이학박사 학위논문

Multiscale Representation of Directional Scattered Data: Use of Anisotropic Radial Basis Functions

비등방성 방사기저함수를 이용한 산재된 방향성 자료의 다중척도 표현

2021년 8월

서울대학교 대학원

통계학과

권 준 현

Multiscale Representation of Directional Scattered Data: Use of Anisotropic Radial Basis Functions

지도교수 오희석

이 논문을 이학박사 학위논문으로 제출함

2021년 4월

서울대학교 대학원

통계학과

권 준 현

권준현의 이학박사 학위논문을 인준함

2021년 6월

위	원 장	이	재 용
부 위 원 장		오	희 석
위	원	장 원	철
위	원	임	채 영
위	원	성	병 찬

Multiscale Representation of Directional Scattered Data: Use of Anisotropic Radial Basis Functions

by

Junhyeon Kwon

A Thesis
submitted in fulfillment of the requirement
for the degree of
Doctor of Philosophy
in
Statistics

The Department of Statistics
College of Natural Sciences
Seoul National University

August, 2021

ABSTRACT

Multiscale Representation of Directional Scattered Data: Use of Anisotropic Radial Basis Functions

Junhyeon Kwon
The Department of Statistics
The Graduate School
Seoul National University

Spatial inhomogeneity along the one-dimensional curve makes two-dimensional data non-stationary. Curvelet transform, first proposed by Candes and Donoho (1999), is one of the most well-known multiscale methods to represent the directional singularity, but it has a limitation in that the data needs to be observed on equally-spaced sites. On the other hand, radial basis function interpolation is widely used to approximate the underlying function from the scattered data. However, the isotropy of the radial basis functions lowers the efficiency of the directional representation. This thesis proposes a new multiscale method that uses anisotropic radial basis functions to efficiently represent the direction from the noisy scattered data in two-dimensional Euclidean space. Basis functions are orthogonalized across the scales so that each scale can represent a global or local directional structure separately. It is shown that the proposed method is remarkable for representing directional scattered data through numerical experiments. Convergence property and practical issues in implementation are discussed as well.

Keywords: Anisotropic radial basis functions, Directional scattered data, Multiscale analysis, Nonparametric function estimation.
Student Number: 2015–20286

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Nonstationary data analysis is a big challenge in statistics because of their inhomogeneous dependence structure. Multiscale methods can help us to understand the nature of the nonstationary data by representing it in different resolutions. One can investigate the global trend from the representation in coarse resolution and identify the local activities from the newly added features in finer resolution. Wavelet analysis is one of the most popular branches of multiscale approaches, and many researchers have demonstrated the great utility of the method over the last few decades.

Wavelets are localized waveform functions that vary in their scales by dilation and contraction and vary in locations through translation shift. Collection of these wavelets at each scale span an orthogonal subspace of $L_2(\mathbb{R})$, and the summation of these subspaces becomes $L_2(\mathbb{R})$. Thus, the wavelet methods allow us to represent nonstationary phenomena such as peaks, jumps, and varying frequency of oscillation by separating the information in the data at different scales and locations. Furthermore, coefficients of the wavelet transform have a sparse structure, and we can exploit this property for the denoising and compression of the data.

When the wavelet methods are extended to the two-dimensional data,

we need to consider a new type of nonstationarity, which is present along the curve on a two-dimensional domain. We name this singularity by the term 'directional structure.' The tensor product of the one-dimensional wavelets is the simplest extension to the two-dimensional case, but it cannot efficiently capture the singularity unless it is vertical or horizontal. To overcome this defect, Candes and Donoho (1999) proposed a curvelet transform and Candes et al. (2006) improved the first version of the curvelet transform to be less redundant.

This thesis focuses on representing two-dimensional data with directional structure when observed at irregularly scattered sites. Although there are wavelet methods for the scattered data including the lifting scheme (Sweldens, 1996, 1998) and the spherical wavelets (Narcowich and Ward, 1996, Li, 1999, Oh, 1999), they also have limitations in reflecting the directional structure. On the other hand, radial basis approximation is a meshfree interpolation method to find underlying functions from the scattered data. Floater and Iske (1996) suggested a multiscale approximation framework using radial basis functions, and its convergence to the true function has been proved by Wendland (2010). However, circular supports of the radial functions are not adequate to reflect the directional structure. Thus, this research suggests using anisotropic basis functions with elliptic supports for multiscale approximation and proves its converging property. Furthermore, the proposed method can be coupled with the shrinkage method as in ordinary wavelet transform so that we can remove the noise when the data is contaminated by noise.

This thesis is organized as follows. Chapter 2 introduces basic concepts of wavelet transform and its extension to the directional data and scattered data separately. Chapter 3 briefly introduces a theoretical background of the radial basis approximation that will be the cornerstone of the proposed method. In Chapter 4, we propose to adopt the anisotropic basis functions

in a multiscale radial basis approximation scheme. We also discuss its theoretical properties and practical issues in implementation. Chapter 5 tests the applicability of the proposed method by applying it to the simulation datasets and temperature data in South Korea. Finally, Chapter 6 summarizes and emphasizes the contributions of this thesis again and discusses a few topics of future research.

Chapter 2

Multiscale Analysis

This chapter reviews the basic concepts of multiscale analysis, focusing mainly on the wavelet transform. Wavelets are basis functions that can decompose $L_2(\mathbb{R})$ into multiple orthogonal levels. Although wavelets are efficient tools for representing nonstationary data in virtue of their localized property, they have limitations in representing the nonstationarity of the two-dimensional data, such as singularity along with the directional structure. Furthermore, most wavelet methods require the data to be observed at equally-spaced grids and even dyadic.

This chapter is organized as follows. Section 2.1 introduces the classical wavelet transform of the regularly spaced data both for one-dimensional and two-dimensional data. Section 2.2 is about two-dimensional wavelet transforms that overcome the limitation of the classical ones by reflecting the directional structures. Section 2.3 discusses the wavelet methods for the scattered data.

2.1 Classical wavelet transform

The term 'wavelet' was first introduced in the domain of geophysics by Morlet et al. (1982). It is a compound of the noun 'wave' and the postfix '-let' which means 'little', so the wavelets are little waves in that they decay rapidly to zero. A wavelet $\psi \in L_2(\mathbb{R})$ is a function which has zero average,

$$\int_{-\infty}^{\infty} \psi(x)dx = 0,$$

is normalized,

$$\int_{-\infty}^{\infty} |\psi(x)|^2 dx = 1,$$

and meets the admissibility condition,

$$\int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < \infty,$$

where $\hat{\psi}$ is a Fourier transform of ψ. A wavelet can be dilated and translated to form a basis in $L_2(\mathbb{R})$. If it is dilated with a scale s and translated by u, we denote it by

$$\psi_{s,u}(x) = \frac{1}{\sqrt{s}} \psi(\frac{x-u}{s}).$$

2.1.1 Continuous wavelet transform

Continuous wavelet transform of the function f at a scale s and a location u is

$$Wf(s,u) = \int_{-\infty}^{\infty} f(x)\psi^*_{s,u}(x)dx.$$

Under the admissibility condition, we can reconstruct the given function $f(x)$ by

$$f(x) = \left[\int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega\right]^{-1} \int_0^{\infty} \int_{-\infty}^{\infty} s^{-5/2} Wf(s,u)\psi_{s,u}(x)duds. \quad (2.1)$$

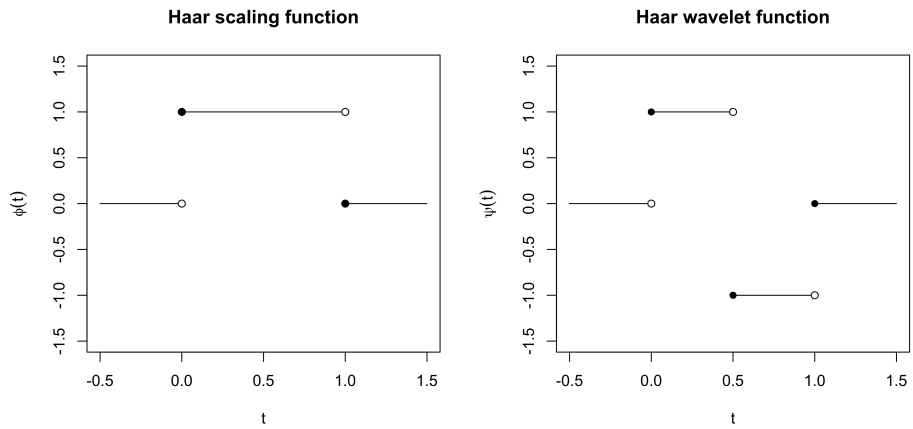

Figure 2.1: Haar scaling function and wavelet function

Continuous Haar wavelet transform

Haar wavelet function is one of the most well-known and intuitive wavelets. Figure 2.1 depicts the Haar scaling and wavelet functions, and their square shape helps us appreciate the meaning of wavelet transform. Haar scaling function is defined as

$$\phi(x) = \begin{cases} 1, & \text{if } x \in [0,1), \\ 0, & \text{otherwise,} \end{cases}$$

and Haar wavelet function is defined as

$$\psi(x) = \begin{cases} 1, & \text{if } x \in [0,1/2), \\ -1, & \text{if } x \in [1/2,1), \\ 0, & \text{otherwise.} \end{cases}$$

Thus, the continuous Haar wavelet transform at scale s and location u, $Wf(s,u)$, gives us the difference between the average values of $f(x)$ on $[u, u+s/2)$ and $[u+s/2, u+s)$. In other words, we can extract the information about the variation of f at any scale $s \in (0,\infty)$ and any location $u \in \mathbb{R}$.

2.1.2 Multiresolution analysis

The concept of multiresolution analysis (MRA) was first suggested by Mallat (1989) for the construction of wavelets and the approximation of a function in $L_2(\mathbb{R})$. MRA projects a function $f \in L_2(\mathbb{R})$ orthogonally on a sequence of approximation spaces $\{\mathbb{V}_\ell\}_{\ell\in\mathbb{Z}}$ that differs in resolution. Hence, the approximation space $\mathbb{V}_\ell$ contains all the approximants at the ℓ-th resolution, and we obtain the approximant of f in $\mathbb{V}_\ell$ by minimizing $\|f - f_\ell\|$ for all $f_\ell \in \mathbb{V}_\ell$. Below is the formal definition of multiresolution approximation from Mallat (1999).

Definition 2.1.1. *A sequence of closed subspaces $\{\mathbb{V}_\ell\}_{\ell\in\mathbb{Z}}$ of $L_2(\mathbb{R})$ is a multiresolution approximation if the following properties holds:*

(1) Nestedness

$$\mathbb{V}_\ell \subset \mathbb{V}_{\ell+1},\ \forall \ell \in \mathbb{Z}$$

(2) Completeness

$$\overline{\bigcup_{\ell=-\infty}^{\infty} \mathbb{V}_\ell} = L_2(\mathbb{R}) \text{ and } \bigcap_{\ell=-\infty}^{\infty} \mathbb{V}_\ell = \{0\}$$

(3) Dilation invariance

$$f(x) \in \mathbb{V}_\ell \Leftrightarrow f(2x) \in \mathbb{V}_{\ell+1},\ \forall \ell \in \mathbb{Z}$$

(4) Translation invariance

$$f(x) \in \mathbb{V}_\ell \Leftrightarrow f(x - 2^{-\ell}k) \in \mathbb{V}_\ell,\ \forall (\ell, k) \in \mathbb{Z}^2$$

(5) Existence of Riesz basis

There exists θ such that $\{\theta(x-k)\}_{k\in\mathbb{Z}}$ is a Riesz basis of $\mathbb{V}_0$.

To give a brief explanation for each property, nestedness means that more functions will be contained in the approximation space $\mathbb{V}_\ell$ as ℓ gets larger. Completeness implies that the function approximation converges to the original function as ℓ goes to the positive infinity in the nested relation, and to 0 as ℓ goes to the negative infinity. Invariance in dilation ensures that dilating a function by 2 leads to an approximation at a one-level coarser scale, and invariance in translation of implies that $\mathbb{V}_\ell$ is invariant by any translation proportional to $2^{-\ell}$. Riesz basis is a relaxed concept of orthogonal basis. We can orthogonalize and normalize it to construct the orthonormal basis $\{\phi(x-k)\}_{k\in\mathbb{Z}}$ of $\mathbb{V}_0$.

Let us abuse the notation and denote the function $\phi(t)$ dilated at scale $s = 2^{-\ell}$ and translated by $u = 2^{-\ell}k$ with

$$\phi_{\ell,k}(t) = 2^{\ell/2}\phi(2^\ell t - k),\ \ell, k \in \mathbb{Z}.$$

Based on this expression, we will sometimes use the term 'level' for the integer ℓ. Then, from the properties above, we observe that $\{\phi_{\ell,k}(x)\}_{k\in\mathbb{Z}}$ becomes an orthonormal basis of $\mathbb{V}_\ell$ through the dilation at level ℓ, or scale $2^{-\ell}$. Thus the orthonormal basis $\{\phi(x-k)\}_{k\in\mathbb{Z}}$ satisfies the two-scale relation

$$\phi(x) = \sum_{k=-\infty}^{\infty} h_k \phi_{1,k}(x) \tag{2.2}$$

because $\phi(x) \in \mathbb{V}_0 \subset \mathbb{V}_1$, and $\{\phi_{1,k}(x)\}_{k\in\mathbb{Z}}$ is a basis of $\mathbb{V}_1$.

On the other hand, let us consider an orthogonal complement $\mathbb{W}_\ell = \mathbb{V}_{\ell+1} \cap \mathbb{V}_\ell^\perp$ for $\ell \in \mathbb{Z}$ and call it a detail space at the ℓ-th level. From the nested structure of the approximation space $\mathbb{V}_\ell$, we have

$$\mathbb{V}_L = \mathbb{V}_{-L} \oplus \bigcup_{\ell=-L}^{L-1} \mathbb{W}_\ell.$$

Then, from the completeness condition of MRA, we have

$$\bigoplus_{\ell\in\mathbb{Z}} \mathbb{W}_\ell = L_2(\mathbb{R}).$$

Since we can find an orthonormal basis of the detail space $\mathbb{W}_\ell$, and the collection of these basis functions over the entire levels becomes an orthonormal basis of $L_2(\mathbb{R})$. These are called wavelets, and Daubechies (1992) presented the existence and construction of them. If $\{\psi(x-k)\}_{k\in\mathbb{Z}}$ is a basis of $\mathbb{W}_0$, then it also satisfies the two-scale relation

$$\psi(x) = \sum_{k=-\infty}^{\infty} g_k \phi_{1,k}(x),\ g_k = (-1)^{k-1} h_{1-k} \tag{2.3}$$

because $\psi(x) \in \mathbb{W}_0 \subset \mathbb{V}_1$, and $\{\phi_{1,k}(x)\}_{k\in\mathbb{Z}}$ is a basis of $\mathbb{V}_1$.

Let $\psi_{\ell,k}(x)$ denote a wavelet function $\psi(t)$ dilated at scale $s = 2^{-\ell}$ and translated by $u = 2^{-\ell}k$. Then $\{\psi_{\ell,k}\}_{k\in\mathbb{Z}}$ is an orthonormal basis of $\mathbb{W}_\ell$ while $\{\phi_{\ell',k}\}_{k\in\mathbb{Z}}$ is an orthonormal basis of $\mathbb{V}_{\ell'}$. Thus, for any $f \in L_2(\mathbb{R}^2)$, we have

$$f(x) = \sum_{\ell=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} d_{\ell,k} \psi_{\ell,k}(x), \tag{2.4}$$

and

$$P_\ell f(x) = \sum_{k=-\infty}^{\infty} c_{\ell,k} \phi_{\ell,k}(x),$$

where $d_{\ell,k} = \langle f, \psi_{\ell,k} \rangle$, $c_{\ell,k} = \langle f, \phi_{\ell,k} \rangle$, and P_ℓ is an operator that projects on $\mathbb{V}_\ell$. As the square-integrable space is decomposed as

$$L_2(\mathbb{R}) = \mathbb{V}_{\ell_0} \oplus \bigcup_{\ell=\ell_0}^{\infty} \mathbb{W}_\ell,$$

we can express any $f \in L_2(\mathbb{R})$ as

$$f(x) = \sum_{k=-\infty}^{\infty} c_{\ell_0,k} \phi_{\ell_0,k}(x) + \sum_{\ell=\ell_0}^{\infty} \sum_{k=-\infty}^{\infty} d_{\ell,k} \psi_{\ell,k}(x). \tag{2.5}$$

The first term of (2.5) can be thought of as a coarse, but overall, approximation of the function f at the ℓ_0-th level, and the second term is sequentially added detail of the function f from the level ℓ_0 to infinity.

2.1.3 Discrete wavelet transform

Frame

The frame is a concept suggested by Duffin and Schaeffer (1952) for the reconstruction of band-limited function, and it ensures a complete and stable representation of a function f in Hilbert space. If the wavelet $\psi(x)$ is a frame, then we can represent a function $f \in L_2(\mathbb{R})$ through the summation form in (2.4). This implies that the integral form (2.1) is redundant compared to (2.4). The formal definition of the frame is given below.

Definition 2.1.2. *Let* $\mathbf{H}$ *be a Hilbert space and* $\{\theta_k\}_{k\in\mathbb{Z}}$ *be a set of vectors in* $\mathbf{H}$*. Then* $\{\theta_k\}_{k\in\mathbb{Z}}$ *is a frame if there exist* $A, B \in (0,\infty)$ *such that for any* $f \in \mathbf{H}$

$$A\|f\|^2 \leq \sum_k |\langle f, \theta_k\rangle|^2 \leq B\|f\|^2.$$

We say that the frame $\{\theta_k\}_{k\in\mathbb{Z}}$ *is tight if* $A = B$ *holds.*

We can further remove the redundancy to the extent that the basis is said to be exactly enough if the frame becomes linearly independent. This frame is called Riesz basis, and we give the formal definition of it below.

Definition 2.1.3. *A sequence of linearly independent vectors* $\{\theta_k\}_{k\in\mathbb{Z}}$ *is a Riesz basis of a Hilbert space* $\mathbf{H}$ *if there exists* $A, B \in (0,\infty)$ *such that any* $f \in \mathbf{H}$ *is represented by*

$$f(x) = \sum_{k=-\infty}^{\infty} a_k \theta_k(x),$$

where $\{a_k\}_{k\in\mathbb{Z}}$ *satisfies*

$$A\|f\|^2 \leq \sum_{k=-\infty}^{\infty} |a_k|^2 \leq B\|f\|^2.$$

Forward and inverse transform

Coefficients at consecutive levels also have two-scale relations

$$c_{0,0} = \sum_{k=-\infty}^{\infty} h_k c_{1,k},$$

and

$$d_{0,0} = \sum_{k=-\infty}^{\infty} g_k c_{1,k},$$

which are derived from the two-scale relations of the orthonormal bases $\{\phi_{\ell,k}\}_{k\in\mathbb{Z}}$ and $\{\psi_{\ell,k}\}_{k\in\mathbb{Z}}$. Thus, if we can extend the two-scale relation of the coefficients to general levels and translations, one only needs the scaling coefficients at the finest level, and the filters $\{h_k\}_{k\in\mathbb{Z}}$ and $\{g_k\}_{k\in\mathbb{Z}}$ for the (forward) discrete wavelet transform.

