Appl. Math. J. Chinese Univ.
2013, 28(4): 505-530

Mathematical theory of signal analysis vs. complex analysis method of harmonic analysis

QIAN Tao[1] ZHANG Li-ming[2]

Abstract. We present recent work of harmonic and signal analysis based on the complex Hardy space approach.

§1 Introduction

Professor Gong Sheng shared with the first author his view: "On the unit circle it is harmonic analysis, and inside the unit circle it is complex analysis" ([16]). This is, in fact, common sense among analysis. In general, we regard the following as what we mean by complex method of harmonic analysis. Suppose that one is to study analysis on a closed and finite-dimensional manifold. One can then imbed the manifold into a space of one more (or several more) dimension (dimensions) with a complex analysis structure, and, in such way, one treats the manifold understudy as a co-dimension 1 (or co-dimension p) space. By a complex structure it means that there exist a Cauchy kernel and a Cauchy formula, and the related complex analysis objects. With the complex structure one can define Hardy spaces of good complex holomorphic functions in the regions enclosed by the manifold. Functions on the manifold then can be split into a sum of the boundary limits of the related Hardy space functions. Those boundary limits constitute boundary Hardy spaces. Such idea was, in fact, taught by M.-T. Cheng and D.-G. Deng when the first author was in his Ph.D. program in Beijing University around 1980. It is also hinted by the book of Gorusin translated by Jian-Gong Chen ([17]). The author also learned this idea from the works by C. Kenig and, separately, by A. McIntosh on complex Hardy spaces and singular integrals on Lipschitz curves and surfaces. This article gives a survey on the results that the authors and their collaborators obtained by implementing the complex analysis approach to signal analysis.

It has been a controversial issue to the present time about what is *instantaneous frequency* (IF), or, in brief, *frequency*. People tend to believe for a general signal there is a certain frequency

Received: 2013-10-16.
MR Subject Classification: 42A50, 32A30, 32A35, 46J15.
Keywords: 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.
Digital Object Identifier(DOI): 10.1007/s11766-013-3225-4.
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.

at any moment of time. This belief is hinted and supported by sinusoidal signals that possess the constant frequencies. To justify this idea is to define what is frequency. For a general signal a well acceptable and reasonable way to define instantaneous frequency has not been found. Our view is that there does not exist a frequency concept for a general signal. Our work proposes a definition under which some signals have well defined instantaneous frequencies, and some not. Signals possessing IF are called *mono-components*, and otherwise, *multi-components*. For multi-components one seeks for mono-component decompositions. We now review the story started from Gabor.

In 1946 Gabor proposed his *analytic signal* approach ([15]). Throughout this article we restrict ourselves to only signals with finite energy, or L^2-functions. Let $s(t)$ be a real-valued signal of finite energy. The associated analytic signal, denoted by $s^+(t)$, is defined as

$$s^+(t) = \frac{1}{2}\left(s(t) + iHs(t)\right),$$

where H is the Hilbert transformation. We note that analytic signals are non-tangential boundary limits (the Plemelj formula) of Hardy space functions in the related domain, the latter consisting of holomorphic functions with good control close to the boundary. For the real line case the related domain is the upper-half complex plane. The Hardy space functions in the domain are given by the Cauchy integral of the signal $s(t)$ on the boundary. We note that through out the paper, except in the final section for multivariate signals, we use the terminology Hardy space only for the complex Hardy H^2 spaces in either the contexts inside or outside of the unit disc, or the contexts of the upper- or lower-half complex planes. There are essentially parallel theories in the four contexts. We will feel free in below to switch from one context to another. From paragraph to paragraph we will make sure that we give clear indication to which context we are referring. We will use the notations $L^2(\mathbf{R}), H^2(\mathbf{C}^+), L^2(\partial\mathbf{D}), H^2(\mathbf{D})$, etc., where $\mathbf{R}, \mathbf{C}, \mathbf{D}$ and $\mathbf{C}^+$ denote, respectively, the real line, the complex plane, the unit disc and the upper-half- complex plane. When we use H^2 and L^2 we mean that we refer to all the four contexts.

On the real line, the Fourier transform of Hs is $-i\mathrm{sgn}(\cdot)\hat{s}(\cdot)$. As consequence, the Fourier transform of s^+ is supported on $[0,\infty)$, and, in particular,

$$s^+(t) = \frac{1}{2\pi}\int_0^\infty e^{it\xi}\hat{s}(\xi)d\xi.$$

This shows that s^+ is a "linear combination" of some terms of non-negative frequencies, viz., of those $e^{it\xi}$ with $\xi \geq 0$. It hence has reason to believe that in a single-term-amplitude-phase representation, viz., $s^+(t) = \rho(t)e^{i\theta(t)}$, one should have that the *phase derivative of* s^+, or alternatively, the *analytic phase derivative of* s, satisfies $\theta'(t) \geq 0$. But, unfortunately, this is not true. To make clear the terminology, we note that if $s = s^+$, then $s^{++} = s$. This amounts that analytic phase derivatives of boundary values of the Hardy space functions coincide with the phase derivatives of the functions. It is a fact that phase derivatives of any non-trivial analytic outer function are negative in a set of positive Lebesgue measure [26]. Such examples can be simply constructed as follows. Consider a fractional linear transform, f, in the complex plane that maps the unit disc centered at the origin onto a disc that dose not contain the origin in its topological closure. Restricted to the unit circle the mapping has an amplitude-phase representation $f(e^{it}) = \rho(t)e^{i\theta(t)}$. One observes that $\theta(t)$ is not a monotone function in the range $[0, 2\pi]$ when t traverses from 0 to 2π. But, instead, in an interval the phase derivative strictly increases and in an adjacent interval strictly decreases. This implies that the phase derivative $\theta'(t)$ changes sign in a pair of adjacent open intervals, and, in particular, negative in an open interval. Such functions can be constructed from their corresponding real parts by Gabor's analytic signal method.

Why positivity of frequency of a signal is important? The primary importance is that the frequency concept is generated from physical practice: it is an extension of vibrating frequency. Secondly, the positivity has a great significance in signal analysis. For instance, the mean of frequency of a real valued signal is zero if we do not restrict to positive frequencies, that makes the mean concept to be useless.

We note that the Hilbert transformation in various contexts play a crucial role when dealing with boundary values of holomorphic functions. We adopt the definition of Hilbert transformation from S. Bell ([3]). Suppose that we deal with a simply-connected domain Ω in the complex plane. Let f be a holomorphic function in the domain. Assume that f has a non-tangential boundary limit, while this is always true when f is in the Hardy space of the domain. Denoting the boundary limit of f by $f = u + iv$, where u and v are scalar-valued. Then the Hilbert transformation H is defined through $H\colon u \to v$, or $v = Hu$. This mapping in some cases should modulo a constant. In different contexts the Hilbert transformation has different representation. On the real line it is given by the principal value singular integral with the kernel $\frac{1}{\pi}\frac{1}{t}$, or, in the inverse Fourier transform formulation, given by the Fourier multiplier $-i\text{sgn}(\xi)$. In the unit circle case it is the so called *circular Hilbert transformation*, or induced by the same Fourier multiplier via Fourier series expansion. The general Hilbert transformation on manifolds in higher dimensional Euclidean spaces may be defined similarly via the Clifford algebra formulation ([1], [46]).

The first task of the study is to find a pool of the functions that have *non-negative analytic phase derivative*. Precisely, we are to find signals $s(t) \in L^2$ such that $s^+(t) = \rho(t)e^{i\theta(t)}$ has non-negative phase derivative, viz., $\theta'(t) \geq 0$. We call such signals s *mono-components* or *real-mono-component*, and, without ambiguity, call s^+ *mono-component*, too, and sometimes *complex-mono-component* [27], [28].

Definition 1. (Tao Qian 2006) *Let s be a real- or complex-valued signal with finite energy. We call s a mono-component if its analytic signal, or its projection into the Hardy space H^2, viz., $s^+(t) = \frac{1}{2}\left(s(t) + iHs(t)\right)$, in its phase-amplitude representation $s^+(t) = \rho(t)e^{i\theta(t)}$ satisfies $\theta'(t) \geq 0$, where the phase derivative $\theta'(t)$ is defined through the non-tangential limit of the same quantity but inside the region. Precisely, in the unit circle case,*

$$\theta'(t) = \lim_{r\to 1-} \theta_r'(t), \quad s^+(re^{it}) = \rho_r(t)e^{i\theta_r(t)}, \quad 0 < r < 1;$$

and, in the upper plane case,

$$\theta'(t) = \lim_{y\to 0+} \theta_y'(t), \quad s^+(t + iy) = \rho_y(t)e^{i\theta_y(t)}, \quad y > 0.$$

s is called a normal mono-component if it is mono-component and, further more, there exists $1 > \delta > 0$ such that $\theta_r'(t) \geq 0$, if for all t and $1 > r > 1 - \delta$, in the unit disc case; or exists

$h > 0$ *such that* $\theta'_y(t) \geq 0$, *if for all* t *and* $0 < y < h$, *in the upper-half plane case.*

We note that if s is a real-valued signal, then $s = 2\text{Re}s^+$ or $s = 2\text{Re}s^+ - c_0$ in the respective two cases. This, in particular, shows that an approximation to s^+ leads to an approximation of s.

