Boundary Orthogonal Wavelets for Sparse Functional Data Analysis

Quanhan Li[1], Ming Li[2,3], Xiaosheng Zhuang[1]
[1]*Department of Mathematics*, *City University of Hong Kong*, Hong Kong, China
[2]*Zhejiang Key Laboratory of Intelligent Education Technology and Application*, *Zhejiang Normal University*, Jinhua, China
[3] *Zhejiang Institute of Optoelectronics*, Jinhua, China
quanhanli2-c@my.cityu.edu.hk, mingli@zjnu.edu.cn, xzhuang7@cityu.edu.hk

***Abstract*—This paper presents the construction of boundary orthogonal wavelet (BoW) systems derived from the Daubechies-N family of orthogonal wavelets. The proposed BoW construction yields a minimal number of boundary wavelets while retaining as many interior scaling functions as possible. Such boundary wavelet systems offer an effective remedy for the limitations of conventional wavelet methods, which are constrained by boundary effects. We apply the constructed BoW system to the analysis of sparse functional data and demonstrate its advantages over traditional techniques when the underlying functions exhibit local abrupt changes, such as jump discontinuities.**

***Index Terms*—Wavelets on the interval, Boundary orthogonal wavelets, Sparse functional data analysis, Soft thresholding**

I. Introduction

Functional data analysis (FDA) is a fundamental branch of statistics concerned with the processing and interpretation of intrinsically infinite-dimensional functional data. In contrast to conventional multivariate techniques that treat discrete observations as independent units, FDA regards each function or curve as a single observational entity, thereby capturing the continuity and dynamic structure inherent in the data. A central problem in FDA is estimating the mean and covariance functions [1], which are essential for characterizing overall trends, quantifying variability, and constructing predictive models. Accurate estimation of these quantities not only reduces dimensionality but also provides a principled foundation for downstream tasks such as regression.

A variety of conventional methods exist for functional data, including functional principal component analysis (FPCA) [2], local linear smoothers [3], and smoothing splines [4]. Compared with these general approaches, the wavelet method offers distinct advantages when the data exhibit localized features or abrupt changes [5]. Most notably, the multi-resolution property of wavelets enables a decomposition of the signal into components at different scales, thereby capturing global trends and fine local details simultaneously.

Despite these merits, traditional wavelet methods encounter serious difficulties when applied to functions defined on a bounded interval. A major challenge is boundary treatment: widely used techniques such as symmetric or periodic extensions often introduce artifacts or systematic biases near the boundaries, compromising estimation accuracy. Moreover, classical wavelet bases tend to become unstable at the interval endpoints, which severely limits their ability to represent local features precisely. These boundary effects are particularly problematic in the FDA and often limit the practical utility of wavelet-based approaches.

To overcome these limitations, the concept of boundary wavelets has been developed. Such kind of wavelets are intrinsically defined on a finite interval, thus circumventing the boundary biases induced by extension methods [6]. In this paper, we go further by providing a construction method for boundary orthogonal wavelets from the Daubechies-N family and integrating them into the estimation of mean and covariance functions for sparse functional data. The objective is to fully retain the multiscale advantages of wavelet analysis—such as sparse representation and local adaptivity—while mitigating boundary effects. This framework is expected to address the boundary issues that have hindered conventional wavelet-based FDA.

The remainder of this paper is organized as follows. In Section II, we present the construction of a boundary orthogonal wavelet basis for $L_2([0,1])$ adapted from an orthogonal wavelet basis for $L_2(\mathbb{R})$ with illustrated examples from the Daubechies-2 and Daubechies-4 wavelets. In Section III, we apply our constructed boundary orthogonal wavelet (BoW) system to the analysis of sparse functional data, including numerical simulations on synthetic data and numerical comparisons on real-world CD4 data. Conclusion and further remarks are given in Section IV.

II. Boundary Orthogonal Wavelet Systems

A. Construction

Our goal is to construct an orthogonal wavelet system on $[0,1]$ from the Daubechies-N orthogonal wavelet family [7] supported on $[-N+1, N]$. For sufficiently large scales, the boundary function on one side will not interfere with the one on the other. Therefore, we can focus on constructing an orthogonal basis $\mathcal{W}_{[0,\infty)}$ of the half space $L_2([0,\infty))$. Similar approach can be done for the orthogonal basis $\mathcal{W}_{(-\infty,1]}$ of the

X. Zhuang was supported in part by the Research Grants Council of Hong Kong (Project no. CityU 11302023, CityU 11301224, and CityU 11300825) and the National Natural Science Foundation of China (NSFC 12471400). M. Li acknowledged the support from the "Pioneer" and "Leading Goose" R&D Program of Zhejiang (No. 2024C03262), and the Jinhua Science and Technology Plan (No. 2023-3-003a).

979-8-3195-1509-4/26/$31.00 ©2026 IEEE
DOI 10.1109/ICKG70389.2026.00035

half space $L_2((-\infty,1])$ and then the final orthogonal basis $\mathcal{W}_{[0,1]}$ is given by $\mathcal{W}_{[0,\infty)} \cap \mathcal{W}_{(-\infty,1]}$.

This idea was first proposed by Meyer [8]. We provide a sketch for the construction of the left scaling function (vector) $\phi^L = [\phi^{L,1}, \ldots, \phi^{L,2N-2}]^\top$ given the Daubechies orthogonal scaling function ϕ supported on $[-N+1, N]$. The right boundary scaling function (vector) ϕ^R can be obtained similarly. Note that the support of $\phi(\cdot - k)$ will "cross" 0 when $-N+1 \le k \le N-2$. We can obtain ϕ^L simply by applying Gram-Schmidt orthogonalization to $\{\phi(\cdot - k)\chi_{[0,\infty)} : -N+1 \le k \le N-2\}$. This approach preserves the polynomial reproduction of ϕ using $2N-2$ left boundary functions.

Cohen et al. [9] and later Han and Michelle [10] show that one only needs N boundary scaling functions on one side to preserve the polynomial reproduction property of ϕ. Our construction is based on the boundary scaling function ϕ^{p} with polynomial reproduction given by Han and Michelle [10]. Let us be more specific about the procedure.

