Banach J. Math. Anal. (2024) 18:2
https://doi.org/10.1007/s43037-023-00312-y

Tusi Mathematical Research Group

ORIGINAL PAPER

Orthogonalization in Clifford Hilbert modules and applications

Jinxun Wang[1] · Tao Qian[2]

Received: 24 April 2023 / Accepted: 4 October 2023 / Published online: 31 October 2023

Abstract
We prove that the Gram–Schmidt orthogonalization process can be carried out in Hilbert modules over Clifford algebras, in spite of the un-invertibility and the un-commutativity of general Clifford numbers. Then, we give two crucial applications of the orthogonalization method. One is to give a constructive proof of existence of an orthonormal basis of the inner spherical monogenics of order k for each $k \in \mathbb{N}$. The second is to formulate the Clifford Takenaka–Malmquist systems, or in other words, the Clifford rational orthogonal systems, as well as to define Clifford Blaschke product functions, in both the unit ball and the half space contexts. The Clifford TM systems then are further used to establish an adaptive rational approximation theory for L^2 functions in $\mathbb{R}^m$ and on the sphere.

Keywords Takenaka–Malmquist system · Adaptive approximation · Clifford algebra · Monogenic Hardy space

Mathematics Subject Classification 30G35 · 41A20 · 46S10

1 Introduction

Due to importance of orthonormal bases in both theoretical analysis and practical applications, for a system of functions, $\mathscr{F}$, in a Hilbert space, the questions of exis-

Communicated by Deguang Han.

✉ Tao Qian
tqian@must.edu.mo

Jinxun Wang
wjx@gdufs.edu.cn

1 School of Mathematics and Statistics, Guangdong University of Foreign Studies, Guangzhou 510006, China

2 Macau University of Science and Technology, Macau, China

tence, and explicit composition if existing, of an orthogonal system equivalent to $\mathscr{F}$ naturally arise. If the functions in $\mathscr{F}$ are complex-valued, existence of an orthonormal basis of $\mathscr{F}$ is guaranteed by the Gram–Schmidt (GS) orthogonalization process, but in the case of Clifford number-valued functions it is not obvious, since Clifford algebras are non-commutative and Clifford numbers are usually un-invertible. In Clifford analysis there is a problem: How to find an orthonormal basis for the Fueter polynomials of degree k (i.e., the inner spherical monogenics of order k)? This problem appears, because there are more than one Fueter polynomials of degree k and they are not mutually orthogonal. In [6] the existence was proved by induction on dimensions, but no explicit forms were given. In [3] an explicit form was constructed in three dimensions using the Gelfand–Tsetlin bases, but the construction is too complicated for higher dimensions. The mentioned construction is also applicable to the Hermitean Clifford analysis and to some other systems as well ([5, 7]).

In this paper we show that the GS orthogonalization process can be applied to general Clifford module Hilbert spaces. This is through proving that the orthogonal projection of a function onto the subspace spanned by some other function exists. We present here a direct construction of an orthonormal basis for a system of Clifford number-valued functions. When we consider some common and familiar functions, such as Fueter polynomials and parameterized Szegö kernel functions, the construction has a concrete expression involving inverse of the Clifford-valued inner product. As applications, we give a constructive proof of existence of an orthonormal basis of the inner spherical monogenics of order k for each $k \in \mathbb{N}$, and generalize the Takenaka–Malmquist (TM) systems into higher dimensions. It is well-known that the TM, or the rational orthogonal systems in one complex variable have attracted, and being attracting as well, great interest among analysts due to their theoretical involvements and applications.

The extended orthogonalization method has potential to establish approximation methods to Clifford number-valued functions of finite energy on manifolds of $\mathbb{R}^n$. In this paper we emphasize a direct extension of the recently established adaptive approximation by TM systems in the one complex variable case ([9, 12, 14]). The type of adaptive approximation theory has also been generalized to some several complex variables and matrix-valued contexts with applications, see the works [1, 2, 17] by Alpay et al.

The paper is organized as follows. In Sect. 2 we review Clifford algebra and Clifford analysis. In Sect. 3 we study orthogonalization of function systems in the right $\mathscr{A}_m$-module inner product space. In Sect. 4 we give definitions for TM systems and Blaschke products in general higher dimensions. In the last section, we study adaptive approximation by Clifford TM systems in the unit ball and half space in higher dimensions.

2 Preliminaries

In this paper we work on the real Clifford algebra $\mathscr{A}_m$ that generated by an orthonormal basis $\{e_1, \dots, e_m\}$ of $\mathbb{R}^m$ with the (non-commutative) multiplication rule:

$$e_i e_j + e_j e_i = -2\delta_{ij}, \quad i, j = 1, \dots, m,$$

where δ_{ij} equals 1 if $i = j$ and 0 otherwise. Each element x in $\mathscr{A}_m$ is of the form:

$$x = \sum_{T \in \mathcal{P}N} x_T e_T,$$

where $x_T \in \mathbb{R}$, $e_T = e_{i_1,\dots,i_l} := e_{i_1} e_{i_2} \dots e_{i_l}$ is the basic element of $\mathscr{A}_m$, $T = \{i_1, \dots, i_l\}$, $1 \le i_1 < \cdots < i_l \le m$, $\mathcal{P}N$ is the set consisting of all the ordered subsets of $\{1, \dots, m\}$. In addition we set $x_\emptyset = x_0$, $e_\emptyset = e_0$, e_0 is identified with the multiplication unit "1". The multiplication of Clifford numbers is determined by the multiplication of the basic elements through linearity and the law of distribution. Let e_A, e_B be any two basic elements in $\mathscr{A}_m$, their multiplication is defined by

$$e_A e_B = (-1)^{\#(A \bigcap B)} (-1)^{p(A,B)} e_{A \triangle B},$$

where $p(A, B) = \sum_{j \in B} p(A, j)$, $p(A, j) = \#\{i \in A : i > j\}$, $A \triangle B = (A \setminus B) \bigcup (B \setminus A)$ is the symmetric difference of A and B. Together with the multiplication, $\mathscr{A}_m$ is an associative algebra of dimension 2^m.

For $x = \sum_T x_T e_T \in \mathscr{A}_m$, we call x_0 the real part or scalar part of x, denote it by Sc x. NSc $x := x -$ Sc x is then the non-scalar part of x. The norm and the conjugate of x are defined by $|x| = (\sum_T x_T^2)^{1/2}$ and $\overline{x} = \sum_T x_T \overline{e_T}$, respectively, where $\overline{e_T} = \overline{e_{i_l}} \cdots \overline{e_{i_2}}\, \overline{e_{i_1}}$, and $\overline{e_0} = e_0$, $\overline{e_i} = -e_i$ for $i \neq 0$. For any $x, y, z \in \mathscr{A}_m$, there hold $\overline{xy} = \overline{y}\,\overline{x}$, $(xy)z = x(yz)$, and $|xy| \le 2^{m/2}|x||y|$. The real numbers, complex numbers and quaternions are special cases of Clifford algebra, i.e., we have $\mathscr{A}_0 = \mathbb{R}$, $\mathscr{A}_1 = \mathbb{C}$, and $\mathscr{A}_2 = \mathbb{H}$.

For any $x \in \mathscr{A}_m$, we have $\mathrm{Sc}(\overline{x}x) = \mathrm{Sc}(x\overline{x}) = |x|^2$. If $x \in \mathscr{A}_m$ is of vector form, i.e., $x = \sum_{i=0}^m x_i e_i \in \mathbb{R}^{m+1}$, then obviously $\overline{x}x = x\overline{x} = |x|^2$. Consequently, in such a vector case, the inverse of x is given by $x^{-1} = \overline{x}/|x|^2$ when $x \neq 0$. However, for a general Clifford number x, the inverse of x may not exist. That is to say, the Clifford algebra $\mathscr{A}_m$ is not a division algebra. Here we give a criterion for a Clifford number being invertible or not.

Proposition 2.1 *Let $a \in \mathscr{A}_m$, the following conclusions are equivalent:*

1. *The equation $ax = 0$ (or $xa = 0$) has only zero solution $x = 0$.*
2. *a is invertible, i.e., there exists a unique $b \in \mathscr{A}_m$, such that $ab = ba = 1$.*
3. *there exists $b \in \mathscr{A}_m$, such that $ab = 1$ (or $ba = 1$).*

Proof (1) $\Rightarrow$ (2): Note that the equation $ax = 0$ can be written in the matrix form $AX = 0$, where A is a $2^m \times 2^m$ matrix associated with a, $X = (x_0, x_1, \dots)^\top$ is the column vector whose components correspond to those of its algebraic representation.