There was a march to find various types of mono-components by several research groups. The group led by Yue-Sheng Xu at Syracuse University was among the first to study the related subjects. Many of the researchers, including Xu's group, were motivated by Norden Huang's algorithm EMD (*Empirical Mode Decomposition*) ([18]). The clearly exposed idea of adaptive decomposition of signals into basic signals of positive frequency in the paper is brilliant, and, as a matter of fact, has been motivating many studies. From the mathematical point of view, however, there are mainly two obstacles with EMD: One is the convergence of the algorithm, and the other is analytic property of the obtained "basid signal" IMF (*Intrinsic Mode Function*): It is claimed by [18] that IMFs are what we defined mono-components, but actually not the case, proved by [48]. In [48] it is shown that an IMF does not necessarily possess a.e. non-negative analytic phase derivative. In principle, since the algorithm throws away errors of unknown types, the convergence is a problem, and the resulted signals cannot be guaranteed to have good analytic properties. It is an engineering algorithm to obtain a discrete signal from discrete data. Besides Syracuse University, researchers in Beijing University, Beijing Normal University, the Chinese Academy of Sciences, Hubei University, University of Macau, Zhongshan University, etc., also participated to the march of finding mono-components.

To the author's knowledge it was Yuesheng Xu who first proposed, around 2005, to characterize all the amplitude functions $\rho \geq 0$ in a mono-component signal $\rho e^{i\theta}$ in which the phase signal itself, viz., $e^{i\theta}$, is already a mono-component ([58], [24]). The characterization is in terms of the phase function θ. This can be interpreted as amplitude retrieving problem (See Subsection 5 below). The product form of the amplitude-phase representation leads to a new phase of the study of Bedrosin identity that deals with the conditions on f and g to ensure

$$H(fg) = fHg,$$

where H is the Hilbert transformation of the context. Indeed. a signals f is an analytic signals if and only if $Hf = -if$. If we already have $H(e^{i\theta}) = -ie^{i\theta}$, in order to also have $H(\rho e^{i\theta}) = -i\rho e^{i\theta}$, a sufficient condition is $H(\rho e^{i\theta}) = \rho H(e^{i\theta})$. This question now is better understood, and a complete solution can be drown when the phase signal $e^{i\theta}$ is from a Blaschke product. For general inner functions this is still open. Qian's result in 2009 [26] shows that if $e^{i\theta}$ is the non-tangential boundary of an inner function, then $\theta' > 0$, a.e. The proof is an application of the Julia-Wolff-Carathéodory Theory. We have

Theorem 1. (Tao Qian 2009) *Let* θ *be a Lebesgue measurable function. Then the phase function* $e^{i\theta}$ *is a mono-component if and only if* $e^{i\theta}$ *is the non-tangential boundary limit of an inner function, or, equivalently, if and only if* $H(e^{i\theta}) = -ie^{i\theta}$.

The above theorem is valid in both the unit circle and the real contexts.

Instead of listing all the important references in relation to the recent developments of Bedrosian identity studies I list the main relevant authors. An incomplete list includes, in the alphabetical order, Q.H. Chen, T. Qian, L.H. Tan, R. Wang, S.L. Wang, Y.S. Xu, D.Y. Yan, L.X. Yan, L.H. Yang, B. Yu, H. Z. Zhang. However, only a moderate percentage of the relevant literature devote to solve the mentioned mono-component problem ([59], [45], [50], etc.). Most, in fact, develop their own interests in finding out conditions for the Bedrosian identity. We have the following basic result: If the phase function part is from a finite or infinite Blaschke product, then the desired amplitude part has to be in the closure of the linear span of the related Takenaka-Malmquist (TM) system, that is a backward-shift invariant subspace of H^2 ([54]). Precisely, we have

Theorem 2. *If*

$$e^{i\theta(t)} = \prod_{k=1}^{\infty} \frac{-\overline{a}_k}{|a_k|} \frac{e^{it} - a_k}{1 - \overline{a}_k e^{it}},$$

then $\rho(t)e^{i\theta(t)} \in H^p(\mathbf{D}), 1 \le p \le \infty$, *if and only if*

$$\rho \in \overline{\text{span}^p}\{B_k\}_{k=1}^{\infty},$$

where

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

Note that a TM system is orthonormal under the inner product

$$< f, g >= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(e^{it})\overline{g}(e^{it})dt$$

([56]). It becomes a basis in H^2 if and only if the non-separable hyperbolic distance condition

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

holds. A basic function B_k in a TM system, called a *weighted Blaschke product*, consists of two parts of which one is a Blaschke product $\prod_{l=1}^{k-1} \frac{z-a_l}{1-\overline{a}_l z}$, being a product of $k-1$ Möbius transforms that is a mono-component, and the other part is the classical Szegö kernel, being an outer function. Szegö kernel is the reproducing kernel of the Hardy H^2 space. A detailed analysis show that if $a_l = 0$, then all the basic functions $B_l, B_{l+1}, \ldots, B_{l+k}, \ldots$ are mono-components.

The stated result for infinite Blaschke products gives rise to sufficient and necessary conditions on real and non-negative amplitudes ρ for $\rho e^{i\theta}$ being new mono-components ([54]). The relation between weighted Blaschke products and p-starlike function in complex analysis functions are indicated in [55].

A great variety of mono-components can be identified based on Theorem 1 and the Bedrosian identity results. The next question is how to express a signal in terms of mono-components. Before giving our answer we first draw the reader's attention to some observations and arguments. First, if a function f in the Hardy space $H^2(\mathbf{D})$, then we can show, for any given $\epsilon > 0$, there exist two mono-components, m_1 and m_2 such that([34])

$$\|f - (m_1 - m_2)\|_2 < \epsilon.$$

We show that m_1 and m_2 can be chosen as boundary values of starlike functions. The construction of the approximating starlike functions do not reflect any frequency nature of the function f being approximated. This shows that seeking for "arbitrary fastest" decomposition without restricting to a certain type of mono-components does not contribute to understanding of frequency. On the other hand, if a decomposition is restricted to a certain type of mono-component and requires to have a small number of terms, then the decomposition is stable: the extremal case is, if a function itself is a mono-component of such type, then the decomposition into a single mono-component term will be unique. Based on such observation one would restrict to certain type of mono-components, and at the same time seek for fast decomposition. For a practical multi-component a best decomposition up to some fixed n terms should be unique.

Below we give three types of mono-component decompositions. They are all based on weighted Blaschke product type, and give rise to fast decompositions. They are of of individual merits in signal analysis. Approximating by using rational functions cannot avoid TM systems. TM systems have the advantages of simplicity and adaptability. A weighted Blaschke product B_k is often a mono-component; and, if not, can be easily modified to become a mono-component. It is based on the fact that the analytic phase derivative of each of its factor Möbius transforms is a Poisson kernel ([25], [32]).

Division of work between the two authors is as follows. The first author is responsible for the basic writing of the paper, while the second author is responsible for all the experiments with diagrams.

§2 Mono-component decompositions of the AFD type

1. **AFD (Core AFD)**

We use the abbreviation AFD for *Adaptive Fourier Decomposition.* 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 being multiplied an exponential function of the form $\exp iMt$ with some $M > 0$. For any parameters $a_1, \dots, a_l$ in the unit disc, all $B_k, k = 1, \dots, l$, are pre-mono-components. If $a_k = 0$, then $B_k, B_{k+1}, \dots,$ are mono-components. If all $a_1, \dots, a_l, \dots$ are zero, then the corresponding TM system becomes Fourier.

Suppose we are given a signal f in the Hardy $H^2(\mathbf{D})$ space, that is $f(z) = \sum_{l=1}^\infty c_l z^l$, $\sum_{l=1}^\infty |c_l|^2 < \infty$. Now we seek for a decomposition into a TM system with adaptively selected parameters. We will use the collection of the functions

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

which are normalized Szegö kernels of the unit disc in which a is a parameter. Set $f = f_1$. First write the identity

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

We note that in this step a_1 can be any complex number in the unit disc. The above can be further written as

$$f(z) =< f_1, e_{a_1} > e_{a_1}(z) + f_2(z)\frac{z - a_1}{1 - \overline{a}_1 z} \tag{2}$$

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 f_2 *the generalized backward shift transform of* f_1 *via* a_1. The notion is related to the classical backward shift operator

$$S(f)(z) = a_1 + a_2 z + \cdots + c_{k+1}z^k + \cdots = \frac{f(z) - f(0)}{z}.$$

Recognizing that $f(0) =< f, e_0 > e_0(z)$, the operator S is generalized backward shift operator vis 0.

In the decomposition f_2 is called a *reduced reminder.* The purpose now is to extract the maximal energy portion from the term $< f_1, e_{a_1} > e_{a_1}(z)$. The energy of the latter, due to the reproducing kernel property of e_a, is given by

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

The orthogonality between the two term in the right hand side of (2) and the unimodular property of Möbius transform on the circle imply

$$\|f\|^2 = (1 - |a_1|^2)|f_1(a_1)|^2 + \|f_2\|^2.$$

This shows that to minimize the remainder $\|f_2\|^2$ is to maximize $(1 - |a_1|^2)|f_1(a_1)|^2$.

Fortunately, one can show that there exists a_1 in the open disc $\mathbf{D}$ such that

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

([36]). The existence of such maximal selection is called *Maximal Selection Principle.* Under such a maximal selection of a_1 we call the decomposition (2) a *maximum sifting.* Selecting such a_1 and repeating the process for f_2, and so on, we obtain

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

where for $k = 1, \ldots, n$,

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

and, for $k = 2, \ldots, n + 1$,

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

It can be shown that

$$\lim_{n\to\infty} \|f_{n+1}(\exp(i\cdot)) \prod_{k=1}^{n} \frac{\exp(i\cdot) - a_k}{1 - \overline{a}_k \exp(i\cdot)}\| = 0.$$

There holds the following theorem.

