Recognizing that $f(0) = \langle f, e_0 \rangle e_0(z)$, the operator S corresponds to our generalized backward shift operator via 0.

Due to the obvious orthogonality between the two terms on the right hand side of (7) and the unimodular property of Möbius transform, we have

$$\|f\|^2 = \|\langle f_1, e_{a_1} \rangle e_{a_1}\|^2 + \|f_2\|^2.$$

We are to extract the maximal energy portion from the term $\langle f_1, e_{a_1} \rangle e_{a_1}(z)$. Due to the reproducing kernel property of e_a, we have

$$\|\langle f_1, e_{a_1} \rangle e_{a_1}\|^2 = (1 - |a_1|^2)|f_1(a_1)|^2.$$

Therefore, to maximize $(1-|a_1|^2)|f_1(a_1)|^2$ is to minimize the remainder $\|f_2\|^2$. One can show without much difficulty that there exists a_1 in the open disc $\mathbf{D}$ such that

$$a_1 = \arg\max\{(1-|a|^2)|f_1(a)|^2 \ : \ a \in \mathbf{D}\}$$

([14]). The existence of such maximal selection is called the *Maximal Selection Principle*. Under a maximal selection of a_1 we call the decomposition (7) a *maximal sifting*. Having selected such a_1, repeating the same process for f_2, and so on, we obtain, after the n-th step,

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

where for $k = 1, ..., n$,

$$a_k = \arg\max\{(1-|a|^2)|f_k(a)|^2 \ : \ a \in \mathbf{D}\},$$

and, for $k = 2, ..., n+1$,

$$f_k(z) = \frac{f_{k-1}(z) - < f_{k-1}, e_{a_{k-1}} > e_{a_{k-1}}(z)}{\frac{z-a_{k-1}}{1-\overline{a}_{k-1} z}}.$$

It can be shown that

$$\lim_{n\to\infty} \|f_{n+1}\| = 0.$$

Thus we have

$$f(z) = \sum_{k=1}^{\infty} < f_k, e_{a_k} > B_k(z)$$

([14]).

Remark 1 We note that the parameters selected under the maximal selection principle do not necessarily satisfy the condition (**??**), and thus $\{B_k\}$ does not have to be a basis. In some applications, however, one may be interested only in fast expanding of a given signal.

Remark 2 If we choose $a_1 = 0$, then all B_k are mono-components, and AFD offers a mono-component decomposition. For arbitrary selections of $a_1, ..., a_n, ...$, the B_k are pre-mono-component decomposition, and after being multiplied by e^{it} all the entries in the infinite sum become mono-components.

Remark 3 AFD is different from greedy algorithm or orthogonal greedy algorithm in relation to the dictionary consisting of Szegö kernels. Besides a maximal selection at each step dictionary words can be repeatedly selected in order to guarantee effective approximation in the process.

Remark 4 The convergence rate for the n-th AFD partial sum is $1/\sqrt{n}$. This is a convergence rate for rather general functions in the Hardy space.

3 Best Approximation by Rational Functions of Order Not Larger Than n

Core AFD and Unwending AFD offer fast decomposition of signals into mono-components. There is a third one comparable with them called *Cyclic AFD* that offers best approximation to Hardy space functions by rational functions of orders less than or equal to a given integer n. Such approximation has a better stability or uniqueness. The best rational approximation is equivalent with simultaneous selection of n-parameters $a_1, ..., a_n$, in a Blaschke form, viz.,

$$\sum_{k=1}^{n} < f, B_k > B_k(z).$$

In AFD, however, it is one by one selection of parameters but not simultaneous. The existence of such approximation has long been proved, but a practical algorithm of it is still an open problem until now.

Let p and q denote polynomials of one complex variable. We say that (p, q) is an *n-pair* if p and q are co-prime whose degrees both are less than or equal to n. We further require that the zeros of q are all outside the closed unit disc. Denote the set of all n-pairs by $\mathcal{R}(n)$. If $(p, q) \in \mathcal{R}(n)$, then the rational function p/q is said to be a rational function of degree less or equal n. Let f be a function in the Hardy H^2 space in the unit disc. The best n-rational approximation is to find an n-pair (p_1, q_1) such that

$$\|f - p_1/q_1\| = \min\{\|f - p/q\| \ : \ (p, q) \in \mathcal{R}_n\}.$$

What we propose, called *Cyclic* AFD *algorithm*, can get a solution of the best n-rational approximation when there is only one critical point for the problem ([10]). We call such a solution a *conditional solution*. Compared with the existing RARL2 algorithm, that offers also a conditional solution ([1], [2], [3]), Cyclic AFD is more explicit, and can directly find the poles of the approximating rational function.