Suppose an orthogonal wavelet basis $\mathcal{W}_{j_0}(\{\phi,\psi\})$, e.g., the Daubechies orthogonal wavelet basis, for $L_2(\mathbb{R})$ is already given by

$$\mathcal{W}_{j_0}(\{\phi,\psi\}) := \{\phi_{j_0;k}; k \in \mathbb{Z}\} \cup \{\psi_{j;k} : k \in \mathbb{Z}\}_{j=j_0}^{\infty}, \quad (1)$$

where for a function f, $f_{j;k} := 2^{j/2} f(2^j \cdot - k)$. This system is associated with an orthogonal wavelet filter bank $\{a, b\}$ by the following refinement equation (two-scale relation):

$$\phi = 2\sum_k a(k)\phi(2\cdot - k),\ \psi = 2\sum_k b(k)\phi(2\cdot - k). \quad (2)$$

We say that a has order r sum rules, i.e., $\mathrm{sr}(a) = r$ if

$$\sum_{k\in\mathbb{Z}} a(k)(-1)^k k^m = 0,\ m = 0, \ldots, r-1. \quad (3)$$

Furthermore, there exists a matching filter $v \in \ell_0(\mathbb{Z})$ of a with $\mathrm{sr}(a) = r$ such that for all $\mathsf{p} \in \mathscr{P}_{r-1}(\mathbb{R})$ of polynomials of degree up to $r-1$, we have

$$\mathsf{p} = \sum_{k\in\mathbb{Z}} \mathsf{p}_v(k)\phi(\cdot - k) \text{ with } \mathsf{p}_v := \mathsf{p} * v. \quad (4)$$

Then we can construct a boundary orthogonal wavelet system $\mathcal{W}_{j_0}^b(\{\Phi,\Psi\})$ for $L_2([0,1])$ with $\Phi = \{\phi^L, \phi, \phi^R\}$ and $\Psi = \{\psi^L, \psi, \psi^R\}$ by the following steps.

S1) **Construction of ϕ^{p} with polynomial reproduction on the left boundary.** Since the linearity of the polynomial space, it is sufficient to consider the monomial vector $\mathsf{p} = [x^0, \ldots, x^{r-1}]^\top$. With (4), we can define

$$\phi^{\mathsf{p}} := \mathsf{p}(x) - \sum_{k=n_\phi}^{\infty} \mathsf{p}_v(k)\phi(x-k),\ x \in [0,\infty), \quad (5)$$

where $n_\phi \ge \max\{-l_\phi, -l_a\}$ and $[l_f, h_f] = \mathrm{supp}(f)$ denotes the support of a function (or filter) f. Then $\{\phi^{\mathsf{p}}\} \cup \{\phi(\cdot - k) : k \ge n_\phi\}$ can reproduce polynomial $\mathsf{p}\chi_{[0,\infty)}$ for $\mathsf{p} \in \mathscr{P}_{r-1}$. Let $A_{\mathsf{p}} := \mathrm{diag}(2^{-1}, \ldots, 2^{-r})$, then we have $\mathsf{p}(x) = 2A_{\mathsf{p}}\mathsf{p}(2x)$, together with (5), we can deduce that

$$\phi^{\mathsf{p}} = 2A_{\mathsf{p}}\phi^{\mathsf{p}}(2\cdot) + 2\sum_{k=n_\phi}^{\infty} a_{\mathsf{p}}(k)\phi(2\cdot - k), \quad (6)$$

where $a_{\mathsf{p}}(k) := A_{\mathsf{p}}\mathsf{p}_v(k) - \sum_{m=n_\phi}^{\infty} \mathsf{p}_v(m)a(k-2m)$. In fact, $a_{\mathsf{p}}(k) = 0$ for all $k \ge h_a + 2n_\phi$.

S2) **Construction of left boundary scaling function (vector) ϕ^L with orthogonality from ϕ^{p}.** It is easy to check that $V(k) := \langle \phi^{\mathsf{p}}, \phi(\cdot - k)\rangle = 0$ for $k \ge N_\phi := h_{\phi^{\mathsf{p}}} - l_\phi$. Then we can define $\breve{\phi}^{\mathsf{p}} := \phi^{\mathsf{p}} - \sum_{k=n_\phi}^{N_\phi - 1} V(k)\phi(\cdot - k)$ so that $\langle \breve{\phi}^{\mathsf{p}}, \phi(\cdot - k)\rangle = 0$ for all $k \ge n_\phi$. We can obtain ϕ^L by (Gram-Schmidt) orthogonalizing $\breve{\phi}^{\mathsf{p}}$, that is, $\phi^L = G\breve{\phi}^{\mathsf{p}}$ for some matrix G. Furthermore, G is a lower triangular matrix and is necessarily invertible. Then, we can deduce that

$$\phi^{\mathsf{p}} = G^{-1}\phi^L + \sum_{k=n_\phi}^{N_\phi - 1} V(k)\phi(\cdot - k). \quad (7)$$

Let $\phi^C := [\phi(\cdot - n_\phi), \ldots, \phi(\cdot - N_\phi + 1)]^\top$ and $\phi^F := [\phi(2\cdot - n_\phi), \ldots, \phi(2\cdot - h_a - 2N_\phi + 2)]^\top$, then based on (2), there exists a matrix H such that $\phi^C = H\phi^F$. Therefore, denoting $Q := [V(n_\phi), \ldots, V(N_\phi - 1)]$, we have

$$\phi^L = 2A_L\phi^L(2\cdot) + 2\sum_{k=n_\phi}^{\infty} a_L(k)\phi(2\cdot - k), \quad (8)$$

where $A_L = GA_{\mathsf{p}}G^{-1}$ and $a_L = GA_{\mathsf{p}}Q + Ga_{\mathsf{p}} - GQH$.

S3) **Construction of left boundary wavelet function (vector) ψ^L with vanishing moments.** We say that a function $\zeta \in L_2([0,1]^d)$ has r-th order vanishing moments if $\int_{[0,1]^d} \zeta(x)\mathsf{p}(x)\mathrm{d}x = 0$ for all $\mathsf{p}(x) \in \mathscr{P}_{r-1}([0,1]^d)$. Denoting $m_\phi := \max(2n_\phi + h_a, 2n_\psi + h_b)$ with $n_\psi := \max(-l_\psi, \frac{n_\phi - l_b}{2}, h_{\phi^L})$, one can show that $\{\phi(2\cdot - k) : k \ge m_\phi\}$ can be represented by $\{\phi(\cdot - k) : k \ge n_\phi\} \cup \{\psi(\cdot - k) : k \ge n_\psi\}$. Then, there exists $\psi_b \subseteq \mathrm{span}(\{\phi^L(2\cdot)\} \cup \{\phi(2\cdot - k) : n_\phi \le k < m_\phi\})$ such that ψ_b and $\{\phi^L\} \cup \{\phi(\cdot - k) : k \ge n_\phi\}$ are orthogonal, which implies ψ_b has r vanishing moments. We can obtain ψ^L from ψ_b by following a similar orthogonalization process to that of ϕ^{p}. Therefore, we have

$$\psi^L = 2B_L\phi^L(2\cdot) + 2\sum_{k=n_\phi}^{\infty} b_L(k)\phi(2\cdot - k). \quad (9)$$