Theorem 3. *For any give function f in the Hardy H^2 space, under the maximum sifting process we have([36])*

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

The following relations are noted:

$$< f_k, e_{a_k} >=< g_k, B_k >=< f, B_k >, \tag{3}$$

where g_k is the standard reminder:

$$g_k(z) = f - \sum_{i=1}^{k-1} < f_i, e_{a_i} > B_i(z).$$

Remark 1. We note that the selected parameters $a_1, \ldots, a_n, \ldots$ in AFD do not have to satisfy the condition (1), and the induced TM system $\{B_k\}$ may not be a basis. The decomposition process exhibit that one is not interested in whether the resulted system is a basis, but interested in whether it can expand the given signal f. One is indeed able to do so, and, in fact, achieves fast convergence.

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 $a_1, \ldots, a_n, \ldots,$ we arrive a pre-mono-component decomposition, of which after multiplying e^{it} all entries in the infinite sum become mono-components.

Remark 3. Based on the same dictionary of Szegö kernels AFD is different from greedy algorithm or orthogonal greedy algorithm in two aspects. One is that at every step we can get a maximal energy portion but not α time of it as in greedy algorithm situation. The second is that we can repeatedly choose the same parameter if necessary to get best possible approximation. Using a cyclic AFD algorithm we can get a conditional solution to the n-best rational function approximation (see subsection 3).

Remark 4. The convergence rate for AFD is $1/\sqrt{n}$ where n is the order of the approximating AFD partial sum. One has to note that this is a convergence rate for bad functions, including those being discontinuous. This convergence estimation, therefore, has a different nature compared with the traditional convergence theorems: the latter are for smooth functions.

Several other mono-component decompositions are based on AFD. To emphasis this fundamental role we sometimes call the above defined AFD as Core AFD.

2. Unwending AFD

In DSP there is the following assertion: If $f = hg$, where f, g are Hardy $H^2(\mathbf{D})$ functions,

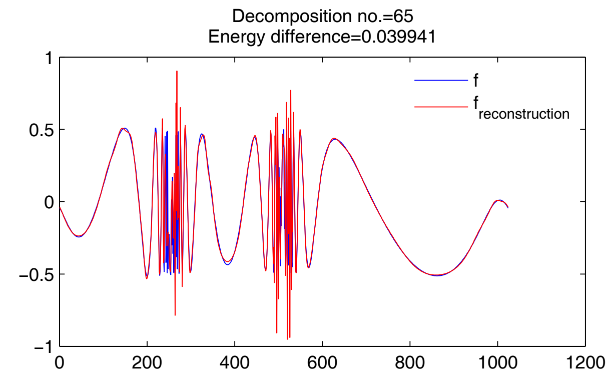

Figure 1: Original signal and reconstructed signal of order 65 AFD decomposition

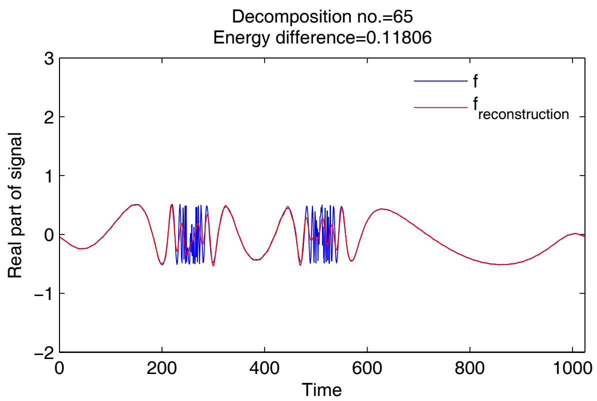

Figure 2: Original signal and reconstructed signal of order 65 FD decomposition

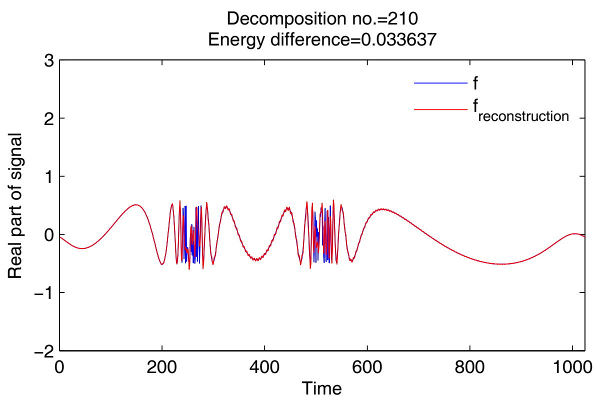

Figure 3: Original signal and reconstructed signal of order 210 FD decomposition

and h is an inner function. Let f and g be expanded into their respective Fourier series, viz.,

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

Then one has, for any n,

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

(see, for instance [8], [5]).

This amounts to say that after factorizing out an inner function factor the remaining Hardy space function series converges faster. This suggests that in the above AFD process if one incorporates a factorization process then the convergence becomes faster. This is reasonable: when a signal by its nature is of high frequency, one should first "unwending" it but not extract from it a maximal portion of lower frequency. We proceed it as follows ([29], [35]). First we do factorization $f=f_1=I_1O_1$, where I_1 and O_1 are, respectively the inner and outer function factors of f. The factorization is based on Nevanlinna's factorization theorem. 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 computation we find the boundary value of the outer function by using the boundary value of f_1 in which the above integral is taken to be of the principal integral sense. The imaginary part of the integral reduces to the circular Hilbert transform of $\log|f_1(e^{it})|$. Next, we do a maximum sifting to O_1. This gives

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

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 factors, $f_2=I_1O_2$, 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 sifting to O_2, and so on. In such way we obtain the decomposition

Theorem 4. *Under the assumptions as in Theorem 2, we have the unwending AFD decomposition*

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

where $f_{k+1}=I_{k+1}O_{k+1}$ *is the backward shift of* O_k *via* $a_k, k=1,\dots,n,$ *and* I_{k+1} *and* O_{k+1}

are respectively the inner and outer functions of f_{k+1}. Furthermore,

$$f(z) = \sum_{k=1}^{\infty} \prod_{l=1}^{k} I_l(z) < O_k, e_{a_k} > B_k(z).$$

Remark 5. In most cases unwending AFD is automatically a mono-component decomposition because of the inner function factors generated in the process. As in AFD we can manually set $a_1 = 0$ to guarantee that all the terms of the unwending AFD are mono-components. Unwending AFD converges very fast. This is shown through comparison of the performances of AFD, unwending AFD, double-sequence AFD, as well as Fourier series on singular inner functions ([35]).

Remark 6. There are other AFD variations that first extract factor signals of high frequencies. Those include a *double-sequence unwending AFD* ([39]) and one using what we call high-order Szegö kernels ([40]). With a similar effectiveness as unwending AFD the algorithm of double-sequence unwending AFD is, however, more complicated. The high-order Szegö kernel method in [40] can get a maximal energy portion as AFD but with a suitable frequency level. It, however, does not have a generalized backward shift structure.

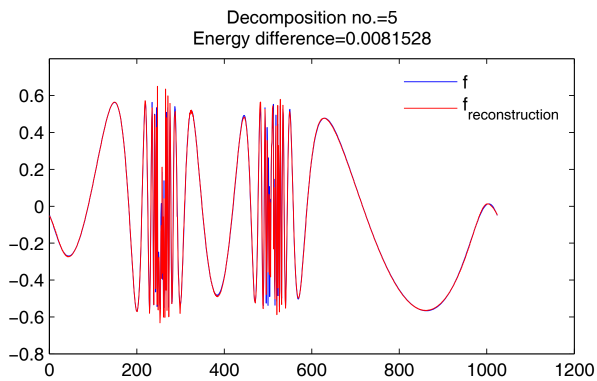

Figure 4: Original signal and reconstructed signal of order 5 Unwending AFD decomposition

3. Optimal approximation by rational functions of order not larger than n

AFD and unwending AFD offer fast decomposition of signals into mono-components. They, however, are not fastest. They are of certain uniqueness only from the algorithm. Due to these obstacles the question on simultaneous selection of n-parameters $a_1, \ldots, a_n$, in an approximating n-Blaschke form, viz.,

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

arises. Simultaneous selection of the parameters but not one by one in a sequel certainly offers better approximate to the given signal. Simultaneous selection of the parameters in an approximating n-Blaschke form is equivalent with the so called *optimal approximation by rational functions of order not larger than* n. We phrase the problem as *best* n*-rational approximation.* It is a long standing open problem till now, formulated as follows.

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, both of degrees less than or equal to n. 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. To find a best n-rational approximation to f is to find an n-pair $p_1, q_1)$ such that

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

By using a so called *cyclic* AFD *algorithm* we can get a solution of the above mentioned problem if there is only one critical point for the objective function ([30]). We call such a solution a *conditional solution*. Besides cyclic AFD, to the author's knowledge, there exists another algorithm, RARL2, by the French institute INRIA, that also can only get a conditional solution [2]. The theory and algorithm of cyclic AFD are both explicit. It directly finds out the poles of the approximating rational function. The other rational approximation models all use the coefficients of p and q as parameters in order to set up and solve the optimization problem. Using coefficients of polynomials involves tedious analysis and computation. The ultimate solution of the optimization problem lays on optimal selection of an initial status to start with. Finding an optimal initial status itself is, however, an NP hard problem.

