

# Greedy Algorithms and Rational Approximation in One and Several Variables

Laurent Baratchart, Wei-Xiong Mai, Tao Qian

► **To cite this version:**

Laurent Baratchart, Wei-Xiong Mai, Tao Qian. Greedy Algorithms and Rational Approximation in One and Several Variables. Swanhild Bernsetin and Uwe Kähler and Irene Sabadini and Franciscus Sommen ISAAC 2015, Aug 2015, Macau, China. Birkäuser, Trends in Mathematics, pp.19 - 33, 2016, Modern Trends in Hypercomplex Analysis. <<http://www.fst.umac.mo/conference/isaac2015/>>. <10.1007/978-3-319-42529-0\_2>. <hal-01421072>

**HAL Id: hal-01421072**

**<https://hal.inria.fr/hal-01421072>**

Submitted on 21 Dec 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# GREEDY ALGORITHMS AND RATIONAL APPROXIMATION IN ONE AND SEVERAL VARIABLES

LAURENT BARATCHART, AND WEI-XIONG MAI AND TAO QIAN

ABSTRACT. We will review the recent development of rational approximation in one and several real and complex variables. The concept rational approximation is closely related to greedy algorithms, based on a dictionary consisting of Szegő kernels in the present context.

## CONTENTS

1. Introduction	1
2. Preliminaries on Greedy algorithm in Hilbert Spaces	1
3. The Hardy $H^2(\mathbf{D})$ case	4
4. Quaternionic and Clifford Contexts for Functions of Several Real Variables	7
5. Several Complex Variables	8
5.1. Functions Defined on $n$ -Torus	8
5.2. Functions Defined on $R^n$ in the Setting of Hardy Spaces on Tubes	9
5.3. Matrix-valued Signals Defined in the Unit Disc	9
5.4. Adaptive Decomposition: the Case of the Drury-Arveson Space	9
5.5. Matrix-valued Signals Defined on the Polydisc	9
6. AFD and Aveiro Method in Reproducing kernel Hilbert Spaces	9
References	10

## 1. INTRODUCTION

Traditionally, rational approximation is mostly restricted to approximation to functions of one complex variable in various domains. Recently a number of adaptive or sparse representation methods were developed, including in principle the greedy algorithms and those based on learning theory, including the SVM method ([25, 26]), that all fit into the concept of rational approximation. This article mainly concerns the greedy algorithm method in relation to a dictionary in the context at hand. Under the greedy algorithm method, rational approximation is generalized to include approximation in several complex and real variables. This idea, in particular, is applicable to function spaces with a Cauchy structure, as well as to reproducing kernel Hilbert spaces. Below we discuss this concept in a number of individual contexts.

## 2. PRELIMINARIES ON GREEDY ALGORITHM IN HILBERT SPACES

Let  $\mathbb{H}$  be a Hilbert space with a dictionary, where, by a dictionary, we mean a set  $\mathbb{D} \subset \mathbb{H}$  satisfying (i)  $e \in \mathbb{D}$  implies  $\|e\| = 1$ ; and (ii)  $\overline{\text{span}\{\mathbb{D}\}} = \mathbb{H}$ .

---

2000 *Mathematics Subject Classification.* 45P05, 30H.

*Key words and phrases.* Rational Orthogonal System, Takenake-Malmquist System, Szegő Kernel, Hardy Spaces.

Supported in part by MYRG116(Y1-L3)-FST13- QT; and FDCT 098/2012/A3.

The most basic greedy algorithm would be the following. Let  $\rho \in (0, 1]$  be fixed. Let  $f \in \mathbb{H}$  and  $g_1 = f$ . Choose  $e_1 \in \mathbb{H}$  such that

$$|\langle g_1, e_1 \rangle| \geq \rho \sup\{|\langle g_1, e \rangle| \mid e \in \mathbb{D}\}.$$

When  $\rho \in (0, 1)$ , a desired  $e_1$  always exists. While for  $\rho = 1$ , a desired  $e_1$  satisfying the above requirement may not exist. In the context of the present paper, where a Cauchy structure prevails, the desired  $e_1$  for  $\rho = 1$  always exists. Our discussion, however, is for the general case  $\rho \in (0, 1]$ . Subsequently, we have the decomposition

$$f = \langle g_1, e_1 \rangle e_1 + g_2,$$

where  $g_2$  is the *standard remainder*:

$$g_2 = f - \langle g_1, e_1 \rangle e_1.$$

Obviously,  $g_2$  is orthogonal with  $\langle g_1, e_1 \rangle e_1$ , and hence,

$$\|f\|^2 = \|\langle g_1, e_1 \rangle e_1\|^2 + \|g_2\|^2 = |\langle g_1, e_1 \rangle|^2 + \|g_2\|^2.$$

Therefore, to minimize  $\|g_2\|^2$  is to maximize  $|\langle g_1, e_1 \rangle|^2$ . If we apply the same reduction to  $g_2$ , we obtain the standard remainder  $g_3$ , where

$$f = \langle g_1, e_1 \rangle e_1 + \langle g_2, e_2 \rangle e_2 + g_3,$$

and

$$\|f\|^2 = |\langle g_1, e_1 \rangle|^2 + |\langle g_2, e_2 \rangle|^2 + \|g_3\|^2,$$

where  $e_2$  is chosen so that

$$|\langle g_2, e_2 \rangle| \geq \rho \sup\{|\langle g_2, e \rangle| \mid e \in \mathbb{D}\}.$$

Repeating the same procedure on  $g_3$  we get  $g_4$ , and so on. Inductively we obtain

$$f = \sum_{k=1}^n \langle g_k, e_k \rangle e_k + g_{k+1}$$

and

$$\|f\|^2 = \sum_{k=1}^n |\langle g_k, e_k \rangle|^2 + \|g_{k+1}\|^2,$$

where  $e_k$  is chosen to make

$$|\langle g_k, e_k \rangle| \geq \rho \sup\{|\langle g_k, e \rangle| \mid e \in \mathbb{D}\},$$

while  $e_1, \dots, e_{k-1}$  were previously chosen.