From this viewpoint, $ax = 0$ has only zero solution $x = 0$ means that the linear system of equations $AX = 0$ has only zero solution $X = 0$. Therefore, the matrix A is invertible, and the equation $AX = (1, 0, \dots, 0)^\top$ has a unique solution, given by $X = A^{-1}(1, 0, \dots, 0)^\top$, which also gives the unique $b \in \mathscr{A}_m$, such that $ab = 1$. To prove $ba = 1$, consider the equation $xa = 0$, we get $x = xab = 0$, so $xa = 0$ has only zero solution $x = 0$. Similarly, we get a unique $c \in \mathscr{A}_m$, such that $ca = 1$, and $c = cab = b$, hence $ba = 1$.

(2) $\Rightarrow$ (3): It is obvious.

(3) $\Rightarrow$ (1): Similar to the proof of (1) $\Rightarrow$ (2). □

Clifford analysis was founded by Brackx et al. ([4]). As a generalization of complex analysis and quaternionic analysis into higher dimensional spaces, Clifford analysis is a theory on Clifford monogenic functions. A function $f = \sum_{T \in \mathcal{P}N} f_T e_T$, defined on an open subset Ω of $\mathbb{R}^{m+1}$, taking values in $\mathscr{A}_m$, is said to be left monogenic on Ω if it satisfies the generalized Cauchy–Riemann equation:

$$Df = \sum_{i=0}^{m} e_i \frac{\partial f}{\partial x_i} = \sum_{i=0}^{m} \sum_{T \in \mathcal{P}N} \frac{\partial f_T}{\partial x_i} e_i e_T = 0,$$

for all $x \in \Omega$, where the Dirac operator D is defined by

$$D = \frac{\partial}{\partial x_0} + \nabla = \sum_{i=0}^{m} e_i \frac{\partial}{\partial x_i}.$$

If f is left monogenic, then $\triangle f = \overline{D}(Df) = 0$, so f is harmonic. The set of all left monogenic functions on Ω constitutes a right $\mathscr{A}_m$-module.

If f is left monogenic on Ω and continuous on $\overline{\Omega}$, then there holds Cauchy's integral formula:

$$f(x) = \frac{1}{\omega_m} \int_{y \in \partial\Omega} E(y - x) n(y) f(y) dS, \quad x \in \Omega,$$

where $E(x) = \frac{\overline{x}}{|x|^{m+1}}$ is the Cauchy kernel, $\omega_m = 2\pi^{\frac{m+1}{2}} / \Gamma(\frac{m+1}{2})$ is the area of the unit sphere in $\mathbb{R}^{m+1}$, $n(y)$ is the outward-pointing unit normal vector and dS is the surface area element on $\partial\Omega$.

For right monogenic functions there is a parallel theory.

3 Orthogonalization in Hilbert modules over Clifford algebras

In this section, we discuss the orthogonalization problem of a right $\mathscr{A}_m$-module inner product space (for the case of left $\mathscr{A}_m$-modules one can similarly formulate). First, we give some definitions (cf. [4]).

Definition 3.1 A space $\mathscr{H}$ is called a right $\mathscr{A}_m$-module if the following conditions are fulfilled:

1. $(\mathscr{H}, +)$ is an abelian group.
2. A multiplication $(f, \lambda) \to f\lambda$ from $\mathscr{H} \times \mathscr{A}_m$ to $\mathscr{H}$ is defined such that for all $\lambda, \mu \in \mathscr{A}_m$ and $f, g \in \mathscr{H}$ there holds

   (1) $f(\lambda + \mu) = f\lambda + f\mu$.
   (2) $f(\lambda\mu) = (f\lambda)\mu$.
   (3) $(f + g)\lambda = f\lambda + g\lambda$.
   (4) $fe_0 = f$.

Definition 3.2 A space $\mathscr{H}$ is called a right $\mathscr{A}_m$-module normed space if the following conditions are fulfilled:

1. $\mathscr{H}$ is a right $\mathscr{A}_m$-module.
2. A norm $\|\cdot\|$ is defined on $\mathscr{H}$, such that

   (1) $\|f\| \geq 0$ for all $f \in \mathscr{H}$, and $\|f\| = 0$ if and only if $f = 0$.
   (2) There is a real positive constant C, such that $\|f\lambda\| \leq C|\lambda|\|f\|$ for all $\lambda \in \mathscr{A}_m$, $f \in \mathscr{H}$, and $\|f\lambda\| = |\lambda|\|f\|$ for all $\lambda \in \mathbb{R}$, $f \in \mathscr{H}$.
   (3) $\|f + g\| \leq \|f\| + \|g\|$ for all $f, g \in \mathscr{H}$.

Definition 3.3 A space $\mathscr{H}$ (in which the element is also named "function") is called a right $\mathscr{A}_m$-module inner product space if the following conditions are fulfilled:

1. $\mathscr{H}$ is a right $\mathscr{A}_m$-module.
2. An inner product $(f, g) \to \langle f, g\rangle$ from $\mathscr{H} \times \mathscr{H}$ to $\mathscr{A}_m$ is defined such that for all $\lambda, \mu \in \mathscr{A}_m$ and $f, g, h \in \mathscr{H}$ there holds

   (1) $\langle f, g\rangle = \overline{\langle g, f\rangle}$.
   (2) $\langle f\lambda + g\mu, h\rangle = \langle f, h\rangle\lambda + \langle g, h\rangle\mu$.
   (3) $\mathrm{Sc}\langle f, f\rangle \geq 0$, and $\mathrm{Sc}\langle f, f\rangle = 0$ if and only if $f = 0$.
   (4) $|\mathrm{Sc}\langle f, g\rangle| \leq \sqrt{\mathrm{Sc}\langle f, f\rangle}\sqrt{\mathrm{Sc}\langle g, g\rangle}$.

We have the following propositions for the right $\mathscr{A}_m$-module inner product space.

Proposition 3.4 *Let $\mathscr{H}$ be a right $\mathscr{A}_m$-module inner product space, then for any $f, g \in \mathscr{H}$*

$$|\langle f, g\rangle| \leq 2^{\frac{m}{2}}\sqrt{\mathrm{Sc}\langle f, f\rangle}\sqrt{\mathrm{Sc}\langle g, g\rangle}.$$

In particular

$$|\langle f, f\rangle| \leq 2^{\frac{m}{2}}\mathrm{Sc}\langle f, f\rangle.$$

Proof Writing $\langle f, g\rangle = \sum_{T\in\mathcal{P}N}\langle f, g\rangle_T e_T$, for every $T \in \mathcal{P}N$ we get

$$\begin{aligned}
\langle f, g\rangle_T^2 &= (\mathrm{Sc}(\overline{e_T}\langle f, g\rangle))^2 \\
&= (\mathrm{Sc}\langle f, ge_T\rangle)^2 \\
&\leq (\mathrm{Sc}\langle f, f\rangle)(\mathrm{Sc}\langle ge_T, ge_T\rangle) \\
&= (\mathrm{Sc}\langle f, f\rangle)(\mathrm{Sc}(\overline{e_T}\langle g, g\rangle e_T)) \\
&= (\mathrm{Sc}\langle f, f\rangle)(\mathrm{Sc}\langle g, g\rangle),
\end{aligned}$$

so $|\langle f, g\rangle| = (\sum_{T\in\mathcal{P}N}\langle f, g\rangle_T^2)^{1/2} \le 2^{\frac{m}{2}}\sqrt{\mathrm{Sc}\langle f, f\rangle}\sqrt{\mathrm{Sc}\langle g, g\rangle}$. □

Proposition 3.5 *Every right $\mathscr{A}_m$-module inner product space $\mathscr{H}$ is a right $\mathscr{A}_m$-module normed space with the induced norm $\|f\| := \sqrt{\mathrm{Sc}\langle f, f\rangle}$ for $f \in \mathscr{H}$.*

Proof For any $\lambda \in \mathscr{A}_m$ and $f, g \in \mathscr{H}$,

$$\begin{aligned}\|f\lambda\| &= \sqrt{\mathrm{Sc}\langle f\lambda, f\lambda\rangle}\\ &= \sqrt{\mathrm{Sc}(\overline{\lambda}\langle f, f\rangle\lambda)}\\ &= \sqrt{\mathrm{Sc}(\lambda\overline{\lambda}\langle f, f\rangle)}\\ &\le \sqrt{|\lambda\overline{\lambda}||\langle f, f\rangle|}\\ &\le \sqrt{2^{\frac{m}{2}}|\lambda|^2\cdot 2^{\frac{m}{2}}\|f\|^2}\\ &= 2^{\frac{m}{2}}|\lambda|\|f\|,\end{aligned}$$

and $\|f+g\|^2 = \mathrm{Sc}\langle f+g, f+g\rangle = \mathrm{Sc}(\langle f, f\rangle + 2\langle f, g\rangle + \langle g, g\rangle) \le (\|f\| + \|g\|)^2$. □

A complete right $\mathscr{A}_m$-module normed space is called a right $\mathscr{A}_m$-module Banach space, and a complete right $\mathscr{A}_m$-module inner product space is called a right $\mathscr{A}_m$-module Hilbert space. The case for the left $\mathscr{A}_m$-module can be similarly formulated.

