WAVELET FRAMES ON THE SPHERE, HIGH ANGULAR RESOLUTION DIFFUSION IMAGING AND ℓ_1-REGULARIZED OPTIMIZATION ON STIEFEL MANIFOLDS

CHEN WEIQIANG

(*M.Sc.*, NUS)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF MATHEMATICS

NATIONAL UNIVERSITY OF SINGAPORE

2015

Declaration

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all sources of information which has been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Chen Weiqiang
14th December 2015

Acknowledgements

First and foremost, I would like to express my gratitude towards my supervisor, Professor Goh Say Song for his continued patience and guidance throughout these dedicated years of supervision. Next, I would like to thank my co-supervisor Professor Shen Zuowei for sharing his inspirational research insights and career advice.

I am thankful to Associate Professor Ji Hui and You Yanfei for a meaningful and enriching research collaboration. I have learned much about optimization and writing papers about Computational Mathematics through our collaboration.

I wish to thank Dr. Alvina Goh for her guidance in HARDI and her generous support in many ways. She often provides refreshing and candid advice to me.

I want to show my appreciation to all friends who have offered support in my PhD years, especially Dr. Wang Fei, Dr. Ku Cheng Yeaw and Dr. Jeffrey Pang. The PhD programme would be a lonelier journey if not for their companionship.

Finally, I wish to say a very big thank you to my family and in-laws for their understanding of my career switch. I also wish to express my heartfelt thanks to my dear wife, Jiawen, for her unwavering love and support towards my aspirations ever since we have been together; otherwise it would not be possible for me to devote fully to the PhD programme.

Contents

Summary

In the past two decades, wavelet frames are preferred over wavelet bases in image and signal processing applications as they yield redundant and flexible representations of square-integrable functions. As such, Chapter 1 provides preliminaries on MRA-based wavelet frames for $L^2(\mathbb{R})$, wavelet frame-based image restoration models. To facilitate the discussion in subsequent chapters, introductions to the spherical harmonic functions and Sobolev spaces on the sphere are also provided.

Building upon the wavelet bases constructions for Hilbert spaces in [**46**], Chapter 2 constructs tight wavelet frames for the space of (symmetric) square-integrable real-valued functions defined on the unit sphere, by considering special linear and weighted combinations of (modified) spherical harmonics.

In Chapter 3, we describe how these wavelet frames can be applied to denoise signals in High Angular Resolution Diffusion Imaging (HARDI) [**83**], a relatively recent non-invasive brain imaging technique. Tight framelet filters can also be used to impose spatial regularization of HARDI signals to improve denoising performances. The proposed wavelet frame-based approach generally denoises highly corrupted HARDI signals more cost-effectively than the spherical harmonics-based and spherical ridgelets-based approaches.

In Chapter 4, the HARDI denoising performances are further improved through adaptive spatial regularization, which can be modelled by optimization on Stiefel manifolds, i.e., orthogonality constrained problems. The resulting optimization problems are solved by the proximal alternating minimized augmented Lagrangian (PAMAL) method, which is a hybridization of the augmented Lagrangian method and the proximal alternating minimization method. Convergence analysis is also provided for the PAMAL method.

In Chapter 5, the PAMAL method is applied to a class of ℓ_1-regularized optimization problems with orthogonality constraints, which includes the compressed modes problem [**69**]. Convergence analysis of the PAMAL method is also provided in this case. Numerical results illustrate that the PAMAL method is noticeably faster than the splitting of orthogonality constraints (SOC) method [**53**] in producing compressed modes with comparable quality.

List of Tables

List of Figures

Notations

$\mathbb{S}^2$	The unit sphere in $\mathbb{R}^3$.
$\mathbb{Z}$	The set of all integers.
$\mathbb{N}_0$	The set of all natural numbers and the integer 0.
$\mathbb{R}^N_+$	The set of all $N \times 1$ column vectors with positive entries.
$\mathbb{R}^N_{\geq 0}$	The set of all $N \times 1$ column vectors with non-negative entries.
δ_{jk}	The Kronecker delta function.
$\equiv$	Equality of two functions pointwise up to a set of Lebesgue measure zero.
$L^2(\mathbb{S}^2)$	The space of all complex-valued square-integrable functions on $\mathbb{S}^2$.
$\ell^2(\mathbb{Z})$	The space of all complex square-summable sequences.
T^1_λ	The soft-thresholding operator with threshold vector λ.
$\operatorname{diag}(x)$	The diagonal matrix defined by $x \in \mathbb{R}^n$, i.e., $[\operatorname{diag}(x)]_{i,i} = x_i$, $i = 1, 2, \ldots, n$.
$\|X\|_F$	The Frobenius norm of $X \in \mathbb{R}^{m \times n}$, i.e., $\|X\|_F := \sqrt{\sum_{i,j} \|X_{i,j}\|^2}$.
$\|X\|_1$	The ℓ_1-norm of $X \in \mathbb{R}^{m \times n}$, i.e., $\|X\|_1 := \sum_{i,j} \|X_{i,j}\|$.
$\|X\|_\infty$	The max-norm of $X \in \mathbb{R}^{m \times n}$, i.e., $\|X\|_\infty := \max_{i,j} \|X_{i,j}\|$.
$A \succeq B$	The matrix $A - B$ satisfies $x^\top (A - B)x \geq 0$, $\forall x \in \mathbb{R}^n \backslash \{0\}$.
$A \succ B$	The matrix $A - B$ satisfies $x^\top (A - B)x > 0$, $\forall x \in \mathbb{R}^n \backslash \{0\}$.
$A \preceq B$	The matrix $B - A$ satisfies $x^\top (B - A)x \geq 0$, $\forall x \in \mathbb{R}^n \backslash \{0\}$.
$A \prec B$	The matrix $B - A$ satisfies $x^\top (B - A)x > 0$, $\forall x \in \mathbb{R}^n \backslash \{0\}$.
$\operatorname{Vec}(X)$	The $mn \times 1$ vector formed by concatenating column vectors of $X \in \mathbb{R}^{m \times n}$.
$X \otimes Y$	The Kronecker product of matrices X and Y.
$\operatorname{clos}_Y(X)$	The closure of a set X in the normed space Y.
χ_S	The characteristic function on a set $S \subseteq \mathbb{R}^n$, $n \in \mathbb{N}$.
δ_S	The indicator function on a set $S \subseteq \mathbb{R}^n$, $n \in \mathbb{N}$.
$N_S(x)$	The normal cone to a set S at the point x.
$\partial F(x)$	The Fréchet subdifferential of a function F at the point x.

CHAPTER 1

Preliminaries

To facilitate understanding of subsequent chapters, some background is provided in this chapter which is organized as follows. Section 1.1 briefly surveys how the multiresolution analysis (MRA) framework introduced by Mallat in [**59**] and the extension principles by [**75**] can be used to obtain (quasi-affine) wavelet frames for $L^2(\mathbb{R})$ and their corresponding wavelet filters for image processing.

Section 1.2 explains how the derived wavelet filters can be used in wavelet frame-based image restoration models. Due to its suitability for bio-imaging applications in subsequent chapters, this thesis focuses on the analysis-based model [**26**]. Its resulting optimization problem is then solved by the split Bregman method developed by Goldstein and Osher in [**49**].

In section 1.3, we give an introduction to the properties of the classical and modified spherical harmonics (SH), which form orthonormal bases for $L^2(\mathbb{S}^2)$ and $L^2_{\text{sym}}(\mathbb{S}^2)$ respectively, where $L^2(\mathbb{S}^2)$ is the space of square-integrable functions defined on $\mathbb{S}^2$, and $L^2_{\text{sym}}(\mathbb{S}^2)$ is the space of square-integrable antipodally symmetric functions defined on $\mathbb{S}^2$. These properties of the classical and modified SH will be useful to the construction of wavelet frames on the sphere in Chapters 2 and 3.

Finally, section 1.4 provides preliminaries and useful properties of the respective Sobolev spaces $H^s(\mathbb{S}^2)$ and $H^s_{\text{sym}}(\mathbb{S}^2)$, $s \in \mathbb{R}$, and their dual spaces. This background will be instrumental to establish that the wavelet systems used to represent HARDI signals in Chapter 3 are actually frame systems for some Sobolev space with a particular exponent.

1.1. Multiresolution Analysis-based Wavelet Frames for $L^2(\mathbb{R})$

In order to enhance understanding of the construction of wavelet frames on the unit sphere, in this section, we briefly review the concepts of multiresolution analysis (MRA) and wavelet frames for $L^2(\mathbb{R})$. The notion of an MRA was first introduced by Mallat [**59**] and Meyer [**63**] as a framework to yield easier constructions of orthonormal wavelet bases

for $L^2(\mathbb{R})$. In 1988, Daubechies [**36**] used the (orthonormal) MRA to construct families of compactly supported orthonormal wavelet bases.

MRA-based wavelet frames, especially MRA-based tight wavelet frames, can be viewed as a generalization of the MRA-based orthonormal wavelets of [**36**, **59**]. In [**11**], the notion of a frame MRA was formulated and it gave rise to the construction of bandlimited wavelet frames with narrow frequency bands. In [**75**], starting from a generalized MRA (we shall use this version in the current chapter), Ron and Shen developed extension principles to construct compactly supported tight wavelet frames for $L^2(\mathbb{R})$.

The MRA framework also enables efficient decomposition and reconstruction of signals and images, making wavelets a very powerful tool in image processing. In the past decade, it is used successfully in image denoising [**21**], image deblurring [**25**], and image inpainting [**22**], etc.

1.1.1. Multiresolution analysis (MRA).

A sequence of subspaces $\{V_k\}_{k\in\mathbb{Z}}$ in $L^2(\mathbb{R})$ is said to form a *multiresolution analysis* (MRA) for $L^2(\mathbb{R})$, if the following conditions are satisfied:

(MR1) $V_k \subset V_{k+1}$, $k \in \mathbb{Z}$;
(MR2) $\operatorname{clos}_{L^2(\mathbb{R})}\left(\bigcup_{k\in\mathbb{Z}} V_k\right) = L^2(\mathbb{R})$;
(MR3) $\bigcap_{k\in\mathbb{Z}} V_k = \{0\}$.

A common way of obtaining an MRA is to seek a *scaling function* $\phi \in L^2(\mathbb{R})$ and define its corresponding family of 2^{-k}-shift invariant subspaces

$$V_k := \operatorname{clos}_{L^2(\mathbb{R})}\left(\operatorname{span}\left\{\phi(2^k\cdot - j) \,:\, j\in\mathbb{Z}\right\}\right), \quad k\in\mathbb{Z}. \tag{1.1.1}$$

The problem then becomes finding conditions on the scaling function ϕ that allows $\{V_k\}_{k\in\mathbb{Z}}$ to satisfy (MR1), (MR2) and (MR3).

To ensure that (MR1) holds, note that the generator ϕ lies in V_0. Therefore, to have $V_0 \subset V_1$, it is natural for ϕ to be refinable, i.e., a function $\phi \in L^2(\mathbb{R})$ is *refinable* if

$$\phi = 2\sum_{j\in\mathbb{Z}} h_0[j]\phi(2\cdot - j),$$

for some sequence $h_0 \in \ell^2(\mathbb{Z})$. The sequence h_0 is called a *refinement mask* of ϕ. In the Fourier domain, the refinability of ϕ can be expressed as

$$\hat{\phi}(2\cdot) = \hat{h}_0\hat{\phi},$$

where $\hat{\phi}$ denotes the Fourier transform of ϕ, and $\hat{h}_0$ denotes the Fourier series of the sequence h_0. The Fourier transform of a function $f \in L^1(\mathbb{R})$ is defined by

$$\hat{f}(\xi) = \int_{\mathbb{R}} f(x) \exp(-\mathrm{i}\xi x)\, dx, \quad \xi \in \mathbb{R},$$

which can be extended to $L^2(\mathbb{R})$ in the usual manner. The Fourier series of a sequence $h_0 \in \ell^2(\mathbb{Z})$ is defined as

$$\hat{h}_0(\xi) = \sum_{j\in\mathbb{Z}} h_0[j] \exp(-\mathrm{i}j\xi), \quad \xi \in \mathbb{R}.$$

Examples of refinable functions include B-splines. A (centered) *B-spline* of order m, denoted as B_m, is defined in the Fourier domain as

$$\hat{B}_m(0) = 1,\ \hat{B}_m(\xi) = \exp(-\mathrm{i}\sigma_m\xi/2) \left(\frac{\sin(\xi/2)}{(\xi/2)}\right)^m, \quad \xi \in \mathbb{R}\backslash\{0\},$$

where $\sigma_m = 0$ when m is even, and $\sigma_m = 1$ when m is odd. The Fourier series of the corresponding refinement mask of B_m is

$$\hat{h}_0(\xi) = \exp(-\mathrm{i}\sigma_m\xi/2) \cos^m(\xi/2), \quad \xi \in \mathbb{R}.$$

The B-spline B_m is a compactly supported function in $C^{m-2}(\mathbb{R})$ with the length of its support equals to m. The interested reader may refer to [**19**] for more detailed discussions about B-splines.

As for (MR2), we recall an important theorem first proved in [**20**].

THEOREM 1.1.1. *Let $\{V_k\}_{k\in\mathbb{Z}}$ be a nested sequence of subspaces in $L^2(\mathbb{R})$, i.e.,* (MR1) *is satisfied. Then,*

$$\mathrm{clos}_{L^2(\mathbb{R})}\left(\bigcup_{k\in\mathbb{Z}} V_k\right) = L^2(\mathbb{R})$$

if and only if $\Omega_0 := \bigcup_{k\in\mathbb{Z}} \mathrm{supp}\, \hat{\phi}(2^k\cdot) = \mathbb{R}$ (modulo a null set).

Note that any refinable $\phi \in L^2(\mathbb{R})$ with $\hat{\phi}$ continuous and non-zero at the origin satisfies the above theorem. This includes compactly supported refinable functions ϕ with $\hat{\phi}(0) = 1$, e.g., B-splines.

It turns out that the (MR3) condition can be automatically satisfied, as shown in the following theorem from [**20**].

THEOREM 1.1.2. *Given any $\phi \in L^2(\mathbb{R})$, for $\{V_k\}_{k\in\mathbb{Z}}$ defined by ϕ as in* (1.1.1)*, there holds*

$$\bigcap_{k\in\mathbb{Z}} V_k = \{0\}\,.$$

Daubechies [**36**] started with an MRA generated by a class of orthonormal refinable functions which she constructed, and thereafter obtained families of compactly supported orthonormal wavelet bases for $L^2(\mathbb{R})$, now commonly known as Daubechies' wavelets. The Daubechies' wavelet bases have been used extensively in applications including image compression, image and signal denoising, etc. However, in applications such as image and signal denoising, it is actually even more advantageous to use more functions than necessary to represent the image or signal. This prompted intense research efforts in the development of alternative function representation systems such as the wavelet frames, which are explained in the next subsection.

1.1.2. Tight wavelet (affine) frames for $L^2(\mathbb{R})$.

In this subsection, we first introduce the idea of a (tight) wavelet frame for $L^2(\mathbb{R})$, with other basic concepts and notations. We then describe how the MRA could be used to construct tight wavelet frames for $L^2(\mathbb{R})$.

We say that a countable set of functions X forms a *frame* for $L^2(\mathbb{R})$ with frame bounds C_1 and C_2 if the inequality

$$C_1\|f\|^2 \leq \sum_{g\in X} |\langle f, g\rangle|^2 \leq C_2\|f\|^2 \tag{1.1.2}$$

holds for all $f \in L^2(\mathbb{R})$. When $C_1 = C_2 = 1$, X is said to form a *normalized tight frame* for $L^2(\mathbb{R})$, which also yields the Parseval's identity

$$f = \sum_{g\in X} \langle f, g\rangle g, \tag{1.1.3}$$

for all $f \in L^2(\mathbb{R})$. From (1.1.3), it is evident that an orthonormal basis for $L^2(\mathbb{R})$ is also a normalized tight frame for $L^2(\mathbb{R})$.

For a given set of functions $\Psi := \{\psi_1, \ldots, \psi_r\} \subset L^2(\mathbb{R})$, define the *wavelet system* (or *affine system*) as

$$X(\Psi) := \{\psi_{l,k,j} \,:\, 1 \leq l \leq r;\, k, j \in \mathbb{Z}\}, \tag{1.1.4}$$

where $\psi_{l,k,j} := 2^{k/2}\psi_l(2^k \cdot -j)$ is a dilated and translated version of a single function f (with $f = \psi_l$), which also can be denoted as

$$f_{k,j} = 2^{k/2} f(2^k \cdot -j), \quad k, j \in \mathbb{Z}.$$

The system $X(\Psi)$ is called a *wavelet frame* (or also commonly known as *an affine frame* or *framelets*) for $L^2(\mathbb{R})$ if $X(\Psi)$ satisfies (1.1.2). The system $X(\Psi)$ is a *normalized tight wavelet*

frame (or also commonly known as *tight affine frame* or *tight framelets*) if the frame bounds in (1.1.2) are both equal to one.

Compactly supported tight framelets constructed from an MRA are very handy to use, because of the existence of fast decomposition and reconstruction algorithms. This motivates the study of MRA-based tight wavelet frames in [**75**] as we shall present next, where the key lies in extension principles, e.g., the *unitary extension principle* (UEP).

Let $\{V_k\}_{k\in\mathbb{Z}}$ be the MRA generated by a refinable function ϕ with refinement mask h_0. The construction of normalized tight wavelet frame systems starts with the constructon of $\Psi \subset L^2(\mathbb{R})$. The idea of MRA-based construction of normalized tight wavelet frames is to find $\Psi = \{\psi_1, \ldots, \psi_r\} \subset V_1$ such that $X(\Psi)$ forms a normalized tight frame for $L^2(\mathbb{R})$. Since V_1 is a $\frac{1}{2}$-shift invariant subspace generated by $\phi(2\cdot)$, finding $\Psi \subset V_1$ is the same as finding sequences h_l, $l = 1, \ldots, r$ such that

$$\psi_l = 2\sum_{j\in\mathbb{Z}} h_l[j]\phi(2\cdot - j), \quad l = 1, \ldots, r. \tag{1.1.5}$$

The sequences $h_1, \ldots, h_r$ are called *wavelet masks*, or the high-pass filters of the system, and the refinement mask h_0 is also known as the low-pass filter. In the Fourier domain, (1.1.5) can be written as

$$\hat{\psi}_l(2\cdot) = \hat{h}_l\hat{\phi}, \quad l = 1, \ldots, r, \tag{1.1.6}$$

where $\hat{h}_1, \ldots, \hat{h}_r$ are 2π-periodic functions.

The univariate version of the UEP of [**75**] can be stated as follows.

THEOREM 1.1.3. (Unitary Extension Principle (UEP) [**75**]). *Let $\phi \in L^2(\mathbb{R})$ be a refinable function with refinement mask h_0 and $\{h_1, \ldots, h_r\}$ be a set of wavelet masks. Assume that the refinable function ϕ and the masks $\{h_0, h_1, \ldots, h_r\}$ satisfy the following conditions:*

(a) *Each mask in $\{h_l : l = 0, 1, \ldots, r\}$ is a sequence in $\ell^2(\mathbb{Z})$ and its Fourier series $\hat{h}_l$ is measurable and (essentially) bounded.*

(b) *The refinement mask h_0 satisfies $|\hat{h}_0(\xi) - 1| \le C|\xi|$, $\xi \in \mathbb{R}$.*

(c) *The function $[\hat{\phi}, \hat{\phi}] := \sum_{l\in\mathbb{Z}} |\hat{\phi}(\cdot + 2\pi l)|^2$ is essentially bounded.*

Then the system $X(\Psi)$ given by (1.1.4), where $\Psi = \{\psi_1, \ldots, \psi_r\}$ defined in (1.1.6), forms a normalized tight frame for $L^2(\mathbb{R})$ provided that the equations

$$\sum_{l=0}^{r} |\hat{h}_l(\xi)|^2 = 1 \quad \textit{and} \quad \sum_{l=0}^{r} \hat{h}_l(\xi)\overline{\hat{h}_l(\xi+\pi)} = 0 \tag{1.1.7}$$

hold for almost all $\xi \in \sigma(V_0)$, *where*

$$\sigma(V_0) := \left\{\xi \in \mathbb{R} : [\hat{\phi}, \hat{\phi}](\xi) \neq 0\right\}.$$

Furthermore, if $r = 1$ *and* $\|\phi\|_{L^2(\mathbb{R})} = 1$, *then* $X(\Psi)$ *is an orthonormal wavelet basis for* $L^2(\mathbb{R})$.

In practice, a given function is usually decomposed to a certain level instead of down to negative infinity. The following corollary states that when one only decomposes a function down to some given level K, then the system

$$X(\phi, \Psi; K) := \{\phi_{K,j}, \psi_{l,k,j} : 1 \leq l \leq r, k \geq K, j \in \mathbb{Z}\} \tag{1.1.8}$$

forms a normalized tight frame for $L^2(\mathbb{R})$. The proof of the corollary follows directly from the well known result that

$$\sum_{j\in\mathbb{Z}} \langle f, \phi_{K,j}\rangle \phi_{K,j} = \sum_{l=1}^{r} \sum_{k<K} \sum_{j\in\mathbb{Z}} \langle f, \psi_{l,k,j}\rangle \psi_{l,k,j},$$

and that $X(\Psi)$ is a normalized tight frame for $L^2(\mathbb{R})$.

COROLLARY 1.1.4. *Let* $\Psi = \{\psi_l : 1 \leq l \leq r\}$ *be the set of functions constructed from the UEP with* ϕ *as the corresponding refinable function. Then for any given* $K \in \mathbb{Z}$, *the system* $X(\phi, \Psi; K)$ *defined in* (1.1.8) *forms a normalized tight frame for* $L^2(\mathbb{R})$, *i.e., for any* $f \in L^2(\mathbb{R})$,

$$f = \sum_{j\in\mathbb{Z}} \langle f, \phi_{K,j}\rangle \phi_{K,j} + \sum_{l=1}^{r} \sum_{k\geq K} \sum_{j\in\mathbb{Z}} \langle f, \psi_{l,k,j}\rangle \psi_{l,k,j}.$$

The UEP condition (1.1.7) means that the matrix

$$M(\xi) := \begin{pmatrix} \hat{h}_0(\xi) & \hat{h}_1(\xi) & \dots & \hat{h}_r(\xi) \\ \hat{h}_0(\xi+\pi) & \hat{h}_1(\xi+\pi) & \dots & \hat{h}_r(\xi+\pi) \end{pmatrix}$$

has its two rows orthonormal to each other, for a.e. $\xi \in \mathbb{R}$.

Next, we present how B-spline-based tight wavelet frames can be constructed. Consider a (centered) B-spline of order m. Recall that the corresponding refinement mask $\hat{h}_0$ is given by

$$\hat{h}_0(\xi) = \exp(-\mathrm{i}\sigma_m \xi/2) \cos^m(\xi/2),$$

with $\sigma_m = 0$ when m is even and $\sigma_m = 1$ when m is odd. We then define m wavelet masks as

$$\hat{h}_l(\xi) := -\mathrm{i}^l \exp(-\mathrm{i}\sigma_m \xi/2) \sqrt{\binom{m}{l}} \sin^l(\xi/2) \cos^{m-l}(\xi/2), \quad l = 1, \dots, m.$$

It can be shown that the scaling function as well as the refinement and wavelet masks satisfy the conditions of the UEP. Therefore, by (1.1.6), the m wavelets ψ_l, $l = 1, \ldots, m$, are given respectively in the Fourier domain as

$$\hat{\psi}_l(\xi) = -\mathrm{i}^l \exp(-\mathrm{i}\sigma_m \xi/2) \sqrt{\binom{m}{l}} \frac{\cos^{m-l}(\xi/4) \sin^{m+l}(\xi/4)}{(\xi/4)^m}, \ \xi \neq 0, \ \hat{\psi}_l(0) = 0,$$

and they generate a normalized tight frame for $L^2(\mathbb{R})$. The cases for $m = 1$ and $m = 2$ are given in the following two examples. Note that in the wavelet literature, a *finite filter* $h = \{h(n)\}_{n\in\mathbb{Z}} \in \ell^2(\mathbb{Z})$ is commonly defined with a vector $v := [v_0 \ldots v_N]$ for some $N \in \mathbb{N}$ as follows:

$$h(n) = v_n, \ 0 \leq n \leq N, \quad h(n) = 0, \text{ otherwise.} \tag{1.1.9}$$

EXAMPLE 1.1.5. (Haar orthonormal wavelet basis) Let h_0 be a finite filter defined by $[\frac{1}{2}, \frac{1}{2}]$ according to (1.1.9). Note that h_0 is the refinement mask of the characteristic function on $[0, 1]$, i.e., $B_1(x) = \chi_{[0,1]}(x)$, $x \in \mathbb{R}$. Let h_1 be another finite filter defined by $[\frac{1}{2}, -\frac{1}{2}]$ according to (1.1.9). Then h_0 and h_1 satisfy the conditions of the UEP. Since $r = 1$, with $\Psi := \{\psi_1\}$ defined in (1.1.5) by h_1 and ϕ, the system $X(\Psi)$ is an orthonormal wavelet basis for $L^2(\mathbb{R})$ by Theorem 1.1.3. This system is also widely known as the *Haar orthonormal wavelet basis*. See Figure 1.1.1 for an illustration of the scaling function ϕ and wavelet function ψ_1.

EXAMPLE 1.1.6. Let h_0 be a finite filter defined by $[\frac{1}{4}, \frac{1}{2}, \frac{1}{4}]$ according to (1.1.9). Note that here, h_0 is the refinement mask of the piecewise linear B-spline $B_2(x) = \max(1 - |x|, 0)$, $x \in \mathbb{R}$. Define respectively the wavelet masks h_1 and h_2 according to (1.1.9) by the vectors

$$[-\frac{1}{4}, \frac{1}{2}, -\frac{1}{4}] \quad \text{and} \ [\frac{\sqrt{2}}{4}, 0, -\frac{\sqrt{2}}{4}].$$

Then h_0, h_1 and h_2 satisfy the conditions of the UEP. Hence, the corresponding system $X(\Psi)$ is a normalized tight wavelet frame for $L^2(\mathbb{R})$, where $\Psi := \{\psi_1, \psi_2\}$ defined in (1.1.5) with h_1, h_2 and ϕ. See Figure 1.1.2 for an illustration of the scaling function ϕ and wavelet functions ψ_1, ψ_2.

1.1.3. Quasi-affine wavelet systems.

Notice that the wavelet (affine) system $X(\Psi)$ defined by (1.1.4) is not shift invariant. Recall that a system X, that contains countably many elements, is *τ-shift invariant* with $\tau \in \mathbb{R}$, if for any $k \in \mathbb{Z}$, $g \in X$, we have $g(\cdot - \tau k) \in X$. In particular, if a system is 1-shift invariant, it is simply called *shift invariant*.

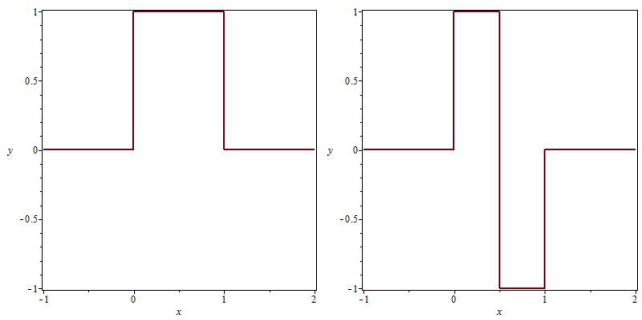

FIGURE 1.1.1. Piecewise constant refinable spline $\phi = B_1$ and corresponding wavelet generator ψ_1.

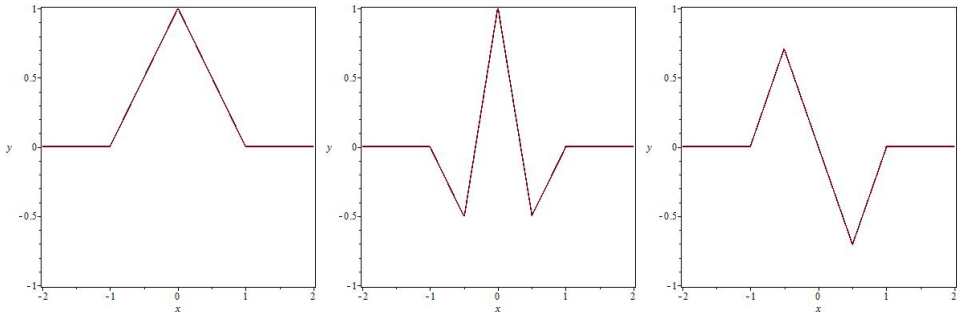

FIGURE 1.1.2. Piecewise linear refinable spline $\phi = B_2$ and corresponding wavelet frame generators ψ_1, ψ_2.

However, in the context of signal and image processing, it is usually preferred to use wavelet systems that are shift invariant. For instance, the associated wavelet filters enable more effective capture of adjacent spatial information around each pixel of processed images. In order to achieve shift invariance, we need to over-sample the affine system $X(\Psi)$ below level 0. This over-sampled system is called a quasi-affine system, which was first introduced in [**75**]. Its resultant wavelet transform is also known as the undecimated wavelet transform which is shown to be very effective in image restoration.

DEFINITION 1.1.7. Let $\Psi := \{\psi_1, \ldots, \psi_r\}$ be a set of functions. A *quasi-affine system* from level K is defined as

$$X^q(\Psi; K) := \left\{ \psi^q_{l,k,j} : 1 \le l \le r, k, j \in \mathbb{Z} \right\},$$

where $\psi^q_{l,k,j}$ is defined by

$$\psi^q_{l,k,j} := \begin{cases} \mathcal{D}^k T_j \psi_l, & k \geq K, \\ 2^{\frac{k-K}{2}} T_{2^{-K}j} \mathcal{D}^k \psi_l, & k < K, \end{cases} \tag{1.1.10}$$

where for a given function $f \in L^2(\mathbb{R})$, the dyadic dilation operator $\mathcal{D}$ is defined by $\mathcal{D}f := \sqrt{2} f(2\cdot)$ and the translation operator T_t is defined by $T_t f := f(\cdot - t)$ for $t \in \mathbb{R}$. Note that in contrast to the quasi-affine system $X^q(\Psi; K)$, the affine system $X(\Psi)$ is of the form $\{\mathcal{D}^k T_j \psi_l : 1 \leq l \leq r; j, k \in \mathbb{Z}\}$. Similar to the quasi-affine system $X^q(\Psi; K)$ which is spanned by wavelet functions defined in (1.1.10), we also define $V_k^{q,K} := \operatorname{span}\{\phi^q_{k,j} : j \in \mathbb{Z}\}$, where

$$\phi^q_{k,j} := \begin{cases} \mathcal{D}^k T_j \phi, & k \geq K, \\ 2^{\frac{k-K}{2}} T_{2^{-K}j} \mathcal{D}^k \phi, & k < K. \end{cases} \tag{1.1.11}$$

The quasi-affine system is obtained by over-sampling the wavelet frame system starting from level $K-1$ and below. Hence, the entire quasi-affine system is a 2^{-K}-shift invariant system. The quasi-affine system from level 0 was first introduced in [**75**] to convert a non-shift invariant system to a shift invariant system. Furthermore, it was shown in [**75**, Theorem 5.5] that a wavelet system $X(\Psi)$ is a tight frame for $L^2(\mathbb{R})$ if and only if its corresponding quasi-affine counterpart $X^q(\Psi; K)$ is a tight frame for $L^2(\mathbb{R})$, for any $K \in \mathbb{Z}$. We also have the following result.

THEOREM 1.1.8. *Let $X(\Psi)$, where $\Psi := \{\psi_1, \ldots, \psi_r\}$, be the affine tight frame system obtained from the UEP with the corresponding refinable function ϕ. Then, for any given $K \in \mathbb{Z}$, the quasi-affine system $X^q(\Psi; K)$ is a normalized tight frame for $L^2(\mathbb{R})$. Furthermore,*

$$\sum_{j\in\mathbb{Z}} \langle f, \phi^q_{K,j}\rangle \phi^q_{K,j} = \sum_{l=1}^{r} \sum_{k<K} \sum_{j\in\mathbb{Z}} \langle f, \psi^q_{l,k,j}\rangle \psi^q_{l,k,j},$$

where $\phi^q_{K,j}$, $j \in \mathbb{Z}$, is given by (1.1.11) *and thus*

$$X^q(\phi, \Psi; K) := \left\{\phi^q_{K,j}, \psi^q_{l,k,j} : 1 \leq l \leq r, k \geq K, j \in \mathbb{Z}\right\}$$

is also a normalized tight frame for $L^2(\mathbb{R})$, i.e., for any $f \in L^2(\mathbb{R})$,

$$f = \sum_{j\in\mathbb{Z}} \langle f, \phi^q_{K,j}\rangle \phi^q_{K,j} + \sum_{l=1}^{r} \sum_{k\geq K} \sum_{j\in\mathbb{Z}} \langle f, \psi^q_{l,k,j}\rangle \psi^q_{l,k,j},$$

for all $f \in L^2(\mathbb{R})$.

REMARK 1.1.9. (Haar Framelets) When the Haar orthonormal wavelet basis $X(\Psi)$ in Example 1.1.5 is chosen as the affine tight frame system in Theorem 1.1.3, then both its corresponding quasi-affine wavelet systems $X^q(\Psi;K)$ and $X^q(\phi,\Psi;K)$ form normalized tight wavelet frame systems (but not orthonormal wavelet bases!) for $L^2(\mathbb{R})$. These quasi-affine wavelet systems are also known as *Haar framelets*.

In practical calculations, K is usually set as 0, and the decomposition of the function $\sum_{j\in\mathbb{Z}}\langle f,\phi^q_{0,j}\rangle\phi^q_{0,j}$ (approximating $f\in L^2(\mathbb{R})$) is performed L times, $L\in\mathbb{N}_0$. Then the decomposition and reconstruction formula can be written as

$$\sum_{j\in\mathbb{Z}}\langle f,\phi^q_{0,j}\rangle\phi^q_{0,j}=\sum_{j\in\mathbb{Z}}\langle f,\phi^q_{-L,j}\rangle\phi^q_{-L,j}+\sum_{l=1}^{r}\sum_{k=-L}^{-1}\sum_{j\in\mathbb{Z}}\langle f,\psi^q_{l,k,j}\rangle\psi^q_{l,k,j}.$$

This process corresponds to the so-called undecimated wavelet decomposition and reconstruction which we will briefly describe in the next subsection.

1.1.4. Algorithms for quasi-affine tight frames.

For simplicity, in this thesis, we consider only framelet algorithms designed for periodically extended signal $\tilde{v}$ from a finite length signal v, i.e.,

$$\tilde{v}[j]:=v[j \bmod N],$$

where N is the length of the finite signal v, $N\in\mathbb{N}$. The framelet filters $\{h_0,h_1,\dots,h_r\}$ are then applied to perform decomposition on the extended signal $\tilde{v}$, by defining the dilated filters $h_{l,k}$, $k\le 0$,

$$h_{l,k}[j]=\begin{cases}h_l[2^k j], & \text{if } j\in 2^{-k}\mathbb{Z},\\ 0, & \text{if } j\in\mathbb{Z}\backslash 2^{-k}\mathbb{Z},\end{cases}$$

and the corresponding circulant filter matrices $H_{l,k}$

$$H_{l,k}:=(H_{l,k}[j,j']):=(\overline{h_{l,k}[j'-j]} \bmod N),$$

where the (j,j')-th entry in $H_{l,k}$ is fully determined by the $(j-j')$th entry in $h_{l,k}$. Then for any N-periodic vector $\tilde{v}$, we have

$$(H_{l,k}\tilde{v})[j \bmod N]=\sum_{j'\in\mathbb{Z}}\overline{h_{l,k}[j'-j]}\tilde{v}[j].$$

It can be shown that

$$\sum_{l=0}^{r}\sum_{j\in\mathbb{Z}}\overline{h_{l,k}[j]}h_{l,k}[j-j']=\delta_{0,j'},\quad j'\in\mathbb{Z},$$

which is equivalent to

$$\sum_{l=0}^{r} H_{l,k}^* H_{l,k} = I \quad \text{or} \quad W_k^* W_k = I,$$

where

$$W_k := \begin{bmatrix} H_{0,k} \\ H_{1,k} \\ \vdots \\ H_{r,k} \end{bmatrix} \in \mathbb{R}^{(r+1)N \times N}.$$

In the case where the Haar filters h_0, h_1 are used, then

$$W_0 = \frac{1}{2} \begin{pmatrix} 1 & & & & 1 \\ 1 & 1 & & & \\ & \ddots & \ddots & & \\ & & & 1 & 1 \\ -1 & & & & 1 \\ 1 & -1 & & & \\ & \ddots & \ddots & & \\ & & & 1 & -1 \end{pmatrix} \in \mathbb{R}^{2N \times N}.$$

Therefore, a single-level decomposition process on the signal $\tilde{v}$ by the matrix W_k yields a single low-pass component $H_{0,k}\tilde{v}$ and r high-pass components $H_{1,k}\tilde{v}, \ldots, H_{r,k}\tilde{v}$. As for the corresponding single-level reconstruction step, simply multiply the matrix $H_{l,k}^*$ to the component $H_{l,k}\tilde{v}$ for each l and sum the resultant vectors together to restore $\tilde{v}$, i.e.,

$$\tilde{v} = \sum_{l=0}^{r} H_{l,k}^* (H_{l,k}\tilde{v}).$$

We now provide the L-level quasi-affine framelet decomposition and reconstruction algorithm.

Quasi-affine Framelet Decomposition and Reconstruction Algorithm

---

Given a signal $v \in \mathbb{R}^N$ with $N \in \mathbb{N}$, denote $v_{0,0} := \tilde{v}$ with $\tilde{v}$ being the N-periodic extension of v. Then the L-level quasi-affine framelet decomposition and reconstruction are given as follows:

(1) **Decomposition**: For each $k = 1, 2, \ldots, L$,

(a) Obtain low frequency approximation to $\tilde{v}$ at level k:

$$\tilde{v}_{0,k} = H_{0,1-k}\tilde{v}_{0,k-1};$$

(b) Obtain framelet coefficients of $\tilde{v}$ at level k:

$$\tilde{v}_{l,k} = H_{l,1-k}\tilde{v}_{0,k-1}, \quad l = 1, 2, \ldots, r;$$

(2) **Reconstruction**: For each $k = L, L-1, \ldots, 1$,

$$\tilde{v}_{0,k-1} = \sum_{l=0}^{r} H^*_{l,1-k}\tilde{v}_{l,k}.$$

REMARK 1.1.10. In the ideal scenario, the above algorithm performs a perfect recovery of the signal $\tilde{v}$ from its decomposition components $\tilde{v}_{l,k}$, $l = 1, 2, \ldots, r$, $k = 1, 2, \ldots, L$. However, if the signal $\tilde{v}$ is corrupted with noise, a thresholding step, involving a thresholding operator is usually performed on the decomposition components $\tilde{v}_{l,k}$ (to filter out noise) before the reconstruction step. In this thesis, we focus on the soft-thresholding operator T^1_α which is applied to a $n \times m$ matrix X with threshold level vector $\alpha := (\alpha_1, \ldots, \alpha_n)^\top$, defined by

$$T^1_\alpha(X) := [T^1_{\alpha_i}(X_{i,j})]_{i,j}, \text{ where } T^1_{\alpha_i}(X_{i,j}) = \text{sign}(X_{i,j}) \cdot \max(|X_{i,j}| - \alpha_i, 0), \forall i, j. \quad (1.1.12)$$

However, during the past decade, in order to achieve more effective signal/image denoising, the above algorithm is gradually replaced by optimization models (such as the wavelet frame-based restoration models) which are equipped with an in-built thresholding mechanism. We shall discuss these optimization models for image restoration in the next section.

1.2. ℓ_1-Regularized Wavelet Frame-based Image Restoration

Image restoration is often formulated as an inverse problem. For ease of notation, images are denoted as vectors in $\mathbb{R}^n$ with n being the total number of pixels. The aim is to approximate as best as possible the unknown true image/signal $u \in \mathbb{R}^n$ from an observed image (or measurements) $f \in \mathbb{R}^l$ defined by

$$f = Au + e, \quad (1.2.1)$$

where e is a vector of white Gaussian noise (each entry e_i is normally distributed with mean 0) with variance σ^2, and $A \in \mathbb{R}^{l \times n}$ is a linear operator, typically a convolution operator for image deblurring problems, a projection operator for image inpainting and the identity operator for image denoising.