S4) **Construction of right boundary scaling function (vector) ϕ^R and right boundary wavelet function (vector) ψ^R.** Denote $\mathring{\phi} := \phi(-\cdot)$ and $\mathring{\psi} := \psi(-\cdot)$, we can repeat the process of constructing the left boundary to obtain $\mathring{\phi}^L$ and $\mathring{\psi}^L$ with

$$\mathring{\phi}^L = 2A_R\mathring{\phi}^L(2\cdot) + 2\sum_{k=n_{\mathring{\phi}}}^{\infty} a_R(k)\mathring{\phi}(2\cdot - k) \quad (10)$$

and

$$\mathring{\psi}^L = 2B_R\mathring{\phi}^L(2\cdot) + 2\sum_{k=n_{\mathring{\phi}}}^{\infty} b_R(k)\mathring{\phi}(2\cdot - k). \qquad (11)$$

Then, the right boundary scaling function ϕ^R and right boundary wavelet function ψ^R can be obtained by $\phi^R = \mathring{\phi}^L(1-\cdot)$ and $\psi^R = \mathring{\psi}^L(1-\cdot)$.

S5) **Final system.** Define

$$\Phi_j := \{\phi^L_{j;0}\} \cup \{\phi_{j;k} : n_\phi \le k \le 2^j - n_{\mathring{\phi}}\} \cup \{\phi^R_{j;2^j-1}\}$$

and

$$\Psi_j := \{\psi^L_{j;0}\} \cup \{\psi_{j;k} : n_\psi \le k \le 2^j - n_{\mathring{\psi}}\} \cup \{\psi^R_{j;2^j-1}\}.$$

Then $\mathcal{W}^b_{j_0}(\{\Phi,\Psi\}) := \{\Phi_{j_0}\} \cup \{\Psi_j\}_{j\ge j_0}$ is a boundary orthogonal wavelet basis of $L_2([0,1])$ for some j_0.

B. Examples

In this section, we provide two examples to illustrate the general construction methods for boundary orthogonal wavelet systems on $[0,1]$ from the Daubechies-N family of orthogonal wavelets.

Example 1: **(Daub2).** Consider the Daubechies-2 compactly supported orthogonal wavelet system $\mathcal{W}_{j_0}(\{\phi,\psi\})$ with $\int_{\mathbb{R}}\phi(x)\mathrm{d}x = 1$ and an orthogonal wavelet filter bank $\{a,b\}$ given by

$$a = \{\frac{1+\sqrt{3}}{8}, \frac{3+\sqrt{3}}{8}, \frac{3-\sqrt{3}}{8}, \frac{1-\sqrt{3}}{8}\}_{[-1,2]},$$
$$b = \{\frac{1-\sqrt{3}}{8}, \frac{-3+\sqrt{3}}{8}, \frac{3+\sqrt{3}}{8}, \frac{-1-\sqrt{3}}{8}\}_{[-1,2]}.$$

Then $\mathrm{sr}(a) = 2$. Following the above procedure of the construction and denoting $s_1 = (4-\sqrt{3})^{-\frac{1}{2}}$ and $s_2 = (4+\sqrt{3})^{-\frac{1}{2}}$, the left boundary elements ϕ^L, ψ^L can be represented as

$$\phi^L = \begin{bmatrix} 1 & 0 \\ \frac{2-\sqrt{3}}{6} & \frac{1}{2}\end{bmatrix}\phi^L(2\cdot) + s_1\begin{bmatrix}\frac{3-\sqrt{3}}{2}\\ \frac{\sqrt{3}-3}{3}\end{bmatrix}\phi(2\cdot-1) + s_1\begin{bmatrix}\frac{1-\sqrt{3}}{2}\\ \frac{\sqrt{3}-1}{3}\end{bmatrix}\phi(2\cdot-2),$$
$$\psi^L = s_1\begin{bmatrix}\frac{\sqrt{6}-9\sqrt{2}}{13} & \frac{19\sqrt{2}-5}{13}\end{bmatrix}\phi^L(2\cdot) + \frac{12\sqrt{2}+3\sqrt{6}}{26}\phi(2\cdot-1) - \frac{4\sqrt{6}+3\sqrt{2}}{26}\phi(2\cdot-2).$$

With $\phi^R = \mathring{\phi}^L(1-\cdot)$ and $\psi^R = \mathring{\psi}^L(1-\cdot)$, the right boundary elements can be determined by

$$\mathring{\phi}^L = \begin{bmatrix} 1 & 0 \\ -\frac{2+\sqrt{3}}{6} & \frac{1}{2}\end{bmatrix}\mathring{\phi}^L(2\cdot) + s_2\begin{bmatrix}\frac{3+\sqrt{3}}{2}\\ \frac{3+\sqrt{3}}{3}\end{bmatrix}\mathring{\phi}(2\cdot-2) + s_2\begin{bmatrix}\frac{1+\sqrt{3}}{2}\\ \frac{1+\sqrt{3}}{3}\end{bmatrix}\mathring{\phi}(2\cdot-3),$$
$$\mathring{\psi}^L = -s_2\begin{bmatrix}\frac{9\sqrt{2}+\sqrt{6}}{13} & \frac{19\sqrt{2}+5\sqrt{6}}{13}\end{bmatrix}\mathring{\phi}^L(2\cdot) + \frac{12\sqrt{2}-3\sqrt{6}}{26}\mathring{\phi}(2\cdot-2) + \frac{4\sqrt{6}-3\sqrt{2}}{26}\mathring{\phi}(2\cdot-3).$$

Note that $\#\phi^L = \#\phi^R = 2$ and $\#\psi^L = \#\psi^R = 1$. Define the generators at scale j by

$$\Phi_j = \{\phi^L_{j;0}\} \cup \{\phi_{j;k} : 1 \le k \le 2^j - 2\} \cup \{\phi^R_{j;2^j-1}\},$$
$$\Psi_j = \{\psi^L_{j;0}\} \cup \{\psi_{j;k} : 1 \le k \le 2^j - 2\} \cup \{\psi^R_{j;2^j-1}\}.$$

Then $\mathcal{W}^b_{j_0}(\{\Phi,\Psi\})$ for any $j_0 \ge 2$ is a boundary orthogonal wavelet basis of $L_2([0,1])$. See Fig. 1 for $\Phi = \{\phi^L, \phi, \phi^R\}$ and $\Psi = \{\psi^L, \psi, \psi^R\}$.