The theory of greedy algorithms ([17], [40]) asserts that

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

and

$$\|f\|^2 = \sum_{k=1}^{\infty} |\langle g_k, e_k \rangle|^2.$$

The above described algorithm is the so called *General Greedy Algorithm*. A refinement of the general greedy algorithm is the so called *Orthogonal Greedy Algorithm*. The difference is only that the  $g_k$  are replaced by the *orthogonal remainders*  $\tilde{g}_k$ , defined through the relation

$$(2.1) \quad f = \sum_{k=1}^n \langle \tilde{g}_k, B_k \rangle B_k + \tilde{g}_{k+1},$$

where  $\{B_1, \dots, B_{k-1}, B_k\}$  is the G-S orthogonalization of  $\{B_1, \dots, B_{k-1}, e_k\}$ , and  $e_k$  is chosen so that

$$|\langle \tilde{g}_k, e_k \rangle| \geq \rho \sup\{|\langle \tilde{g}_k, e \rangle| \mid e \in \mathbb{D}\}.$$

Among the well known greedy algorithms, the most effective one would be the just described Orthogonal Greedy Algorithm. We note that neither the General Greedy Algorithm, nor the Orthogonal Greedy Algorithm can repeatedly select dictionary elements: Repeated selections or even selecting one in the linear span of the already selected will give nil contribution to the energy approximation.

When the space is a reproducing kernel Hilbert space, consisting of functions defined, say, in a set  $\mathbf{D}$ , and if the dictionary consists of the normalized reproducing kernels  $e_a$ , where  $a$  ranges over a set  $\mathbf{D} \subset C^N$  for some  $N$ , then for any  $f \in \mathbb{H}$ ,

$$\langle f, e_a \rangle = r(a)f(a),$$

where  $r(a) > 0$  is the normalizing constant that makes  $\|e_a\| = 1$ .

In such a case, at each recursive step of the General Greedy and Orthogonal Greedy Algorithms, one seeks a suitable  $e_a$  such that

$$r(a)|f(a)| \geq \rho \sup\{r(b)|f(b)| \mid b \in \mathbf{D}\}.$$

Numerically this is easy to achieve through computation based on the information on the known function  $f$ .

We will show that under certain assumption there exists a variation of the Orthogonal Greedy Algorithm that allows repeated selection of the variable  $a$ . Repeating the selection of the variable  $a$  corresponds to selecting directional derivatives, of order one and higher, of the dictionary elements  $e_a$ .

This new greedy algorithm proposed in [35], called *Pre-Orthogonal Greedy Algorithm* (P-OGA), is formulated as follows.

Let  $\{e_{a_1}, \dots, e_{a_{n-1}}\}$  be the  $(n-1)$ -tuple of the previously selected dictionary elements, and  $\{B_1, \dots, B_{n-1}\}$  its G-S orthogonalization. Sometimes  $B_k = B_{\{a_1, \dots, a_k\}}$  is more precisely written as  $B_{\{a_1, \dots, a_{k-1}\}}^{a_k}$ . The selection criterion for  $a_n$  is

$$(2.2) \quad |\langle f_n, B_{\{a_1, \dots, a_{n-1}\}}^{a_n} \rangle| \geq \rho \sup\{|\langle f_n, B_{\{a_1, \dots, a_{n-1}\}}^a \rangle| \mid a \in \mathbf{D}\},$$

where  $f_n$  is the standard remainder with respect to the orthonormal system  $\{B_1, \dots, B_{n-1}\}$ . We note that under such a selection criterion  $f_k$  is different from  $\tilde{g}_k$  defined through (2.1).

Now we add an assumption under which the machinery that we design will work. Namely, we assume for any  $f \in \mathbb{H}$  and  $a_1, \dots, a_{n-1} \in \mathbf{D}$  that

$$(2.3) \quad \lim_{a \rightarrow \partial \mathbf{D}} |\langle f, B_{\{a_1, \dots, a_{n-1}\}}^a \rangle| = 0,$$

where  $\partial \mathbf{D}$  indicates the boundary of  $\mathbf{D}$  in  $C^N \cup \infty$ . We show that under this assumption, if the dictionary is suitably extended, then the threshold  $\rho = 1$  can be reached, and the following equality holds:

$$(2.4) \quad |\langle f_n, B_{\{a_1, \dots, a_{n-1}\}}^{a_n} \rangle| = \max\{|\langle f_n, B_{\{a_1, \dots, a_{n-1}\}}^a \rangle| \mid a \in \mathbf{D}\}.$$

The extension of the dictionary consists in adjoining the directional derivatives of the kernels  $e_a$  with respect to  $a$ , in all directions and for all  $a \in \mathbf{D}$ . This of course only makes sense if  $\mathbf{D}$  is open and  $e_a$  has some smoothness with respect to  $a$ . Hereafter we assume that  $a \rightarrow e_a$  is smooth as a function  $\mathbf{D} \rightarrow \mathbb{H}$ , in particular the derivatives of  $e_a$  again lie in  $\mathbb{H}$  as limits of divided differences in  $\mathbb{H}$ . In all examples that we shall deal with,  $e_a$  is even analytic with respect to  $a$ , and this warrants the discussion below.

In fact, the Cauchy-Schwarz inequality gives

$$|\langle f_n, B_{\{a_1, \dots, a_{n-1}\}}^a \rangle| \leq \|f_n\|.$$

Thus, there exists a sequence of points,  $a^{(l)}$ , converging to an interior or a boundary point of  $\mathbf{D}$ ,  $\lim_{l \rightarrow \infty} a^{(l)} = a_n$ , such that

$$(2.5) \quad \lim_{l \rightarrow \infty} |\langle f_n, B_{\{a_1, \dots, a_{n-1}\}}^{a^{(l)}} \rangle| = \sup\{|\langle f_n, B_{\{a_1, \dots, a_{n-1}\}}^a \rangle| \mid a \in \mathbf{D}\}.$$

With the assumption (2.3) the limiting point  $a_n$  of  $a^{(l)}$  must be an interior point of  $\mathbf{D}$  unless  $f_n \equiv 0$ . In the latter case, our contention trivially holds. Otherwise, the relation (2.5) becomes

$$(2.6) \quad \lim_{l \rightarrow \infty} |\langle f_n, B_{\{a_1, \dots, a_{n-1}\}}^{a^{(l)}} \rangle| = \sup\{|\langle f_n, B_{\{a_1, \dots, a_{n-1}\}}^a \rangle| \mid a \in \mathbf{D}\}$$

Next, we compute

$$\lim_{l \rightarrow \infty} B_{\{a_1, \dots, a_{n-1}\}}^{a^{(l)}}.$$

Denote by  $P_{\{a_1, \dots, a_{n-1}\}}$  the projection operator from  $\mathbb{H}$  to  $\text{span}\{B_1, \dots, B_{n-1}\}$ . Now, there are two cases.