We will call

$$\sum_{k=1}^{n} c_k B_k(z)$$

an n-Blaschke form where the parameters $a_1, ..., a_n$ of B_n are arbitrary complex numbers in $\mathbf{D}$. An n-Blaschke form is said to be non-degenerate if $c_n \neq 0$. It is easy to see that a non-degenerate n-Blaschke form is either an n-rational function or an $(n-1)$-rational function, depending on whether 0 is one of the parameters defining B_n. This shows a little inconsistence with n-rational functions. But if we work with the parallel context the region outside the unit disc, then the class of the n-Blaschke forms corresponds exactly with the class of the n-rational functions. To simplify the writing we ignore the little inconsistence and still work inside the unit disc.

For any given natural number n the objective function for the problem is set to be

$$A(f; a_1, ..., a_n) = \|f\|^2 - \sum_{k=1}^{n} |\langle f, B_k\rangle|^2. \quad (8)$$

We need a few more new terminology. Assume that $f \in H^2$ and f is not an m-Blaschke form for any $m < n$. Then $a_1, ..., a_n$ is said to be a *coordinate-minimum point* of $A(f; z_1, ..., z_n)$ if for any permutation P whenever fixing $n-1$ points $(z_1, ..., z_{n-1}) = (Pa_1, ..., Pa_{n-1})$ and performing the Maximal Selection Principle to $|\langle f_n, e_{z_n}\rangle|$ for the remaining complex variable z_n, then the missing point Pa_n is one of the optimal choices for z_n.

In the AFD algorithm we proceed the procedure with the increasing k : Along with choosing $a_1, ..., a_{k-1}$ in $\mathbf{D}$, we produce the reduced remainders $f_2, ..., f_k$. Then to f_k we apply the Maximal Selection Principle to find an a_k giving rise to $\max\{|\langle f_k, e_a\rangle| : a \in \mathbf{D}\}$. The Cyclic AFD Algorithm repeats such procedure at the n-th step: Whenever $a_1', ..., a_{n-1}'$ are fixed from previous steps we inductively obtain the reduced remainders $f_2, ..., f_n$, and then use the Maximal Selection Principle to select an optimal a_n'.

Denote by *LMP* a local minimum points, by *CMP* a coordinate-minimum point, and *CP* a critical point of the objective function. Denote by $\mathcal{LM}, \mathcal{CM}$ and $\mathcal{C}$ the sets, of, respectively, all LMPs, CMPs and CPs. Then we have the inclusion relations

$$\mathcal{LM} \subset \mathcal{CM} \subset \mathcal{C}. \quad (9)$$

The proposed cyclic AFD algorithm is contained in the following procedure.

Suppose that f is not an m-Blaschke form for any $m < n$. Let $s_0 = \{b_1^{(0)}, ..., b_n^{(0)}\}$ be any n-tuple of parameters inside $\mathbf{D}$. Fix some $n-1$ parameters of s_0 and make an optimal selection of the single remaining parameter according to the Maximum Selection Principle. Denote the obtained new n-tuple of parameters by s_1. We repeat this process and make cyclic optimal selections over the n parameters. We thus obtain a sequence of n-tuples $s_0, s_1, ..., s_l, ...,$ with decreasing objective function values d_l that tend to a limit $d \geq 0$, where

$$d_l = A(f; b_1^{(l)}, ..., b_n^{(l)}) = \|f\|^2 - \sum_{k=1}^{n} (1 - |b_k^{(l)}|^2) |f_k^{(l)}(b_k^{(l)})|^2. \quad (10)$$

Then, (i) If $\overline{s}$, as an n-tuple, is a limit of a subsequence of $\{s_l\}_{l=0}^{\infty}$, then $\overline{s}$ is in $\mathbf{D}$; (ii) $\overline{s}$ is a CMP of $A(f; \cdots)$; (iii) If the correspondence between a CMP and the corresponding value of $A(f; \cdots)$ is one to one, then the sequence $\{s_l\}_{l=0}^{\infty}$

itself converges to the CMP, being dependent of the initial n-tuple s_0; (iv) If $A(f;...)$ has only one CMP, then $\{s_l\}_{l=0}^{\infty}$ converges to a limit $\overline{s}$ in $\mathbf{D}$ at which $A(f;\cdots)$ attains its global minimum value.