Lemma 3.6 *If $\mathscr{H}$ is a right $\mathscr{A}_m$-module inner product space, then for any function $f \in \mathscr{H}$, $\{fc : c \in \mathscr{A}_m\}$ is a close subspace of $\mathscr{H}$.*

Proof Our goal is to show that if $\|fc_N - fc_M\| \to 0\,(N, M \to \infty)$, then there exists $c \in \mathscr{A}_m$ such that $\|fc_N - fc\| \to 0$ as $N \to \infty$. Because

$$\begin{aligned}\|fc_N - fc_M\|^2 &= \mathrm{Sc}\langle f(c_N - c_M), f(c_N - c_M)\rangle\\ &= \mathrm{Sc}(\overline{(c_N - c_M)}\langle f, f\rangle(c_N - c_M)) \ge 0,\end{aligned}$$

$\|fc_N - fc_M\|^2$ can be seen as a positive semidefinite quadratic form of $c_N - c_M$. Now, we treat $c_N - c_M$ as a column vector whose ith component coincides with the ith component of its algebraic form, and denote by A the real symmetric matrix associated with the quadratic form $\|fc_N - fc_M\|^2$, which is determined by $\langle f, f\rangle$. Then, we have

$$\|fc_N - fc_M\|^2 = (c_N - c_M)^\top A(c_N - c_M).$$

Let Γ be the orthogonal matrix, such that $\Gamma^\top A\Gamma$ is a diagonal matrix. Without loss of generality, we assume that $\Gamma^\top A\Gamma = \mathrm{diag}(\lambda_1, \ldots, \lambda_k, 0, \ldots, 0)$ with $\lambda_1, \ldots, \lambda_k$ being positive, and write $c_N - c_M = \Gamma(d_N - d_M)$, then

$$\begin{aligned}\|fc_N - fc_M\|^2 &= (d_N - d_M)^\top\Gamma^\top A\Gamma(d_N - d_M)\\ &= \lambda_1(d_{N_1} - d_{M_1})^2 + \cdots + \lambda_k(d_{N_k} - d_{M_k})^2 \to 0,\end{aligned}$$

which means $d_{N_1} - d_{M_1} \to 0, \dots, d_{N_k} - d_{M_k} \to 0$ as $N, M \to \infty$. By the completeness of the real numbers, there exists $d_1, \dots, d_k$, such that $d_{N_1} - d_1 \to 0, \dots, d_{N_k} - d_k \to 0$ as $N \to \infty$. Now, let $d = (d_1, \dots, d_k, 0, \dots, 0)^\top$, $c = \Gamma d$, then

$$\begin{aligned} \|fc_N - fc\|^2 &= (d_N - d)^\top \Gamma^\top A \Gamma (d_N - d) \\ &= \lambda_1 (d_{N_1} - d_1)^2 + \cdots + \lambda_k (d_{N_k} - d_k)^2 \to 0 \end{aligned}$$

as $N \to \infty$. □

Lemma 3.7 *If $\mathscr{H}$ is a right $\mathscr{A}_m$-module inner product space, then for any functions $\alpha, \beta \in \mathscr{H}$, the orthogonal projection of α onto the subspace spanned by β uniquely exists, denoted by $\mathscr{P}_{\overline{span}\{\beta\}}\alpha$.*

Proof The purpose is to show that there exists a unique βc such that

$$\|\alpha - \beta c\| = \inf_{c' \in \mathscr{A}_m} \|\alpha - \beta c'\|,$$

and such βc satisfies

$$\langle \alpha - \beta c, \beta \rangle = 0.$$

Let $d = \inf_{c' \in \mathscr{A}_m} \|\alpha - \beta c'\|$, then for any $N \in \mathbb{N}_+$ there exists $c_N \in \mathscr{A}_m$, such that $d \le \|\alpha - \beta c_N\| \le d + \frac{1}{N}$. By the parallelogram identity,

$$\begin{aligned} \|\beta c_N - \beta c_M\|^2 &= \|(\alpha - \beta c_N) - (\alpha - \beta c_M)\|^2 \\ &= 2(\|\alpha - \beta c_N\|^2 + \|\alpha - \beta c_M\|^2) - 4 \left\| \alpha - \beta \frac{c_N + c_M}{2} \right\|^2 \\ &\le 2\left(\left(d + \frac{1}{N} \right)^2 + \left(d + \frac{1}{M} \right)^2 \right) - 4d^2 \to 0 \end{aligned}$$

as $N, M \to \infty$. By Lemma 3.6, $\{\beta c_N\}_{N=1}^{\infty}$ has a limit βc, so by the continuity of the norm we get $\|\alpha - \beta c\| = d$. To prove the uniqueness, suppose there is another $\beta \widetilde{c}$ satisfying $\|\alpha - \beta \widetilde{c}\| = d$, then

$$\begin{aligned} \|\beta c - \beta \widetilde{c}\|^2 &= 2(\|\alpha - \beta c\|^2 + \|\alpha - \beta \widetilde{c}\|^2) - 4 \left\| \alpha - \beta \frac{c + \widetilde{c}}{2} \right\|^2 \\ &\le 4d^2 - 4d^2 = 0, \end{aligned}$$

which implies $\beta c = \beta \widetilde{c}$. Finally, we turn to show that $\langle \alpha - \beta c, \beta \rangle = 0$. For any $x \in \mathbb{R}$, we have

$$\begin{aligned} d^2 &\le \|\alpha - \beta c - \beta x\|^2 \\ &= \mathrm{Sc}\langle \alpha - \beta c - \beta x, \alpha - \beta c - \beta x \rangle \\ &= \|\alpha - \beta c\|^2 - 2x\mathrm{Sc}\langle \alpha - \beta c, \beta \rangle + x^2\|\beta\|^2 \\ &= d^2 - 2x\mathrm{Sc}\langle \alpha - \beta c, \beta \rangle + x^2\|\beta\|^2. \end{aligned}$$

Therefore

$$-2x\mathrm{Sc}\langle \alpha - \beta c, \beta \rangle + x^2\|\beta\|^2 \ge 0,$$

for all $x \in \mathbb{R}$, which implies

$$\mathrm{Sc}\langle \alpha - \beta c, \beta \rangle = 0.$$

After replacing βx by $\beta e_T x$ for each $T \in \mathcal{P}N$ and repeating the above discussions, we see that every component of $\langle \alpha - \beta c, \beta \rangle$ equals 0. Hence, $\langle \alpha - \beta c, \beta \rangle = 0$. □

Remark 3.8 In the above proof the orthogonal projection βc is unique, but $c \in \mathscr{A}_m$ may not be unique. This is different from the case of complex inner product space.

As a consequence of Lemma 3.7 we have

Theorem 3.9 *Let $\{\alpha_n\}_{n=1}^{\infty}$ be a sequence of functions in a right $\mathscr{A}_m$-module inner product space $\mathscr{H}$. Set*

$$\begin{aligned} \beta_1 &= \alpha_1, \\ \beta_2 &= \alpha_2 - \mathscr{P}_{\overline{span}\{\beta_1\}}\alpha_2, \\ &\vdots \\ \beta_n &= \alpha_n - \sum_{i=1}^{n-1} \mathscr{P}_{\overline{span}\{\beta_i\}}\alpha_n, \\ &\vdots \end{aligned}$$

then $\{\beta_n\}_{n=1}^{\infty}$ is an orthogonal system of functions in $\mathscr{H}$.

As an application, we now consider the inner spherical monogenics of order k ($k \in \mathbb{N}$) in Clifford analysis (playing an analogous role as the powers of the complex variable z), denoted by

$$\mathcal{M}_k = \{V_{l_1,\ldots,l_k} : (l_1, \ldots, l_k) \in \{1, \ldots, m\}^k\},$$

where by definition $V_0(x) = e_0$,

$$V_{l_1,\ldots,l_k} = \frac{1}{k!} \sum_{\pi(l_1,\ldots,l_k)} z_{l_1} \ldots z_{l_k},$$

in which the sum runs over all distinguishable permutations of $l_1, \ldots, l_k$, and the hyper-complex variables

$$z_l = x_l e_0 - x_0 e_l, \quad l = 1, \ldots, m.$$

For $f, g \in \bigcup_{k\in\mathbb{N}} \mathcal{M}_k$, the inner product is defined by

$$\langle f, g\rangle := \frac{1}{\omega_m} \int_{S^m} \overline{g} f dS,$$

with the induced norm

$$\|f\| := (\mathrm{Sc}\langle f, f\rangle)^{1/2} = \left(\frac{1}{\omega_m} \int_{S^m} |f|^2 dS\right)^{1/2},$$

where S^m is the unit sphere in $\mathbb{R}^{m+1}$ centered at the origin, dS is the surface area element on S^m.