We will call

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

an n-Blaschke form where the parameters $a_1, \ldots, a_n$ defining B_n are arbitrary complex numbers in $\mathbf{D}$. As shown in the literature, with an abuse of terminology, the parameters $a_1, \ldots, a_n$ are often called "poles" although they are, in fact, zeros. 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 a pole of B_n. This shows a little inconsistence with the notion of n-rational functions. If, instead, we work on the parallel context outside the unit disc, then the set of all n-Blaschke forms coincides with the set of all n-rational functions. Some researchers, including L. Baratchart, choose to work in the context outside the unit disc. To simplify the writing we ignore the inconsistence and still work on inside the unit disc.

For any given natural number n the objective function for the optimization problem is

$$A(f; a_1, \ldots, a_n) = \|f\|^2 - \sum_{k=1}^{n} |\langle f, B_k\rangle|^2. \tag{4}$$

Definition 2. *An n-tuple $(a_1, \ldots, a_n)$ is said to be a coordinate-minimum point of an objective function $A(f; z_1, \ldots, z_n)$ if for any chosen k among 1,…,n, whenever we fix the rest $n-1$ variables, being $z_1 = a_1, \ldots, z_{k-1} = a_{k-1}, z_{k+1} = a_{k+1}, \ldots, z_n = a_n$, and select the kth variable z_k to minimize the objective function, we have*

$$a_k = \arg\max\{A(f; a_1, \ldots, a_{k-1}, z_k, a_{k+1}, \ldots, a_n) \ : \ z_k \in \mathbf{D}\}.$$

In the AFD algorithm we repeat the following procedure: Along with choosing $a_1, \dots, a_{k-1}$ in $\mathbf{D}$, we produce the reduced remainders $f_2, \dots, f_k$. 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 for $k = n$: For any permutation P of $1, \dots, n$, whenever $a_{P(1)}, \dots, a_{P(n-1)}$ are fixed from previous steps we accordingly and inductively obtain the reduced remainders $f_2, \dots, f_n$, and, next, use the Maximal Selection Principle to select an optimal $a_{P(n)}$.

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

Proposition 5.

$$\mathcal{LM} \subset \mathcal{CM} \subset \mathcal{C}. \tag{5}$$

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

Theorem 6. *Suppose that f is not an m-Blaschke form for any $m < n$. Let $s_0 = \{b_1^{(0)}, \dots, b_n^{(0)}\}$ 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 Maximal Selection Principle based on the objective function (4). 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 $s_0, s_1, \dots, s_l, \dots,$ with decreasing objective function values d_l that tend to a limit $d \geq 0$, where, in the notation and formulation of (4) and (3),*

$$d_l = A(f; b_1^{(l)}, \dots, b_n^{(l)}) = \|f\|^2 - \sum_{k=1}^{n} (1 - |b_k^{(l)}|^2) |f_k^{(l)}(b_k^{(l)})|^2. \tag{6}$$

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 $A(f; \dots)$ 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 [30].

Remark 7. Various types of AFD related signal expansions have applications in practices of different areas. For applications in system identification, for instance, see ([22], [21]).

Experiments.

The experimental function is in the Hardy H^2 space with non-trivial singular inner part given by

$$f = \left(\frac{1 - x^2}{(x - \frac{3}{2})(x + \frac{5}{2})} - \frac{1}{(x+2)(x+3)}\right) e^{\left(\frac{x-1}{x+1} + \frac{x+i}{x+i}\right)}.$$

With this example we compare performances of Fourier series decomposition (FD), Core AFD, Unwending AFD and n-Best AFD. We also include comparison between their corresponding time-frequency distributions (with FD replaced by Short Time Fourier Transform or STFT). It is well known that convergence of Fourier series of Hardy space functions with non-trivial singular inner parts is very slow. The experiments show that to reach a similar accuracy of

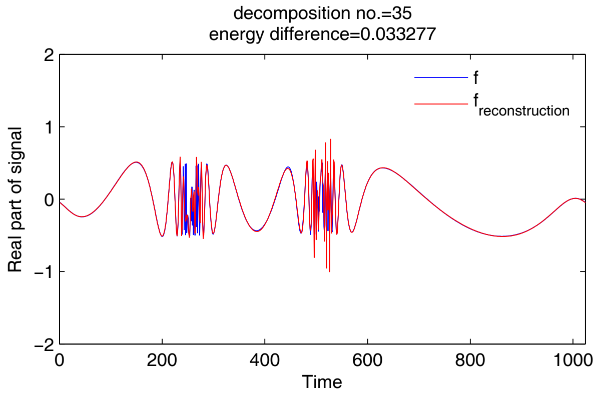

Figure 5: Original signal and reconstructed signal of n-Best AFD decomposition for n=35

approximation of 65 iterations of Core AFD one needs to run 210 iterations of FD. To reach almost the same accuracy one needs to run n-Best AFD for $n = 35$, and only run 5 iterations of Unwending AFD (We note that electronic version of the pictures gives considerably better illustrations of the accuracy comparison due to the easily seen colors). In the time-frequency aspect we see that all of those give two peaks at almost the same time instants. They, however, have different carrier frequencies. This can be understood. In fact, decomposition is not unique. For instance, a mono-component can have many different decompositions into other mono-components. By considering both aspects we recommend Unwending AFD: If a signal is essentially of a certain frequency (or say to have a certain carrier frequency), then Unwending AFD can first factorize the corresponding inner function. The approximation accuracy and time-frequency representation of Unwending AFD can be further improved if computation of the Hilbert transform can be improved: Circular Hilbert transform is encountered to work out the corresponding outer function.

§3 The mathematical theory of phase derivative and its impacts to digital signal processing

Signal analysis practice has been longing to have a frequency theory. Many signal analysts tend to believe that frequency, or instantaneous frequency, should be defined as the phase derivative of a complex signal associated with the given real-valued signal. While this seems to be reasonable, there, however, exist a number of obstacles associated with this idea. First, how to associate a real-valued signal with a complex-valued signal so that we can well define a frequency concept as the phase derivative of the complex signal. Gabor proposed his analytic phase derivative method through Hilbert transform of the signal. Ever since then, signal analysts have been justifying the analytic signal approach through enormous experiments and arguments, but till now there has been no major progress with this method. Gabor's approach, in fact, cannot be easily implemented due to at least two problems. One is that given a general signal, how to mathematically define the analytic phase derivative. The point is that a signal itself is usually not smooth, and neither is the associated analytic signal. A general signal should be assumed to be only a function in the Lebesgue L^2 space. L^2 functions are not considered as functions precisely and pointwisely defined. Functions that defer in their values in a Lebesgue null set are considered to be the same L^2 function. This completely rules out smoothness of L^2 functions. The phase derivative approach in such case amounts to get smooth objects from non-smooth objects. When dealing with a discrete signal one methodology is to treat the data as the Fourier coefficients through Z-transformation; and the other methodology is to treat the data as sampling of a continuous or smooth signal. Both of these methodologies run into the same smoothness verses non-smoothness problem. The following is a concrete example concerning definition of such phase derivative. In L. Cohen's book [7] he proves the following formula

$$\int_{-\infty}^{\infty} \omega |\hat{s}(\omega)|^2 d\omega = \int_{-\infty}^{\infty} \theta'(t)|s(t)|^2 dt \tag{7}$$

under the assumptions that $s(t) = \rho(t)e^{i\theta(t)}$, and s, $\rho(t)$ and $\theta(t)$ are all smooth in t and $\|s\|_2 = 1$. This formula gives a reason why phase derivative should, in general, be considered as instantaneous frequency. The question is, for non-smooth signals s, for which the phase derivative θ' may not exist, in what capacity the above relation or a similar one would still hold true?

The second problem is that although non-negativity of analytic phase derivative is desired, it is not always available. The non-negativity of analytic phase derivative is necessary not only because the physical meaning of the frequency concept, but also because that analysis of positive and negative frequencies together sometimes does not give meaningful results. This can be seen, for instance, from the mean of frequencies. If one does not require non-negativity, the mean of Fourier frequency of any real-valued signal is zero. The task is to construct a coherent analytic signal theory that solves the mentioned and not mentioned problems. The author and his collaborators have built up such a theory. In the previous sections we introduced the mono-component and mono-component decomposition theory. In this section we give a brief summary on phase derivative vs. frequency and some related results.

From our study there are two cases in which we can define phase derivative. One is for some classes of functions defined in complex analysis and the other is based on Sobolev space conditions.

1. Inner and outer functions

If f is a function in the Hardy H^2 space, then f has the following canonical factorization decomposition, called Nevanlinna factorization:

$$f = OBS,$$

where O, B and S are, respectively, the outer, the Blaschke product and the singular inner function part of f. Such classes of functions and the related factorizations are available for all the concerned contexts: the unit disc and outside the closed unit disc, and the upper- and lower half complex planes. The function $I = BS$ is called the inner function part of f. For any inner function Theorem 1 shows that its phase derivative as the limit of the same quantity but from inside of the disc always exists, and be positive, if non-trivial. For the outer function part, under certain conditions it exists, and has the zero-mean property. This shows that it should be sometimes positive and sometimes negative. Here "sometimes positive (sometimes negative)" means that it is positive (negative) in a set of Lebesgue positive measure. There exist results for amplitude derivatives, too. For details we refer the reader to [26].