Since $c_{\ell,k} = \langle f, \phi_{\ell,k}\rangle$ and $c_{\ell-1,k} = \langle f, \phi_{\ell-1,k}\rangle$, we first need to relate $\{\phi_{\ell,k}\}_{k\in\mathbb{Z}}$ and $\{\phi_{\ell-1,k}\}_{k\in\mathbb{Z}}$. From the nested structure of the approximation spaces $\{\mathbb{V}\}_{\ell\in\mathbb{Z}}$ and the two-scale relation of $\{\phi_{\ell,k}\}$, we obtain

$$\begin{aligned}
\phi_{\ell,k}(x) &= 2^{\ell/2}\phi(2^\ell x - k) \\
&= 2^{\ell/2}\sum_{n\in\mathbb{Z}} h_n \phi_{1,n}(2^\ell x - k) \\
&= 2^{(\ell+1)/2}\sum_{n\in\mathbb{Z}} h_n \phi(2^{\ell+1} x - 2k - n) \\
&= \sum_{n\in\mathbb{Z}} h_n \phi_{\ell+1,n+2k}(x),
\end{aligned}$$

and this leads to

$$c_{\ell,k} = \sum_{n\in\mathbb{Z}} h_{n-2k} c_{\ell+1,n}. \tag{2.6}$$

In the same way, we can derive the two-scale relation of the detail coefficients, which is

$$d_{\ell,k} = \sum_{n\in\mathbb{Z}} g_{n-2k} c_{\ell+1,n}. \tag{2.7}$$

Mallat (2009) showed that the relation that helps us to calculate the finer scaling coefficients from the one-level coarser coefficients. It is used for the inverse discrete wavelet transform and expressed as

$$c_{\ell,k} = \sum_n h_{k-2n} c_{\ell-1,n} + \sum_n g_{k-2n} d_{\ell-1,n},$$

with the same filters in (2.6) and (2.7).

Discrete Haar wavelet transform

Suppose that we want to approximate $f \in L_2(\mathbb{R})$ in the form of (2.5). But, if we limit the details to be added up to the $(L-1)$-th level, then the approximant $f_L \in \mathbb{V}_L$ will have the form

$$f_L(x) = \sum_{k=-\infty}^{\infty} c_{\ell_0,k}\phi_{\ell_0,k}(x) + \sum_{\ell=\ell_0}^{L-1}\sum_{k=-\infty}^{\infty} d_{\ell,k}\psi_{\ell,k}(x).$$

We can obtain the coefficients of this discrete wavelet transform from the scaling coefficients at the finest level. However, calculating $\{c_{L,k}\}_k$ is not simple unless wavelet transform is based on the Haar basis functions.

Let $\{x_k\}_{k=1}^N$, $N = 2^L$, be equally-spaced data locations and $y_k = f(x_k)$ be an observation on each data site. Then we can regard $\{y_k\}_{k=1}^N$ as the finest level (the L-th level) approximation with the Haar scaling function.

Note that the Haar scaling function $\phi(x)$ satisfies the two-scale relation

$$\phi(x) = \phi(2x) + \phi(2x-1) = \frac{1}{\sqrt{2}}\left(\phi_{1,0}(x) + \phi_{1,1}(x)\right)$$

and the Haar wavelet function $\psi(x)$ satisfies

$$\psi(x) = \phi(2x) - \phi(2x-1) = \frac{1}{\sqrt{2}}\left(\phi_{1,0}(x) - \phi_{1,1}(x)\right).$$

Hence, the filter $\{h_k\}_{k\in\mathbb{Z}}$ in (2.6) satisfies

$$h_k = \begin{cases} 1/\sqrt{2}, & \text{for } k = 0, 1 \\ 0, & \text{otherwise,} \end{cases}$$

and the filter $\{g_k\}_{k\in\mathbb{Z}}$ in (2.7) satisfies

$$g_k = \begin{cases} 1/\sqrt{2}, & \text{for } k = 0, \\ -1/\sqrt{2}, & \text{for } k = 1, \\ 0, & \text{otherwise.} \end{cases}$$

Thus, we obtain the coefficients at the coarser level $\ell = L-2, \cdots, 0$ by

$$c_{\ell,k} = \frac{1}{\sqrt{2}}(c_{\ell+1,2k} + c_{\ell+1,2k-1}),$$

and

$$d_{\ell,k} = \frac{1}{\sqrt{2}}(c_{\ell+1,2k} - c_{\ell+1,2k-1}),$$

for $k = 1, 2, \cdots, 2^\ell$. Original observations $\{y_k\}_{k=1}^N$ can be reconstructed by the inversion relations

$$c_{\ell+1,2k} = (c_{\ell,k} + d_{\ell,k})/\sqrt{2},$$

and

$$c_{\ell+1,2k-1} = (c_{\ell,k} - d_{\ell,k})/\sqrt{2}.$$

2.1.4 Two-dimensional wavelet transform

Note that a tensor product of the bases, $\{\theta_k^{(1)} \otimes \theta_{k'}^{(2)}\}_{(k,k')\in\mathbb{Z}^2}$, becomes a Riesz basis of a tensor product space $\mathbf{H} = \mathbf{H}_1 \otimes \mathbf{H}_2$ if $\{\theta_k^{(1)}\}_{k\in\mathbb{Z}}$ and $\{\theta_k^{(2)}\}_{k\in\mathbb{Z}}$ are Riesz bases of the Hilbert spaces $\mathbf{H}_1$ and $\mathbf{H}_2$, respectively. Based on this fact, we can define a multiresolution approximation $\{\mathbb{V}_\ell^2\}_{\ell\in\mathbb{Z}}$ of $L_2(\mathbb{R}^2)$ as in the one-dimensional case and find an orthonormal basis $\{\phi_{\ell,k}(x_1, x_2)\}_{(\ell,k)\in\mathbb{Z}^3}$ of a tensor product space $\mathbb{V}_\ell^2$, which is

$$\phi_{\ell,k}(x_1, x_2) = \phi_{\ell,k_1}(x_1) \otimes \phi_{\ell,k_2}(x_2).$$

Here, $\{\phi_{\ell,k}\}_{k\in\mathbb{Z}}$ is an orthonormal basis of $\mathbb{V}_\ell^2$, and we say that it is separable.

On the other hand, let $\mathbb{W}_\ell^2$ be the detail space which is an orthogonal complement

$$\mathbb{W}_\ell^2 = \mathbb{V}_{\ell+1}^2 \cap \mathbb{V}_\ell^{2\perp}.$$

Then the next theorem from Mallat (1999) shows how we can construct a separable wavelet basis for $\mathbb{W}_\ell^2$.

Theorem 2.1.1. *Let $\{\phi_{\ell,k}\}_{(\ell,k)\in\mathbb{Z}^2}$ and $\{\psi_{\ell,k}\}_{(\ell,k)\in\mathbb{Z}^2}$ be orthonormal bases of the approximation space $\mathbb{V}_\ell$ and the detail space $\mathbb{W}_\ell$ of $L_2(\mathbb{R})$, respectively. Define three types of wavelets*

$$\psi^{(1)} = \phi \otimes \psi,\ \psi^{(2)} = \psi \otimes \phi,\ \psi^{(3)} = \psi \otimes \psi.$$

Then $\{\psi_{\ell,k}^{(1)}, \psi_{\ell,k}^{(2)}, \psi_{\ell,k}^{(3)}\}_{k\in\mathbb{Z}^2}$ is an orthonormal basis of $\mathbb{W}_\ell^2$, and $\{\psi_{\ell,k}^{(1)}, \psi_{\ell,k}^{(2)}, \psi_{\ell,k}^{(3)}\}_{(\ell,k)\in\mathbb{Z}^3}$ is an orthonormal basis of $L_2(\mathbb{R}^2)$.

2.2 Wavelets for equally-spaced directional data

Data observed on a two-dimensional domain can have a relatively complicated type of singularity compared to the data on a one-dimensional domain. In one-dimensional space, the classical wavelet transform can detect and represent pointwise jump discontinuity of the underlying function effectively. However, in two-dimensional space, discontinuity can occur along a line or a curve on the domain, and the simple tensor product of one-dimensional classical wavelets is not capable of representing this in a sparse manner. We use the term 'directional structure' to refer to the discontinuity along a curve, and we call the data with directional structure 'directional data.'

2.2.1 Ridgelets

Ridgelets

Ridgelet transform was proposed in Candes (1998) to detect the discontinuity along a line on a two-dimensional domain. Ridgelet transform of a function f is defined as

$$C_f(a,b,\theta) = \int\int f(x_1,x_2)\psi_{a,b,\theta}(x_1,x_2)dx_1dx_2,$$

where

$$\psi_{a,b,\theta}(x_1,x_2) = a^{-1/2}\psi((x_1\cos\theta + x_2\sin\theta - b)/a)$$

is a ridgelet function at scale a which has constant value along the line parallel to

$$x_1\cos\theta + x_2\sin\theta = b.$$

Multiscale ridgelets

Orthonormal ridgelets were proposed by Donoho (2000). Let these ridgelets have length 1 and arbitrarily finer width, and partition the domain smoothly into 2^{2j} squares with side lengths 2^{-j}. Based on this, we will define the multiscale ridgelets by windowing, transporting, and normalizing the ridgelets for each square and each scale. Then the resulting ridgelets can have arbitrary values for the lengths as well.

However, multiscale ridgelets are not orthogonal to each other if these are derived from overlapping partitioning squares from different scales. In other words, multiscale ridgelets are too massive to be frameable and suffer from 'energy blow-up' problem

$$\sum_{\mu\in\mathcal{M}} \langle\psi_\mu, f\rangle^2 = \infty,$$

where $\mathcal{M}$ is a dictionary of multiscale ridgelets.

2.2.2 Curvelets

Curvelet transform was first proposed in Candes and Donoho (1999). It is a combination of multiscale ridgelets and bandpass filtering to represent the discontinuity along a curve on a two-dimensional domain. Let Δ_j be a subband filter which extracts the frequency $|\omega| \in [2^j, 2^{j+1}]$ and P_0 be a filter for the frequency $|\omega| \leq 1$. Then we have an equation

$$\|f\|_2^2 = \|P_0 f\|_2^2 + \sum_j \|\Delta_j f\|_2^2.$$

Candes and Donoho (1999) overcame the 'energy blow-up' problem by decomposing f into multiple subbands and applying monoscale ridgelet dictionary for each subband. We have to note that, among these curvelets, only the ones with approximate length $2^{-j/2}$ and width 2^{-j} have non-negligible norms. Thus, we observe that the width of a curvelet is approximately the square of the length.

Candes et al. (2006) suggested second-generation curvelets by further developing the original ones to be faster and less redundant. We will consider windows in the frequency domain based on a polar coordinate system. Let $W(r)$ be a radial window supported on $r \in (1/2, 2)$ and $V(t)$ be an angular window supported on $t \in [-1, 1]$, and assume that these windows meet some certain admissibility conditions. Then, for $j \geq j_0$, we can define a wedge-shaped frequency window

$$U_j(r, \theta) = 2^{-3j/4} W(2^{-j} r) V(\lfloor j/2 \rfloor \theta / \pi). \tag{2.8}$$

We rotate this window by $\theta_\ell, \ell = 0, 1, \cdots$, for each level j to make a tiling of two-dimensional frequency domain based on concentric circles. Figure 2.2 is an illustrative example of this concept.

Waveform of the curvelet $\psi_j(x_1, x_2)$ is obtained by the rotation and translation of the inverse Fourier transform of U_j. Let the rotation angles be denoted by a sequence $\theta_\ell = 2\pi \cdot 2^{-\lfloor j/2 \rfloor} \ell \in [0, 2\pi)$ for $\ell = 0, 1, \cdots$, and the

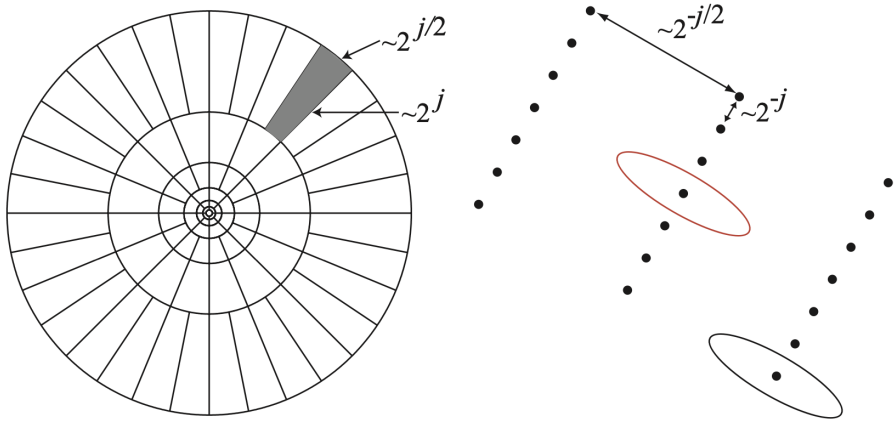

Figure 2.2: Illustrative example of tiling in the frequency domain. The shaded wedge on the left plot is at level j and makes the corresponding waveforms in the original Cartesian plane to have approximate supports with length $2^{-j/2}$ and width 2^{-j} as in the right plot. This figure is originated from Candes et al. (2006).

translation parameters be denoted by $k = (k_1, k_2) \in \mathbb{Z}^2$. Then the curvelet function at scale 2^{-j}, angle θ_ℓ, and position $x_k^{(j,\ell)} = R_{\theta_\ell}^{-1}(2^{-j}k_1, 2^{-j/2}k_2)$ is defined as

$$\psi_{j,\ell,k}(x) = \psi_j(R_{\theta_\ell}(x - x_k^{(j,\ell)})),$$

where R_{θ_ℓ} is a matrix which rotates a vector by θ_ℓ. Thus, we can obtain curvelet coefficients by

$$\begin{aligned} c(j,\ell,k) &:= \langle f, \psi_{j,\ell,k} \rangle = \int_{\mathbb{R}^2} f(x)\overline{\psi_{j,\ell,k}(x)}dx \\ &= (2\pi)^{-2} \int_{\mathbb{R}^2} \hat{f}(\omega)\overline{\hat{\psi}_{j,\ell,k}(\omega)}d\omega \\ &= (2\pi)^{-2} \int_{\mathbb{R}^2} \hat{f}(\omega) U_j(R_{\theta_\ell}\omega) \exp(i\omega^T x_k^{(j,\ell)})d\omega. \end{aligned}$$

On the other hand, tiling of the frequency domain in (2.8) uses concentric circles, and it is difficult to directly use it to compute the curvelet co-

efficients from discrete Cartesian observations. Hence, Candes et al. (2006) uses concentric squares instead to make Cartesian coronae as in Figure 2.3. Based on this tiling, Candes et al. (2006) proposed two different implementations of discrete curvelet transform. One is based on unequally-spaced fast Fourier transform (USFFT), and it follows the original curvelet transform faithfully by tilting a rectangular grid for each scale and angle. The other one is based on wrapping a data into a rectangular grid, making the implementation easier.

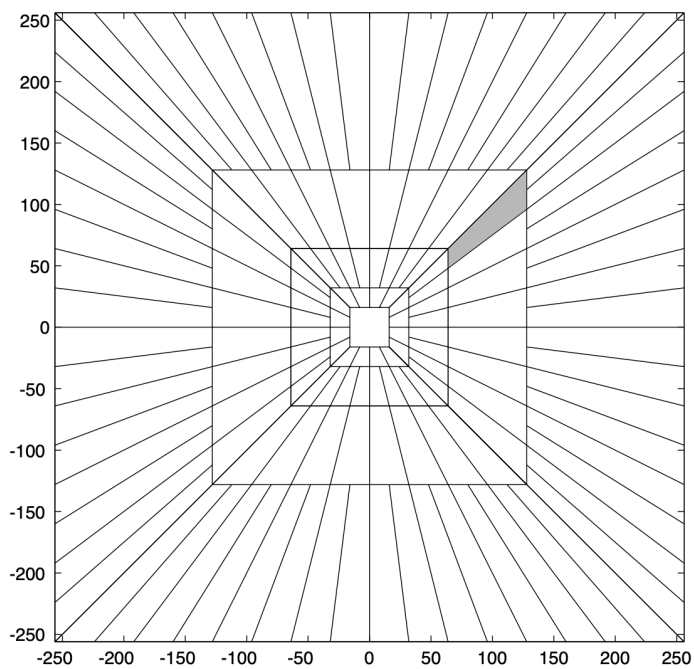

Figure 2.3: A modified version of frequency domain tiling for discrete Cartesian observations. This figure is originated from Candes et al. (2006).

2.3 Wavelets for scattered data

Many wavelet methods assume that the data is regularly observed, and one can consider two simple cases of extension when the data is observed on irregularly scattered locations. One is applying classical wavelets on the scattered data as if it is observed on an equally-spaced grid. The other is constructing basis functions on a regular grid and evaluating them on each observation site. However, Jansen and Oonincx (2005) pointed out that both of these approaches may not work well as they are for the equally-spaced data.

Let the observation sites $\{x_j : 1, 2, \cdots, N\} \subset [0, 1]$ be expressed as

$$x_j = x(u_j),\ j = 1, 2, \cdots, N,$$

for a regular grid $u_j = j/N$ and increasing function $x(u)$. If we consider a heavy sine function

$$f(x) = 4\sin(4\pi x) - \text{sign}(x - 0.3) - \text{sign}(0.72 - x)$$

observed on these sites with random measurement errors, then applying the wavelet shrinkage method on this data is the same as applying it on $g(u) = f(x(u))$. Jansen and Oonincx (2005) compared the wavelet shrinkage results both for equally-spaced data and scattered data. We reproduce this simulation for 512 observations on an interval $[0, 1]$ when the errors are independently and identically from $N(0, 0.5^2)$. The wavelet coefficients are from the Daubechies least-asymmetric phase wavelet with 10 vanishing moments, and we have replaced the ones below universal threshold with zeroes. Figure 2.4 shows the plots of heavy sine function and wavelet shrinkage results. A plot in the middle is from the data which is equally-spaced, and a plot in the bottom is from the data which is scattered while being regarded as it is not. Based on Figure 2.4, we observe that regarding the scattered data as equally-spaced one results in unintended wiggly

output, and this is because the basis function $\phi(x)$ is not as smooth as it is supposed to be after it is remapped from $u \mapsto x(u)$.