For further details including examples on Cyclic AFD we refer the reader to [10].

4 Unwending AFD

Assume that $f = hg$, where f, g are functions in the Hardy H^2 space, h is an inner function. Expand f and g into their individual Fourier series:

$$f(z) = \sum_{k=1}^{\infty} c_k z^k, \qquad g(z) = \sum_{k=1}^{\infty} d_k z^k.$$

Under such circumstance, in digital signal processing (DSP), there is the following assertion: For any n,

$$\sum_{k=n}^{\infty} |c_k|^2 \geq \sum_{k=n}^{\infty} |d_k|^2,$$

see, for instance [5]. This amounts to say that, after factorizing an inner function factor the remaining Hardy space function converges faster. This suggests that in the above Core AFD procedure if one combines a factorization process then the convergence becomes faster. This is reasonable: when a signal by its nature is of high frequency, one should "unwending" it but not try first to get maximal portions corresponding to the lowest frequencies. When this idea is implemented the AFD is amended as follows ([10], [13]). First we do the factorization $f = f_1 = I_1 O_1$, where I_1 and O_1 are, respectively the inner and outer function parts of f. The factorization is based on Nevanlinna's factorization theorem and the outer function has the explicit integral representation

$$O_1(z) = e^{\frac{1}{2\pi}\int_0^{2\pi} \frac{e^{it}+z}{e^{it}-z} \log|f_1(e^{it})|dt}.$$

In the algorithm we need to use the boundary value of O_1 to compute the above integral in the principal integral sense and the imaginary part reduces to the circular Hilbert transform of $\log|f_1(e^{it})|$. Next we do a maximum sift to O_1 that means the decomposition of O_1 under the maximal selection principle:

$$f(z) = I_1(z)[< O_1, e_{a_1} > e_{a_1}(z) + f_2(z)\frac{z - a_1}{1 - \overline{a}_1 z}],$$

where f_2 is the backward shift of O_1 via a_1 :

$$f_2(z) = \frac{O_1(z) - < O_1, e_{a_1} > e_{a_1}(z)}{\frac{z-a_1}{1-\overline{a}_1 z}}.$$

By factorizing f_2 into its inner and outer parts, $f_2 = I_1 O_2$, we have

$$f(z) = I_1(z)[< O_1, e_{a_1} > e_{a_1}(z) + I_2(z) O_2(z) \frac{z - a_1}{1 - \overline{a}_1 z}].$$

Next, we do a maximum sift to O_2, and so on. In such way we obtain the Unwending AFD decomposition ([9])

$$\begin{aligned} f(z) &= \sum_{k=1}^{n} \prod_{l=1}^{k} I_l(z) < O_k, e_{a_k} > B_k(z) + f_{n+1}(z) \prod_{k=1}^{n} \frac{z - a_k}{1 - \overline{a}_k z} \prod_{l=1}^{n} I_l(z) \\ &= \sum_{k=1}^{\infty} \prod_{l=1}^{k} I_l(z) < O_k, e_{a_k} > B_k(z), \end{aligned}$$

where $f_{k+1} = I_{k+1} O_{k+1}$ is the backward shift of O_k via $a_k, k = 1, ..., n$, and I_{k+1} and O_{k+1} are respectively the inner and outer functions of f_{k+1}.

Remark 5 In most cases an Unwending AFD is automatically a mono-component decomposition due to the fact that inner functions has positive phase derivatives. In general, we set $a_1 = 0$ to guarantee that Unwending AFD gives rise to a mono-component decomposition. Unwending AFD converges very fast, and considerably faster than Core AFD. This is shown through the relevant examples especially on singular inner functions ([13]).

Remark 6 There are other AFD-variations that first extract factor signals of high frequencies. Those include *Double-sequence unwending AFD* ([16] and one using what we call high-order Szegö kernels ([18]). The algorithm of double-sequence Unwending AFD is more complicated than that of Unwending AFD but with a similar performance as the latter. The high-order Szegö kernel method does the selection following the principle of greedy algorithm that does not have a generalized backward shift feature.