Example 2: **(Daub4).** Consider the Daubechies-4 compactly supported orthogonal wavelet system $\mathcal{W}_{j_0}(\{\phi,\psi\})$ with $\int_{\mathbb{R}}\phi(x)\mathrm{d}x = 1$ and an orthogonal wavelet filter bank $\{a,b\}$ given by

$$a = \{-7.4935\text{E-}3, 2.3252\text{E-}2, 2.1808\text{E-}2, -1.3225\text{E-}1, -1.9788\text{E-}2, 4.4610\text{E-}1, 5.0547\text{E-}1, 1.6290\text{E-}1\}_{[-3,4]},$$
$$b = \{-1.6290\text{E-}1, 5.0547\text{E-}1, -4.4610\text{E-}1, -1.9788\text{E-}2, 1.3225\text{E-}1, 2.1808\text{E-}2, -2.3252\text{E-}2, -7.4935\text{E-}3\}_{[-3,4]}.$$

Then $\mathrm{sr}(a) = 4$. Following the procedure of the construction, the left boundary scaling function ϕ^L can be represented as

$$\phi^L = 2A_L\phi^L(2\cdot) + 2a_L\cdot[\phi(2\cdot-3),\ldots,\phi(2\cdot-8)]^\top,$$

where

$$A_L = \begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 \\ -4.3573\text{E-}1 & \frac{1}{4} & 0 & 0 \\ -1.1014\text{E-}2 & -4.9698\text{E-}1 & \frac{1}{8} & 0 \\ 1.7173\text{E-}1 & 2.7575\text{E-}1 & -3.9423\text{E-}1 & \frac{1}{16}\end{bmatrix}$$

and

$$a_L^\top = \begin{bmatrix} 2.1650\text{E-}1 & 4.1573\text{E-}2 & -2.2627\text{E-}1 & -9.0871\text{E-}2 \\ 2.0339\text{E-}1 & 8.2263\text{E-}2 & -2.2403\text{E-}1 & -3.2684\text{E-}1 \\ 2.0720\text{E-}1 & 1.6004\text{E-}1 & -9.5777\text{E-}2 & -2.8492\text{E-}1 \\ 2.5981\text{E-}1 & 3.2848\text{E-}1 & 2.1246\text{E-}1 & 6.6386\text{E-}2 \\ 2.1564\text{E-}1 & 3.0934\text{E-}1 & 2.7183\text{E-}1 & 1.7674\text{E-}1 \\ 6.9496\text{E-}2 & 9.9691\text{E-}2 & 8.7604\text{E-}2 & 5.6959\text{E-}2\end{bmatrix}.$$

The left boundary wavelet function ψ^L can be represented as

$$\psi^L = 2B_L\phi^L(2\cdot) + 2b_L\cdot[\phi(2\cdot-3),\ldots,\phi(2\cdot-8)]^\top,$$

where

$$B_L = \begin{bmatrix} -8.1882\text{E-}2 & -2.4005\text{E-}1 & -5.1613\text{E-}1 & 1.9965\text{E-}1 \\ -6.0622\text{E-}2 & -1.2290\text{E-}1 & -2.3373\text{E-}1 & -6.3455\text{E-}1 \\ -1.4192\text{E-}1 & -2.0432\text{E-}1 & -8.9108\text{E-}2 & 2.3148\text{E-}1\end{bmatrix}$$

and

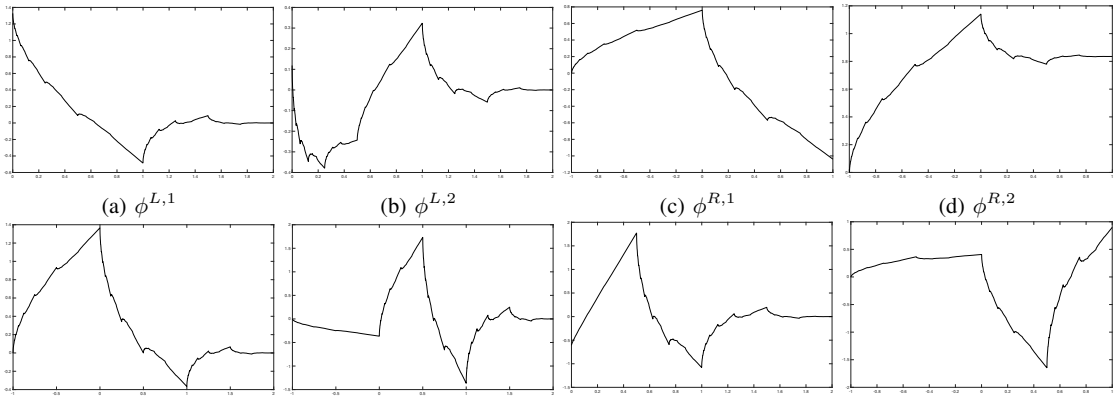
$$b_L^\top = \begin{bmatrix} -1.5097\text{E-}1 & 1.4994\text{E-}1 & 5.6575\text{E-}1 \\ 1.9351\text{E-}1 & 1.6688\text{E-}2 & -3.2908\text{E-}2 \\ 2.4877\text{E-}1 & -3.3020\text{E-}2 & -2.2815\text{E-}1 \\ 7.6895\text{E-}3 & -2.9158\text{E-}3 & -3.2549\text{E-}2 \\ -8.0973\text{E-}2 & 8.6307\text{E-}3 & 4.5777\text{E-}2 \\ -2.6096\text{E-}2 & 2.7815\text{E-}3 & 1.4753\text{E-}2\end{bmatrix}.$$

Since $\phi^R = \mathring{\phi}^L(1-\cdot)$, the right boundary scaling function ϕ^R can be represented with

$$\mathring{\phi}^L = 2A_R\mathring{\phi}^L(2\cdot) + 2a_R\cdot[\mathring{\phi}(2\cdot-4),\ldots,\mathring{\phi}(2\cdot-9)]^\top,$$

where

$$A_R = \begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 \\ -4.5537\text{E-}1 & \frac{1}{4} & 0 & 0 \\ 1.0966\text{E-}1 & 1.6376\text{E-}1 & \frac{1}{8} & 0 \\ 1.6479\text{E-}1 & 5.9924\text{E-}1 & 8.5089\text{E-}2 & \frac{1}{16}\end{bmatrix}$$



(a) $\phi^{L,1}$ (b) $\phi^{L,2}$ (c) $\phi^{R,1}$ (d) $\phi^{R,2}$

(e) ϕ (f) ψ (g) ψ^L (h) ψ^R

Fig. 1: The generators of the boundary orthogonal wavelet basis of $L_2([0,1])$ from Daubechies-2 wavelet family. (e) and (f): scaling function ϕ and wavelet function ψ; (a)-(b) and (c)-(d): left boundary scaling function $\phi^L = [\phi^{L,1}, \phi^{L,2}]^\top$ and right boundary scaling function $\phi^R = [\phi^{R,1}, \phi^{R,2}]^\top$; (g) and (h): left boundary wavelet function ψ^L and right boundary wavelet function ψ^R.