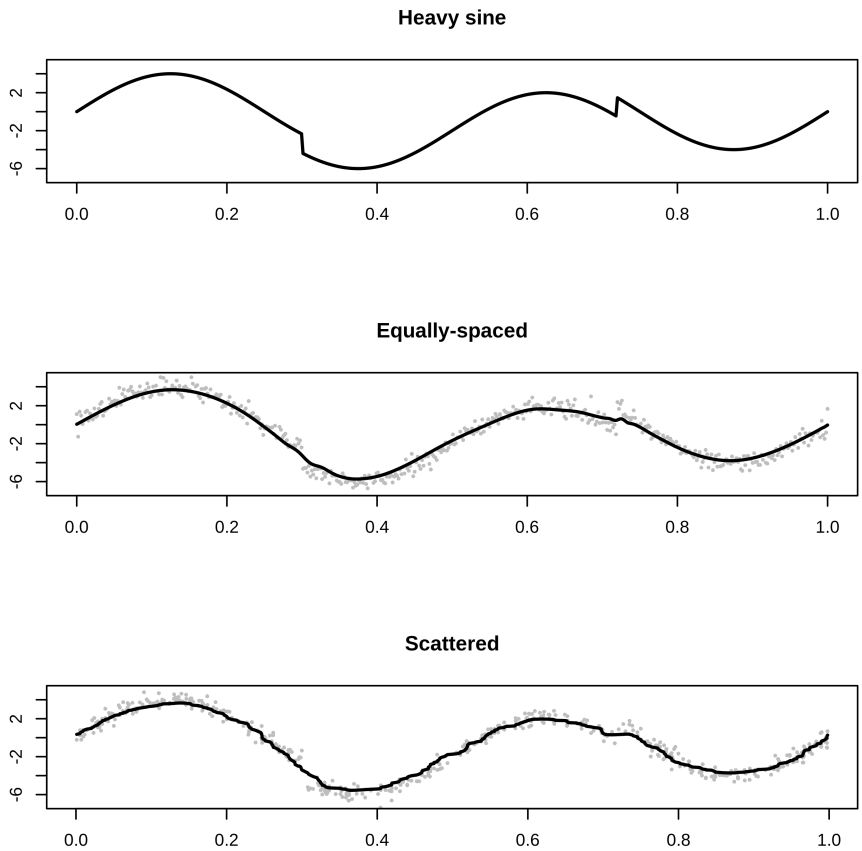

Figure 2.4: Top: Plot of the heavy sine function, Middle: Equally-spaced data and wavelet shrinkage result, Bottom: Irregularly scattered data and wavelet shrinkage result

On the other hand, according to Jansen and Oonincx (2005), evaluating equally-spaced basis functions on the scattered sites can affect the independence (or dependence) of the data and make it impossible to use the fast algorithms for the discrete wavelet transform.

2.3.1 Lifting scheme

Lifting scheme, or second-generation wavelets, was first introduced by Sweldens (1996). It is capable of multiscale representation of scattered data and does not require the data to be dyadic. It is a generalization of classical (first generation) wavelets which do not require dilation and translation of wavelet functions, and it preserves powerful properties of the first generation. Sweldens (1998) summarized those preserved properties in four aspects.

1. The wavelets are Riesz basis of $L_2(\mathbb{R})$, and unconditional basis of Lebesgue, Lipschitz, Sobolev, and Besov spaces.

2. The wavelets are (bi-)orthogonal, and this leads to explicit information about the synthesizing wavelet $\tilde{\psi}(x)$.

3. The wavelets have localized property both in the time and frequency domain.

4. The wavelets are constructed by the multiresolution analysis framework, which makes a fast algorithm feasible for the corresponding transform.

On the other hand, Sweldens (1998) listed three desirable properties that the second generation wavelets should possess through the generalization.

1. More generalized domain (e.g., curves, surfaces, or manifolds) where the wavelets are defined on.

2. Generalized basis adapted to weighted measures.

3. Applicability to the data with irregularly scattered observation sites.

Sweldens (1998) observed that translation and dilation of the wavelet functions are not essential for the second generation wavelets, which preserves four powerful characteristics of the first generation and possesses three new desirable properties. So, the Fourier transform cannot be used, and this leads to the need for another approach for the construction of the wavelets.

The term 'lifting scheme' is used as a framework for the construction of second-generation wavelets or as a second-generation wavelet transform itself. It achieves a sparse representation of the data by removing the redundancy through the prediction of one half by the other half. There are three steps in the lifting scheme: split, predict (dual lifting), and update (primal lifting). Figure 2.5 shows a diagram of these three steps. In the split step, we split the input values $\{s_{\ell+1,k}\}_{k\in\mathbb{Z}}$ at level $\ell+1$ based on whether the index k is odd or even. In the prediction step, each input with an odd index is predicted by using the nearby input with even indexes. Then we calculate how much the odd-indexed observations are apart from the predicted values and regard these differences as details at level ℓ. Finally, we update the even-indexed observations to reflect large-scale features. In this step, we use detail coefficients $d_{\ell,k}$'s to make coarser level scaling coefficients $s_{\ell,k}$'s reflect the local averages. The inverse transform is straightforward, and it is done by undoing the steps in the lifting scheme in reversed order.

When the data is observed on two-dimensional space or other generalized domains, it becomes difficult to split the observations based on their indexes to determine which are to be predicted and which are used for prediction. Thus, we usually need to find neighbors of an observation that needs to be predicted, and we use an approach based on triangulation when the data is scattered on its domain. We do not cover the splitting stage of the two-dimensional lifting scheme in detail in this thesis, and you can refer to Jansen and Oonincx (2005) if needed.

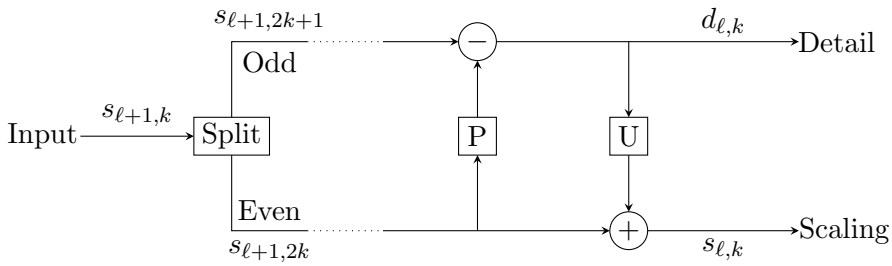

Figure 2.5: Diagram of lifting scheme procedure. 'P' stands for prediction, and 'U' stands for update.

2.3.2 Spherical wavelets

There are wavelets developed for the data observed on a sphere, which we call spherical wavelets. These are based on the construction of basis functions that considers the geometrical characteristics of its domain. Schröder and Sweldens (1995) suggested a spherical wavelet using the approach of lifting scheme. It uses triangulation-based construction, which results in non-smooth representation. On the other hand, Narcowich and Ward (1996) proposed spherical wavelets based on spherical basis functions of the Legendre series

$$G(n_1, n_2) = G(n_1 \cdot n_2) = \sum_{n=0}^{\infty} g_n P_n(n_1 \cdot n_2),$$

where P_n is a Legendre polynomial of degree n such that

$$\int_{-1}^{1} P_n^2(x)dx = 1,$$

and $n_1 = (\cos\phi_1 \cos\theta_1, \cos\phi_1 \sin\theta_1, \sin\phi_1)$ is a vector on a unit sphere for latitude θ_1 and longitude ϕ_1. Although both of these two methods can represent scattered data on a sphere, the method proposed by Narcowich and Ward (1996) has advantages in theoretical aspects and smooth representation of the data.

Li (1999) further improved this method by determining the nested network of observations

$$\mathcal{N}_1 \subset \mathcal{N}_2 \subset \cdots \subset \mathcal{N}_L,$$

where $\mathcal{N}_\ell = \bigcup_{k=1}^{\ell} \{n_{k,j}\}_{j=1}^{N_k}$, based on geodesic distance through bottom-up design procedure, and adjusted the bandwidth of each spherical basis function based on the level it belongs to. Since the bandwidth of spherical basis function changes as the level (or density of observations) increases, we find that the level has spatial interpretation. Initialization of spherical basis function representation

$$f_L(n) = \sum_{k=1}^{L} \sum_{j=1}^{N_k} \beta_{k,j} G_k(n \cdot n_{k,j}),$$

and implementation of the bottom-up design procedure are discussed in detail at Li (2001). On the other hand, spherical field at level ℓ is from the spanned space

$$\mathbb{V}_\ell = \text{span}\left\{G_k(n \cdot n_{k,j}) : j = 1, \cdots, N_k;\ k = 1, \cdots, \ell\right\},$$

and this space also has a nested structure

$$\mathbb{V}_1 \subset \mathbb{V}_2 \subset \cdots \subset \mathbb{V}_L.$$

Let us define an inner product between the spherical fields $U(n)$ and $V(n)$ as

$$\langle U(n), V(n) \rangle = \int U(n) V(n) d\Omega(n)$$

for $\int d\Omega(n) = 1$. Then, for each level ℓ, we can obtain an orthogonal decomposition

$$\mathbb{V}_\ell = \mathbb{V}_{\ell-1} \oplus \mathbb{W}_{\ell-1},$$

where $\mathbb{W}_\ell$ is a detail space at level ℓ, and $f_{L-1}(n)$ is obtained from the minimization of

$$\int |f_L(n) - \sum_{k=1}^{L-1} \sum_{j=1}^{N_k} \beta_{k,j} G_k(n \cdot n_{k,j})|^2 d\Omega(n).$$

Thus, we have

$$f_L(n) = f_{L-1}(n) + \sum_{j=1}^{N_{L-1}} \gamma_{L-1,j} W_{L-1}(n \cdot n_{L-1,j}), \tag{2.9}$$

where W_{L-1} is a spherical wavelet whose linear combination becomes the detail space at level $L-1$, and recursive application of (2.9) leads to the representation of a spherical field $f_L(n)$ with spherical wavelets at multiple scales as

$$f_L(n) = f_1(n) + \sum_{k=1}^{L-1} \sum_{j=1}^{N_k} \gamma_{k,j} W_{L-1}(n \cdot n_{k,j}).$$

In this way, we can express the global field as a summation of coarse-level approximant and local activities at different levels.

Chapter 3

Radial Basis Function Approximation

Interpolation with radial basis functions (RBFs) is one of the most commonly used meshfree methods to approximate a given multivariate function when the data is observed on irregularly scattered sites. We place a basis function on each data site and evaluate it based on the Euclidean distance from the center. Then the linear combination will superpose these basis functions upon each other and interpolate the given data. Its applications include geodesy, solving partial differential equations, computer graphics, and neural networks.

Fasshauer (2007) discussed the related mathematical theory and `MATLAB` implementation concisely while Wendland (2004) focused on the theoretic parts in detail. This chapter is mostly based on these two monographs and organized as follows. In Section 3.1, we introduce the RBF interpolation with some examples and present the upper bound of the interpolation error. Section 3.2 illustrates the multiscale approximation method with RBFs and its error bounds.

3.1 Radial basis function interpolation

3.1.1 Radial basis functions and scattered data interpolation

Suppose that we have a continuous function f on a bounded domain $\Omega \subset \mathbb{R}^2$, and its noise-free scattered observations $y_j = f(x_j)$ for $j = 1, 2, \cdots, N$. We can approximate f by interpolating with certain basis functions. Radial functions are one of the most commonly used basis functions, and they are symmetric about their centers so that every point has the same value if it is same distance apart from the center. The formal definition of a radial function is given below.

Definition 3.1.1. *A function $G : \mathbb{R}^d \to \mathbb{R}$ is radial if there exists a univariate function $g : [0, \infty) \to \mathbb{R}$ such that*

$$G(x) = g(\|x\|_2), \ x \in \mathbb{R}^d,$$

where $\|x\|_2 = (x^T x)^{1/2}$.

Now, we can define an interpolant $\hat{f}(x)$ using RBF G in the form

$$\hat{f}(x) = \sum_{j=1}^{N} \hat{\beta}_j G(x - x_j), \tag{3.1}$$

which satisfies $\hat{f}(x_j) = y_j$ for $j = 1, \cdots, N$. Thus, the observations y_j's can be expressed as below.

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} G(x_1 - x_1) & G(x_1 - x_2) & \cdots & G(x_1 - x_N) \\ G(x_2 - x_1) & G(x_2 - x_2) & \cdots & G(x_2 - x_N) \\ \vdots & \vdots & \ddots & \vdots \\ G(x_N - x_1) & G(x_N - x_2) & \cdots & G(x_N - x_N) \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_N \end{bmatrix}$$

Coefficients $\hat{\beta}_j$'s are determined uniquely if the matrix $\mathbf{G} = (G(x_i - x_j))_{1 \le i,j \le N}$ is non-singular. For better theoretical characterization, we further require G to be positive definite.

Definition 3.1.2. *A function $G : \mathbb{R}^d \to \mathbb{R}$ is positive definite if*

$$\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j G(x_i - x_j) > 0, \ \forall \alpha \in \mathbb{R}^N \setminus \{\mathbf{0}\}$$

for all sets of pairwise distinct $\{x_1, \cdots, x_N\} \subset \mathbb{R}^d$ and for all $N \in \mathbb{N}$.

From Bochner's theorem, we can characterize positive definite functions as below.

Theorem 3.1.1. *If $h \in L_1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ is nonnegative and nonvanishing, then its d-dimensional Fourier transform*

$$G(x) = \int_{\mathbb{R}^d} h(\omega) \exp(-ix^T\omega) d\omega, \ x \in \mathbb{R}^d,$$

is positive definite.

For example, suppose that we observe a function defined in Franke (1979)

$$\begin{aligned} f(x_1, x_2) = &\frac{3}{4}\exp\left(-\frac{(9x_1-2)^2 + (9x_2-2)^2}{4}\right) \\ &+ \frac{3}{4}\exp\left(-\frac{(9x_1+1)^2}{49} - \frac{(9x_2+1)}{10}\right) \\ &+ \frac{1}{2}\exp\left(-\frac{(9x_1-7)^2 + (9x_2-3)^2}{4}\right) \\ &- \frac{1}{5}\exp\left(-(9x_1-4)^2 - (9x_2-7)^2\right) \end{aligned}$$

on 20 scattered sites on $[0,1] \times [0,1]$. It is called 'Franke function' and used as a testbed for scattered data interpolation. We can approximate f by interpolation with positive definite RBFs, and the Gaussian function $G(x) = \exp(-a\|x\|_2^2), \ a > 0$, is one of the most well-known functions. We can show its positive definiteness with Theorem 3.1.1 because

$$G(x) = \frac{1}{(4\pi a)^{d/2}} \int_{\mathbb{R}^d} \exp(-\|\omega\|_2^2/(4a)) \exp(-ix^T\omega) d\omega.$$

Figure 3.1 shows perspective plots of a Franke function and its interpolant with Gaussian RBF $G(x) = \exp(-16\|x\|_2^2)$. Black dots represent 20 observations used for interpolation.

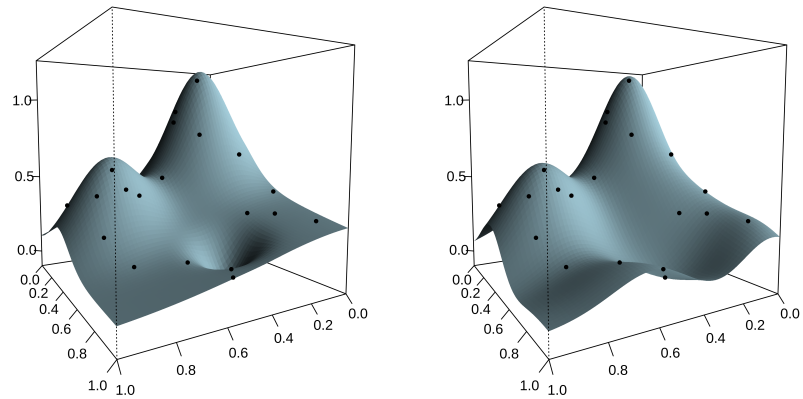

Figure 3.1: Perspective plots of a Franke function (left) and its interpolant (right). Black dots represent observations.

3.1.2 Compactly supported radial basis functions

We use RBFs with compact support to reflect the local activities only. Moreover, the advantages of using compactly supported radial basis functions (CSRBFs) are a sparse interpolation matrix $\mathbf{G} = (G(x_i - x_j))_{1 \le i,j \le N}$ and the chance of fast interpolant evaluation (Wendland, 2004). Among many CSRBFs (Wendland, 1995, Wu, 1995, Gneiting, 2002), we will introduce piecewise polynomial functions with compact support and call them Wendland functions.

Positive definiteness on $\mathbb{R}^d$

The positive definiteness of Wendland functions depends on the dimension of the space where the data sites reside. Thus, we will characterize the positive definiteness of RBFs on a specific dimension before defining Wendland functions. It is known that an integrable continuous function is positive definite if and only if it is bounded and its Fourier transform is nonnegative and nonvanishing. Since we consider a bounded, integrable, and continuous radial function G, we present a useful theorem for identi-

fying the positive definiteness of an RBF in a specific dimension.

Theorem 3.1.2. *Let us say that univariate function $g : [0, \infty) \to \mathbb{R}$ is positive definite on $\mathbb{R}^d$ if the corresponding radial function $G(x) = g(\|x\|_2)$ is positive definite for $x \in \mathbb{R}^d$. This condition holds if and only if the d-dimensional Fourier transform*

$$\begin{aligned} \mathcal{F}_d g(\|\omega\|_2) &= \int_{\mathbb{R}^d} G(x) \exp(-ix^T\omega)dx \\ &= \|\omega\|_2^{-(d-2)/2} \int_0^\infty g(t)t^{d/2} J_{(d-2)/2}(t\|\omega\|_2)dt. \end{aligned} \tag{3.2}$$

is nonnegative and nonvanishing.

We skip the proof of Theorem 3.1.2 for now and discuss it later in the proof of Theorem 4.1.2 of Section 4.

Dimension walk

We need to define two function operators and discuss their properties to define Wendland functions and calculate their multidimensional Fourier transforms.

Definition 3.1.3. *Let $\mathcal{I}$ and $\mathcal{D}$ be function operators which map the nonnegative parts of a given function as below.*

(a) For a function g such that $t \mapsto tg(t)$ is integrable on $[0, \infty)$,

$$(\mathcal{I}g)(r) = \int_r^\infty tg(t)dt, \text{ for } r \geq 0.$$

(b) For an even function g in $C^2(\mathbb{R})$,

$$(\mathcal{D}g)(r) = -\frac{1}{r}g'(r), \text{ for } r \geq 0.$$

We call $\mathcal{I}$ integral operator and $\mathcal{D}$ differential operator. The integral operator works as an inverse of the differential operator and vice versa.

We usually extend the resultants of these operators to be even functions. Based on this relation, dimension can vary for multidimensional Fourier transform of a univariate function g such that $G = g(\|\cdot\|_2)$. This is called dimension walk, and the precise statement below is from Wendland (2004).

Proposition 3.1.3. *Suppose that g is continuous.*

(a) $\mathcal{F}_d(g) = \mathcal{F}_{d-2}(\mathcal{I}g)$ *for* $d \geq 3$ *if* $t \mapsto g(t)t^{d-1} \in L_1[0, \infty)$.

(b) $\mathcal{F}_d(g) = \mathcal{F}_{d+2}(\mathcal{D}g)$ *if* $g \in C^2(\mathbb{R})$ *is even and* $t \mapsto g'(t)t^d \in L_1[0, \infty)$.

Wendland functions

Wendland functions are piecewise polynomial functions that have nonzero values on $[0, 1)$. We derive the Wendland functions from the truncated power functions

$$g_\ell(r) = (1 - r)_+^\ell, \ \ell \in \mathbb{N},$$

where $x_+ = \max\{x, 0\}$, with the integral operator $\mathcal{I}$ as below.