1. The limiting point  $e_{a_n}$  is not in  $\text{span}\{B_1, \dots, B_{n-1}\}$ . In such a case,  $a_n$ , in particular, does not coincide with any of  $a_1, \dots, a_{n-1}$ , and  $B_{\{a_1, \dots, a_{n-1}\}}^{a_n}$  is just given by

$$(2.7) \quad B_{\{a_1, \dots, a_{n-1}\}}^{a_n} = \frac{e_{a_n} - P_{\{a_1, \dots, a_{n-1}\}} e_{a_n}}{\|e_{a_n} - P_{\{a_1, \dots, a_{n-1}\}} e_{a_n}\|},$$

and  $\{B_1, \dots, B_{n-1}, B_{\{a_1, \dots, a_{n-1}\}}^{a_n}\}$  is the G-S orthogonalization of  $\{B_1, \dots, B_{n-1}, e_{a_n}\}$ .

2. The limiting point  $e_{a_n}$  lies in  $\text{span}\{B_1, \dots, B_{n-1}\}$ . In particular, if  $a_n$  coincides with one of  $a_1, \dots, a_{n-1}$ , we are in this case. We note that none of the  $e_{a^{(l)}}$  is in  $\text{span}\{B_1, \dots, B_{n-1}\}$ , for, otherwise,  $B_{\{a_1, \dots, a_{n-1}\}}^{a^{(l)}} = 0$ , and  $|\langle f_n, B_{\{a_1, \dots, a_{n-1}\}}^{a^{(l)}} \rangle|$  will have no contribution towards the maximum. We consequently have

$$\begin{aligned} B_{\{a_1, \dots, a_{n-1}\}}^{a^{(l)}} &= \frac{e_{a^{(l)}} - P_{\{a_1, \dots, a_{n-1}\}} e_{a^{(l)}}}{\|e_{a^{(l)}} - P_{\{a_1, \dots, a_{n-1}\}} e_{a^{(l)}}\|} \\ &= \frac{(e_{a^{(l)}} - P_{\{a_1, \dots, a_{n-1}\}} e_{a^{(l)}}) - (e_{a_n} - P_{\{a_1, \dots, a_{n-1}\}} e_{a_n})}{\|(e_{a^{(l)}} - P_{\{a_1, \dots, a_{n-1}\}} e_{a^{(l)}}) - (e_{a_n} - P_{\{a_1, \dots, a_{n-1}\}} e_{a_n})\|} \\ &= \frac{(e_{a^{(l)}} - e_{a_n}) - P_{\{a_1, \dots, a_{n-1}\}} (e_{a^{(l)}} - e_{a_n})}{\|(e_{a^{(l)}} - e_{a_n}) - P_{\{a_1, \dots, a_{n-1}\}} (e_{a^{(l)}} - e_{a_n})\|} \\ &= \frac{\left(\frac{e_{a^{(l)}} - e_{a_n}}{\|e_{a^{(l)}} - e_{a_n}\|}\right) - P_{\{a_1, \dots, a_{n-1}\}} \left(\frac{e_{a^{(l)}} - e_{a_n}}{\|e_{a^{(l)}} - e_{a_n}\|}\right)}{\left\|\left(\frac{e_{a^{(l)}} - e_{a_n}}{\|e_{a^{(l)}} - e_{a_n}\|}\right) - P_{\{a_1, \dots, a_{n-1}\}} \left(\frac{e_{a^{(l)}} - e_{a_n}}{\|e_{a^{(l)}} - e_{a_n}\|}\right)\right\|}. \end{aligned}$$

Extracting a subsequence if necessary, we may suppose that  $(e_{a^{(l)}} - e_{a_n})/\|e_{a^{(l)}} - e_{a_n}\|$  converges to some unit vector  $\nu \in \mathbb{H}$ . If  $\nu$  is not in  $\text{span}\{B_1, \dots, B_{n-1}\}$ , then we can take the limit in the above expression as  $a^{(l)} \rightarrow a_n$ , to obtain that  $B_{\{a_1, \dots, a_{n-1}\}}^{a^{(l)}}$  converges to  $B_{\{a_1, \dots, a_{n-1}\}}^{\partial_\nu a_n}$ , where the notation  $\partial_\nu$  indicates that we computed the directional derivative of  $e_a$  in the direction  $\nu$ . When  $\nu \in \text{span}\{B_1, \dots, B_{n-1}\}$ , then the above expression has indeterminate limiting form  $0/0$ , and higher order derivatives must be computed that we will not discuss further. Observe there may be several directions  $\nu$ , accounting for the fact that  $a_n$  needs not be unique. In particular, the directions along the real or purely imaginary axes induce partial derivatives of the reproducing kernels ([35]).

The just described theory mainly grows out from the study of the Hardy space  $H^2$  on the open unit disc. It is then applied, at least in part or adaptively, to other contexts in one or several complex variables, as well as in the Clifford algebra setting. That helps to form a concept of rational approximation in various contexts. In the following sections we will briefly review the particulars of each individual context.

### 3. THE HARDY $H^2(\mathbf{D})$ CASE

The so called AFD and Pre-Orthogonal Greedy Algorithm were originated from this context. In this section  $\mathbf{D}$  represents the open unit disc. Among the equivalent definitions of the Hardy  $H^2(\mathbf{D})$  space we will cite only

$$H^2(\mathbf{D}) = \{f(z) = \sum_{k=0}^{\infty} c_k z^k \mid c_k \in \mathbf{C}, \|f\|_{H^2} := \sum_{k=0}^{\infty} |c_k|^2 < \infty, |z| < 1\}.$$

$f \in H^2(\mathbf{D})$  implies that  $f$  is holomorphic in  $\mathbf{D}$ , and, as an important property, the partial sum of the infinite series on the unit circle has a  $L^2(\partial\mathbf{D})$ -limit that equals to the non-tangential boundary limit  $\lim_{r \rightarrow 1^-} f(re^{it})$  almost everywhere. Those boundary limits form a closed subspace of the  $L^2$  space on the boundary circle, denoted by  $H^2(\partial\mathbf{D})$ . The mapping from  $H^2(\mathbf{D})$  to  $H^2(\partial\mathbf{D})$  is an isometry. We hence identify the space  $H^2(\partial\mathbf{D})$  with  $H^2(\mathbf{D})$ . If we start from a function  $f$  in  $L^2(\partial\mathbf{D})$  with Fourier coefficients  $c_k, k = 0, \pm 1, \pm 2, \dots$ , then we have the so called Hardy spaces decomposition  $f = f^+ + f^-$ , where

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

If  $f$  is real-valued on the circle, then we have

$$f = 2\operatorname{Re}f^+ - c_0.$$

The last relation shows that approximation of functions in  $L^2(\partial\mathbf{D})$  can be reduced to that of functions in the Hardy class. In other contexts we have analogous relations, so the case of Hardy spaces on which we concentrate below will be a prototypical example.