Inner spherical monogenics of different orders are mutually orthogonal, but for a fixed order k, there are $\binom{m+k-1}{k}$ elements in $\mathcal{M}_k$ being not necessarily mutually orthogonal. Therefore, it is natural to ask for the construction of the orthonormal basis of $\mathcal{M}_k$. The existence of the orthonormal basis of $\mathcal{M}_k$ was proved in [4] by induction, but with no concrete expressions. By Theorem 3.9 we can now immediately give the explicit orthogonal formulas. More precisely, we have

Theorem 3.10 *Rearrange the elements in $\mathcal{M}_k$ by writing*

$$\mathcal{M}_k = \{V_1, V_2, \ldots, V_n\},$$

where $n = \binom{m+k-1}{k}$, then $\langle V_1, V_1\rangle$ is invertible. Let $U_1 = V_1$, then

$$\mathscr{P}_{\overline{span}\{U_1\}} V_2 = U_1 \langle U_1, U_1\rangle^{-1} \langle V_2, U_1\rangle.$$

Let

$$U_2 = V_2 - \mathscr{P}_{\overline{span}\{U_1\}} V_2,$$

then $\langle U_2, U_2\rangle$ is invertible and $\langle U_2, U_1\rangle = 0$. In general, let

$$U_j = V_j - \sum_{i=1}^{j-1} \mathscr{P}_{\overline{span}\{U_i\}} V_j = V_j - \sum_{i=1}^{j-1} U_i \langle U_i, U_i\rangle^{-1} \langle V_j, U_i\rangle, \text{ for } j \le n,$$

then $\langle U_j, U_j\rangle$ *is invertible for each* $j \le n$*, and* $\langle U_j, U_l\rangle = 0$ *for* $j \ne l$*. Therefore,* $\{U_1, \dots, U_n\}$ *consists an orthogonal basis of* $\mathcal{M}_k$*.*

Proof Consider the equation $\langle U_j, U_j\rangle c = 0$ in c, then

$$\overline{c}\langle U_j, U_j\rangle c = \langle U_j c, U_j c\rangle = 0,$$

so $\|U_j c\|^2 = \mathrm{Sc}\langle U_j c, U_j c\rangle = 0$, which gives $U_j c = 0$. Note that $U_j c$ is a linear combination of $V_1, \dots, V_j$ with the coefficient of V_j being c, by the uniqueness of the Taylor series ($V_1, \dots, V_n$ are basic functions constituting the Taylor series) we get $c = 0$. By Proposition 2.1, we conclude that $\langle U_j, U_j\rangle$ is invertible. The orthogonality $\langle U_i, U_j\rangle = 0$ for $i \ne j$ can be directly verified. □

4 Takenaka–Malmquist systems in higher dimensions

Denote by B^{m+1} the unit ball in $\mathbb{R}^{m+1}$ centered at the origin, $B^{m+1} = \{x \in \mathbb{R}^{m+1} : |x| < 1\}$, $S^m = \partial B^{m+1}$. The monogenic Hardy space $\mathcal{H}^2(B^{m+1})$ consists of all left monogenic functions f on B^{m+1} that satisfy

$$\|f\| := \sup_{0<r<1}\left(\frac{1}{\omega_m}\int_{\eta\in S^m} |f(r\eta)|^2 dS\right)^{1/2} < \infty.$$

For $f, g \in \mathcal{H}^2(B^{m+1})$, their Clifford number-valued inner product is defined by

$$\langle f, g\rangle := \frac{1}{\omega_m}\int_{\eta\in S^m} \overline{g(\eta)} f(\eta) dS,$$

where $f(\eta)$ and $g(\eta)$ ($\eta \in S^m$) are, respectively, the non-tangential boundary limit of f and g. We have

$$\|f\| = (\mathrm{Sc}\langle f, f\rangle)^{1/2} = \left(\frac{1}{\omega_m}\int_{\eta\in S^m} |f(\eta)|^2 dS\right)^{1/2}.$$

$\mathcal{H}^2(B^{m+1})$ is a right $\mathscr{A}_m$-module Hilbert space.

Let $a \in B^{m+1}$,

$$S_a(x) = \frac{\overline{1 - \overline{a}x}}{|1 - \overline{a}x|^{m+1}} \quad (x \in B^{m+1})$$

be the Szegö kernel for B^{m+1}. For any multi-index $k = (k_0, k_1, \dots, k_m) \in \mathbb{N}^{m+1}$ and any $f \in \mathcal{H}^2(B^{m+1})$, by Cauchy's integral formula and exchanging order of integration and differentiation ([19]), we have

$$\langle f, \partial_a^k S_a\rangle = (\partial_x^k f)(a), \tag{4.1}$$

where $\partial_x^k f = \frac{\partial^{|k|} f}{\partial x_0^{k_0} \partial x_1^{k_1} \cdots \partial x_m^{k_m}}$, $|k| = \sum_{i=0}^m k_i$.

Let $\{a_n\}_{n=1}^\infty$ be a sequence of Clifford numbers taking values in B^{m+1}. If a_n ($n \in \mathbb{N}_+$) are distinct from each other, then we have

Theorem 4.1 *The GS orthogonalization process*

$$\begin{cases} T_{a_1} := S_{a_1}, \\ T_{a_1,\ldots,a_n} := S_{a_n} - \sum_{i=1}^{n-1} T_{a_1,\ldots,a_i} \langle T_{a_1,\ldots,a_i}, T_{a_1,\ldots,a_i} \rangle^{-1} \langle S_{a_n}, T_{a_1,\ldots,a_i} \rangle, \; n \geq 2 \end{cases}$$

is realizable.

Proof To show that $\langle T_{a_1,\ldots,a_n}, T_{a_1,\ldots,a_n} \rangle$ is invertible, consider the equation:

$$\langle T_{a_1,\ldots,a_n}, T_{a_1,\ldots,a_n} \rangle c = 0.$$

By the same argument as that in the proof of Theorem 3.10, we have

$$T_{a_1,\ldots,a_n}(x)c = S_{a_n}(x)c + \sum_{i=1}^{n-1} S_{a_i}(x)c_i \equiv 0 \tag{4.2}$$

for some Clifford numbers $c_1, \ldots, c_{n-1} \in \mathscr{A}_m$ and $x \in B^{m+1}$. Since $S_{a_1}, \ldots S_{a_n}$ are of different poles outside the unit sphere, we can show that

$$c = c_1 = \cdots = c_{n-1} = 0.$$

To be specific, first by the uniqueness theorem of monogenic functions we can extend the identity (4.2) to $\mathbb{R}^{m+1} \setminus \{\frac{a_1}{|a_1|^2}, \ldots, \frac{a_n}{|a_n|^2}\}$. After multiplying (4.2) by

$$(1 - \overline{a_n} x)|1 - \overline{a_n} x|^{m-1}$$

from the left-hand side, we get

$$c + (1 - \overline{a_n} x)|1 - \overline{a_n} x|^{m-1} \sum_{i=1}^{n-1} S_{a_i}(x)c_i \equiv 0$$

for all $x \in \mathbb{R}^{m+1} \setminus \{\frac{a_1}{|a_1|^2}, \ldots, \frac{a_n}{|a_n|^2}\}$. Letting $x \to \frac{a_n}{|a_n|^2}$ we obtain $c = 0$, which implies that $\langle T_{a_1,\ldots,a_n}, T_{a_1,\ldots,a_n} \rangle^{-1}$ exists by Proposition 2.1. □

Remark 4.2 We have checked by calculations that $\langle T_{a_1,\ldots,a_n}, T_{a_1,\ldots,a_n} \rangle$ is a positive real number for $n \leq 5$. We conjecture that it holds for all $n \in \mathbb{N}_+$.

Hence, $\{B_n\} := \{B_{a_1,\ldots,a_n}\} := \{\frac{T_{a_1,\ldots,a_n}}{||T_{a_1,\ldots,a_n}||}\}_{n=1}^{\infty}$ becomes an orthonormal system for $\mathcal{H}^2(B^{m+1})$.