The inner and outer function theory given by Theorem 1 gives a great impact to the theory of all-pass filters, energy delay and signals of minimum phase [8]. It is noticeable that textbooks of DSP claim that inner functions have positive phase derivative (see [5]), but this fact had never been rigorously proved until the publication of Theorem 1 in [26].

2. Signals in the Sobolev spaces

For a function in the L^2 space on the boundary one can proceed the Hardy space decomposition, viz.,

$$s = s^+ + s^-.$$

The functions $s^\pm$ are holomorphic functions in the respective domains in which they are defined. In the real line case the respective domains are the upper and lower half planes, and we have, as a basic and important property of Hardy space functions,

$$\lim_{\pm y \to 0+} s^\pm(x+y) = s^\pm(x), \qquad \text{a.e.}$$

In order to make

$$\lim_{\pm y \to 0+} (s^\pm)'(x+iy)$$

also exist a.e. it suffices that $(s^\pm)'$ belong to the Hardy space $H^2(\mathbf{C}^\pm)$. Fourier analysis shows that a sufficient and necessary condition for $(s^\pm)'$ belonging to the Hardy space $H^2(\mathbf{C}^\pm)$ is that s belongs to the Sobolev space

$$L_1^2 = \{s \in L^2(\mathbf{R}) \ : \ \frac{d^* s}{dt} \in L^2(\mathbf{R})\},$$

where $\frac{d^*}{dt}$ stands for the distributive derivative [11]. Under such condition we have non-tangentially

$$\lim_{\pm y \to 0+} s^\pm(x+iy), \qquad \lim_{\pm y \to 0+} (s^\pm)'(x+iy)$$

both exist a.e. and the limits are a.e. non-zero. Therefore, by definition,

$$(\theta^\pm)'(t) = \lim_{\pm y \to 0+} (\theta_y^\pm)'(t) = \lim_{\pm y \to 0+} \mathrm{Im}\left(\frac{(s^\pm)'(x+iy)}{s^\pm(x+iy)}\right)$$

exist and being non-zero and finite a.e. Under the Sobolev condition the formula (7) is generalized to

$$\int_{-\infty}^{\infty} \omega|\hat{s}(\omega)|^2 = \int_{-\infty}^{\infty} (\theta^+)'(t)|s^+(t)|^2 dt + \int_{-\infty}^{\infty} (\theta^-)'(t)|s^-(t)|^2 dt$$

(see [11]).

What is amazing is that under the same assumption we are able to define the so called Hardy-Sobolev phase derivative

$$\theta^*(t) = \chi_{\{s^+ + s^- \neq 0\}}(t)\mathrm{Im}\left(\frac{(s^+)'(t) + (s^-)'(t)}{s^+(t) + s^-(t)}\right),$$

where $s^{\pm}$ are understood as a.e. determined through non-tangential boundary limits. Under the Hardy-Sobolev phase derivative notion we can show the ultimate relation (see [8])

$$\int_{-\infty}^{\infty} \omega|\hat{s}(\omega)|^2 = \int_{-\infty}^{\infty} \theta^*(t)|s^+(t)|^2 dt.$$

It is to a great satisfaction to notice that if $s(t) = \rho(t)e^{i\theta(t)}$ and the classical derivatives $s'(t_0), \rho'(t_0)$ and $\theta'(t_0)$ all exist, then $\theta'(t_0) = \theta^*(t_0)$. Similar generalizations for higher order moments and deviations are available under the notion of Hardy-Sobolev derivatives ([11], [8]).

We finally note that with various notions of the phase and amplitude derivatives we are able to prove truly stronger uncertainty principle for the classical setting, the LCT setting, as well as for the general self-adjoint operator setting ([9], [10]).

§4 Time-frequency distribution based on mono-components

Suppose that $m(t) = \rho(t)e^{i\theta(t)}$ is a complex signal, then the ideal time-frequency distribution is one of the Dirac type defined as a function of two variables, viz.,

$$P(t, \omega) = \rho(t)\delta(\omega - \theta'(t)).$$

Here the Dirac function δ is understood as being with value 1 at the zero point and value zero otherwise. The implementation of this idea, however, brings in great controversies. One is that a practical signal, no matter complex-valued or real-valued, cannot be simply expressed in such form with well defined phase derivative, let alone the requirement for $\theta'(t) \geq 0$, a.e. Only in the latter case such time-frequency distribution has properties like a probability distribution for $\omega > 0$.

This formulation is valid and practical only for mono-component signals: Although this has been desired by many signal analysts for many years, it can never be implemented, for there were no clearly defined notions of mono-component and instantaneous frequency. Signal analysts generally admit that a signal is said to be a monocomponent (with a little difference in spelling from what we defined mono-component), "if for this signal, there is only one frequency or a narrow range of frequencies varying as a function of time; and, it is a multicomponent if it is not a monocomponent" (Boashash, [4]). The ambiguity of such definition of monocomponent lays on the fact that it is based on the notion of frequency and its narrow range. What they call "frequency" and "narrow range", however, are not defined in signal processing knowledge system. In such way signal analysis has been established based on intuition with undefined concepts but not on rigorous mathematics. This not only restricts signal analysis practice, but also restrict the theoretical and concept development. There has been no agreement among signal analysts on what is frequency or instantaneous frequency: it depends on personal discrete understanding. Most signal analysts tend to believe that something called frequency objectively exist, and what human being can do is just to "estimate" the frequency. Boashash proposed the above definition in 1990's which adopted the idea of Gabor. Until the present time, however, there has been no progress and the situation stays as the same as that more than a half century ago.

Our way to get out of the frequency paradoxes is to define frequency (instantaneous frequency) as a function to be the analytic phase derivative if the latter can be well defined and non-negative almost everywhere in the Lebesgue measure sense. Signals that possess instantaneous frequency are mono-components. In such way not every signal has instantaneous frequency. For general signals that do not have a global instantaneous frequency function, or, not a mono-component, one seeks for appropriate mono-component decompositions. In such way we have, for a complex-valued signal s in the Hardy space,

$$s(t) = \sum_{k=1}^{\infty} m_k(t),$$

where for each k, m_k is a mono-component. In the AFD decomposition case, for instance, we have $m_k(t) =< f, B_k > B_k = \rho_k(t)e^{i\theta_k(t)}$ with $\theta'_k(t) \geq 0$, a.e. If s is real-valued in L^2, we use the relation $s = 2\text{Re}s^+ - c_0$, and get $m_k(t) = \rho_k(t)\cos\theta_k(t), \theta'(t) \geq 0$, a.e. In both cases we define $P(t,\omega) = \sum_{k=1}^{\infty} P_k(t,\omega), P_k(t,\omega) = \sum_{k=1}^{\infty} \rho_k(t)\delta(\omega - \theta'_k(t))$. Such defined time-frequency distributions enjoy almost all of the commonly desired properties for time-frequency distributions. For details we refer the reader to paper [60].

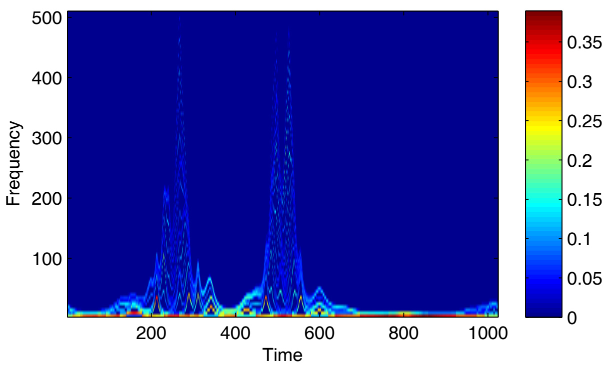

Figure 6: Time frequency distribution of order 65 AFD decomposition

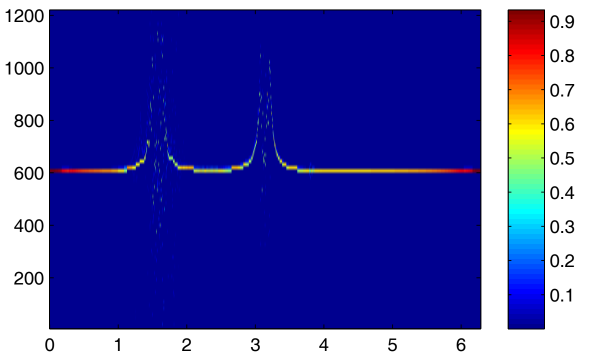

Figure 7: Time frequency distribution of order 5 Unwending AFD decomposition

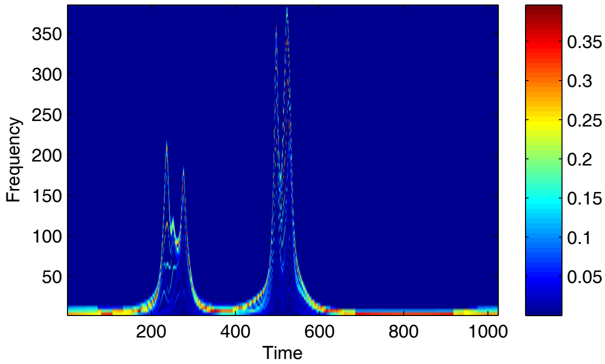

Figure 8: Time frequency distribution of n-Best AFD decomposition for n=35

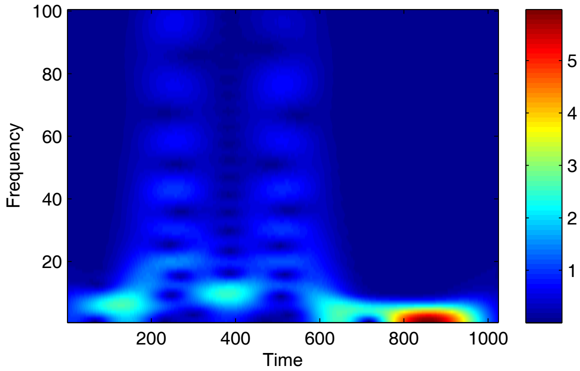

Figure 9: Time frequency distribution of STFT

§5 Further study on TM systems and backward shift operator invariant spaces

Shift and backward shift operator analysis was developed with the cornerstone Beurling Theorem and Beurling-Lax Theorem. Many of the studies are related to Russian mathematicians. The proposed AFD has a close relationship with backward shift invariant spaces. For any sequence of complex numbers in the unit disc, $a_1, \ldots, a_n, \ldots$, there exist two classifications.