In  $H^2(\mathbf{D})$ , the normalized reproducing kernels (also known as Szegő kernels) are the rational functions

$$(3.1) \quad e_a(z) = \frac{\sqrt{1-|a|^2}}{1-\bar{a}z}, \quad a \in \mathbf{D}.$$

Then  $\{e_a\}_{a \in \mathbf{D}}$  is a dictionary of  $H^2(\mathbf{D})$ . Let  $A = (a_1, \dots, a_n) \in \mathbf{D}^n$  be a  $n$ -tuple. If the  $a_k$  are all distinct, we associate to  $A$  the  $n$ -tuple  $(e_{a_1}, \dots, e_{a_n})$ , of normalized reproducing kernels. More generally, if  $A$  consists of  $m < n$  distinct points  $b_1, \dots, b_m$  where  $b_k$  is repeated  $l_k$  times with  $l_1 + \dots + l_m = n$ , then we associate to  $A$  the  $n$ -tuple

$$(3.2) \quad (E_{\{b_1,1\}}, \dots, E_{\{b_1,l_1\}}, E_{\{b_2,1\}}, \dots, E_{\{b_2,l_2\}}, \dots, E_{\{b_m,1\}}, \dots, E_{\{b_m,l_m\}})$$

defined as follows. We set  $E_{\{0,j\}}(z) = z^j$ , and if  $a_k \neq 0$  then  $E_{\{a_k,j\}}(z) = \frac{c(a_k,j)}{(1-\bar{a}_k z)^j}$ , where  $c(a_k, j)$  is the constant making  $\|E_{\{a_k,j\}}\| = 1$ . The orthogonalization of (3.2) is the so called Takenaka-Malmquist (TM-) system, or orthogonal rational system,  $(B_1, \dots, B_n)$ , where

$$B_k(z) = B_{\{a_1, \dots, a_k\}}(z) = \frac{\sqrt{1-|a_k|^2}}{1-\bar{a}_k z} \prod_{l=1}^{k-1} \frac{z-a_l}{1-\bar{a}_l z}$$

[36, Lecture V]. We call each  $B_k$  a *modified Blaschke product*. One recognizes that the rational function  $\frac{\sqrt{1-|a_n|^2}}{1-\bar{a}_n z}$  in front is the Szegő kernel and the product thereafter is the Blaschke product with the zeros  $a_1, \dots, a_{k-1}$ . The claim is: when studying rational approximation in  $H^2(\mathbf{D})$ , TM systems are unavoidable. This is no wonder, because every rational function is a linear combination of  $E_{b_k,j}$  for some  $b_k$ . Existing studies on TM systems include Laguerre and Kautz systems. The traditional works on TM systems deal with the condition

$$(3.3) \quad \sum_{k=1}^{\infty} (1-|a_k|) = \infty.$$

This is called the *hyperbolic non-separability condition*. For any  $p \in [1, \infty)$ , it is necessary and sufficient for the corresponding TM system to be complete in  $H^p(\mathbf{D})$  ([15]). Also, for  $1 < p < \infty$  and any sequence  $(a_1, \dots, a_n, \dots)$ , the corresponding TM system is a Schauder basis of the closure of  $\operatorname{span}\{B_n\}_{n=1}^{\infty}$  in  $H^p(\mathbf{D})$  ([31]).

The difference with the current study is that the parameters  $a_1, \dots, a_n, \dots$  used to approximate a given function are not fixed before hand, nor are they required to satisfy the hyperbolic non-separability condition (3.3), and, correspondingly, the induced TM-system  $\{B_n\}$  is not necessarily a basis. Instead, we adaptively select the parameters  $a_n$ , as in greedy algorithms, and thus formulate expansions of given signals with fast convergence. Below we will introduce our main algorithm in the unit disc case called *Adaptive Fourier Decomposition*, abbreviated as AFD ([32]).

Let  $f$  be any function in  $H^2(\mathbf{D})$ . Letting  $f_1 = f$ , recursively and for any  $n$  complex numbers  $a_1, \dots, a_n$  in  $\mathbf{D}$ , we have

$$f(z) = \langle f_1, e_{a_1} \rangle e_{a_1}(z) + f_2(z) \frac{z - e_{a_1}}{1 - \bar{a}_1 z},$$

and

$$f_2(z) = \langle f_2, e_{a_2} \rangle e_{a_2}(z) + f_3(z) \frac{z - e_{a_2}}{1 - \bar{a}_2 z},$$

etc. so that we arrive at

$$(3.4) \quad f(z) = \sum_{k=1}^n \langle f_k, e_{a_k} \rangle B_{\{a_1, \dots, a_k\}}(z) + f_{n+1}(z) \prod_{k=1}^n \frac{z - a_k}{1 - \bar{a}_k z}.$$

This identity gives rise to an interpolating rational function to  $f$  for the interpolating points  $a_1, \dots, a_n$  where repeated points correspond to interpolation with derivatives of the function.

The identity, furthermore, gives rise to fast approximation in energy if one selects  $a_k$ , once  $a_1, \dots, a_{k-1}$  have been fixed, according to the formula:

$$(3.5) \quad a_k = \arg \max \{ |\langle f_k, e_a \rangle|^2 \mid a \in \mathbf{D} \}.$$

The following energy relation is to be noted:

$$\|f\|^2 = \sum_{k=1}^n |\langle f_k, e_{a_k} \rangle|^2 + \|f_{n+1}\|^2.$$

Under the selection criterion (3.5) we can show that in the energy sense ([32]), namely in the sense of strong convergence in  $H^2(\mathbf{D})$ ,

$$f = \sum_{k=1}^{\infty} \langle f_k, e_{a_k} \rangle B_k.$$

The above described AFD, or *Core AFD* algorithm was published in 2012 and lately, in 2015, found to be equivalent with the Pre-Orthogonal Greedy Algorithm ([35]). The motivation of AFD is characterizing positive-instantaneous-frequency decomposition, or mono-component decomposition of signals. There followed two elaborations on AFD of which one is the called unwinding AFD ([24, 28, 16]) and the other is geablack towards  $n$ -best rational approximation ([33, 29]).