But if at least two of the parameters are the same, for example, a_2 equals a_1, then obviously $T_{a_1,a_2} = T_{a_1,a_1} = 0$. At this case we interpret B_2 as $\lim_{\rho\to 0^+} B_{a_1,b}$ (cf. [15]), where $b = a_1 + \rho\omega$, $\omega = \cos\theta_1 + \sin\theta_1\cos\theta_2 e_1 + \sin\theta_1\sin\theta_2\cos\theta_3 e_2 + \ldots + \sin\theta_1\sin\theta_2\cdots\sin\theta_m e_m$, and $\theta_1, \theta_2, \ldots, \theta_{m-1} \in [0,\pi]$, $\theta_m \in [0, 2\pi]$. More precisely,

$$
\begin{aligned}
B_2 &:= \lim_{\rho\to 0^+} B_{a_1,b} \\
&= \lim_{\rho\to 0^+} \frac{T_{a_1,b}}{\|T_{a_1,b}\|} \\
&= \lim_{\rho\to 0^+} \frac{T_{a_1,b} - T_{a_1,a_1}}{\|T_{a_1,b} - T_{a_1,a_1}\|} \\
&= \lim_{\rho\to 0^+} \frac{\frac{T_{a_1,b} - T_{a_1,a_1}}{\rho}}{\|\frac{T_{a_1,b} - T_{a_1,a_1}}{\rho}\|} \\
&= \frac{\nabla_\omega T_{a_1,y}|_{y=a_1}}{\|\nabla_\omega T_{a_1,y}|_{y=a_1}\|} \\
&= \frac{\nabla_\omega S_y|_{y=a_1} - T_{a_1}\langle T_{a_1}, T_{a_1}\rangle^{-1}\langle \nabla_\omega S_y|_{y=a_1}, T_{a_1}\rangle}{\|\nabla_\omega S_y|_{y=a_1} - T_{a_1}\langle T_{a_1}, T_{a_1}\rangle^{-1}\langle \nabla_\omega S_y|_{y=a_1}, T_{a_1}\rangle\|},
\end{aligned}
$$

where $\nabla_\omega S_y = \frac{\partial S_y}{\partial y_0}\cos\theta_1 + \frac{\partial S_y}{\partial y_1}\sin\theta_1\cos\theta_2 + \frac{\partial S_y}{\partial y_2}\sin\theta_1\sin\theta_2\cos\theta_3 + \cdots + \frac{\partial S_y}{\partial y_m}\sin\theta_1\sin\theta_2\cdots\sin\theta_m$ is the directional derivative of S_y with respect to y. In other words, when $a_2 = a_1$, B_2 is interpreted as the orthonormalization of T_{a_1} and $\nabla_\omega S_y|_{y=a_1}$ (see also [11]).

We further note that as a function of y, S_y satisfies $S_y\overline{D} = 0$, which implies that $\frac{\partial S_y}{\partial y_0}, \frac{\partial S_y}{\partial y_1}, \ldots, \frac{\partial S_y}{\partial y_m}$ are linear dependent in $\mathcal{H}^2(B^{m+1})$. Hence, if the multiplicity of the parameter a_n (we call the cardinal number of the set $\{j : a_j = a_n, j \le n\}$ the multiplicity of a_n and denote it by $m(a_n)$) is greater than $m+1$, then the second-order partial derivatives of S_y at the point a_n should be involved in the orthogonalization process. In general, when $m(a_n) > \sum_{i=0}^{k-1}\binom{i+m-1}{m-1} = \binom{k+m-1}{m}$, then the kth-order partial derivatives of S_y at the point a_n must appear.

Observe that in complex analysis the TM systems for the unit disc and upper half space can be generated by Szegö or higher order Szegö kernels through GS orthogonalization process ([20]). Heuristically, we propose the following definition.

Definition 4.3 We call $\{B_n\}_{n=1}^{\infty}$ the Takenaka–Malmquist system for B^{m+1}. If the kth parameter $a_k = 0$, then B_k is called a Blaschke product of order $k-1$ for B^{m+1}.

By the orthogonality of $\{B_n\}_{n=1}^{\infty}$ and the reproducing property of the Szegö kernel we easily get the following property similar to the complex TM systems.

Proposition 4.4 *For any $B_{a_1,\ldots,a_n}$ in the TM system, a_i ($i \le n-1$) is a zero point of $B_{a_1,\ldots,a_n}$ with multiplicity $m(a_i)$.*

For the cases of half space and general domains (provided that the Szegö kernels exist) we have similar results. Let $\mathbb{R}^{m+1}_+ := \{x \in \mathbb{R}^{m+1} : \mathrm{Sc}\, x > 0\}$ be the half space in $\mathbb{R}^{m+1}$, the Szegö kernel we use for $\mathbb{R}^{m+1}_+$ is

$$S_a(x) = \frac{\overline{x + \overline{a}}}{|x + \overline{a}|^{m+1}}, \quad x, a \in \mathbb{R}^{m+1}_+.$$

5 Adaptive Clifford TM system approximation

Let us first have a brief review of adaptive TM system approximation. Consider the complex Hardy space $\mathcal{H}^2(\mathbf{D})$, where $\mathbf{D}$ denotes the unit disc in the complex plane. For $f \in \mathcal{H}^2(\mathbf{D})$, in adaptive TM system approximation f is associated with the expansion:

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

where $\{B_k(z)\}_{k=1}^{\infty}$ is the Takenaka–Malmquist (TM) system on $\mathbf{D}$ determined by a sequence $\{a_k\}_{k=1}^{\infty}$ in $\mathbf{D}$ specially selected according to the Maximal Selection Principle (see below) of the context,

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

f_k is the kth *standard remainder*, defined by

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

and g_l is the lth *reduced remainder*, defined by

$$g_l(z) = f_l(z) \prod_{j=1}^{l-1} \frac{1 - \overline{a_j} z}{z - a_j},$$

and

$$e_{a_l}(z) = \frac{\sqrt{1 - |a_l|^2}}{1 - \overline{a_l} z}$$

being the normalized Szegö kernel of $\mathbf{D}$, that plays the role as reproducing kernel of the Hilbert space $\mathcal{H}^2(\mathbf{D})$. When $a_1, a_2, \ldots$ are mutually different, $B_1, B_2, \ldots$ are consecutively GS orthonormalizations of $e_{a_1}, e_{a_2}, \ldots$; and if $a_1, a_2, \ldots$ have multiples, in the GS process $e_{a_1}, e_{a_2}, \ldots$ are replaced by the so called higher order Szegö kernels involving derivatives of the Szegö kernels (see below).

The approximation is said to be adaptive, because the parameters in TM system are adaptively chosen to best match the kth reduced reminder at each step according to the Maximum Selection Principle:

$$a_k = \arg\max_{a\in D} |\langle g_k, e_a\rangle|^2 = \arg\max_{a\in D}(1-|a|^2)|g_k(a)|^2.$$

Note that if all the parameters are zero, then the TM system reduces to the half Fourier system. If $a_k = 0$, then B_k becomes a Blaschke product. If the first parameter a_1 is chosen to be zero, then we get an adaptive mono-components decomposition, i.e., every B_k is a mono-component that possesses a non-negative analytic instantaneous frequency function. The case for the upper half plane is similar, and was discussed in [9].

The advantages of adaptive TM system approximation compared with the usual greedy algorithms ([18]) include that at each step the former attains the optimal energy pursuit and at the same time contributes a term possessing positive analytic frequency. Adaptive TM system approximation has been extended to the general type Hilbert spaces with a dictionary satisfying the boundary vanishing condition ([10]).

Generalization of adaptive TM system approximation into higher dimensions now has two routes. One is based on several complex variables ([1, 10]), the other is the quaternionic ([15]) and Clifford analysis direction. In the context of several complex variables, in [1] the Drury–Arveson space of functions analytic in the unit ball of $\mathbb{C}^N$ was discussed. In [10] two different approaches were discussed, of which one used product-TM systems in the context of the n-torus T^n, and the other used the pre-orthogonal adaptive TM system method on product-Szegö dictionaries. The several complex variables contexts that have been achieved were inspired by the fact that complex TM systems can be generated by the Szegö and higher order Szegö kernels through GS orthogonalization process. The case for matrix-valued functions was studied in [2].

Since the Euclidean space $\mathbb{R}^n$ can be naturally embedded into quaternions or a Clifford algebra, it is natural to perform quaternionic or Clifford GS orthogonalization process in constructing an analogous adaptive approximation theory. Due to unavailability of Clifford GS orthogonalization process in the earlier times the previous studies in this direction were mostly execution of greedy algorithm ([16, 19]). The significance of an orthogonalization process in general Clifford module Hilbert spaces, and therefore, a TM systems theory in the Clifford algebra setting is that they give rise to rational and orthogonal approximations in monogenic functional spaces.

As in the one complex variable case the adaptive Clifford TM system approximation proposed below gives rise to sparse, therefore, fast approximations to monogenic functions.