As discussed in the previous subsection, the framelet decomposition and reconstruction algorithms involve linear operators. When signals are considered to be in $\mathbb{R}^n$, these linear operators have matrix representations. Here we simply denote the framelet decomposition

as a matrix $W \in \mathbb{R}^{m\times n}$ with $m \geq n$, and reconstruction as $W^\top$. Recall from the previous subsection that $W^\top W = I$, thus for every vector $u \in \mathbb{R}^n$,

$$u = W^\top (Wu).$$

The components of the vector Wu are called the canonical coefficients representing u. The matrix W is generated from the masks $\{h_l : l = 0, 1, \ldots, r\}$ constructed from the univariate UEP. It should be emphasized that the image is viewed as a column vector and the corresponding tight wavelet frame transform as the matrix W merely for simplicity.

This section is organized as follows. Subsection 1.2.1 introduces wavelet frame-based image restoration models where the tight wavelet frame transform plays a central role. Particular emphasis is placed on the analysis-based model [**26**] as it is well suited for bio-imaging applications in subsequent chapters of this thesis. In order to solve the resulting optimization problems from the analysis-based model, subsection 1.2.2 provides details on the split Bregman method, which is developed by Goldstein and Osher in [**49**].

1.2.1. Wavelet frame-based image restoration models.

Since tight wavelet frame systems are redundant systems (i.e., $m \geq n$), the mapping from the image u to its coefficients is not one-to-one, i.e., the representation of u in the frame domain is not unique. Therefore, there are three formulations for the sparse approximation of the underlying image, namely the analysis-based approach, the synthesis-based approach and the balanced approach, which can be integrated into a single minimization problem as follows:

$$\min_{\alpha\in\mathbb{R}^m} \frac{1}{2}\|AW^\top\alpha - f\|_D^2 + \frac{\kappa}{2}\|(I - WW^\top)\alpha\|_2^2 + \|\mathrm{diag}(\lambda)\alpha\|_1, \tag{1.2.2}$$

where $0 \leq \kappa \leq \infty$, $\|x\|_D := \sqrt{x^\top Dx}$, D is some appropriately chosen symmetric positive definite matrix and λ is a given vector defined as

$$\lambda := (\lambda_1, \ldots, \lambda_m)^\top.$$

The model (1.2.2) is called the (*single-system*) *general balanced approach.*

When $0 < \kappa < \infty$, the system (1.2.2) is called the *balanced approach.* When $\kappa = 0$, the system (1.2.2) is reduced to a *synthesis-based approach:*

$$\min_{\alpha\in\mathbb{R}^m} \frac{1}{2}\|AW^\top\alpha - f\|_D^2 + \|\mathrm{diag}(\lambda)\alpha\|_1.$$

When $\kappa = \infty$, the problem (1.2.2) is reduced to an analysis-based approach. Note that the term $\|(I - WW^\top)\alpha\|$ must be zero when $\kappa = \infty$. This implies that α is in the range of

W, i.e., $\alpha = Wu$ for some $u \in \mathbb{R}^n$, so we can rewrite (1.2.2) as

$$\min_{u\in\mathbb{R}^n} \frac{1}{2}\|Au - f\|_D^2 + \|\mathrm{diag}(\lambda)Wu\|_1, \tag{1.2.3}$$

which is referred as the *analysis-based approach*. We remark that the analysis-based approach does not require the use of the left-inverse of W. Hence, it can be generalized to any linear transform W.

Note that for redundant tight frame system W, the analysis-based, synthesis-based and balanced approaches cannot be derived from one another. Indeed, all the approaches have their own favourable datasets and applications. In general, it is difficult to draw definitive conclusions on which approach is better without specifying the applications and datasets. For frame-based image restoration, the synthesis-based and balanced approaches tend to explore more on the sparse representation of the underlying solution in terms of the given frame system by utilizing the redundancy. Thus the synthesis-based and balanced approaches usually enhance and sharpen edges, though they may introduce some artifacts as shown in [**22**].

Numerical simulations in [**26**] show that the analysis-based approach tends to capture the geometrical orientations of the objects in the image better than the other two approaches. This is because the coefficient Wu is often linked to the geometrical orientations of the objects in the image. Due to this reason, we focus on the analysis-based approach, which is applied to bio-image denoising in Chapter 3 of this thesis where we shall see that, the bio-images have rich geometrical structures.

Furthermore, the split Bregman method was used to develop a fast algorithm for the analysis-based approach in frame-based image restoration in [**26**], where numerical simulations showed that the split Bregman method is efficient for image deblurring, inpainting and denoising. The split Bregman method was first proposed in [**49**] which was shown to be powerful in [**49**, **89**] when it is applied to various PDE based image restoration approaches, e.g., total variation (TV) models. Convergence analysis of the split Bregman method was provided in [**26**].

1.2.2. The split Bregman method.

The ℓ_1-term involved in the analysis-based model (1.2.3) is non-smooth and non-separable, making direct optimization of (1.2.3) challenging. To overcome this, the main idea of the split Bregman method is that one can transfer (1.2.3) to a relatively simpler problem involving

only separable non-smooth terms. In particular, one introduces an auxiliary variable $d = Wu$ for (1.2.3), leading to the equivalent problem

$$\min_{u,d} H(u) + \|\mathrm{diag}(\lambda)d\|_1 \quad \text{subject to } d = Wu, \tag{1.2.4}$$

where $H(u) := \frac{1}{2}\|Au - f\|_D^2$ is a convex function. In order to solve (1.2.4), an iterative algorithm based on the Bregman distance with an inexact solver was proposed in [**49**]. The split Bregman method can also be understood as a special case of the alternating direction of multipliers (ADMM) [**43**], which is a variant of the augmented Lagrangian method (see e.g., [**45**]) when applied to (1.2.4). Here we shall derive the split Bregman algorithm based on an augmented Lagrangian of (1.2.4). Note that (1.2.4) is equivalent to the following optimization problem

$$\min_{u,d} H(u) + \|\mathrm{diag}(\lambda)d\|_1 + \frac{\mu}{2}\|Wu - d\|_2^2 \quad \text{subject to } d = Wu, \tag{1.2.5}$$

for some pre-defined positive constant μ. The Lagrangian for (1.2.5), also known as the augmented Lagrangian for (1.2.4), is given as

$$L_\mu(u, d, p) := H(u) + \|\mathrm{diag}(\lambda)d\|_1 + \langle p, d - Wu\rangle + \frac{\mu}{2}\|Wu - d\|_2^2.$$

The saddle points of $L_\mu(u, d, p)$ can be obtained by the following iterative procedure:

$$\begin{cases} (u^{k+1}, d^{k+1}) = \mathrm{argmin}_{u,d}\, L_\mu(u, d, p^k), \\ p^{k+1} = p^k + \mu(d^{k+1} - d^k), \end{cases}$$

which consists of one step of joint optimization of the variables (u, d) followed by one update of the Lagrange multiplier p. Now letting $b^k = -p^k/\mu$, we then have the equivalent problem

$$\begin{cases} (u^{k+1}, d^{k+1}) = \mathrm{argmin}_{u,d}\, H(u) + \|\mathrm{diag}(\lambda)d\|_1 + \frac{\mu}{2}\|Wu - d + b^k\|_2^2, \\ b^{k+1} = b^k + (Wu^{k+1} - d^{k+1}). \end{cases}$$

Now if one alternately optimizes the variables u and d in the first equation above, we will have the split Bregman algorithm as follows:

$$\begin{cases} u^{k+1} = \mathrm{argmin}_u\, H(u) + \frac{\mu}{2}\|Wu - d^k + b^k\|_2^2, \\ d^{k+1} = \mathrm{argmin}_d \|\mathrm{diag}(\lambda)d\|_1 + \frac{\mu}{2}\|d - Wu^{k+1} - b^k\|_2^2, \\ b^{k+1} = b^k + (Wu^{k+1} - d^{k+1}). \end{cases}$$

Since $H(u)$ is convex and differentiable, the subproblem in the first line is easy to solve explicitly. The second subproblem above can be solved analytically by soft-thresholding given in (1.1.12). Both of these analytical solutions to the two subproblems make the iteration efficient

and fast for many problems that are difficult to solve by other means. Besides its speed, the split Bregman method has several advantages. It generally performs well at the early stages with satisfactory results for applications in imaging sciences, where highly accurate solutions are not really needed. The method is also easy to code. Both these characteristics make the split Bregman method a practical algorithm for large scale problems in imaging sciences.

Split Bregman Algorithm

---

(1) Set initial guess d^0 and b^0.

(2) For $k = 0, 1, \ldots,$ perform the following iterations till convergence:

$$\begin{cases} u^{k+1} = (A^\top DA + \mu I)^{-1}(A^\top Df + \mu W^\top (d^k - b^k)); \\ d^{k+1} = T^1_{\lambda/\mu}(Wu^{k+1} + b^k); \\ b^{k+1} = b^k + Wu^{k+1} - d^{k+1}. \end{cases}$$

Note that the first equation in the split Bregman algorithm can be solved efficiently by the fast Fourier transform (FFT) when A is diagonalizable by the discrete Fourier transform, e.g., a convolution matrix. A deblurring result of the image "barbara" is displayed in Figure 1.2.1, where the matrix A in (1.2.1) is taken to be a convolution matrix with corresponding kernel a Gaussian function (generated in MATLAB by "fspecial('gaussian',15,1.5);") and the noise vector e is generated from a zero mean Gaussian distribution with $\sigma = 3$.

As a final remark, the convergence proof for the split Bregman method in [**49**] remains true in the case where the function H is merely convex (not necessarily differentiable) and the matrix W does not satisfy $W^\top W = I$, assuming the existence of saddle points for (1.2.5) and solutions to all subproblems arising within the split Bregman algorithm.

1.3. Spherical Harmonics

One way of constructing wavelet frames to represent square-integrable functions defined on the unit sphere involves the use of spherical harmonics (the spherical analogue of Fourier bases), which we shall see in Chapter 2. Therefore, this section presents some useful properties of the spherical harmonics (SH) and modified spherical harmonics (modified SH) respectively in subsections 1.3.1 and 1.3.2.

FIGURE 1.2.1. Deblurring results for the test image "barbara". Columns 1 to 3 contain respectively the original image, blurred image and deblurred image.

1.3.1. Spherical harmonics on $L^2(\mathbb{S}^2)$.

Let $\mathbb{S}^2$ denote the unit sphere centered at the origin in $\mathbb{R}^3$, parametrized as follows:

$$\mathbb{S}^2 := \{(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta) \,|\, \theta \in [0,\pi],\ \phi \in [0,2\pi)\}, \tag{1.3.1}$$

where ϕ, θ represent the azimuth and zenith respectively. The space $L^2(\mathbb{S}^2)$ denotes the set of all square-integrable functions on $\mathbb{S}^2$, i.e.,

$$L^2(\mathbb{S}^2) := \left\{ f : \int_0^{2\pi}\int_0^{\pi} |f(\theta,\phi)|^2 \sin\theta \, d\theta d\phi < \infty \right\},$$

and it is also a Hilbert space endowed with the inner product

$$\langle f,\, g\rangle_{L^2(\mathbb{S}^2)} := \int_0^{2\pi}\int_0^{\pi} f(\theta,\phi)\overline{g(\theta,\phi)} \sin\theta \, d\theta d\phi, \quad f,\, g \in L^2(\mathbb{S}^2). \tag{1.3.2}$$

The classical *spherical harmonics* (SH) $\{Y_l^m\}_{m,l}$ are defined as

$$Y_l^m(\theta,\phi) = N_{m,l} P_l^m(\cos\theta)\exp(\mathrm{i}m\phi),\ \theta \in [0,\pi],\ \phi \in [0,2\pi), \tag{1.3.3}$$

where m is the *degree* of the SH, l is the *order* of the SH,

$$N_{m,l} := (-1)^m \sqrt{\frac{(2l+1)\cdot(l-m)!}{4\pi\cdot(l+m)!}},\ |m| \le l,\ l \in \mathbb{N}_0, \tag{1.3.4}$$

$P_l^m(x)$ are the associated Legendre functions, defined in terms of derivatives of *Legendre polynomials* P_l on $[-1,1]$:

$$P_l^m(x) = (-1)^m (1-x^2)^{m/2} \frac{d^m}{dx^m} P_l(x), \quad 0 \le m \le l, \quad l \in \mathbb{N}_0,$$

$$P_l^{-m}(x) = (-1)^m \frac{(l-m)!}{(l+m)!} P_l^m(x), \quad 0 \le m \le l, \quad l \in \mathbb{N}_0, \tag{1.3.5}$$

and $\{P_l\}_{l\in\mathbb{N}_0}$ are the ordinary Legendre polynomials which can be expressed as the following by Rodrigues' formula

$$P_l(x) = \frac{1}{2^l l!}\frac{d^l}{dx^l}(x^2-1)^l, \quad x \in [-1,1], \quad l \in \mathbb{N}_0. \tag{1.3.6}$$

The normalization constant $N_{m,l}$ in (1.3.4) has a handy relation:

$$\frac{N_{m,l}}{N_{-m,l}} \cdot \frac{(l+m)!}{(l-m)!} = 1, \quad |m| \leq l,\, l \in \mathbb{N}_0, \tag{1.3.7}$$

which can be used to show that the spherical harmonics also have a useful property of complex conjugation, namely

$$\overline{Y_l^m} = (-1)^m Y_l^{-m}, |m| \leq l,\, l \in \mathbb{N}_0. \tag{1.3.8}$$

An alternative expression for the spherical harmonics that will be useful is

$$Y_l^m(\theta,\phi) := \Theta_l^m(\cos\theta)\Phi^m(\phi),\, \theta \in [0,\pi],\, \phi \in [0,2\pi),\, |m| \leq l,\, l \in \mathbb{N}_0, \tag{1.3.9}$$

where

$$\Theta_l^m(\cos\theta) := (-1)^m \sqrt{\frac{(2l+1)(l-m)!}{2(l+m)!}} P_l^m(\cos\theta),\, \Phi^m(\phi) := \frac{1}{\sqrt{2\pi}}\exp(\mathrm{i}m\phi), \tag{1.3.10}$$

It can be shown [**5**, Pages 69-71] that the SH $\{Y_l^m\}_{m,l}$ actually form an orthonormal basis for $L^2(\mathbb{S}^2)$. To see why they are orthonormal, observe that after a change of variables $x = \cos\theta$, the inner product $\langle Y_l^m,\, Y_{l'}^{m'}\rangle_{L^2(\mathbb{S}^2)}$ can be expressed as a product of two integrals given by

$$\langle Y_l^m,\, Y_{l'}^{m'}\rangle_{L^2(\mathbb{S}^2)} = \left\{\int_0^{2\pi} \Phi^m(\phi)\overline{\Phi^{m'}(\phi)}\, d\phi\right\}\left\{\int_{-1}^{1} \Theta_l^m(x)\Theta_{l'}^m(x)\, dx\right\}. \tag{1.3.11}$$

It is easy to see from the exponential definition of Φ^m in (1.3.10) and a standard calculation that

$$\int_0^{2\pi} \Phi^m(\phi)\overline{\Phi^{m'}(\phi)}\, d\phi = \delta_{m,m'},$$

where $\delta_{m,m'}$ denotes the Kronecker delta function, i.e., $\delta_{m,m'}$ equals 1 if $m = m'$, and zero otherwise. Therefore, to see that $\{Y_l^m\}_{m,l}$ is an orthonormal family, i.e.,

$$\langle Y_l^m,\, Y_{l'}^{m'}\rangle_{L^2(\mathbb{S}^2)} = \delta_{m,m'}\delta_{l,l'},$$

we consider $m = m'$ in the second integral in (1.3.11) and obtain

$$\int_{-1}^{1} \Theta_l^m(x)\Theta_{l'}^m(x)\, dx = \delta_{l,l'}, \tag{1.3.12}$$

where the above orthonormality relation can be established by proving that $\{\Theta_l^m(\cos\cdot)\}_l$ are eigenfunctions of the associated Legendre differential equation ($l \in \mathbb{N}_0$, and a fixed integer m such that $|m| \leq l$) given by

$$\frac{d^2\Theta}{d\theta^2} + \frac{\cos\theta}{\sin\theta}\frac{d\Theta}{d\theta} - \frac{m^2}{\sin^2\theta}\Theta = \lambda_l\Theta,$$

with eigenvalues $\lambda_l = -l(l+1)$. It can be shown [**37**, Theorems 5.9.7 to 5.9.10] that this differential equation is a special case of a Sturm-Liouville system, where its eigenfunctions $\{\Theta_l^m\}_l$ are orthonormal.

It is also known [**5**, Proposition 3.5] that the spherical harmonics $\{Y_l^m\}_{m,l}$ are eigenfunctions of the *Laplace-Beltrami operator* on $\mathbb{S}^2$ (also known as the Laplacian operator on the unit sphere) defined as

$$\Delta_b := \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta\,\frac{\partial}{\partial\theta}) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}, \tag{1.3.13}$$

with eigenvalues $\lambda_l = -l(l+1)$, i.e.,

$$\Delta_b Y_l^m = -l(l+1)Y_l^m, \quad |m| \leq l,\, l \in \mathbb{N}_0. \tag{1.3.14}$$

This relation will be helpful in the remainder of the thesis when spherical harmonics are applied to computational problems on the sphere.

One last useful property of the spherical harmonics is their antipodally (anti)-symmetric behaviour: for $\theta \in [0,\pi]$, $\phi \in [0,2\pi)$,

$$Y_l^m(\pi-\theta,\phi+\pi) = (-1)^l Y_l^m(\theta,\phi), \quad |m| \leq l,\, l \in \mathbb{N}_0, \tag{1.3.15}$$

i.e., for $\mathbf{u} \in \mathbb{S}^2$ with the standard spherical parametrization in (1.3.1), we have

$$Y_l^m(-\mathbf{u}) = (-1)^l Y_l^m(\mathbf{u}), \quad |m| \leq l,\, l \in \mathbb{N}_0. \tag{1.3.16}$$

This means that the spherical harmonics of even/odd order are antipodally symmetric/anti-symmetric respectively. To visualize spherical harmonics, Figure 1.3.1 provides a 3D-surface plot of $\mathrm{Re}[Y_l^m(\theta,\phi)]$ of the spherical harmonics $\{Y_l^m\}_{m,l}$ up to order 3.

REMARK 1.3.1. In this thesis, visualization of a bounded function F on the sphere is made as follows: we first express F in terms of spherical coordinates, say $F(\theta,\phi)$. Then F is visualized by means of a 3D-surface plot, whose Cartesian coordinates are given by $(R\sin\theta\,\cos\phi,\, R\sin\theta\,\sin\phi,\, R\cos\theta)$, where $R = |F(\theta,\phi)|$. This 3D-surface plot projects away from the origin of $\mathbb{R}^3$ in the directions along which the function $|F|$ has large relative values, while staying close to the origin in the directions where the function $|F|$ has small relative values near zero.

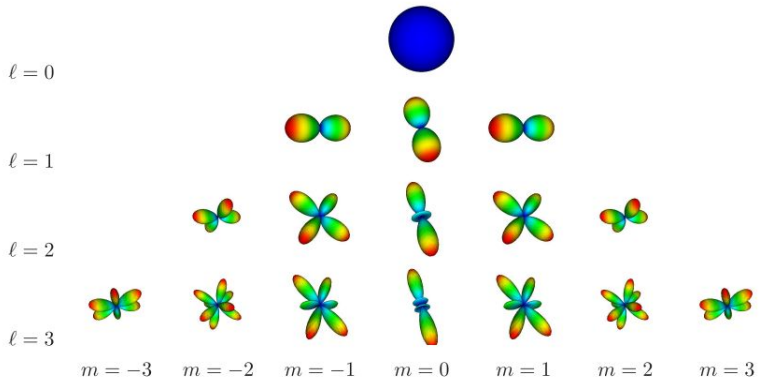

FIGURE 1.3.1. A 3D-surface plot of the real part $\mathrm{Re}[Y_l^m(\theta, \phi)]$ of the spherical harmonics up to order 3 (see Remark 1.3.1 on how to visualize this plot).

1.3.2. Modified spherical harmonics on $L^2_{\mathbf{sym}}(\mathbb{S}^2)$.

In certain applications where the functions or signals are defined on the sphere such as High Angular Resolution Diffusion Imaging (HARDI) [**83**], the signal is assumed to be real and antipodally symmetric. Therefore, it is preferable to obtain a modified spherical harmonic basis that is also real and antipodally symmetric. To obtain the symmetric property in the modified basis, only spherical harmonics of even order are considered. As explained earlier, only spherical harmonics of even order are antipodally symmetric. To obtain the real-valued property, real and imaginary parts of the spherical harmonics are selected accordingly to the degree m.

For $l \in 2\mathbb{N}_0$, $|m| \leq l$, the *modified spherical harmonics* (modified SH) are defined by

$$\tilde{Y}_l^m := \begin{cases} \sqrt{2} \cdot \mathrm{Re}[Y_l^m], & \text{if } m < 0, \\ Y_l^m, & \text{if } m = 0, \\ \sqrt{2} \cdot \mathrm{Im}[Y_l^m], & \text{if } m > 0, \end{cases} \tag{1.3.17}$$

where $\mathrm{Re}[Y_l^m]$ and $\mathrm{Im}[Y_l^m]$ are respectively the real and imaginary parts of Y_l^m. The normalization factor $\sqrt{2}$ in (1.3.17) makes this modified basis orthonormal. Plots of some selected modified SH are provided in Figure 1.3.2. Modified SH are used to represent HARDI signals in [**38**].

Let $L^2_{\mathrm{sym}}(\mathbb{S}^2)$ be the set of antipodally symmetric, square-integrable functions defined on the sphere, i.e.,

$$L^2_{\mathrm{sym}}(\mathbb{S}^2) := \left\{ f \in L^2(\mathbb{S}^2) \,:\, f(-\mathbf{u}) = f(\mathbf{u}),\, \mathbf{u} \in \mathbb{S}^2 \text{ a.e.} \right\}. \tag{1.3.18}$$

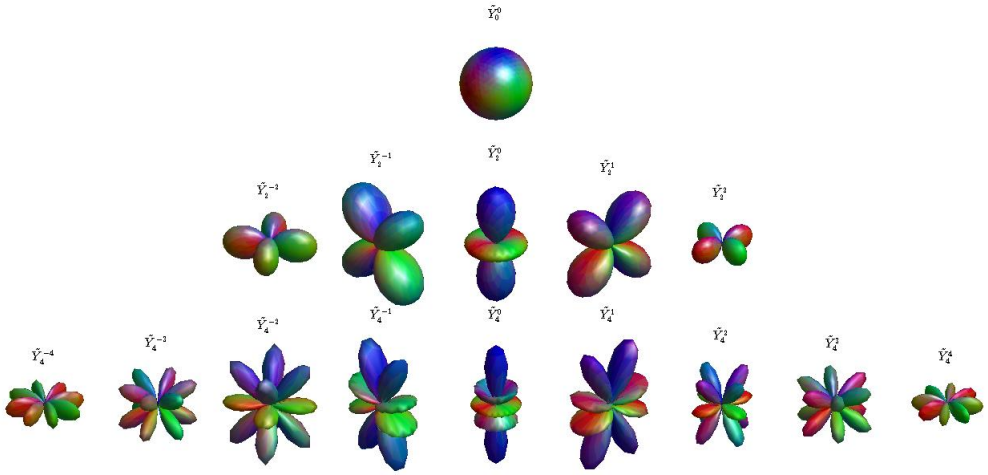

FIGURE 1.3.2. 3D-surface plots of the modified SH, $\tilde{Y}_l^m(\theta,\phi)$, $|m| \leq l$, $l = 0, 2, 4$ (see Remark 1.3.1 on how to visualize these plots).

We shall show that the modified spherical harmonics $\left\{\tilde{Y}_l^m\right\}_{m,l}$ forms an orthonormal basis of $L^2_{\mathrm{sym}}(\mathbb{S}^2)$, which also implies that

$$L^2_{\mathrm{sym}}(\mathbb{S}^2) = \mathrm{clos}_{L^2(\mathbb{S}^2)}\left(\mathrm{span}\left\{\tilde{Y}_l^m : |m| \leq l,\, l \in 2\mathbb{N}_0\right\}\right).$$

Consider any continuous function g which is antipodally symmetric on the sphere, i.e., $g(-\mathbf{u}) = g(\mathbf{u})$, $\mathbf{u} \in \mathbb{S}^2$. Clearly, $g \in L^2(\mathbb{S}^2)$ and we have the expansion in spherical harmonics

$$g = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} \hat{g}_{m,l} Y_l^m, \quad \hat{g}_{m,l} = \int_{\mathbb{S}^2} g(\mathbf{u})\overline{Y_l^m(\mathbf{u})}\, d\sigma(\mathbf{u}),$$

where $d\sigma(\mathbf{u}) := \sin\theta d\theta d\phi$, and $\mathbf{u} \in \mathbb{S}^2$ has corresponding spherical co-ordinates (θ,ϕ) as defined in (1.3.1). We wish to show that the coefficients $\hat{g}_{m,l}$ are zero when $|m| \leq l$, l is an odd positive integer. To do so, we first describe the sphere $\mathbb{S}^2$ as a union of its two hemispheres (up to a null set with respect to the measure σ), i.e., $\mathbb{S}^2 = \mathbb{S}^2_1 \cup \mathbb{S}^2_2$, and then consider the following

$$\begin{aligned}
\hat{g}_{m,l} &= \int_{\mathbb{S}^2_1} g(\mathbf{u})\overline{Y_l^m(\mathbf{u})}\, d\sigma(\mathbf{u}) + \int_{\mathbb{S}^2_2} g(\mathbf{u})\overline{Y_l^m(\mathbf{u})}\, d\sigma(\mathbf{u}) \\
&= \int_{\mathbb{S}^2_1} g(\mathbf{u})\overline{Y_l^m(\mathbf{u})}\, d\sigma(\mathbf{u}) + \int_{\mathbb{S}^2_1} g(-\mathbf{v})\overline{Y_l^m(-\mathbf{v})}\, d\sigma(\mathbf{v}) \\
&= \int_{\mathbb{S}^2_1} g(\mathbf{u})\overline{Y_l^m(\mathbf{u})}\, d\sigma(\mathbf{u}) - \int_{\mathbb{S}^2_1} g(\mathbf{v})\overline{Y_l^m(\mathbf{v})}\, d\sigma(\mathbf{v}) = 0,
\end{aligned}$$

where we have used the fact that g is antipodally symmetric and Y_l^m is antipodally anti-symmetric from (1.3.16) when $|m| \leq l$, l is an odd positive integer. Therefore,

$$g \in \mathrm{clos}_{L^2(\mathbb{S}^2)} \left(\mathrm{span} \left\{ Y_l^m : |m| \leq l,\, l \in 2\mathbb{N}_0 \right\} \right),$$

for any continuous function g which is antipodally symmetric on the sphere. Using a standard density argument, we can show that

$$L^2_{\mathrm{sym}}(\mathbb{S}^2) = \mathrm{clos}_{L^2(\mathbb{S}^2)} \left(\mathrm{span} \left\{ Y_l^m : |m| \leq l,\, l \in 2\mathbb{N}_0 \right\} \right).$$

Finally, to show $\left\{ \tilde{Y}_l^m \right\}_{m,l}$ is an orthonormal basis for $L^2_{\mathrm{sym}}(\mathbb{S}^2)$, we note that $\left\{ \tilde{Y}_l^m \right\}_{m,l}$ is an orthonormal family, so it remains to establish

$$\mathrm{clos}_{L^2(\mathbb{S}^2)} \left(\mathrm{span} \left\{ Y_l^m : |m| \leq l,\, l \in 2\mathbb{N}_0 \right\} \right) = \mathrm{clos}_{L^2(\mathbb{S}^2)} \left(\mathrm{span} \left\{ \tilde{Y}_l^m : |m| \leq l,\, l \in 2\mathbb{N}_0 \right\} \right),$$

which can be shown by using the following relations between Y_l^m and $\tilde{Y}_l^m$: for $|m| \leq l$, $l \in 2\mathbb{N}_0$,

$$\tilde{Y}_l^m = \begin{cases} 2^{-1/2}[Y_l^m + (-1)^m Y_l^{-m}], & \text{if } m < 0, \\ Y_l^m, & \text{if } m = 0, \\ -2^{-1/2}\mathrm{i}[Y_l^m - (-1)^m Y_l^{-m}], & \text{if } m > 0, \end{cases} \tag{1.3.19}$$

and

$$Y_l^m = \begin{cases} 2^{-1/2}[\tilde{Y}_l^m - \mathrm{i}(-1)^m \tilde{Y}_l^{-m}], & \text{if } m < 0, \\ \tilde{Y}_l^m, & \text{if } m = 0, \\ 2^{-1/2}[(-1)^m \tilde{Y}_l^{-m} + \mathrm{i}\tilde{Y}_l^m], & \text{if } m > 0. \end{cases} \tag{1.3.20}$$

Equation (1.3.19) follows from the definition of modified SH in (1.3.17) and the complex conjugation property (1.3.8). Equation (1.3.20) is obtained using the identity $\exp(\mathrm{i}m\phi) = \cos(m\phi) + \mathrm{i}\sin(m\phi)$, $\forall m$, ϕ, on (1.3.3) with the earlier properties (1.3.5) and (1.3.7).

Throughout the next two chapters, this set of modified spherical harmonics $\left\{ \tilde{Y}_l^m \right\}_{m,l}$ and its corresponding space $L^2_{\mathrm{sym}}(\mathbb{S}^2)$ will referred to frequently in our new wavelet frame constructions pertaining to HARDI applications.

1.4. Sobolev Spaces on the Sphere

This section serves to provide basic properties of Sobolev spaces of functions defined on the unit sphere, which are instrumental in establishing that the wavelet systems constructed to represent bio-images in Chapter 3, are frame systems for some Sobolev space of a particular exponent. In subsection 1.4.1, some fundamentals regarding Sobolev spaces $H^s(\mathbb{S}^2)$ and their dual spaces $H^{-s}(\mathbb{S}^2)$ are presented with detailed proofs. Analogous properties are then

presented in subsection 1.4.2 regarding their symmetric counterparts $H^s_{\text{sym}}(\mathbb{S}^2)$ and their dual spaces $H^{-s}_{\text{sym}}(\mathbb{S}^2)$.

1.4.1. Sobolev spaces on the sphere $H^s(\mathbb{S}^2)$ and SH.

As mentioned in the previous section, the spherical harmonics $\{Y_l^m\}_{m,l}$ defined in (1.3.9) form an orthonormal basis for $L^2(\mathbb{S}^2)$. In fact, spherical harmonics have often been referred as the 'Fourier basis on the sphere'. Similar to the classical Fourier expansions, if $f \in L^2(\mathbb{S}^2)$, then f has a corresponding spherical harmonics series

$$\sum_{l=0}^{\infty}\sum_{m=-l}^{l} \hat{f}_{m,l} Y_l^m, \quad \text{where } \hat{f}_{m,l} = \int_0^{2\pi}\int_0^{\pi} f(\theta,\phi)\overline{Y_l^m(\theta,\phi)}\sin\theta\, d\theta d\phi, \tag{1.4.1}$$

Here we are interested in Sobolev spaces on the unit sphere. For $s \in [0,\infty)$, the Sobolev spaces $H^s(\mathbb{S}^2)$ are subspaces of $L^2(\mathbb{S}^2)$ which are determined by the decay of the spherical harmonics coefficients. For our purposes, the Sobolev space results that we use in this subsection are spherical analogues of results found in [**27**, section 1.4]. We provide detailed proofs in this subsection because these proofs are analogous in the next subsection, which involves the symmetric Sobolev spaces. Results on Sobolev spaces on domains including spheres, are also available in, for instance, [**2**].

Definition 1.4.1. Let $s \in [0,\infty)$. Then by $H^s(\mathbb{S}^2)$ we denote the space

$$H^s(\mathbb{S}^2) := \left\{ f \in L^2(\mathbb{S}^2) \,:\, \sum_{l=0}^{\infty}\sum_{m=-l}^{l} (1+l^2)^s |\hat{f}_{m,l}|^2 < \infty \right\},$$

where $\left\{\hat{f}_{m,l}\right\}_{m,l}$ are the spherical harmonic coefficients given in (1.4.1). The space $H^s(\mathbb{S}^2)$ is called a *Sobolev space* on $\mathbb{S}^2$ of order s. Note that $H^0(\mathbb{S}^2) = L^2(\mathbb{S}^2)$.

Theorem 1.4.2. (*c.f.* [**27**, Theorem 1.29]) *The space $H^s(\mathbb{S}^2)$ is a Hilbert space with the inner product*

$$\langle f, g\rangle_s := \sum_{l=0}^{\infty}\sum_{m=-l}^{l} (1+l^2)^s \hat{f}_{m,l}\overline{\hat{g}_{m,l}}, \tag{1.4.2}$$

for f, $g \in H^s(\mathbb{S}^2)$ with spherical harmonic coefficients $\left\{\hat{f}_{m,l}\right\}_{m,l}$, $\{\hat{g}_{m,l}\}_{m,l}$ respectively. The induced norm is then given by

$$\|f\|_s = \Big(\sum_{l=0}^{\infty}\sum_{m=-l}^{l} (1+l^2)^s |\hat{f}_{m,l}|^2\Big)^{1/2}.$$

Furthermore, the linear span of the spherical harmonics is dense in $H^s(\mathbb{S}^2)$.

PROOF. It is easy to observe that $H^s(\mathbb{S}^2)$ is a linear space and that $\langle\cdot,\cdot\rangle_s$ is an inner product. Indeed, $\langle\cdot,\cdot\rangle_s$ is well defined because by the Cauchy-Schwarz inequality,

$$\left|\sum_{l=0}^{\infty}\sum_{m=-l}^{l}(1+l^2)^s\hat{f}_{m,l}\overline{\hat{g}_{m,l}}\right| \leq \sqrt{\sum_{l=0}^{\infty}\sum_{m=-l}^{l}(1+l^2)^s|\hat{f}_{m,l}|^2}\sqrt{\sum_{l=0}^{\infty}\sum_{m=-l}^{l}(1+l^2)^s|\hat{g}_{m,l}|^2},$$

which implies that $\langle f,g\rangle_s$ exists for every f, $g\in H^s(\mathbb{S}^2)$.

To see that $H^s(\mathbb{S}^2)$ is complete, let $\{f^n\}_n$ be a Cauchy sequence in $H^s(\mathbb{S}^2)$, i.e., for every $\epsilon>0$, there exists an N_ϵ such that

$$\|f^n-f^k\|_s<\epsilon,\quad n,k>N_\epsilon,$$

or equivalently, denoting $\left\{\hat{f}^n_{m,l}\right\}_{m,l}$,$\left\{\hat{f}^k_{m,l}\right\}_{m,l}$ as the respective spherical harmonic coefficients of f^n, f^k,

$$\sum_{l=0}^{\infty}\sum_{m=-l}^{l}(1+l^2)^s|\hat{f}^n_{m,l}-\hat{f}^k_{m,l}|^2<\epsilon^2,\ n,k>N_\epsilon.$$

From this, we can see that for all non-negative integers L,

$$\sum_{l=0}^{L}\sum_{m=-l}^{l}(1+l^2)^s|\hat{f}^n_{m,l}-\hat{f}^k_{m,l}|^2<\epsilon^2,\ n,k>N_\epsilon. \tag{1.4.3}$$

Since $\mathbb{C}$ is complete, there is a sequence $\left\{\hat{f}_{m,l}\right\}_{m,l}$ such that for each fixed pair of m and l, where $l\in\mathbb{N}_0$ and $-l\leq m\leq l$, $\hat{f}^n_{m,l}\to\hat{f}_{m,l}$ as $n\to\infty$. Now passing to the limit in (1.4.3) as k goes to infinity, we have

$$\sum_{l=0}^{L}\sum_{m=-l}^{l}(1+l^2)^s|\hat{f}^n_{m,l}-\hat{f}_{m,l}|^2\leq\epsilon^2,\ n>N_\epsilon.$$

Next, letting L approach infinity, we obtain $\sum\limits_{l=0}^{\infty}\sum\limits_{m=-l}^{l}(1+l^2)^s|\hat{f}^n_{m,l}-\hat{f}_{m,l}|^2\leq\epsilon^2$, $n>N_\epsilon$. It then follows that

$$f:=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\hat{f}_{m,l}Y_l^m\in H^s(\mathbb{S}^2),$$

since for $n>N_\epsilon$,

$$\begin{aligned}\sum_{l=0}^{\infty}\sum_{m=-l}^{l}(1+l^2)^s|\hat{f}_{m,l}|^2&=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}(1+l^2)^s|\hat{f}_{m,l}-\hat{f}^n_{m,l}+\hat{f}^n_{m,l}|^2\\&\leq 2\sum_{l=0}^{\infty}\sum_{m=-l}^{l}(1+l^2)^s|\hat{f}_{m,l}-\hat{f}^n_{m,l}|^2+2\sum_{l=0}^{\infty}\sum_{m=-l}^{l}(1+l^2)^s|\hat{f}^n_{m,l}|^2\\&<\infty.\end{aligned}$$

Consequently, this also shows that $\|f - f^n\|_s \to 0$ as $n \to \infty$.

Finally let $f \in H^s(\mathbb{S}^2)$ with spherical harmonic coefficients $\left\{\hat{f}_{m,l}\right\}_{m,l}$ and set

$$f^L = \sum_{l=0}^{L} \sum_{m=-l}^{l} \hat{f}_{m,l} Y_l^m.$$

Then we have

$$\|f - f^L\|_s^2 = \sum_{l=L+1}^{\infty} \sum_{m=-l}^{l} (1+l^2)^s |\hat{f}_{m,l}|^2 \to 0, \quad L \to \infty,$$

since $\sum_{l=0}^{\infty} \sum_{m=-l}^{l} (1+l^2)^s |\hat{f}_{m,l}|^2 < \infty$. Thus the linear span of the spherical harmonics is dense in $H^s(\mathbb{S}^2)$. □

We now consider the case of Sobolev spaces with negative exponent, i.e., $H^{-s}(\mathbb{S}^2)$ with $s > 0$. Much of what has already been written holds in this case as well but some care must be exercised. For instance, for $s > 0$, sums of the form

$$\sum_{l=0}^{\infty} \sum_{m=-l}^{l} (1+l^2)^{-s} |\hat{f}_{m,l}|^2$$

are more likely to converge even if the spherical harmonic coefficients $\left\{\hat{f}_{m,l}\right\}_{m,l}$ do not tend to zero as l tends to infinity. This allows for the study of distributions or generalized functions [**14**].

Definition 1.4.3. For $0 < s < \infty$, we denote by $H^{-s}(\mathbb{S}^s)$, the dual space of $H^s(\mathbb{S}^2)$, i.e., the space of bounded linear functionals on $H^s(\mathbb{S}^2)$.

Theorem 1.4.4. (*cf.* [**27**, Theorem 1.33]) *Let G be a bounded linear functional defined on $H^s(\mathbb{S}^2)$. Then*

$$\|G\|_{-s} = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} (1+l^2)^{-s} |\hat{g}_{m,l}|^2, \tag{1.4.4}$$

where we define

$$\|G\|_{-s} := \sup_{f \in H^s(\mathbb{S}^2), \|f\|_s = 1} |G[f]|, \quad \hat{g}_{m,l} := G[Y_l^m],\ |m| \le l,\ l \in \mathbb{N}_0.$$

Conversely, for each sequence $\{\hat{g}_{m,l}\}_{m,l}$ satisfying

$$\sum_{l=0}^{\infty} \sum_{m=-l}^{l} (1+l^2)^{-s} |\hat{g}_{m,l}|^2 < \infty,$$

there exists a bounded linear functional G defined on $H^s(\mathbb{S}^2)$ such that $\hat{g}_{m,l} = G[Y_l^m]$, $|m| \le l$, $l \in \mathbb{N}_0$.