Unwinding AFD dwells on Core AFD and the principle of energy front-loading for Nevanlinna outer functions in one complex variable. The latter principle addresses the following fact: if  $F(z) = \sum_{k=0}^{\infty} c_k z^k$  is a  $H^2(\mathbf{D})$ -function with the Nevanlinna factorization (see [19], [43]):  $F(z) = I(z)O(z)$ , where  $I(z)$  and  $O(z)$  are, respectively, its inner and outer factors, and if we write  $O(z) = \sum_{k=0}^{\infty} d_k z^k$ , then there holds for any positive integer  $N$  that

$$\sum_{k=0}^N |c_k|^2 \leq \sum_{k=0}^N |d_k|^2.$$

Since

$$\sum_{k=0}^{\infty} |c_k|^2 = \|F\|^2 = \|O\|^2 = \sum_{k=0}^{\infty} |d_k|^2,$$

the above inequality amounts to saying that the outer part of a Hardy function has best polynomial approximation of degree  $N$  in  $H^2(\mathbf{D})$  that converges faster than that of the original function. This suggests that when decomposing a function in  $H^2(\mathbf{D})$ , it may be a good strategy to perform the Nevanlinna factorization and then decompose the outer part instead of the original function, to finally multiply back by the inner factor which is a finite Blaschke product, at least when the function is continuous on  $\partial\mathbf{D}$ . Of course, performing the Nevanlinna factorization is not such an easy business as it essentially involves estimating conjugate functions, and fair judgement should be used in each case.

The related theory is developed in [24, 28, 16]. Experiments show that Unwinding AFD is indeed among the best positive frequency decomposition methods ([30]). The authors became aware late 2015 that the Ph. D. thesis of M. Nahon ([24]) at Yale University, 2000, under the guidance of R. Coifman, develops an analogous unwinding algorithm based on the Nevanlinna factorization ([19, 43]). In a recent paper by Coifman and Steinerberger, theoretical aspects of the algorithm are further developed ([16]).

Cyclic AFD was designed to approach the problem of  $n$ -best rational approximation in  $H^2(\mathbf{D})$ . The problem is formulated as follows. Given  $f \in H^2(\mathbf{D})$ , find a rational function of the form  $p/q$ , with  $\deg\{p\}$  and  $\deg\{q\}$  not exceeding  $n$ , and  $q$  zero-free inside the unit disc, such that  $\|f - p/q\|$  is minimum among all possible rational functions of the same kind. The latter are just the rational functions of degree no larger than  $n$  in  $H^2(\mathbf{D})$ . Existence of such minimizing rational functions was proved a long time ago, but a theoretical algorithm for finding  $p/q$  to give rise to the minimum is still an open issue. A detailed account of the problem may be found in [4, 11, 5, 12, 14]. Both the RARL2 algorithm (which extends to the matrix-valued case, see [www-sop.inria.fr/apics/RARL2/rarl2.html](http://www-sop.inria.fr/apics/RARL2/rarl2.html) for a description and tutorial as well as [7, 9, 18] for further references) and the one through Cyclic AFD ([29]) provide practical algorithms. RARL2 is a descent algorithm using Schur parameters to describe Blaschke products of given degree along with a compactification thereof to ensure convergence to a local optimum. It is used in identification and design of microwave devices, see [27, 39]. The algorithm using Cyclic AFD is parameterized by the zeros of the denominator polynomial, and uses the fact that the expansion as a sum of modified Blaschke products

$$\sum_{k=1}^n \langle f_k, e_{a_k} \rangle B_{\{a_1, \dots, a_k\}}(z)$$

is a rational function of degree no larger than  $n$  by construction. The theory and algorithmic scheme of Cyclic AFD are definitely simpler (though the 1-D search over  $a \in \mathbf{D}$  iterated at each step to reach a fixed point is nontrivial), but convergence to a local minimum is still an issue. Cyclic AFD corresponds to simultaneous optimal selection of  $n$  parameters, while Core AFD corresponds to sequential selection of  $n$  parameters ([29]). For some related studies we refer to [5, 8, 14, 13]. Other algorithms based on fixed point heuristics or balanced truncation of Hankel operators to approach rational  $H^2(\mathbf{D})$ -approximation can be found, *e.g.* in [41, 22].

The above considerations and algorithms extend to the context of the half-plane rather than the disk, by means of conformal mapping. The reason is that a conformal map from the disk to the half-plane is a rational function of degree 1 (*i.e.* a Möbius transform) and therefore it preserves rationality and the degree, see [19, Ch. I]. There is also a parallel approach by using TM systems and the corresponding maximal principle in the half-plane. More generally, the relevant extension of what precedes to more general domains is that of best *meromorphic approximation with  $n$  poles*, see [10].

For functions of multivalent variables, finding a basis is equivalent to finding a uniqueness set. That is a fundamental task and therefore of great interest in mathematical analysis. It is, however, in many cases difficult to achieve. On the other hand, Szegő kernels are usually simple rational functions, and fast representation of signals as linear combinations of Szegő kernels has great significance in relation to applications. The principles outlined in the last two sections are valid and the results available in a number of contexts for several real and complex variables, and with functions valued in vectors and matrices. We give a brief introduction to this circle of ideas in the following sections.

#### 4. QUATERNIONIC AND CLIFFORD CONTEXTS FOR FUNCTIONS OF SEVERAL REAL VARIABLES

Denote by  $\mathbf{Q}$  the quaternion algebra, and by  $\mathcal{A}_m$  the Clifford algebra of linear dimension  $2^m$ . The analogues of AFD have been formulated for Hardy spaces in the setting of  $\mathbf{Q}$  and of  $\mathcal{A}_m$  in [34] [42] respectively. The first setting is very much in the spirit of AFD.



The second, however, is more in the spirit of General Greedy Algorithm with dictionary consisting of the higher order Szegő kernels, due to the fact that the inner product of Clifford algebra-valued functions is not necessarily scalar-valued. That prevents the G-S orthogonalization process from being applied.

Define  $R_+^{m+1} = \{x = (x_0, \underline{x}) \in R^{m+1}; x_0 > 0, \underline{x} \in R^m\}$ . We briefly introduce the AFD (General Greedy Algorithm) in the Hardy space  $H^2(R_+^4)$  of the upper-half space (resp.  $H^2(R_+^{m+1})$ ), which consists of  $\mathbf{Q}$ -valued (resp.  $\mathcal{A}_m$ -valued) left monogenic functions [21]. For the parallel theory in the Hardy spaces on the unit ball, see [34, 42]. Denote by  $\phi_a(x)$  the non-normalized Szegő kernel of  $H^2(R_+^4)$  (resp.  $H^2(R_+^{m+1})$ ), i.e.  $\phi_a(x) = \frac{\overline{x+\bar{a}}}{|x+\bar{a}|^4}$  (resp.  $\phi_a(x) = \frac{\overline{x+\bar{a}}}{|x+\bar{a}|^{m+1}}$ ). Regarding  $H^2(R_+^4)$ , although  $\mathbf{Q}$  is a non-commutative algebra, it is not difficult to apply the P-OGA to  $\{\phi_a(x), a, x \in R_+^4\}$  to obtain an orthonormal system  $\{\mathcal{B}_n\}_{n=1}^\infty$  parameterized by the selected sequence  $\{a^{(n)}\}_{n=1}^\infty$  in  $R_+^4$ . Similar to AFD, one has  $\lim_{n \rightarrow \infty} \|f - \sum_{k=1}^n \mathcal{B}_k \langle f, \mathcal{B}_k \rangle\| = 0$ . We note that the study of AFD in the setting of  $\mathbf{Q}$  predates that of P-OGA: the latter is a generalization of the former.