Let $f \in \mathcal{H}^2(B^{m+1})$. We associate f with the Fourier-type series:

$$f(x) \sim \sum_{n=1}^{\infty} B_n(x)c_n,$$

where the coefficients c_n's are given by

$$c_1 = \langle B_1, B_1\rangle^{-1}\langle f, B_1\rangle = (1-|a_1|^2)^{\frac{m}{2}} f(a_1),$$

and for $n \geq 2$,

$$\begin{aligned} c_n =& \langle B_n, B_n\rangle^{-1}\langle f, B_n\rangle \\ =& \langle B_n, B_n\rangle^{-1}\left\langle f, \frac{S_{a_n} - \sum_{i=1}^{n-1} T_{a_1,\ldots,a_i}\langle T_{a_1,\ldots,a_i}, T_{a_1,\ldots,a_i}\rangle^{-1}\langle S_{a_n}, T_{a_1,\ldots,a_i}\rangle}{\|T_{a_1,\ldots,a_n}\|}\right\rangle \\ =& \langle B_n, B_n\rangle^{-1}\frac{\left\langle f - \sum_{i=1}^{n-1} T_{a_1,\ldots,a_i}\langle T_{a_1,\ldots,a_i}, T_{a_1,\ldots,a_i}\rangle^{-1}\langle f, T_{a_1,\ldots,a_i}\rangle, S_{a_n}\right\rangle}{\|T_{a_1,\ldots,a_n}\|}. \end{aligned}$$

Let

$$\begin{aligned} f_n(x) &= f(x) - \sum_{i=1}^{n-1} T_{a_1,\ldots,a_i}(x)\langle T_{a_1,\ldots,a_i}, T_{a_1,\ldots,a_i}\rangle^{-1}\langle f, T_{a_1,\ldots,a_i}\rangle \\ &= f(x) - \sum_{i=1}^{n-1} B_i(x)\langle B_i, B_i\rangle^{-1}\langle f, B_i\rangle. \end{aligned}$$

If $m(a_n) = 1$, then

$$c_n = \frac{\langle B_n, B_n\rangle^{-1}}{\|T_{a_1,\ldots,a_n}\|} f_n(a_n) = \|T_{a_1,\ldots,a_n}\|\langle T_{a_1,\ldots,a_n}, T_{a_1,\ldots,a_n}\rangle^{-1} f_n(a_n), \tag{5.1}$$

$$\begin{aligned} \|B_n c_n\|^2 &= \mathrm{Sc}\langle B_n c_n, B_n c_n\rangle \\ &= \mathrm{Sc}\left\langle B_n \frac{\langle B_n, B_n\rangle^{-1}}{\|T_{a_1,\ldots,a_n}\|} f_n(a_n), B_n \frac{\langle B_n, B_n\rangle^{-1}}{\|T_{a_1,\ldots,a_n}\|} f_n(a_n)\right\rangle \\ &= \mathrm{Sc}(\overline{f_n(a_n)}\frac{\langle B_n, B_n\rangle^{-1}}{\|T_{a_1,\ldots,a_n}\|^2} f_n(a_n)) \\ &= \mathrm{Sc}(\overline{f_n(a_n)}\langle T_{a_1,\ldots,a_n}, T_{a_1,\ldots,a_n}\rangle^{-1} f_n(a_n)) \\ &= \mathrm{Sc}((1-|a_n|^2)^m\overline{f_n(a_n)}((1-|a_n|^2)^m\langle T_{a_1,\ldots,a_n}, T_{a_1,\ldots,a_n}\rangle)^{-1} f_n(a_n)), \end{aligned} \tag{5.2}$$

and

$$
\begin{aligned}
&(1-|a_n|^2)^m \langle T_{a_1,\ldots,a_n}, T_{a_1,\ldots,a_n}\rangle \\
&\quad = (1-|a_n|^2)^m \left(\langle S_{a_n}, S_{a_n}\rangle - \sum_{i=1}^{n-1} \langle T_{a_1,\ldots,a_i}, S_{a_n}\rangle \langle T_{a_1,\ldots,a_i}, T_{a_1,\ldots,a_i}\rangle^{-1} \overline{\langle T_{a_1,\ldots,a_i}, S_{a_n}\rangle} \right) \\
&\quad = 1-(1-|a_n|^2)^m \sum_{i=1}^{n-1} T_{a_1,\ldots,a_i}(a_n) \langle T_{a_1,\ldots,a_i}, T_{a_1,\ldots,a_i}\rangle^{-1} \overline{T_{a_1,\ldots,a_i}(a_n)}. \qquad (5.3)
\end{aligned}
$$

If $m(a_n) > 1$, c_n and $\|B_n c_n\|^2$ are taken in the limit sense as before.

Lemma 5.1 *Let $a_1, \ldots, a_{n-1} \in B^{m+1}$ be fixed, $a = |a|\xi = r\xi$, then*

$$
\lim_{r\to 1^-} \|B_{a_1,\ldots,a_{n-1},a} \langle B_{a_1,\ldots,a_{n-1},a}, B_{a_1,\ldots,a_{n-1},a}\rangle^{-1} \langle f, B_{a_1,\ldots,a_{n-1},a}\rangle\|^2 = 0
$$

holds uniformly in $|\xi| = 1$.

Proof Note that when $r \to 1^-$, a must be different from a_i $(i \le n-1)$, then (5.3) clearly shows that

$$
\lim_{r\to 1^-} (1-|a|^2)^m \langle T_{a_1,\ldots,a_{n-1},a}, T_{a_1,\ldots,a_{n-1},a}\rangle = 1.
$$

On the other hand, according to Lemma 3.2 in [16] we have

$$
\lim_{r\to 1^-} (1-|a|^2)^{\frac{m}{2}} f_n(a) = 0
$$

uniformly in $|\xi| = 1$. Therefore, from (5.2), we immediately get the desired result. □

Lemma 5.1 implies

Theorem 5.2 (Maximum Selection Principle) *For any $f \in \mathcal{H}^2(B^{m+1})$ and any fixed $a_1, \ldots, a_{n-1} \in B^{m+1}$, there exist an $a_n \in B^{m+1}$, such that*

$$
\begin{aligned}
&\|B_{a_1,\ldots,a_{n-1},a_n} \langle B_{a_1,\ldots,a_{n-1},a_n}, B_{a_1,\ldots,a_{n-1},a_n}\rangle^{-1} \langle f, B_{a_1,\ldots,a_{n-1},a_n}\rangle\| \\
&\quad = \sup_{a\in B^{m+1}} \|B_{a_1,\ldots,a_{n-1},a} \langle B_{a_1,\ldots,a_{n-1},a}, B_{a_1,\ldots,a_{n-1},a}\rangle^{-1} \langle f, B_{a_1,\ldots,a_{n-1},a}\rangle\|. \qquad (5.4)
\end{aligned}
$$

The maximum selection principle enables us to obtain the best approximation to f step by step, by choosing a suitable parameter a_n at the nth step, such that the energy of the nth term $B_n\langle B_n, B_n\rangle^{-1}\langle f, B_n\rangle$ attains its maximum, or equivalently, making the energy of the residue f_n attain its minimum, so that the adaptive Fourier series associated with f converges in a fast way. Note that the choice of a_n in (5.4) may not be unique.

We now proceed to prove the convergence of the adaptive Fourier series. First we show a technical lemma.

Lemma 5.3 *For any $a_1, \ldots, a_n \in B^{m+1}$, we have*

$$\begin{aligned}
&\|B_{a_1,\ldots,a_n}\langle B_{a_1,\ldots,a_n}, B_{a_1,\ldots,a_n}\rangle^{-1}\langle f, B_{a_1,\ldots,a_n}\rangle\| \\
&\quad\geq \|B_{a_n}\langle B_{a_n}, B_{a_n}\rangle^{-1}\langle f_n, B_{a_n}\rangle\| \\
&\quad= |\langle f_n, B_{a_n}\rangle| \\
&\quad= (1-|a_n|^2)^{\frac{m}{2}}|f_n(a_n)|.
\end{aligned}$$

Proof Since f_n and B_n are both orthogonal to $B_1, B_2, \ldots, B_{n-1}$, we have

$$\begin{aligned}
&\|B_{a_1,\ldots,a_n}\langle B_{a_1,\ldots,a_n}, B_{a_1,\ldots,a_n}\rangle^{-1}\langle f, B_{a_1,\ldots,a_n}\rangle\|^2 \\
&\quad= \|B_{a_1,\ldots,a_n}\langle B_{a_1,\ldots,a_n}, B_{a_1,\ldots,a_n}\rangle^{-1}\langle f_n, B_{a_1,\ldots,a_n}\rangle\|^2 \\
&\quad= \|B_{a_1}\langle B_{a_1}, B_{a_1}\rangle^{-1}\langle f_n, B_{a_1}\rangle\|^2 + \|B_{a_1,a_2}\langle B_{a_1,a_2}, B_{a_1,a_2}\rangle^{-1}\langle f_n, B_{a_1,a_2}\rangle\|^2 \\
&\qquad+\cdots+\|B_{a_1,\ldots,a_n}\langle B_{a_1,\ldots,a_n}, B_{a_1,\ldots,a_n}\rangle^{-1}\langle f_n, B_{a_1,\ldots,a_n}\rangle\|^2.
\end{aligned} \tag{5.5}$$