Definition 3.1.4. *We define the Wendland function $g_{d,k}$ to be*

$$g_{d,k}(r) = (\mathcal{I}^k g_{\lfloor d/2 \rfloor + k + 1})(r).$$

Apart from the compactness of supports, Wendland functions are positive definite and piecewise polynomial with minimal degrees. These are advantages in theoretical and numerical aspects, respectively, and Wendland (1995) states these properties as in the theorem below.

Theorem 3.1.4. *Wendland functions $g_{d,k}$ are positive definite on $\mathbb{R}^d$, and their derivatives are continuous up to order $2k$. They are piecewise polynomial functions with minimal degrees for given space dimension d and smoothness $2k$.*

Wendland functions are uniquely determined up to a constant multiplication. Table 3.1 gives a few examples of Wendland functions when $d = 2$ and Figure 3.2 plots each of these functions simultaneously.

Table 3.1: Examples of Wendland functions $g_{2,k}$ for $k = 0, 1, \cdots, 3$.

Wendland functions	Smoothness
$g_{2,0} = (1-r)_+^2$	C^0
$g_{2,1} = (1-r)_+^4(4r+1)$	C^2
$g_{2,2} = (1-r)_+^6(35r^2+18r+3)$	C^4
$g_{2,3} = (1-r)_+^8(32r^3+25r^2+8r+1)$	C^6

3.1.3 Error bounds

The approximation error of the interpolant (3.1) depends on data sites, underlying function, and the RBF. To find the upper bound of this error, Wu and Schaback (1993) suggested expressing the interpolant in Lagrange form

$$\hat{f}(x) = \sum_{j=1}^{N} f(x_j) u_j^*(x),$$

where u_j^* is a cardinal basis function such that $u_j^*(x_i) = \delta_{ij}$. If the RBF G is positive definite, we can obtain the cardinal basis functions from the system of linear equations

$$G(x - x_i) = \sum_{j=1}^{N} G(x_i - x_j) u_j^*(x), \ i = 1, 2, \cdots, N.$$

On the other hand, the interpolant (3.1) has RBFs centered on each data site. Thus, with abuse of notation, we can consider G as a kernel function such that $G(\cdot, \cdot) = G(\cdot - \cdot)$. Furthermore, note that every positive definite kernel is associated with reproducing kernel Hilbert space (RKHS), and we call it a native space of the given kernel. In other words, if G is a positive definite kernel, its native space is the completion of the spanned space of $\{G(\cdot, x) : x \in \Omega\}$ with respect to the norm $\langle \cdot, \cdot \rangle_G$ such that

$$\langle \sum_{i=1}^{N_1} \alpha_i G(\cdot, x_{1i}), \sum_{j=1}^{N_2} \alpha_j G(\cdot, x_{2j}) \rangle_G = \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \alpha_i \alpha_j G(x_{1i}, x_{2j}).$$

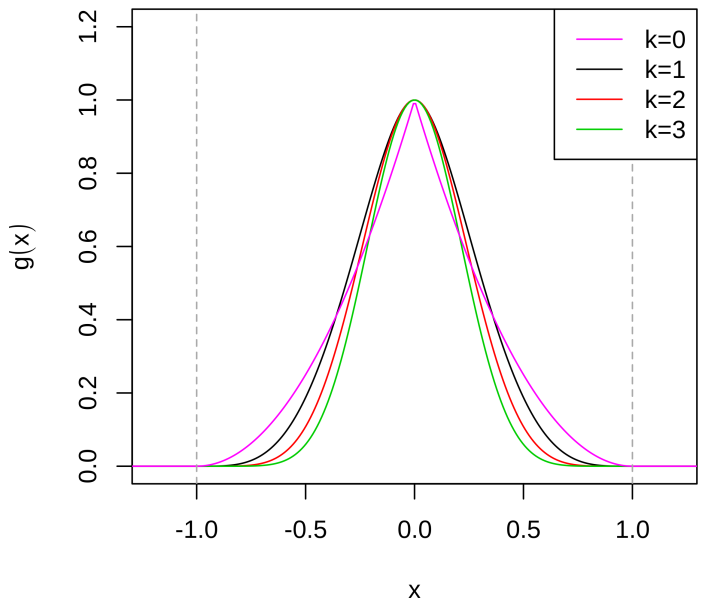

Figure 3.2: Plots of Wendland functions $g_{2,k}$ for $k = 0, 1, \cdots, 3$.

We denote this space by $\mathcal{N}_G(\Omega)$ and its corresponding norm by $\langle \cdot, \cdot \rangle_{\mathcal{N}_G(\Omega)}$, which satisfies $\|f\|_{\mathcal{N}_G(\Omega)} = \|f\|_G$ for all $f \in \text{span}\{G(\cdot, x) : x \in \Omega\}$. Thus, if f belongs to $\mathcal{N}_G(\Omega)$, we have

$$\begin{aligned}
\hat{f}(x) &= \sum_{j=1}^{N} f(x_j) u_j^*(x) \\
&= \sum_{j=1}^{N} u_j^*(x) \langle f, G(\cdot, x_j) \rangle_{\mathcal{N}_G(\Omega)} \\
&= \langle f, \sum_{j=1}^{N} u_j^*(x) G(\cdot, x_j) \rangle_{\mathcal{N}_G(\Omega)},
\end{aligned}$$

from the definition of the RKHS. This leads to the inequality

$$|f(x) - \hat{f}(x)| = |\langle f, G(\cdot, x) - \sum_{j=1}^{N} u_j^*(x) G(\cdot, x_j)\rangle_{\mathcal{N}_G(\Omega)}|$$

$$\leq \|f\|_{\mathcal{N}_G(\Omega)} \|G(\cdot, x) - \sum_{j=1}^{N} u_j^*(x) G(\cdot, x_j)\|_{\mathcal{N}_G(\Omega)}. \qquad (3.3)$$

We can decompose the right-hand side of (3.3) into two parts. One is the effect of the underlying true function, $\|f\|_{\mathcal{N}_G(\Omega)}$, and the other is the effect of the data sites, $\|G(\cdot, x) - \sum_{j=1}^{N} u_j^*(x) G(\cdot, x_j)\|_{\mathcal{N}_G(\Omega)}$.

Once the kernel G is specified, one can further bound the second term of the inequality (3.3) with respect to the fill distance

$$h_{X,\Omega} = \sup_{x \in \Omega} \min_{x_j \in X} \|x - x_j\|_2.$$

Fill distance implies the largest radius of a ball in the domain Ω which does not have any observation $X = \{x_1, x_2, \cdots, x_N\}$ inside, and it measures how densely X is scattered over Ω. Wendland (1998) has shown that the native space of Wendland function $G_{d,k}$ is a Sobolev space $H^\tau(\Omega)$ of degree $\tau = d/2 + k + 1/2$ and found the upper bound of (3.3).

Proposition 3.1.5. *Suppose that integrable and continuous function G on $\mathbb{R}^d$ satisfies an inequality*

$$0 < c_1 \leq (1 + \|\omega\|_2^2)^\tau \widehat{G}(\omega) \leq c_2, \ \omega \in \mathbb{R}^d,$$

for $\tau > d/2$. Then the native space $\mathcal{N}_G(\mathbb{R}^d)$ corresponding to G coincides with the Sobolev space $H^\tau(\mathbb{R}^d)$, and their respective norms are equivalent.

Wendland (1998) has shown that the Fourier transform of the Wendland function $g_{d,k}$ decays with the rate of $(1 + \|\cdot\|_2^2)^{-d/2-k-1/2}$, and this leads to the fact that $\mathcal{N}_{G_{d,k}}(\Omega)$ is norm equivalent to $H^\tau(\Omega)$ for $\tau = d/2 + k + 1/2$. Now we can bound the error of the RBF interpolant based on the Wendland functions if the underlying function is from the associated native space.

Theorem 3.1.6. *Suppose that $\Omega \subseteq \mathbb{R}^d$ is bounded and satisfies an interior cone condition, and $X = \{x_1, x_2, \ldots, x_N\} \subseteq \Omega$ is a set of pairwise distinct observations on Ω. Let $G_{d,k} : \mathbb{R}^d \to \mathbb{R}$ be a Wendland function in space dimension d with smoothness $2k$. If f is from its associated native space and $\hat{f}$ is the interpolant of X based on $G_{d,k}$, there exist constants $h_0, C > 0$ such that*

$$|f(x) - \hat{f}(x)| \leq C h_{X,\Omega}^{k+1/2} |f|_{\mathcal{N}_G(\Omega)}$$

holds for every $x \in \Omega$ provided that $h_{X,\Omega} \leq h_0$.

Note that the interior cone condition in Theorem 3.1.6 is necessary for bounding the remainder term of the Taylor expansion. It implies that we can find a cone of a certain radius and angle contained in Ω for every point in Ω. The formal definition of this condition is given below.

Definition 3.1.5. *A domain Ω satisfies an interior cone condition if there exist radius $r > 0$, angle $\theta \in (0, \pi/2)$, and a unit vector $\xi(x)$ for every point $x \in \Omega$ so that the cone*

$$C = \{x + \lambda y : \ x \in \Omega, \ y \in \mathbb{R}^d, \ \|y\|_2 = 1, \ y^T \xi(x) \geq \cos\theta, \ \lambda \in [0, r]\}$$

is contained in Ω.

3.2 Multiscale representation with radial basis functions

3.2.1 Multiscale approximation

Consider interpolating the data with CSRBFs with a fixed bandwidth. This guarantees the convergence of an interpolant, but the interpolation matrix will not be sparse anymore and become numerically unstable as the data gets denser. On the other hand, the interpolant will not converge to

the underlying function if one varies the bandwidth based on how densely the data is observed. One can find a detailed discussion on this trade-off relation in Schaback (1997) both theoretically and numerically.

Multiscale RBF approximation for the scattered data was first suggested by Floater and Iske (1996) to overcome the trade-off relation and represent the nature of the underlying function. As a result, it has a well-conditioned interpolation matrix, not a too heavy computational burden, and convergence property. Furthermore, different areas of the basis function supports can extract different scales of patterns. In other words, one can identify the global behavior of the data with large basis functions and the local features with small ones.

Suppose that X_1, X_2, $\cdots$, X_L are data sites in $\Omega \subseteq \mathbb{R}^d$ whose fill distance $h_\ell = h_{X_\ell,\Omega}$ decreases as ℓ gets larger. Note that we will assume that X_1, X_2, ..., X_L are mutually disjoint even if it was not required in Floater and Iske (1996). We illustrate the multiscale approximation algorithm below. At the ℓ-th level, it interpolates the residuals from the previous level, $f - \sum_{k=1}^{\ell-1} \hat{s}_k$, at the points in $X_\ell = \{x_{\ell,1}, x_{\ell,2}, \cdots, x_{\ell,N_\ell}\}$.

Algorithm 1: Multiscale approximation of scattered data

Set $\hat{f}_0 = 0$, $e_0 = f$.
for $\ell = 1, 2, \cdots, L$ **do**
    Obtain an interpolant $\hat{s}_\ell$ to $e_{\ell-1}$ at X_ℓ.
    Set $\hat{f}_\ell = \hat{f}_{\ell-1} + \hat{s}_\ell$.
    Set $e_\ell = e_{\ell-1} - \hat{s}_\ell$.
end

As an example, let us consider a modified Franke function $\breve{f}$ that exaggerates bumps and dips compared to the original one. It is defined as

$$\breve{f}(x_1, x_2) = 15 \exp\left(-\frac{1}{1-(2x_1-1)^2} - \frac{1}{1-(2x_2-1)^2}\right) f(x_1, x_2),$$

where

$$
\begin{aligned}
f(x_1, x_2) = \frac{3}{4}\exp\left(-\frac{(9x_1-2)^2+(9x_2-2)^2}{4}\right) \\
+ \frac{3}{4}\exp\left(-\frac{(9x_1+1)^2}{49} - \frac{(9x_2+1)^2}{10}\right) \\
+ \frac{1}{2}\exp\left(-\frac{(9x_1-7)^2+(9x_2-3)^2}{4}\right) \\
- \frac{1}{5}\exp\left(-(9x_1-4)^2-(9x_2-7)^2\right).
\end{aligned}
$$

Suppose that we approximate $\breve{f}$ with Algorithm 1 from the uniformly scattered 9, 25, and 81 observations from X_1, X_2, and X_3, respectively. Figure 3.3 shows a perspective plot of the modified Franke function, and Figure 3.4 illustrates the approximant of the modified Franke function at each level. We can observe from Figure 3.4 that the approximant gets more similar to the true function $\breve{f}$ as the level increases.

3.2.2 Error bounds

There had been numerous researches on the convergence property of multiscale RBF approximation. Fasshauer and Jerome (1999) discussed the converging property after the additional smoothing step, and Narcowich et al. (1999) obtained the convergence rate when the smoother radial functions are used as the scale gets coarser. Hales and Levesley (2002) proved the convergence theorem when the scattered data is approximated with multilevel polyharmonic splines. For the methods based on the Wendland functions, Gia et al. (2010) and Wendland (2010) derived the convergence rate for the domains of sphere and plane, respectively.

Before introducing the convergence theorem of multiscale approximant and its assumptions, we need to know when the bounded extension operator can be defined on the Sobolev space. Consider two Sobolev spaces of degree τ defined on $\mathbb{R}^2$, $H^\tau(\mathbb{R}^2)$, and its open subset Ω, $H^\tau(\Omega)$. If Ω

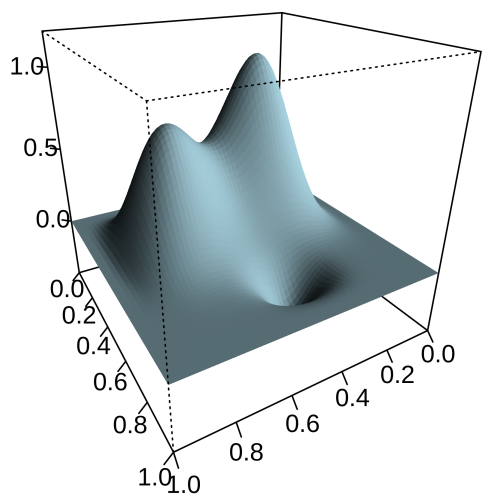

Figure 3.3: A perspective plot of the modified Franke function $\breve{f}$

has a Lipschitz boundary, then there exists a bounded extension operator $E : H^\tau(\Omega) \to H^\tau(\mathbb{R}^2)$ such that, for all $f \in H^\tau(\Omega)$, (i) $Ef|_\Omega = f|_\Omega$, and (ii) $\|Ef\|_{H^\tau(\mathbb{R}^2)} \leq C_\tau \|f\|_{H^\tau(\Omega)}$.

In Wendland (2010), three assumptions are required for the convergence of the approximant

$$\hat{f}_L(x) = \sum_{\ell=1}^{L} \sum_{j=1}^{N_\ell} \hat{\beta}_{\ell,j} G_\ell(x, x_{\ell,j}), \tag{3.4}$$

which is obtained from the Algorithm 1.

Assumption 1. *Domain $\Omega \subset \mathbb{R}^2$ is bounded and has Lipschitz boundary.*

Assumption 2. *Assume that the function $f \in H^\tau(\Omega)$ is observed at a se-*

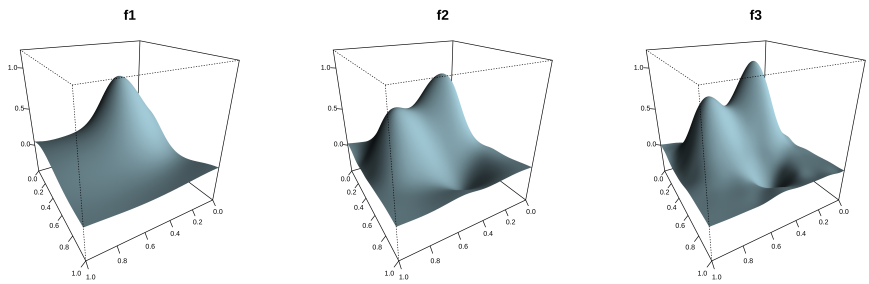

Figure 3.4: Approximants of the modified Franke function from the multi-scale radial basis approximation algorithm.

quence of data sites $X_1, X_2, \cdots$ *on* Ω *whose fill distances* $h_1, h_2, \cdots$ *satisfy*

$$c\mu h_\ell \leq h_{\ell+1} \leq \mu h_\ell$$

for fixed $\mu \in (0,1)$ *and* $c \in (0,1)$.

Assumption 3. *For each* X_ℓ*, we will use different bandwidth* δ_ℓ *that is proportional to* h_ℓ*. Denote the scaled Wendland functions by*

$$G_\ell(x,y) = \delta_\ell^{-2} G_{2,k}((x-y)/\delta_\ell),\ \ell = 1,\ 2,\ \cdots,$$

for $\delta_1 \leq 1$*, and* $h_\ell/\delta_\ell = 1/\nu \leq \mu/\gamma$ *for a fixed* $\gamma > 0$.

Assumption 1 assures the existence of the bounded extension operator as stated before, and Assumption 2 states that the data sites are getting denser within the specified range as the level increases. Assumption 3 scales the bandwidths of the RBFs at each level to be proportional to the corresponding fill distance. Now, we state the convergence theorem.

Theorem 3.2.1. *Let the target function* f *belong to* $H^\tau(\Omega)$*,* $\tau = k + 3/2$*. Then, under the Assumptions 1 - 3, there exists* $C_1 = C_1(\gamma) > 0$ *such that*

$$\|Ee_\ell\|_{G_{\ell+1}} \leq \alpha \|Ee_{\ell-1}\|_{G_\ell} \text{ for } \ell = 1, 2, \cdots,$$

where $\alpha = C_1\mu^\tau$ *and* $\|\cdot\|_{G_\ell}$ *is a norm in the native space associated with* $G_\ell : \mathbb{R}^2 \to \mathbb{R}$. *Hence, there exists* $C > 0$ *such that*

$$\|f - \hat{f}_n\|_{L_2(\Omega)} \leq C\alpha^n \|f\|_{H^\tau(\Omega)} \text{ for } \ell = 1, 2, \cdots .$$

Thus the multiscale approximant $\hat{f}_n$ *converges to* f *in* L_2 *norm if* $\alpha = C_1\mu^\tau < 1$.

Chapter 4

Multiscale Representation of Directional Scattered Data

This chapter introduces a multiscale method to represent the noisy scattered data with a directional structure. Most of the RBF interpolation methods are based on isotropic functions, and this makes the directional representation inefficient. To overcome this limitation, we define and employ a compactly supported anisotropic radial basis function (ARBF) for a new multiscale representation.

Section 4.1 extends the methods introduced in Chapter 3 by making the basis functions anisotropic. Section 4.2 decomposes the approximation space into orthogonal detail spaces to denoise the contaminated noisy data.

4.1 Anisotropic radial basis function approximation

This section modifies the Wendland functions to define compactly supported ARBFs and prove their positive definiteness. We then suggest an interpolation method based on these ARBFs and discuss its theoretical

properties.