PROOF. Let G be a bounded linear functional on $H^s(\mathbb{S}^2)$ and $\hat{g}_{m,l} = G[Y_l^m]$, $|m| \leq l$, $l \in \mathbb{N}_0$. Given any $\psi \in H^s(\mathbb{S}^2)$ with the expansion $\psi = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \hat{\psi}_{m,l} Y_l^m$, for a non-negative integer L, consider

$$\psi_L := \sum_{l=0}^{L} \sum_{m=-l}^{l} \hat{\psi}_{m,l} Y_l^m.$$

Then we have

$$G[\psi_L] = \sum_{l=0}^{L} \sum_{m=-l}^{l} \hat{\psi}_{m,l} G[Y_l^m] = \sum_{l=0}^{L} \sum_{m=-l}^{l} \hat{\psi}_{m,l} \hat{g}_{m,l}.$$

By Cauchy-Schwartz's inequality, we obtain

$$\begin{aligned} |G[\psi_L]| &\leq (\sum_{l=0}^{L} \sum_{m=-l}^{l} (1+l^2)^s |\hat{\psi}_{m,l}|^2)^{1/2} (\sum_{l=0}^{L} \sum_{m=-l}^{l} (1+l^2)^{-s} |\hat{g}_{m,l}|^2)^{1/2} \\ &\leq \|\psi_L\|_s (\sum_{l=0}^{\infty} \sum_{m=-l}^{l} (1+l^2)^{-s} |\hat{g}_{m,l}|^2)^{1/2}. \end{aligned}$$

Since the linear span of the spherical harmonics is dense in $H^s(\mathbb{S}^2)$, the above inequality remains true if we replace ψ_L with $\psi \in H^s(\mathbb{S}^2) \backslash \{0\}$ and set $f := \frac{\psi}{\|\psi\|_s}$, which implies that

$$\|G\|_{-s} = \sup_{f \in H^s(\mathbb{S}^2), \|f\|_s = 1} |G[f]| \leq (\sum_{l=0}^{\infty} \sum_{m=-l}^{l} (1+l^2)^{-s} |\hat{g}_{m,l}|^2)^{1/2}.$$

To complete the first part of the proof, we shall now show that

$$\|G\|_{-s} \geq (\sum_{l=0}^{\infty} \sum_{m=-l}^{l} (1+l^2)^{-s} |\hat{g}_{m,l}|^2)^{1/2}.$$

Define

$$\varphi_L := \sum_{l=0}^{L} \sum_{m=-l}^{l} (1+l^2)^{-s} \overline{\hat{g}_{m,l}} Y_l^m.$$

Then we obtain

$$\begin{aligned} \|\varphi_L\|_s^2 &= \sum_{l=0}^{L} \sum_{m=-l}^{l} (1+l^2)^s |(1+l^2)^{-s} \overline{\hat{g}_{m,l}}|^2 \\ &= \sum_{l=0}^{L} \sum_{m=-l}^{l} (1+l^2)^{-s} |\hat{g}_{m,l}|^2. \end{aligned}$$

Therefore, using the theory of bounded linear functionals in Hilbert spaces, for all non-negative integers L,

$$\begin{aligned}\|G\|_{-s} &\geq \left|G\left[\frac{\varphi_L}{\|\varphi_L\|_s}\right]\right| \\ &= \left|\sum_{l=0}^{L}\sum_{m=-l}^{l}(1+l^2)^{-s}\overline{\hat{g}_{m,l}}G[Y_l^m]\right| / (\sum_{l=0}^{L}\sum_{m=-l}^{l}(1+l^2)^{-s}|\hat{g}_{m,l}|^2)^{1/2} \\ &= (\sum_{l=0}^{L}\sum_{m=-l}^{l}(1+l^2)^{-s}|\hat{g}_{m,l}|^2)^{1/2},\end{aligned}$$

where we have used the facts that G is a linear functional and $\hat{g}_{m,l} = G[Y_l^m]$, $|m| \leq l$, $l \in \mathbb{N}_0$. The above gives

$$\|G\|_{-s} \geq (\sum_{l=0}^{\infty}\sum_{m=-l}^{l}(1+l^2)^{-s}|\hat{g}_{m,l}|^2)^{1/2}.$$

Thus we conclude that

$$\|G\|_{-s} = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}(1+l^2)^{-s}|\hat{g}_{m,l}|^2.$$

Conversely, we note that given any sequence $\{\hat{g}_{m,l}\}_{m,l}$ satisfying

$$\sum_{l=0}^{\infty}\sum_{m=-l}^{l}(1+l^2)^{-s}|\hat{g}_{m,l}|^2 < \infty,$$

we can define a functional G on $H^s(\mathbb{S}^2)$ by

$$G[\psi] := \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\hat{\psi}_{m,l}\hat{g}_{m,l}, \quad \psi \in H^s(\mathbb{S}^2),$$

and the above estimates show that G is a bounded linear functional on $H^s(\mathbb{S}^2)$ and (1.4.4) holds. Furthermore, taking $\psi = Y_l^m$, we obtain $G[Y_l^m] = \hat{g}_{m,l}$, $|m| \leq l$, $l \in \mathbb{N}_0$. □

From Theorem 1.4.4, we have an elaboration on the duality between $H^s(\mathbb{S}^2)$ and $H^{-s}(\mathbb{S}^2)$.

THEOREM 1.4.5. (*cf.* [**27**, Theorem 1.34]) *For $g \in L^2(\mathbb{S}^2)$, the duality pairing*

$$G[\varphi] := \langle \varphi, g\rangle_0, \quad \varphi \in H^s(\mathbb{S}^2), \tag{1.4.5}$$

defines a bounded linear functional on $H^s(\mathbb{S}^2)$, i.e., $G \in H^{-s}(\mathbb{S}^2)$. In particular, $L^2(\mathbb{S}^2)$ can be viewed as a subspace of the dual space $H^{-s}(\mathbb{S}^2)$, $0 \leq s < \infty$, and thus the linear span of the spherical harmonics is dense in $H^{-s}(\mathbb{S}^2)$.

PROOF. Let $\{\hat{g}_{m,l}\}_{m,l}$ be the spherical harmonic coefficients of g. Since $G[Y_l^m] = \hat{g}_{m,l}$, $|m| \leq l$, $l \in \mathbb{N}_0$, by the second part of Theorem 1.4.4, we have $G \in H^{-s}(\mathbb{S}^2)$. Now let $F \in H^{-s}(\mathbb{S}^2)$ with $\hat{F}_{m,l} = F[Y_l^m]$, $|m| \leq l$, $l \in \mathbb{N}_0$, and define $F_L \in H^{-s}(\mathbb{S}^2)$ by

$$F_L[\varphi] := \langle \varphi, g_L \rangle_0$$

where

$$g_L := \sum_{l=0}^{L} \sum_{m=-l}^{l} \hat{F}_{m,l} Y_l^m.$$

Then

$$\|F - F_L\|_{-s}^2 = \sum_{l=L+1}^{\infty} \sum_{m=-l}^{l} (1+l^2)^{-s} |\hat{F}_{m,l}|^2$$

tends to zero as L goes to infinity, which implies that the linear span of the spherical harmonics is dense in $H^{-s}(\mathbb{S}^2)$. □

REMARK 1.4.6. By using a density argument, the above duality pairing (1.4.5) can be extended to bounded linear functionals in $H^{-s}(\mathbb{S}^2)$. In particular, for $\varphi \in H^s(\mathbb{S}^2)$ and $g \in H^{-s}(\mathbb{S}^2)$, we define

$$g[\varphi] := \langle \varphi, g \rangle_0$$

with corresponding norm

$$\|g\|_{-s} := \sup_{\varphi \in H^s(\mathbb{S}^2), \|\varphi\|_s = 1} \langle \varphi, g \rangle_0.$$

Note that $H^{-s}(\mathbb{S}^2)$ becomes a Hilbert space by extending the inner product (1.4.2) previously defined in Theorem 1.4.2 for $s \geq 0$ to $s < 0$.

1.4.2. Sobolev spaces on the sphere $H^s_{\mathbf{sym}}(\mathbb{S}^2)$ and modified SH.

This subsection is analogous to the previous subsection, so we shall forgo the proofs here. The modified spherical harmonics $\left\{\tilde{Y}_l^m\right\}_{m,l}$ defined in (1.3.17) form an orthonormal basis of $L^2_{\mathrm{sym}}(\mathbb{S}^2)$. If $f \in L^2_{\mathrm{sym}}(\mathbb{S}^2)$, then f has a corresponding modified spherical harmonic series

$$\sum_{l \in 2\mathbb{N}_0} \sum_{m=-l}^{l} \tilde{f}_{m,l} \tilde{Y}_l^m, \quad \text{where } \tilde{f}_{m,l} = \int_{\mathbb{S}^2} f \overline{\tilde{Y}_l^m} \, d\sigma.$$

We are now interested in Sobolev spaces of symmetric functions on the sphere. For $s \in [0, \infty)$, the Sobolev spaces $H^s_{\mathrm{sym}}(\mathbb{S}^2)$ are subspaces of $L^2_{\mathrm{sym}}(\mathbb{S}^2)$ which are determined by the decay of the modified spherical harmonics coefficients.

DEFINITION 1.4.7. Let $s \in [0, \infty)$. By $H^s_{\mathrm{sym}}(\mathbb{S}^2)$, we denote the space

$$H^s_{\mathrm{sym}}(\mathbb{S}^2) = \left\{ f \in L^2_{\mathrm{sym}}(\mathbb{S}^2) \, : \, \sum_{l \in 2\mathbb{N}_0} \sum_{m=-l}^{l} (1+l^2)^s |\tilde{f}_{m,l}|^2 < \infty \right\},$$

where $\left\{\tilde{f}_{m,l}\right\}_{m,l}$ are the coefficients of the modified spherical harmonics. The space $H^s_{\mathrm{sym}}(\mathbb{S}^2)$ is called a *symmetric Sobolev space* on $\mathbb{S}^2$ of order s. Note that $H^0_{\mathrm{sym}}(\mathbb{S}^2) = L^2_{\mathrm{sym}}(\mathbb{S}^2)$.

Note that due to (1.3.19) and (1.3.20), the triangle inequality gives the equivalence of the SH coefficients $\left\{\hat{f}_{m,l}\right\}_{m,l}$ and modified SH coefficients $\left\{\tilde{f}_{m,l}\right\}_{m,l}$ in the following sense:

$$|\hat{f}_{m,l}|^2 \leq |\tilde{f}_{m,l}|^2 + |\tilde{f}_{-m,l}|^2, \quad |\tilde{f}_{m,l}|^2 \leq |\hat{f}_{m,l}|^2 + |\hat{f}_{-m,l}|^2, \, |m| \leq l, \, l \in 2\mathbb{N}_0.$$

Thus we obtain

$$\frac{1}{2} \sum_{l \in 2\mathbb{N}_0} \sum_{m=-l}^{l} (1+l^2)^s |\hat{f}_{m,l}|^2 \leq \sum_{l \in 2\mathbb{N}_0} \sum_{m=-l}^{l} (1+l^2)^s |\tilde{f}_{m,l}|^2 \leq 2 \sum_{l \in 2\mathbb{N}_0} \sum_{m=-l}^{l} (1+l^2)^s |\hat{f}_{m,l}|^2,$$

which justifies our intent of defining $H^s_{\mathrm{sym}}(\mathbb{S}^2)$ in terms of the decay of the modified SH coefficients $\left\{\tilde{f}_{m,l}\right\}_{m,l}$ (instead of the SH coefficients $\left\{\hat{f}_{m,l}\right\}_{m,l}$).

We now state the results on $H^s_{\mathrm{sym}}(\mathbb{S}^2)$ corresponding to those on $H^s(\mathbb{S}^2)$ in the previous subsection.

THEOREM 1.4.8. *The space $H^s_{\mathrm{sym}}(\mathbb{S}^2)$ is a Hilbert space with the inner product*

$$\langle f, g \rangle_{\star,s} = \sum_{l \in 2\mathbb{N}_0} \sum_{m=-l}^{l} (1+l^2)^s \tilde{f}_{m,l} \overline{\tilde{g}_{m,l}}, \tag{1.4.6}$$

for f, $g \in H^s_{\mathrm{sym}}(\mathbb{S}^2)$ with modified spherical harmonic coefficients $\{\tilde{f}_{m,l}\}_{m,l}$, $\{\tilde{g}_{m,l}\}_{m,l}$ respectively. The induced norm is then given by

$$\|f\|_{\star,s} = \Big(\sum_{l \in 2\mathbb{N}_0} \sum_{m=-l}^{l} (1+l^2)^s |\tilde{f}_{m,l}|^2 \Big)^{1/2}.$$

Furthermore, the linear span of the modified spherical harmonics is dense in $H^s_{\mathrm{sym}}(\mathbb{S}^2)$.

DEFINITION 1.4.9. For $0 < s < \infty$, we denote by $H^{-s}_{\mathrm{sym}}(\mathbb{S}^2)$, the dual space of $H^s_{\mathrm{sym}}(\mathbb{S}^2)$, i.e., the space of bounded linear functionals on $H^s_{\mathrm{sym}}(\mathbb{S}^2)$.

THEOREM 1.4.10. *Let G be a bounded linear functional defined on $H^s_{\mathrm{sym}}(\mathbb{S}^2)$. Then*

$$\|G\|_{\star,-s} = \sum_{l \in 2\mathbb{N}_0} \sum_{m=-l}^{l} (1+l^2)^{-s} |\tilde{g}_{m,l}|^2$$

where we define

$$\|G\|_{\star,-s} := \sup_{f\in H^s_{\mathrm{sym}}(\mathbb{S}^2),\|f\|_{\star,s}=1} |G[f]|, \quad \tilde{g}_{m,l} := G[\tilde{Y}_l^m],\ |m|\le l,\ l\in 2\mathbb{N}_0.$$

Conversely, for each sequence $\{\tilde{g}_{m,l}\}_{m,l}$ *satisfying*

$$\sum_{l\in 2\mathbb{N}_0}\sum_{m=-l}^{l}(1+l^2)^{-s}|\tilde{g}_{m,l}|^2<\infty,$$

there exists a bounded linear functional G *defined on* $H^s_{\mathrm{sym}}(\mathbb{S}^2)$ *such that* $\tilde{g}_{m,l} = G[\tilde{Y}_l^m]$, $|m|\le l$, $l\in 2\mathbb{N}_0$.

From Theorem 1.4.10, we have an elaboration on the duality between $H^s_{\mathrm{sym}}(\mathbb{S}^2)$ and $H^{-s}_{\mathrm{sym}}(\mathbb{S}^2)$.

THEOREM 1.4.11. *For* $g\in L^2_{\mathrm{sym}}(\mathbb{S}^2)$, *the duality pairing*

$$G[\varphi] := \langle\varphi, g\rangle_{\star,0}, \quad \varphi\in H^s_{\mathrm{sym}}(\mathbb{S}^2), \tag{1.4.7}$$

defines a bounded linear functional on $H^s_{\mathrm{sym}}(\mathbb{S}^2)$, *i.e.,* $G\in H^{-s}_{\mathrm{sym}}(\mathbb{S}^2)$. *In particular,* $L^2_{\mathrm{sym}}(\mathbb{S}^2)$ *can be viewed as a subspace of the dual space* $H^{-s}_{\mathrm{sym}}(\mathbb{S}^2)$, $0\le s<\infty$, *and thus the linear span of the modified spherical harmonics is dense in* $H^{-s}_{\mathrm{sym}}(\mathbb{S}^2)$.

REMARK 1.4.12. By using a density argument, the above duality pairing (1.4.7) can be extended to bounded linear functionals in $H^{-s}_{\mathrm{sym}}(\mathbb{S}^2)$. In particular, for $\varphi\in H^s_{\mathrm{sym}}(\mathbb{S}^2)$ and $g\in H^{-s}_{\mathrm{sym}}(\mathbb{S}^2)$, we define

$$g[\varphi] := \langle\varphi, g\rangle_{\star,0}$$

with corresponding norm

$$\|g\|_{\star,-s} := \sup_{\varphi\in H^s_{\mathrm{sym}}(\mathbb{S}^2),\|\varphi\|_{\star,s}=1}\langle\varphi, g\rangle_{\star,0}.$$

We also note that $H^{-s}_{\mathrm{sym}}(\mathbb{S}^2)$ becomes a Hilbert space by extending the inner product (1.4.6) defined in Theorem 1.4.8 for $s\ge 0$ to $s<0$.

CHAPTER 2

Wavelet Frames on the Sphere

In this chapter, we begin by describing a framework to construct wavelet frames for Hilbert spaces in section 2.1. We then apply the framework to create tight wavelet frames for $L^2(\mathbb{S}^2)$ and $L^2_{\text{sym}}(\mathbb{S}^2)$ in sections 2.2 and 2.3 respectively. They are in turn used to obtain dual pairs of wavelet frames for the corresponding dual pairs of Sobolev spaces in sections 2.4 and 2.5 respectively.

2.1. Construction of Wavelet Frames for Hilbert Spaces

This section is organized as follows: subsection 2.1.1 presents basics behind the framework employed to construct wavelet frames for a subspace of a given Hilbert space; this framework is then combined with an MRA to yield wavelet frames for this Hilbert space in subsection 2.1.2.

2.1.1. Basic ideas of wavelet frame constructions on Hilbert spaces.

A general method to build wavelet bases for Hilbert spaces of functions was introduced in [**46**]. Examples constructed in [**46**] include wavelets for periodic functions over the real line, analytic functions on the unit disk and functions generated by Chebyshev polynomials. In this section, we extend the results of [**46**] to a frame setting.

Let $\mathcal{H}$ be a separable Hilbert space of real- or complex-valued functions defined on a set S. As $\mathcal{H}$ is separable, it has an orthonormal basis $\{e_j\}_{j\in\Lambda}$ of $\mathcal{H}$ for some countable index set Λ. In this section, the inner product and induced norm that $\mathcal{H}$ is equipped with are denoted as $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$ respectively. Recall that a countable collection of functions $\{f_\nu\}_{\nu\in I}$ in $\mathcal{H}$ is a *frame* for $\mathcal{H}$ if there exist positive constants C and D such that for every $f\in\mathcal{H}$,

$$C\|f\|^2 \le \sum_{\nu\in I} |\langle f, f_\nu\rangle|^2 \le D\|f\|^2. \tag{2.1.1}$$

The constants C and D, which are not unique, are called the *bounds* of the frame. In the case that $C = D = 1$, we say that $\{f_\nu\}_{\nu\in I}$ is a *normalized tight frame* for $\mathcal{H}$. As opposed to bases, frames provide overcomplete, yet stable, representations of functions. Due to this

property, frames usually offer superior denoising performances than bases in image and signal processing. The reader can find more details on the basic concepts of frames in [**34**].

Similar to the approach described in [**46**], a general method of constructing wavelet frames for $\mathcal{H}$ builds upon the orthonormal basis $\{e_j\}_{j\in\Lambda}$. The wavelet frame in this thesis is a family of functions $\{\psi_k(\cdot;\eta_{k,\nu}) : \nu \in \mathcal{J}'_k,\, k \geq -1\}$ that satisfies (2.1.1) and for every $k \geq -1$, has its functions taking the form

$$\psi_k(\cdot;\eta_{k,\nu}) = \mu_{k,\nu} \sum_{j\in\mathcal{J}_k} \gamma_{k,j}\, \overline{e_j(\eta_{k,\nu})}\, e_j, \quad \nu \in \mathcal{J}'_k,$$

where $\mathcal{J}_k$ is a finite subset of Λ, and $\eta_{k,\nu} \in S$, $\mu_{k,\nu}$, $\gamma_{k,j} \in \mathbb{C}\backslash\{0\}$ for all $j \in \mathcal{J}_k$, $\nu \in \mathcal{J}'_k$, with $\mathcal{J}'_k$ being a finite set such that $|\mathcal{J}_k| \leq |\mathcal{J}'_k|$. In the classical wavelet literature, for a fixed k, $\psi_k(\cdot;\eta_{k,\nu})$, $\nu \in \mathcal{J}'_k$, represent translates of a function. In general Hilbert spaces, for a fixed level k, we regard $\psi_k(\cdot;\eta_{k,\nu})$ as the generalized translates of a single function $\sum_{j\in\mathcal{J}_k} \gamma_{k,j} e_j$ by the operator $T_{\eta_{k,\nu}}$ (then multiplied by constants), where T_ζ is defined as the *generalized translation operator* acting on functions f of the form $f = \sum_{l=1}^{N} \alpha_{j_l} e_{j_l}$ such that

$$T_\zeta f \;:= \sum_{l=1}^{N} \alpha_{j_l} \overline{e_{j_l}(\zeta)}\, e_{j_l}. \tag{2.1.2}$$

To see the motivation behind the definition of this generalized translation operator (2.1.2), consider the special case where

$$\mathcal{H} := L^2[0, 2\pi), \quad S := [0, 2\pi), \quad \Lambda := \mathbb{Z},$$

$$e_j(x) := \exp(\mathrm{i}jx), \quad x \in S, \quad j \in \Lambda.$$

We next consider a function f formed by a finite weighted sum of the orthonormal functions, i.e., $f(x) := \sum_{j=0}^{N-1} \alpha_j e_j(x)$, $x \in S$, for some positive integer N. If we select a collection of nodes $\{\zeta_\nu\}_\nu$ for (2.1.2) where in this case, $\zeta_\nu := 2\pi\nu/N$, $\nu = 0, 1, \ldots, N-1$, then we have

$$\begin{aligned} T_{\zeta_\nu} f\,(x) &= \sum_{j=0}^{N-1} \alpha_j \overline{e_j(\zeta_\nu)}\, e_j(x) \\ &= \sum_{j=0}^{N-1} \alpha_j \exp(\mathrm{i}j(x-\zeta_\nu)) = f(x-\zeta_\nu), \end{aligned}$$

which is a translate of the function f in this special case. Furthermore, note that the matrix

$$A = [e_j(\zeta_\nu)]_{j,\nu=0}^{N-1} = [\exp(\mathrm{i}2\pi j\nu/N)]_{j,\nu=0}^{N-1}$$

has orthogonal rows and columns. Indeed, the basic setup of the construction method in [**46**] hinges on the existence of a finite-dimensional *square* matrix $[e_j(\zeta)]_{j\in\Lambda,\zeta\in S}$ which possesses orthogonal rows or orthogonal columns.

Let us now generalize the basic setup in [**46**, section 1] to the frame setting. Consider finite sets $\mathcal{I}$, $\mathcal{I}'$, $\mathcal{I}\subset\Lambda$ with $|\mathcal{I}|\leq|\mathcal{I}'|$. Suppose that there exist nodes $\zeta_\nu\in S$, $\nu\in\mathcal{I}'$, and a $|\mathcal{I}'|\times|\mathcal{I}'|$ diagonal weight matrix $D:=\operatorname{diag}(d_\nu)_{\nu\in\mathcal{I}'}$ with positive diagonal entries such that the $|\mathcal{I}|\times|\mathcal{I}'|$ *rectangular* matrix $A:=[e_j(\zeta_\nu)]_{j\in\mathcal{I},\nu\in\mathcal{I}'}$ satisfies

$$ADA^*=I. \tag{2.1.3}$$

In other words, we require the rows of the rectangular matrix A are orthonormal under appropriate weighting. Note that when $|\mathcal{I}|=|\mathcal{I}'|$, $A^*A=D^{-1}=\operatorname{diag}(d_\nu^{-1})_{\nu\in\mathcal{I}}$, which illustrates that (2.1.3) is a generalization of [**46**, equation 3]. We shall just focus on this generalization as it yields our desired constructions of wavelet frames on the sphere later.

Let us define

$$F:=A\,D^{1/2},$$

and in view of (2.1.3), the rows of matrix F are orthonormal, i.e.,

$$FF^*=I.$$

Now consider a function $\varphi\in\mathcal{H}$ of the form

$$\varphi=\sum_{j\in\mathcal{I}}a_je_j, \tag{2.1.4}$$

with $a_j\in\mathbb{C}$, $j\in\mathcal{I}$. With $[\varphi(\cdot;\zeta_\nu)]_{\nu\in\mathcal{I}'}$ and $[a_je_j]_{j\in\mathcal{I}}$ as column vectors, set

$$[\varphi(\cdot;\zeta_\nu)]_{\nu\in\mathcal{I}'}:=F^*[a_je_j]_{j\in\mathcal{I}}. \tag{2.1.5}$$

By (2.1.5),

$$\varphi(\cdot;\zeta_\nu)=\sqrt{d_\nu}\sum_{j\in\mathcal{I}}a_j\overline{e_j(\zeta_\nu)}\,e_j,\quad \nu\in\mathcal{I}'. \tag{2.1.6}$$

Note that $\varphi(\cdot;\zeta_\nu)$, $\nu\in\mathcal{I}'$, in (2.1.6) can be written respectively in the form

$$\varphi(\cdot;\zeta_\nu)=\sqrt{d_\nu}T_{\zeta_\nu}\varphi,\quad \nu\in\mathcal{I}'.$$

Remark 2.1.1. Denoting

$$U:=\operatorname{span}\left\{\varphi(\cdot;\zeta_\nu):\ \nu\in\mathcal{I}'\right\},$$

it is of interest to study such subspaces (when $\{a_j\}_{j\in\mathcal{I}}$ is a constant sequence) for the functions

$$\varphi(\cdot;\zeta_\nu) = \mu_\nu \sum_{j\in\mathcal{I}} \overline{e_j(\zeta_\nu)}\, e_j, \quad \nu \in \mathcal{I}',$$

as they act as kernel functions to provide the subspace U with the reproducing property, i.e., for all $f \in U$,

$$\langle f,\, \mu_\nu^{-1}\varphi(\cdot;\zeta_\nu)\rangle = f(\zeta_\nu), \quad \nu \in \mathcal{I}'. \tag{2.1.7}$$

Indeed, since $f \in U$ and $U \subseteq \text{span}\,\{e_j : j \in \mathcal{I}\}$, $f = \sum_{j\in\mathcal{I}} c_j e_j$ for some $[c_j]_{j\in\mathcal{I}} \in \mathbb{C}^{|\mathcal{I}|}$. Then for $\nu \in \mathcal{I}'$,

$$\begin{aligned}\langle f\,,\, \mu_\nu^{-1}\varphi(\cdot;\zeta_\nu)\rangle &= \sum_{j\in\mathcal{I}} c_j \sum_{j'\in\mathcal{I}} e_{j'}(\zeta_\nu)\langle e_j, e_{j'}\rangle \\ &= \sum_{j\in\mathcal{I}} c_j e_j(\zeta_\nu) = f(\zeta_\nu).\end{aligned}$$

The following result provides characterizations of a subspace spanned by finitely many e_j, $j \in \mathcal{I}$, in terms of $\{a_j\}_{j\in\mathcal{I}}$.

PROPOSITION 2.1.2. *Let V be a subspace defined by*

$$V := \text{span}\,\{e_j : j \in \mathcal{I}\}.$$

Then the following characterizations hold.

(i) $\text{span}\,\{\varphi(\cdot;\zeta_\nu) : \nu \in \mathcal{I}'\} = V$ if and only if $a_j \neq 0$, $j \in \mathcal{I}$.

(ii) $\{\varphi(\cdot;\zeta_\nu) : \nu \in \mathcal{I}'\}$ forms a frame for V with frame bounds C and D if and only if $C \leq |a_j|^2 \leq D$, $j \in \mathcal{I}$.

(iii) $\{\varphi(\cdot;\zeta_\nu) : \nu \in \mathcal{I}'\}$ forms a normalized tight frame for V if and only if $|a_j| = 1$, $j \in \mathcal{I}$.

PROOF. Since that $[\varphi(\cdot;\zeta_\nu)]_{\nu\in\mathcal{I}'} = F^*[a_j e_j]_{j\in\mathcal{I}}$ and $[a_j e_j]_{j\in\mathcal{I}} = F[\varphi(\cdot;\zeta_\nu)]_{\nu\in\mathcal{I}'}$ as $FF^* = I$,

$$\text{span}\left\{\varphi(\cdot;\zeta_\nu) : \nu \in \mathcal{I}'\right\} = \text{span}\,\{a_j e_j : j \in \mathcal{I}\} = V,$$

if and only if $a_j \neq 0$, $j \in \mathcal{I}$, proving statement (i).

For statements (ii) and (iii), consider $f \in V$. Then $f = \sum_{j\in\mathcal{I}} c_j e_j$ for some $[c_j]_{j\in\mathcal{I}} \in \mathbb{C}^{|\mathcal{I}|}$ and

$$\|f\|^2 = \sum_{j\in\mathcal{I}} |c_j|^2 = \|[c_j]_{j\in\mathcal{I}}\|_2^2. \tag{2.1.8}$$

Note that

$$[\langle f,\, \varphi(\cdot;\zeta_\nu)\rangle]_{\nu\in\mathcal{I}'} = [\sqrt{d_\nu}\sum_{j\in\mathcal{I}} e_j(\zeta_\nu)\overline{a_j}c_j]_{\nu\in\mathcal{I}'} = \overline{F}^*[\overline{a_j}c_j]_{j\in\mathcal{I}}.$$

Thus, since $FF^* = I$,

$$\sum_{\nu\in\mathcal{I}'} |\langle f,\, \varphi(\cdot;\zeta_\nu)\rangle|^2 = \|\overline{F}^*[\overline{a_j}c_j]_{j\in\mathcal{I}}\|_2^2 = \|[a_jc_j]_{j\in\mathcal{I}}\|_2^2. \tag{2.1.9}$$

We now only prove statement (ii) as statement (iii) is a special case of it when $C = D = 1$. Suppose that $C \le |a_j|^2 \le D$, $j \in \mathcal{I}$. Then we have for all $[c_j]_{j\in\mathcal{I}} \in \mathbb{C}^{|\mathcal{I}|}$,

$$C\sum_{j\in\mathcal{I}} |c_j|^2 \le \|[a_jc_j]_{j\in\mathcal{I}}\|_2^2 \le D\sum_{j\in\mathcal{I}} |c_j|^2.$$

By (2.1.8) and (2.1.9), the above inequalities imply that for all $f \in V$,

$$C\|f\|^2 \le \sum_{\nu\in\mathcal{I}'} |\langle f,\, \varphi(\cdot;\zeta_\nu)\rangle|^2 \le D\|f\|^2.$$

Conversely, suppose that $\{\varphi(\cdot;\zeta_\nu) : \nu \in \mathcal{I}'\}$ forms a frame for V with frame bounds C and D. Using the same arguments as above with (2.1.8) and (2.1.9),

$$C\|f\|^2 \le \sum_{\nu\in\mathcal{I}'} |\langle f,\, \varphi(\cdot;\zeta_\nu)\rangle|^2 \le D\|f\|^2, \quad \forall f \in V,$$

is equivalent to

$$C\sum_{j\in\mathcal{I}} |c_j|^2 \le \|[a_jc_j]_{j\in\mathcal{I}}\|_2^2 \le D\sum_{j\in\mathcal{I}} |c_j|^2, \quad \forall [c_j]_{j\in\mathcal{I}} \in \mathbb{C}^{|\mathcal{I}|}.$$

Selecting $f = e_{j'}$, $j' \in \mathcal{I}$ (i.e., $\{c_j\}_{j\in\mathcal{I}} = \{\delta_{j,j'}\}_{j\in\mathcal{I}}$), gives $C \le |a_{j'}|^2 \le D$, $j' \in \mathcal{I}$. This completes the proof. □

As we shall see, the reproducing property (2.1.7) shows that functions of the form

$$\varphi(\cdot;\zeta_\nu) = \mu_\nu \sum_{j\in\mathcal{I}} \overline{e_j(\zeta_\nu)}\, e_j, \quad \nu \in \mathcal{I}', \tag{2.1.10}$$

are well localized in some sense according to the following proposition. This localization property will prove useful for constructing wavelet functions which are localized at particular nodes on the sphere.

Proposition 2.1.3. *Define $\varphi(\cdot;\zeta_\nu)$ according to* (2.1.10) *for some collection of nodes $\{\zeta_\nu\}_{\nu\in\mathcal{I}'}$. If $V := \operatorname{span}\{e_j : j \in \mathcal{I}\}$, then the function $\varphi(\cdot;\zeta_\nu)$, $\nu \in \mathcal{I}'$, is well localized in the sense that*

$$\frac{\|\varphi(\cdot;\zeta_\nu)\|}{|\varphi(\zeta_\nu;\zeta_\nu)|} = \min\{\|f\| :\, f \in V,\, f(\zeta_\nu) = 1\}, \quad \nu \in \mathcal{I}'. \tag{2.1.11}$$

PROOF. For a fixed $\nu \in \mathcal{I}'$, by applying the Cauchy-Schwartz's inequality on the reproducing property (2.1.7), we obtain

$$1 = f(\zeta_\nu) = \langle f, \mu_\nu^{-1}\varphi(\cdot;\zeta_\nu)\rangle \leq \|f\|\|\mu_\nu^{-1}\varphi(\cdot;\zeta_\nu)\|,$$

and thus

$$\|\mu_\nu^{-1}\varphi(\cdot;\zeta_\nu)\|^{-1} \leq \|f\|, \quad \forall f \in V,\, f(\zeta_\nu) = 1.$$

To establish (2.1.11), we show that the above lower bound on $\|f\|$ can be attained when we select

$$f(\cdot) = \frac{\varphi(\cdot;\zeta_\nu)}{\varphi(\zeta_\nu;\zeta_\nu)}. \tag{2.1.12}$$

Note that $f \in V$ and $f(\zeta_\nu) = 1$, by the reproducing property,

$$\langle f,\, \mu_\nu^{-1}\varphi(\cdot;\zeta_\nu)\rangle = f(\zeta_\nu) = 1, \quad \nu \in \mathcal{I}'.$$

Using the expression of f in (2.1.12), we attain the required lower bound

$$\|f\| = \frac{\|\varphi(\cdot;\zeta_\nu)\|}{|\varphi(\zeta_\nu;\zeta_\nu)|} = \|\mu_\nu^{-1}\varphi(\cdot;\zeta_\nu)\|^{-1}.$$

□

2.1.2. Multiresolution analysis and wavelets on Hilbert spaces.

We shall now describe a strategy for obtaining a multiresolution analysis of $\mathcal{H}$. Let $\{\mathcal{I}_k\}_{k\geq 0}$ be an increasing sequence of finite subsets of Λ, i.e., $\mathcal{I}_k \subset \mathcal{I}_{k+1}$, $k \geq 0$. Suppose that for every $k \geq 0$, there exist $\zeta_{k,\nu} \in S$, $\nu \in \mathcal{I}'_k$, where $\mathcal{I}'_k$ is a finite set with $|\mathcal{I}_k| \leq |\mathcal{I}'_k|$, and a diagonal $|\mathcal{I}'_k| \times |\mathcal{I}'_k|$ matrix $D_k := \operatorname{diag}(d_{k,\nu})_{\nu\in\mathcal{I}'_k}$ with positive diagonal entries such that the $|\mathcal{I}_k| \times |\mathcal{I}'_k|$ matrix $A_k := [e_j(\zeta_{k,\nu})]_{j\in\mathcal{I}_k,\nu\in\mathcal{I}'_k}$ satisfies

$$A_k D_k A_k^* = I.$$

Note that $F_k F_k^* = I$, when we define

$$F_k := A_k\, D_k^{1/2}.$$

For $a_{k,j} \in \mathbb{C}$, $j \in \mathcal{I}_k$, set $\varphi_k := \sum_{j\in\mathcal{I}_k} a_{k,j}e_j$ and define similarly $\varphi_k(\cdot;\zeta_{k,\nu})$, $\nu \in \mathcal{I}'_k$, by (2.1.6), i.e., $[\varphi_k(\cdot;\zeta_{k,\nu})]_{\nu\in\mathcal{I}'_k} := F_k^*[a_{k,j}e_j]_{j\in\mathcal{I}_k}$, or

$$\varphi_k(\cdot;\zeta_{k,\nu}) = \sqrt{d_{k,\nu}} \sum_{j\in\mathcal{I}_k} a_{k,j}\overline{e_j(\zeta_{k,\nu})}\, e_j, \quad \nu \in \mathcal{I}'_k. \tag{2.1.13}$$

A sequence of finite-dimensional subspaces $\{V_k\}_{k\geq 0}$ in $\mathcal{H}$ is a *multiresolution analysis* (MRA) of $\mathcal{H}$ if

(MRA1) For every $k \geq 0$, there exist functions $\varphi_k(\cdot;\zeta_{k,\nu})$, $\nu \in \mathcal{J}_k'$, of the form (2.1.13) such that $\{\varphi_k(\cdot;\zeta_{k,\nu}) : \nu \in \mathcal{I}_k'\}$ spans V_k.

(MRA2) For all $k \geq 0$, $V_k \subset V_{k+1}$.

(MRA3) $\text{clos}_{\mathcal{H}}\left(\bigcup_{k\geq 0} V_k\right) = \mathcal{H}$.

We call the functions $\varphi_k(\cdot;\zeta_{k,\nu})$, $\nu \in \mathcal{I}_k'$, $k \geq 0$, *scaling functions*. For $k \geq 0$, define

$$V_k := \text{span}\,\{e_j : j \in \mathcal{I}_k\}.$$

Then (MRA1) can be obtained by letting $a_{k,j} \neq 0$, $j \in \mathcal{I}_k$, $k \geq 0$, so that by Proposition 2.1.2, $V_k = \text{span}\,\{\varphi_k(\cdot;\zeta_{k,\nu}) : \nu \in \mathcal{I}_k'\}$. Characterizations of (MRA2) and (MRA3) are given in terms of the index sets $\mathcal{I}_k$, $k \geq 0$, as follows.

PROPOSITION 2.1.4. ([**46**, Propositions 4,5]) *The condition in* (MRA2) *is equivalent to*

$$\mathcal{I}_k \subset \mathcal{I}_{k+1}, \quad k \geq 0.$$

The condition in (MRA3) *holds if and only if*

$$\bigcup_{k\geq 0} \mathcal{I}_k = \Lambda.$$

PROOF. The reader may refer to the proofs of [**46**, Propositions 4,5]. □

Having an MRA of $\mathcal{H}$, we are in a position to construct wavelets. Define the wavelet subspace W_k as the orthogonal complement of V_k in V_{k+1}. Since $V_k = \text{span}\,\{e_j : j \in \mathcal{I}_k\}$, defining $\mathcal{J}_k := \mathcal{I}_{k+1}\backslash\mathcal{I}_k$, it then follows that

$$W_k = \text{span}\,\{e_j : j \in \mathcal{J}_k\}. \tag{2.1.14}$$

Suppose similarly that for every $k \geq 0$, there exist $\eta_{k,\nu} \in S$, $\nu \in \mathcal{J}_k'$, where $\mathcal{J}_k'$ is a finite set with $|\mathcal{J}_k| \leq |\mathcal{J}_k'|$, and a diagonal $|\mathcal{J}_k'| \times |\mathcal{J}_k'|$ matrix $\hat{D}_k := \text{diag}(\hat{d}_{k,\nu})_{\nu\in\mathcal{J}_k'}$ with positive diagonal entries, such that the $|\mathcal{J}_k| \times |\mathcal{J}_k'|$ matrix $B_k := [e_j(\eta_{k,\nu})]_{j\in\mathcal{J}_k,\nu\in\mathcal{J}_k'}$ satisfies

$$B_k\hat{D}_kB_k^* = I.$$

Note that $G_kG_k^* = I$, when we define

$$G_k := B_k\,\hat{D}_k{}^{1/2}.$$

For $b_{k,j} \in \mathbb{C}$, $j \in \mathcal{J}_k$, define the function $\psi_k := \sum_{j\in\mathcal{J}_k} b_{k,j}e_j$ as in (2.1.4). Then putting $\mathcal{I} = \mathcal{J}_k$, $\mathcal{I}' = \mathcal{J}_k'$, $A = B_k$, $D = \hat{D}_k$ and $a_j = b_{k,j}$ for $j \in \mathcal{J}_k$, we obtain $\psi_k(\cdot;\eta_{k,\nu})$, $\nu \in \mathcal{J}_k'$,

from (2.1.6), i.e., define $[\psi_k(\cdot;\eta_{k,\nu})]_{\nu\in\mathcal{J}_k'} := G_k^*[b_{k,j}e_j]_{j\in\mathcal{J}_k}$, or alternatively

$$\psi_k(\cdot;\eta_{k,\nu}) = \sqrt{\hat{d}_{k,\nu}} \sum_{j\in\mathcal{J}_k} b_{k,j}\overline{e_j(\eta_{k,\nu})}\, e_j, \quad \nu \in \mathcal{J}_k'.$$

The functions $\psi_k(\cdot;\eta_{k,\nu})$, $\nu \in \mathcal{J}_k'$, are known as *wavelets*. As seen in (2.1.6), such functions can be regarded as generalized translates of a single function.