1. The hyperbolic non-separable condition (1) holds

In such case the related TM system is dense in the H^p spaces for $1 \le p \le \infty$, viz.,

$$H^p = \overline{\text{span}}\{B_k\}_{k=1}^{\infty}.$$

On the other hand, if the span of $\{B_k\}$ is dense in H^p for any $p \in [1, \infty]$, then the complex numbers in the sequence must satisfy (1).

2. The hyperbolic non-separable hyperbolic condition (1) does not hold

In such case a Blaschke product ϕ with the complex numbers $a_1, \ldots, a_n, \ldots$ as zeros (together with the multiplicities if there are repeatings) can be defined, and

$$H^2(\mathbf{D}) = \overline{\text{span}}\{B_n\} \oplus \phi H^2(\mathbf{D}).$$

This is to say that a sequence $a_1, \ldots, a_n, \ldots$ obtained from an AFD process may not satisfy the non-separable condition, and the span of the related TM system $\{B_k\}$ may not be dense. In such case for the signal s based on which the AFD process is done there holds the relation $s \in \overline{\text{span}}\{B_n\}$.

For general $H^p(\mathbf{D})$ spaces we have [53]

Theorem 7. Assume that $\sum_{k=1}^{\infty}(1-|a_k|) < \infty$, and ϕ is the Blaschke product defined by the a_k's in the sequence, then for any $p \in (1, \infty)$ there holds

$$H^p(\mathbf{D}) \bigcap \phi\overline{H^p(\mathbf{D})} = (\phi H^{p'}(\mathbf{R}))^{\perp} = \overline{\text{span}}\{B_n\}_{n=1}^{\infty}, \tag{8}$$

where the closure $\overline{\text{span}}$ is in the $L^p(\partial\mathbf{D})$ topology and $(\phi H^{p'}(\mathbf{D}))^{\perp} = \{f \in H^p(\mathbf{D}) | \langle f, \phi g\rangle = 0, \ \forall \ g \in H^{p'}(\mathbf{D})\}$.

In [51] we give pointwise convergence results of TM system corresponding to the classic Dini, Dirichlet type results in the Fourier series case. In [6] we give a constructive proof of Beurling-Lax Theorem. In [31] we prove that a TM system $\{B_k\}$ is a Schauder basis in the closure of the span$\{B_k\}$.

Restricted to bandlimited functions we satisfactorily characterized the solutions of the band preserving, phase and amplitude retrieving problems, as well as solutions of the Bedrosian equation when one of the product function is bandlimited. The band preserving and phase retrieving problems mainly arise from optics, and the amplitude retrieving problem and the Bedrosian equation problems are related to signal analysis. These problems are described as follows.

Let $A > 0$.

(i) Knowing supp$\hat{f} \subset [0, A]$, characterize all functions g in $L^p, 1 \le p \le \infty$, such that supp$(\widehat{fg}) \subset [0, A]$. We phase this as **band preserving problem**.

(ii) Knowing supp$\hat{f} \subset [0, A]$, characterize all functions g such that $|g| = 1$, a.e., and supp$(\widehat{fg}) \subset [0, A]$. We phase this as **phase retrieving problem**.

(iii) Let $f(t) = \rho(t)e^{i\theta(t)}$ be a complex mono-component. Characterize all real-valued function g such that fg is again a complex mono-component. We call this as **amplitude retrieving problem**.

(iv) Characterize all solutions for the Bedrosian equation $H(fg) = fHg$, where f or g is bandlimited. We call this as **bandlimited Bedrosian equation**.

Particular cases of amplitude retrieving problem (iii) have been addressed in the previous sections. It is obvious that Problem (ii) can be considered as a particular case of Problem (i). Both problems have been attacked by researchers. The results that the other researchers obtained are based on Weierstrass' infinite product formula for entire functions. Our approach to Problem (i) and (ii), as well as to problem (iv) are based on backward shift invariant spaces that is more explicit and computable.

Denote by $\mathrm{FH}^q[A,B]$ the space of functions in $L^q(\mathbf{R})$ whose distributional Fourier transform is supported in $[A,B]$; by $\mathrm{L}^q_H(\mathbf{R})$ the space of functions in $L^q(\mathbf{R})$ whose Hilbert transform is also in $L^q(\mathbf{R})$. The notation $\partial^{-1}f$ denotes the Laplace transform of f whenever definable. The bandlimited Bedrosian problem (iv) is treated in the following three theorems ([53]).

Theorem 8. *Let $f \in L^p_H(\mathbf{R})$ and $g \in L^q_H(\mathbf{R})$, where $1 \le p, q \le \infty$ and $0 \le r^{-1} = p^{-1}+q^{-1} \le 1$. Then the following assertions are equivalent.*

(1) $H(fg) = fHg$;

(2) $H(f_-g_+) = -if_-g_+$ *and* $H(f_+g_-) = if_+g_-$;

(3) $f_-g_+ \in \mathrm{H}^r(\mathbf{R})$ *and* $\overline{f_+g_-} \in \mathrm{H}^r(\mathbf{R})$;

(4) $f_-g_+ \in \mathrm{FH}^r(\mathbf{R}^+)$ *and* $\overline{f_+g_-} \in \mathrm{FH}^r(\mathbf{R}^+)$;

(5) $\overline{f_-} \in \mathrm{H}^p(\mathbf{R}) \cap \mathbf{I}_{\mathbf{g}_+}\overline{\mathrm{H}^p(\mathbf{R}^+)}$ *and* $f_+ \in \mathrm{H}^p(\mathbf{R}) \bigcap \mathbf{I}_{\overline{\mathbf{g}_-}}\overline{\mathrm{H}^p(\mathbf{R})}$ *if g_+ and g_- are nonzero functions.*

(6) $g_+ \in \varphi_1\mathrm{H}^q(\mathbf{R})$, $\overline{g_-} \in \varphi_1'\mathrm{H}^q(\mathbf{R})$ *and* $\frac{O_{\overline{f_-}}}{\overline{O_{\overline{f_-}}}}I_{\overline{f_-}} = \frac{\varphi_1}{\varphi_2}$, $\frac{O_{f_+}}{\overline{O_{f_+}}}I_{f_+} = \frac{\varphi_1'}{\varphi_2'}$, *where f_+ and f_- are nonzero functions, φ_1 and φ_2 is a pair of co-prime inner functions, φ_1' and φ_2' is also a pair of co-prime inner functions.*

Theorem 9. *Let $f \in \mathrm{FH}^p[A,B]$ and $g \in \mathrm{H}^p(\mathbf{R})$ be nonzero functions. If $A, B \in \operatorname{supp}\widehat{f}$ and $A < 0 < B$. Then $H(fg) = fHg$ if and only if $g_+ \in e^{-iAx}\mathrm{H}^q(\mathbf{R})$ and $\overline{g_-} \in e^{iBx}\mathrm{H}^q(\mathbf{R})$.*

Theorem 10. *Let $g \in \mathrm{FH}^q[A,B]$, where $A, B \in \operatorname{supp}\widehat{g}$, and $f \in \mathrm{L}^p_H(\mathbf{R})$ be nonzero functions. Then $H(fg) = fHg$ if and only if*

$$f \in \overline{\operatorname{span}}^p \left\{ \frac{1}{(x-\lambda)^j}, \quad \lambda \in E_1 \cup E_2, \quad j = 1, \cdots, m(\lambda) \right\},$$

where E_1 is the set of all different zeros of $G_+(z) := (\partial^{-1}g_+)(z)$ in the upper half plane, E_2 is the set of all different zeros of $G_-(z) := (\partial^{-1}g_-)(z)$ in the lower half plane and $m(\lambda)$ be the multiplicity at λ. The above representation is with the convention that if one of g_+ and g_- is zero, then the corresponding set of zeros is the empty set.

The next two theorems treat the band preserving problem (i) ([52], [53]).