4.1.1 Representation of a single linear directional structure

Wendland functions $G_{d,k}(x) = g_{d,k}(\|x\|_2)$ have isotropic supports, and this results in inefficient representation of the directional structure. However, one can reflect the direction of an unknown function f with fewer basis functions if the supports of them become anisotropic. Thus, we define ARBF which has an elliptic support whose major axis reflects the dominant direction. It has the form $G_S(x) = g_{d,k}(\|x\|_S)$, where $\|x\|_S = (x'S^{-1}x)^{1/2}$ is a Mahalanobis distance and S is a symmetric positive definite matrix with $|S| = 1$.

Suppose that we have a continuous function f with directional structure on a bounded domain $\Omega \subset \mathbb{R}^2$, and its noise-free scattered observations $y_j = f(x_j)$ for $j = 1, 2, \cdots, N$. Assume further that it has a single linear direction, and the first eigenvector of S reflects it. Then we can approximate f along with its directional structure by interpolating the data with G_S. The interpolant is expressed as

$$\hat{f}(x) = \sum_{j=1}^{N} \hat{\beta}_j G_S(x - x_j), \tag{4.1}$$

which satisfies $\hat{f}(x_j) = y_j$ for $j = 1, 2, \cdots, N$.

Positive definiteness on $\mathbb{R}^d$

We can determine the coefficients of (4.1) uniquely if the ARBF G_S is positive definite on $\mathbb{R}^2$. Before introducing an anisotropic version of Theorem 3.1.2, we need to note the positive definiteness of a truncated power function from the lemma below.

Lemma 4.1.1. *The truncated power function*

$$\phi_\ell(r) = (1 - r)_+^\ell, \ \ell \geq \lfloor d/2 \rfloor + 1,$$

is positive definite on $\mathbb{R}^d$.

To prove this lemma, one needs to check whether $\mathcal{F}_{2n+1}\phi_{n+1}$ is non-negative and nonvanishing when the space dimension is $d = 2n + 1$, and use the fact that $\phi_{\lfloor 2n/2 \rfloor + 1} = \phi_{\lfloor (2n+1)/2 \rfloor + 1}$. We will skip the detailed proof of the lemma in this thesis, and you can see Wendland (2004) if interested.

The next theorem states the positive definiteness of the anisotropic Wendland functions, and the proof of it is based on the modification of the results in Wendland (1995) and Wendland (1998).

Theorem 4.1.2. *The anisotropic radial basis function* $G_S(x) = g_{d,k}(\|x\|_S)$ *is positive definite on* $\mathbb{R}^d$.

Proof. We prove the positive definiteness of $G_S(x) = g_{d,k}(\|x\|_S)$ based on the dimension walk from Proposition 3.1.3

$$\mathcal{F}_d g_{d,k} = \mathcal{F}_{d+2k}\phi_{\lfloor d/2 \rfloor + k + 1},$$

where $\phi_\ell(r) = (1-r)_+^\ell$ is a truncated power function. To this end, we will show that the Fourier transform of anisotropic truncated power function $\Phi_S(x) = \phi_\ell(\|x\|_S)$ has the form

$$\widehat{\Phi}_S(\omega) = \|\omega\|_{S^{-1}}^{-(d-2)/2} \int_0^\infty \phi_\ell(t) t^{d/2} J_{(d-2)/2}(t\|\omega\|_{S^{-1}}) dt, \tag{4.2}$$

where $\|\cdot\|_S$ is a Mahalanobis norm, and S is a symmetric positive definite matrix whose determinant is unity. Then the Lemma 4.1.1 will complete the proof.

Fourier transform of Φ_S can be simply obtained when $d = 1$. In this

case, we have $\Phi_S(x) = \phi_\ell(|x|)$ so that

$$\begin{aligned}
\widehat{\Phi}_S(\omega) &= \left(\frac{1}{2\pi}\right)^{1/2} \int_{-\infty}^{\infty} \Phi_S(t) \exp(-i\omega t) dt \\
&= \left(\frac{2}{\pi}\right)^{1/2} \int_0^{\infty} \phi_\ell(t) \cos(\omega t) dt \\
&= \omega^{1/2} \int_0^{\infty} \phi_\ell(t) t^{1/2} \left(\frac{2}{\pi\omega t}\right)^{1/2} \cos(\omega t) dt \\
&= \omega^{1/2} \int_0^{\infty} \phi_\ell(t) t^{1/2} J_{-1/2}(\omega t) dt,
\end{aligned}$$

where J_ν is a Bessel function of the first kind of order ν. If $d \geq 2$, we have

$$\begin{aligned}
\widehat{\Phi}_S(\omega) &= (2\pi)^{-d/2} \int_{\mathbb{R}^d} \Phi_S(x) \exp(-i\omega' x) dx \\
&= (2\pi)^{-d/2} \int_0^{\infty} t^{d-1} \int_{E^{d-1}} \phi_\ell(t\|x\|_S) \exp(-it\omega' x) dE(x) dt \\
&= (2\pi)^{-d/2} \int_0^{\infty} \phi_\ell(t) t^{d-1} \int_{E^{d-1}} \exp(-it\omega' x) dE(x) dt
\end{aligned}$$

for a unit ellipsoid $E^{d-1} = \{x \in \mathbb{R}^d : \|x\|_S = 1\}$. If we eigendecompose the matrix $S = VDV'$ and let $P = VD^{-1/2}$, the ellipsoid E^{d-1} can be transformed to a unit sphere

$$S^{d-1} = \{y \in \mathbb{R}^d : y = P'x,\ x \in E^{d-1}\}.$$

This enables the inner integration to be calculated as

$$\begin{aligned}
\int_{E^{d-1}} \exp(-it\omega' x) dE(x) &= \int_{S^{d-1}} \exp(i(-tP^{-1}\omega)' y) dS(y) \\
&= \frac{2\pi^{(d-1)/2}}{\Gamma((d-1)/2)} \int_0^{\pi} \exp(itr\cos\theta) \sin^{d-2}\theta d\theta \\
&= \frac{2\pi^{(d-1)/2}}{\Gamma((d-1)/2)} \left\{ \sum_{k=0}^{\infty} \frac{(-1)^k (tr)^{2k}}{(2k)!} \int_0^{\pi} \cos^{2k}\theta \sin^{d-2}\theta d\theta \right\},
\end{aligned}$$

where $r = \|P^{-1}\omega\|_2 = \|\omega\|_{S^{-1}}$, by change of variables and Taylor expansion. The Equation (4.2) is then derived by the additional fact that

$$\begin{aligned}
\int_0^{\pi} \cos^{2k}\theta \sin^{d-2}\theta d\theta &= \frac{\Gamma((2k+1)/2)\Gamma((d-1)/2)}{\Gamma(k+d/2)} \\
&= \frac{(2k)!\Gamma(1/2)}{2^{2k}k!} \frac{\Gamma((d-1)/2)}{\Gamma(k+d/2)},
\end{aligned}$$

and the definition of a Bessel function of the first kind. □

Native spaces

Anisotropic Wendland function G_S has its associated native space because it is positive definite. Recall that we determine the order of the associated Sobolev space from the Fourier transform of the basis function. However, $\widehat{G}_S$ is expressed as an equation of $\|\omega\|_{S^{-1}}$, and it satisfies the inequality

$$0 < c_1 \leq \widehat{G}_S(\omega)(1 + \|\omega\|_{S^{-1}}^2)^{d/2+k+1/2} \leq c_2.$$

Since we have

$$\lambda_2 \leq \frac{1 + \omega^T S\omega}{1 + \omega^T \omega} \leq \lambda_1,$$

where λ_1 and λ_2 are the first and the second eigenvalues of S, we can easily derive that $\widehat{G}_S$ decays at the rate of $(1 + \|\cdot\|_2^2)^{-d/2-k-1/2}$, and the native space of G_S is $H^\tau(\Omega)$ for $\tau = d/2 + k + 1/2$.

Error bounds

We can obtain an error bound of the interpolant (4.1) as in Theorem 3.1.6. Note that the convergence order is the same with the interpolant based on the isotropic Wendland functions.

Theorem 4.1.3. *Suppose that $\Omega \subseteq \mathbb{R}^d$ is bounded and satisfies an interior cone condition, and $X = \{x_1, x_2, \ldots, x_N\} \subseteq \Omega$ is a set of pairwise distinct observations on Ω. Let $G_S : \mathbb{R}^d \to \mathbb{R}$ be an anisotropic Wendland function in space dimension d with smoothness $2k$ such that $G_S(x) = g_{d,k}(\|x\|_S)$ where S is a symmetric positive definite matrix with unit determinant. If f is from its associated native space and $\hat{f}(x)$ is the interpolant of X based on G_S, there exist constants h_0, $C > 0$ such that*

$$|f(x) - \hat{f}(x)| \leq C h_{X,\Omega}^{k+1/2} |f|_{\mathcal{N}_{G_S}(\Omega)}$$

holds for every $x \in \Omega$ provided that $h_{X,\Omega} \leq h_0$.

4.1.2 Representation of complex directional structure

In general, the directional structure in the data can be multiple lines, curves, or circles. Hence, it is not a good approach to fix the direction of the basis functions and represent the data. However, the convergence of the approximation is not guaranteed if we use the basis functions of different directions for different data sites. In this thesis, we suggest dividing the domain Ω into several non-overlapping subdomains $\Omega_1,\ \Omega_2, \cdots, \Omega_M$ based on the similarity of the directional structure, and using the anisotropic basis of the most representative direction for each subdomain. Then, on each subdomain Ω_m, the convergence of an approximant

$$\hat{f}^{(m)}(x) = \sum_{j=1}^{N^{(m)}} \hat{\beta}_j^{(m)} G_{S_m}(x - x_j^{(m)})$$

to the restriction of f to Ω_m is retained by Theorem 4.1.3, and so is it for the approximant on the global domain

$$\hat{f}(x) = \sum_{m=1}^{M} \hat{f}^{(m)}(x)\mathbb{I}(x \in \Omega_m). \tag{4.3}$$

Here, $N^{(m)}$ is number of the elements of the set $X^{(m)} = \Omega_m \cap X = \{x_1^{(m)}, x_2^{(m)}, \cdots, x_{N^{(m)}}^{(m)}\}$, and G_{S_m} is an anisotropic Wendland basis which best reflects the local direction on Ω_m with symmetric positive definite matrix S_m whose determinant is unity. This still has the same convergence order with the interpolant based on isotropic Wendland functions with respect to $\max\{h_{X^{(m)},\Omega_m} :\ m = 1, 2, \ldots, M\}$.

4.1.3 Multiscale representation of the directional structure

As explained in Section 3.2, multiscale representation can have sparse interpolation matrix and convergence property at the same time by overcoming the trade-off relation. It also has the advantage of reflecting the inhomogeneous patterns both globally and locally. One can obtain global behavior

if the basis functions with large supports are employed, while the local features are identified by the functions with small supports.

In this subsection, we construct a multiscale approximant based on compactly supported ARBFs to represent the complex directional structures with inhomogeneous nature. To this end, we couple the Algorithm 1 with the directional approximant (4.3). The entire data locations X will be partitioned into $X^{(1)}, X^{(2)}, \cdots, X^{(M)}$ based on the directional similarity, and each $X^{(m)}$ will be divided into a disjoint sequence of data sites $X_1^{(m)}, X_2^{(m)}, \cdots, X_L^{(m)}$ whose fill distance $h_\ell^{(m)} = h_{X_\ell^{(m)},\Omega_m}$ decreases as ℓ gets larger.

On the other hand, isotropic Wendland functions in Algorithm 1 is replaced by anisotropic ones for the directional representation. Then we approximate an unknown function on each subdomain Ω_m using elliptic Wendland functions with multiple scales. The global approximant using multiscale ARBFs can be expressed as

$$\hat{f}_L(x) = \sum_{m=1}^{M} \hat{f}_L^{(m)}(x) = \sum_{m=1}^{M} \sum_{\ell=1}^{L} \sum_{j=1}^{N_\ell^{(m)}} \beta_{\ell,j}^{(m)} G_{\ell,S_m}(x, x_{\ell,j}^{(m)}), \tag{4.4}$$

where $N_\ell^{(m)}$ is number of the elements of the set $X_\ell^{(m)} = \Omega_m \cap X_\ell$, and $G_{\ell,S_m} = G_{\ell,d,k,S_m}$ is a basis which best reflects the local direction on Ω_m in the ℓ-th scale. In addition, we can easily derive the convergence property of (4.4) when $M = 1$ because the basis G_{ℓ,S_m} is positive definite and the RBF-based approximant (3.4) converges to f in L_2-norm.

4.2 Directional wavelets for scattered data

Let f be a function on $\Omega \subseteq \mathbb{R}^2$ with a directional structure which is observed at uniformly, but irregularly, distributed data sites over Ω. Assume

further that the data is observed with independent measurement error as

$$y_j = f(x_j) + \epsilon_j,\ j = 1, 2, \ldots, N.$$

In this section, we propose a method to estimate the underlying true function f in the presence of noise. We use compactly supported ARBFs at L different scales and orthogonalize them across the scales to separate the local activities from the global trend. The shrinkage method is suggested for the noise removal, and the practical implementation issues are discussed as well.

4.2.1 Directional wavelets

Without loss of generality and for the simplicity of notation, let us assume that the function f has a single linear directional structure, and let $X_1 = \{x_{1,j}\}_{j=1}^{N_1}, X_2 = \{x_{2,j}\}_{j=1}^{N_2}, \cdots, X_L = \{x_{L,j}\}_{j=1}^{N_L}$ be a sequence of data sites on Ω which is getting denser. Then, in the context of multiresolution analysis, an approximation space at the ℓ-th level is defined as

$$\mathbb{V}_\ell = \text{span}\{G_{k,S}(\cdot, x) : x \in X_k,\ k = 1, 2, \cdots, \ell\},$$

and it contains the approximant

$$\hat{f}_\ell(x) = \sum_{k=1}^{\ell} \sum_{j=1}^{N_k} \hat{\beta}_{k,j} G_{k,S}(x, x_{k,j}),$$

where $G_{\ell,S} = \delta_\ell^{-2} G_S(\cdot/\delta_\ell)$, δ_ℓ is a bandwidth of the basis functions at the ℓ-th level, and G_S is an anisotropic Wendland function which best reflects the directional structure. The approximation spaces have a nested structure

$$\mathbb{V}_1 \subset \mathbb{V}_2 \subset \cdots \subset \mathbb{V}_L,$$

and we can show the completeness, $\bigcup_{\ell=1}^{\infty} \mathbb{V}_\ell = L_2(\Omega)$, as in Gia et al. (2010). On the other hand, we can characterize the detail space $\mathbb{W}_\ell =$

$\mathbb{V}_{\ell+1} \cap \mathbb{V}_\ell^\perp$ by the span of $W_{\ell,S}(x, x_{\ell,j})$ which satisfy

$$\int W_{\ell,S}(x, x_{\ell,j}) W_{\ell',S}(x, x_{\ell',j'}) dx = 0, \ \ell \neq \ell', \ \forall x_{\ell,j} \in X_\ell, \ \forall x_{\ell',j} \in X_{\ell'}.$$

Since $W_{\ell,S}$ are orthogonal across the levels by definition, we call these functions the directional wavelets and obtain them by the Gram-Schmidt process

$$W_{\ell,S}(x, x_{\ell,j}) = G_{\ell,S}(x, x_{\ell,j}) - \sum_{\ell'=1}^{\ell-1} \sum_{j'=1}^{N_{\ell'}} e_\ell(\ell', j', j) G_{\ell'}(x, x_{\ell',j'}),$$

where $e_\ell(\ell', j', j)$ can be seen as a filter in classical wavelet transform. Thus, we can decompose the approximant at the ℓ-th level as

$$\hat{f}_\ell(x) = \hat{f}_{\ell-1}(x) + \sum_{j=1}^{N_\ell} \hat{\gamma}_{\ell,j} W_{\ell,S}(x, x_{\ell,j}),$$

for $\ell = 2, \cdots, L$, so that the approximant at the finest level can be expressed as

$$\hat{f}_L(x) = \sum_{j=1}^{N_1} \hat{\beta}_{1,j} G_{1,S}(x, x_{1,j}) + \sum_{\ell=2}^{L} \sum_{j=1}^{N_\ell} \hat{\gamma}_{\ell,j} W_{\ell,S}(x, x_{\ell,j}). \tag{4.5}$$

4.2.2 Estimation of coefficients

We use the least squares approach to estimate the coefficients in (4.5), and this leads to the interpolation of the entire data observed at $X_1, X_2, \cdots, X_L$. However, since the measurement errors contaminate the observations, we cannot directly make an inference about the underlying function f from the interpolation. The estimated coefficient of the directional wavelet $W_{\ell,S}(\cdot, x_{\ell,j})$ from the least squares method are expected to contain the noise as

$$\hat{\gamma}_{\ell,j} = \gamma_{\ell,j} + \eta_{\ell,j}.$$

Since the localized direction of $W_{\ell,S}$ causes $\gamma_{\ell,j}$'s to be sparse, we can threshold $\hat{\gamma}_{\ell,j}$'s with small values and remove the noise. Thus, it is important to determine the rule of thresholding because optimal values of the

threshold vary for the nature of the underlying function of interest, noise level, and the type of basis functions in use. In our case, the noise $\eta_{\ell,j}$'s might be correlated with one another, even if $\epsilon_{\ell,j}$'s are not, because the directional structure of f is represented only with the elliptic basis functions. In this respect, we suggest conducting the level-dependent thresholding of Johnstone and Silverman (1997) and determining the thresholding value at each level based on cross-validation.

4.2.3 Practical issues in implementation

This subsection will introduce practical approaches that we suggest to reflect the local directional structure, divide the basis functions into multiple levels disjointly, avoid the multicollinearity problem, and remove the noise.

Construction of anisotropic Wendland functions

In practice, it is challenging to explicitly divide the domain Ω based on the directional similarity and specify the representative direction for each subdomain. So, we will instead set the direction of each basis based on the neighboring observations around the center in $\Omega \times \mathbb{R}$ to reflect local directional information implicitly.

Let $Y_\ell = \{y_{\ell,j}\}_{j=1}^{N_\ell}$ and $X_\ell = \{x_{\ell,j}\}_{j=1}^{N_\ell}$ denote the observations and data sites at the ℓ-th level, respectively. In this research, for each $(x_{\ell,j}, y_{\ell,j}) \in (X_\ell, Y_\ell)$, we find seven nearest neighbors in $\Omega \times \mathbb{R}$, project them on Ω, and obtain the symmetric positive definite matrix $S_{\ell,j}$ whose determinant is scaled to unity from the covariance matrix of projected observations. Then we can define Mahalanobis distance from $x_{\ell,j}$ and make anisotropic Wendland function

$$G_{\ell,S}(\cdot, x_{\ell,j}) = \delta_\ell^{-2} g_{2,k}\left(\frac{\sqrt{(\cdot - x_{\ell,j})' S_{\ell,j}^{-1}(\cdot - x_{\ell,j})}}{\delta_\ell}\right),$$

where $g_{2,k}$ is a Wendland function in space dimension 2 and smoothness $2k$, and δ_ℓ is a bandwidth which is proportional to the fill distance

$$h_{X_\ell,\Omega} = \sup_{x\in\Omega} \min_{x_{\ell,j}\in X_\ell} \|x - x_{\ell,j}\|_2.$$

Level division

Since the basis functions are centered on observation sites, dividing them into multiple levels implies dividing the corresponding observations. We expect the data sites X_ℓ at each level to be uniformly distributed so that the basis functions defined on them can cover Ω with close to minimal bandwidth while reflecting directional structure. To this end, we will remove the observations progressively using non-adaptive and adaptive thinning algorithms sequentially and make a nested structure of the data sites as

$$X = \bigcup_{\ell=1}^{L} X_\ell \supset \bigcup_{\ell=1}^{L-1} X_\ell \supset \cdots \supset X_1.$$

Non-adaptive thinning (Floater and Iske, 1998) is an algorithm that removes observations while preserving uniformity of data sites as much as possible, and adaptive thinning (Dyn et al., 2001) is an algorithm that removes less significant observations first based on 2-dimensional linear interpolation on Delaunay triangulation. Detailed descriptions of these two algorithms can be found in Iske (2004).