Note that for any $f \in \mathcal{H}^2(B^{m+1})$, the orthogonal projection of f onto the space spanned by $B_1, B_2, \ldots, B_n$ is uniquely determined by $a_1, \ldots, a_n$, regardless of their orders. Therefore

$$\begin{aligned}
&B_{a_1}\langle B_{a_1}, B_{a_1}\rangle^{-1}\langle f_n, B_{a_1}\rangle + B_{a_1,a_2}\langle B_{a_1,a_2}, B_{a_1,a_2}\rangle^{-1}\langle f_n, B_{a_1,a_2}\rangle \\
&\qquad+\cdots+B_{a_1,\ldots,a_n}\langle B_{a_1,\ldots,a_n}, B_{a_1,\ldots,a_n}\rangle^{-1}\langle f_n, B_{a_1,\ldots,a_n}\rangle \\
&\quad= B_{a_n}\langle B_{a_n}, B_{a_n}\rangle^{-1}\langle f_n, B_{a_n}\rangle + B_{a_n,a_1}\langle B_{a_n,a_1}, B_{a_n,a_1}\rangle^{-1}\langle f_n, B_{a_n,a_1}\rangle \\
&\qquad+\cdots+B_{a_n,a_1,\ldots,a_{n-1}}\langle B_{a_n,a_1,\ldots,a_{n-1}}, B_{a_n,a_1,\ldots,a_{n-1}}\rangle^{-1}\langle f_n, B_{a_n,a_1,\ldots,a_{n-1}}\rangle,
\end{aligned}$$

and (5.5) equals

$$\begin{aligned}
&\|B_{a_n}\langle B_{a_n}, B_{a_n}\rangle^{-1}\langle f_n, B_{a_n}\rangle\|^2 + \|B_{a_n,a_1}\langle B_{a_n,a_1}, B_{a_n,a_1}\rangle^{-1}\langle f_n, B_{a_n,a_1}\rangle\|^2 \\
&\qquad+\cdots+\|B_{a_n,a_1,\ldots,a_{n-1}}\langle B_{a_n,a_1,\ldots,a_{n-1}}, B_{a_n,a_1,\ldots,a_{n-1}}\rangle^{-1}\langle f_n, B_{a_n,a_1,\ldots,a_{n-1}}\rangle\|^2 \\
&\quad\geq \|B_{a_n}\langle B_{a_n}, B_{a_n}\rangle^{-1}\langle f_n, B_{a_n}\rangle\|^2.
\end{aligned}$$

□

Theorem 5.4 *Subject to the maximum selection principle (5.4) we have*

$$\left\|\sum_{n=1}^{N} B_n\langle B_n, B_n\rangle^{-1}\langle f, B_n\rangle - f\right\| \to 0 \quad (N\to\infty). \tag{5.6}$$

Proof From Bessel's inequality, we have

$$\sum_{n=1}^{\infty}\left\|B_n\langle B_n, B_n\rangle^{-1}\langle f, B_n\rangle\right\|^2 \leq \|f\|^2,$$

which implies that there exists a function $g \in \mathcal{H}^2(B^{m+1})$, such that

$$\sum_{n=1}^{\infty} B_n \langle B_n, B_n\rangle^{-1} \langle f, B_n\rangle = g$$

holds in the sense of $\mathcal{H}^2(B^{m+1})$. If (5.6) is not true, then

$$h := f - g \neq 0,$$

so there exists a point $a \in B^{m+1} \setminus \bigcup_{i=1}^{\infty}\{a_i\}$, such that

$$\|B_a \langle B_a, B_a\rangle^{-1} \langle h, B_a\rangle\| = |\langle h, B_a\rangle| = (1 - |a|^2)^{\frac{m}{2}} |h(a)| = \delta > 0.$$

Let

$$f_N = f - \sum_{n=1}^{N-1} B_n \langle B_n, B_n\rangle^{-1} \langle f, B_n\rangle, \quad r_N = -\sum_{n=N}^{\infty} B_n \langle B_n, B_n\rangle^{-1} \langle f, B_n\rangle.$$

When N is large enough,

$$\begin{aligned} |\langle r_N, B_a\rangle| &= \|B_a \langle B_a, B_a\rangle^{-1} \langle r_N, B_a\rangle\| \\ &\leq \|r_N\| = \left(\sum_{n=N}^{\infty} \|B_n \langle B_n, B_n\rangle^{-1} \langle f, B_n\rangle\|^2\right)^{1/2} < \delta/2. \end{aligned} \tag{5.7}$$

Therefore

$$|\langle f_N, B_a\rangle| = |\langle h - r_N, B_a\rangle| \geq |\langle h, B_a\rangle| - |\langle r_N, B_a\rangle| > \delta/2.$$

By Lemma 5.3 we get

$$\begin{aligned} &\|B_{a_1,\ldots,a_{N-1},a} \langle B_{a_1,\ldots,a_{N-1},a}, B_{a_1,\ldots,a_{N-1},a}\rangle^{-1} \langle f, B_{a_1,\ldots,a_{N-1},a}\rangle\| \\ &\quad \geq \|B_a \langle B_a, B_a\rangle^{-1} \langle f_N, B_a\rangle\| = |\langle f_N, B_a\rangle| > \delta/2. \end{aligned}$$

On the other hand, from (5.7) we know that

$$\begin{aligned} &\|B_{a_1,\ldots,a_{N-1},a_N} \langle B_{a_1,\ldots,a_{N-1},a_N}, B_{a_1,\ldots,a_{N-1},a_N}\rangle^{-1} \langle f, B_{a_1,\ldots,a_{N-1},a_N}\rangle\| \\ &\quad = \|B_N \langle B_N, B_N\rangle^{-1} \langle f, B_N\rangle\| < \delta/2. \end{aligned}$$

Therefore, we arrive at

$$\begin{aligned} &\|B_{a_1,\ldots,a_{N-1},a_N} \langle B_{a_1,\ldots,a_{N-1},a_N}, B_{a_1,\ldots,a_{N-1},a_N}\rangle^{-1} \langle f, B_{a_1,\ldots,a_{N-1},a_N}\rangle\| \\ &\quad < \|B_{a_1,\ldots,a_{N-1},a} \langle B_{a_1,\ldots,a_{N-1},a}, B_{a_1,\ldots,a_{N-1},a}\rangle^{-1} \langle f, B_{a_1,\ldots,a_{N-1},a}\rangle\|, \end{aligned}$$

which contradicts with the maximum selection principle that we should not have chosen a_N at the Nth step. □

Next we consider a convergence rate for adaptive Clifford TM system approximation. To deal with this, as in [8] we introduce a subclass of $\mathcal{H}^2(B^{m+1})$:

$$\mathcal{H}^2(B^{m+1}, M) := \left\{ f \in \mathcal{H}^2(B^{m+1}) : f = \sum_{k=1}^{\infty} B_{b_k} c_k \text{ with } \sum_{k=1}^{\infty} |c_k| \leq M < \infty \right\}.$$

We also need the following lemma.

Lemma 5.5 ([8]) *Let $\{d_n\}_{n=l}^{\infty}$ be a sequence of non-negative numbers satisfying the inequalities*

$$d_1 \leq A, \quad d_{n+1} \leq d_n(1 - d_n/A), \quad n = 1, 2, \ldots.$$

Then, we have for each n

$$d_n \leq A/n.$$

Now, we can prove a convergence rate result.

Theorem 5.6 *If $f \in \mathcal{H}^2(B^{m+1}, M)$, then*

$$\|f_N\| \leq \frac{M}{\sqrt{N}},$$

where f_N is the residue produced from the adaptive TM system approximation of f at the Nth step.