Theorem 11. *Suppose that $0 \not\equiv g \in \mathrm{FH}^q[A,B]$ and $f \in \mathrm{L}^p_H(\mathbf{R})$. Then $fg \in \mathrm{FH}^r[A,B]$ if and only if*

$$\overline{f_-} \in \mathrm{H}^p(\mathbf{R}) \quad \bigcap \quad I_{g_1}(x)\overline{\mathrm{H}^p(\mathbf{R})} = \mathrm{FH}^p[0,a_2] \bigoplus \bigoplus e^{ia_2x}\overline{\operatorname{span}}^p \left\{ \frac{1}{(x-\overline{\lambda})^j}, \quad \lambda \in E_1, \quad j = 1, \cdots, m(\lambda) \right\},$$

and

$$f_+ \in \mathrm{H}^p(\mathbf{R}) \quad \bigcap \quad I_{g_2}(x)\overline{\mathrm{H}^p(\mathbf{R})} = \mathrm{FH}^p[0,a_1] \bigoplus \bigoplus e^{ia_1x}\overline{\operatorname{span}}^p \left\{ \frac{1}{(x-\lambda)^j}, \quad \lambda \in E_2, \quad j = 1, \cdots, m(\lambda) \right\},$$

where $I_{g_1}(x)$ and $I_{g_2}(x)$ are, respectively,

$$I_{g_1}(x) := e^{ia_1x} \prod_{\lambda \in E_1} \left(\frac{|\lambda^2+1|}{\lambda^2+1} \cdot \frac{z-\lambda}{z-\overline{\lambda}} \right)^{m(\lambda)}, I_{g_2}(x) := e^{ia_2x} \prod_{\lambda \in E_2} \left(\frac{|\lambda^2+1|}{\lambda^2+1} \cdot \frac{z-\lambda}{z-\overline{\lambda}} \right)^{m(\lambda)}.$$

Theorem 11 gives a characterization of the solutions $f \in L^p(\mathbf{R}), \mathbf{1} < \mathbf{p} < \infty$, in terms of backward shift invariant subspaces. It, however, does not cover the cases $p = 1$ and $p = \infty$ due to the failure of the relevant Hardy spaces decomposition. Below we will treat the two exceptional cases by using an alternative approach.

Theorem 12. *Suppose that function $0 \not\equiv g \in \mathrm{FH}^q[A, B]$ and $f \in L^p(\mathbf{R}), \mathbf{1} \leq \mathbf{p} \leq \infty$, be nonzero functions. Then $fg \in \mathrm{FH}^r[A, B]$ if and only if*

$$f \in \overline{I_{g_1}}\mathrm{H}^p(\mathbf{R}) \bigcap \mathbf{I_{g_2}}\overline{\mathrm{H}^{\mathbf{p}}(\mathbf{R})} = \overline{\mathbf{I_{g_1}}}\left[\mathrm{H}^{\mathbf{p}}(\mathbf{R}) \bigcap \mathbf{I_{g_1}}\mathbf{I_{g_2}}\overline{\mathbf{H}^{\mathbf{p}}(\mathbf{R})}\right],$$

where $I_{g_1}(x) := e^{i(a_1x+b_1)}B_1(x)$ is the inner function of $g_1(x) := e^{-iAx}g(x)$ and $I_{g_2}(x) := e^{i(a_2x+b_2)}B_2(x)$ is the inner function of $g_2(x) := e^{iBx}\overline{g(x)}$.

The next three theorems treat phase retrieving problem (ii) ([52]).

Theorem 13. *Assume that $0 \neq f \in H^p(\mathbf{R})$ and its Laplace transform $f(z)$ is holomorphic across $\mathbf{R}$. Then there exists an analytic signal $g(x) \in H^p(\mathbf{R})$ whose Laplace transform $g(z)$ is holomorphic across $\mathbf{R}$ such that $|f(x)| = |g(x)|$ if and only if*

$$g(x) = e^{iax+ib}\left(\prod_{n=1}^{\infty}\frac{|\beta_n'^2+1|}{\beta_n'^2+1}\frac{x-\beta_n'}{x-\overline{\beta_n'}}\right)\left(\prod_{n=1}^{\infty}\frac{\alpha_n'^2+1}{|\alpha_n'^2+1|}\frac{x-\overline{\alpha_n'}}{x-\alpha_n'}\right)B_f(x)f(x), \tag{9}$$

where a and b are some real constants, $\{\alpha_n'\}_{n=1}^{\infty}$ are partial zeros of $f(z)$ in the upper half plane, $\{\beta_n'\}_{n=1}^{\infty}$ is a complex sequence satisfying $\sum\limits_{n=1}^{\infty}\frac{2\mathrm{Im}(\beta_n')}{1+|\beta_n'|^2} < \infty$, $\{\beta_n'\}_{n=1}^{\infty}$ can only have an accumulation point at ∞ and $\{\beta_n'\}\bigcap\{\alpha_n'\} = \emptyset$.

Theorem 14. *Assume that analytic signal $0 \neq f \in H^p(\mathbf{R})$ and its Laplace transform $f(z)$ be an entire function. Then there exists an analytic signal $g(x) \in H^p(\mathbf{R})$ whose Laplace transform $g(z)$ is an entire function such that $|f(x)| = |g(x)|$ if and only if*

$$g(x) = e^{iax+ib}\left(\prod_{n=1}^{\infty}\frac{|\beta_n'^2+1|}{\beta_n'^2+1}\frac{x-\beta_n'}{x-\overline{\beta_n'}}\right)\left(\prod_{n=1}^{\infty}\frac{\alpha_n'^2+1}{|\alpha_n'^2+1|}\frac{x-\overline{\alpha_n'}}{x-\alpha_n'}\right)f(x),$$

where a and b are some real constants, $\{\alpha_n'\}_{n=1}^{\infty}$ are partial zeros of $f(z)$ in the upper half plane, $\{\beta_n'\}_{n=1}^{\infty}$ are partial zeros of $f(z)$ satisfying $\sum\limits_{n=1}^{\infty}\frac{2\mathrm{Im}(\beta_n')}{1+|\beta_n'|^2} < \infty$,and $\{\beta_n'\}\bigcap\{\alpha_n'\} = \emptyset$.

The following theorem completely solves the phase retrieving problem for bandlimited signals.

Theorem 15. *Let nonzero analytic signals $f \in H^p(\mathbf{R})$ and $Supp\widehat{f} \subseteq [0, A]$. Then there exists an analytic signal $g \in H^p(\mathbf{R})$ with $Supp\widehat{g} \subseteq [0, A]$ such that $|f(x)| = |g(x)|$ if and only if*

$$g(x) = e^{ib+iax}\prod_{\alpha_k'}\frac{\overline{\alpha_k'}^2+1}{|\overline{\alpha_k'}^2+1|}\cdot\frac{x-\overline{\alpha_k'}}{x-\alpha_k'}\prod_{\beta_k'}\frac{|\overline{\beta_k'}^2+1|}{\overline{\beta_k'}^2+1}\cdot\frac{x-\overline{\beta_k'}}{x-\beta_k'}f(x), \tag{10}$$

where b and a are some real constants,$\{\alpha_k'\}$ are partial zeros of $f(z)$ in the upper half plane, $\{\beta_k'\}$ are partial zeros of $g(z)$ in the lower half plane and $\{\beta_n'\}\bigcap\{\alpha_n'\} = \emptyset$.

§6 Higher dimensional generalizations

In higher dimensional spaces there are mainly two types of complex structures of which one is several complex variables which is essentially of the tensor form and the other is Clifford algebra that treats a vector variable as a complex variable. Quaternionic algebra is a Clifford algebra. It has a particular position not because it is the only commutative or non-commutative field of a finite dimension apart from the real and the complex number fields (Frobenius) but because that the quaternionic space offers an "non-canonical" imbedding of $\mathbf{R}^3$ into the span of the Clifford algebra generated by two Clifford basis elements e_1 and e_2 satisfying $e_1^2 = e_2^2 = -1, e_1e_2 = -e_2e_1$. The canonical imbedding of $\mathbf{R}^3$ is the one that identifies $\mathbf{R}^3$ with the set $\{x_1e_1 + x_2e_2 + x_3e_3\}$ in which e_1, e_2, e_3 are independent Clifford basic elements. By complex structure we mean a Cauchy type structure, including at least a Cauchy kernel, a Cauchy theorem and a Cauchy formula. There may also be a related Hardy space in the context, and correspondingly a Szegö kernel as reproducing kernel in the boundary L^2-space, etc. In principle, one can develop an approximation theory by using linear combinations of sampled Szegö kernels. We in below mention some particulars in each of the concerned contexts.

With the several complex variable setting we have been studying two contexts, the n-torus and the tubes in the sense of Stein and Weiss [49]. In the n-torus we use the direct product of the TM systems associated with each of the complex variables. Taking $n = 2$, we show that for any function f in the Hardy space of the two-disc, we can choose adaptively two sequences $\mathbf{a} = \{a_n\}$ and $\mathbf{b} = \{b_n\}$ such that the corresponding direct product of the two TM systems, $\mathcal{B}^{\mathbf{a}} \oplus \mathcal{B}^{\mathbf{b}} = \{B_k^{\mathbf{a}} \oplus B_k^{\mathbf{b}}\}$ offer fast decomposition of the given signal [33]. For the tubes case there is a parallel theory.

In the Clifford algebra setting, for the quaternionic case we can construct a theory very similar to the one complex variable case. Although there is no backward shift operator in the context we can show a mechanism similar to (3) ([41]). For a general Clifford algebra this has not been done. The obstacle is that in the Clifford case in the Gram-Schmidt scalar-valued. This gives rise to a technical difficulty. Instead, in the general Clifford algebra setting we use an improved greedy algorithm to get one of the best parameters at each selection ([44]). In the multi-dimensional case we also developed a higher order Szegö kernel method to extract suitable basic functions of high frequency with the maximal energy [57].