We will first apply non-adaptive thinning algorithm and remove N_L, N_{L-1}, ..., N_2 observations sequentially, and denote their observation sites by X_L, X_{L-1}, ..., X_2, respectively. The remaining N_1 observations are located on X_1, and these will be the centers of basis functions at the coarsest level. Though non-adaptive thinning makes X_ℓ uniformly distributed in Ω, we cannot guarantee that the basis functions at each level, especially the ones on the coarsest level, are reflecting the local directional structure because only a few nearest observations around the center are used to determine the direction of the basis functions.

To overcome this limitation, we use adaptive thinning locally to switch the center of each basis at the coarse level with the one at the finer level and adjust the direction of the basis function accordingly. New centers of basis functions at the coarse level should be close enough to the original centers and are able to represent the local directional structure. The detailed description is stated in Algorithm 2.

For example, suppose that we observe two-dimensional directional data as in the top middle plot in Figure 4.1. The number of observations is 115, and the non-adaptive thinning divides this into three levels. First, 81 observations will be removed, and these will be allocated to level 3 (bottom right plot in Figure 4.1). Next, we remove 25 observations in the next step, and these will be the data at level 2 (bottom middle plot in Figure 4.1). Finally, the remaining 9 observations will naturally be at level 1 (bottom right plot in Figure 4.1).

We will then exchange the data at level 1 with the other observations at the finer levels if it can improve the directional representation at the coarsest level (Algorithm 2). We want the exchange of the observations locally to keep the data uniformly scattered. Hence, we will restrict the domain of exchange by the circular neighborhood of each observation at level 1. We consider that the adequate area of the circle should be slightly larger than $Area(\Omega)/N_1 = 1/9$. To this end, we will make a circle whose diameter is $\sqrt{2Area(\Omega)/N_1} = \sqrt{2}/3$ for each $x_{1,j}$, and this will be the domain of exchange in this specific setting. Since the area of the neighborhoods is $\pi/18$, there will be approximately 18 observations for each of them. Hence, for each neighborhood, we will remove 9 observations by the adaptive thinning and find a medoid of the remaining ones. If this medoid belongs to level 2 or 3, we will exchange the level it belongs with the original center of the circular neighborhood. If it belongs to level 1, nothing will happen. Figure 4.2 plots the supports of anisotropic basis functions at each level after local

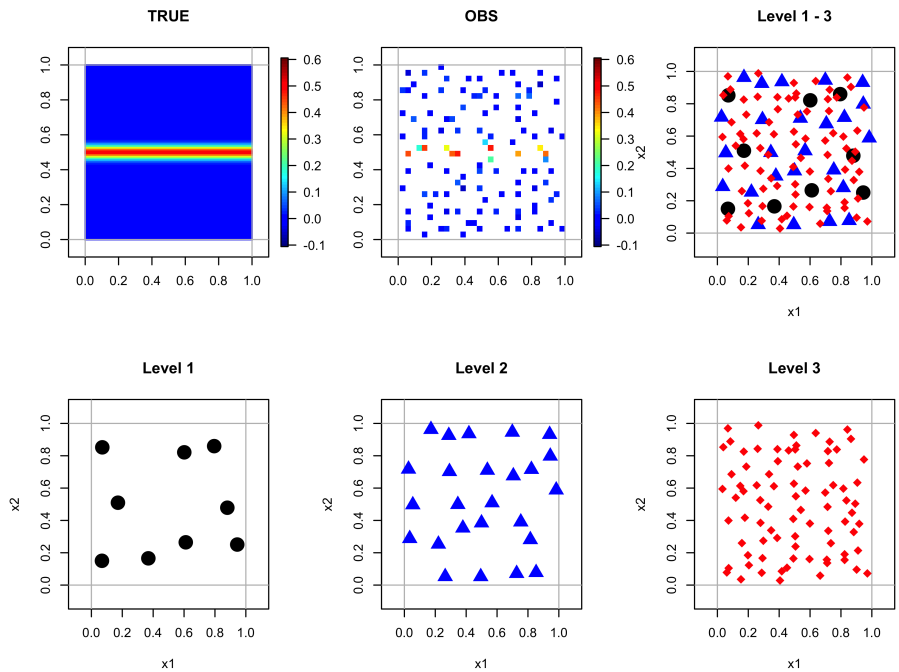

Figure 4.1: An illustrative example of non-adaptive thinning. First row presents, from left to right, the plots of a function with simple directional structure, scattered observations of the function, and the result of non-adaptive thinning. Second row shows the data at each level.

adaptive thinning. We can observe that the basis functions at every level reflect the underlying directional structure very well.

Rank deficiency problem

We orthogonalize the basis functions across the levels using the Gram-Schmidt process after the construction. Thus, the coefficients in (4.5) can be estimated at each level separately. However, one should not estimate

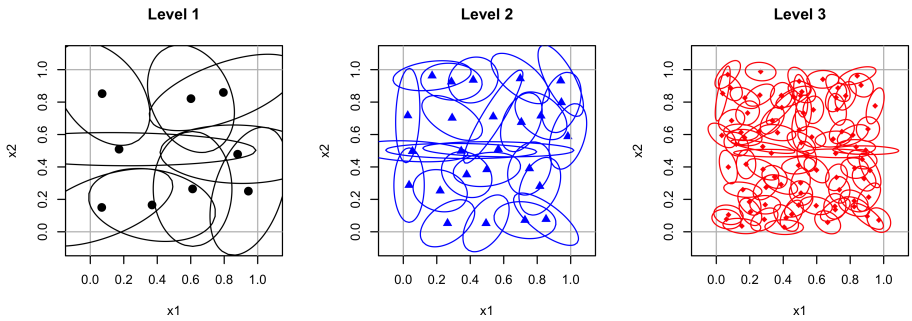

Figure 4.2: Supports of anisotropic basis functions at each level after local adaptive thinning.

the coefficients by directly minimizing

$$\|Y - \sum_{j=1}^{N_1} \hat{\beta}_{1,j} G_{1,S}(X - x_{1,j})\|_2,$$

or

$$\|Y - \sum_{j=1}^{N_\ell} \hat{\gamma}_{\ell,j} W_{\ell,S}(X - x_{\ell,j})\|_2,$$

for $\ell = 2,\ 3,\ \ldots,\ L$, because linear dependency among basis functions results in unstable predicted values on the points far from the observation sites. In this research, we examine multicollinearity by the ratio between the smallest and the largest eigenvalues of the covariance matrix of predictors at each level. If the ratio is less than 0.01, we use principal component regression instead of least square regression to alleviate the effect of multicollinearity.

For example, let $\boldsymbol{\beta} = (\beta_{1,1}, \beta_{1,2}, \ldots, \beta_{1,N_1})^T$,

$$\boldsymbol{G} = \begin{bmatrix} \{G_S(x_{1,j}, x_{1,1})\}_{j=1}^{N_1} & \{G_S(x_{1,j}, x_{1,2})\}_{j=1}^{N_1} & \cdots & \{G_S(x_{1,j}, x_{1,N_1})\}_{j=1}^{N_1} \\ \{G_S(x_{2,j}, x_{1,1})\}_{j=1}^{N_2} & \{G_S(x_{2,j}, x_{1,2})\}_{j=1}^{N_2} & \cdots & \{G_S(x_{2,j}, x_{1,N_1})\}_{j=1}^{N_2} \\ \vdots & \vdots & \ddots & \vdots \\ \{G_S(x_{L,j}, x_{1,1})\}_{j=1}^{N_L} & \{G_S(x_{L,j}, x_{1,2})\}_{j=1}^{N_L} & \cdots & \{G_S(x_{L,j} - x_{1,N_1})\}_{j=1}^{N_L} \end{bmatrix},$$

Algorithm 2: Center adjustment of basis functions in the coarse level to reflect the local directional structure

for $\ell = 1,\ 2,\ \ldots,\ \ell_0$ **do**
  for $j = 1,\ 2,\ \ldots,\ N_\ell$ **do**
    Find a circle whose center is $x_{\ell,j}$ and diameter is the length of a square's diagonal whose area is $Area(\Omega)/N_\ell$.
    Remove half of the observations in the circle by adaptive thinning.
    Find a medoid of remaining observations and identify the level it belongs to, denoting it by ℓ'.
    Exchange the levels of the basis functions if $\ell' > \ell$.

$$x_{\ell,j} \leftrightarrow x_{\ell',j'}$$

  end
end

and $\boldsymbol{V}$ be a matrix of eigenvectors of the covariance matrix

$$\frac{1}{N_1}(\boldsymbol{G} - \frac{1}{N_1}\boldsymbol{1}\boldsymbol{1}'\boldsymbol{G})'(\boldsymbol{G} - \frac{1}{N_1}\boldsymbol{1}\boldsymbol{1}'\boldsymbol{G})$$

whose corresponding eigenvalues are larger than one-hundredth of the largest eigenvalue. If we let

$$\boldsymbol{Z} = (\boldsymbol{G} - \frac{1}{N_1}\boldsymbol{1}\boldsymbol{1}'\boldsymbol{G})\boldsymbol{V}$$

and estimate the corresponding coefficients by least squares criterion as $\hat{\boldsymbol{\alpha}} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'Y$, estimator of $\boldsymbol{\beta}$ from the principal component regression is obtained by

$$\hat{\boldsymbol{\beta}} = \boldsymbol{V}\hat{\boldsymbol{\alpha}}.$$

Similar method can be applied to estimate $\boldsymbol{\gamma}_\ell = \{\gamma_{\ell,j}\}_{j=1}^{N_\ell}$ from

$$\boldsymbol{W}_\ell = \begin{bmatrix} \{W_{\ell,1}(x_{1,j})\}_{j=1}^{N_1} & \{W_{\ell,2}(x_{1,j})\}_{j=1}^{N_1} & \cdots & \{W_{\ell,L_1}(x_{1,j})\}_{j=1}^{N_1} \\ \{W_{\ell,1}(x_{2,j})\}_{j=1}^{N_2} & \{W_{\ell,2}(x_{2,j})\}_{j=1}^{N_2} & \cdots & \{W_{\ell,L_1}(x_{2,j})\}_{j=1}^{N_2} \\ \vdots & \vdots & \ddots & \vdots \\ \{W_{\ell,1}(x_{L,j})\}_{j=1}^{N_L} & \{W_{\ell,2}(x_{L,j})\}_{j=1}^{N_L} & \cdots & \{W_{\ell,L_1}(x_{L,j})\}_{j=1}^{N_L} \end{bmatrix}$$

for $\ell = 2, 3, \ldots, L$.

Chapter 5

Numerical Experiments

5.1 Simulation study

Let's consider a domain $\Omega = \{(x_1, x_2) : (x_1, x_2) \in [0, 1] \times [0, 1]\}$ and five different directional structures lying on it including straight line, sine curve, circle, crossing lines, and the Greek letter ϕ. We constitute this set of structures for the simulation to consider various forms of lines and curves. Sine curve and circle are to cover curves with different curvatures. Crossing lines and ϕ represent directions that cross each other. Figure 5.1 illustrates these structures, which are generated from the bivariate functions below.

- Straight line

$$f_0(x_1, x_2) = \exp\left(-\frac{(x_2 - (0.6x_1 + 0.2))^2}{2w^2}\right)$$

- Sine curve

$$f_0(x_1, x_2) = \exp\left(-\frac{(x_2 - (0.25\sin(2\pi x_1) + 0.5))^2}{2w^2}\right)$$

- Circle

$$f_0(x_1, x_2) = \exp\left(-\frac{((x_1 - 0.5)^2 + (x_2 - 0.5)^2 - 1/9)^2}{0.5w^2}\right)$$

- Crossing lines

$$f_0(x_1, x_2) = \max\left\{\exp\left(-\frac{(x_1 - x_2)^2}{2w^2}\right), \exp\left(-\frac{(x_1 + x_2 - 1)^2}{2w^2}\right)\right\}$$

- ϕ

$$f_0(x_1, x_2) = \max\{\exp\left(-\frac{(x_1 - (0.6x_2 + 0.2))^2}{w^2}\right), \exp\left(-\frac{((x_1 - 0.5)^2 + (x_2 - 0.5)^2 - 1/9)^2}{0.5w^2}\right)\}$$

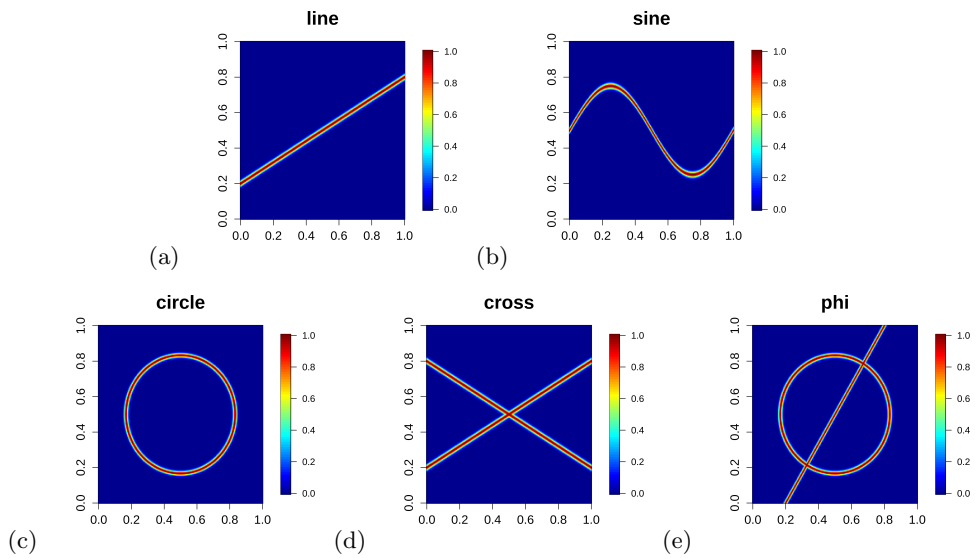

Figure 5.1: Five different directional structure of simulation data: (a) straight line, (b) sine curve, (c) circle, (d) crossing lines, (e) ϕ.

We further assume that the structure has intensity $\alpha(x_1, x_2)$ and consider two scenarios of it: constant intensity $\alpha(x_1, x_2) = 0.5$ and non-constant intensity $\alpha(x_1, x_2) = 0.5 + 0.15\cos(2\pi x_1)$. Hence, there are 10

different types of underlying functions and we observe them in the form of

$$y_j = f(x_{1j}, x_{2j}) + \epsilon_j, \ j = 1, \ 2, \ \ldots, \ N, \tag{5.1}$$

where $f = \alpha f_0$ and ϵ_j is a Gaussian white noise from $N(0, \sigma^2)$. Figure 5.2 illustrates the surface of underlying bivariate functions with perspective plots. These functions have a directional structure of the shape ϕ and have constant and non-constant intensity, respectively.

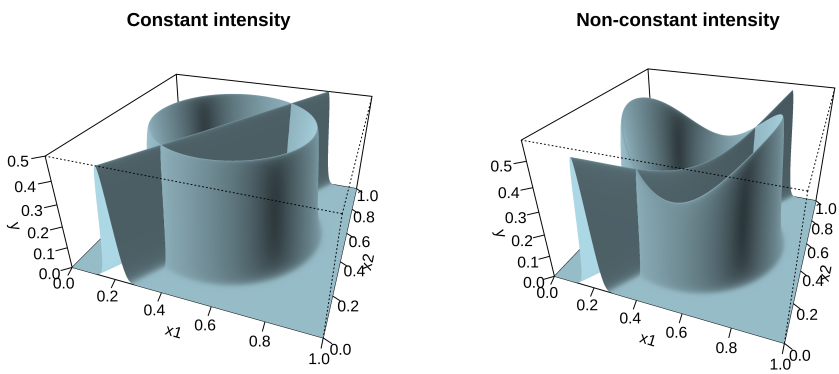

Figure 5.2: Perspective plots of ϕ-shaped bivariate functions with constant and non-constant directional intensity.

As a specific simulation setting, we set the width of direction by $w = 1/90$, and the standard deviation of noise to be $\sigma = 0.025$ or 0.05. The proposed method, directional wavelets, works not only on the irregularly scattered sites but also on the equally-spaced grid. Hence, in this section, we generate the simulation data for both designs and prove the applicability and efficiency of the proposed method by the comparison with existing methods. These methods are listed below.

- 2-dimensional wavelet transform: We use the tensor product of one-dimensional Daubechies least-asymmetric orthonormal compactly sup-

ported wavelets with 10 vanishing moments. For the noise removal, empirical Bayes thresholding is used. This can be easily implemented by `wavethresh` and `EbayesThresh` packages in `R`, respectively.

- Curvelet transform: Candes and Donoho (1999) generalized the ridgelet transform to analyze the singularity along the curves efficiently. `CurveLab` Toolbox (Candes et al., 2005) in `MATLAB` implements the second generation of this transform (Candes et al., 2006) and we use wrapping-based one for our simulation experiments.

- Adaptive weights smoothing (AWS): Polzehl and Spokoiny (2000) suggested a smoothing method that gradually expands and smooths the local constant model. This method can be easily implemented by `aws` package in `R`.

- Patch-wise adaptive weights smoothing (PAWS): Polzehl et al. (2020) suggested a method that overcomes the limitation of local constant assumption in AWS and can represent a smooth function. This method can also be easily implemented by `aws` package in `R`.

- Thin-plate smoothing splines (TPS): TPS is a minimizer of

$$\frac{1}{N}\sum_{j=1}^{N}(y_i - f(x_{1i}, x_{2i}))^2 + \lambda \int\int \left(\frac{d^2 f}{dx_1^2} + 2\frac{d^2 f}{dx_1 dx_2} + \frac{d^2 f}{x_2^2}\right) dx_1 dx_2.$$

  It penalizes the roughness of the data measured by the sum of second-order derivatives. We implement TPS by the `fields` package in `R`.