As to  $H^2(R_+^{m+1})$ , we introduce the completion of the Szegő kernel dictionary given by

$$\tilde{\mathbf{D}} = \left\{ \frac{\psi_{\beta,a}(x)}{\|\psi_{\beta,a}\|}, \beta = (\beta_1, \dots, \beta_{m+1}), a, x \in R_+^{m+1} \right\},$$

where  $\psi_{\beta,a}(x) = \frac{\partial^{|\beta|}}{\partial a_0^{\beta_1} \dots \partial a_{m+1}^{\beta_{m+1}}} \phi_a(x)$ . For each  $f \in H^2(R_+^{m+1})$ , applying the General Greedy Algorithm with  $\tilde{\mathbf{D}}$  one has  $\lim_{n \rightarrow \infty} \|f - \sum_{k=1}^n \frac{\psi_{\beta^{(k)}, a^{(k)}}}{\|\psi_{\beta^{(k)}, a^{(k)}}\|} \langle g_k, \frac{\psi_{\beta^{(k)}, a^{(k)}}}{\|\psi_{\beta^{(k)}, a^{(k)}}\|} \rangle\| = 0$ , where  $g_k$  is the standard remainder defined in §2.

The sphere cases are also considered in the quaternionic and Clifford algebra settings ([34, 42]).

In both settings it is shown that a global maximal selection of the parameter is attainable at each step of the recursive process (i.e.  $\rho = 1$ ). In particular, one can obtain rational approximations of functions in  $L^2(R^4)$  by applying the corresponding AFD and the well-known Sokhotskiy-Plemelj formula (e.g. [21, 34]).

## 5. SEVERAL COMPLEX VARIABLES

Similar approximation schemes in settings involving several complex variables have also been studied. In fact, one can consider functions defined on various classical domains, with values in  $\mathbf{C}^N$ , or  $\mathbf{C}^{N \times M}$  matrices, etc.

**5.1. Functions Defined on  $n$ -Torus.** Denote by  $T^n$  the  $n$ -torus, where  $T = \partial \mathbf{D}$ . There are two generalizations of AFD in the Hardy space of the  $n$ -torus  $H^2(T^n)$  ([35]). One merely consists in processing P-OGA in this context (i.e. it is shown that P-OGA is applicable to  $H^2(T^n)$ ). The other is based on the product-TM system. As to the former, we omit the details as it should be clear from the previous section already how to perform. Below, we give a brief introduction to the latter. For simplicity, we consider only the case  $n = 2$ .

Denote by  $B_k^{\mathbf{a}}(z)$  the modified Blaschke product (a member of the TM system) associated with the sequence  $\mathbf{a} = \{a_k\}_{k=1}^\infty$  in  $\mathbf{D}$ . We introduce the *tensor product type modified Blaschke product*  $\{B_k^{\mathbf{a}}(z) \otimes B_l^{\mathbf{b}}(w)\}$ , where  $\mathbf{a}, \mathbf{b} \subset \mathbf{D}$  and  $(z, w) \in \mathbf{D}^2 = \mathbf{D} \times \mathbf{D}$ . For  $f \in H^2(T^2)$ , we look for a rational approximation of  $f$  of separable type given by  $f = \lim_{m \rightarrow \infty} \sum_{1 \leq k, l \leq m} \langle f, B_k^{\mathbf{a}} \otimes B_l^{\mathbf{b}} \rangle B_k^{\mathbf{a}} \otimes B_l^{\mathbf{b}} = \lim_{m \rightarrow \infty} S_m(f)$  in the  $H^2$ -norm. Denote by  $D_m(f) = S_m(f) - S_{m-1}(f)$  the  *$m$ -partial sum difference*. The main step is to select  $(a_{m+1}^*, b_{m+1}^*) \in \mathbf{D}^2$  according to the maximal problem

$$(5.1) \quad (a_{m+1}^*, b_{m+1}^*) := \arg \sup_{(a_{m+1}, b_{m+1}) \in \mathbf{D}^2} \|D_{m+1}(f)\|^2,$$

where  $\{a_1, \dots, a_m\}$  and  $\{b_1, \dots, b_m\}$  are previously fixed. The existence of  $(a_{m+1}^*, b_{m+1}^*)$  is proved in [35] through a technical discussion. In a way similar to previously described

AFD,  $S_m(f)$  converges to  $f$  in the  $H^2$ -norm if each  $(a_k, b_k)$  is selected according to criterion (5.1). As an application, one can obtain rational approximations of functions in  $L^2(T^2)$ .

**5.2. Functions Defined on  $R^n$  in the Setting of Hardy Spaces on Tubes.** As mentioned in the previous sections, one can obtain rational approximations of functions in  $L^2$  of the boundary of a domain by applying AFD in the domain. Since  $R^n$  can be written as the union of  $2^n$  octants, seeking rational approximations of functions in  $L^2(R^n)$  motivates the study of AFD in Hardy spaces on tubes over octants ([23]). For the purpose of illustration, it suffices here to investigate AFD in the Hardy space of the tube over the first octant  $H^2(T_{\Gamma_1})$ , where  $T_{\Gamma_1} = \mathbf{C}_+ \times \cdots \times \mathbf{C}_+$ .

Denote by  $S_z(w)$  the Cauchy- Szegő kernel of  $H^2(T_{\Gamma_1})$ , i.e.  $S_z(w) = \prod_{j=1}^n \frac{-1}{2\pi i(w_j - \bar{z}_j)}$ . By using the methodology P-OGA, one can obtain an orthonormal system  $\{\mathcal{B}_k\}_{k=1}^\infty$  parameterized by the sequence  $\{z^{(k)}\}_{k=1}^\infty \subset T_{\Gamma_1}$ . Indeed,  $\{\mathcal{B}_k\}_{k=1}^\infty$  is the G-S orthogonalization of the selected Cauchy- Szegő kernels and, if necessary, their higher order directional derivatives. It is concluded in [23] that the attainability of a global maximal selection at each step follows from a certain kind of boundary behavior of functions in  $H^2(T_{\Gamma_1})$ , called “boundary vanishing condition (BVC)” (also see [35]). After verifying the BVC in  $H^2(T_{\Gamma_1})$ , the convergence follows from the general theory given in [35]. In [23] the P- OGA generalization of AFD in the Hardy spaces on tubes *over regular cones* is also given.