Similar to Proposition 2.1.2, we next have a characterization of frame properties of $\{\psi_k(\cdot;\eta_{k,\nu}) : \nu \in \mathcal{J}_k'\}$ for W_k in terms of the values of $b_{k,j}$, $j \in \mathcal{J}_k$.

THEOREM 2.1.5. *For $k \geq 0$, let W_k be defined by* (2.1.14)*. Then we have the following characterizations.*

(i) $\operatorname{span}\{\psi_k(\cdot;\eta_{k,\nu}) : \nu \in \mathcal{J}_k'\} = W_k$ if and only if $b_{k,j} \neq 0$, $j \in \mathcal{J}_k$.

(ii) $\{\psi_k(\cdot;\eta_{k,\nu}) : \nu \in \mathcal{J}_k'\}$ is a frame for W_k with frame bounds C and D if and only if

$$C \leq |b_{k,j}|^2 \leq D, \quad j \in \mathcal{J}_k.$$

(iii) $\{\psi_k(\cdot;\eta_{k,\nu}) : \nu \in \mathcal{J}_k'\}$ is a normalized tight frame for W_k if and only if

$$|b_{k,j}| = 1, \quad j \in \mathcal{J}_k.$$

PROOF. The result follows from a direct application of Proposition 2.1.2. □

Thus, starting from an MRA $\{V_k\}_{k\geq 0}$, for every $k \geq 0$, with an appropriate choice of $b_{k,j}$, $j \in \mathcal{J}_k$, we obtain (normalized tight) frames for the wavelet subspaces W_k. The scaling functions $\phi_k(\cdot;\zeta_{k,\nu})$, $\nu \in \mathcal{I}_k'$, and the wavelets $\psi_k(\cdot;\eta_{k,\nu})$, $\nu \in \mathcal{J}_k'$, aim to provide a good representation of functions in V_k and W_k respectively. Such a representation is used when we decompose, by virtue of W_k being the orthogonal complement of V_k in V_{k+1}, a function $f_{k+1} \in V_{k+1}$ into $f_{k+1} = f_k + g_k$, where $f_k \in V_k$, $g_k \in W_k$. The function f_{k+1} can also be recovered perfectly from the functions f_k and g_k.

After obtaining (normalized tight) frames for all the wavelet subspaces as above, it remains to show that the entire collection of these (normalized tight) frames forms a (normalized tight) frame for $\mathcal{H}$. Due to W_k being the orthogonal complement of V_k in V_{k+1}, we have the orthogonal decomposition $\mathcal{H} = V_0 \oplus^\perp W_0 \oplus^\perp W_1 \oplus^\perp \ldots$. To simplify notations, we denote $W_{-1} := V_0$, $\mathcal{J}_{-1} := \mathcal{I}_0$, $\mathcal{J}_{-1}' := \mathcal{I}_0'$, $\eta_{-1,\nu} := \zeta_{0,\nu}$ and $\psi_{-1}(\cdot;\eta_{-1,\nu}) := \varphi_0(\cdot;\zeta_{0,\nu})$ for $\nu \in \mathcal{J}_{-1}'$.

THEOREM 2.1.6. *If for every $k \geq -1$, $\{\psi_k(\cdot;\eta_{k,\nu}) : \nu \in \mathcal{J}_k'\}$ forms a frame for W_k with common frame bounds, then the collection $\{\psi_k(\cdot;\eta_{k,\nu}) : \nu \in \mathcal{J}_k',\ k \geq -1\}$ forms a frame for $\mathcal{H}$ with the same frame bounds. In particular, if for every $k \geq -1$, $\{\psi_k(\cdot;\eta_{k,\nu}) : \nu \in \mathcal{J}_k'\}$ forms*

a normalized tight frame for W_k, then the collection $\{\psi_k(\cdot;\eta_{k,\nu}) : \nu \in \mathcal{J}_k', k \geq -1\}$ forms a normalized tight frame for $\mathcal{H}$.

PROOF. By Proposition 2.1.4, we have $\Lambda = \bigcup_{k\geq 0} \mathcal{I}_k$. Since $\mathcal{J}_k = \mathcal{I}_{k+1}\backslash\mathcal{I}_k$ for every $k \geq 0$, Λ can be formed by the disjoint union of all $\mathcal{J}_k$'s, i.e.,

$$\Lambda = \bigcup_{k\geq -1} \mathcal{J}_k. \tag{2.1.15}$$

For each $k \geq -1$, by definition, $\{e_j : j \in \mathcal{J}_k\}$ is an orthonormal basis for W_k, so

$$\|g_k\|^2 = \sum_{j\in\mathcal{J}_k} |\langle g_k, e_j\rangle|^2, \quad g_k \in W_k.$$

The subspace W_k is finite-dimensional and hence closed. So any $f \in \mathcal{H}$ can be written as $f = g_k + h_k$, where $g_k \in W_k$, $h_k \in W_k^{\perp}$. Thus

$$\sum_{j\in\mathcal{J}_k} |\langle f, e_j\rangle|^2 = \sum_{j\in\mathcal{J}_k} |\langle g_k, e_j\rangle|^2 = \|g_k\|^2.$$

Since $\{e_j : j \in \Lambda\}$ is an orthonormal basis for $\mathcal{H}$, and applying (2.1.15), for any $f \in \mathcal{H}$,

$$\|f\|^2 = \sum_{k\geq -1}\sum_{j\in\mathcal{J}_k} |\langle f, e_j\rangle|^2 = \sum_{k\geq -1} \|g_k\|^2.$$

If for every $k \geq -1$, $\{\psi_k(\cdot;\eta_{k,\nu}) : \nu \in \mathcal{J}_k'\}$ forms a frame for W_k with common frame bounds, say C and D, we then obtain

$$C\sum_{k\geq -1} \|g_k\|^2 \leq \sum_{k\geq -1}\sum_{\nu\in\mathcal{J}_k'} |\langle g_k, \psi_k(\cdot;\eta_{k,\nu})\rangle|^2 \leq D\sum_{k\geq -1} \|g_k\|^2.$$

Noting that $\|f\|^2 = \sum_{k\geq -1} \|g_k\|^2$, and also $f = g_k + h_k$, $g_k \in W_k$, $h_k \in W_k^{\perp}$, for every $k \geq -1$,

$$\sum_{k\geq -1}\sum_{\nu\in\mathcal{J}_k'} |\langle f, \psi_k(\cdot;\eta_{k,\nu})\rangle|^2 = \sum_{k\geq -1}\sum_{\nu\in\mathcal{J}_k'} |\langle g_k, \psi_k(\cdot;\eta_{k,\nu})\rangle|^2,$$

we obtain the desired result that for any $f \in \mathcal{H}$,

$$C\|f\|^2 \leq \sum_{k\geq -1}\sum_{\nu\in\mathcal{J}_k'} |\langle f, \psi_k(\cdot;\eta_{k,\nu})\rangle|^2 \leq D\|f\|^2.$$

□

REMARK 2.1.7. We like to point out that [**47**] (an upcoming sequel to [**46**]) also constructs wavelet frames for Hilbert spaces of functions. One main difference between [**47**] and this work lies in their respective basic setups. In [**47**], square matrices with orthogonal rows or columns are used to construct the scaling functions, whereas as seen in (2.1.3), rectangular matrices with (weighted) orthogonal rows are considered for our construction of scaling functions. In

fact, the motivation behind the rectangular setup is to enable the construction of wavelet frames on the sphere using spherical harmonics. This shall be explained in the subsequent sections.

2.2. Constructing Tight Wavelet Frames for $L^2(\mathbb{S}^2)$ with SH

This section illustrates how the theoretical framework in section 2.1 can be applied to construct wavelet frames for $L^2(\mathbb{S}^2)$. Subsection 2.2.1 presents how to select the so-called Gauss-Legendre quadrature nodes so that the corresponding matrix of discretized spherical harmonics (SH) satisfies the framework with weighted orthogonal rows. Building upon that, subsection 2.2.2 presents how MRA-based tight wavelet frames for $L^2(\mathbb{S}^2)$ are constructed.

2.2.1. Gauss-Legendre quadrature nodes for SH.

Let us now see how to apply our basic setup in the earlier section to construct wavelet frames from spherical harmonics. In this case, we set

$$\mathcal{H} = L^2(\mathbb{S}^2), \quad S = \mathbb{S}^2, \quad \Lambda = \{(m,l) \in \mathbb{Z} \times \mathbb{N}_0 \,:\, |m| \leq l\}\,,$$
$$e_{\mathbf{j}} := Y_{\mathbf{j}} := Y_l^m, \quad \mathbf{j} := (m,l) \in \Lambda, \tag{2.2.1}$$

where the inner product on $L^2(\mathbb{S}^2)$ was defined earlier in (1.3.2) and Y_l^m, $(m,l) \in \Lambda$, are spherical harmonics which can be expressed as

$$Y_l^m(\theta,\phi) := \Theta_l^m(\cos\theta)\Phi^m(\phi),\ \theta \in [0,\pi],\ \phi \in [0,2\pi),\ (m,l) \in \Lambda, \tag{2.2.2}$$

where

$$\Theta_l^m(\cos\theta) := (-1)^m \sqrt{\frac{(2l+1)(l-m)!}{2(l+m)!}} P_l^m(\cos\theta),\ \Phi^m(\phi) := \frac{1}{\sqrt{2\pi}} \exp(\mathrm{i}m\phi).$$

For convenience, we use interchangeably the notations $Y_{\mathbf{j}}(\mathbf{u})$, $Y_l^m(\mathbf{u})$ $Y_l^m(\theta,\phi)$ in (2.2.1), where it is understood that (θ,ϕ) are the corresponding spherical co-ordinates of the vector $\mathbf{u} \in \mathbb{S}^2$ according to the parametrization given in (1.3.1).

We consider

$$\mathcal{I} := \{(m,l) \in \mathbb{Z} \times \mathbb{N}_0 \,:\, |m| \leq l,\, L_0 \leq l \leq L\}\,, \tag{2.2.3}$$

and define its corresponding spanning subspace V as

$$\begin{aligned} V :&= \operatorname{span}\,\{e_{\mathbf{j}} \,:\, \mathbf{j} \in \mathcal{I}\} \\ &= \operatorname{span}\,\{Y_l^m \,:\, |m| \leq l,\, L_0 \leq l \leq L\}\,. \end{aligned} \tag{2.2.4}$$

Recall that to construct wavelet frames to span the subspace V, given $\mathcal{I}$ in (2.2.3), we seek nodes $\zeta_\nu \in S$, $\nu \in \mathcal{I}'$ (with $\mathcal{I}'$ a finite set with $|\mathcal{I}| \leq |\mathcal{I}'|$), and a $|\mathcal{I}'| \times |\mathcal{I}'|$ diagonal weight

matrix $D := \text{diag}(d_\nu)_{\nu\in\mathcal{I}'}$ with positive diagonal entries such that the $|\mathcal{I}| \times |\mathcal{I}'|$ rectangular matrix $A := [Y_{\mathbf{j}}(\zeta_\nu)]_{\mathbf{j}\in\mathcal{I},\nu\in\mathcal{I}'}$ satisfies

$$ADA^* = I,$$

which is equivalent to

$$\sum_{\nu\in\mathcal{I}'} Y_{\mathbf{j}'}(\zeta_\nu)\overline{Y_{\mathbf{j}}(\zeta_\nu)}d_\nu = \delta_{\mathbf{j},\mathbf{j}'}, \quad \mathbf{j},\mathbf{j}' \in \mathcal{I}. \tag{2.2.5}$$

To accomplish (2.2.5), an appropriate choice of nodes $\{\zeta_\nu\}_{\nu\in\mathcal{I}'}$ is given by a tensor product of azimuth and zenith points:

$$\zeta_\nu := (\theta_\alpha, \phi_\beta), \quad \nu := (\alpha,\beta) \in \mathcal{I}', \tag{2.2.6}$$

where the azimuth points are given as

$$\theta_\alpha := \cos^{-1}(x_\alpha), \quad \alpha = 0, 1, \ldots, L, \tag{2.2.7}$$

with $\{x_\alpha\}_{\alpha=0}^L$ being the zeros of the Legendre polynomial P_{L+1} and the zenith points are given as

$$\phi_\beta := \frac{2\pi\beta}{2L+1}, \quad \beta = 0, 1, \ldots, 2L. \tag{2.2.8}$$

Thus the set $\mathcal{I}'$ is given by

$$\mathcal{I}' := \{(\alpha,\beta) : \alpha = 0, 1, \ldots, L, \quad \beta = 0, 1, \ldots, 2L\}.$$

If we enumerate the nodes in the following manner:

$$\{(\theta_\alpha, \phi_\beta)\}_{(\alpha,\beta)} := \{(\theta_0,\phi_0), \ldots, (\theta_0,\phi_{2L}), \ldots\ldots, (\theta_L,\phi_0), \ldots, (\theta_L,\phi_{2L})\}, \tag{2.2.9}$$

then we shall see that the required diagonal weight matrix D given by

$$D = \frac{2\pi}{2L+1}\text{diag}(w_0,\ldots,w_0,\, w_1,\ldots,w_1,\ldots\ldots,w_L,\ldots,w_L), \tag{2.2.10}$$

where $\{w_\alpha\}_{\alpha=0}^L$ are the corresponding Gauss-quadrature weights of the Legendre polynomial P_{L+1} defined in (1.3.6), with each unique value of w_α appearing $2L+1$ times in the matrix D.

To see why the above choice of nodes and weight matrix is appropriate, we consider the separable structure of $Y_j = Y_l^m$ as in (1.3.9) and (1.3.10) and obtain that

$$\begin{aligned}\sum_{\nu\in\mathcal{I}'} Y_{\mathbf{j}'}(\zeta_\nu)\overline{Y_{\mathbf{j}}(\zeta_\nu)}d_\nu &= \sum_{(\alpha,\beta)\in\mathcal{I}'} Y_{l'}^{m'}(\theta_\alpha,\phi_\beta)\overline{Y_l^m(\theta_\alpha,\phi_\beta)}w_\alpha\cdot\frac{2\pi}{2L+1}\\ &= \left\{\sum_{\alpha=0}^{L}\Theta_l^m(x_\alpha)\Theta_{l'}^{m'}(x_\alpha)w_\alpha\right\}\cdot\left\{\sum_{\beta=0}^{2L}\Phi^{m'}(\phi_\beta)\overline{\Phi^m(\phi_\beta)}\cdot\frac{2\pi}{2L+1}\right\}\\ &= \left\{\sum_{\alpha=0}^{L}\Theta_l^m(x_\alpha)\Theta_{l'}^{m}(x_\alpha)w_\alpha\right\}\cdot\delta_{m,m'}\\ &= \delta_{l,l'}\cdot\delta_{m,m'} = \delta_{\mathbf{j},\mathbf{j}'},\end{aligned}$$

for all $\mathbf{j}$, $\mathbf{j}'\in\mathcal{I}$, where we have made use of the following results:

$$\sum_{\beta=0}^{2L}\overline{\Phi^m(\phi_\beta)}\Phi^{m'}(\phi_\beta)\cdot\frac{2\pi}{2L+1} = \delta_{m,m'},\quad \forall|m|,\,|m'|\le L, \tag{2.2.11}$$

$$\sum_{\alpha=0}^{L}\Theta_l^m(x_\alpha)\Theta_{l'}^m(x_\alpha)w_\alpha = \delta_{l,l'},\quad \forall|m|\le\min(l,l'),\quad 0\le l,\,l'\le L. \tag{2.2.12}$$

It is easy to verify (2.2.11) by using standard geometric sum arguments. To show why (2.2.12) is true, recall from (1.3.12) that for any fixed m with $|m|\le\min(l,l')$,$\int_{-1}^{1}\Theta_l^m(x)\Theta_{l'}^m(x)\,dx = \delta_{l,l'}$, $0\le l,\,l'\le L$. So we need to establish for any fixed m with $|m|\le\min(l,l')$,

$$\sum_{\alpha=0}^{L}\Theta_l^m(x_\alpha)\Theta_{l'}^m(x_\alpha)w_\alpha = \int_{-1}^{1}\Theta_l^m(x)\Theta_{l'}^m(x)\,dx,\quad 0\le l,\,l'\le L. \tag{2.2.13}$$

The reason behind the validity of (2.2.13) stems from the following classical Gaussian quadrature result in numerical integration.

THEOREM 2.2.1. ([**79**]) *Let $\{p_l\}_{l=0}^n$ be a set of orthogonal polynomials on $[-1,1]$ with respect to the inner product*

$$\langle f,g\rangle_w = \int_{-1}^{1} f(x)g(x)w(x)\,dx,$$

where w is a non-negative function defined on $[-1,1]$. Let $\{x_\alpha\}_{\alpha=0}^{n-1}$ be the zeros of the polynomial p_n. Then the quadrature rule

$$I(h) := \int_{-1}^{1} h(x)w(x)\,dx = \sum_{\alpha=0}^{n-1} h(x_\alpha)w_\alpha, \tag{2.2.14}$$

holds exactly for all polynomials h of degree at most $2n-1$, where

$$w_\alpha = \int_{-1}^{1} L_\alpha(x) w(x)\, dx, \quad L_\alpha(x) := \prod_{k=0, k\neq\alpha}^{n-1} \frac{(x - x_k)}{(x_\alpha - x_k)}.$$

Note that when $w(x) \equiv 1$, $\{p_l\}_{l=0}^n$ are exactly the Legendre polynomials $\{P_l\}_{l=0}^n$. Given $0 \le l,\, l' \le L$, and a fixed m with $|m| \le \min(l, l')$, by the definition of Θ_l^m, the product $\Theta_l^m \Theta_{l'}^m$ is a polynomial of degree $l + l' \le 2L$. Thus by (2.2.14), if we set

$$h := \Theta_l^m \Theta_{l'}^m, \quad w \equiv 1, \quad n = L + 1,$$

$\{x_\alpha\}_{\alpha=0}^L$ as the zeros of P_{L+1} and $\{w_\alpha\}_{\alpha=0}^L$ as the corresponding quadrature weights, we finally obtain (2.2.13).

REMARK 2.2.2. We note that the cardinality of the set of nodes $\{\zeta_\nu\}_{\nu\in\mathcal{I}'}$ defined in (2.2.6) is $(L+1)\cdot(2L+1)$, which is only dependent on the largest order L of the spherical harmonics in the subspace V defined in (2.2.4). Actually, there are many other choices of nodes $\{\zeta_\nu\}_{\nu\in\mathcal{I}'}$ to achieve (2.2.11) and (2.2.12). For instance, we may replace θ_α in (2.2.7) by $\theta_\alpha := \cos^{-1}(x_\alpha)$, with $\{x_\alpha\}_{\alpha=0}^N$ being the zeros of the Legendre polynomial P_{N+1}, $N \ge L$. We may also replace ϕ_β in (2.2.8) with $\phi_\beta = \frac{2\pi\beta}{M+1}$, $\beta = 0, 1, \ldots, M$, $M \ge 2L$. Thus the corresponding set $\mathcal{I}'$ is given by

$$\mathcal{I}' := \{(\alpha, \beta) \,:\, \alpha = 0, 1, \ldots, N, \quad \beta = 0, 1, \ldots, M\}.$$

If we enumerate $\{(\theta_\alpha, \phi_\beta)\}_{(\alpha,\beta)}$ as in (2.2.9), then the corresponding diagonal weight matrix D is given by

$$D = \frac{2\pi}{M+1} \mathrm{diag}(w_0, \ldots, w_0,\, w_1, \ldots, w_1, \ldots\ldots, w_N, \ldots, w_N),$$

where $\{w_\alpha\}_{\alpha=0}^N$ are the corresponding Gauss-quadrature weights of the Legendre polynomial P_{N+1}, $N \ge L$, with each unique value of w_α appearing $M+1$ times in the matrix D, $M \ge 2L$. For ease of reference, we will refer such nodes $\{\zeta_\nu\}_{\nu\in\mathcal{I}'}$ defined in (2.2.6) and its above variations as the *Gauss-Legendre nodes for the spherical harmonics.*

2.2.2. MRA of tight wavelet frames for $L^2(\mathbb{S}^2)$ with SH.

In this subsection, we will use the idea of Gauss-Legendre nodes for the spherical harmonics to obtain an MRA of tight wavelet frames for $L^2(\mathbb{S}^2)$ based on the spherical harmonics. We consider

$$\mathcal{I}_k := \{(m, l) \in \mathbb{Z} \times \mathbb{N}_0 \,:\, |m| \le l,\ 0 \le l \le L_k\}, \quad k \ge 0, \tag{2.2.15}$$

where $\{L_k\}_{k\geq 0}$ is an increasing sequence of non-negative integers increasing to infinity. Define its corresponding spanning subspace V_k as

$$\begin{aligned} V_k :&= \text{span } \{e_{\mathbf{j}} : \mathbf{j} \in \mathcal{I}_k\} \\ &= \text{span } \{Y_l^m : |m| \leq l, 0 \leq l \leq L_k\}. \end{aligned} \tag{2.2.16}$$

To construct scaling functions $\{\varphi_k(\cdot;\zeta_{k,\nu}) : \nu \in \mathcal{I}_k'\}$ to span the above subspaces V_k, we seek nodes $\zeta_{k,\nu} \in S$, $\nu \in \mathcal{I}_k'$, (with $\mathcal{I}_k'$ a finite set with $|\mathcal{I}_k| \leq |\mathcal{I}_k'|$) and a $|\mathcal{I}_k'| \times |\mathcal{I}_k'|$ diagonal weight matrix $D_k := \text{diag}(d_{k,\nu})_{\nu\in\mathcal{I}_k'}$ with positive diagonal entries such that the $|\mathcal{I}_k| \times |\mathcal{I}_k'|$ rectangular matrix $A_k := [Y_{\mathbf{j}}(\zeta_{k,\nu})]_{\mathbf{j}\in\mathcal{I}_k,\nu\in\mathcal{I}_k'}$ satisfies

$$A_k D_k A_k^* = I.$$

Replacing ζ_ν by $\zeta_{k,\nu}$ in (2.2.6), an appropriate choice of nodes $\{\zeta_{k,\nu}\}_{\nu\in\mathcal{I}_k'}$ is given by

$$\zeta_{k,\nu} := (\theta_{k,\alpha}, \phi_{k,\beta}), \quad \nu := (\alpha,\beta) \in \mathcal{I}_k', \tag{2.2.17}$$

where $\{\theta_{k,\alpha}\}_\alpha$, $\{\phi_{k,\beta}\}_\beta$ and D_k replace $\{\theta_\alpha\}_\alpha$, $\{\phi_\beta\}_\beta$ and D in (2.2.7), (2.2.8) and (2.2.10) respectively by setting L as L_k. For each $k \geq 0$, the scaling functions $\varphi_k(\cdot;\zeta_{k,\nu})$, $\nu \in \mathcal{I}_k'$, of V_k are then defined as

$$\varphi_k(\cdot;\zeta_{k,\nu}) := \sqrt{d_{k,\nu}} \sum_{\mathbf{j}\in\mathcal{I}_k} \overline{Y_{\mathbf{j}}(\zeta_{k,\nu})} Y_{\mathbf{j}}, \quad \nu \in \mathcal{I}_k',$$

which yield a normalized tight frame for V_k by Proposition 2.1.2.

For $k \geq 0$, since $W_k = V_{k+1}\backslash V_k$, $\mathcal{J}_k = \mathcal{I}_{k+1}\backslash\mathcal{I}_k$, it follows that

$$\begin{aligned} W_k :&= \text{span } \{e_{\mathbf{j}} : \mathbf{j} \in \mathcal{J}_k\} \\ &= \text{span } \{Y_l^m : |m| \leq l, L_k < l \leq L_{k+1}\}. \end{aligned}$$

To construct wavelets $\{\psi_k(\cdot;\eta_{k,\nu}) : \nu \in \mathcal{J}_k'\}$ to span the above subspace W_k, we seek nodes $\eta_{k,\nu} \in S$, $\nu \in \mathcal{J}_k'$, (with $\mathcal{J}_k'$ a finite set with $|\mathcal{J}_k| \leq |\mathcal{J}_k'|$) and a $|\mathcal{J}_k'| \times |\mathcal{J}_k'|$ diagonal weight matrix $\hat{D}_k := \text{diag}(\hat{d}_{k,\nu})_{\nu\in\mathcal{J}_k'}$ with positive diagonal entries such that the $|\mathcal{J}_k| \times |\mathcal{J}_k'|$ rectangular matrix $B_k := [Y_{\mathbf{j}}(\eta_{k,\nu})]_{\mathbf{j}\in\mathcal{J}_k,\nu\in\mathcal{J}_k'}$ satisfies

$$B_k \hat{D}_k B_k^* = I.$$

Replacing $\zeta_{k,\nu}$ by $\eta_{k,\nu}$ in (2.2.17), an appropriate choice of nodes $\{\eta_{k,\nu}\}_{\nu\in\mathcal{J}_k'}$ is given by

$$\eta_{k,\nu} := (\hat{\theta}_{k,\alpha}, \hat{\phi}_{k,\beta}), \quad \nu := (\alpha,\beta) \in \mathcal{J}_k',$$

where $\left\{\hat{\theta}_{k,\alpha}\right\}_\alpha$, $\left\{\hat{\phi}_{k,\beta}\right\}_\beta$ and $\hat{D}_k$ replace $\{\theta_\alpha\}_\alpha$, $\{\phi_\beta\}_\beta$ and D in (2.2.7), (2.2.8) and (2.2.10) respectively by setting L as L_{k+1}. For each $k \geq 0$, the wavelet functions $\psi_k(\cdot;\eta_{k,\nu})$, $\nu \in \mathcal{J}'_k$, of W_k are then defined as

$$\psi_k(\cdot;\eta_{k,\nu}) := \sqrt{\hat{d}_{k,\nu}} \sum_{\mathbf{j}\in\mathcal{J}_k} \overline{Y_{\mathbf{j}}(\zeta_{k,\nu})} Y_{\mathbf{j}}, \quad \nu \in \mathcal{J}'_k,$$

which yield a normalized tight frame for W_k by Proposition 2.1.2. Therefore, by Theorem 2.1.6, $\{\psi_k(\cdot;\eta_{k,\nu}) : \nu \in \mathcal{J}'_k, k \geq -1\}$ forms a normalized tight wavelet frame for $L^2(\mathbb{S}^2)$.

2.3. Constructing Tight Wavelet Frames for $L^2_{\mathbf{sym}}(\mathbb{S}^2)$ Using Modified SH

In this section, we discuss how to analogously design normalized tight wavelet frames for $L^2_{\mathrm{sym}}(\mathbb{S}^2)$ using the modified spherical harmonics $\left\{\tilde{Y}_l^m\right\}_{m,l}$ as defined in (1.3.17). Recall from (1.3.18) that

$$L^2_{\mathrm{sym}}(\mathbb{S}^2) = \mathrm{clos}_{L^2(\mathbb{S}^2)}\left(\mathrm{span}\left\{\tilde{Y}_l^m : |m| \leq l,\, l \in 2\mathbb{N}_0\right\}\right),$$

and from (1.3.17) that for $|m| \leq l$, $l \in 2\mathbb{N}_0$, the modified spherical harmonics $\tilde{Y}_l^m$ can be expressed in its cosine ($m \leq 0$) and sine ($m > 0$) components as follows:

$$\tilde{Y}_l^m(\theta,\phi) := \begin{cases} \Theta_l^m(\cos\theta)\cdot\tilde{\Phi}^{1,m}(\phi), & \text{if } -l \leq m \leq 0, \\ \Theta_l^m(\cos\theta)\cdot\tilde{\Phi}^{2,m}(\phi), & \text{if } 0 < m \leq l, \end{cases}$$

where

$$\tilde{\Phi}^{1,m}(\phi) := \begin{cases} \frac{1}{\sqrt{\pi}}\cos(m\phi), & \text{if } -l \leq m < 0, \\ \frac{1}{\sqrt{2\pi}}, & \text{if } m = 0, \end{cases} \tag{2.3.1}$$

$$\tilde{\Phi}^{2,m}(\phi) := \frac{1}{\sqrt{\pi}}\sin(m\phi), \quad 0 < m \leq l, \tag{2.3.2}$$

and $\Theta_l^m(\cos\theta)$, $|m| \leq l$, $l \in 2\mathbb{N}_0$ are given by (1.3.10), $\theta \in [0,\pi]$, $\phi \in [0,2\pi)$. Due to the two cosine and sine components present in the modified spherical harmonics, in order to design normalized tight wavelet frames for $L^2_{\mathrm{sym}}(\mathbb{S}^2)$ using modified spherical harmonics, we consider $L^2_{\mathrm{sym}}(\mathbb{S}^2)$ as an orthogonal direct sum of two corresponding spanning subspaces (also Hilbert spaces) $\mathcal{H}^1$, $\mathcal{H}^2$, i.e.,

$$L^2_{\mathrm{sym}}(\mathbb{S}^2) = \mathcal{H}^1 \oplus^\perp \mathcal{H}^2, \tag{2.3.3}$$

where

$$\mathcal{H}^i := \mathrm{clos}_{L^2_{\mathrm{sym}}(\mathbb{S}^2)}\left(\mathrm{span}\left\{\tilde{Y}_l^m : (m,l) \in \Lambda^i\right\}\right), \quad i = 1,2,$$

$$\Lambda^1 := \{(m,l) \in \mathbb{Z}\times 2\mathbb{N}_0 : -l \leq m \leq 0\}, \quad \Lambda^2 := \{(m,l) \in \mathbb{Z}\times 2\mathbb{N}_0 : 1 \leq m \leq l\}.$$

Thus, to achieve our objective, it suffices to construct normalized tight wavelet frames for the orthogonal subspaces $\mathcal{H}^1$ and $\mathcal{H}^2$ respectively. To this end, consider the following MRAs of $\mathcal{H}^1$ and $\mathcal{H}^2$ formed respectively by the two sequences of subspaces $\left\{V_k^1\right\}_{k\geq 0}$ and $\left\{V_k^2\right\}_{k\geq 0}$ defined by

$$V_k^i := \operatorname{span}\left\{\tilde{Y}_l^m \,:\, (m,l)\in \mathcal{I}_k^i\right\}, \quad k\geq 0,\, i=1,2,$$

where for $i=1,2$, $k\geq 0$, the corresponding index sets $\mathcal{I}_k^i$ for V_k^i are given as

$$\mathcal{I}_k^i := \left\{(m,l)\in \Lambda^i \,:\, 0\leq l\leq L_k\right\}, \tag{2.3.4}$$

and $\{L_k\}_{k\geq 0}$ is an increasing sequence of non-negative integers increasing to infinity.

For $i=1,2$, $k\geq 0$, to construct scaling functions $\left\{\varphi_k^i(\cdot;\zeta_{k,\nu}^i) \,:\, \nu\in\mathcal{I}_k'\right\}$ to span the above subspaces V_k^i, we seek nodes $\zeta_{k,\nu}^i\in S$, $\nu\in\mathcal{I}_k'$, (with $\mathcal{I}_k'$ a finite set with $|\mathcal{I}_k^i|\leq|\mathcal{I}_k'|$) and $|\mathcal{I}_k'|\times|\mathcal{I}_k'|$ diagonal weight matrices $D_k := \operatorname{diag}(d_{k,\nu})_{\nu\in\mathcal{I}_k'}$ with positive diagonal entries such that the $|\mathcal{I}_k^i|\times|\mathcal{I}_k'|$ rectangular matrices $A_k^i := [Y_{\mathbf{j}}(\zeta_{k,\nu})]_{\mathbf{j}\in\mathcal{I}_k^i,\nu\in\mathcal{I}_k'}$ satisfy

$$A_k^i D_k A_k^{i*} = I,$$

which is equivalent to

$$\sum_{\nu\in\mathcal{I}_k'} \tilde{Y}_{\mathbf{j}'}(\zeta_{k,\nu})\overline{\tilde{Y}_{\mathbf{j}}(\zeta_{k,\nu})}d_{k,\nu} = \delta_{\mathbf{j},\mathbf{j}'}, \quad \mathbf{j},\mathbf{j}'\in\mathcal{I}_k^i. \tag{2.3.5}$$

To accomplish (2.3.5), for $i=1,2$, $k\geq 0$, an appropriate choice of nodes $\left\{\zeta_{k,\nu}^i\right\}_{\nu\in\mathcal{I}_k'}$ is given by a tensor product of azimuth and zenith points:

$$\zeta_{k,\nu}^i := (\theta_{k,\alpha},\phi_{k,\beta}^i), \quad \nu := (\alpha,\beta)\in\mathcal{I}_k',$$

where the azimuth points are given by

$$\theta_{k,\alpha} := \cos^{-1}(x_{k,\alpha}), \quad \alpha = 0,1,\ldots,L_k,$$

with $\{x_\alpha\}_{\alpha=0}^{L_k}$ being the zeros of the Legendre polynomial P_{L_k+1} and the zenith points are given as

$$\phi_{k,\beta}^1 = \frac{\pi(\beta+1/2)}{L_k+1}, \quad \phi_{k,\beta}^2 = \frac{\pi(\beta+1)}{L_k+2}, \quad \beta = 0,1,\ldots,L_k.$$

Thus the sets $\mathcal{I}_k'$ are given by

$$\mathcal{I}_k' := \{(\alpha,\beta) :\, \alpha,\beta = 0,1,\ldots,L_k\}.$$

If we enumerate the nodes in the following manner

$$\left\{(\theta_{k,\alpha},\phi_{k,\beta}^i)\right\}_{(\alpha,\beta)} := \left\{(\theta_{k,0},\phi_{k,0}^i),\ldots,(\theta_{k,0},\phi_{k,L_k}^i),\ldots\ldots,(\theta_{k,L_k},\phi_{k,0}^i),\ldots,(\theta_{k,L_k},\phi_{k,L_k}^i)\right\},$$

then the required diagonal weight matrices D_k are given by

$$D_k = \frac{2\pi}{L_k+1}\mathrm{diag}(w_0,\ldots,w_0,\, w_1,\ldots,w_1,\ldots\ldots,w_{L_k},\ldots,w_{L_k}),$$

where $\{w_\alpha\}_{\alpha=0}^{L_k}$ are the corresponding Gauss-quadrature weights of the Legendre polynomial P_{L_k+1}, with each unique value of w_α appearing L_k+1 times in the matrix D_k. For $i=1,2$, $k\geq 0$, the scaling functions $\varphi_k^i(\cdot;\zeta_{k,\nu}^i)$, $\nu\in\mathcal{I}_k'$, of the subspaces V_k^i are then defined as

$$\varphi_k^i(\cdot;\zeta_{k,\nu}^i) := \sqrt{d_{k,\nu}}\sum_{\mathbf{j}\in\mathcal{I}_k^i}\overline{\tilde{Y}_{\mathbf{j}}(\zeta_{k,\nu}^i)}\tilde{Y}_{\mathbf{j}},\quad \nu\in\mathcal{I}_k',$$

which yield a normalized tight frame for V_k^i by Proposition 2.1.2.

To see why the above choices of nodes and weight matrices are appropriate, we consider the separable structure of $\tilde{Y}_{\mathbf{j}}=\tilde{Y}_l^m$ as in (2.2.2), (2.3.1) and (2.3.2), obtaining

$$\begin{aligned}\sum_{\nu\in\mathcal{I}_k'}\tilde{Y}_{\mathbf{j}'}(\zeta_{k,\nu}^i)\overline{\tilde{Y}_{\mathbf{j}}(\zeta_{k,\nu}^i)}d_{k,\nu}^i &= \sum_{(\alpha,\beta)\in\mathcal{I}_k'}\tilde{Y}_{l'}^{m'}(\theta_{k,\alpha},\phi_{k,\beta}^i)\overline{\tilde{Y}_l^m(\theta_\alpha,\phi_\beta)}w_{k,\alpha}^i\cdot\frac{2\pi}{L_k+1}\\ &= \left\{\sum_{\alpha=0}^{L_k}\Theta_l^m(x_{k,\alpha})\Theta_{l'}^{m'}(x_{k,\alpha})w_{k,\alpha}\right\}\cdot\left\{\sum_{\beta=0}^{L_k}\tilde{\Phi}^{i,m'}(\phi_{k,\beta}^i)\overline{\tilde{\Phi}^{i,m}(\phi_{k,\beta}^i)}\cdot\frac{2\pi}{L_k+1}\right\}\\ &= \delta_{l,l'}\cdot\delta_{m,m'}=\delta_{\mathbf{j},\mathbf{j}'},\end{aligned}$$

for all $\mathbf{j},\mathbf{j}'\in\mathcal{I}_k^i$, $i=1,2$, where we have made use of the results of (2.2.12) and

$$\sum_{\beta=0}^{L_k}\tilde{\Phi}^{i,m'}(\phi_{k,\beta}^i)\overline{\tilde{\Phi}^{i,m}(\phi_{k,\beta}^i)}\cdot\frac{2\pi}{L_k+1}=\delta_{m,m'},\quad k\geq 0,\, i=1,2. \tag{2.3.6}$$

The orthogonality result in (2.3.6) holds because the matrices $[\tilde{\Phi}^{i,m}(\phi_{k,\beta})]_{m,\beta}$ corresponds to the discrete cosine transform-II (DCT-II) and discrete sine transform-I (DST-I) matrices respectively for $i=1,2$. As reference, the orthogonal $N\times N$ DCT-II matrix C_N^{II} is given by

$$\left[C_N^{\mathrm{II}}\right]_{n,p}=\sqrt{\frac{2}{N}}\varepsilon_p\cos\left(\frac{\pi(2n+1)p}{2N}\right),\quad n,p=0,1,\ldots,N-1,$$

where $\varepsilon_p=\frac{1}{\sqrt{2}}$, when $p=0$ and takes on the value 1 otherwise. The orthogonal $(N-1)\times(N-1)$ DST-I matrix S_{N-1}^{I} is given as

$$\left[S_{N-1}^{\mathrm{I}}\right]_{n,p}=\sqrt{\frac{2}{N}}\sin\left(\frac{\pi(n+1)(p+1)}{N}\right),\quad n,p=0,1,\ldots,N-2.$$

Now we construct the corresponding wavelets. For $i=1,2$, $k\geq 0$, since $W_k^i=V_{k+1}^i\backslash V_k^i$ and $\mathcal{J}_k^i=\mathcal{I}_{k+1}^i\backslash\mathcal{I}_k^i$, it follows that

$$W_k^i := \mathrm{span}\left\{Y_{\mathbf{j}}:\mathbf{j}\in\mathcal{J}_k^i\right\},$$

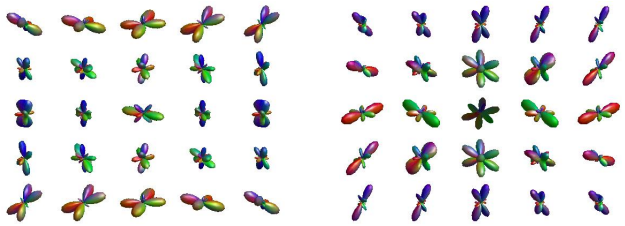

FIGURE 2.3.1. The left subplot displays the scaling functions $\varphi_0^1(\cdot;\zeta_{0,\nu}^1), \nu \in \mathcal{I}_0'$, for the subspace $V_0^1 := \text{span}\{\tilde{Y}_l^m \,:\, -l \le m \le 0,\, 0 \le l \le 4\}$. The right subplot displays the scaling functions $\varphi_0^2(\cdot;\zeta_{0,\nu}^2), \nu \in \mathcal{I}_0'$, for the subspace $V_0^2 := \text{span}\{\tilde{Y}_l^m \,:\, 1 \le m \le l,\, 0 \le l \le 4\}$.

where

$$\mathcal{J}_k^i := \left\{(m,l) \in \Lambda^i \,:\, L_k < l \le L_{k+1}\right\}. \tag{2.3.7}$$

To construct wavelets $\left\{\psi_k^i(\cdot;\eta_{k,\nu}^i) \,:\, \nu \in \mathcal{J}_k'\right\}$ to span the above subspaces W_k^i, we seek nodes $\eta_{k,\nu}^i \in S$, $\nu \in \mathcal{J}_k'$ (with $\mathcal{J}_k'$ a finite set with $|\mathcal{J}_k^i| \le |\mathcal{J}_k'|$), and a $|\mathcal{J}_k'| \times |\mathcal{J}_k'|$ diagonal weight matrix $\hat{D}_k := \text{diag}(\hat{d}_{k,\nu})_{\nu \in \mathcal{J}_k'}$ with positive diagonal entries such that the $|\mathcal{J}_k^i| \times |\mathcal{J}_k'|$ rectangular matrix $B_k^i := [Y_{\mathbf{j}}(\zeta_{k,\nu})]_{\mathbf{j} \in \mathcal{J}_k^i, \nu \in \mathcal{J}_k'}$ satisfies

$$B_k^i \hat{D}_k B_k^{i*} = I, \quad k \ge 0,\, i = 1,2.$$

This can be done by selecting $\eta_{k,\nu}^i := \zeta_{k+1,\nu}^i$, $i = 1,2$, $\nu \in \mathcal{J}_k' := \mathcal{I}_{k+1}'$, $\hat{D}_k := D_{k+1}$, $k \ge 0$.