5.1.1 Scattered observation sites

We consider two cases for the number of observation sites ($N = 1024, 4096$). Each coordinate of observation sites follows uniform distribution on $[0, 1]$ independently so that the observations are randomly scattered over the

domain Ω. We generate the simulation data and estimate the underlying function f 50 times for each combination of direction and intensity. Since curvelet transform, discrete wavelet transform, and PAWS requires equally-spaced observation sites, we consider 128×128 equally-spaced grid points $\{z_1, z_2, \ldots, z_{16384}\}$ on Ω and linearly interpolate these points based on Delaunay triangulation before applying the aforementioned methods. Evaluation on the grid points which is not inside of any Delaunay triangle is substituted by the average of interpolated values on the nearest grid points. Figure 5.3 shows an example of scattered observations and their 2-dimensional interpolated values on 128×128 grid for the function of sine curve direction with constant intensity.

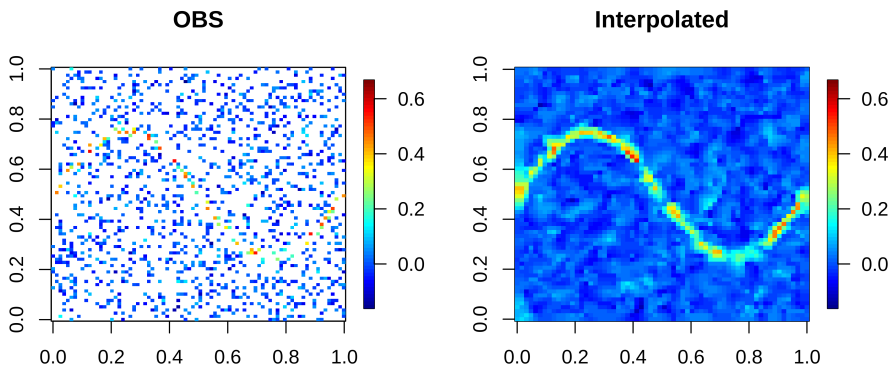

Figure 5.3: Scattered observations and their 2-dimensional interpolated values on 128×128 grid for the function of sine curve direction with constant intensity

To estimate the underlying true function f by the directional wavelets, we apply the non-adaptive thinning first to divide the observed values and sites into multiple levels $\{y_{\ell,j}, x_{\ell,j}\}_{j=1}^{N_\ell}$ for $\ell = 1, \cdots, L$. We let $L = 3$ when $N = 1024$, and $L = 4$ when $N = 4096$. We describe the detailed number

of observations at each level for each N in Table 5.1. Then, we switch the

Table 5.1: The number of observations at each level in the simulation.

	N_1	N_2	N_3	N_4
$N = 1024$	64	196	764	-
$N = 4096$	64	225	784	3023

centers of basis functions at level 1 ($\ell = 1$) with the ones in the other levels by adaptive thinning to reflect the local directional structure in the respective supporting areas. Anisotropic Wendland basis functions and wavelets are constructed based on this division of observations. To estimate the coefficients, principal component regression is used in the case of multicollinearity as in the subsection 4.2.3. Figure 5.4 illustrates an example of the disposition of basis functions and the consequently estimated structure at each level. We observe that the basis functions and their resultant estimates reflect the underlying directional structure as we intended.

Coefficients of the wavelets are soft thresholded based on predetermined numbers, which vary between the shapes of directions, not by their intensities. To alleviate the computation burden, we have determined the values of thresholds on an ad hoc basis and stated them in Table 5.2. These numbers are obtained by a grid search for the first 10 simulation data. The Figure 5.5 shows the denoised result of estimation at each level and their cumulative sums for the function of sine curve direction with constant intensity. We find that the thresholded estimate at each level has a much more distinct direction than before the denoising, and their cumulative sums become more and more precise as the details are added.

Figure 5.6 is an example of resultant estimates from the aforementioned methods for the function of sine curve direction with constant intensity. Each method is evaluated on 128×128 equally-spaced grid points, and we

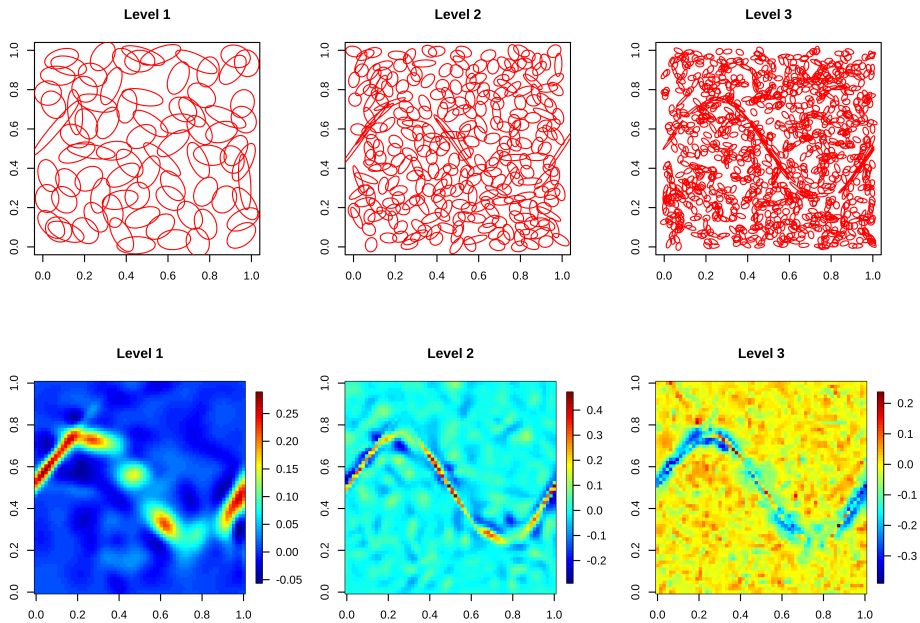

Figure 5.4: Disposition of bases and estimated structure from them on each level

can compare the results qualitatively from this. The proposed method represents a sharp directional structure more efficiently and provides the most accurate estimate of the underlying function compared to the other methods. Curvelet transform, PAWS, and TPS express the directional structure by and large but not as precisely as a directional wavelet. On the other hand, AWS provides the most smooth and accurate estimate for the background region with no directional structure. However, it has a limitation on a continuous representation of the direction because it approximates the function based on a locally constant model. The two-dimensional wavelet transform expresses the upper left part of the directional structure quite well, but it fails for the lower right part. Since the transform is applied only horizontally and vertically, we conjecture that this is because of vague inter-

Table 5.2: Thresholding value at each level for the scattered data simulation. Values are based on directional intensity, shape, and noise level.

Intensity	Direction	Noise level					
		$\sigma = 0.025$			$\sigma = 0.05$		
		Level 2	Level 3	Level 4	Level 2	Level 3	Level 4
Constant	line	.0045	.0025	.0011	.0130	.0065	.0056
	sine	.0075	.0030	.0007	.0160	.0060	.0024
	circle	.0050	.0025	.0005	.0110	.0050	.0021
	cross	.0015	.0020	.0007	.0080	.0050	.0033
	ϕ	.0050	.0025	.0005	.0090	.0055	.0016
Non-constant	line	.0050	.0020	.0011	.0145	.0065	.0051
	sine	.0075	.0025	.0007	.0170	.0070	.0022
	circle	.0055	.0025	.0005	.0120	.0055	.0023
	cross	.0030	.0020	.0007	.0070	.0045	.0030
	ϕ	.0045	.0025	.0005	.0090	.0050	.0017

polation results on the lower right part of the direction due to the random observation sites.

For the quantitative comparison, we compute the mean squared error on 128×128 equally-spaced grid points in Ω as

$$\frac{1}{16384} \sum_{j=1}^{16384} \left(f(z_j) - \hat{f}(z_j) \right)^2,$$

where $\hat{f}$ is the estimated function. Table 5.3 shows the simulation results for the data with constant directional intensity, and Table 5.4 shows the ones with non-constant directional intensity. We find that the proposed method approximates the underlying true function f more accurately than the other methods for almost every combination of direction and its intensity variation.

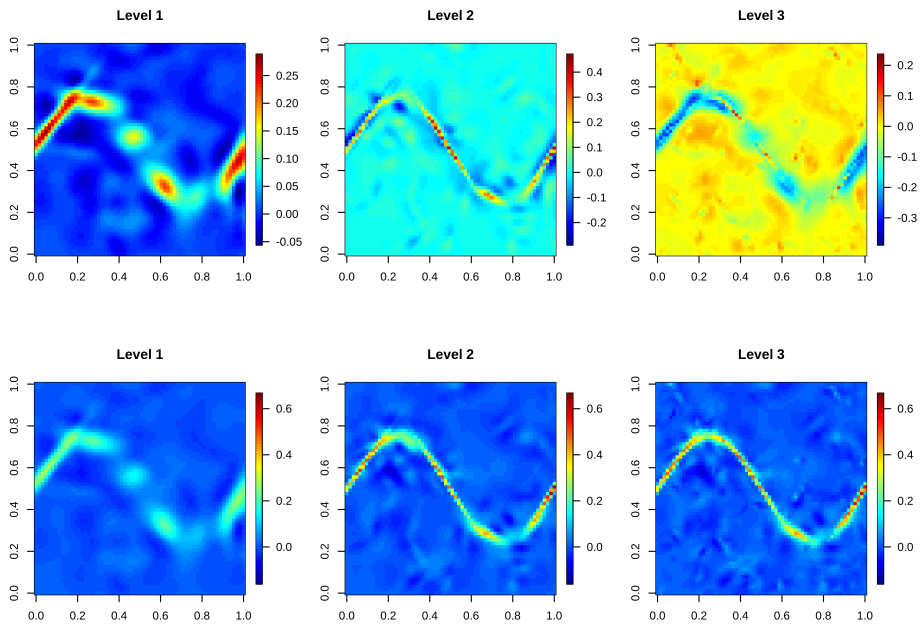

Figure 5.5: Denoised result of estimation on each level and their cumulative sums

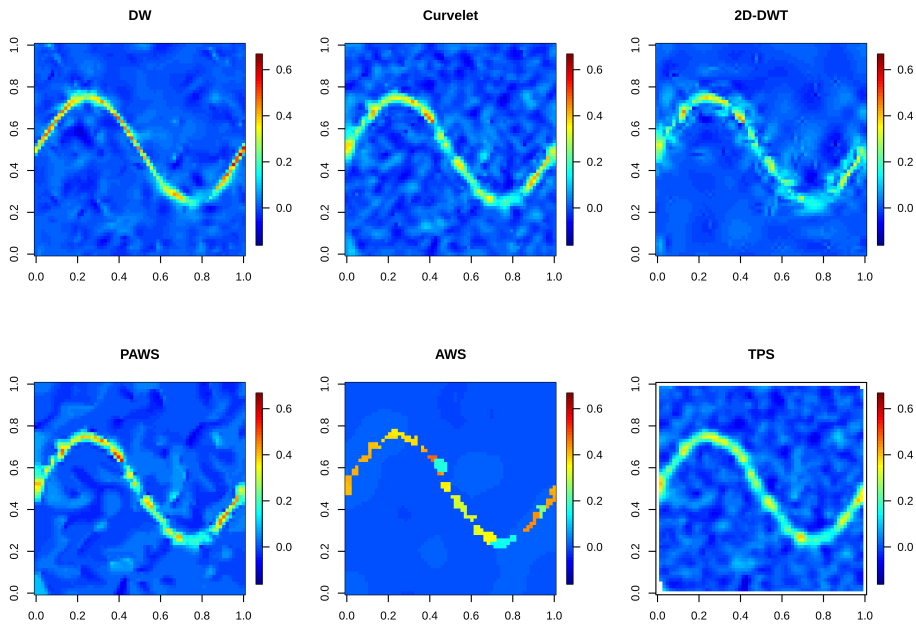

Figure 5.6: Resultant estimates from the comparing methods for the function of sine curve direction with constant intensity.

Table 5.3: Simulation results for scattered observations from each directional structure with constant intensity. Values presented in the table are averages (standard deviations in the parentheses) of mean squared error multiplied by 1000.

N	Method	Noise level									
		$\sigma = 0.025$					$\sigma = 0.05$				
		line	sine	circle	cross	phi	line	sine	circle	cross	phi
1024	DirW	**0.84 (0.31)**	**1.96 (0.39)**	**2.43 (0.49)**	**2.04 (0.48)**	**4.03 (0.52)**	**1.45 (0.33)**	**2.6 (0.38)**	**3.28 (0.45)**	**2.96 (0.55)**	**4.92 (0.60)**
	CURVELET	2.07 (0.25)	2.65 (0.27)	3.58 (0.31)	3.94 (0.31)	5.46 (0.40)	2.60 (0.26)	3.20 (0.29)	4.18 (0.34)	4.52 (0.31)	6.07 (0.41)
	2-D DWT	2.29 (0.25)	2.89 (0.26)	4.24 (0.29)	4.84 (0.38)	6.81 (0.35)	2.91 (0.26)	3.49 (0.26)	4.98 (0.29)	5.69 (0.40)	7.51 (0.33)
	PAWS	2.43 (0.27)	3.03 (0.32)	3.50 (0.32)	4.54 (0.35)	4.92 (0.37)	3.12 (0.27)	3.71 (0.30)	4.60 (0.36)	5.04 (0.32)	6.50 (0.43)
	AWS	3.04 (0.33)	3.86 (0.31)	4.96 (0.36)	5.94 (0.43)	7.03 (0.40)	3.14 (0.28)	3.89 (0.28)	5.60 (0.33)	6.02 (0.38)	8.50 (0.50)
	TPS	2.22 (0.23)	2.88 (0.24)	3.97 (0.28)	4.04 (0.38)	6.02 (0.42)	2.83 (0.21)	3.37 (0.21)	4.66 (0.27)	4.88 (0.36)	6.74 (0.38)
4096	DirW	**0.32 (0.07)**	**0.71 (0.09)**	**0.90 (0.10)**	**0.61 (0.09)**	**1.27 (0.14)**	**0.66 (0.12)**	**1.24 (0.12)**	1.57 (0.13)	**1.15 (0.18)**	**2.11 (0.20)**
	CURVELET	0.56 (0.06)	0.95 (0.09)	1.17 (0.09)	1.05 (0.10)	1.89 (0.14)	0.84 (0.07)	1.25 (0.10)	**1.55 (0.10)**	1.38 (0.11)	2.31 (0.14)
	2-D DWT	0.78 (0.10)	1.25 (0.12)	1.87 (0.15)	2.39 (0.35)	3.95 (0.15)	1.30 (0.12)	1.82 (0.12)	2.61 (0.17)	3.17 (0.37)	4.84 (0.20)
	PAWS	0.69 (0.07)	1.06 (0.10)	1.20 (0.10)	1.20 (0.11)	1.93 (0.15)	1.26 (0.08)	1.62 (0.11)	1.77 (0.11)	1.75 (0.12)	2.48 (0.16)
	AWS	1.11 (0.08)	1.74 (0.11)	2.34 (0.13)	2.20 (0.12)	3.67 (0.22)	1.28 (0.10)	1.92 (0.14)	2.58 (0.16)	2.47 (0.13)	3.99 (0.25)
	TPS	0.64 (0.05)	1.02 (0.08)	1.11 (0.08)	0.98 (0.07)	1.73 (0.11)	1.17 (0.06)	1.60 (0.08)	1.83 (0.08)	1.70 (0.09)	2.58 (0.13)

Table 5.4: Simulation results for scattered observations from each directional structure with fluctuating intensity. Values presented in the table are averages (standard deviations in the parentheses) of mean squared error multiplied by 1000.

N	Method	Noise level									
		$\sigma = 0.025$					$\sigma = 0.05$				
		line	sine	circle	cross	phi	line	sine	circle	cross	phi
1024	DirW	**0.90 (0.29)**	**2.22 (0.50)**	**2.38 (0.46)**	**1.98 (0.48)**	**3.55 (0.41)**	**1.59 (0.42)**	**2.88 (0.37)**	**3.21 (0.41)**	**2.86 (0.57)**	**4.44 (0.42)**
	CURVELET	2.23 (0.27)	2.82 (0.31)	3.35 (0.31)	4.28 (0.35)	4.74 (0.35)	2.76 (0.27)	3.37 (0.31)	3.94 (0.34)	4.85 (0.37)	5.34 (0.37)
	2-D DWT	2.48 (0.27)	3.06 (0.30)	3.94 (0.28)	5.38 (0.41)	5.94 (0.32)	3.09 (0.29)	3.66 (0.28)	4.64 (0.31)	6.20 (0.46)	6.60 (0.31)
	PAWS	2.43 (0.27)	3.03 (0.32)	3.50 (0.32)	4.54 (0.35)	4.92 (0.37)	3.28 (0.28)	3.89 (0.34)	4.36 (0.35)	5.40 (0.38)	5.76 (0.39)
	AWS	3.04 (0.33)	3.86 (0.31)	4.96 (0.36)	5.94 (0.43)	7.03 (0.40)	3.30 (0.33)	4.10 (0.31)	5.24 (0.36)	6.38 (0.42)	7.38 (0.45)
	TPS	2.33 (0.26)	3.01 (0.25)	3.72 (0.29)	4.25 (0.42)	5.20 (0.35)	2.94 (0.22)	3.49 (0.21)	4.39 (0.28)	5.12 (0.40)	5.89 (0.33)
4096	DirW	**0.33 (0.08)**	**0.73 (0.11)**	**0.86 (0.10)**	**0.61 (0.12)**	**1.16 (0.14)**	**0.69 (0.15)**	1.33 (0.14)	1.51 (0.14)	**1.10 (0.19)**	**1.96 (0.20)**
	CURVELET	0.62 (0.08)	1.01 (0.12)	1.10 (0.09)	1.16 (0.13)	1.67 (0.12)	0.90 (0.08)	**1.32 (0.12)**	**1.46 (0.09)**	1.49 (0.13)	2.08 (0.12)
	2-D DWT	0.86 (0.11)	1.32 (0.13)	1.77 (0.15)	2.48 (0.32)	3.62 (0.18)	1.40 (0.13)	1.87 (0.14)	2.49 (0.16)	3.39 (0.40)	4.49 (0.19)
	PAWS	0.74 (0.09)	1.13 (0.12)	1.13 (0.10)	1.30 (0.15)	1.70 (0.12)	1.31 (0.10)	1.69 (0.13)	1.69 (0.10)	1.86 (0.16)	2.25 (0.13)
	AWS	1.16 (0.08)	1.91 (0.16)	2.19 (0.13)	2.32 (0.14)	3.19 (0.21)	1.34 (0.11)	2.09 (0.16)	2.41 (0.15)	2.61 (0.17)	3.50 (0.21)
	TPS	0.65 (0.06)	1.06 (0.10)	1.06 (0.08)	1.02 (0.08)	1.52 (0.09)	1.20 (0.07)	1.65 (0.09)	1.75 (0.08)	1.76 (0.10)	2.33 (0.11)

5.1.2 Equally-spaced observation sites

Directional wavelets can be constructed for the equally-spaced data as well and give comparable results with curvelets and AWS. We verify this by another simulation data whose observation sites are all the points of 32×32 grid over Ω. Based on the model (5.1), we generate 50 simulation datasets for the direction ϕ with both intensity types and estimate these underlying functions by the same methods used above. Figure 5.7 shows the true function and an example noise-contaminated ($\sigma = 0.05$) data with constant directional intensity.