**5.3. Matrix-valued Signals Defined in the Unit Disc.** We denote by  $H_2^{p \times q}$  the space of  $p \times q$  matrices with entries in  $H^2(\mathbf{D})$ . In a recent paper of D. Alpay, F. Colombo, T. Qian and I. Sabadini they show that it is possible, as in the scalar case, to decompose those functions as linear combinations of suitably modified matrix-valued Blaschke product, in an adaptive way. The procedure is based on a generalization to the matrix-valued case of the maximum selection principle of 1-D AFD, which involves not only selections of suitable points in the unit disc but also suitable orthogonal projections. It can be shown that the maximum selection principle again gives rise to a convergent algorithm ([1]). The analogous parametrization in terms of Schur analysis and tangential interpolation directions was given earlier in [6], and has been used to design a matrix-valued version of the RARL2 algorithm, see [18].

**5.4. Adaptive Decomposition: the Case of the Drury-Arveson Space.** Blaschke factors and products have counterparts in the unit ball of  $C^N$ , and this fact allows us to extend the maximum selection principle to the case of functions in the Drury-Arveson space of functions analytic in the unit ball of  $C^n$ . This gives rise to an algorithm which is a variation of the higher dimensional AFD. In the corresponding paper of D. Alpay, F. Colombo, T. Qian and I. Sabadini they also introduce infinite Blaschke products in this setting and study their convergence ([2]).

**5.5. Matrix-valued Signals Defined on the Polydisc.** The polydisc case has been given special attention, due to its connection with image processing. It develops in the context of multi-trigonometric series, like 2-D AFD treated in an earlier subsection. An alternative setting is given in the third paper of D. Alpay, F. Colombo, T. Qian and I. Sabadini where they develop interpolation theory as well as an operator-valued Blaschke product method that offers an adaptive expansion of holomorphic functions in the Hardy space over the polydisc corresponding to signals on the n-torus ([3]).

## 6. AFD AND AVEIRO METHOD IN REPRODUCING KERNEL HILBERT SPACES

We first note that the Hardy  $H^2$  space is a reproducing kernel Hilbert space, where the reproducing kernel is given by the Szegő kernel. Subsequently, P-OGA was proposed as an expansion algorithm in reproducing kernel Hilbert spaces (see §2) although in the previous sections we restrict ourselves to AFD in various Hardy spaces. The key of AFD (or

P-OGA) is the construction of an orthonormal system by applying the G-S orthogonalization process to the selected reproducing kernels and their higher order derivatives. Such a construction ensures that the approximating function and its derivatives meet interpolation conditions to the approximated function at the selected points.

The study of interpolating functions is closely related to interpolation and sampling problems in reproducing kernel Hilbert spaces, some prototypical aspects of which may be found in [38]. We use the notation  $\mathbb{H}_K$  to indicate that  $\mathbb{H}$  is a reproducing kernel Hilbert space admitting a reproducing kernel  $K(q, \bar{p})$ . Suppose that  $\mathbb{H}_K$  consists of holomorphic functions defined in an open set  $E \subset \mathbb{C}$ . Let further  $\{p_k\}_{k=1}^{\infty} \subset E$  be a sequence of distinct points. The so-called Aveiro Method, proposed by S. Saitoh et al in [37], aims at constructing an approximating function to  $f \in \mathbb{H}_K$  involving a finite number of sampling points  $\{p_1, \dots, p_n\}$ .

Based on this work, the authors of [20] propose the so-called ‘‘Aveiro Method under complete dictionary (AMUCD)’’ by combining the ideas of P-OGA with Aveiro Method. Roughly speaking, AMUCD enhances the power of Aveiro Method in that the approximating function given by AMUCD does not require all elements of  $\{p_k\}_{k=1}^{\infty}$  to be distinct. As in AFD, the representation and its derivatives enjoy interpolation properties at  $\{p_k\}_{k=1}^{\infty}$ . It is shown in [20] that AMUCD is applicable to the classical Hardy spaces and Paley-Wiener spaces. It turns out that AMUCD is, in fact, an alternative representation of AFD. Nevertheless, AMUCD has the advantage not to require working out the related orthonormal system, whereas in many instances of P-OGA one does not know explicit formulas for the related orthonormal system of functions.

#### REFERENCES

- [1] D. Alpay, F. Colombo, T. Qian, I. Sabadini, *Adaptive orthonormal systems for matrix-valued functions*, preprint, arXiv:1511.07834.
- [2] D. Alpay, F. Colombo, T. Qian, I. Sabadini, *Adaptive Decomposition: the Case of the Drury-Arveson Space*, preprint, arXiv:
- [3] D. Alpay, F. Colombo, T. Qian, I. Sabadini, *Adaptive Decomposition: the Polydisc Case*, preprint.
- [4] L. Baratchart, *Existence and Generic Properties of  $L^2$  Approximants for Linear Systems*, I.M.A. Journal of Math. Control and Information, Vol. 3, pp 89-101, 1986.
- [5] 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.
- [6] L. Baratchart and D. Alpay and A. Gombani, *On the differential structure of matrix-valued inner functions*, Operator Theory: Advances and Applications, vol. 73, pp 30-68, 1994.
- [7] L. Baratchart, M. Cardelli, M. Olivi, *Identification and rational  $L^2$  approximation, a gradient algorithm*, Automatica, 27(1991), pp. 413-418.
- [8] Laurent Baratchart and Sylvain Chevillard and Tao Qian, *Minimax principle and lower bounds in  $H^2$ -rational approximation*, Journal of Approximation Theory, vol. 206, pp. 17-47, 2016.
- [9] L. Baratchart, M. Olivi, *Critical points and error rank in matrix  $H^2$  rational approximation and the strong differential consistency of output error identification from white noise inputs*, Constructive Approximation 14, pp. 273-300 (1998).
- [10] L. Baratchart and F. Mandrea and E.B. Saff and F. Wielonsky, *2-D inverse problems for the Laplacian: A meromorphic approximation approach*, J. de Mathématiques Pures et Appl., vol. 86, pp. 1-41, 2006.
- [11] L. Baratchart and E. B. Saff and F. Wielonsky, *A criterion for uniqueness of a critical point in  $H^2$  rational approximation*, J. Analyse Mathématique, vol. 70, pp. 225-266, 1996.
- [12] L. Baratchart and H. Stahl and F. Wielonsky, *Asymptotic uniqueness of best rational approximations of given degree to Markov functions in  $L^2$  of the circle*, Constructive Approximation, vol. 17, pp 103-138, 2001.
- [13] L. Baratchart, H. Stahl and M. Yattselev, *Weighted extremal domains and rational approximation*, Advances in Maths, vol. 229, pp. 357-407, 2012.
- [14] L. Baratchart, M. Yattselev, *Asymptotic uniqueness of best rational approximations to complex Cauchy transforms in  $L^2$  of the circle*, Recent Trends in Orthogonal Polynomials and Approximation Theory, volume 507 of Contemporary Mathematics, page 87-111, Amer. Math. Soc., Providence, RI, 2010.
- [15] A. Bultheel, P. Gonzalez-Vera, E. Hendriksen, O. Njåstad, *Orthogonal Rational Functions*,
- [16] R. Coifman, S. Steinerberger, *Nonlinear phase unwinding of functions*, preprinted on ArXiv.
- [17] R.A. DeVore and V.N. Temlyakov, *Some remarks on Greedy Algorithms*, Adv. Comput. Math. 5 (1996), 173-187.
- [18] P. Fulcheri and M. Olivi, *Matrix rational  $H^2$  approximation: a gradient algorithm based on Schur analysis*, SIAM J. Control Optim., 36(6), pp. 2103-2127, 1998.