Similarly, $\left\{\psi_k^i(\cdot;\eta_{k,\nu}^i) \,:\, \nu \in \mathcal{J}_k'\right\}$ forms a normalized tight frame for W_k^i, $k \ge 0$, $i = 1,2$, where

$$\psi_k^i(\cdot;\eta_{k,\nu}^i) = \sqrt{\hat{d}_{k,\nu}} \sum_{(m,l) \in \mathcal{J}_k^i} \overline{\tilde{Y}_l^m(\eta_{k,\nu}^i)} \tilde{Y}_l^m, \quad \nu \in \mathcal{J}_k'. \tag{2.3.8}$$

Therefore, by (2.3.3) and Theorem 2.1.6, $\left\{\psi_k^i(\cdot;\eta_{k,\nu}^i) \,:\, \nu \in \mathcal{J}_k',\, k \ge -1,\, i = 1,2\right\}$ forms a normalized tight wavelet frame for $L^2_{\text{sym}}(\mathbb{S}^2)$.

As an illustration, plots of some constructed scaling functions and wavelet functions are provided in Figure 2.3.1, Figure 2.3.2 and Figure 2.3.3. We shall see how this modified SH-based wavelet frame construction can be used to yield sparse representations of antipodally symmetric and real-valued HARDI orientation diffusion functions in Chapter 3 (see Figure 3.2.1 for examples of such functions).

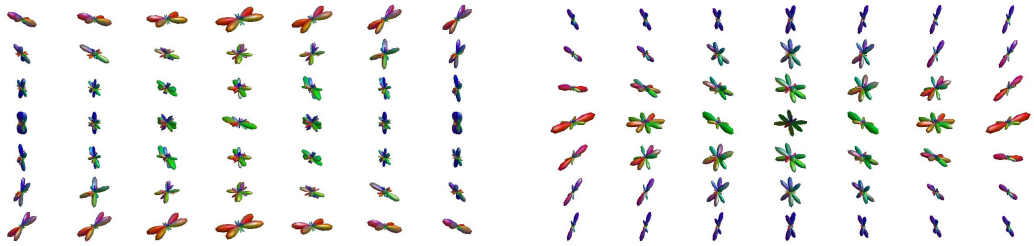

FIGURE 2.3.2. The left subplot displays the scaling functions $\varphi_1^1(\cdot;\zeta_{1,\nu}^1), \nu \in \mathcal{I}_1'$, for the subspace $V_1^1 := \text{span}\{\tilde{Y}_l^m \,:\, -l \le m \le 0,\, 0 \le l \le 6\}$. The right subplot displays the scaling functions $\varphi_1^2(\cdot;\zeta_{1,\nu}^2), \nu \in \mathcal{I}_1'$, for the subspace $V_1^2 := \text{span}\{\tilde{Y}_l^m \,:\, 1 \le m \le l,\, 0 \le l \le 6\}$.

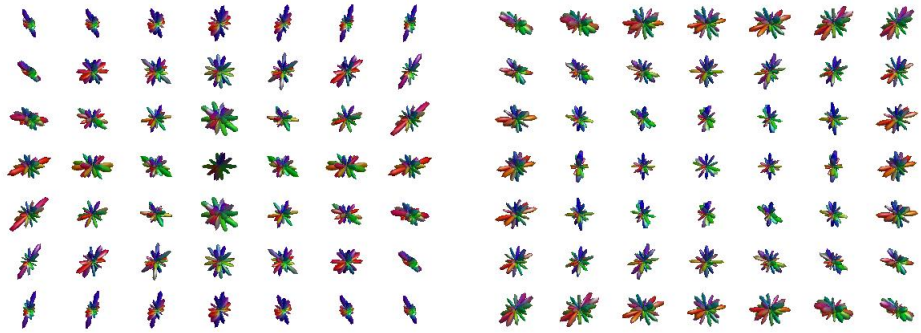

FIGURE 2.3.3. The left subplot displays the wavelet functions $\psi_0^1(\cdot;\eta_{0,\nu}^1), \nu \in \mathcal{J}_0'$, for the subspace $W_0^1 := \text{span}\{\tilde{Y}_l^m \,:\, -l \le m \le 0,\, l = 6\}$. The right subplot displays the wavelet functions $\psi_0^2(\cdot;\eta_{0,\nu}^2), \nu \in \mathcal{J}_0'$, for the subspace $W_0^2 := \text{span}\{\tilde{Y}_l^m \,:\, 1 \le m \le l,\, l = 6\}$.

2.4. Constructing Dual Pairs of Wavelet Frames for $(H^s(\mathbb{S}^2), H^{-s}(\mathbb{S}^2))$

Recall from our earlier work that $\{\psi_k(\cdot;\eta_{k,\nu}) \,:\, \nu \in \mathcal{J}_k',\, k \ge -1\}$ forms a tight wavelet frame for $L^2(\mathbb{S}^2)$, where

$$\psi_k(\cdot;\eta_{k,\nu}) := \sqrt{\hat{d}_{k,\nu}} \sum_{(l,m)\in\mathcal{J}_k} \overline{Y_l^m(\eta_{k,\nu})} Y_l^m, \quad \nu \in \mathcal{J}_k', \tag{2.4.1}$$

$\{\eta_{k,\nu}\}_{\nu\in\mathcal{J}_k'}$ are the Gauss-Legendre quadrature nodes, $\left\{\hat{d}_{k,\nu}\right\}_{\nu\in\mathcal{J}_k'}$ are obtained from the corresponding Gauss-Legendre quadrature weights, and the index set $\mathcal{J}_k$ was given earlier as

$$\mathcal{J}_k = \{(m,l) \in \Lambda \,:\, L_k < l \le L_{k+1}\}, \tag{2.4.2}$$

with $\{L_k\}_{k\geq -1}$ an increasing sequence of positive integers and

$$\Lambda = \{(m,l) \in \mathbb{Z} \times \mathbb{N}_0 \,:\, |m| \leq l\},$$

the index set of the spherical harmonics. Note that $|\mathcal{J}_k| \leq |\mathcal{J}_k'|$, $k \geq -1$. From our earlier work, the disjoint union of $\mathcal{J}_k$'s gives $\bigcup_{k=-1}^{K} \mathcal{J}_k = \mathcal{I}_{K+1}$, where

$$\mathcal{I}_k = \{(m,l) \in \Lambda \,:\, 0 \leq l \leq L_k\}$$

The union of the nested index sets $\mathcal{I}_k$ gives $\bigcup\limits_{k=-1}^{\infty} \mathcal{I}_k = \Lambda$.

Renumerating the spherical harmonics with $\mathbf{j} := (l,m)$, $Y_{\mathbf{j}} := Y_l^m$, $(m,l) \in \Lambda$, (2.4.1) gives

$$\psi_k(\cdot;\eta_{k,\nu}) := \sqrt{\hat{d}_{k,\nu}} \sum_{\mathbf{j}\in\mathcal{J}_k} \overline{Y_{\mathbf{j}}(\eta_{k,\nu})} Y_{\mathbf{j}}, \quad \nu \in \mathcal{J}_k', \tag{2.4.3}$$

which we will from time to time take reference to simplify notations. Recall that by setting $\hat{D}_k = \operatorname{diag}(\hat{d}_{k,\nu})_\nu$, the matrix $B_k := [Y_{\mathbf{j}}(\eta_{k,\nu})]_{\mathbf{j}\in\mathcal{J}_k, \nu\in\mathcal{J}_k'}$ satisfies

$$B_k \hat{D}_k B_k^* = I, \quad k \geq -1.$$

Note that this is equivalent to

$$\sum_{\nu\in\mathcal{J}_k'} \overline{Y_{\mathbf{j}}^m(\eta_{k,\nu})} Y_{\mathbf{j}'}(\eta_{k,\nu}) \hat{d}_{k,\nu} = \delta_{\mathbf{j},\mathbf{j}'}, \quad \mathbf{j},\mathbf{j}' \in \mathcal{J}_k,\, k \geq -1. \tag{2.4.4}$$

Let $s \in \mathbb{R}$ and $\{p^s[m,l]\}_{(m,l)\in\Lambda}$ be a sequence of non-zero real numbers such that there exist positive constants C_1 and C_2 such that

$$C_1(1+l^2)^{-s/2} \leq |p^s[m,l]| \leq C_2(1+l^2)^{-s/2},\, \forall (m,l) \in \Lambda, \tag{2.4.5}$$

which means that asymptotically, its decay/growth rate is only dependent on the order l and exponent s. For subsequent ease of notation, we set

$$p_{l_{\mathbf{j}}}^s := p^s[m,l], \quad \mathbf{j} = (l,m) \in \Lambda.$$

Define

$$\psi_{k,s}^\sharp(\cdot;\eta_{k,\nu}) := \sqrt{\hat{d}_{k,\nu}} \sum_{\mathbf{j}\in\mathcal{J}_k} p_{l_{\mathbf{j}}}^s \overline{Y_{\mathbf{j}}(\eta_{k,\nu})} Y_{\mathbf{j}}, \quad \nu \in \mathcal{J}_k', \tag{2.4.6}$$

$$\tilde{\psi}_{k,-s}(\cdot;\eta_{k,\nu}) := \sqrt{\hat{d}_{k,\nu}} \sum_{\mathbf{j}\in\mathcal{J}_k} (p_{l_{\mathbf{j}}}^s)^{-1} \overline{Y_{\mathbf{j}}(\eta_{k,\nu})} Y_{\mathbf{j}}, \quad \nu \in \mathcal{J}_k'. \tag{2.4.7}$$

REMARK 2.4.1. The motivation behind defining the two wavelet systems in (2.4.6) and (2.4.7) is that in section 3.3, the tight frame wavelet system in (2.4.3) actually undergoes a linear transformation under some linear operator T to yield either the system in (2.4.6) or (2.4.7) to represent a signal more effectively. We shall see that the eigenfunctions of T are either the SH or the modified SH, i.e.,

$$T(Y_{\mathbf{j}}) = p_{l_{\mathbf{j}}} Y_{\mathbf{j}} \quad \text{or} \quad T(\tilde{Y}_{\mathbf{j}}) = p_{l_{\mathbf{j}}} \tilde{Y}_{\mathbf{j}},$$

where the sequence $\{p_{l_{\mathbf{j}}}\}_{\mathbf{j}}$ satisfies (2.4.5) for some $s \in \mathbb{R}$. Therefore, it is natural to investigate whether the pair of wavelet systems (2.4.6) and (2.4.7) actually form a dual pair of frames for a dual pair of Sobolev spaces for some order $s \in \mathbb{R}$. This basic idea also motivates the investigation in the next section that under the influence of the operator T, the resultant pair of wavelet systems form a pair of dual frames for a dual pair of symmetric Sobolev spaces for some order $s \in \mathbb{R}$.

Using the definitions (2.4.1) and (2.4.5), we show below that for $k \geq -1$, $\nu \in \mathcal{J}'_k$, $\psi^\sharp_{k,s}(\cdot;\eta_{k,\nu}) \in H^s(\mathbb{S}^2)$: for some positive constant C,

$$\begin{aligned}
\|\psi^\sharp_{k,s}(\cdot;\eta_{k,\nu})\|_s^2 &= \hat{d}_{k,\nu} \sum_{(m,l)\in\mathcal{J}_k} (1+l^2)^s |p^s[m,l]|^2 |Y_l^m(\eta_{k,\nu})|^2 \\
&\leq \hat{d}_{k,\nu} \sum_{(m,l)\in\mathcal{J}_k} C|Y_l^m(\eta_{k,\nu})|^2 \\
&= C\|\psi_k(\cdot;\eta_{k,\nu})\|^2 < \infty.
\end{aligned}$$

Similarly, for $k \geq -1$, $\nu \in \mathcal{J}'_k$, $\tilde{\psi}_{k,-s}(\cdot;\eta_{k,\nu}) \in H^{-s}(\mathbb{S}^2)$. Note that $\psi^\sharp_{k,s}(\cdot;\eta_{k,\nu})$, $\tilde{\psi}_{k,-s}(\cdot;\eta_{k,\nu})$ lie respectively in Sobolev spaces that are dual to each other. Indeed, our aim here is to show that the two systems $(\Gamma^s(\Psi^\sharp), \Gamma^{-s}(\tilde{\Psi}))$, defined by

$$\Gamma^s(\Psi^\sharp) := \left\{\psi^\sharp_{k,s}(\cdot;\eta_{k,\nu}) \,:\, \nu \in \mathcal{J}'_k,\, k \geq -1\right\}, \tag{2.4.8}$$

$$\Gamma^{-s}(\tilde{\Psi}) := \left\{\tilde{\psi}_{k,-s}(\cdot;\eta_{k,\nu}) \,:\, \nu \in \mathcal{J}'_k,\, k \geq -1\right\}, \tag{2.4.9}$$

form a pair of dual wavelet frames for $H^s(\mathbb{S}^2)$ and $H^{-s}(\mathbb{S}^2)$ respectively in the following sense:

(1) The identity

$$\langle f, g\rangle = \sum_{k\geq -1,\, \nu\in\mathcal{J}'_k} \langle g, \psi^\sharp_{k,s}(\cdot;\eta_{k,\nu})\rangle \langle \tilde{\psi}_{k,-s}(\cdot;\eta_{k,\nu}), f\rangle \tag{2.4.10}$$

holds for all $f \in H^s(\mathbb{S}^2)$, $g \in H^{-s}(\mathbb{S}^2)$, where $\langle\cdot,\cdot\rangle$ is the $L^2(\mathbb{S}^2)$ inner product.

(2) The two systems $\Gamma^s(\Psi^\sharp)$ and $\Gamma^{-s}(\tilde{\Psi})$ are wavelet frames for $H^s(\mathbb{S}^2)$ and $H^{-s}(\mathbb{S}^2)$ respectively, i.e., there exist positive constants C_1 and C_2 such that

$$C_1\|f'\|_s^2 \le \sum_{k\ge -1,\,\nu\in\mathcal{J}_k'} |\langle f',\, \psi_{k,s}^\sharp(\cdot;\eta_{k,\nu})\rangle_s|^2 \le C_2\|f'\|_s^2, \tag{2.4.11}$$

$$C_1\|g'\|_{-s}^2 \le \sum_{k\ge -1,\,\nu\in\mathcal{J}_k'} |\langle g',\, \tilde{\psi}_{k,-s}(\cdot;\eta_{k,\nu})\rangle_{-s}|^2 \le C_2\|g'\|_{-s}^2, \tag{2.4.12}$$

hold for all $f' \in H^s(\mathbb{S}^2)$, $g' \in H^{-s}(\mathbb{S}^2)$, where $\langle\cdot,\cdot\rangle_s$ is the inner product that the Sobolev space $H^s(\mathbb{S}^2)$, $s \in \mathbb{R}$, is equipped with.

To ease our analysis, we show that (2.4.11) and (2.4.12) are equivalent to the following.

PROPOSITION 2.4.2. *Let $s \in \mathbb{R}$, $\Gamma^s(\Psi^\sharp)$ and $\Gamma^{-s}(\tilde{\Psi})$ defined in* (2.4.8) *and* (2.4.9) *are wavelet frames for $H^s(\mathbb{S}^2)$ and $H^{-s}(\mathbb{S}^2)$ respectively if and only if there exist positive constants C_1 and C_2 such that*

$$C_1\|g\|_{-s}^2 \le \sum_{k\ge -1,\,\nu\in\mathcal{J}_k'} |\langle g,\, \psi_{k,s}^\sharp(\cdot;\eta_{k,\nu})\rangle|^2 \le C_2\|g\|_{-s}^2, \tag{2.4.13}$$

$$C_1\|f\|_s^2 \le \sum_{k\ge -1,\,\nu\in\mathcal{J}_k'} |\langle f,\, \tilde{\psi}_{k,-s}(\cdot;\eta_{k,\nu})\rangle|^2 \le C_2\|f\|_s^2, \tag{2.4.14}$$

hold for all $f \in H^s(\mathbb{S}^2)$, $g \in H^{-s}(\mathbb{S}^2)$.

PROOF. For $s \in \mathbb{R}$, define an operator $\theta_s : H^s(\mathbb{S}^2) \mapsto H^{-s}(\mathbb{S}^2)$ acting on $h \in H^s(\mathbb{S}^2)$ via spherical harmonic coefficients, i.e.,

$$\widehat{\theta_s h}[l,m] := (1+l^2)^s \hat{h}[m,l], \quad (m,l) \in \Lambda.$$

Then it is easy to see that $\|\theta_s h\|_{-s}^2 = \|h\|_s^2$ and thus θ_s is an isometric and onto mapping between $H^s(\mathbb{S}^2)$ and $H^{-s}(\mathbb{S}^2)$. On the other hand, for f', $v_1 \in H^s(\mathbb{S}^2)$, g', $v_2 \in H^{-s}(\mathbb{S}^2)$,

$$\langle f', v_1\rangle_s = \langle \theta_s f', v_1\rangle, \quad \langle g', v_2\rangle_{-s} = \langle \theta_{-s} g', v_2\rangle.$$

This implies that (2.4.11) and (2.4.12) are equivalent to (2.4.13) and (2.4.14) by setting $g = \theta_s f'$ and $f = \theta_{-s} g'$. □

We are now ready to state the main result of this section.

THEOREM 2.4.3. *Let $s \in \mathbb{R}$. Then the systems $\Gamma^s(\Psi^\sharp)$ and $\Gamma^{-s}(\tilde{\Psi})$ defined by (2.4.8) and (2.4.9) form a pair of dual wavelet frames for $H^s(\mathbb{S}^2)$ and $H^{-s}(\mathbb{S}^2)$ respectively.*

PROOF. To establish the result, we shall first prove the identity in (2.4.10). Consider for a fixed $K \in \mathbb{N}$, $f \in H^s(\mathbb{S}^2)$, $g \in H^{-s}(\mathbb{S}^2)$, noting that $\mathbf{j} = (m, l)$, $\mathbf{j}' = (m', l')$, $\hat{f}[\mathbf{j}] := \hat{f}[m, l]$, $\hat{g}[\mathbf{j}] := \hat{g}[m, l]$, $(m, l) \in \Lambda$, we have

$$\begin{aligned}
S_K :&= \sum_{k=-1}^{K} \sum_{\nu \in \mathcal{J}_k'} \langle f, \psi_{k,s}^\sharp(\cdot; \eta_{k,\nu})\rangle \langle \tilde{\psi}_{k,-s}(\cdot; \eta_{k,\nu}), g\rangle \\
&= \sum_{k=-1}^{K} \sum_{\nu \in \mathcal{J}_k'} \left\{ \sqrt{\hat{d}_{k,\nu}} \sum_{\mathbf{j} \in \mathcal{J}_k} \hat{f}[\mathbf{j}] Y_{\mathbf{j}}(\eta_{k,\nu}) \cdot (p_{l_{\mathbf{j}}}^s) \right\} \left\{ \sqrt{\hat{d}_{k,\nu}} \sum_{\mathbf{j}' \in \mathcal{J}_k} \overline{\hat{g}[\mathbf{j}'] Y_{\mathbf{j}'}(\eta_{k,\nu}) \cdot \left(p_{l_{\mathbf{j}'}}^s\right)^{-1}} \right\} \\
&= \sum_{k=-1}^{K} \sum_{\mathbf{j}, \mathbf{j}' \in \mathcal{J}_k} \hat{f}[\mathbf{j}] \overline{\hat{g}[\mathbf{j}']} (p_{l_{\mathbf{j}'}}^s)^{-1} p_{l_{\mathbf{j}}}^s \left\{ \sum_{\nu \in \mathcal{J}_k'} Y_{\mathbf{j}}(\eta_{k,\nu}) \overline{Y_{\mathbf{j}'}(\eta_{k,\nu})} \hat{d}_{k,\nu} \right\} \\
&= \sum_{k=-1}^{K} \sum_{\mathbf{j} \in \mathcal{J}_k} \hat{f}[\mathbf{j}] \overline{\hat{g}[\mathbf{j}]} = \sum_{\mathbf{j} \in \mathcal{I}_{K+1}} \hat{f}[\mathbf{j}] \overline{\hat{g}[\mathbf{j}]},
\end{aligned}$$

where we have invoked (2.4.4) in the third line, and that the disjoint union of $\mathcal{J}_k$ gives $\bigcup_{k=-1}^{K} \mathcal{J}_k = \mathcal{I}_{K+1}$ at the fourth line. Recall that $\bigcup_{k=-1}^{\infty} \mathcal{I}_k = \Lambda$ and thus we have for all $K \in \mathbb{N}$, $|S_K| \leq \|f\|_s \|g\|_{-s} < \infty$, for all $f \in H^s(\mathbb{S}^2)$, $g \in H^{-s}(\mathbb{S}^2)$. The desired result (2.4.10) is then obtained by letting K go to infinity in the above.

To complete our proof, we establish (2.4.13) and (2.4.14) by showing that the wavelet systems have the following Bessel properties:

$$\sum_{k \geq -1, \nu \in \mathcal{J}_k'} |\langle g, \psi_{k,s}^\sharp(\cdot; \eta_{k,\nu})\rangle|^2 \leq C \|g\|_{-s}^2, \tag{2.4.15}$$

$$\sum_{k \geq -1, \nu \in \mathcal{J}_k'} |\langle f, \tilde{\psi}_{k,-s}(\cdot; \eta_{k,\nu})\rangle|^2 \leq C \|f\|_s^2, \tag{2.4.16}$$

for all $f \in H^s(\mathbb{S}^2)$, $g \in H^{-s}(\mathbb{S}^2)$. We shall prove only (2.4.15), for the proof of (2.4.16) is similar. For a fixed $K \in \mathbb{N}$, $g \in H^{-s}(\mathbb{S}^2)$,

$$\begin{aligned}
&\sum_{k=-1}^{K} \sum_{\nu \in \mathcal{J}_k'} |\langle g, \psi_{k,s}^\sharp(\cdot;\eta_{k,\nu})\rangle|^2 \\
&= \sum_{k=-1}^{K} \sum_{\nu \in \mathcal{J}_k'} \left\{ \sqrt{\hat{d}_{k,\nu}} \sum_{\mathbf{j} \in \mathcal{J}_k} \hat{g}[\mathbf{j}] Y_{\mathbf{j}}(\eta_{k,\nu}) \cdot p_{l_{\mathbf{j}}}^s \right\} \left\{ \sqrt{\hat{d}_{k,\nu}} \sum_{\mathbf{j}' \in \mathcal{J}_k} \overline{\hat{g}[\mathbf{j}'] Y_{\mathbf{j}'}(\eta_{k,\nu}) \cdot p_{l_{\mathbf{j}'}}^s} \right\} \\
&= \sum_{k=-1}^{K} \sum_{\mathbf{j},\mathbf{j}' \in \mathcal{J}_k} \hat{g}[\mathbf{j}] \overline{\hat{g}[\mathbf{j}']} p_{l_{\mathbf{j}}}^s p_{l_{\mathbf{j}'}}^s \left\{ \sum_{\nu \in \mathcal{J}_k'} Y_{\mathbf{j}}(\eta_{k,\nu}) \overline{Y_{\mathbf{j}'}(\eta_{k,\nu})} \hat{d}_{k,\nu} \right\} \\
&= \sum_{k=-1}^{K} \sum_{\mathbf{j} \in \mathcal{J}_k} |\hat{g}[\mathbf{j}]|^2 \cdot |p_{l_{\mathbf{j}}}^s|^2 \\
&\le C \sum_{(m,l) \in \mathcal{I}_{K+1}} |\hat{g}[m,l]|^2 (1+l^2)^{-s} \le C \|g\|_{-s}^2,
\end{aligned}$$

where we have invoked (2.4.4) and (2.4.5) respectively at the third and fourth lines. The desired result (2.4.15) is then obtained by letting K go to infinity.

It remains to establish the existence of lower frame bounds in (2.4.13) and (2.4.14). Starting from a standard trick in wavelet frame analysis, we apply the Cauchy-Schwartz's inequality on the identity (2.4.10) and then invoke the Bessel property as shown below:

$$\begin{aligned}
\|g\|_{-s} &= \sup_{\|f\|_s=1} |\langle f, g\rangle| \\
&\le \sup_{\|f\|_s=1} \left\{ \sum_{k \ge -1,\ \nu \in \mathcal{J}_k'} |\langle f,\ \tilde{\psi}_{k,-s}(\cdot;\eta_{k,\nu})\rangle|^2 \right\}^{1/2} \left\{ \sum_{k \ge -1,\ \nu \in \mathcal{J}_k'} |\langle g,\ \psi_{k,s}^\sharp(\cdot;\eta_{k,\nu})\rangle|^2 \right\}^{1/2} \\
&\le \sup_{\|f\|_s=1} C\|f\|_s \left\{ \sum_{k \ge -1,\ \nu \in \mathcal{J}_k'} |\langle g,\ \psi_{k,s}^\sharp(\cdot;\eta_{k,\nu})\rangle|^2 \right\}^{1/2} \\
&= C \left\{ \sum_{k \ge -1,\ \nu \in \mathcal{J}_k'} |\langle g,\ \psi_{k,s}^\sharp(\cdot;\eta_{k,\nu})\rangle|^2 \right\}^{1/2},
\end{aligned}$$

giving the required lower frame bound in (2.4.13). A similar argument also yields the lower frame bound in (2.4.14), completing the proof that $\Gamma^s(\Psi^\sharp)$ and $\Gamma^{-s}(\tilde{\Psi})$ defined in (2.4.8) and (2.4.9) form a pair of dual wavelet frames for $H^s(\mathbb{S}^2)$ and $H^{-s}(\mathbb{S}^2)$ respectively. □

2.5. Constructing Dual Pairs of Wavelet Frames for $(H^s_{\mathbf{sym}}(\mathbb{S}^2),\, H^{-s}_{\mathbf{sym}}(\mathbb{S}^2))$

In this section, we construct instead dual pairs of wavelet frames for $(H^s_{\mathrm{sym}}(\mathbb{S}^2),\, H^{-s}_{\mathrm{sym}}(\mathbb{S}^2))$ in an analogous fashion to section 2.4. Many of the proofs for this section follow very similarly to those in section 2.4, so we shall only outline its key ideas and results here.

Recall from section 2.3 that $\left\{\psi^i_k(\cdot;\eta^i_{k,\nu}) \,:\, \nu\in\mathcal{J}'_k,\, k\geq -1,\, i=1,2\right\}$ forms a tight wavelet frame for $L^2_{\mathrm{sym}}(\mathbb{S}^2)$, where for $i=1,2$, $k\geq -1$,

$$\psi^i_k(\cdot;\eta^i_{k,\nu}) := \sqrt{\hat{d}_{k,\nu}} \sum_{(m,l)\in\mathcal{J}^i_k} \overline{\tilde{Y}^m_l(\eta^i_{k,\nu})}\tilde{Y}^m_l, \quad \nu\in\mathcal{J}'_k, \tag{2.5.1}$$

where $\left\{\eta^i_{k,\nu}\right\}_{\nu\in\mathcal{J}'_k}$ are the Gauss-Legendre quadrature nodes for the modified spherical harmonics, $\left\{\hat{d}_{k,\nu}\right\}_{\nu\in\mathcal{J}'_k}$ are obtained from the corresponding Gauss-Legendre quadrature weights, and the index sets $\mathcal{J}^i_k$, $i=1,2$, are given earlier in (2.3.7). From our earlier work, the disjoint union of $\mathcal{J}^i_k$'s gives $\bigcup_{k=-1}^{K}\mathcal{J}^i_k = \mathcal{I}^i_{K+1}$, where the index sets $\mathcal{I}^i_k$, $i=1,2$, are given in (2.3.4). The union of the nested index sets $\mathcal{I}^i_k$ produces $\bigcup_{i=1,2}\bigcup_{k\geq 0}\mathcal{I}^i_k = \Lambda$, the index set of the modified spherical harmonics. Renumerating the modified spherical harmonics with $\mathbf{j}=(l,m)$, $\tilde{Y}_{\mathbf{j}} := \tilde{Y}^m_l$, $(m,l)\in\Lambda$, (2.5.1) gives

$$\psi^i_k(\cdot;\eta^i_{k,\nu}) := \sqrt{\hat{d}_{k,\nu}} \sum_{\mathbf{j}\in\mathcal{J}^i_k} \overline{\tilde{Y}_{\mathbf{j}}(\eta^i_{k,\nu})}\tilde{Y}_{\mathbf{j}}, \quad \nu\in\mathcal{J}'_k.$$

Recall that by setting $\hat{D}_k = \mathrm{diag}(\hat{d}_{k,\nu})_{\nu\in\mathcal{J}'_k}$, the matrix $B^i_k := [\tilde{Y}_{\mathbf{j}}(\eta^i_{k,\nu})]_{\mathbf{j}\in\mathcal{J}^i_k,\nu\in\mathcal{J}'_k}$ satisfies

$$B^i_k\hat{D}_kB^{i*}_k = I, \quad k\geq -1,\, i=1,2.$$

Note that the above is equivalent to

$$\sum_{\nu\in\mathcal{J}'_k} \overline{\tilde{Y}^m_{\mathbf{j}}(\eta^i_{k,\nu})}Y_{\mathbf{j}'}(\eta^i_{k,\nu})\hat{d}_{k,\nu} = \delta_{\mathbf{j},\mathbf{j}'}, \quad \mathbf{j},\mathbf{j}'\in\mathcal{J}^i_k,\, k\geq -1,\, i=1,2.$$

Let $s\in\mathbb{R}$, $\{p^s[m,l]\}_{(m,l)\in\Lambda}$ be the same sequence of non-zero real numbers that satisfy the decay rate in (2.4.5). For subsequent ease of notation in this section, we define

$$p^s_{l_{\mathbf{j}}} := p^s[l,m], \quad \mathbf{j}:=(m,l)\in\Lambda, \quad \langle\cdot,\cdot\rangle_\star := \langle\cdot,\cdot\rangle_{\star,0}, \quad \|\cdot\|_\star := \|\cdot\|_{\star,0}.$$

Define

$$\psi^{i,\sharp}_{k,s}(\cdot;\eta^i_{k,\nu}) := \sqrt{\hat{d}_{k,\nu}} \sum_{\mathbf{j}\in\mathcal{J}^i_k} p^s_{l_{\mathbf{j}}} \overline{\tilde{Y}_{\mathbf{j}}(\eta^i_{k,\nu})}\tilde{Y}_{\mathbf{j}}, \quad \nu\in\mathcal{J}'_k,$$

$$\tilde{\psi}^i_{k,-s}(\cdot;\eta^i_{k,\nu}) := \sqrt{\hat{d}_{k,\nu}} \sum_{\mathbf{j}\in\mathcal{J}^i_k} (p^s_{l_{\mathbf{j}}})^{-1} \overline{\tilde{Y}_{\mathbf{j}}(\eta^i_{k,\nu})}Y_{\mathbf{j}}, \quad \nu\in\mathcal{J}'_k,$$

Using similar arguments as in section 2.4, we can show that for $k\geq -1$, $\nu\in\mathcal{J}'_k$, $\psi^{i,\sharp}_{k,s}(\cdot;\eta^i_{k,\nu})\in H^s_{\mathrm{sym}}(\mathbb{S}^2)$. Similarly, for $k\geq -1$, $\nu\in\mathcal{J}'_k$, $\tilde{\psi}^i_{k,-s}(\cdot;\eta^i_{k,\nu})\in H^{-s}_{\mathrm{sym}}(\mathbb{S}^2)$, which is dual to $H^{-s}_{\mathrm{sym}}(\mathbb{S}^2)$. Our aim here is to show that the two systems $(\Gamma^s(\Psi^\sharp),\Gamma^{-s}(\tilde{\Psi}))$, defined by

$$\Gamma^s(\Psi^\sharp) := \left\{\psi^{i,\sharp}_{k,s}(\cdot;\eta^i_{k,\nu}) \,:\, \nu\in\mathcal{J}'_k,\, k\geq -1,\, i=1,2\right\}, \tag{2.5.2}$$

$$\Gamma^{-s}(\tilde{\Psi}) := \left\{\tilde{\psi}^i_{k,-s}(\cdot;\eta^i_{k,\nu}) \,:\, \nu\in\mathcal{J}'_k,\, k\geq -1,\, i=1,2\right\}, \tag{2.5.3}$$

form a pair of dual wavelet frames for $H^s_{\mathrm{sym}}(\mathbb{S}^2)$ and $H^{-s}_{\mathrm{sym}}(\mathbb{S}^2)$ respectively in the following sense:

- The identity

$$\langle f,\, g\rangle = \sum_{i=1,2,\, k\geq -1,\, \nu\in\mathcal{J}'_k} \langle g,\, \psi^{i,\sharp}_{k,s}(\cdot;\eta^i_{k,\nu})\rangle_\star \langle\tilde{\psi}^i_{k,-s}(\cdot;\eta^i_{k,\nu}),\, f\rangle_\star$$

  holds for all $f\in H^s_{\mathrm{sym}}(\mathbb{S}^2)$, $g\in H^{-s}_{\mathrm{sym}}(\mathbb{S}^2)$.

- The two systems $\Gamma^s(\Psi^\sharp)$ and $\Gamma^{-s}(\tilde{\Psi})$ are wavelet frames for $H^s_{\mathrm{sym}}(\mathbb{S}^2)$ and $H^{-s}_{\mathrm{sym}}(\mathbb{S}^2)$ respectively, i.e., there exist positive constants C_1 and C_2 such that

$$C_1\|f'\|^2_{\star,s} \leq \sum_{i=1,2,\, k\geq -1,\, \nu\in\mathcal{J}'_k} |\langle f',\, \psi^{i,\sharp}_{k,s}(\cdot;\eta^i_{k,\nu})\rangle_{\star,s}|^2 \leq C_2\|f'\|^2_{\star,s}, \tag{2.5.4}$$

$$C_1\|g'\|^2_{\star,-s} \leq \sum_{i=1,2,\, k\geq -1,\, \nu\in\mathcal{J}'_k} |\langle g',\, \tilde{\psi}^i_{k,-s}(\cdot;\eta^i_{k,\nu})\rangle_{\star,-s}|^2 \leq C_2\|g'\|^2_{\star,-s}, \tag{2.5.5}$$

  hold for all $f'\in H^s_{\mathrm{sym}}(\mathbb{S}^2)$, $g'\in H^{-s}_{\mathrm{sym}}(\mathbb{S}^2)$.

Similar to section 2.4, (2.5.4) and (2.5.5) can be shown to be equivalent to the following.

PROPOSITION 2.5.1. *Let $s \in \mathbb{R}$, $\Gamma^s(\Psi^\sharp)$ and $\Gamma^{-s}(\tilde{\Psi})$ are wavelet frames for $H^s_{\mathrm{sym}}(\mathbb{S}^2)$ and $H^{-s}_{\mathrm{sym}}(\mathbb{S}^2)$ respectively if and only if there exist positive constants C_1 and C_2 such that*

$$C_1\|g\|^2_{\star,-s} \leq \sum_{i=1,2,\, k\geq -1,\, \nu\in\mathcal{J}'_k} |\langle g,\, \psi^{i,\sharp}_{k,s}(\cdot;\eta^i_{k,\nu})\rangle_\star|^2 \leq C_2\|g\|^2_{\star,-s},$$

$$C_1\|f\|^2_{\star,s} \leq \sum_{i=1,2,\, k\geq -1,\, \nu\in\mathcal{J}'_k} |\langle f,\, \tilde{\psi}^{i}_{k,-s}(\cdot;\eta^i_{k,\nu})\rangle_\star|^2 \leq C_2\|f\|^2_{\star,s},$$

hold for all $f \in H^s_{\mathrm{sym}}(\mathbb{S}^2)$, $g \in H^{-s}_{\mathrm{sym}}(\mathbb{S}^2)$.

PROOF. The proof is analogous to that of Proposition 2.4.2 by replacing $H^s(\mathbb{S}^2)$, $\|\cdot\|_s$ and $\langle\cdot,\cdot\rangle_s$ with $H^s_{\mathrm{sym}}(\mathbb{S}^2)$, $\|\cdot\|_{\star,s}$ and $\langle\cdot,\cdot\rangle_{\star,s}$ respectively. □

THEOREM 2.5.2. *Let $s \in \mathbb{R}$. Then the systems $\Gamma^s(\Psi^\sharp)$ and $\Gamma^{-s}(\tilde{\Psi})$ defined by (2.5.2) and (2.5.3) form a pair of dual wavelet frames for $H^s_{\mathrm{sym}}(\mathbb{S}^2)$ and $H^{-s}_{\mathrm{sym}}(\mathbb{S}^2)$ respectively.*

PROOF. The proof is analogous to that of Theorem 2.4.3 with substitutions of appropriate summations and notations as mentioned in the proof of Proposition 2.5.1. □

CHAPTER 3

Application of Wavelet Frames to High Angular Resolution Diffusion Imaging (HARDI)

In this chapter, we shall describe how the wavelet frames constructed in subsection 2.5 can be applied to High Angular Resolution Diffusion Imaging (HARDI), a relatively recent brain imaging technique. We will demonstrate that, when applied to denoise highly corrupted HARDI signals, the proposed wavelet frame-based approach has a more cost-effective performance over two other closely related approaches, namely the spherical harmonics-based and spherical ridgelets-based approaches. It also turns out that our proposed method yields wavelet frame systems in some appropriately chosen Sobolev spaces described in sections 2.4 and 2.5.

This chapter is organised as follows. Section 3.1 briefly reviews some background of HARDI and a classical spherical harmonics-based approach to perform Q-ball imaging (QBI), a special technique for HARDI. Section 3.2 focuses on an alternative spherical ridgelets-based approach to perform QBI. In section 3.3, the proposed wavelet frame-based approach for QBI is described in detail. Note that in each of the above sections, each approach has two variants (described in different subsections), which differ in whether a constant solid angle (CSA) reconstruction model is assumed.

In section 3.4, we describe how spatial regularization of HARDI signals can be imposed on the above approaches by tight framelet filters based on B-splines (e.g., the Haar framelet filters defined in (3.4.4)) and conclude from numerical experiments that with this spatial regularization, the proposed wavelet frame-based approach generally denoises highly corrupted HARDI signals more cost-effectively than the other two approaches.

3.1. Review of HARDI

This section is organized in the following manner. In subsection 3.1.1, we review some research milestones in medical imaging that eventually led to the development of HARDI,

which is an effective method to detect multiple neuronal fibers. However, the original approach in implementing HARDI is only computationally feasible provided the HARDI signal is modelled after a multi-Gaussian mixture.

In subsection 3.1.2, we review a model-independent method called Q-ball imaging to implement HARDI. The computational speed of Q-ball imaging can be accelerated when (modified) spherical harmonics are used to represent the HARDI signals and their diffusion orientation distribution functions (ODFs) with analytic expressions. Details are then given about this variant of Q-ball imaging, which is also commonly known as spherical harmonics-based Q-ball imaging.

Finally, subsection 3.1.3 reviews another variant of Q-ball imaging based on spherical harmonics that yields sharper reconstruction of the HARDI ODFs. Details are then given about this variant, which is also known as the spherical harmonics-based constant solid angle Q-ball imaging.

3.1.1. Introduction to HARDI.

Diffusion Magnetic Resonance Imaging (d-MRI) is a non-invasive method to determine the directionality of neural fiber bundles through the diffusion of water molecules in brain tissues. d-MRI works on the assumption that the water molecules tend to diffuse along fibers in white matter. In turn, the directionality of neural fiber bundles results in accurate description of the geometry of brain microstructure. Applications of d-MRI include the characterization of neuro-degenerative diseases and surgical planning, etc.