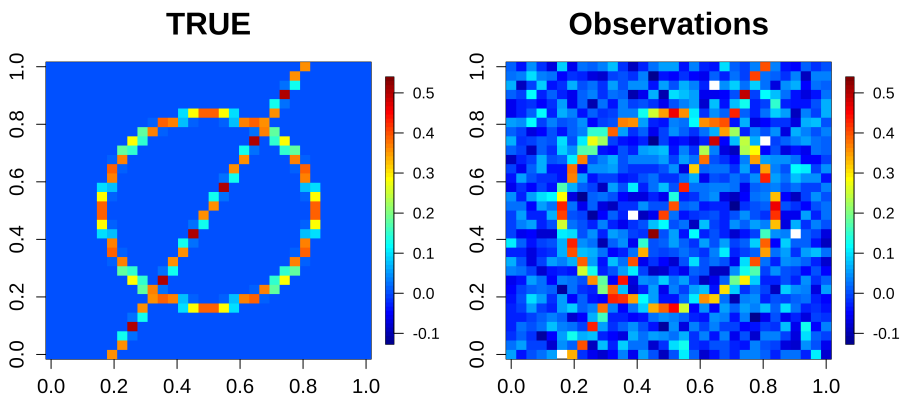

Figure 5.7: Plots of an underlying function (left) and its noise-contaminated observations (right). Directional shape: ϕ, Intensity: constant, Noise level: $\sigma = 0.05$

To define directional wavelets, we divide 1024 observations into three levels with sizes $N_1 = 64$, $N_2 = 196$ and $N_3 = 764$ using non-adaptive and adaptive thinning methods sequentially. Coefficients are estimated by least-square regression (or principal component regression in the case of multi-collinearity) and thresholded based on the predetermined values. When the intensity is constant, thresholding values are 0.0105 and 0.0040 for $\ell = 2$

and $\ell = 3$, respectively. When the intensity is non-constant, thresholding values are 0.0110 and 0.0040 for $\ell = 2$ and $\ell = 3$, respectively.

Table 5.5 shows the simulation results for constant and non-constant directional intensity, respectively, with directional wavelets and the other comparing methods. Though the state-of-the-art method PAWS gives the best result, the proposed method works better than 2-dimensional DWT, curvelet transform, AWS, and TPS. Figure 5.8 shows an example of the resultant estimate from each method when the intensity is constant.

Table 5.5: Simulation results for equally-spaced 32×32 observations from ϕ-shaped directional structure with two intensity types. Values presented in the table are averages (standard deviation in the parentheses) of mean squared error multiplied by 1000.

Intensity	Method					
	DirW	CURVELET	2-D DWT	PAWS	AWS	TPS
Constant	1.08 (0.08)	1.60 (0.09)	3.17 (0.29)	0.79 (0.14)	1.17 (0.13)	5.55 (0.28)
Non-constant	1.04 (0.07)	1.56 (0.08)	2.99 (0.28)	0.99 (0.24)	1.17 (0.12)	4.78 (0.27)

5.2 Real data analysis

5.2.1 Temperature data in South Korea

A large proportion of South Korea is covered with mountains, and because of this, a variety of climates appear despite the small territory. Taebaek Mountains, one of the major mountain ranges in South Korea, stretch along the east coast, and this geological factor makes this region have a distinctive climate from other regions. In particular, the mountainous regions have a cool climate in the summer, making temperature data have a directional structure. We have collected daily low temperatures of June, July, and August in 2018–2020 from 392 mainland (or near mainland) sta-

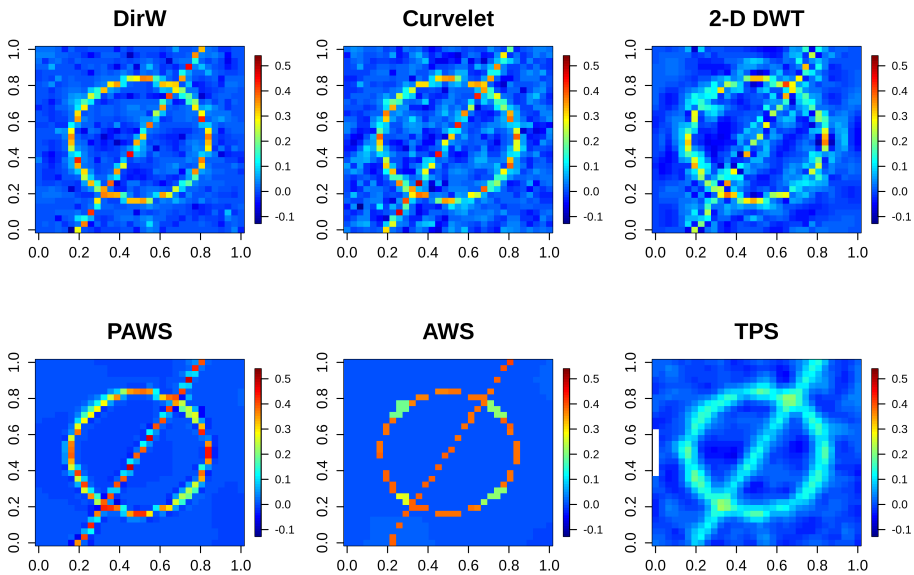

Figure 5.8: An example of the resultant estimate from each method when the intensity is constant

tions operated by the Automatic Weather System of Korea Meteorological Administration. Figure 5.9 illustrates the observatory location and the average value of collected temperature data on each site over the map of South Korea.

As expected, the northeastern region of the country shows a directional structure along with the mountain range. We intend to represent this through the proposed method and compare it with TPS based on a cross-validation approach. We consider TPS as a comparing method because it gives one of the best estimates other than directional wavelets in the simulation study and does not require equally-spaced observations on a rectangular domain as curvelets and PAWS do. For the comparison based on 10-fold cross-validation, we remove 12 observations near Seoul via non-adaptive thinning and permute the remaining 380 observations.

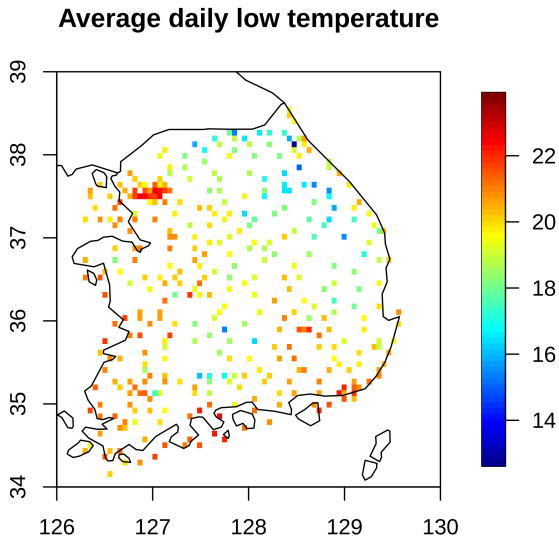

Figure 5.9: Three-year average of daily low temperature from June to August in South Korea

Then every 38 observations become the validation data, and the residuals are calculated from the estimates based on the training data. We repeat this process 20 times and obtain the cross-validation errors of the proposed method and TPS.

To construct the directional wavelets, we divide the 342 stations in the training data into two levels and construct the basis functions with different scales depending on the level they belong to. Locations of observatories are represented by longitude and latitude and we denote them by $\{x_{\ell,j}\}_{j=1}^{N_\ell}$ for the level $\ell = 1, 2$, where $N_1 = 75$, $N_2 = 267$. We construct the anisotropic Wendland functions $G_{\ell,S}(\cdot, x_{\ell,j})$ as illustrated in subsection 4.2.3 and orthogonalize them between the levels. Estimated coefficients will be denoised by the same thresholding value for each permutation, and we

use these denoised coefficients to represent the underlying structure with less noise.

Figure 5.10 shows the fitted result when both methods are applied to the entire data, and we can observe that the proposed method represents a more distinct direction. However, after 20 repetitions of the 10-fold cross-validation, we find that mean-squared errors of the proposed method and TPS are 1.077 and 0.915, respectively. (Standard deviations are 0.032 and 0.014, respectively.) We conjecture that the proposed method works poorly because it captures insignificant direction excessively and the temperature distribution does not have dramatic directional contrast as in the simulation setting. Based on this analysis, we conclude that our method still has room for improvement and the basis functions need to be constructed more elaborately and data-adaptively.

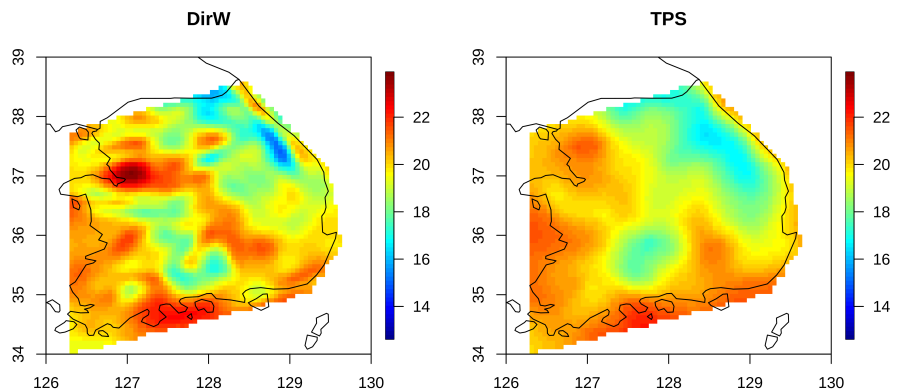

Figure 5.10: Fitted results from the proposed method and the thin-plate smoothing spline.

Chapter 6

Concluding Remarks

6.1 Summary of results

This thesis studies the problem of estimating a bivariate function on a plane when it has a directional structure and is observed on scattered sites with random noise. The proposed method employs the basis functions with anisotropic compact supports which are orthogonal across multiple resolution levels. The convergence property is derived for the interpolant when it is based on a single type of anisotropic Wendland functions. For implementation, we divide the scattered observations into several levels and construct basis functions while keeping them to reflect the directional structure and not to be clustered. Numerical study demonstrates the applicability and efficiency of the proposed method regardless of the directional structure and design of observation sites.

6.2 Future research

The proposed method has a few limitations, and we will introduce them as future research topics in this section. First, we choose some parameters of

the proposed method in an ad hoc sense in the proposed method. Thus, we need to identify an adaptive rule for choosing appropriate level-dependent threshold values and bandwidths.

Second, although the proposed method represents the directional structure very well, we find that it performs relatively poorly for the planar regions. We find this phenomenon through the representation of the temperature data in South Korea. We assume the main reason for this is the unnecessary and excessive reflection of the directional structure. To improve our method, we consider setting a threshold for the ratio of eigenvalues of a local covariance matrix. We will make the basis function anisotropic if the ratio exceeds the threshold and leave it as an isotropic radial basis function if the ratio does not exceed. We also consider calculating the local covariance matrix with weights based on the distance from the center. We expect these modifications to improve the current method by the less excessive and more accurate reflection of the direction.

Third, it would improve the directional representation if we construct a basis function that reflects a more general local direction. Though the proposed method can estimate a function with a nonlinear curved direction, it is natural to expect that we can represent it more efficiently if we bend the major axes of elliptic basis functions. So, we consider aligning the basis along the nonlinear direction by using the principal curve proposed by Hastie and Stuetzle (1989). We expect it will help us to determine how much the basis should be bent using the principal curve. However, it is hard to identify the convergence property when the curved basis functions are employed. Thus, in this thesis, we will leave this topic as future research.

Lastly, we intend to generalize the domain where the function lies. Our current method is limited to estimating a function on a domain in $\mathbb{R}^2$. This can further be extended to a sphere or even a Riemannian manifold. Once the method is generalized, we may be able to represent the directional

structure of the data on a broad domain without ignoring the Earth's curvature.

Bibliography

Candes, E., Demanet, L., Donoho, D., and Ying, L. (2005). Curvelab toolbox, version 2.0. *CIT*.

Candes, E., Demanet, L., Donoho, D., and Ying, L. (2006). Fast discrete curvelet transforms. *Multiscale Modeling & Simulation*, 5(3):861–899.

Candes, E. J. (1998). *Ridgelets: theory and applications*. Stanford University.

Candes, E. J. and Donoho, D. L. (1999). Curvelets: A surprisingly effective nonadaptive representation for objects with edges. Technical report, Stanford Univ Ca Dept of Statistics.

Daubechies, I. (1992). *Ten lectures on wavelets*. SIAM.

Donoho, D. L. (2000). Orthonormal ridgelets and linear singularities. *SIAM Journal on Mathematical Analysis*, 31(5):1062–1099.

Duffin, R. J. and Schaeffer, A. C. (1952). A class of nonharmonic fourier series. *Transactions of the American Mathematical Society*, 72(2):341–366.

Dyn, N., Floater, M. S., and Iske, A. (2001). Univariate adaptive thinning. In *Mathematical Methods for Curves and Surfaces: Oslo 2000*, pages 123–134. Vanderbilt University Press Nashville.

Fasshauer, G. E. (2007). *Meshfree approximation methods with MATLAB*, volume 6. World Scientific.

Fasshauer, G. E. and Jerome, J. W. (1999). Multistep approximation algorithms: Improved convergence rates through postconditioning with smoothing kernels. *Advances in Computational Mathematics*, 10(1):1–27.

Floater, M. S. and Iske, A. (1996). Multistep scattered data interpolation using compactly supported radial basis functions. *Journal of Computational and Applied Mathematics*, 73(1-2):65–78.

Floater, M. S. and Iske, A. (1998). Thinning algorithms for scattered data interpolation. *BIT Numerical Mathematics*, 38(4):705–720.

Franke, R. (1979). A critical comparison of some methods for interpolation of scattered data. Technical report, NAVAL POSTGRADUATE SCHOOL MONTEREY CA.

Gia, Q. L., Sloan, I. H., and Wendland, H. (2010). Multiscale analysis in sobolev spaces on the sphere. *SIAM journal on numerical analysis*, 48(6):2065–2090.

Gneiting, T. (2002). Compactly supported correlation functions. *Journal of Multivariate Analysis*, 83(2):493–508.

Hales, S. and Levesley, J. (2002). Error estimates for multilevel approximation using polyharmonic splines. *Numerical Algorithms*, 30(1):1–10.

Hastie, T. and Stuetzle, W. (1989). Principal curves. *Journal of the American Statistical Association*, 84(406):502–516.

Iske, A. (2004). *Multiresolution methods in scattered data modelling*, volume 37. Springer Science & Business Media.

Jansen, M. H. and Oonincx, P. J. (2005). *Second generation wavelets and applications*. Springer Science & Business Media.

Johnstone, I. M. and Silverman, B. W. (1997). Wavelet threshold estimators for data with correlated noise. *Journal of the royal statistical society: series B (statistical methodology)*, 59(2):319–351.

Li, T.-H. (1999). Multiscale representation and analysis of spherical data by spherical wavelets. *SIAM Journal on Scientific Computing*, 21(3):924–953.

Li, T.-H. (2001). Multiscale wavelet analysis of scattered spherical data: design and estimation. *Environmetrics: The official journal of the International Environmetrics Society*, 12(2):179–202.

Mallat, S. (1999). *A wavelet tour of signal processing*. Elsevier.

Mallat, S. G. (1989). Multiresolution approximations and wavelet orthonormal bases of ^{2} (). *Transactions of the American mathematical society*, 315(1):69–87.

Mallat, S. G. (2009). A theory for multiresolution signal decomposition: the wavelet representation. In *Fundamental Papers in Wavelet Theory*, pages 494–513. Princeton University Press.

Morlet, J., Arens, G., Fourgeau, E., and Glard, D. (1982). Wave propagation and sampling theory—part i: Complex signal and scattering in multilayered media. *Geophysics*, 47(2):203–221.

Narcowich, F. J., Schaback, R., and Ward, J. D. (1999). Multilevel interpolation and approximation. *Applied and Computational Harmonic Analysis*, 7(3):243–261.

Narcowich, F. J. and Ward, J. D. (1996). Nonstationary wavelets on themsphere for scattered data. *Applied and Computational Harmonic Analysis*, 3(4):324–336.

Oh, H.-S. (1999). *Spherical wavelets and their statistical analysis with applications to meteorological data*. Texas A&M University.

Polzehl, J., Papafitsoros, K., and Tabelow, K. (2020). Patch-wise adaptive weights smoothing in r. *Journal of Statistical Software*, 95(1):1–27.

Polzehl, J. and Spokoiny, V. G. (2000). Adaptive weights smoothing with applications to image restoration. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(2):335–354.

Schaback, R. (1997). On the efficiency of interpolation by radial basis functions.

Schröder, P. and Sweldens, W. (1995). Spherical wavelets: Efficiently representing functions on the sphere. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 161–172.

Sweldens, W. (1996). The lifting scheme: A custom-design construction of biorthogonal wavelets. *Applied and computational harmonic analysis*, 3(2):186–200.

Sweldens, W. (1998). The lifting scheme: A construction of second generation wavelets. *SIAM journal on mathematical analysis*, 29(2):511–546.

Wendland, H. (1995). Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Advances in computational Mathematics*, 4(1):389–396.

Wendland, H. (1998). Error estimates for interpolation by compactly supported radial basis functions of minimal degree. *Journal of approximation theory*, 93(2):258–272.

Wendland, H. (2004). *Scattered data approximation*, volume 17. Cambridge university press.

Wendland, H. (2010). Multiscale analysis in sobolev spaces on bounded domains. *Numerische Mathematik*, 116(3):493–517.

Wu, Z. (1995). Compactly supported positive definite radial functions. *Advances in Computational Mathematics*, 4(1):283.

Wu, Z.-m. and Schaback, R. (1993). Local error estimates for radial basis function interpolation of scattered data. *IMA journal of Numerical Analysis*, 13(1):13–27.

국문초록

2차원 공간에서 관측되는 비정상 자료는 그 공간적 비동질성이 1차원 곡선을 따라 나타난다. 이러한 방향적 특이성을 표현하기 위한 다중척도 방법론으로는 Candes and Donoho (1999)가 처음 제시한 커브렛 변환이 널리 알려져 있지만 이는 자료가 일정한 간격으로 관측되어야 한다는 제약이 있다. 한편 산재된 자료에 내재된 함수를 근사하기 위해서는 방사기저함수를 이용한 내삽법이 흔히 이용되지만 등방성이 있는 방사기저함수로는 방향성을 효율적으로 표현할 수 없다. 본 학위논문에서는 2차원 유클리드 공간에서 잡음과 함께 산재되어 관측되는 방향성 자료의 효율적인 표현을 위해 비등방성 방사기저함수를 이용한 새로운 다중척도 방법론을 제안한다. 이때 각 스케일에서 전반적인 방향성 구조와 국소적인 방향성 구조를 분리하여 표현하기 위해 기저함수의 스케일 간 직교화가 이루어진다. 제안된 방법이 산재된 방향성 자료를 표현하는 데 있어 우수함을 보이기 위해 모의실험과 실제 자료에 대한 수치실험을 한 결과를 제시하였다. 한편 제안된 방법의 수렴성과 실제 구현 방법에 관한 사안들도 다루었다.

주요어 : 다중척도 방법론, 방향성 자료, 비등방성 방사기저함수, 산재된 자료

학 번 : 2015–20286