and

$$a_R^\top = \begin{bmatrix} 4.7418\text{E-}1 & 4.5413\text{E-}1 & 3.8535\text{E-}2 & -1.9999\text{E-}1 \\ -7.6984\text{E-}3 & -8.9076\text{E-}2 & -4.4292\text{E-}1 & 1.2189\text{E-}1 \\ -1.5333\text{E-}1 & -7.9489\text{E-}2 & 4.7717\text{E-}1 & -9.7924\text{E-}2 \\ 2.0136\text{E-}2 & 7.7258\text{E-}2 & -1.4129\text{E-}1 & 1.4403\text{E-}1 \\ 3.2707\text{E-}2 & -5.7689\text{E-}2 & -1.3954\text{E-}2 & -1.2564\text{E-}1 \\ -1.0541\text{E-}2 & 1.8592\text{E-}2 & 4.4970\text{E-}3 & 4.0492\text{E-}2 \end{bmatrix}.$$

Similarly with $\psi^R = \mathring{\psi}^L(1-\cdot)$, the right boundary wavelet function ψ^R can be represented with

$$\mathring{\psi}^L = 2B_R\mathring{\phi}^L(2\cdot) + 2b_R \cdot [\mathring{\phi}(2\cdot - 4), \ldots, \mathring{\phi}(2\cdot - 9)]^\top,$$

where

$$B_R = \begin{bmatrix} 2.1163\text{E-}3 & -1.0752\text{E-}1 & 4.2393\text{E-}1 & -4.7840\text{E-}1 \\ -3.7985\text{E-}2 & 1.7772\text{E-}1 & 1.2856\text{E-}1 & -2.7313\text{E-}1 \\ 4.4798\text{E-}2 & 9.1936\text{E-}2 & -5.2999\text{E-}1 & -4.3889\text{E-}1 \end{bmatrix}$$

and

$$b_R^\top = \begin{bmatrix} -7.1726\text{E-}3 & 9.3601\text{E-}3 & -2.6487\text{E-}2 \\ -1.2556\text{E-}2 & 1.4566\text{E-}1 & -5.9604\text{E-}2 \\ -3.6357\text{E-}2 & -4.8624\text{E-}2 & 5.5945\text{E-}2 \\ 1.8819\text{E-}1 & -3.7571\text{E-}1 & 4.9191\text{E-}2 \\ -1.9714\text{E-}1 & 4.3722\text{E-}1 & -7.5094\text{E-}2 \\ 6.3533\text{E-}2 & -1.4091\text{E-}1 & 2.4201\text{E-}2 \end{bmatrix}.$$

Note that $\#\phi^L = \#\phi^R = 4$ and $\#\psi^L = \#\psi^R = 3$. Define the generators at scale j by

$$\Phi_j = \{\phi^L_{j;0}\} \cup \{\phi_{j;k} : 3 \le k \le 2^j - 4\} \cup \{\phi^R_{j;2^j-1}\},$$
$$\Psi_j = \{\psi^L_{j;0}\} \cup \{\psi_{j;k} : 3 \le k \le 2^j - 4\} \cup \{\psi^R_{j;2^j-1}\}.$$

Then $\mathcal{W}^b_{j_0}(\{\Phi, \Psi\})$ for any $j_0 \ge 3$ is a boundary orthogonal wavelet basis of $L_2([0,1])$. See Fig. 2 for $\Phi = \{\phi^L, \phi, \phi^R\}$ and $\Psi = \{\psi^L, \psi, \psi^R\}$.

III. Numerical Studies

In this section, we apply our constructed boundary orthogonal wavelet (BoW) system to the analysis of sparse functional data. We first briefly discuss the functional data analysis for the mean estimator $\hat{\mu}_n$ and the covariance estimator $\hat{\gamma}_n$ from our BoW system. Then, we run the numerical simulation and comparisons on synthetic data. Finally, numerical comparisons are done on the real-world CD4 data.

A. Stochastic model

Let $\{X(t), t \in [0,1]\}$ be an L_2 stochastic process defined on a fixed interval $[0,1]$. The mean and covariance functions of $X(t)$ are defined as

$$\mu(t) = \mathrm{E}[X(t)] \text{ and } \gamma(t,t') = \mathrm{Cov}(X(t), X(t')),$$

respectively. Decompose $X(t)$ into

$$X(t) = \mu(t) + U(t), \tag{12}$$

where $U(t)$ is the stochastic part of $X(t)$ with $\mathrm{E}[U(t)] = 0$. We can infer that

$$\gamma(t,t') = \mathrm{Cov}(X(t), X(t')) = \mathrm{Cov}(U(t), U(t')).$$

Let $X_i, i \in \{1, \ldots, n\}$ be i.i.d. copies of $X(t)$, which are not observable in practice. We observe the i-th object X_i with random errors at N_i random points $T_{i,m}, m \in \{1, \ldots, N_i\}$, where N_i's are i.i.d. copies of random variable N. Thus, we model the sparse functional data as

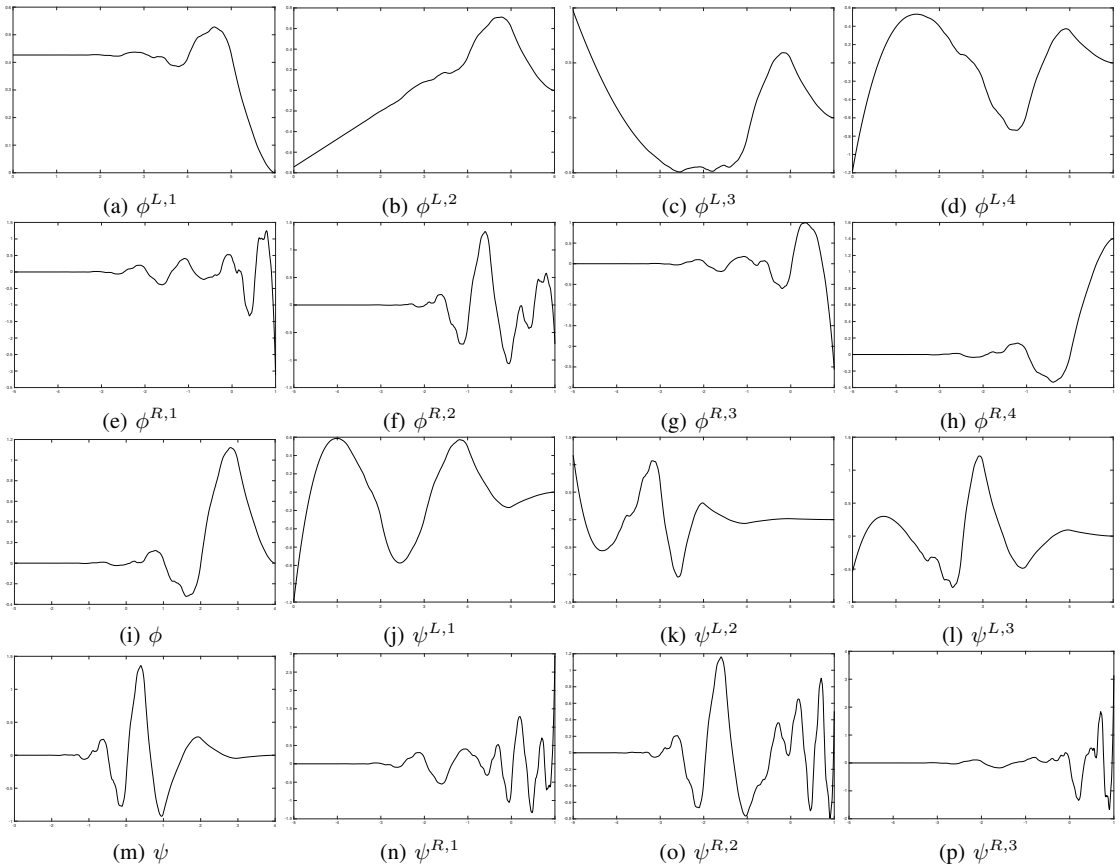
$$Y_{i,m} = \mu(T_{i,m}) + U_i(T_{i,m}) + \epsilon_{i,m}, \tag{13}$$