In 1965, Stejskal and Tanner [**78**] developed a model to measure the diffusion strengths of water molecules in brain tissues. According to their model, for a given diffusion sensitizing gradient $\mathbf{q}$, the signal attenuation of the magnetic resonance (MR) signal $\mathbf{s}(\mathbf{q})$ can be expressed as the three-dimensional (3-D) Fourier transform $\mathcal{F}$ of the probability density function (PDF) $\mathbb{P}(\mathbf{r})$ of the average diffusion of water molecules, i.e.,

$$\mathbf{s}(\mathbf{q}) = \int_{\mathbb{R}^3} \mathbb{P}(\mathbf{r}) \exp(-\mathrm{i}\mathbf{q}^\top \mathbf{r})\, d\mathbf{r}, \quad \mathbf{q} \in \mathbb{R}^3, \tag{3.1.1}$$

where $\mathbf{r}$ represents the displacement vector of the water molecules over an experiment diffusion time. In d-MRI, the goal is to reconstruct the diffusion PDF $\mathbb{P}$. However, in practice, it is computationally expensive to obtain the diffusion PDF $\mathbb{P}$ using the Fourier transform as it requires huge measurements of $\mathbf{s}(\mathbf{q})$ over in $\mathbf{q}$-space, i.e., a wide range of $\mathbf{q} \in \mathbb{R}^3$.

Therefore, researchers came up with alternative acquisition and reconstruction techniques such as Diffusion Tensor Imaging (DTI) [**72**]. In 1992, Basser et al. [**9**] proposed a second

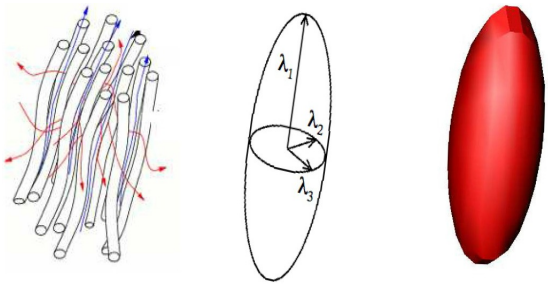

FIGURE 3.1.1. Diffusion Tensor Representation ([**38**]): the left figure displays water Brownian motion along the neural fibers; the middle figure shows the eigen-decomposition of the DT; the right figure gives an ellipsoidal visualization of the DT.

order symmetric positive-definite diffusion tensor D to model diffusion properties of biological tissues. Indeed, DTI implicitly assumes that the PDF of the displacement of water diffusion $\mathbb{P}$ is Gaussian. For DTI, by (3.1.1), we have the following correspondence between the signal attenuation of the MR signal $\mathbf{s}$, given by

$$\mathbf{s}(\mathbf{q}) = \exp(-\tau \mathbf{q}^\top D\mathbf{q}), \quad \mathbf{q} \in \mathbb{R}^3, \tau > 0, \tag{3.1.2}$$

and the PDF of the displacement of water diffusion $\mathbb{P}$, defined by

$$\mathbb{P}(\mathbf{r}) = (4\pi\tau|D|)^{-1/2} \exp(-\mathbf{r}^\top D^{-1}\mathbf{r}/4\tau), \quad \mathbf{r} \in \mathbb{R}^3,$$

where $|D|$ is the determinant of the diffusion tensor D, which is a symmetric positive definite matrix of the form:

$$D = R^\top \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3)R, \quad R^\top R = I,$$

with the eigenvalues of D, λ_1, λ_2, λ_3 satisfying $\lambda_1 > \lambda_2 \geq \lambda_3 > 0$, and their corresponding eigenvectors are given by e_1, e_2, e_3 respectively. The largest eigenvalue λ_1 corresponds to the principal direction of the diffusion tensor e_1 and the span of the other two eigenvectors form the orthogonal plane to it, as illustrated in Figure 3.1.1.

As we shall see later in this section, for the purposes of identifying the principal direction of the dominant diffusion tensor and computational efficiency for HARDI applications, it suffices to consider the HARDI signal defined on the unit spherical shell. Note that from (3.1.2), for some $\tau > 0$, by setting

$$\mathbf{g} := \frac{\mathbf{q}}{\|\mathbf{q}\|_2}, \quad b := \tau\|\mathbf{q}\|_2^2, \quad \mathbf{q} \in \mathbb{R}^3 \backslash \{0\},$$

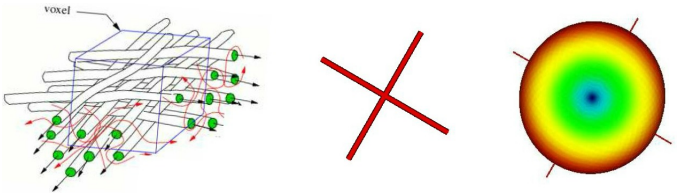

FIGURE 3.1.2. Limitation of DTI in voxels with crossing configuration. DTI cannot resolve imaging voxels containing multiple fiber crossings ([**38**]). The left picture displays water diffusion in a fiber crossing voxel; the middle picture shows the corresponding multiple fiber distributions; the right picture illustrates the inability of DTI to reconstruct the diffusion tensors correctly.

the corresponding Gaussian MR signal $\mathbf{s}$ (with a slight abuse of notation) defined on the unit sphere $\mathbb{S}^2$ can be expressed as

$$\mathbf{s}(\mathbf{g}) = \exp(-b\mathbf{g}^\top D\mathbf{g}), \quad \mathbf{g} \in \mathbb{S}^2.$$

For large b-values, the signal $\mathbf{s}$ decays quickly, while background noise is relatively unaffected, resulting in very noisy data measurements. For small b-values, this results in a very low signal attenuation. Thus it is important to have an appropriate trade-off in the choice of the value of b.

However, DTI is limited when imaging voxels with multiple fibers (e.g., a multi-Gaussian model, see (3.1.7) and (3.1.8)) which cross or branch, due to the single Gaussian PDF assumption. This limitation is illustrated for two orthogonal crossing fibers as in Figure 3.1.2. Note that the expected PDF has two maxima whereas the reconstructed DT profile is planar-like with no preferred diffusion direction.

In order to overcome these difficulties, Tuch et al. [**72**] proposed *High Angular Resolution Diffusion Imaging* (HARDI) as a method to resolve multiple fibers. The idea is that radial information of the diffusion PDF can be discarded if one is mainly interested in fiber directions. In HARDI, the *diffusion orientation distribution function* (ODF) is computed instead and it contains the full angular information of the PDF:

$$\Psi_1(\mathbf{u}) = \frac{1}{Z}\int_0^\infty \mathbb{P}(r\mathbf{u})\,dr, \quad \mathbf{u} \in \mathbb{S}^2, \tag{3.1.3}$$

where Z is a normalization constant such that $\int_{\mathbb{S}^2} \Psi_1 = 1$.

Suppose that the signal $\mathbf{s}$ is Gaussian (or its corresponding PDF $\mathbb{P}$ is Gaussian) with the following analytical forms:

$$\mathbf{s}(\mathbf{g}) = \exp(-b\mathbf{g}^\top D\mathbf{g}), \quad \mathbf{g} \in \mathbb{S}^2,$$
$$\mathbb{P}(\mathbf{u}) = K \exp(-\mathbf{u}^\top D^{-1}\mathbf{u}/4b), \quad \mathbf{u} \in \mathbb{S}^2, \tag{3.1.4}$$

where $K := [(4\pi b)^3|D|]^{-1/2}$. Then the corresponding ODF Ψ_1 can be calculated analytically as follows. For $\mathbf{u} \in \mathbb{S}^2$,

$$\begin{aligned}\Psi_1(\mathbf{u}) &= \frac{1}{Z}\int_0^\infty \mathbb{P}(r\mathbf{u})\,dr = \frac{K}{Z}\int_0^\infty \exp(-r^2\mathbf{u}^\top D^{-1}\mathbf{u}/4b)\,dr \\ &= \frac{K}{Z}\cdot\frac{1}{2}\sqrt{\frac{4b\pi}{\mathbf{u}^\top D^{-1}\mathbf{u}}} = \frac{1}{Z}\frac{1}{8\pi b|D|^{1/2}}(\mathbf{u}^\top D^{-1}\mathbf{u})^{-1/2},\end{aligned} \tag{3.1.5}$$

where we have used the following Gaussian integration result on the second line of (3.1.5): for $\alpha > 0$,

$$\int_0^\infty \exp(-\alpha r^2)\,dr = \frac{1}{2}\sqrt{\frac{\pi}{\alpha}}. \tag{3.1.6}$$

For the multi-Gaussian setting, the ODF can also be calculated analytically for M fibers. Indeed, suppose that the signal $\mathbf{s}$ is multi-Gaussian (or its corresponding PDF $\mathbb{P}$ is multi-Gaussian) with the analytical forms

$$\mathbf{s}(\mathbf{g}) = \sum_{i=1}^M p_i \exp(-b\mathbf{g}^\top D_i\mathbf{g}), \quad \mathbf{g} \in \mathbb{S}^2, \tag{3.1.7}$$

$$\mathbb{P}(\mathbf{u}) = \sum_{i=1}^M p_i K_i \exp(-\mathbf{u}^\top D_i^{-1}\mathbf{u}/4b), \quad \mathbf{u} \in \mathbb{S}^2, \quad \sum_{i=1}^M p_i = 1, \tag{3.1.8}$$

where $p_i \geq 0$ and $K_i = [(4\pi b)^3|D_i|]^{-1/2}$, $i = 1, 2, \ldots, M$. Then the corresponding ODF Ψ_1 can be calculated analytically: for $\mathbf{u} \in \mathbb{S}^2$,

$$\Psi_1(\mathbf{u}) = \frac{1}{Z}\sum_{i=1}^M p_i \frac{1}{8\pi b|D_i|^{1/2}}(\mathbf{u}^\top D_i^{-1}\mathbf{u})^{-1/2}.$$

In HARDI applications, the number of fibers M ranges from 1 to 4, the coefficients $\{p_i\}_{i=1}^M$ are typically equally weighted. This multi-Gaussian setting is usually a reasonable model for voxels possessing two or more fibers interacting with one another. Furthermore, it has analytical PDF and ODF expressions and therefore it can be used to generate synthetic datasets for validation against ground truth. In the last section of this chapter, the multi-Gaussian model will also be used in our synthetic datasets and experiments

HARDI actually reduces computational burden as sampling is only needed on a single spherical shell (as opposed to sampling on a 3D Cartesian grid). Typically, 60 to 200 samples

are required on a single spherical shell. Furthermore, significantly weaker imaging gradients, i.e., lower b-values ($1000 \leq b \leq 3000$) are needed in HARDI, improving the signal-to-noise ratio (SNR) measurements. This is in contrast with older HARDI-related techniques such as Diffusion Spectrum Imaging (DSI) [**84**], which usually requires much more than 200 samples in 3D Cartesian grid and very strong imaging gradients ($3000 \leq b \leq 20000$), leading to worse SNR measurements and reconstructions.

3.1.2. SH-based Q-ball imaging (SH-based QBI).

One limitation of the approach described in (3.1.5) to reconstruct the ODF, is that one needs to estimate the PDF $\mathbb{P}$ of the HARDI signal, which is generally not known or given in practical situations. However, in experiments, usually only information about the signal $\mathbf{s}$ is given, and model-independent (i.e., the HARDI signal and PDF need not be Gaussian) methods that utilize the signal $\mathbf{s}$ directly to reconstruct the ODF are desired.

Thus, one such model-independent method called *Q-Ball Imaging* (QBI) was developed by Tuch [**82**]. Tuch showed that one can use the *Funk Radon Transform* (FRT) to reconstruct a smooth approximation of the ODF from HARDI samples on a single shell as follows:

$$\Psi_1(\mathbf{v}) \approx \mathcal{R}[\mathbf{s}](\mathbf{u}) := \int_{\mathbf{v} \in C(\mathbf{u})} \mathbf{s}(\mathbf{v})\, d\sigma(\mathbf{v}), \tag{3.1.9}$$

where $\mathbf{s}$ is the HARDI signal, σ rotation invariant measure on $\mathbb{S}^2$ and $C(\mathbf{u})$ is the great circle with pole $\mathbf{u}$, i.e., the circle formed by the intersection of $\mathbb{S}^2$ with a 3D-plane containing the origin O, with normal $\mathbf{u}$. Although the approximation of the ODF proved to be reasonably good in practical situations, computationally expensive numerical integration along many great circles are needed.

To avoid the computational burden, Descoteaux et al. [**38**] introduced a method called *Analytical Q-Ball Imaging* (Analytical QBI) using spherical harmonic representations. They first supposed that the HARDI signal $\mathbf{s}$ can be approximated by a modified spherical harmonic representation of the form

$$\mathbf{s}(\mathbf{w}) = \sum_{j=1}^{R} c_j \tilde{Y}_j(\mathbf{w}), \quad \mathbf{w} \in \mathbb{S}^2,$$

then the FRT $\mathcal{R}$ is applied onto $\mathbf{s}$, yielding the following approximation

$$\Psi_1(\mathbf{u}) \approx \mathcal{R}[\mathbf{s}](\mathbf{u}) = \sum_{j=1}^{R} 2\pi P_{l_j}(0) c_j \tilde{Y}_j(\mathbf{u}), \quad \mathbf{u} \in \mathbb{S}^2, \tag{3.1.10}$$

where l_j is the order of the modified spherical harmonic function $\tilde{Y}_j$, and $P_{l_j}(0)$ is the Legendre polynomial of degree l_j evaluated at 0, i.e.,

$$P_{l_j}(0) = \begin{cases} 0, & \text{if } l_j \text{ is odd,} \\ (-1)^{l_j/2} \frac{1\cdot 3\cdot 5 \ldots (l_j-1)}{2\cdot 4\cdot 6 \ldots l_j}, & \text{if } l_j \text{ is even.} \end{cases} \tag{3.1.11}$$

The above derivation is actually a corollary of the following Funk-Hecke theorem.

THEOREM 3.1.1. (Funk-Hecke theorem [**14**]) *Let $f \in C[-1,1]$ and Y_l^m be any spherical harmonic of degree $m \in [-l,l]$ and order $l \in \mathbb{N}_0$. Then, given $\mathbf{u} \in \mathbb{S}^2$,*

$$\int_{\mathbf{w}\in\mathbb{S}^2} f(\mathbf{u}^\top \mathbf{w}) Y_l^m(\mathbf{w})\, d\mathbf{w} = \lambda_l Y_l^m(\mathbf{u}), \quad |m| \le l,\, l \in \mathbb{N}_0,$$

where $\lambda_l = 2\pi \int_{-1}^{1} f(t) P_l(t)\, dt$ and P_l is the Legendre polynomial of degree l.

The reader may refer to [**14**] for a proof of the Funk-Hecke theorem. It should be noted that a simple calculation reveals that the Funk-Hecke theorem remains true when the classical spherical harmonics Y_l^m are replaced with the modified spherical harmonics $\tilde{Y}_l^m$.

COROLLARY 3.1.2. ([**38**]) *The (modified) spherical harmonics are eigenfunctions of the FRT with eigenvalues given as follows: for $l \in \mathbb{N}_0$, $|m| \le l$,*

$$\mathcal{R}[\tilde{Y}_l^m](\mathbf{u}) = 2\pi P_l(0) \tilde{Y}_l^m(\mathbf{u}), \quad \mathbf{u} \in \mathbb{S}^2,$$

where $P_l(0)$ is the Legendre polynomial of degree l evaluated at 0, given explicitly in (3.1.11).

The proof of Corollary 3.1.2 is given in the Appendix.

In practice, we usually discretize the input HARDI signal $\mathbf{s}$ with K gradient vectors $\{g_i\}_{i=1}^K$. A naive but natural approach to reconstruct the ODF would be the classical least-squares method. Setting $f = [\mathbf{s}(g_i)]_{i=1}^K$, the least-squares approach seeks the modified SH coefficient vector $c = [c_j]_{j=1}^R \in \mathbb{R}^R$ such that it is optimal to the following problem:

$$\min_{c\in\mathbb{R}^R} \frac{1}{2} \|Bc - f\|_2^2,$$

where the $K \times R$ (assume at this moment $K \ge R$ for simplicity) matrix B is given by

$$B = \begin{bmatrix} \tilde{Y}_1(g_1) & \tilde{Y}_2(g_1) & \ldots & \tilde{Y}_R(g_1) \\ \tilde{Y}_1(g_2) & \tilde{Y}_2(g_2) & \ldots & \tilde{Y}_R(g_2) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{Y}_1(g_K) & \tilde{Y}_2(g_K) & \ldots & \tilde{Y}_R(g_K) \end{bmatrix}. \tag{3.1.12}$$

Then the optimal modified SH coefficient vector c is given by

$$c = (B^\top B)^{-1} B^\top f,$$

provided that $B^\top B$ is invertible (in practice, this can be made readily possible by choosing a set of sampling gradients $\{g_i\}_{i=1}^K$ which are relatively evenly spaced out on the unit sphere). Then, by (3.1.10), the discretized ODF can be approximated by

$$\Psi_{1,d} \approx BPc,$$

where

$$P := \mathrm{diag}[2\pi P_{l_j}(0)]_{j=1}^R. \tag{3.1.13}$$

To improve reconstruction results of the ODF, Descotaux et al. [**38**] proposed Laplace-Beltrami (L-B) regularization upon the optimization problem as follows:

$$\min_{c\in\mathbb{R}^R} \frac{1}{2}\|Bc - f\|_2^2 + \frac{\lambda}{2}\|Lc\|_2^2, \tag{3.1.14}$$

where λ is a pre-defined positive parameter and

$$L := \mathrm{diag}[-l_j(l_j+1)]_{j=1}^R. \tag{3.1.15}$$

Note that the model in (3.1.14) is actually a discretized version of the following variational model:

$$\min_{c\in\mathbb{R}^R} \frac{1}{2}\int_{\mathbb{S}^2} (\sum_{j=1}^R c_j \tilde{Y}_j - \mathbf{s})^2 + \frac{\lambda}{2}\int_{\mathbb{S}^2} |\Delta_b(\sum_{j=1}^R c_j \tilde{Y}_j)|^2,$$

where we note that

$$\Delta_b \tilde{Y}_l^m = -l(l+1)\tilde{Y}_l^m, \quad |m| \le l,\, l \in 2\mathbb{N}_0,$$

and Δ_b is the Laplace-Beltrami operator on $\mathbb{S}^2$ defined as

$$\Delta_b f := \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta\frac{\partial}{\partial\theta}f) + \frac{1}{\sin^2\theta}\frac{\partial^2 f}{\partial\phi^2}. \tag{3.1.16}$$

This regularization is done to improve denoising effects as it tends to reduce the magnitude of SH coefficients $\{c_j\}_{j=1}^R$ corresponding to higher order SH, which amplify noise. Then the optimal modified SH coefficient vector c is given by

$$c = [B^\top B + \lambda L^\top L]^{-1} B^\top f, \tag{3.1.17}$$

and as before, one can reconstruct the discretized ODF

$$\Psi_{1,d} \approx BPc.$$

This ODF reconstruction method is known as the *spherical harmonics-based Q-ball imaging* (SH-based QBI).

3.1.3. SH-based constant solid angle QBI (SH-based CSA QBI).

In [**3**], it was pointed out that, if we represent the orientation of the unit vector $\mathbf{u}$ using the spherical coordinate (θ, ϕ), then

$$\begin{aligned}\int_{\mathbb{R}^3} \mathbb{P}(\mathbf{r})\, d\mathbf{r} &= \int_0^{\pi}\int_0^{2\pi}\int_0^{\infty} \mathbb{P}(r\mathbf{u}) r^2 \sin(\theta) dr d\phi d\theta \\ &= \int_{\mathbb{S}^2} \Psi_2(\mathbf{u})\, d\sigma(\mathbf{u}),\end{aligned} \tag{3.1.18}$$

where σ is the rotation-invariant measure on the sphere and Ψ_2 is the ODF on the unit sphere represented by

$$\Psi_2(\mathbf{u}) = \int_0^{\infty} P(r\mathbf{u}) r^2 dr, \quad \mathbf{u} \in \mathbb{S}^2. \tag{3.1.19}$$

Remark 3.1.3. In (3.1.18), we may regard $\Psi_2(\mathbf{u})\, d\sigma(\mathbf{u})$ as the probability of diffusion direction $\mathbf{u}$ through a "very small constant solid" (infinitesimal) angle $d\sigma(\mathbf{u})$. We note that in the HARDI literature, the above version of the ODF definition, i.e., Ψ_2 in (3.1.19), is actually mathematically correct as it can be shown that Ψ_2 is a probability distribution function, which does not require any normalization factor Z, as opposed to Ψ_1 in (3.1.3). This in turn leads to sharper reconstruction of the ODF and better resolution of multiple fiber orientations.

It was also shown in [**3**] that if the PDF $\mathbb{P}$ is Gaussian with the same analytical form in (3.1.4), then the corresponding ODF Ψ_2 can be calculated analytically using integration by parts and the integration techniques in (3.1.5) and (3.1.6):

$$\begin{aligned}\Psi_2(\mathbf{u}) &= \int_0^{\infty} \mathbb{P}(r\mathbf{u}) r^2\, dr \\ &= K \int_0^{\infty} \exp(-r^2 \mathbf{u}^\top D^{-1}\mathbf{u}/4b)\, r^2 dr \\ &= \frac{1}{8\pi b |D|^{1/2}} (\mathbf{u}^\top D^{-1}\mathbf{u})^{-3/2}.\end{aligned}$$

For the multi-Gaussian setting, when (3.1.19) is applied to (3.1.8), its corresponding ODF Ψ_2 can also be calculated analytically: for $\mathbf{u} \in \mathbb{S}^2$,

$$\Psi_2(\mathbf{u}) = \sum_{i=1}^{M} p_i \frac{1}{8\pi b |D_i|^{1/2}} (\mathbf{u}^\top D_i^{-1}\mathbf{u})^{-3/2}. \tag{3.1.20}$$

Agani et al. [**3**] showed that we can approximate the ODF Ψ_2 by the modified HARDI signal $\tilde{\mathbf{s}}$ in the following manner:

$$\Psi_2(\mathbf{u}) \approx \frac{1}{4\pi} + \frac{1}{16\pi^2}\mathcal{R}[\Delta_b \tilde{\mathbf{s}}](\mathbf{u}), \quad \mathbf{u} \in \mathbb{S}^2, \tag{3.1.21}$$

where

$$\tilde{\mathbf{s}}(\mathbf{u}) = \log(-\log \mathbf{s}(\mathbf{u})), \quad \mathbf{u} \in \mathbb{S}^2, \tag{3.1.22}$$

with $0 < \mathbf{s}(\mathbf{u}) < 1$, $\mathbf{u} \in \mathbb{S}^2$, $\mathcal{R}$ and Δ_b represent respectively the Funk-Radon transform and the spherical Laplace-Beltrami operator defined in (3.1.9) and (3.1.16). If the *modified signal* $\tilde{\mathbf{s}}$ in (3.1.22) is represented using the modified spherical harmonics, i.e.,

$$\tilde{\mathbf{s}}(\mathbf{u}) = \sum_{j=1}^{R} c_j \tilde{Y}_j(\mathbf{u}), \quad \mathbf{u} \in \mathbb{S}^2,$$

then the ODF Ψ_2 is given as

$$\begin{aligned}
\Psi_2(\mathbf{u}) &\approx \frac{1}{4\pi} + \frac{1}{16\pi^2}\mathcal{R}[\Delta_b \tilde{\mathbf{s}}](\mathbf{u}) \\
&= \frac{1}{4\pi} + \frac{1}{16\pi^2}\sum_{j=1}^{R} c_j \mathcal{R}[\Delta_b \tilde{Y}_j](\mathbf{u}) \\
&= \frac{1}{4\pi} + \frac{1}{16\pi^2}\sum_{j=1}^{R} c_j 2\pi P_{l_j}(0)[-l_j(l_j+1)]\tilde{Y}_j(\mathbf{u}).
\end{aligned}$$

If $\tilde{\mathbf{s}}$ is discretized with K gradient vectors $\{g_i\}_{i=1}^{K}$, we set $f = [\tilde{\mathbf{s}}(g_i)]_{i=1}^{K}$. Using the Laplace-Beltrami (L-B) regularization in (3.1.14) to obtain the SH coefficient vector $c = [c_j]_{j=1}^{R}$ as in (3.1.17), then the discretized ODF is given by

$$\Psi_{2,d} \approx \frac{1}{4\pi} + \frac{1}{16\pi^2} BLPc,$$

where B, L, and P are the matrices described in (3.1.12), (3.1.15) and (3.1.13) respectively. In view of Remark 3.1.3, this ODF reconstruction method is also known as the *SH-based Constant Solid Angle QBI* (SH-based CSA QBI).

3.2. Review of Spherical Ridgelets (SR) for HARDI

This section is organized in the following manner. In subsection 3.2.1, we review how spherical ridgelets (SR) are being constructed by Michailovich et al. [**64**] as an alternative set of functions (for representing HARDI signals) to the modified spherical harmonics. Due to its localization properties, spherical ridgelets provide sparser representations of HARDI

signals than the modified spherical harmonics, and thus compressed sensing (CS) techniques can be now exploited to perform analytical Q-ball imaging. This is also known as spherical ridgelets-based Q-ball imaging (SR-based QBI).

In subsection 3.2.2, in order to yield sharper reconstruction of the HARDI ODFs, we develop a 'constant solid angle' (CSA) variant of the SR-based QBI, which we will name as SR-based CSA QBI. This subsection is analogous to the earlier subsection 3.1.3 on SH-based CSA QBI.

3.2.1. SR-based QBI.

Although SH-based QBI techniques for HARDI performs much better than DTI for the reconstruction of multiple crossing fibers, HARDI requires a substantially larger number (60-200) of diffusion-encoding gradients K, as compared to $K \in [25, 35]$ in the case of DTI. As the total scanning time increases linearly with K, HARDI may be too slow to be effectively used in clinical applications involving children or patients with dementia as accurate HARDI measurements require them to stay still for a prolonged period.

The above deficiency of HARDI can be overcome using the theory of compressed sensing (CS) [**29**, **30**, **40**] which predicts that sparse signals and images can be reconstructed from what was previously believed to be incomplete information. Furthermore, efficient algorithms such as ℓ_1-minimization can be used for recovery. It should be noted that there already exists much work in which the theory of CS has been used for reconstruction of grayscale MR images from their subcritical samples, e.g., [**50**, **55**, **58**].

Although the spherical harmonics provide a reasonably stable representation of the HARDI ODFs, they generally do not yield sparse representations of HARDI ODFs. This is because the supports of the HARDI ODFs tend to be localized along a few (typically one or two) pairs of radial directions as illustrated in Figure 3.2.1, whereas the the supports of the (modified) spherical harmonics are rather globalized or localized along fixed and incompatible pairs of radial directions as seen in Figure 1.3.2.

Motivated by the need to provide sparse representations of HARDI signals, Michailovich et al. [**64**] constructed *spherical ridgelets* (SR). The basic construction idea is that, since the FRT $\mathcal{R}$ is used on representation functions $\{\psi_{j,\mathbf{v}}\}_{j\in\mathcal{I},\mathbf{v}\in\mathbb{S}^2}$ to approximate the ODFs as in (3.1.10) and the ODFs of the HARDI signals are localized along a few radial directions, then $\{\mathcal{R}[\psi_{j,\mathbf{v}}]\}_{j\in\mathcal{I},\mathbf{v}\in\mathbb{S}^2}$ should also have similar localization behavior.

Therefore, Michailovich et al. [**64**] started out by designing localized spherical kernels intended to represent the ODF. Specifically, let $\rho \in (0, 1)$ be a positive scaling parameter.

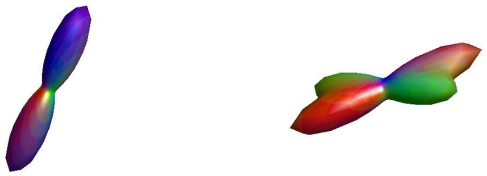

FIGURE 3.2.1. Examples of HARDI ODFs: (left subplot) an ODF with a single fiber direction; (right subplot) an ODF with two fiber directions.

They defined the Gaussian-Weiestrass scaling kernel $\chi_{j,\mathbf{v}} : \mathbb{S}^2 \to \mathbb{R}$ at resolution $j \in \mathbb{N}_0$ and orientation $\mathbf{v} \in \mathbb{S}^2$ as follows:

$$\chi_{j,\mathbf{v}}(\mathbf{u}) := \sum_{l=0}^{\infty} \frac{2l+1}{4\pi} \kappa_j[l] P_l(\mathbf{u}^\top \mathbf{v}), \quad \mathbf{u} \in \mathbb{S}^2, \tag{3.2.1}$$

where $\kappa_j[l] := \exp\left\{-\rho \frac{l}{2^j}(\frac{l}{2^j}+1)\right\}$ and P_l denote the Legendre polynomial of order l, $j, l \in \mathbb{N}_0$. The L^2-energy of $\chi_{j,\mathbf{v}}$ is concentrated around the point $\mathbf{v}$, and $\chi_{j,\mathbf{v}}$ gets more localized as j increases.

The spherical ridgelets are then constructed using the FRT $\mathcal{R}$, where the spherical ridgelet functions $\psi_{j,\mathbf{v}}$, $j \geq -1$, $\mathbf{v} \in \mathbb{S}^2$, are obtained from $\chi_{j,\mathbf{v}}$ according to

$$\psi_{j,\mathbf{v}}(\mathbf{u}) := \frac{1}{2\pi} \mathcal{R}[\chi_{j+1,\mathbf{v}} - \chi_{j,\mathbf{v}}](\mathbf{u}), \quad \mathbf{u} \in \mathbb{S}^2,$$

where $\chi_{-1,\mathbf{v}} \equiv 0$. Using the FRT $\mathcal{R}$ on (3.2.1) yields the following analytical form for the spherical ridgelets

$$\psi_{j,\mathbf{v}}(\mathbf{u}) = \sum_{l=0}^{\infty} P_l(0) \frac{2l+1}{4\pi} [\kappa_{j+1}(l) - \kappa_j(l)] P_l(\mathbf{u}^\top \mathbf{v}), \quad \mathbf{u} \in \mathbb{S}^2, \tag{3.2.2}$$

where $\kappa_{-1}(l) = 0$, $l \in \mathbb{N}_0$, and $P_l(0)$ is the Legendre polynomial of degree l evaluated at 0 as seen in (3.1.11). This is because the spherical harmonics Y_l^m are the eigenfunctions of the FRT $\mathcal{R}$ with eigenvalue $2\pi P_l(0)$ and the *addition theorem for spherical harmonics* [**38**], i.e., for $\mathbf{u}, \mathbf{v} \in \mathbb{S}^2$,

$$P_l(\mathbf{u}^\top \mathbf{v}) = \frac{4\pi}{2l+1} \sum_{m=-l}^{l} \overline{Y_l^m(\mathbf{u})} Y_l^m(\mathbf{v}).$$

The interested reader may refer to [**14**] for its proof. It was then shown in [**64**] that

$$\mathbb{U} := \left\{\psi_{j,\mathbf{v}} : j = -1, 0, \ldots, \mathbf{v} \in \mathbb{S}^2\right\}$$

forms a *semi-discrete frame* in $L^2_{\mathrm{sym}}(\mathbb{S}^2)$, i.e.,

$$A\|f\|_2^2 \leq \sum_{j\geq -1} \int_{\mathbb{S}^2} |\langle f, \psi_{j,\mathbf{v}}\rangle|^2 \, d\sigma(\mathbf{v}) \leq B\|f\|_2^2, \quad \forall f \in L^2_{\mathrm{sym}}(\mathbb{S}^2).$$

However, $\mathbb{U}$ is an infinite-dimensional set and thus infeasible for practical applications. It was recommended in [**65**] to instead use the following discrete counterpart $\mathbb{U}_d(J)$ defined by

$$\mathbb{U}_d(J) := \left\{ \psi_{j,\mathbf{v}_j^i} \,:\, i = 1, 2, \ldots, M_j,\, j = -1, 0, \ldots, J \right\},$$

where $M_j := (2^{j+1} \cdot 3 + 1)^2$, $j = -1, 0, \ldots, J$ and J is a pre-defined positive integer (chosen as $J = 1$ in numerical experiments of [**65**]). The set of all possible orientations $\mathbf{v} \in \mathbb{S}^2$ in $\mathbb{U}$ is discretized by taking a suitably chosen finite set of M_j directions $\left\{\mathbf{v}_j^i\right\}_{i=1}^{M_j}$ on the sphere for each level j. For instance, for each $j \geq -1$, $\left\{\mathbf{v}_j^i\right\}_{i=1}^{M_j}$ can be formed by a tensor product of equidistant azimuth and zenith points of the form

$$\mathbf{v}_j^i := (\cos\theta_{\alpha,j} \sin\phi_{\beta,j}, \sin\theta_{\alpha,j} \sin\phi_{\beta,j}, \cos\phi_{\beta,j}), \quad i := (\alpha, \beta), \quad \alpha, \beta = 0, 1, \ldots, 2^{j+1} \cdot 3,$$

where the azimuth points and zenith points are given respectively as

$$\theta_{\alpha,j} := \frac{\pi\beta}{2^{j+1} \cdot 3 + 1}, \quad \phi_{\beta,j} := \frac{2\pi\beta}{2^{j+1} \cdot 3 + 1}, \quad \alpha, \beta = 0, 1, \ldots, 2^{j+1} \cdot 3.$$

To simplify notations, the spherical ridgelets in $\mathbb{U}_d(J)$ are indexed as $\psi_m(\mathbf{u})$, with combined index $m = 1, 2, \ldots, R$, where $R = \sum_{j=-1}^{J} M_j$. Suppose that we have the following approximation to a HARDI signal $\mathbf{s}$ at a single voxel:

$$\mathbf{s} \approx \sum_{j=-1}^{J} \sum_{i=1}^{M_j} c_{j,i} \psi_{j,\mathbf{v}_j^i}.$$

Since an approximation to the ODF Ψ_1 can be obtained by the applying the FRT onto the HARDI signal $\mathbf{s}$ according to [**64**,**65**], i.e., $\Psi_1 \approx \mathcal{R}[\mathbf{s}]$, it can be reconstructed analytically by the following spherical ridgelet expansion:

$$\begin{aligned}
\Psi_1(\mathbf{u}) &\approx \sum_{j=-1}^{J} \sum_{i=1}^{M_j} c_{j,i} \mathcal{R}[\psi_{j,\mathbf{v}_j^i}](\mathbf{u}), \\
&= \sum_{j=-1}^{J} \sum_{i=1}^{M_j} c_{j,i} \sum_{l=0}^{\infty} [P_l(0)]^2 \cdot \frac{2l+1}{2} [\kappa_{j+1}(l) - \kappa_j(l)] P_l(\mathbf{u}^\top \mathbf{v}_j^i).
\end{aligned}$$

However, in practical situations, one is given a sampling set of K diffusion gradients $\{g_k\}_{k=1}^K$. Then one can use (3.2.2) to compute the values of the spherical ridgelets in $\mathbb{U}_d(J)$

over the sampling set[1]. This discretization gives rise to a $K \times R$ matrix B_2 given by

$$B_2 = \begin{bmatrix} \psi_1(g_1) & \psi_2(g_1) & \dots & \psi_R(g_1) \\ \psi_1(g_2) & \psi_2(g_2) & \dots & \psi_R(g_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(g_K) & \psi_2(g_K) & \dots & \psi_R(g_K) \end{bmatrix}.$$

At a single voxel, if $\mathbf{s}$ is discretized with K gradient vectors $\{g_i\}_{i=1}^K$ we have

$$f = B_2 c + e,$$

where $f = [\mathbf{s}(g_i)]_{i=1}^K$, $c := [c_j]_{j=1}^R$, e is a vector to account for both measurement noise and modelling errors.

In [**65**], the sparse reconstruction of the HARDI ODF at a single voxel was implemented by considering the ℓ_1-minimization model

$$\min_{c \in \mathbb{R}^R} \frac{1}{2}\|f - B_2 c\|_2^2 + \lambda \|c\|_1, \tag{3.2.3}$$

where $\|c\|_1 = \sum_{j=1}^{R} |c_j|$. The *fast iterative soft-thresholding algorithm* (FISTA) [**10**] below was used in [**65**] to solve the CS problem in (3.2.3). Here, T_α^1 is the soft-thresholding operator defined in (1.1.12).

Algorithm 1 FISTA with constant stepsize

Input: Spectral norm of $B_2^\top B_2$, i.e., $\|B_2^\top B_2\|_2$ and tolerance level 'tol',

Output: Updated coefficient vector c^k after the k iteration,

1: **Initialization** $y^1 = c^0$, $t^1 = 1$, set penalty parameter $\rho > \|B_2^\top B_2\|_2$,
2: **while** $\|c^k - c^{k-1}\|_2 >$ 'tol', **do**
3: $c^k = T_{\lambda/\rho}^1[y^k - \rho^{-1} B_2^\top (B_2 y^k - f)]$,
4: $t^{k+1} = 0.5\left\{1 + \sqrt{1 + 4(t^k)^2}\right\}$,
5: $y^{k+1} = c^k + \frac{t^k - 1}{t^{k+1}}(c^k - c^{k-1})$,
6: $k = k + 1$,
7: **end while**

[1]Since the definition in (3.2.2) involves an infinite sum, for practical purposes, in our experiments, we truncate the summation at the 50th term as the magnitude of its summand drops below 10^{-10} for all the chosen orientations $\{\mathbf{v}_j^i\}_{i=1}^{M_j}$, $j = -1, 0, \dots, J$.

After obtaining a satisfactory coefficient vector c from FISTA, according to [**65**], the discrete ODF $\Psi_{1,d}$ at a single voxel is then reconstructed using the formula

$$\Psi_{1,d} \approx B_2 P c,$$

where P is the corresponding matrix defined in (3.1.13). We name this ODF reconstruction method as the *spherical ridgelets-based QBI* (SR-based QBI).

REMARK 3.2.1. We point out that although the set $\mathbb{U}$ forms a semi-discrete frame for $L^2(\mathbb{S}^2)$, it is unknown whether the discrete set of spherical ridgelets in $\mathbb{U}_d := \lim_{J\to+\infty} \mathbb{U}_d(J)$ actually forms a frame for $L^2_{\text{sym}}(\mathbb{S}^2)$ or $H^s_{\text{sym}}(\mathbb{S}^2)$ for some $s \in \mathbb{R}$. A similar situation also holds for the modified spherical ridgelets given in the next subsection.

3.2.2. SR-based CSA QBI.

In the previous subsection 3.1.3, a constant solid angle (CSA) version of analytical QBI was implemented to obtain a sharper reconstruction of the ODF using the formulas (3.1.19) and (3.1.21) involving Ψ_2:

$$\Psi_2(\mathbf{u}) \approx \frac{1}{4\pi} + \frac{1}{16\pi^2}\mathcal{R}[\Delta_b \tilde{\mathbf{s}}](\mathbf{u}), \quad \mathbf{u} \in \mathbb{S}^2,$$

where $\tilde{\mathbf{s}}$ is modified from the HARDI signal $\mathbf{s}$ by

$$\tilde{\mathbf{s}}(\mathbf{u}) = \log(-\log \mathbf{s}(\mathbf{u})), \quad \mathbf{u} \in \mathbb{S}^2,$$

with a mild assumption that $0 < \mathbf{s}(\mathbf{u}) < 1$, $\mathbf{u} \in \mathbb{S}^2$, $\mathcal{R}$ and Δ_b represent respectively the Funk-Radon transform and the spherical Laplace-Beltrami operator defined in (3.1.9) and (3.1.16). Spherical harmonics were then used to obtain analytical representations of the modified signal $\tilde{\mathbf{s}}$ and the ODF Ψ_2.