These results offer a theory of rational function approximation in higher dimensions. Precisely, the related reproducing kernels are not necessarily rational functions but with square root of polynomials. We also developed compressed sensing with Szegö kernels [19], supporting vector machine with Szegö kernels [23]), as well as adaptive Aveiro discretization method by using Szegö kernels [20].

References

[1] A. Axelsson, K.I. Kou, T. Qian. *Hilbert transforms and the Cauchy integral in Euclidean space,* Studia Math, 2009, 193(2): 161-187.

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

[3] S. Bell. *The Cauchy Transform, Potential theory and Conformal Mappings,* CRC Press, Boca, Raton, 1992.

[4] B. Boashash. *Estimating and interpreting the instantaneous frequency of a signal—Part 1: fundamentals,* Proc IEEE, 1992, 80(4): 520-538.

[5] Q.S. Cheng. *Digital Signal Processing,* Peking University Press, 2003, in Chinese.

[6] Q.H. Chen, T. Qian, L.H. Tan. *Constructive proof of Beurling-Lax theorem,* accepted to appear in Chinese Ann Math.

[7] L. Cohen. *Time-Frequency Analysis: Theory and Applications,* Prentice Hall, 1995.

[8] P. Dang, T. Qian. *Analytic phase derivatives, all-pass filters and signals of minimum phase,* IEEE Trans Signal Process, 2011, 59(10): 4708-4718.

[9] P. Dang, G.T. Deng, T. Qian. *A sharper uncertainty principle*, J Funct Anal, 2013, 265: 2239-2266.

[10] P. Dang, G.T. Deng, T. Qian. *A tighter uncertainty principle for linear canonical transform in terms of phase derivative*, accepted to appear in IEEE Trans Signal Process.

[11] P. Dang, T. Qian, Z. You. *Hardy-Sobolev spaces decomposition and applications in signal analysis,* J Fourier Anal Appl, 2011, 17(1): 36-64.

[12] G. Davis, S. Mallet, M. Avellaneda. *Adaptive greedy approximations*, Constr Approx, 1997, 13: 57-98.

[13] P. Fulcheri, M. Olivi. *Matrix rational H^2 approximation: a gradient algorithm based on schur analysis*, SIAM J Control Optim, 1998, 36(6): 2103-2127.

[14] D. Gabor. *Theory of communication,* J IEE, 1946, 93(III): 429-457.

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

[16] S. Gong. *Private comminication*, 2002.

[17] G.M. Gorusin. *Geometrical Theory of Functions of One Complex Variable,* translated by Jian-Gong Chen, 1956.

[18] N.E. Huang, et al. *The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,* Proc Roy Soc London Ser A, 1998, 454: 903-995.

[19] S. Li, T. Qian, W.X. Mai. *Sparse reconstruction of Hardy signal and applications to time-frequency distribution,* accepted to appear in Int J Wavelets Multiresolut Inf Process.

[20] W.X. Mai, T. Qian, S. Saitoh. *Adaptive decomposition of functions with reproducing kernels,* in preparation.

[21] W. Mi, T. Qian. *Frequency domain identification: an algorithm based on adaptive rational orthogonal system*, Automatica, 48(6): 1154-1162.

[22] W. Mi, T. Qian, F. Wan. *A fast adaptive model reduction method based on Takenaka-Malmquist systems*, Systems Control Lett, 2012, 61(1): 223-230.

[23] Y. Mo, T. Qian, W. Mi. *SVM for frequency domain identification,* in preparation.

[24] B. Picinbono. *On instantaneous amplitude and phase of signals,* IEEE Trans Signal Process, 1997, 45(3): 552-560.

[25] T. Qian. *Analytic signals and Harmonic measures,* J Math Anal Appl, 2006, 314: 526-536.

[26] T. Qian. *Boundary derivatives of the phases of inner and outer functions and applications,* Math Methods Appl Sci, 2009, 32: 253-263. DOI: 10.1002/mma.1032.

[27] T. Qian. *Mono-components for decomposition of signals,* Math Methods Appl Sci (Z Anal Anwendungen), 2006, 29: 1187-1198.

[28] T. Qian. *Mono-components for signal decomposition,* In: *Wavelet Analysis and Applications*, Appl Numer Harmon Anal, 2007, 299-319.

[29] T. Qian. *Intrinsic mono-component decomposition of functions: an advance of Fourier theory,* Math Methods Appl Sci, 2010, 33: 880-891. DOI: 10.1002/mma.1214.

[30] T. Qian. *Cyclic AFD algorithm for best approximation by rational functions of given order,* accepted to appear in Math Methods Appl Sci.

[31] T. Qian, Q.H. Chen. *Rational orthogonal systems are Schauder bases,* accepted by Complex Var Elliptic Equ.

[32] T. Qian, Q.H. Chen, L.Q. Li. *Analytic unit quadrature signals with non-linear phase,* Phys D, 2005, 303: 80-87.

[33] T. Qian, J.S. Huang. *AFD on the n-torus,* in preparation.

[34] , T. Qian, I.T. Ho, I.T. Leong, Y.B. Wang. *Adaptive decomposition of functions into pieces of non-negative instantaneous frequencies,* Int J Wavelets, Multiresolut Inf Process, 2010, 8(5): 813-833.

[35] T. Qian, H. Li, M. Stessin. *Comparison of adaptive mono-component decompositions,* Nonlinear Anal Real World Appl, 2013, 14(2): 1055-1074.

[36] T. Qian, Y.B. Wang. *Adaptive Fourier series—a variation of greedy algorithm*, Adv Comput Math, 2011, 34(3): 279-293.

[37] T. Qian, E. Wegert. *Optimal approximation by Blaschke forms,* Complex Var Elliptic Equ, 2013, 58(1): 123-133.

[38] T. Qian, L.H. Tan. *Condtruction of mono-components through Bedrosian identity,* preprint.

[39] T. Qian, L.H. Tan, Y.B. Wang. *Adaptive decomposition by weighted inner functions: a generalization of Fourier serie,* J Fourier Anal Appl, 2011, 17(2): 175-190.

[40] T. Qian, J.X. Wang. *Adaptive decomposition of functions by higher order Szegö kernels I: a method for mono-component decomposition,* submitted to Acta Appl Math.

[41] T. Qian, W. Sproessig, J.X. Wang. *Adaptive Fourier decomposition of functions in quaternionic Hardy spaces,* Math Methods Appl Sci, 2012, 35: 43-64. DOI: 10.1002/mma.1532.

[42] T. Qian, L.H. Qian. *Phase and amplitude of analytic signals,* preprint.

[43] T. Qian, L.H. Qian. *Mono-components and p-starlike functions,* preprint.

[44] T. Qian, J.X. Wang, Y. Yang. *Matching pursuits among shifted Cauchy kernels in higher-dimensional spaces,* accepted to appear in Acta Math Sinica.

[45] T. Qian, R. Wang, Y.S. Xu, H.Z. Zhang. *Orthonormal bases with nonlinear phase,* Adv Comput Math, 2010, 33: 75-95.

[46] T. Qian, Y. Yang. *Hilbert transforms on the sphere with the Clifford algebra setting,* J Fourier Anal Appl, 2009, 15: 753-774. DOI: 10.1007/s00041-009-9062-4.

[47] T. Qian, L.M. Zhang, Z.X. Li. *Algorithm of adaptive Fourier decomposition,* IEEE Trans Signal Process, 2011, 59(12): 5899-5902.

[48] R.C. Sharpley, V. Vatchev. *Analysis of intrinsic mode functions,* Constr Approx, 2006, 24: 17-47.

[49] E.M. Stein, G. Weiss. *Introduction to Fourirer Analysis on Euclidean Spaces,* Princton University Press, Princeton, New Jersey, 1971.

[50] L.H. Tan, L.X. Shen, L.H. Yang. *Rational orthogonal bases satisfying the Bedrosian identity,* Adv Comput Math, 2010, 33: 285-303.

[51] L.H. Tan, T. Qian. *Estimations of convergence rate of rational Fourier series and conjugate rational Fourier series and their applications,* Sci Sinica Math, 2013, 43(6): 541-550.

[52] L.H. Tan, T. Qian. *Backward shift invariant subspaces with applications to band preserving and phase retrieval problems,* preprint.

[53] L.H. Tan, T. Qian, Q.H. Chen. *New aspects of Beurling-Lax shift invariant subspaces,* preprint.

[54] T. Qian, L.H. Qian. *Phase and amplitude of analytic signals,* preprint.

[55] T. Qian, L.H. Qian. *Mono-components and p-starlike functions,* preprint.

[56] J.L. Walsh. *Interpolation and Approximation by Rational Functions in the Complex Plane,* American Mathematical Society, Providence, RI, 1969.

[57] J.X. Wang, T. Qian. *Approximation of monogenic functions by higher order Szegö kernels on the unit ball and the upper half space,* accepted to appear in Sci China Math, DOI: 10.1007/s11425-013-4710-1.

[58] Y.S. Xu. *Private comminication*, 2005.

[59] B. Yu, H.Z. Zhang. *The Bedrosian identity and homogeneous semi-convolution equations,* J Integral Equations Appl, 2008, 20: 527-568.

[60] L.M. Zhang, T. Qian, P. Dang, W.X. Mai. *Diract type time-frequency distribution based on mono-component decomposition,* preprint.

[1] Department of Mathematics, University of Macau, Macao. Email: fsttq@umac.mo

[2] Department of Computer and Information Science, University of Macau, Macao.
Email: lmzhang@umac.mo