References

[1] L. Baratchart, A remark on uniqueness of best rational approximation of degree 1 in L^2 of the circle, Electronic Transactions on Numerical Analysis, Volume 25, pp. 54-66, 2006.

[2] L. Baratchart, M. Cardelli, M. Olivi, Identification and rational L^2 approximation, a gradient algorithm, Automatica, 27(1991), pp. 413-418.

[3] L. Baratchart, E.B. Saff, F. Wielonsky, Non-uniqueness of rational approximations, J. Comput. Appl. Math., 105:141-154, 1999.

[4] P. Dang and T. Qian, *Analytic Phase Derivatives, All-Pass Filters and Signals of Minimum Phase,* IEEE Transactions on Signal Processing, Issue Date: Oct. 2011, Volume: 59 Issue:10, on page(s): 4708 V 4718.

[5] P. Dang, T. Qian and Z. You, *Hardy-Sobolev spaces decomposition and applications in signal analysis,* J. Fourier Anal. Appl. 17 (2011), no. 1, 36V64.

[6] J.B. Garnett, Bounded Analyic Functions, Academic Press, 1981.

[7] W. Mi and T. Qian, *Frequency Domain Identification: An Algorithm Based On Adaptive Rational Orthogonal System,* Automatica, 48(6). pp. 1154-1162.

[8] W. Mi, T. Qian, F. Wan, A Fast Adaptive Model Reduction Method Based on Takenaka-Malmquist Systems, by W. Mi, T. Qian and F. Wan, Systems and Control Letters, Volume 61, Issue 1, January 2012, Page 223-230.

[9] T. Qian, *Intrinsic mono-component decomposition of functions: An advance of Fourier theory,* Mathematical Methods in Applied Sciences, 2010, 33, 880-891, DOI: 10.1002/mma.1214.

[10] T. Qian, *Cyclic AFD Algorithm for Best Approximation by Rational Functions of Given Order,* accepted by Mathematical Methods in the Applied Sciences.

[11] T. Qian, *Adaptive Fourier decomposition on torus,* preprint.

[12] T. Qian, *Mono-components for decomposition of signals,* Mathematical Methods in the Applied Sciences (Zeitschrift fur Analysis and ihre Anwendungen), 2006;29:1187-1198.

[13] T. Qian, H. Li, M. Stessin , *Comparison of Adaptive Mono-component Decompositions,* Nonlinear Analysis: Real World Applications, 2013 **14** (2), 1055-1074.

[14] T. Qian, Yanbo Wang, *Adaptive Fourier Series-A Variation of Greedy Algorithm*, Advances in Computational Mathematics, 34(2011), no.3, 279-293.

[15] T. Qian, W. Sproessig and J-X. Wang, *Adaptive Fourier decomposition of functions in quaternionic Hardy spaces,* Mathematical Methods in the Applied Sciences. (35) 2012, 43V64.DOI: 10.1002/mma.1532.

[16] T. Qian, L.H. Tan and Y.B., *Wang,Adaptive Decomposition by Weighted Inner Functions: A Generalization of Fourier Serie,* J. Fourier Anal. Appl. 17 (2011), no. 2, 175V190.

[17] T. Qian and E. Wegert, *Optimal Approximation by Blaschke Forms,* Complex Variables and Elliptic Equations, Volume 58, Issue 1, 2013, page 123-133.

[18] T. Qian and J-X. Wang, *Adaptive Decomposition of Functions by Higher Order Szegö Kernels I. A Method for Mono-component Decomposition,* preprint.

[19] Tao Qian, Jinxun Wang and Yan Yang, *Matching Pursuits among Shifted Cauchy Kernels in Higher-Dimensional Spaces,* accepted to appear in Acta Mathematica Sinica.

[20] T. Qian, Liming Zhang and Zhi-Xiong Li, *Algorithm of Adaptive Fourier Decomposition,* IEEE Transaction on Signal Processing, Dec., 2011, Volume 59, Issue 12, page 5899-5902.

[21] J.L. Walsh, *Interpolation and Approximation by Rational Functions in the Complex Domain,*