where $\epsilon_{i,m}$ are i.i.d. copies of error ϵ with $\mathrm{E}(\epsilon) = 0$ and $\mathrm{Var}(\epsilon) = \sigma^2$. Here, 'sparse' means that N_i is bounded for all i.

Since our BoW system constitutes an orthogonal basis of $L_2([0,1])$, we can represent any $\mu \in L_2([0,1])$ as

$$\mu = \sum_{\eta \in \Phi_{j_0}} \langle \mu, \eta \rangle \eta + \sum_{j \ge j_0} \sum_{\zeta \in \Psi_j} \langle \mu, \zeta \rangle \zeta. \tag{14}$$

Furthermore, the bivariate boundary orthogonal wavelet basis $\mathcal{W}^b_{j_0}(\{\mathbf{\Phi}, \mathbf{\Psi}\}) = \mathbf{\Phi}_{j_0} \cup \{\mathbf{\Psi}_j\}_{j \ge j_0}$ for $L_2([0,1]^2)$ can be constructed by using the tensor product technique, that is,



(a) $\phi^{L,1}$ (b) $\phi^{L,2}$ (c) $\phi^{L,3}$ (d) $\phi^{L,4}$

(e) $\phi^{R,1}$ (f) $\phi^{R,2}$ (g) $\phi^{R,3}$ (h) $\phi^{R,4}$

(i) ϕ (j) $\psi^{L,1}$ (k) $\psi^{L,2}$ (l) $\psi^{L,3}$

(m) ψ (n) $\psi^{R,1}$ (o) $\psi^{R,2}$ (p) $\psi^{R,3}$

Fig. 2: The generators of the boundary orthogonal wavelet basis of $L_2([0,1])$ from Daubechies-4 wavelet family. (i) and (m): scaling function ϕ and wavelet function ψ; (a)-(d) and (e)-(h): left boundary scaling function $\phi^L = [\phi^{L,1}, \ldots, \phi^{L,4}]^\top$ and right boundary scaling function $\phi^R = [\phi^{R,1}, \ldots, \phi^{R,4}]^\top$; (j)-(l) and (n)-(p): left boundary wavelet function $\psi^L = [\psi^{L,1}, \ldots, \psi^{L,4}]^\top$ and right boundary wavelet function $\psi^R = [\psi^{R,1}, \ldots, \psi^{R,3}]^\top$.

$\boldsymbol{\Phi}_j := \Phi_j \otimes \Phi_j$ and $\boldsymbol{\Psi}_j = \{\Phi_j \otimes \Psi_j, \Psi_j \otimes \Phi_j, \Psi_j \otimes \Psi_j\}$. Then for any $\gamma \in L_2([0,1]^2)$, we have

$$\gamma = \sum_{\boldsymbol{\eta} \in \boldsymbol{\Phi}_{j_0}} \langle \gamma, \boldsymbol{\eta} \rangle \boldsymbol{\eta} + \sum_{j \geq j_0} \sum_{\boldsymbol{\zeta} \in \boldsymbol{\Psi}_j} \langle \gamma, \boldsymbol{\zeta} \rangle \boldsymbol{\zeta}. \tag{15}$$

With given sample data $\{(T_{i,m}, Y_{i,m}) : m = 1, \ldots, N_i, i = 1, \ldots, n\}$, we can provide an empirical version of μ as

$$\check{\mu}_n = \sum_{\eta \in \Phi_{j_0}} \check{\alpha}_{j_0}^{\eta} \eta + \sum_{j=j_0}^{\lfloor \log_2 n \rfloor} \sum_{\zeta \in \Psi_j} \check{\beta}_j^{\zeta} \zeta, \tag{16}$$

where

$$\check{\alpha}_{j_0}^{\eta} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{N_i} \sum_{m=1}^{N_i} Y_{i,m} \eta(T_{i,m}), \tag{17}$$

$$\check{\beta}_j^{\zeta} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{N_i} \sum_{m=1}^{N_i} Y_{i,m} \zeta(T_{i,m}). \tag{18}$$

Note that $\check{\alpha}_{j_0}^{\eta}, \check{\beta}_j^{\zeta}$ are from the noisy observations. To reduce the noise effect, we perform soft thresholding on the wavelet detail coefficients $\check{\beta}_j^{\zeta}$. The threshold is set to $c_j/\sqrt{n}$, where the $1/\sqrt{n}$ rate follows from the standard noise level in the wavelet domain and c_j are scale-dependent constants chosen according to a conventional thresholding rule. Denoting $\hat{\beta}_j^{\zeta}$ as the thresholded results of $\check{\beta}_j^{\zeta}$, the *mean estimator* $\hat{\mu}_n$ can be expressed as

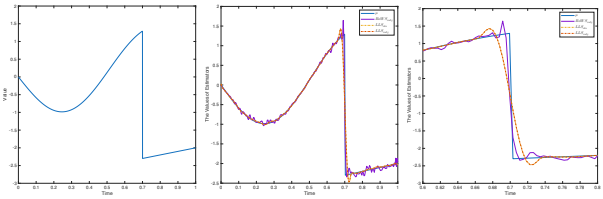
$$\hat{\mu}_n = \sum_{\eta \in \Phi_{j_0}} \check{\alpha}_{j_0}^{\eta} \eta + \sum_{j=j_0}^{\lfloor \log_2 n \rfloor} \sum_{\zeta \in \Psi_j} \hat{\beta}_j^{\zeta} \zeta. \tag{19}$$