In this subsection, to obtain a sharper reconstruction of the ODF, we instead design *modified spherical ridgelets* $\left\{\psi^\dagger_{j,\mathbf{v}}\right\}_{j,\mathbf{v}}$ (modified SR) to obtain analytical representations of the modified signal $\tilde{\mathbf{s}}$ and the ODF Ψ_2. For $\mathbf{v} \in \mathbb{S}^2$, $j = -1, 0, \ldots$, let $\psi^\dagger_{j,\mathbf{v}} : \mathbb{S}^2 \to \mathbb{R}$ be defined by

$$\psi^\dagger_{j,\mathbf{v}}(\mathbf{u}) := \sum_{l=0}^{\infty} \delta_l^{-1} P_l(0) \frac{2l+1}{4\pi} [\kappa_{j+1}(l) - \kappa_j(l)] P_l(\mathbf{u}^\top \mathbf{v}), \, \mathbf{u} \in \mathbb{S}^2, \tag{3.2.4}$$

where

$$\delta_l := \begin{cases} 1, & l = 0, \\ -l(l+1), & l \in \mathbb{N}. \end{cases}$$

Note that the motivation of the term δ_l^{-1}, $l \in \mathbb{N}_0$, in (3.2.4) is that prior to the mappings $\mathcal{R}$ and Δ_b, $\psi^\dagger_{j,\mathbf{v}}(\mathbf{u})$ acts as a pre-image of $\mathcal{R}[\Delta_b \psi^\dagger_{j,\mathbf{v}}]$ which is locally concentrated on the radial direction $\mathbf{v}$, enabling the collection $\{\mathcal{R}[\Delta_b \psi^\dagger_{j,\mathbf{v}}]\}_{j,\mathbf{v}}$ to give a possibly more efficient representation of the ODF.

Similar to the spherical ridgelets in subsection 3.2.1, the infinite-dimensional set $\mathbb{U}^\dagger := \left\{\psi^\dagger_{j,\mathbf{v}} : j = -1, 0, \ldots, \mathbf{v} \in \mathbb{S}^2\right\}$ is replaced by its discrete counterpart $\mathbb{U}^\dagger_d(J)$ defined by

$$\mathbb{U}^\dagger_d(J) := \left\{\psi^\dagger_{j,\mathbf{v}^i_j} : i = 1, 2, \ldots, M_j,\, j = -1, 0, \ldots, J\right\}, \tag{3.2.5}$$

where $J \in \mathbb{N}$. The set of all possible orientations $\mathbf{v} \in \mathbb{S}^2$ in $\mathbb{U}$ is discretized by taking a suitably chosen finite set of M_j directions $\mathbf{v}^i_j$ on the sphere for each level j. As before, in our experiments, we set $M_j = (2^{j+1} \cdot 3 + 1)^2$, $j = -1, 0, \ldots, J$, $J = 1$, and $R := \sum_{j=-1}^{1} M_j$.

Let us now consider the following expansion of the modified HARDI signal $\tilde{\mathbf{s}}$ using the modified spherical ridgelets

$$\tilde{\mathbf{s}} \approx \sum_{j=-1}^{J} \sum_{i=1}^{M_j} c_{j,i} \psi^\dagger_{j,\mathbf{v}^i_j}.$$

Since an approximation to the ODF Ψ_2 can be obtained by applying the Laplace-Beltrami operator followed by the FRT onto the modified HARDI signal $\tilde{\mathbf{s}}$ according to $\Psi_2 \approx \frac{1}{4\pi} + \frac{1}{16\pi^2}\mathcal{R}[\Delta_b \tilde{\mathbf{s}}]$, the ODF can be reconstructed analytically by the following modified spherical ridgelet expansion

$$\Psi_2(\mathbf{u}) \approx \frac{1}{4\pi} + \frac{1}{16\pi^2} \sum_{j=-1}^{J} \sum_{i=1}^{M_j} c_{j,i} \mathcal{R}[\Delta_b \psi^\dagger_{j,\mathbf{v}^i_j}](\mathbf{u}), \quad \mathbf{u} \in \mathbb{S}^2,$$

where for $M_j = (2^{j+1} \cdot 3 + 1)^2$, $j = -1, 0, \ldots, J$,

$$\mathcal{R}[\Delta_b \psi^\dagger_{j,\mathbf{v}^i_j}](\mathbf{u}) = \sum_{l=0}^{\infty} [P_l(0)]^2 \frac{2l+1}{2} [\kappa_{j+1}(l) - \kappa_j(l)] P_l(\mathbf{u}^\top \mathbf{v}^i_j),\ \mathbf{u} \in \mathbb{S}^2,$$

As explained in the previous subsection, for computational purposes, a sampling set of K diffusion gradients $\{g_k\}_{k=1}^K$ is used to compute the values of the modified spherical ridgelets

in $\mathbb{U}_d^\dagger(J)$ over the sampling set[2]. This discretization gives rise to a $K \times R$ matrix B_3 given by

$$B_3 = \begin{bmatrix} \psi_1^\dagger(g_1) & \psi_2^\dagger(g_1) & \dots & \psi_R^\dagger(g_1) \\ \psi_1^\dagger(g_2) & \psi_2^\dagger(g_2) & \dots & \psi_R^\dagger(g_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1^\dagger(g_K) & \psi_2^\dagger(g_K) & \dots & \psi_R^\dagger(g_K) \end{bmatrix}. \tag{3.2.6}$$

If $\tilde{\mathbf{s}}$ is discretized with K gradient vectors $\{g_k\}_{k=1}^K$ we have $f \approx B_3 c$, where $f = [\tilde{\mathbf{s}}(g_k)]_{k=1}^K$, $c = [c_j]_{j=1}^R$, and a sparse reconstruction of the HARDI ODF at a single voxel was implemented by considering the same ℓ_1-minimization model in (3.2.3). Thus the algorithm FISTA could be used as before to obtain a suitable estimate of the coefficient vector c, and an estimate of the discrete ODF $\Psi_{2,d}$ is then reconstructed as follows:

$$\Psi_{2,d} \approx \frac{1}{4\pi} + \frac{1}{16\pi^2} B_3 L P c,$$

where L,P are the corresponding matrices defined respectively in (3.1.15) and (3.1.13). We name this ODF reconstruction method as the *spherical ridgelet-based constant solid angle Q-ball imaging* (SR-based CSA QBI).

3.3. Application of Wavelet Frames (WF) to HARDI

Recall from section 2.3 that the collection

$$\Gamma(\Psi) := \left\{ \psi_k^i(\cdot; \eta_{k,\nu}^i) \,:\, \nu \in \mathcal{J}_k',\, k \geq -1,\, i = 1, 2 \right\} \tag{3.3.1}$$

forms a normalized tight wavelet frame for $L^2_{\mathrm{sym}}(\mathbb{S}^2)$, where for $i = 1, 2$, $k \geq -1$, $\nu \in \mathcal{J}_k'$,

$$\psi_k^i(\cdot; \eta_{k,\nu}^i) = \sqrt{\hat{d}_{k,\nu}} \sum_{(m,l) \in \mathcal{J}_k^i} \overline{\tilde{Y}_l^m(\eta_{k,\nu}^i)} \tilde{Y}_l^m,$$

$\mathcal{J}_{-1}^i := \mathcal{I}_0^i$, $\mathcal{J}_{-1}' := \mathcal{I}_0'$, $\eta_{-1,\nu}^i := \zeta_{0,\nu}^i$, $\psi_{-1}^i(\cdot; \eta_{-1,\nu}^i) := \varphi_0^i(\cdot; \zeta_{0,\nu}^i)$ for $\nu \in \mathcal{J}_{-1}'$, and the scaling functions are given by

$$\varphi_k^i(\cdot; \zeta_{k,\nu}^i) := \sqrt{d_{k,\nu}} \sum_{(m,l) \in \mathcal{I}_k^i} \overline{\tilde{Y}_l^m(\zeta_{k,\nu}^i)} \tilde{Y}_l^m, \quad \nu \in \mathcal{I}_k'.$$

We also note that $\left\{ \varphi_k^i(\cdot; \zeta_{k,\nu}^i) \,:\, \nu \in \mathcal{I}_k' \right\}$ and $\left\{ \psi_k^i(\cdot; \eta_{k,\nu}^i) \,:\, \nu \in \mathcal{J}_k' \right\}$ are also normalized tight frames for V_k^i and W_k^i respectively, $k \geq 0$, $i = 1, 2$.

[2]Since the definition in (3.2.4) involves an infinite sum, for practical purposes, in our experiments, we truncate the summation at the 50th term as the magnitude of its summand drops below 10^{-10} for all the chosen orientations $\{\mathbf{v}_j^i\}_{i=1}^{M_j}$, $j = -1, 0, \dots, J$.

In this section, similar to the rationale behind the construction of spherical ridgelets, we design wavelet frame systems that are pre-images of the normalized tight wavelet frame systems $\Gamma(\Psi)$ in (3.3.1) under the FRT $\mathcal{R}$ and the Laplace-Beltrami operator Δ_b. These wavelet frame systems designed are then used to process HARDI signals for analytical QBI and the CSA version of analytical QBI respectively in subsection 3.3.1 and subsection 3.3.2. The results of section 2.5 also show that these wavelet systems form wavelet frames for a Sobolev space $H^s_{\text{sym}}(\mathbb{S}^2)$ for some appropriately chosen exponent s.

The rationale of the above setup is that by Remark 2.1.3, the functions $\varphi^i_k(\cdot;\zeta^i_{k,\nu})$, $\psi^i_k(\cdot;\eta^i_{k,\nu'})$ are relatively well localized at their respective points $\zeta^i_{k,\nu}$, $\eta^i_{k,\nu'}$, $\nu \in \mathcal{I}'_k$, $\nu' \in \mathcal{J}'_k$, $k \geq 0$, $i = 1, 2$. Thus we select the normalized tight wavelet frame system $\Gamma(\Psi)$ to represent the ODF, which is also well localized at only a few directions in each voxel. For instance, each of the two ODFs given in Figure 3.2.1 can be sparsely represented by (just one or two of) the scaling functions constructed in Figure 2.3.2. Then the above 'inverse' mappings are applied onto $\Gamma(\Psi)$ to obtain wavelet frame systems to represent the corresponding HARDI signals.

3.3.1. Wavelet frames-based QBI (WF-based QBI).

Before constructing wavelet frames for analytical QBI, recall from Corollary 3.1.2 that the FRT $\mathcal{R}$ has eigenfunctions $\tilde{Y}^m_l$, with corresponding eigenvalues $2\pi P_l(0)$, $|m| \leq l$, $l \in 2\mathbb{N}_0$, i.e.,

$$\mathcal{R}[\tilde{Y}^m_l] = 2\pi P_l(0)\tilde{Y}^m_l, \quad |m| \leq l,\, l \in 2\mathbb{N}_0, \tag{3.3.2}$$

where $P_l(0)$ is given explicitly earlier in (3.1.11). Suppose now that for $L \in 2\mathbb{N}_0$,

$$f^L = \sum_{l \in 2\mathbb{N}_0 \cap [0,L]} \sum_{|m| \leq l} \tilde{f}_{m,l}\tilde{Y}^m_l.$$

Thus one can easily see from (3.3.2) that

$$\mathcal{R}[f^L] = \sum_{l \in 2\mathbb{N}_0 \cap [0,L]} \sum_{|m| \leq l} 2\pi P_l(0)\, \tilde{f}_{m,l}\tilde{Y}^m_l.$$

When L is raised to infinity, it is then natural to seek the conditions on the modified SH coefficients $\{\tilde{f}_{m,l}\}_{m,l}$ in order for $f := \lim_{L\to\infty} \mathcal{R}[f^L]$ to lie in $L^2_{\text{sym}}(\mathbb{S}^2)$. Note that this requires

$$\sum_{l \in 2\mathbb{N}_0} \sum_{|m| \leq l} |2\pi P_l(0)|^2\, |\tilde{f}_{m,l}|^2 < \infty. \tag{3.3.3}$$

To check (3.3.3), it is thus important to know the asymptotic decay/growth rate of the sequence $\{P_l(0)\}_{l\in 2\mathbb{N}_0}$. This is answered by the following proposition (by letting $l = 2N$, $N \in \mathbb{N}_0$).

PROPOSITION 3.3.1. *The sequence $\{P_{2N}(0)\}_{N\in\mathbb{N}_0}$ decays asymptotically at the rate of $N^{-1/2}$, i.e., there exist positive constants C_1 and C_2 such that*

$$C_1 N^{-1/2} \leq |P_{2N}(0)| \leq C_2 N^{-1/2}, \quad N \in \mathbb{N}. \tag{3.3.4}$$

Proposition 3.3.1 can be used to establish mapping properties of the FRT $\mathcal{R}$ given in the following theorem.

THEOREM 3.3.2. *The FRT $\mathcal{R}$ is a linear mapping from $H^s_{\mathrm{sym}}(\mathbb{S}^2)$ to $H^{s+1/2}_{\mathrm{sym}}(\mathbb{S}^2)$, $s \geq 0$. Furthermore this mapping $\mathcal{R}$ is invertible with its inverse $\mathcal{R}^{-1}$ given by*

$$\mathcal{R}^{-1}[g] = \sum_{l\in 2\mathbb{N}_0} \sum_{|m|\leq l} \frac{1}{2\pi P_l(0)}\, \tilde{g}_{m,l} \tilde{Y}_l^m,$$

where $g \in H^{s+1/2}_{\mathrm{sym}}(\mathbb{S}^2)$, and $\{\tilde{g}_{m,l}\}_{m,l}$ are the modified spherical harmonic coefficients of g.

We focus here on the description of the wavelet construction method, so the proofs of Proposition 3.3.1 and Theorem 3.3.2 will be provided in the Appendix instead.

In this subsection, we obtain a wavelet frame system

$$\Gamma(\tilde{\Psi}) := \left\{ \tilde{\psi}_k^i(\cdot; \eta_{k,\nu}^i) \,:\, \nu \in \mathcal{J}_k', \, k \geq -1, \, i = 1, 2 \right\} \tag{3.3.5}$$

to represent HARDI signals by applying the inverse FRT $\mathcal{R}^{-1}$ onto the wavelet frame system

$$\Gamma(\Psi) = \left\{ \psi_k^i(\cdot; \eta_{k,\nu}^i) \,:\, \nu \in \mathcal{J}_k', \, k \geq -1, \, i = 1, 2 \right\}$$

defined in (2.3.8). More explicitly, for $i = 1, 2$, $k \geq -1$, $\nu \in \mathcal{J}_k'$,

$$\begin{aligned} \tilde{\psi}_k^i(\cdot; \eta_{k,\nu}^i) :&= \mathcal{R}^{-1}[\psi_k^i(\cdot; \eta_{k,\nu}^i)] \\ &= \sqrt{\hat{d}_{k,\nu}} \sum_{(l,m)\in\mathcal{J}_k^i} [2\pi P_l(0)]^{-1} \cdot \overline{\tilde{Y}_l^m(\eta_{k,\nu}^i)} \tilde{Y}_l^m. \end{aligned}$$

As it turns out, the system $\Gamma(\tilde{\Psi})$ defined in (3.3.5) is dual to

$$\Gamma(\Psi^\sharp) := \left\{ \psi_k^{i,\sharp}(\cdot; \eta_{k,\nu}^i) \,:\, \nu \in \mathcal{J}_k', \, k \geq -1, \, i = 1, 2 \right\} \tag{3.3.6}$$

where the functions in (3.3.6) are given by

$$\begin{aligned}\psi_k^{i,\sharp}(\cdot;\eta_{k,\nu}^i) &:= \mathcal{R}[\psi_k^i(\cdot;\eta_{k,\nu}^i)] \\ &= \sqrt{\hat{d}_{k,\nu}} \sum_{(l,m)\in\mathcal{J}_k^i} 2\pi P_l(0)\cdot\overline{\tilde{Y}_l^m(\eta_{k,\nu}^i)}\tilde{Y}_l^m,\end{aligned}$$

where $\nu \in \mathcal{J}_k'$, $k \geq -1$, $i = 1, 2$.

So $\Gamma(\Psi^\sharp)$ and $\Gamma(\tilde{\Psi})$ satisfy the framework in (2.4.6) and (2.4.7) with

$$p[m,l] = 2\pi P_l(0), \quad |m| \leq l,\, l \in 2\mathbb{N}_0.$$

According to Proposition 3.3.1, since the sequence $\{P_l(0)\}_{l\in 2\mathbb{N}_0}$ decays asymptotically at a rate of $l^{-1/2}$, $\{p[m,l]\}_{(m,l)\in\Lambda}$ satisfies (2.4.5) with exponent $s = 1/2$. Then $(\Gamma(\Psi^\sharp), \Gamma(\tilde{\Psi}))$ form a pair of dual wavelet frames for Sobolev spaces $(H_{\mathrm{sym}}^{1/2}(\mathbb{S}^2), H_{\mathrm{sym}}^{-1/2}(\mathbb{S}^2))$ by the results in section 2.5.

We remark that HARDI signals are typically assumed to be symmetric and smooth (e.g., HARDI test signals are usually derived from a multi-Gaussian mixture), and thus belong to the Sobolev space $H_{\mathrm{sym}}^{-1/2}(\mathbb{S}^2)$. So we may use $\Gamma(\tilde{\Psi})$ to represent HARDI signals. Consider the following approximation of a given HARDI signal

$$\mathbf{s}(\mathbf{w}) \approx \sum_{k=-1}^{K}\sum_{i=1}^{2}\sum_{\nu\in\mathcal{J}_k'} c_{k,\nu}^i\tilde{\psi}_k^i(\mathbf{w};\eta_{k,\nu}^i) =: \sum_{j=1}^{R} c_j\tilde{\psi}_j(\mathbf{w}),\ \mathbf{w}\in\mathbb{S}^2$$

where we have re-enumerated $\left\{\tilde{\psi}_k^i(\cdot;\eta_{k,\nu}^i) : \nu\in\mathcal{J}_k',\, k=-1,0,\ldots,K,\, i=1,2\right\}$ by another set $\left\{\tilde{\psi}_j : j=1,2,\ldots,R\right\}$ for simplicity of notation. Since $\Psi_1(\mathbf{u}) \approx \mathcal{R}[\mathbf{s}](\mathbf{u})$, the ODF Ψ_1 can be reconstructed analytically by

$$\Psi_1(\mathbf{u}) \approx \sum_{j=1}^{R} 2\pi P_{l_j}(0)\cdot c_j\tilde{\psi}_j(\mathbf{u}).$$

Similar to subsection 3.2.1, in practical situations, $\mathbf{s}$ is discretized with a sampling set of K gradient directions $\{g_k\}_{k=1}^K$. The discrete HARDI signal is then given by $f = [\mathbf{s}(g_k)]_{k=1}^K$ with the corresponding discrete representation matrix B_4 given by

$$B_4 = \begin{bmatrix} \tilde{\psi}_1(g_1) & \tilde{\psi}_2(g_1) & \cdots & \tilde{\psi}_R(g_1) \\ \tilde{\psi}_1(g_2) & \tilde{\psi}_2(g_2) & \cdots & \tilde{\psi}_R(g_2) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\psi}_1(g_K) & \tilde{\psi}_2(g_K) & \cdots & \tilde{\psi}_R(g_K) \end{bmatrix}.$$

Consider for sparse reconstruction at a single voxel,

$$\min_c \frac{1}{2}\|f - B_4 c\|_2^2 + \lambda\|c\|_1,$$

where $c = [c_j]_{j=1}^R$ is the coefficient vector. One may then use the FISTA in subsection 3.2.1 to solve the above ℓ_1-minimization model. After obtaining a satisfactory coefficient vector c from FISTA, according to [**65**], the discrete ODF $\Psi_{1,d}$ at a single voxel is then reconstructed using the formula

$$\Psi_{1,d} \approx B_4 P c,$$

where P is the corresponding matrix defined in (3.1.13). We name this ODF reconstruction method as the *wavelet frames-based QBI* (WF-based QBI).

3.3.2. WF-based CSA QBI.

In this subsection, to obtain a sharper reconstruction of the ODF, we design instead wavelet frames for the *CSA version of analytical QBI* with the ODF reconstruction formula given by

$$\Psi_2(\mathbf{u}) \approx \frac{1}{4\pi} + \frac{1}{16\pi^2}\mathcal{R}[\Delta_b \tilde{\mathbf{s}}](\mathbf{u}), \quad \mathbf{u} \in \mathbb{S}^2,$$

where $\tilde{\mathbf{s}}$ is modified from the HARDI signal $\mathbf{s}$ by

$$\tilde{\mathbf{s}}(\mathbf{u}) = \log(-\log \mathbf{s}(\mathbf{u})), \quad 0 < \mathbf{s}(\mathbf{u}) < 1, \quad \mathbf{u} \in \mathbb{S}^2,$$

$\mathcal{R}$ and Δ_b represent respectively the Funk-Radon transform and the spherical Laplace-Beltrami operator defined in (3.1.9) and (3.1.16).

Before constructing wavelet frames for the CSA version of Analytical QBI, recall from Corollary 3.1.2 that the FRT $\mathcal{R}$ and the Laplace-Beltrami operator Δ_b have eigenfunctions $\tilde{Y}_l^m$, with respective eigenvalues $2\pi P_l(0)$ and $-l(l+1)$, $|m| \le l$, $l \in 2\mathbb{N}_0$, i.e.,

$$\mathcal{R}[\tilde{Y}_l^m] = 2\pi P_l(0)\tilde{Y}_l^m, \quad |m| \le l,\ l \in 2\mathbb{N}_0, \tag{3.3.7}$$

$$\Delta_b[\tilde{Y}_l^m] = -l(l+1)\tilde{Y}_l^m, \quad |m| \le l,\ l \in 2\mathbb{N}_0, \tag{3.3.8}$$

where $P_l(0)$ is given explicitly earlier in (3.1.11). In addition, recall from (2.3.8) that the functions $\psi_k^i(\cdot;\eta_{k,\nu}^i) = \sqrt{\hat{d}_{k,\nu}}\sum_{(l,m)\in\mathcal{J}_k^i}\overline{\tilde{Y}_l^m(\eta_{k,\nu}^i)}\tilde{Y}_l^m$, $\nu \in \mathcal{J}_k'$, $k \ge -1$, $i = 1,2$, form a tight frame for $L^2_{\mathrm{sym}}(\mathbb{S}^2)$.

In this subsection, for $i = 1,2$, $k \ge -1$, $\nu \in \mathcal{J}_k'$, define

$$\Gamma_0(\Psi^\sharp) := \left\{\psi_k^{i,\sharp}(\cdot;\eta_{k,\nu}^i) \,:\, \nu \in \mathcal{J}_k',\, k \ge -1,\, i = 1,2\right\}, \tag{3.3.9}$$

where the functions in (3.3.9) are given by

$$\psi_k^{i,\sharp}(\cdot;\eta_{k,\nu}^i) := \sqrt{\hat{d}_{k,\nu}} \sum_{(l,m)\in\mathcal{J}_k^i} \delta_l^{-1}\cdot[2\pi P_l(0)]^{-1}\cdot\overline{\tilde{Y}_l^m(\eta_{k,\nu}^i)}\tilde{Y}_l^m, \tag{3.3.10}$$

and

$$\delta_l := \begin{cases} 1, & l=0,\\ -l(l+1), & l\in 2\mathbb{N}. \end{cases} \tag{3.3.11}$$

We remark that in view of (3.3.7) and (3.3.8), we may regard $\psi_k^{i,\sharp}(\cdot;\eta_{k,\nu}^i)$ in (3.3.10) as an pre-image of $\psi_k^i(\cdot;\eta_{k,\nu}^i)$ (defined in (2.3.8)) under the operators Δ_b and $\mathcal{R}$.

As it turns out, the system $\Gamma_0(\Psi^\sharp)$ is dual to the system $\Gamma_0(\tilde{\Psi})$, defined by

$$\Gamma_0(\tilde{\Psi}) := \left\{\tilde{\psi}_k^i(\cdot;\eta_{k,\nu}^i) \,:\, \nu\in\mathcal{J}_k', k\geq -1, i=1,2\right\} \tag{3.3.12}$$

where the functions in (3.3.12) are given by

$$\tilde{\psi}_k^i(\cdot;\eta_{k,\nu}^i) := \sqrt{\hat{d}_{k,\nu}} \sum_{(l,m)\in\mathcal{J}_k^i} \delta_l\cdot 2\pi P_l(0)\cdot\overline{\tilde{Y}_l^m(\eta_{k,\nu}^i)}\tilde{Y}_l^m,$$

where $\nu\in\mathcal{J}_k'$, $k\geq -1$, $i=1,2$. The systems $\Gamma_0(\Psi^\sharp)$ and $\Gamma_0(\tilde{\Psi})$ satisfy the framework in (2.4.6) and (2.4.7) with

$$p[m,l] = \delta_l^{-1}\cdot[2\pi P_l(0)]^{-1}, \quad |m|\leq l,\, l\in 2\mathbb{N}_0. \tag{3.3.13}$$

According to Proposition 3.3.1 and (3.3.11), $\{p[m,l]\}_{(m,l)\in\Lambda}$ defined in (3.3.13) satisfies (2.4.5) with exponent $s=3/2$. Then $(\Gamma_0(\Psi^\sharp),\Gamma_0(\tilde{\Psi}))$ form a pair of dual wavelet frames for Sobolev spaces $(H_{\mathrm{sym}}^{3/2}(\mathbb{S}^2),\, H_{\mathrm{sym}}^{-3/2}(\mathbb{S}^2))$ by the results in section 2.5.

So we may use $\Gamma_0(\Psi^\sharp)$ to represent HARDI signals, as they can be assumed to be smooth and lying in the Sobolev space $H_{\mathrm{sym}}^{3/2}(\mathbb{S}^2)$. Consider the following approximation of a given modified HARDI signal

$$\tilde{\mathbf{s}}(\mathbf{w}) \approx \sum_{k=-1}^{K}\sum_{i=1}^{2}\sum_{\nu\in\mathcal{J}_k'} c_{k,\nu}^i\psi_k^{i,\sharp}(\mathbf{w};\eta_{k,\nu}^i) =: \sum_{j=1}^{R} c_j\psi_j^\sharp(\mathbf{w}),\ \mathbf{w}\in\mathbb{S}^2,$$

where we have re-enumerated

$$\left\{\psi_k^{i,\sharp}(\cdot;\eta_{k,\nu}^i) \,:\, \nu\in\mathcal{J}_k', k=-1,0,\ldots,K, i=1,2\right\} \tag{3.3.14}$$

by another set $\left\{\psi_j^\sharp : j = 1, 2, \ldots, R\right\}$ for simplicity of notation. Then the ODF Ψ_2 can be reconstructed analytically by

$$\begin{aligned}\Psi_2(\mathbf{u}) &\approx \frac{1}{4\pi} + \frac{1}{16\pi^2}\mathcal{R}[\Delta_b \tilde{\mathbf{s}}](\mathbf{u}) \\ &= \frac{1}{4\pi} + \frac{1}{16\pi^2}\sum_{j=1}^{R} c_j \mathcal{R}[\Delta_b \psi_j^\sharp](\mathbf{u}), \\ &= \frac{1}{4\pi} + \frac{1}{16\pi^2}\sum_{j=1}^{R} c_j 2\pi P_{l_j}(0)[-l_j(l_j+1)]\psi_j^\sharp(\mathbf{u}).\end{aligned}$$

Similar to subsection 3.2.1, in practical situations, $\mathbf{s}$ is discretized with a sampling set of K gradient directions $\{g_k\}_{k=1}^K$. The discrete HARDI signal is then given by $f = [\mathbf{s}(g_k)]_{k=1}^K$ with the corresponding discrete representation matrix B_5 given by

$$B_5 = \begin{bmatrix} \psi_1^\sharp(g_1) & \psi_2^\sharp(g_1) & \cdots & \psi_R^\sharp(g_1) \\ \psi_1^\sharp(g_2) & \psi_2^\sharp(g_2) & \cdots & \psi_R^\sharp(g_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1^\sharp(g_K) & \psi_2^\sharp(g_K) & \cdots & \psi_R^\sharp(g_K) \end{bmatrix}. \tag{3.3.15}$$

For sparse reconstruction at a single voxel, consider the following model:

$$\min_c \frac{1}{2}\|f - B_5 c\|_F^2 + \lambda\|c\|_1,$$

where $c = [c_j]_{j=1}^R$ is the coefficient vector. One may then use FISTA in subsection 3.2.1 to solve this ℓ_1-minimization model. After obtaining a satisfactory coefficient vector c from FISTA, according to [**65**], the discrete ODF $\Psi_{2,d}$ at a single voxel is reconstructed by the formula

$$\Psi_{2,d} \approx \frac{1}{4\pi} + \frac{1}{16\pi^2} B_5 L P c,$$

where P and L are the matrices defined in (3.1.13) and (3.1.15) respectively. We name this ODF reconstruction method as the *wavelet-frame based constant solid angle Q-ball imaging* (WF-based CSA QBI).

3.4. HARDI Spatial Regularization and Numerical Results

In clinical HARDI ODF reconstruction, one is required to process HARDI signals in *multiple* voxels simultaneously. In most practical situations, the HARDI ODFs and signals in neighboring voxels are largely similar. Figure 3.4.1 shows a 2D image (slice) of HARDI ODFs, note that the image can be divided into several *spatial regions* where within each

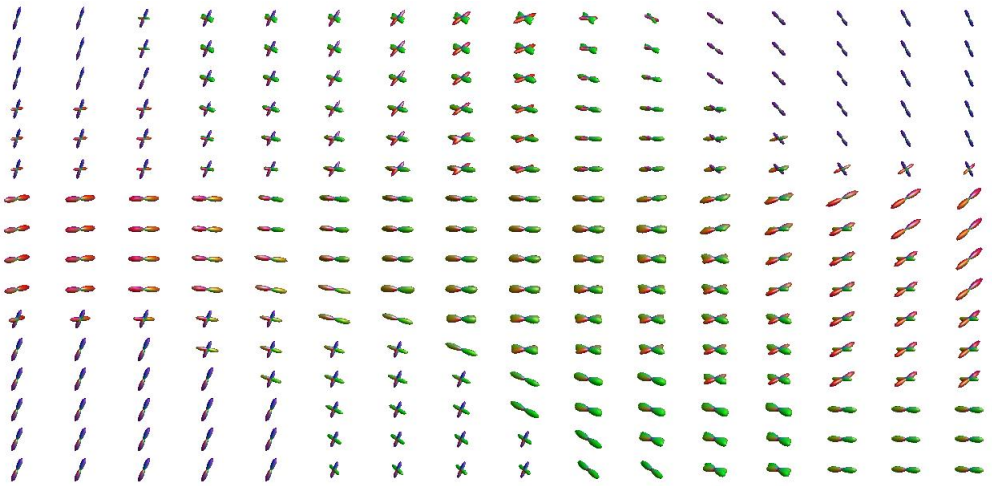

FIGURE 3.4.1. A 2D image (slice) of HARDI ODFs.

region, ODFs have similar diffusion directions. We can exploit this phenomena to improve HARDI reconstruction by modifying our earlier optimization models with HARDI *spatial regularization*.

To describe HARDI spatial regularization, in the following optimization models, we consider Ω a rectangular area of interest within which HARDI diffusion measurements are acquired, i.e., $\Omega := [0, L_x] \times [0, L_y] \subset \mathbb{R}^2$. We remark that the HARDI spatial regularization concepts in this section can be extended to the three-dimensional setting. Let Ω_d be a discrete subset of Ω, which represents the spatial locations at which the diffusion signal is measured, i.e.,

$$\Omega_d := \left\{ r = (x_i, y_j) \,|\, x_i = \frac{i}{N_x} L_x,\, y_j = \frac{j}{N_y} L_y,\, 1 \le i \le N_x,\, 1 \le j \le N_y \right\},$$

where $1 \le i \le N_x$, $1 \le j \le N_y$ are sampling indices in the direction of x, y coordinates respectively. Let K be the number of sampling gradients used for the acquisition of HARDI data, and the corresponding gradient orientations be denoted by $\{g_k\}_{k=1}^K$, where $g_k \in \mathbb{S}^2$. For each sampling gradient g_k, the MRI measurements yield a corresponding $N_x \times N_y$ image of HARDI signals $\mathbf{S}(g_k)$, where its (i,j)-th entry $[\mathbf{S}(g_k)]_{i,j}$ represents the signal intensity of the HARDI signal sampled at the gradient vector g_k at the corresponding voxel/image position (x_i, y_j), $i = 1, 2, \ldots, N_x, \quad j = 1, 2, \ldots, N_y$. In subsequent optimization models, we use a $K \times N_x \times N_y$ HARDI data real matrix F, which is obtained from $\{\mathbf{S}(g_k)\}_{k=1}^K$ by the following

vectorization procedure:

$$F = \begin{bmatrix} \mathbf{S}'(g_1) \\ \mathbf{S}'(g_2) \\ \vdots \\ \mathbf{S}'(g_K) \end{bmatrix}, \tag{3.4.1}$$

where for a fixed sampling gradient g_k, $\mathbf{S}'(g_k)$ denotes the $1 \times N_x \cdot N_y$ row vector formed by concatenating all the columns of the matrix $\tilde{\mathbf{S}}(g_k) := [\tilde{\mathbf{S}}_1(g_k)|\dots|\tilde{\mathbf{S}}_{N_y}(g_k)]$, $\tilde{\mathbf{S}}_j(g_k) \in \mathbb{R}^{N_x}$, $k = 1, 2, \dots K$, i.e.,

$$\mathbf{S}'(g_k) := (\mathrm{Vec}[\tilde{\mathbf{S}}(g_k)])^\top := [\tilde{\mathbf{S}}_1(g_k)^\top, \dots, \tilde{\mathbf{S}}_{N_y}(g_k)^\top],$$

where $\tilde{\mathbf{S}}(g_k)$ denotes a $N_x \times N_y$ modified HARDI signal real matrix defined by $\mathbf{S}(g_k)$ as

$$[\tilde{\mathbf{S}}(g_k)]_{i,j} := \log(-\log[\mathbf{S}(g_k)]_{i,,j}), \quad i = 1, 2, \dots, N_x, \quad j = 1, 2, \dots, N_y.$$

3.4.1. HARDI spatial regularization by wavelet frame filters.

As mentioned previously, CSA QBI generally reconstructs sharper HARDI ODFs. Thus, in this subsection, we will describe how the wavelet framelet filters in [**23**] can be utilized to impose spatial regularization on the optimization models in subsections 3.1.3, 3.2.2, 3.3.2, giving rise to the respective spatially regularized methods:

(a) Spatially regularized, SH-based CSA QBI;
(b) Spatially regularized, SR-based CSA QBI;
(c) Spatially regularized, WF-based CSA QBI.

We shall see from the numerical results in subsection 3.4.2 that spatial regularization greatly increases the effectiveness in denoising HARDI signals. We now describe in detail the spatially regularized optimization models for (a)-(c).

3.4.1.1. *SH-based CSA QBI with spatial regularization.*

We propose imposing spatial regularization by adding the term $\|\mathrm{diag}(\mu)WC^\top\|_1$ to a multi-voxel version of the optimization model in subsection 3.1.3, yielding

$$\min_{C \in \mathbb{R}^{R \times N_x \cdot N_y}} \frac{1}{2}\|F - BC\|_F^2 + \frac{\lambda}{2}\|LC\|_F^2 + \|\mathrm{diag}(\mu)WC^\top\|_1, \tag{3.4.2}$$

where F is the $K \times N_x \cdot N_y$ HARDI data real matrix in (3.4.1), B is the matrix given in (3.1.12), $C := [C(r,n)]_{r,n}$ is an $R \times N_x \cdot N_y$ matrix of SH coefficients, L is the $R \times R$ matrix given in (3.1.15), $\lambda \in \mathbb{R}_+$, $\mu \in \mathbb{R}_{\geq 0}^{4N_x \cdot N_y}$ are tuning parameters that respectively control the

degrees of the Laplace-Beltrami regularization and the spatial regularization by the 2D-tensor Haar tight framelet filter matrix W in [**23**]. More explicitly,

$$W = \begin{bmatrix} H_0^y \otimes H_0^x \\ H_0^y \otimes H_1^x \\ H_1^y \otimes H_0^x \\ H_1^y \otimes H_1^x \end{bmatrix} \in \mathbb{R}^{4N_x \cdot N_y \times N_x \cdot N_y},\ \mu := \mu' \begin{bmatrix} \mathbf{0}_{N_x \cdot N_y} \\ \mathbf{1}_{N_x \cdot N_y} \\ \mathbf{1}_{N_x \cdot N_y} \\ \mathbf{1}_{N_x \cdot N_y} \end{bmatrix} \in \mathbb{R}^{4N_x \cdot N_y},\ \mu' > 0, \tag{3.4.3}$$

where $\mathbf{0}_{N_x \cdot N_y}$ and $\mathbf{1}_{N_x \cdot N_y}$ are the $\mathbb{R}^{N_x \cdot N_y}$ column vectors of zeros and ones respectively, $\otimes$ denote the Kronecker product, H_0^x, H_1^x are $N_x \times N_x$ matrices and H_0^y, H_1^y are $N_y \times N_y$ matrices defined by the Haar framelet filters h_0, h_1 as follows:

$$H_i^x[l,k] := h_i[(l-k) \bmod N_x], \quad H_i^y[l,k] := h_i[(l-k) \bmod N_y], \quad i = 0, 1,$$

$$h_0 := \frac{1}{2}\begin{bmatrix} 1 & 1 \end{bmatrix}, \quad h_1 := \frac{1}{2}\begin{bmatrix} 1 & -1 \end{bmatrix}. \tag{3.4.4}$$

Note that the matrices H_0^x, H_1^x H_0^y, H_1^y are circulant and satisfy the tight frame condition

$$H_0^{x\top} H_0^x + H_1^{x\top} H_1^x = I, \quad H_0^{y\top} H_0^y + H_1^{y\top} H_1^y = I,$$

which leads to W satisfying

$$W^\top W = I.$$

It was proved recently in [**23**] that, in the discrete model, the last three blocks in W (defined by h_1) correspond to (local) finite difference operators in the x, y and diagonal directions respectively, which may also be regarded as certain discretizations of differential operators in the continuous variational models. Indeed, as shown in [**23**], minimizing the term $\|\mathrm{diag}(\mu)WC^\top\|_1$ in the above discrete model is analogous to modelling sharp jumps (whilst smoothing small variations) of a function in the corresponding variational models. Thus, we have included the term $\|\mathrm{diag}(\mu)WC^\top\|_1$ in our optimization model to model sharp changes in ODF diffusion directions from one voxel to another in our HARDI data set, see for example Figure 3.4.1.

Remark 3.4.1. The Haar tight framelet filters (3.4.4) is a special case of B-spline based tight framelet filters constructed in [**75**] and discussed in [**23**]. In principle, we may also define the matrix W using longer B-spline based tight framelet filters instead. However, in this thesis, we have chosen the Haar tight framelet filters (3.4.4) to impose the spatial regularization because our numerical experiences reveal that the Haar tight framelet filters achieve the best balance between denoising performances and computational speed.

We now describe how (3.4.2) can be solved using the split Bregman method [**49**], which has been extremely successful in solving ℓ_1-norm regularized problems in compressed sensing [**26**] and image processing [**24**], etc. We start first by setting the auxiliary variable $Q = WC^\top$ to tackle the non-smooth ℓ_1-norm in (3.4.2), which yields the following equivalent optimization problem:

$$\min_{C,Q} \frac{1}{2}\|F - BC\|_F^2 + \frac{\lambda}{2}\|LC\|_F^2 + \|\mathrm{diag}(\mu)Q\|_1, \quad \text{s.t. } Q = WC^\top. \tag{3.4.5}$$

The corresponding augmented Lagrangian of (3.4.5) is given by

$$L_\rho(C, Q; \Lambda) := \frac{1}{2}\|F - BC\|_F^2 + \frac{\lambda}{2}\|LC\|_F^2 + \|\mathrm{diag}(\mu)Q\|_1 - \rho\langle\Lambda, Q - WC^\top\rangle + \frac{\rho}{2}\|Q - WC^\top\|_F^2$$

where Λ denotes the Lagrange multiplier, ρ is a given positive parameter, and the inner product $\langle\cdot,\cdot\rangle$ is defined by

$$\langle X, Y\rangle := \mathrm{Tr}(X^\top Y),$$

for any given matrices X and Y of the same dimension.