- [19] J. B. Garnett, *Bounded Analytic Functions*, Academic Press 1987.
- [20] W.-X. Mai and T. Qian, *Aveiro method in reproducing kernel hilbert spaces under complete dictionary*, submitted.
- [21] J. E. Gilbert and M. A. Murray, *Clifford algebras and Dirac operators in harmonic analysis*, Cambridge Studies in Advanced Mathematics 26, Cambridge University Press 1991.
- [22] S. Gugercin and A. Antoulas and C. Beattie, *H2 model reduction for large-scale linear dynamical systems*, SIAM J. MATRIX ANAL. APPL., vol. 30 (2), pp. 609-638, 2008.
- [23] W.-X. Mai and T. Qian, *Rational approximation of functions in the hardy spaces on tubes*, submitted.
- [24] M. Nahon, *Phase Evaluation and Segmentation*, Ph.D. Thesis, Yale University, 2000.
- [25] Y. Mo and T. Qian, *Support vector machine adapted Tikhonov regularization method to solve Dirichlet problem*, Applied Mathematics and Computation, 2014, 245: 509-519.
- [26] Y. Mo, T. Qian and W. Mi, *Sparse Representation in Szegő Kernels through Reproducing Kernel Hilbert Space Theory with Applications*, accepted by International Journal of Wavelet, Multiresolution and Information Processing.
- [27] M. Olivi and F. Seyfert and J.P. Marmorat, *Identification of microwave filters by analytic and rational  $H^2$  approximation*, Automatica, vol. 49, pp. 317–325, 2013.
- [28] 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.
- [29] T. Qian, *Cyclic AFD Algorithm for Best Approximation by Rational Functions of Given Order*, Mathematical Methods in the Applied Sciences, Article first published online: 28 JUN 2013 DOI: 10.1002/mma.2843.
- [30] T. Qian, H. Li, M. Stessin, *Comparison of Adaptive Mono-component Decompositions*, Nonlinear Analysis: Real World Applications. Volume 14, Issue 2, April 2013, Pages 1055V1074.
- [31] T. Qian, Q-H. Chen, L-H. Tan, *Rational Orthogonal Systems are Schauder Bases*, Complex Variables and Elliptic Equations, 59 (2014), pp. 841–846.
- [32] T. Qian, Y-B. Wang, *Adaptive fourier series- a variation of greedy algorithm*, Advances in Computation Mathematics, 34 (2011), pp. 279–293.
- [33] T. Qian and E. Wegert, *Optimal Approximation by Blaschke Forms*, Complex Variables and Elliptic Equations, DOI: 10.1080/17476933.2011.557152, Volume 58, Issue 1, 2013, page 123-133.
- [34] T. Qian, W. Sproessig, and J.-X. Wang, *Adaptive fourier decomposition of functions in quaternionic hardy spaces*, Mathematical Methods in Applied Sciences, 35 (2012), pp. 43–64.
- [35] T. Qian, W. Sproessig, *Two-dimensional adaptive fourier decomposition*, arXiv:1406.2008 [math. CV], (2014).
- [36] N. Nikolskii, *Treatise on the shift operator*, Springer 1986.
- [37] L. P. Castro, H. Fujiwara, M. M. Rodrigues, and S. Saitoh, *A new discretization method by means of reproducing kernels*, Science and Technology Publication House, 2012, pp. 185–223.
- [38] K. Seip, *Interpolation and sampling in spaces of analytic functions*. University Lecture Series, vol.33, American Mathematical Society, Providence, RI, 2004.
- [39] F. Seyfert and M. Oldoni and M. Olivi and S. Lefteriu and D. Pacaud, *Deembedding of filters in multiplexers via rational approximation and interpolation*, Internat. Jour. of RF and Microwave Computer-Aided Engineering, Wiley, 2015.
- [40] V. N. Temlyakov, *Greedy algorithm and m-Term trigonometric approximation*, Constr. Approx., 1998, Pages: 569-587.
- [41] P. Van Dooren and K.A. Gallivan and P.-A. Absil, *H2 optimal model reduction of MIMO systems*, Applied Mathematics Letters, vol. 21, pp. 1267–1273, 2008.
- [42] J. X. Wang and T. Qian, *Approximation of monogenic functions by higher order szegő kernels on the unit ball and the upper half space*, Science China: Mathematics, 57 (2014).
- [43] G. Weiss and M. Weiss, *A derivation of the main results of the theory of Hp-spaces*, Rev. Un. Mat. Argentina 20 1962 63-71.

INRIA, 2004 ROUTE DES LUCIOLES, BP 93, 06902 SOPHIA-ANTIPOLIS CEDEX, FRANCE.  
 E-mail address: Laurent.Baratchart@inria.fr

DEPARTMENT OF MATHEMATICS, FACULTY OF SCIENCE AND TECHNOLOGY, UNIVERSITY OF MACAU, MACAU, CHINA.  
 E-mail address: fsttq@umac.mo