Proof Since $\mathcal{H}^2(B^{m+1})$ is a right $\mathscr{A}_m$-module normed space, we have

$$\|f_1\| = \|f\| \leq \sum_{k=1}^{\infty} \|B_{b_k} c_k\| = \sum_{k=1}^{\infty} (\mathrm{Sc}(\overline{c_k} c_k))^{1/2} = \sum_{k=1}^{\infty} |c_k| \leq M,$$

and

$$\begin{aligned} \|f_N\|^2 &= \mathrm{Sc}\langle f_N, f_N \rangle \\ &= \mathrm{Sc}\langle f_N, f \rangle \\ &= \mathrm{Sc}\left\langle f_N, \sum_{k=1}^{\infty} B_{b_k} c_k \right\rangle \end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=1}^{\infty} \mathrm{Sc}\langle f_N, B_{b_k} c_k \rangle \\
&= \sum_{k=1}^{\infty} \mathrm{Sc}(\overline{c_k}\langle f_N, B_{b_k}\rangle) \\
&\leq \sum_{k=1}^{\infty} |c_k||\langle f_N, B_{b_k}\rangle| \\
&\leq M \sup_{k\geq 1} |\langle f_N, B_{b_k}\rangle| \\
&\leq M \sup_{a\in B^{m+1}} |\langle f_N, B_a\rangle|. \qquad (5.8)
\end{aligned}
$$

By Lemma 5.3 we get

$$
\begin{aligned}
&\|B_N\langle B_N, B_N\rangle^{-1}\langle f, B_N\rangle\| \\
&\quad = \sup_{a\in B^{m+1}} \|B_{a_1,\ldots,a_{N-1},a}\langle B_{a_1,\ldots,a_{N-1},a}, B_{a_1,\ldots,a_{N-1},a}\rangle^{-1}\langle f, B_{a_1,\ldots,a_{N-1},a}\rangle\| \\
&\quad \geq \sup_{a\in B^{m+1}} |\langle f_N, B_a\rangle|. \qquad (5.9)
\end{aligned}
$$

So, from (5.8) and (5.9) we obtain

$$
\begin{aligned}
\|f_{N+1}\|^2 &= \|f_N - B_N\langle B_N, B_N\rangle^{-1}\langle f, B_N\rangle\|^2 \\
&= \|f_N\|^2 - \|B_N\langle B_N, B_N\rangle^{-1}\langle f, B_N\rangle\|^2 \\
&\leq \|f_N\|^2\left(1 - \frac{\|f_N\|^2}{M^2}\right).
\end{aligned}
$$

By Lemma 5.5, we conclude the proof. □

Remark 5.7 Let $f \in L^2(S^m)$ (square integrable on S^m), where f in not necessarily monogenic. To get the adaptive approximation of f, without loss of generality we assume that f is real-valued, and take

$$
F(x) := T(f)(x) := \int_{\omega\in S^m} S(x,\omega) f(\omega) dS, \quad |x| < 1,
$$

where

$$
S(x,\omega) = P(x,\omega) + Q(x,\omega)
$$

is the monogenic Schwarz kernel,

$$
P(x,\omega) = \frac{1}{\omega_m}\frac{1-|x|^2}{|x-\omega|^{m+1}}
$$

is the Poisson kernel and

$$Q(x,\omega) = \mathrm{NSc}\left(\int_0^1 t^{m-1}(\overline{D}P)(tx,\omega)x\,dt\right)$$
$$= \left(\frac{1}{\omega_m}\int_0^1 \frac{(m+1)t^{m-1}(1-t^2|x|^2)}{|tx-\omega|^{m+3}}dt\right)\mathrm{NSc}(\overline{\omega}x)$$

is the Cauchy-type harmonic conjugate of $P(x,\omega)$ on the unit sphere S^m, which can be computed out explicitly with an expression in elementary functions. As a consequence of boundedness of Hilbert transform on the sphere ([13]), T is a bounded operator from $L^2(S^m)$ to $\mathcal{H}^2(B^{m+1})$. Therefore, $F \in \mathcal{H}^2(B^{m+1})$. The adaptive approximation of f can be obtained by the adaptive TM system approximation of F through the relation:

$$\lim_{r\to 1^-} \mathrm{Sc}(F(r\xi)) = f(\xi)$$

for a.e. $\xi \in S^m$.

Remark 5.8 The above theory can be similarly formulated in the context of the half space $\mathbb{R}_+^{m+1}$. While for a real-valued function $f \in L^2(\mathbb{R}^m)$ we consider the Cauchy integral of f:

$$F(x) = C(f) := \frac{-1}{\omega_m}\int_{\mathbb{R}^m} \frac{\overline{\underline{y}-x}}{|\underline{y}-x|^{m+1}} f(\underline{y})d\underline{y}, \quad x \in \mathbb{R}_+^{m+1},$$

where $\underline{y} = y_1e_1 + \cdots + y_me_m$, $d\underline{y} = dy_1 \ldots dy_m$. We have $F \in \mathcal{H}^2(\mathbb{R}_+^{m+1})$, and by Sokhotsky–Plemelj formula, we get

$$\lim_{x_0\to 0^+} F(x_0+\underline{x}) = \frac{1}{2}f(\underline{x}) + \frac{1}{2}H(f)(\underline{x}),$$

where $H(f) = \sum_{i=1}^m e_iR_i(f)$, and

$$R_i(f)(\underline{x}) := \frac{2}{\omega_m}\ \mathrm{p.v.}\int_{\mathbb{R}^m} \frac{y_i - x_i}{|\underline{y}-\underline{x}|^{m+1}} f(\underline{y})d\underline{y}$$

is the ith ($1 \le i \le m$) Riesz transform of f. The adaptive approximation of f is then obtained by the adaptive TM system approximation of F through

$$2\lim_{x_0\to 0^+} \mathrm{Sc}\,(F(x_0+\underline{x})) = f(\underline{x})$$

for a.e. $\underline{x} \in \mathbb{R}^m$.

Acknowledgements This work was supported by the National Natural Science Foundation of China (No. 11701105), Research Grant of Macau University of Science and Technology (No. FRG-22-075-MCMS) and Macao Government Research Funding (No. FDCT0128/2022/A). We thank the reviewers for their careful reading of the manuscript and valuable comments and suggestions.

Data availability Not applicable.

References

1. Alpay, D., Colombo, F., Qian, T., Sabadini, I.: Adaptative decomposition: the case of the Drury-Arveson space. J. Fourier Anal. Appl. **23**, 1426–1444 (2017)
2. Alpay, D., Colombo, F., Qian, T., Sabadini, I.: Adaptive orthonormal systems for matrix-valued functions. Proc. Am. Math. Soc. **145**, 2089–2106 (2017)
3. Bock, S., Gürlebeck, K., Lávička, R., Souček, V.: Gelfand-Tsetlin bases for spherical monogenics in dimension 3. Rev. Mat. Iberoam. **28**(4), 1165–1192 (2012)
4. Brackx, F., Delanghe, R., Sommen, F.: Clifford Analysis. Research Notes in Mathematics, vol. 76. Pitman Books Limited, USA (1982)
5. Brackx, F., De Schepper, H., Lávička, R., Souček, V.: Gelfand-Tsetlin bases of orthogonal polynomials in Hermitean Clifford analysis. Math. Methods Appl. Sci. **34**(17), 2167–2180 (2011)
6. Delanghe, R., Sommen, F., Souček, V.: Clifford Algebra and Spinor-Valued Functions: a Function Theory for the Dirac Operator. Mathematics and Its Applications, vol. 53. Kluwer Academic, The Netherlands (1992)
7. Delanghe, R., Lávička, R., Souček, V.: The Gelfand-Tsetlin bases for Hodge-de Rham systems in Euclidean spaces. Math. Methods Appl. Sci. **35**(7), 745–757 (2012)
8. DeVore, R., Temlyakov, V.: Some remarks on greedy algorithm. Adv. Comput. Math. **5**, 173–187 (1996)
9. Qian, T.: Intrinsic mono-component decomposition of functions: an advance of Fourier theory. Math. Methods Appl. Sci. **33**, 880–891 (2010)
10. Qian, T.: Two-dimensional adaptive Fourier decomposition. Math. Methods Appl. Sci. **39**, 2431–2448 (2016)
11. Qian, T.: Positive-instantaneous frequency and approximation. Front. Math. China **17**, 337–371 (2022)
12. Qian, T., Wang, Y.: Adaptive Fourier series-a variation of greedy algorithm. Adv. Comput. Math. **34**, 279–293 (2011)
13. Qian, T., Yang, Y.: Hilbert transforms on the sphere with the Clifford algebra setting. J. Fourier Anal. Appl. **15**, 753–774 (2009)
14. Qian, T., Tan, L., Wang, Y.: Adaptive decomposition by weighted inner functions: a generalization of Fourier series. J. Fourier Anal. Appl. **17**, 175–190 (2011)
15. Qian, T., Sprößig, W., Wang, J.: Adaptive Fourier decomposition of functions in quaternionic Hardy spaces. Math. Methods Appl. Sci. **35**, 43–64 (2012)
16. Qian, T., Wang, J., Yang, Y.: Matching pursuits among shifted Cauchy kernels in higher-dimensional spaces. Acta Math. Scientia **34**, 660–672 (2014)
17. Qian, T., Wang, X., Zhang, L.: MIMO frequency domain system identification using matrix-valued orthonormal functions. Automatica **133**, 109882 (2021)
18. Temlyakov, V.: Greedy Approximation. Cambridge Monographs on Applied and Computational Mathematics, vol. 20. Cambridge University Press, Cambridge (2011)
19. Wang, J., Qian, T.: Approximation of monogenic functions by higher order Szegö kernels on the unit ball and half space. Sci. China Math. **57**, 1785–1797 (2014)
20. Wang, J., Qian, T.: Approximation of functions by higher order Szegö kernels I. Complex variable cases. Complex Var. Elliptic Equ. **60**, 733–747 (2015)