The coefficients $\check{\boldsymbol{\alpha}}_{j_0}^{\boldsymbol{\eta}}, \check{\boldsymbol{\beta}}_j^{\boldsymbol{\zeta}}$ in the bivariate case can be estimated by replacing $Y_{i,m}$ with the "row covariance"

$$C_{i,m_1,m_2} = [Y_{i,m_1} - \hat{\mu}_n(T_{i,m_1})][Y_{i,m_2} - \hat{\mu}_n(T_{i,m_2})]. \tag{20}$$

After applying the soft thresholding to $\check{\boldsymbol{\beta}}_j^{\boldsymbol{\zeta}}$, the *covariance estimator* of the covariance function γ can be expressed as

$$\hat{\gamma}_n = \sum_{\boldsymbol{\eta} \in \boldsymbol{\Phi}_{j_0}} \check{\boldsymbol{\alpha}}_{j_0}^{\boldsymbol{\eta}} \boldsymbol{\eta} + \sum_{j=j_0}^{\lfloor \frac{1}{2} \log_2 n \rfloor} \sum_{\boldsymbol{\zeta} \in \boldsymbol{\Psi}_j} \hat{\boldsymbol{\beta}}_j^{\boldsymbol{\zeta}} \boldsymbol{\zeta}. \tag{21}$$



(a) μ (b) $\hat{\mu}_n$ (c) $\hat{\mu}_n$, Zoom in

Fig. 3: (a) the ground truth mean function μ; (b) Visualization comparison of different approaches with the average curve $\mathbb{E}(\hat{\mu}_n) \approx \frac{1}{P}\sum_{p=1}^{P} \hat{\mu}_n^{[p]}$ for $n = 2^8 = 256$, $\sigma = 0.1$ and $P = 500$; (c) Zoom in on $[0.6, 0.8]$. Our method, BoWS (purple line) outperforms LLS near the discontinuity point at 0.7.

B. Synthesis data

To test the performance of the boundary orthogonal wavelet (BoW) system, we consider a true mean function

$$\mu(t) = \left(t^3 - \sin(2\pi t)\right)\mathbf{I}(t \le 0.7) + (t-3)\mathbf{I}(t > 0.7),$$

where $\mathbf{I}(\cdot)$ is an indicator function and a true covariance function γ given by

$$\gamma(t, t') = \sin(2\pi t)\sin(2\pi t') + \frac{1}{4}\cos(2\pi t)\cos(2\pi t'),$$

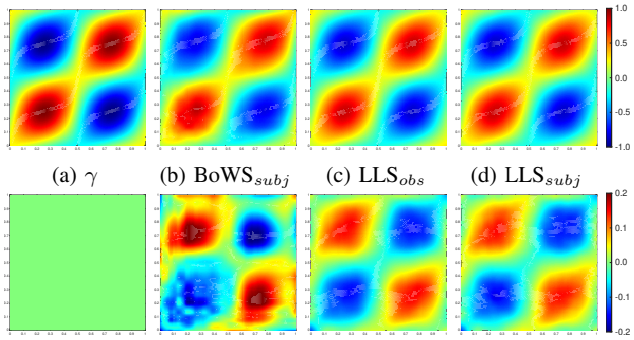
where $t, t' \in [0, 1]$. We compare our boundary orthogonal wavelet method with soft thresholding (denoted as BoWS) against the local linear smoother (LLS) [3]. The system used in the simulation is shown in Example 2. In consistency with Zhang and Wang [3], we use the subscripts "obs" and "subj" to denote the OBS and SUBJ weighting schemes, where the OBS scheme refers to the equal weight per observation $w_i = 1/\sum_{i=1}^{n} N_i$ while the SUBJ scheme refers to the equal weight per subject $w_i = 1/(nN_i)$.

The simulation results of the two methods for the mean function μ are shown in Fig. 3. We can observe that the function μ is smooth overall, except for the jump discontinuity at 0.7. The LLS method fits the smooth parts of the function well, but struggles significantly with the Gibbs phenomenon at the jump discontinuity. In contrast, our method demonstrates its advantages in this situation, as shown in Fig. 3(c).

We have presented the simulation results for the covariance function γ in Fig. 4. One can see that the LLS methods yield a larger overall error than our BoWs method.

C. Real data

We apply our method to a CD4 percentage data set from the Multicenter AIDS Cohort Study [11]. This dataset shows the change in the percentage of CD4 counts among 283 homosexual men who developed HIV between 1984 and 1991. The number of time points per observation ranged from 1 to 14, resulting in 1,817 records. The raw CD4 data is shown in Fig. 5 (c). Using our method and LLS, we present the overall trend in CD4 percentage changes, as shown in Fig. 5 (d).



(a) γ (b) BoWS$_{subj}$ (c) LLS$_{obs}$ (d) LLS$_{subj}$

(e) 0 map (f) Error map (g) Error map (h) Error map

Fig. 4: Visualization comparison of different approaches with the 2D average function $\mathbb{E}(\hat{\gamma}_n) \approx \frac{1}{P}\sum_{p=1}^{P} \hat{\gamma}_n^{[p]}$ from the corresponding $\hat{\mu}_n^{[p]}$ with $P = 500$. (a) the ground truth covariance function γ; (e) 0 map; (b) and (f): BoWS$_{subj}$ result and the corresponding error map between the result and γ; (c) and (g): LLS$_{obs}$ result and the corresponding error map between the result and γ; (d) and (h): LLS$_{subj}$ result and the corresponding error map between the result and γ.

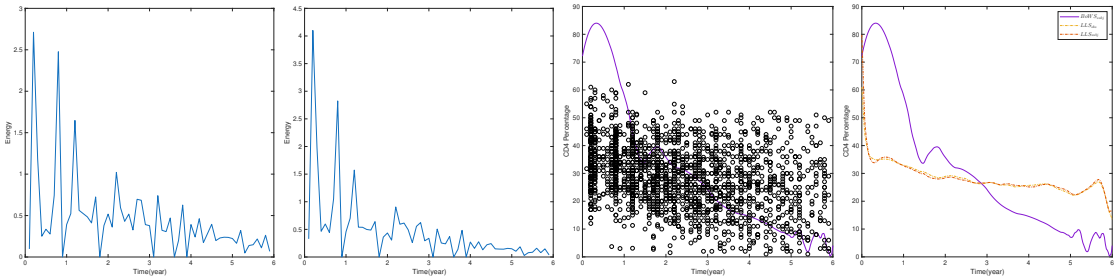
From these summaries, the average observed values appear markedly higher near the left boundary, indicating a concentration of the observed signal in that region. Compared to the LLS method, which exhibits only a gradual decrease, the proposed method shows a more localized increase near the left boundary, consistent with the observed data pattern. However, since the true mean function is unknown and the sampling distribution may be uneven across time, this visual comparison should be interpreted with caution. The estimated covariance function is shown in Fig. 6, where the covariance between $X(t)$ and $X(t')$ decreases as the time gap $|t - t'|$ increases.

IV. Conclusion

In this paper, we present an explicit construction of boundary orthogonal wavelets based on the Daubechies family, following the polynomial reproduction framework of Han and Michelle [10], [12]. Compared to the construction methods of Meyer [8] and Cohen [13], our construction yields exactly N boundary scaling functions—the minimum needed to reproduce polynomials of order N—while preserving all interior scaling functions whose supports are fully contained in the interval. This minimal boundary system is particularly well-suited to sparse functional data analysis, where parsimonious boundary adaptation can improve the stability of mean and covariance estimation. Compared to traditional statistical methods, wavelets can better capture local features such as abrupt changes, as confirmed in our simulation results.

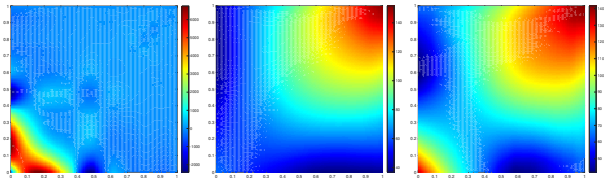
Several directions can be pursued for future work.

1) The present framework is built upon a boundary orthogonal wavelet system. While orthogonality guarantees energy conservation, the associated filters generally lack linear phase. To address this, we can consider extending the framework to boundary biorthogonal wavelet systems,



(a) E_{obs} (b) E_{subj} (c) Data set (d) Result

Fig. 5: (a) "Energy" distribution of CD4 dataset at different time points with OBS scheme; (b) "Energy" distribution of CD4 dataset at different time points with SUBJ scheme; (c) CD4 percentage data from 283 homosexual men; (d) Visualization comparison of our method (BoWS) and local linear smoother (LLS)



(a) BoWS$_{subj}$ (b) LLS$_{obs}$ (c) LLS$_{subj}$

Fig. 6: Visualization comparison of different approaches for the covariance between $X(t)$ and $X(t')$ of CD4 data. (a) our method with SUBJ scheme (BowS$_{subj}$); (b) local linear smoother with OBS scheme (LLS$_{obs}$); (c) local linear smoother with SUBJ scheme (LLS$_{subj}$).

which can simultaneously achieve symmetry and compact support, thereby permitting more natural signal extensions near the boundaries and reduced phase distortion.

2) Regarding thresholding, only soft thresholding has been adopted in this study. Although soft thresholding preserves the continuity of the shrinkage estimate, it tends to introduce bias. We will investigate alternative thresholding strategies, including hard thresholding and data-driven adaptive approaches, to identify more effective thresholding schemes under specific fidelity criteria.
3) For the two-dimensional case, bivariate boundary wavelet bases were constructed simply by the tensor product. This is simple but treats directions separately, limiting the ability to capture anisotropic features or complex geometric structures. Future work will explore non-tensor-product constructions of bivariate boundary wavelet bases, for instance, those based on lifting schemes, subdivision algorithms, or directional frameworks [14]–[17]. Such constructions are expected to yield greater sparsity and flexibility in representing complex boundary regions and multidirectional singularities, thereby enhancing the performance of wavelet methods in applications such as functional data analysis and the numerical solution of partial differential equations.

REFERENCES

[1] L. Horváth and P. Kokoszka, *Inference for Functional Data with Applications*. Springer Science & Business Media, 2012.
[2] J. R. Berrendero, A. Justel, and M. Svarc, "Principal components for multivariate functional data," *Computational Statistics & Data Analysis*, vol. 55, no. 9, pp. 2619–2634, 2011.
[3] X. Zhang and J.-L. Wang, "From sparse to dense functional data and beyond," *The Annals of Statistics*, vol. 44, pp. 2281–2321, 2016.
[4] T. T. Cai and M. Yuan, "Optimal estimation of the mean function based on discretely sampled functional data: Phase transition," *The Annals of Statistics*, vol. 39, no. 5, pp. 2330–2355, 2011.
[5] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Transactions on Information Theory*, vol. 36, no. 5, pp. 961–1005, 2002.
[6] A. Antoniadis, "Wavelet methods in statistics: Some recent developments and their applications," *Statistics Surveys*, vol. 1, pp. 16–55, 2007.
[7] I. Daubechies, *Ten Lectures on Wavelets*. SIAM, 1992.
[8] Y. Meyer, "Ondelettes sur l'intervalle," *Revista Matematica Iberoamericana*, vol. 7, no. 2, pp. 115–133, 1991.
[9] A. Cohen, I. Daubechies, and P. Vial, "Wavelets on the interval and fast wavelet transforms," *Applied and Computational Harmonic Analysis*, vol. 1, pp. 54–81, 1993.
[10] B. Han and M. Michelle, "Wavelets on intervals derived from arbitrary compactly supported biorthogonal multiwavelets," *Applied and Computational Harmonic Analysis*, vol. 53, pp. 270–331, 2021.
[11] R. A. Kaslow, D. G. Ostrow, R. Detels, J. P. Phair, B. F. Polk, C. R. Rinaldo Jr, and M. A. C. Study, "The multicenter aids cohort study: rationale, organization, and selected characteristics of the participants," *American journal of epidemiology*, vol. 126, no. 2, pp. 310–318, 1987.
[12] B. Han and M. Michelle, "Construction of wavelets and framelets on a bounded interval," *Analysis and Applications*, vol. 16, no. 06, pp. 807–849, 2018.
[13] A. Cohen, I. Daubechies, and J.-C. Feauveau, "Biorthogonal bases of compactly supported wavelets," *Communications on Pure and Applied Mathematics*, vol. 45, no. 5, pp. 485–560, 1992.
[14] B. Han, Z. Zhao, and X. Zhuang, "Directional tensor product complex tight framelets with low redundancy," *Applied and Computational Harmonic Analysis*, vol. 41, no. 2, pp. 603–637, 2016.
[15] X. Zhuang, "Digital affine shear transforms: fast realization and applications in image/video processing," *SIAM Journal on Imaging Sciences*, vol. 9, no. 3, pp. 1437–1466, 2016.
[16] Z. Che and X. Zhuang, "Digital affine shear filter banks with 2-layer structure and their applications in image processing," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 3931–3941, 2018.
[17] B. Han, Q. Mo, Z. Zhao, and X. Zhuang, "Directional compactly supported tensor product complex tight framelets with applications to image denoising and inpainting," *SIAM Journal on Imaging Sciences*, vol. 12, no. 4, pp. 1739–1771, 2019.