Using the initialization $C^0 = [B^\top B + \lambda L^\top L]^{-1}B^\top F$, $Q^0 = W(C^0)^\top$, $\Lambda = 0$, for a fixed (C^k, Q^k, Λ^k), we update $(C^{k+1}, Q^{k+1}, \Lambda^{k+1})$ by a Gauss-Seidel scheme on the augmented Lagrangian in the following manner:

Setting X_c^k to be the concatenation of $\mathrm{Vec}(C^k)$, $\mathrm{Vec}(Q^k)$ and $\mathrm{Vec}(\Lambda^k)$,

$$\text{while } \|X_c^{k+1} - X_c^k\|_\infty > \text{`tol'},$$

for $k = 0, 1, \ldots,$

(1) $C^{k+1} = \mathrm{argmin}_C\, L_\rho(C, Q^k; \Lambda^k)$;
(2) $Q^{k+1} = \mathrm{argmin}_Q\, L_\rho(C^{k+1}, Q; \Lambda^k)$;
(3) $\Lambda^{k+1} = \Lambda^k - (Q^{k+1} - W(C^{k+1})^\top)$.

The first two steps amounts to solving the following two optimization subproblems:

(1) $C^{k+1} = \mathrm{argmin}_C\, \frac{1}{2}\|F - BC\|_F^2 + \frac{\lambda}{2}\|LC\|_F^2 + \frac{\rho}{2}\|CW^\top - (Q^k - \Lambda^k)^\top\|_F^2$;
(2) $Q^{k+1} = \mathrm{argmin}_Q\, \|\mathrm{diag}(\mu)Q\|_1 + \frac{\rho}{2}\|Q - (W(C^{k+1})^\top + \Lambda^k)\|_F^2$;
(3) $\Lambda^{k+1} = \Lambda^k - (Q^{k+1} - W(C^{k+1})^\top)$.

Using the fact that $W^\top W = I$, one can show that we have the following analytic solutions to the above subproblems:

(1) $C^{k+1} = [B^\top B + \lambda L^\top L + \rho I]^{-1}[B^\top F + \rho(Q^k - \Lambda^k)^\top W]$,
(2) $Q^{k+1} = T^1_{\mu/\rho}[W(C^{k+1})^\top + \Lambda^k]$,
(3) $\Lambda^{k+1} = \Lambda^k - (Q^{k+1} - W(C^{k+1})^\top)$,

where we recall that $T^1_{\mu/\rho}$ is the soft-thresholding operator defined in (1.1.12).

We note that (3.4.5) is a convex optimization problem which adheres to the framework of optimization problems in [**26**], where a proof of the convergence of the split Bregman method was established.

3.4.1.2. *SR and WF-based CSA QBI with spatial regularization.*

We shall now deal with the spatially regularized SR-based QBI and WF-based CSA QBI at the same time because upon adding the term $\|\mathrm{diag}(\mu)WC^\top\|_1$ to a multi-voxel version of the optimization models in subsections 3.2.2 and 3.3.2, they yield the same optimization model given as follows:

$$\min_C \frac{1}{2}\|F - BC\|_F^2 + \lambda\|C\|_1 + \|\mathrm{diag}(\mu)WC^\top\|_1, \tag{3.4.6}$$

where the matrix C is the corresponding coefficient matrix and

$$B := B_3 \text{ or } B := B_5$$

in (3.2.6) and (3.3.15) respectively for SR-based CSA QBI and WF-based CSA QBI.

We start first by setting the auxiliary variables $Q = WC^\top$ and $P = C$ to tackle the two non-smooth ℓ_1-norms in (3.4.6), yielding the following equivalent optimization problem:

$$\min_{C,P,Q} \frac{1}{2}\|F - BC\|_F^2 + \lambda\|P\|_1 + \|\mathrm{diag}(\mu)Q\|_1, \quad \text{s.t. } Q = WC^\top \text{and } P = C. \tag{3.4.7}$$

Denoting the Lagrange multipliers as Λ_1, Λ_2 and ρ_1, ρ_2 as some given positive parameters, the corresponding augmented Lagrangian of (3.4.7) is then given by

$$\begin{aligned} L_{\rho_1,\rho_2}(C,Q,P;\Lambda_1,\Lambda_2) :=& \frac{1}{2}\|F - BC\|_F^2 + \lambda\|P\|_1 + \mu\|Q\|_1 \\ &- \rho_1\langle\Lambda_1, Q - WC^\top\rangle + \frac{\rho_1}{2}\|Q - WC^\top\|_F^2 \\ &- \rho_2\langle\Lambda_2, P - C\rangle + \frac{\rho_2}{2}\|P - C\|_F^2. \end{aligned}$$

We use the initialization

$$C^0 = [B^\top B + \lambda I]^{-1}B^\top F, Q^0 = W(C^0)^\top, P^0 = C^0, \Lambda_1,\ \Lambda_2 = 0.$$

Then, for a fixed $(C^k, Q^k, P^k, \Lambda_1^k, \Lambda_2^k)$, we update $(C^{k+1}, Q^{k+1}, P^{k+1}, \Lambda_1^{k+1}, \Lambda_2^{k+1})$ by a Gauss-Seidel scheme on the above augmented Lagrangian as follows: Set X_c^k to be the concatenation of $\mathrm{Vec}(C^k)$, $\mathrm{Vec}(Q^k)$, $\mathrm{Vec}(P^k)$, $\mathrm{Vec}(\Lambda_1^k)$ and $\mathrm{Vec}(\Lambda_2^k)$, while $\|X_c^{k+1} - X_c^k\|_\infty >$ 'tol', perform the following till convergence: for $k = 0, 1, \ldots,$

(1) $C^{k+1} = \arg\min_C\ L_{\rho_1,\rho_2}(C, Q^k, P^k; \Lambda_1^k, \Lambda_2^k)$;

(2) $Q^{k+1} = \arg\min_Q\ L_{\rho_1,\rho_2}(C^{k+1}, Q, P^k; \Lambda_1^k, \Lambda_2^k)$;

(3) $P^{k+1} = \arg\min_P L_{\rho_1,\rho_2}(C^{k+1}, Q^{k+1}, P; \Lambda_1^k, \Lambda_2^k)$;
(4) $\Lambda_1^{k+1} = \Lambda_1^k - (Q^{k+1} - W(C^{k+1})^\top)$;
(5) $\Lambda_2^{k+1} = \Lambda_2^k - (P^{k+1} - C^{k+1})$.

Using $W^\top W = I$ again, one can similarly show that we have the following analytic solutions to the above subproblems:

(1) $C^{k+1} = [B^\top B + (\rho_1 + \rho_2)I]^{-1}[B^\top F + \rho_1(Q^k - \Lambda_1^k)^\top W + \rho_2(P^k - \Lambda_2)]$;
(2) $Q^{k+1} = T^1_{\mu/\rho_1}[W(C^{k+1})^\top + \Lambda_1^k]$;
(3) $P^{k+1} = T^1_{\lambda/\rho_2}[C^{k+1} + \Lambda_2^k]$.

Note that (3.4.6) is a convex optimization problem that also adheres to the framework of optimization problems in [**26**].

REMARK 3.4.2. In [**71**], [**65**], a total variational (TV) term $\mu\|C\|_{TV}$ is used (instead of $\|\mathrm{diag}(\mu)WC^\top\|_1$) to impose spatial regularization on SH-based QBI and SR-based QBI. For instance, the optimization model used in [**71**] is given by

$$\min_{C\in\mathbb{R}^{R\times N_x\cdot N_y}} \frac{1}{2}\|F - BC\|_F^2 + \frac{\lambda}{2}\|LC\|_F^2 + \mu\|C\|_{TV}, \tag{3.4.8}$$

where B is the $K \times R$ discrete representation matrix for the SH defined in (3.1.12), $C := [C(r,n)]_{r,n}$ is an $R \times N_x \cdot N_y$ matrix of SH coefficients,

$$\|C\|_{TV} := \sum_{r=1}^{R} \sum_{(x_i,y_j)\in\Omega_d} \left\{ \sum_{(x_{i'},y_{j'})\in\mathcal{N}(x_i,y_j)} |\mathbf{C}_r(i,j) - \mathbf{C}_r(i',j')|^2 \right\}^{1/2},$$

$$\mathbf{C}_r := \begin{bmatrix} C(r,1) & C(r,N_x+1) & \dots & C(r,N_x\cdot(N_y-1)+1) \\ C(r,2) & C(r,N_x+2) & \dots & C(r,N_x\cdot(N_y-1)+2) \\ \vdots & \vdots & \ddots & \vdots \\ C(r,N_x) & C(r,2N_x) & \dots & C(r,N_x\cdot N_y) \end{bmatrix} \in \mathbb{R}^{N_x\times N_y},$$

where $r = 1, 2, \dots, R$, $\mathcal{N}(x_i, y_j)$ represents the set of vertices in the regular two-dimensional lattice Ω_d that are connected to the vertex (x_i, y_j), L is the matrix defined in (3.1.15), and μ is a pre-defined positive parameter that controls the degree of spatial regularization in (3.4.8).

It was shown in [**23**, Remark 3.2] that the wavelet frame ℓ_1-regularization term $\|\mathrm{diag}(\mu)WC^\top\|_1$ yields better spatial regularization than the TV term. This is because the TV term accounts for only spatial differences/similarities along the vertical and horizontal directions, whereas the wavelet frame ℓ_1-regularization term accounts for additional spatial differences/similarities along diagonal directions. We also remark that our numerical experiments give superior results when the ℓ_1-norm is used instead of the $\ell_{1,2}$-norm $\|\cdot\|_{1,2}$, the discretized version of the

TV norm, which was described in [**23**], possibly due to the anisotropic nature of the given HARDI data. Therefore, in this thesis, we just focus on optimization models that are based on the wavelet frame ℓ_1-regularization term $\|\mathrm{diag}(\mu)WC^\top\|_1$.

3.4.2. Computational results.

In this subsection, under different signal-to-noise (SNR) settings, we shall perform HARDI numerical simulations for six methods, namely:

M1a: SH-based CSA QBI,
M2a: SR-based CSA QBI,
M3a: WF-based CSA QBI,
M1b: Spatially regularized, SH-based CSA QBI,
M2b: Spatially regularized, SR-based CSA QBI,
M3b: Spatially regularized, WF-based CSA QBI,

and compare their performances with evaluation criteria (to be introduced below). As we shall see, we can make two main conclusions:

- spatially regularized methods are far superior than their unregularized versions in terms of HARDI denoising and ODF reconstruction, especially under low SNR settings,
- spatially regularized, WF-based CSA QBI achieves the best balance between the quality of ODF reconstruction and computational time, especially under low SNR settings.

3.4.2.1. *Generation of synthetic data.*

We now describe how the HARDI synthetic data used in this thesis is generated. This synthetic data is actually obtained from the challenge website[3] of the *diffusion MRI reconstruction challenge* 2012. This challenge was organized in the context of the 9th IEEE International Symposium on Biomedical Imaging (ISBI) conference, which was held in Barcelona (Spain) in May 2012. The objective of this challenge was to compare different reconstruction methods for reconstructing the HARDI intra-voxel fiber structure, i.e., the number and orientation of the fiber populations present in each voxel, from d-MRI acquisitions, using the same ground-truth data and under controlled conditions. The interested reader may also refer to [**35**] for more details.

[3]http://hardi.epfl.ch/static/events/2012_ISBI/

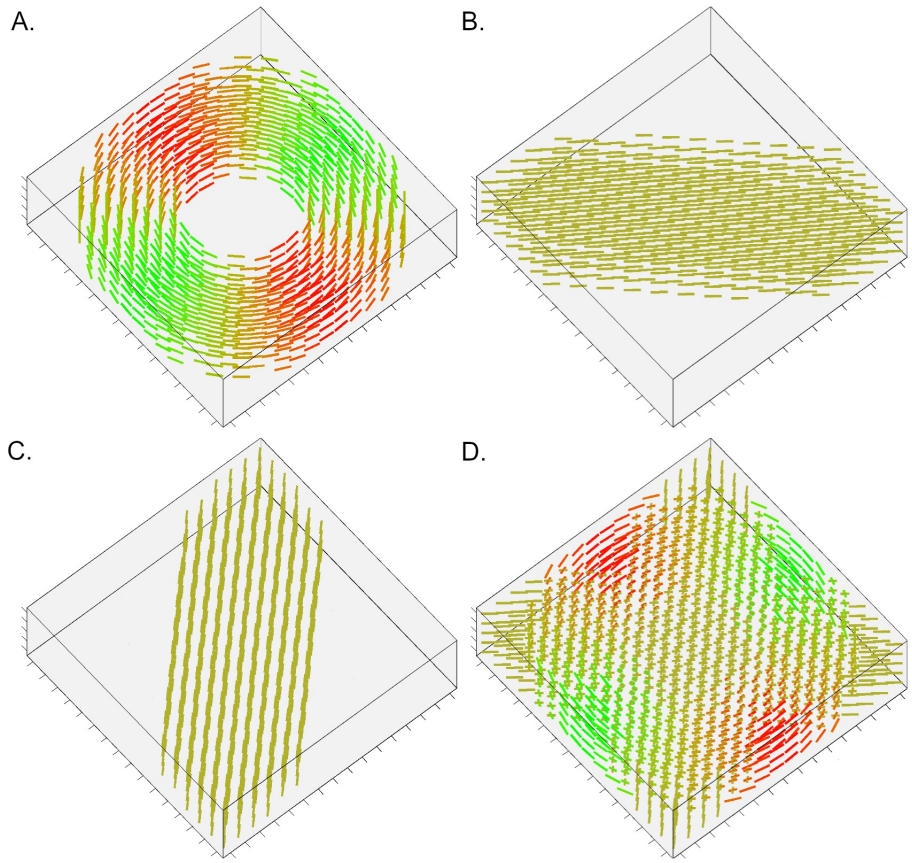

FIGURE 3.4.2. Structured Field Training Dataset: $16 \times 16 \times 5$ voxel set Ω_d. The dataset simulates a realistic 3D configuration of tracts: it comprises 3 different fiber bundles (Subplots A-C), which give rise to non-planar configurations of bending, crossing, kissing tracts. In each voxel, the directions are color-coded based on their orientation (x-axis, y-axis, z-axis). Subplot D shows the structured field training dataset.

The two HARDI data sets we have used from the challenge website are the *structured field training data* and the *structured field testing data* as shown in Figures 3.4.2 and 3.4.3. Each of the structured field datasets consists of a $16 \times 16 \times 5$ volume whose spatial index set $\Omega_d := \{\mathbf{r} = (x, y, z) \in \mathbb{Z}^3 \cap [0, N_x) \times [0, N_y) \times [0, N_z)\}$ (with N_x, $N_y = 16$, $N_z = 5$) simulates a configuration of neural fiber tracts. The training dataset and testing dataset respectively are made up of three and five different fiber bundles as seen in Figures 3.4.2 and 3.4.3.

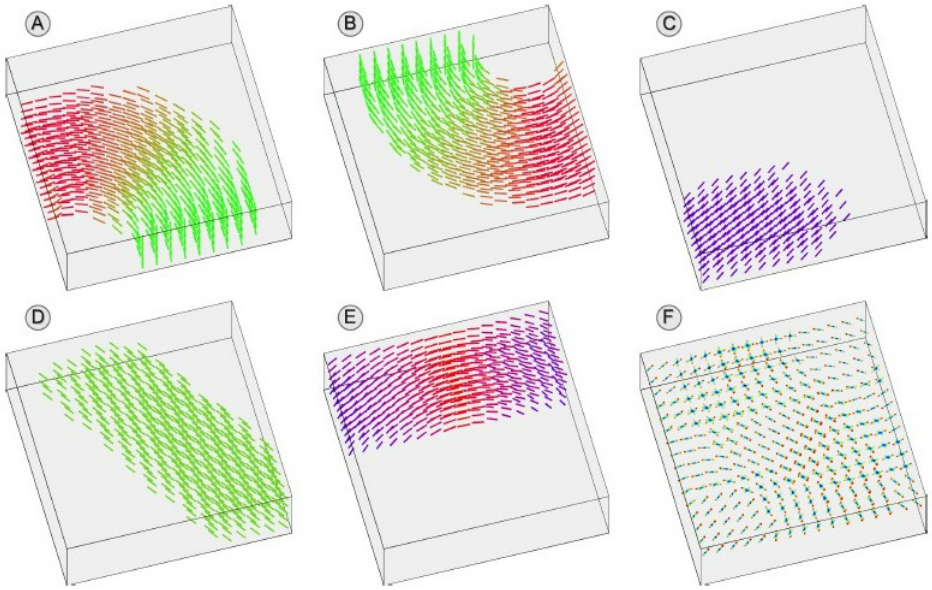

FIGURE 3.4.3. Structured Field Testing Dataset: $16 \times 16 \times 5$ voxel set Ω_d. The dataset simulates a realistic 3D configuration of tracts: it comprises 5 different fiber bundles (Subplots A-E). In each voxel, the directions are color-coded based on their orientation (x-axis, y-axis, z-axis). Subplot F shows a representative slice with the ODF orientations.

REMARK 3.4.3. The HARDI training data is merely a dataset released to the participants for familiarization purposes before they implement their methods on the testing data in the HARDI challenge. The reader should note that the purposes of the training and test datasets in the HARDI challenge are markedly different from those used in machine learning.

In each voxel of the $16 \times 16 \times 5$ volume of the structured field training data and the structured field testing data, the dMRI signal intensity in a voxel with M fiber populations is simulated by using a classical Gaussian mixture model as recalled in (3.1.7):

$$s(\mathbf{g}) = \sum_{i=1}^{M} p_i \exp(-b\mathbf{g}^\top D_i \mathbf{g}), \text{ s.t. } \sum_{i=1}^{M} p_i = 1,\ p_i \geq 0,\ i = 1, 2, \ldots, M, \quad \mathbf{g} \in \mathbb{S}^2,$$

with the b-value satisfying $b \geq 2000$, its corresponding ODF Ψ_2 in the analytic form (shown in the derivation of (3.1.20))

$$\Psi_2(\mathbf{u}) = \sum_{i=1}^{M} p_i \sqrt{\frac{1}{64\pi^2 b^2 |D_i|}} (\mathbf{u}^\top D_i^{-1} \mathbf{u})^{-3/2}, \quad \mathbf{u} \in \mathbb{S}^2.$$

Here, for $i = 1, 2, \dots M$, $|D_i|$ is the determinant of the s.p.d. diffusion tensor D_i, which is a symmetric positive definite matrix and so can be written in the form:

$$D_i = R_i^\top \mathrm{diag}(\lambda_1^i, \lambda_2^i, \lambda_3^i) R_i, \quad R_i^\top R_i = I, \quad i = 1, 2, \dots, M,$$

where $R_i := [e_1^i | e_2^i | e_3^i]$ is the rotation matrix rotating the main axis of D_i to the direction of a given fiber population, with eigenvalues $\lambda_1^i, \lambda_2^i, \lambda_3^i$ ordered in the manner $\lambda_1^i > \lambda_2^i \geq \lambda_3^i > 0$. The diagonal elements of D_i are the diffusivities along the main axis e_1^i of the fiber (λ_1^i) and on the orthogonal plane spanned by $\{e_2^i, e_3^i\}$. In the HARDI challenge, $\lambda_1^i \in [1, 2] \times 10^{-3}\mathrm{mm}^2/\mathrm{s}$ and λ_2^i, $\lambda_3^i \in [0.1, 0.6] \times 10^{-3}\mathrm{mm}^2/s$, $i = 1, 2, \dots, M$, are generated from diffusivities in the human brain.

Adhering to the settings of the HARDI challenge [**35**], we also corrupt the HARDI signal at each voxel with Rician noise in the following manner:

$$s_{\mathrm{noisy}} = \sqrt{(s + \eta_1)^2 + \eta_2^2}, \quad \eta_1, \eta_2 \sim N(0, \sigma^2),$$

where the signal-to-noise ratio (SNR) on the image is given and σ is set to be $\frac{1}{\mathrm{SNR}}$ to control the level of the noise. Our numerical simulations with both datasets will deal with the reconstruction of the HARDI ODF under four different noise levels, i.e., SNR=5, 10, 20, 40.

3.4.2.2. *Evaluation criteria.*

In this thesis, we have adopted the following evaluation criteria for the comparison of numerical performances.

(1) Normalized mean square error (NMSE): The accuracy in the estimation of the ODF, computed as the normalized mean squared error between the estimated ODF, Ψ_{est}, and the one analytically computed from the model used for the simulations, Ψ_{true}, is given by

$$\mathrm{NMSE} := \frac{1}{N_x N_y N_z} \sum_{\mathbf{r} \in \Omega_d} \frac{\|\Psi_{\mathrm{est}}(\mathbf{r}) - \Psi_{\mathrm{true}}(\mathbf{r})\|_2^2}{\|\Psi_{\mathrm{true}}(\mathbf{r})\|_2^2}.$$

(2) Probability of detection (Pd): The probability of false fiber detection, expressed by means of

$$P_d := \frac{1}{N_x N_y N_z} \sum_{\mathbf{r} \in \Omega_d} \frac{|M_{\mathrm{est}}(\mathbf{r}) - M_{\mathrm{true}}(\mathbf{r})|}{|M_{\mathrm{true}}(\mathbf{r})|} \times 100\%,$$

where M_{true} and M_{est} are, respectively, the real and estimated number of fiber compartments inside the voxel.

(3) Angular Deviation (AD): Angular precision of the estimated fiber compartments is assessed by means of the angular error (in degrees) between the estimated fiber direction and the true one inside the voxel given by

$$\bar{\theta} = \frac{180}{\pi} \cos^{-1}(|\mathbf{d}_{\text{est}} \cdot \mathbf{d}_{\text{true}}|),$$

where $\mathbf{d}_{\text{est}}$ is an estimated reconstructed fiber direction and $\mathbf{d}_{\text{true}}$ is a true direction inside the voxel. The final value will be the average of the angular errors computed for all the true fiber compartments.

3.4.2.3. *Numerical experiments.*

In all experiments, we have selected the number of sampling gradients K to be 32 in each voxel and fixed the b-value as 3000. The tolerance level 'tol' for both FISTA and the split Bregman methods described previously is set as 10^{-6}, with the maximum number of iterations for FISTA and the split Bregman methods set as 100 and 15 respectively.

For SH-based methods (M1a, M1b), as recommended in [**64**, **71**], we set the maximum level $l_{\max} = 8$ for the finite collection of modified SH $\left\{\tilde{Y}_l^m\right\}_{l=1,|m|\leq l}^{l_{\max}}$ and thus the number of SH functions/coefficients $R = 45$.

For SR-based methods (M2a, M2b), as recommended in [**65**], we set the maximum level $J = 1$ in the discrete collection of SR $U_d^\dagger$ defined in (3.2.5), so that the resolution levels at levels $j = -1, 0, 1$ have $M_{-1} = 16$, $M_0 = 49$, $M_1 = 169$ spherical ridgelets respectively, yielding a total number of SR functions/coefficients $R = 234$.

For WF-based methods (M3a, M3b), for best effect, the resolution level of the wavelet frames defined in (3.3.14) and (3.3.10) is set at $K = -1$, with $I_{-1} = \{(m, l) \in \mathbb{Z} \times 2\mathbb{N}_0 : l \leq 4\}$ and $I_{-1} = \{(m, l) \in \mathbb{Z} \times 2\mathbb{N}_0 : l \leq 6\}$ in experiments where SNR levels are 5 and in the range 10-40 respectively.

After optimizing all other parameters (such as λ, μ), we present the computational performances (averaged over 50 trials) of methods M1a-M3a, M1b-M3b (spatially regularized) on the HARDI training data in terms of NMSE, Pd, and AD values are shown in Tables 3.4.1, 3.4.2, 3.4.3 respectively. Similarly, computational performances of methods M1a-M3a, M1b-M3b (spatially regularized) on the HARDI testing data in terms of NMSE, Pd, and AD values shown in Tables 3.4.4, 3.4.5, 3.4.6 respectively. For each SNR level in these tables, the best performance is indicated in **bold**.

SNR	M1a	M2a	M3a	M1b	M2b	M3b
40	0.0432	0.0549	0.0417	**0.0330**	0.0414	0.0357
20	0.0781	0.0756	0.0670	**0.0433**	0.0564	0.0473
10	0.1930	0.1300	0.1210	**0.0728**	0.0809	0.0773
5	0.2690	0.2110	0.2040	0.1450	0.1390	**0.1160**

TABLE 3.4.1. NMSE values of methods M1a-M3a, M1b-M3b for HARDI training data under SNR levels 5, 10, 20 and 40.

SNR	M1a	M2a	M3a	M1b	M2b	M3b
40	9.98	13.4	8.41	**3.77**	4.57	4.52
20	18.9	16.2	13.1	6.21	12.9	**5.55**
10	24.9	26.4	32.9	21.8	17.3	**13.4**
5	39.5	34.0	38.8	39.1	**27.9**	28.2

TABLE 3.4.2. False detection rates (Pd) of methods M1a-M3a, M1b-M3b for HARDI training data under SNR levels 5, 10, 20 and 40.

SNR	M1a	M2a	M3a	M1b	M2b	M3b
40	5.52	5.31	5.49	**4.68**	5.03	5.25
20	7.48	7.76	6.82	**5.32**	5.91	5.81
10	12.0	13.2	11.9	**6.62**	8.55	6.81
5	20.6	24.6	22.1	10.9	13.4	**10.6**

TABLE 3.4.3. Angular deviation (AD) values of methods M1a-M3a, M1b-M3b for HARDI training data under SNR levels 5, 10, 20 and 40.

SNR	M1a	M2a	M3a	M1b	M2b	M3b
40	0.0649	0.0731	0.0560	0.0491	0.0505	**0.0396**
20	0.1083	0.1008	0.0844	0.0638	0.0645	**0.0518**
10	0.2243	0.1611	0.1641	0.1286	0.0891	**0.0753**
5	0.3739	0.2996	0.2738	0.2729	0.1517	**0.1436**

TABLE 3.4.4. NMSE values of methods M1a-M3a, M1b-M3b for HARDI testing data under SNR levels 5, 10, 20, 40.

SNR	M1a	M2a	M3a	M1b	M2b	M3b
40	19.2	17.5	18.0	**17.4**	20.8	18.6
20	28.6	21.9	23.8	**20.2**	22.1	21.5
10	44.1	36.8	37.7	36.0	27.0	**25.9**
5	46.2	36.8	39.1	35.9	35.5	**34.5**

TABLE 3.4.5. False detection rate (Pd) of methods M1a-M3a, M1b-M3b for HARDI testing data under SNR levels 5, 10, 20, 40.

SNR	M1a	M2a	M3a	M1b	M2b	M3b
40	5.85	6.01	5.64	**5.24**	6.31	5.45
20	7.69	7.95	7.18	**5.85**	6.92	6.20
10	11.58	12.83	12.68	8.43	8.18	**7.24**
5	19.28	20.39	18.93	11.02	10.59	**9.72**

TABLE 3.4.6. Angular deviation (AD) values of methods M1a-M3a, M1b-M3b for HARDI testing data under SNR levels 5, 10, 20, 40.

3.4.2.4. *Conclusion.*

From Tables 3.4.1 to 3.4.6, we can infer the following points:

- Spatially regularized methods (M1b-M3b) generally have superior denoising performances over their respective non-spatially regularized versions (M1a-M3a) in all three criteria, especially in terms of NMSE values (which is almost halved under low SNR levels).

Data Type	M1b	M2b	M3b
Training Data	0.346	2.19	0.714
Testing Data	0.319	2.13	0.678

TABLE 3.4.7. Average computational time (seconds) for the spatially regularized methods M1b-M3b on HARDI training data and testing data.

- Under medium to high SNR levels (20-40), the spatially regularized SH-based method (M1b) generally performs rather well in all three criteria with a relatively small number of sampling gradients, despite that SH generally do not yield sparse representations of HARDI signals.
- Under low SNR levels (5-10), spatially regularized WF and SR-based methods (M2b-M3b) generally outperform spatially regularized SH-based method (M1b) in all three criteria.
- The spatially regularized WF-based method (M3b) generally performs slightly better than the spatially regularized SR-based method (M2b) in all three criteria.
- As seen in Table 3.4.7, a comparison of average computational speeds of the spatially regularized methods reveals that M1b is roughly twice as fast as M3b, which is almost thrice as fast as M2b.

Taking into account the above points, we may conclude that

(1) Under medium to high SNR levels (20-40), the spatially regularized SH-based method (M1b) is arguably the most cost-effective method, in terms of balancing trade-offs between denoising effectiveness and computational speed.
(2) Under low SNR levels (5-10), the most cost-effective method is the spatially regularized WF-based method (M3b).

REMARK 3.4.4. (Possible Future Research Work/Directions) In [**39**], Descoteaux et al. developed a new deconvolution sharpening transformation from the diffusion ODF to the fiber ODF, where it was shown to improve angular resolution and fiber detection of QBI and thus greatly enhancing tractography results. Under the assumption that the HARDI signal is Gaussian, the (convolution) diffusion ODF kernel was modelled explicitly based on a fixed function with relatively gentle decay, see [**39**, equation (14)]. However, the HARDI signal may not be Gaussian in practice. Thus for future work, we can consider estimating the (convolution) diffusion ODF kernel adaptively through an ℓ_1-regularized optimization

model, which seeks sparse representations of the HARDI signal and the ODF kernel with appropriately chosen systems (possibly wavelet frames and Legendre polynomials).

CHAPTER 4

Adaptive HARDI Denoising by Optimization on Stiefel Manifolds

The HARDI denoising performances may be improved through adaptive spatial regularization, which can be modelled by optimization on Stiefel manifolds, i.e., optimization problems with orthogonality constraints. This chapter is organized in the following manner. In section 4.1, we propose the proximal alternating minimized augmented Lagrangian (PAMAL) method to achieve adaptive spatial regularization on HARDI. The PAMAL method consists of Algorithms I and II, where Algorithm I uses Algorithm II to run a small number of inner iterations to solve each outer iteration for the first step of Algorithm I. Empirical results illustrate that the PAMAL method yields better HARDI denoising performances.

Section 4.2 provide some preliminaries and notations for non-smooth analysis to facilitate the convergence analysis in subsequent sections. Then, we show in section 4.3 that Algorithms I and II are well defined. Finally, in section 4.4, we provide the convergence analysis of Algorithms I and II for the HARDI adaptive spatial regularization problem.

4.1. Adaptive Spatial Regularization for HARDI Denoising

In this section, we illustrate that one can improve HARDI denoising results through imposing spatial regularization adaptively. We shall see that adaptivity can be imposed by optimizing on the *Stiefel manifold* $\mathcal{S}$ where

$$\mathcal{S} := \left\{ P \in \mathbb{R}^{p\times q} \,:\, P^\top P = I_q \right\}. \tag{4.1.1}$$

More explicitly, we consider the following modified optimization model for spatially regularized, SH-based QBI (see Remark 4.1.1)

$$\min_{C,W} \frac{1}{2}\|BC - F\|_F^2 + \frac{\lambda}{2}\|LC\|_F^2 + \mu\|WC^\top\|_1 \quad \text{s.t. } W^\top W = I, \tag{4.1.2}$$

where μ is a pre-defined positive parameter and we no longer require the matrix W to be fixed by the Haar tight framelet filters as described in (3.4.3) and (3.4.4), but instead we seek a matrix $W \in \mathcal{S}$ to yield adaptive spatial regularization of the given HARDI data F.

REMARK 4.1.1. For ease of notation in the accompanying convergence analysis, in this thesis, we consider only modifying the optimization model for spatially regularized, SH-based QBI. Analogous modifications can be made for spatially regularized, SR and WF-based QBI, but it requires more auxiliary variables to be introduced to the optimization models, which makes the notations in the corresponding convergence analysis more complicated.

Furthermore, the adaptive spatial regularization requires greater computational complexity (as the algorithm implemented requires the calculation of singular value decompositions in each iteration), and the spatially regularized SH-based method (M1b in Chapter 3) is more cost-effective for moderate and high SNR levels, so this justifies our sole consideration of an adaptive model for spatially regularized, SH-based QBI in this chapter.

Due to the orthogonality constraint and an ℓ_1-norm term in the objective function of (4.1.2), this is a non-convex and non-smooth optimization problem which is challenging to provide an algorithm with convergence analysis. We note that the convergence analysis of the split Bregman method introduced in Chapter 3 only applies to convex problems. Therefore, we need to devise an alternative algorithm that also has convergence analysis.

We now introduce a method, which we name as the *proximal alternating minimized augmented Lagrangian* (PAMAL) method. As we shall see, the PAMAL method hybridizes the *proximal alternating minimization* (PAM) scheme [**6**] and the augmented Lagrangian method in [**4**]. It consists of Algorithms I and II, where Algorithm I is based on the Augmented Lagrangian method introduced in [**4**], and Algorithm II is actually the PAM scheme in [**6**], which is used to solve Step 1 of Algorithm I.

This section is organized in the following manner. Details of Algorithms I and II will be provided in subsection 4.1.1 and subsection 4.1.2 respectively. Finally, in subsection 4.1.3, some numerical results are given to illustrate how HARDI denoising performance can be improved through adaptivity by Algorithms I and II.

To apply the PAMAL method, we first introduce the auxiliary variables

$$Q = WC^\top, \quad P = W,$$

to split both the non-smooth ℓ_1-term and the non-convex constraint respectively, and then convert (4.1.2) into the following equivalent problem:

$$\min_{C,W,Q,P} \frac{1}{2}\|BC - F\|_F^2 + \frac{\lambda}{2}\|LC\|_F^2 + \mu\|Q\|_1 \quad \text{s.t. } Q = WC^\top,\ P = W,\ P^\top P = I. \tag{4.1.3}$$

Define the indicator function $\delta_{\mathcal{S}}$ on the Stiefel manifold $\mathcal{S}$ defined in (4.1.1) by

$$\delta_{\mathcal{S}}(P) := \begin{cases} 0, & \text{if } P \in \mathcal{S}, \\ +\infty, & \text{if } P \notin S. \end{cases}$$

Denoting the Lagrange multipliers as Λ_1, Λ_2, $\Lambda := (\Lambda_1, \Lambda_2)$ and ρ as a given positive parameter, the corresponding augmented Lagrangian of (4.1.3) is then given by

$$\begin{aligned} L(C, W, Q, P; \Lambda, \rho) :=& \frac{1}{2}\|F - BC\|_F^2 + \frac{\lambda}{2}\|LC\|_F^2 + \mu\|Q\|_1 \\ &+ \langle \Lambda_1, Q - WC^\top \rangle + \frac{\rho}{2}\|Q - WC^\top\|_F^2 \\ &+ \langle \Lambda_2, P - W \rangle + \frac{\rho}{2}\|P - W\|_F^2 + \delta_{\mathcal{S}}(P). \end{aligned} \tag{4.1.4}$$

4.1.1. Algorithm I.

We now describe Algorithm I.

Algorithm I: Method for solving (4.1.4)

Given pre-defined parameters $\{\epsilon^k\}_{k\in\mathbb{N}}$, $\bar{\Lambda}^1 = (\bar{\Lambda}_1^1, \bar{\Lambda}_2^1)$, ρ^1, $\bar{\Lambda}_{p,\min}$, $\bar{\Lambda}_{p,\max}$, τ, γ that satisfy the conditions in Remark 4.1.2, for $k = 1, 2, \ldots$,

(1) Compute (C^k, W^k, Q^k, P^k) such that there exists $\Theta^k \in \partial L(C^k, W^k, Q^k, P^k; \bar{\Lambda}^k, \rho^k)$ satisfying

$$\|\Theta^k\|_\infty \leq \epsilon^k, \quad (P^k)^\top P^k = I, \tag{4.1.5}$$

where $\{\epsilon^k\}_{k\in\mathbb{N}}$ is a sequence of positive tolerance parameters.

(2) Estimate the multipliers Λ_1^{k+1}, Λ_2^{k+1} by

$$\Lambda_1^{k+1} = \bar{\Lambda}_1^k + \rho^k(Q^k - W^k(C^k)^\top), \quad \Lambda_2^{k+1} = \bar{\Lambda}_2^k + \rho^k(P^k - W^k),$$

where $\bar{\Lambda}_p^{k+1}$ is the projection of Λ_p^{k+1} onto the set $\{\Lambda_p : \bar{\Lambda}_{p,\min} \leq \Lambda_p \leq \bar{\Lambda}_{p,\max}\}$, $p = 1, 2$, i.e., for $\forall i, j$,

$$[\bar{\Lambda}_p^{k+1}]_{i,j} = \begin{cases} [\bar{\Lambda}_{p,\min}]_{i,j}, & \text{if } [\Lambda_p^{k+1}]_{i,j} < [\bar{\Lambda}_{p,\min}]_{i,j}, \\ [\Lambda_p^{k+1}]_{i,j}, & \text{if } [\bar{\Lambda}_{p,\min}]_{i,j} \leq [\Lambda_p^{k+1}]_{i,j} \leq [\bar{\Lambda}_{p,\max}]_{i,j}, \\ [\bar{\Lambda}_{p,\max}]_{i,j}, & \text{if } [\Lambda_p^{k+1}]_{i,j} > [\bar{\Lambda}_{p,\max}]_{i,j}. \end{cases} \tag{4.1.6}$$

(3) Update the penalty parameter ρ^{k+1} by

$$\rho^{k+1} = \begin{cases} \rho^k, & \text{if } \|R_i^k\|_\infty \leq \tau \|R_i^{k-1}\|_\infty,\ i = 1, 2, \\ \gamma\rho^k, & \text{otherwise,} \end{cases} \tag{4.1.7}$$

where $R_1^k := Q^k - W^k(C^k)^\top$, $R_2^k := P^k - W^k$, $k \in \mathbb{N}$.

Let us denote $Y := (C, W, Q, P)$ and for given $\bar{\Lambda}$, ρ, define

$$\partial L(Y; \bar{\Lambda}, \rho) := \partial L(C, W, Q, P; \bar{\Lambda}, \rho).$$

The notation $\partial L(Y; \bar{\Lambda}, \rho)$ refers to the Fréchet (limiting) subdifferential of L at point Y (with fixed $\bar{\Lambda}, \rho$). We shall postpone the formalism (see Definition 4.2.2) of the Fréchet subdifferential to section 4.2, where more abstract notations and concepts are frequently used in convergence analysis.

Moreover, the focus of this section is the description of Algorithms I and II to perform adaptive spatial regularization of HARDI data. To allow one to check explicitly whether the ϵ^k-stopping criterion in (4.1.5) is satisfied, we provide within Algorithm II an analytic form of a subgradient element $\Theta^k \in \partial L(C^k, W^k, Q^k, P^k; \bar{\Lambda}^k\rho^k)$ that satisfies (4.1.5) (to be justified instead in section 4.3).

Indeed, to solve Step 1 of Algorithm I, we use Algorithm II which is the PAM scheme to be described in detail later in this subsection. Steps 2 and 3 are straightforward in their implementations. Step 2 updates the multipliers Λ_p^k, which is done by projecting the first order approximation to Λ_p^k to some pre-defined box $(\bar{\Lambda}_{p,\min}, \bar{\Lambda}_{p,\max})$, $p = 1, 2$. Step 3 updates the penalty parameter ρ^k according to the degree of infeasibility. It is noted that the choice of parameters ϵ^k, $\bar{\Lambda}_{p,\min}$, $\bar{\Lambda}_{p,\max}$, ρ^k, γ, τ will impact the convergence property of the proposed algorithm. Remark 4.1.2 below discusses the setting of the parameters.

Remark 4.1.2. (Parameter setting) For Step 1 of Algorithm I to be well defined, the parameters in Algorithm I are set as follows. The sequence of positive tolerance parameters $\{\epsilon^k\}_{k\in\mathbb{N}}$ in (4.1.5) is chosen such that

$$\lim_{k\to\infty} \epsilon^k = 0.$$

The safeguard matrices $\bar{\Lambda}_{p,\min}$, $\bar{\Lambda}_{p,\max}$ are finite-valued matrices satisfying

$$-\infty < [\bar{\Lambda}_{p,\min}]_{i,j} < [\bar{\Lambda}_{p,\max}]_{i,j} < +\infty,\ \forall i, j, \quad p = 1, 2.$$

As we shall see in section 4.3, for Step 1 to be well defined, it suffices to have $\tau \in (0, 1]$ and $\gamma > 1$.

Step 1 is the most crucial and difficult step of Algorithm I. We will later provide an explicit expression for the subgradient element Θ^k in (4.1.10) and (4.1.11) of Algorithm II. Step I of Algorithm I is about finding a point (C^k, W^k, Q^k, P^k) which is an ϵ^k-approximation of a critical point of $L(C, W, Q, P; \bar{\Lambda}^k, \rho^k)$. In line with Definition 4.2.2 (to be given later), we say that $(\bar{C}^k, \bar{W}^k, \bar{Q}^k, \bar{P}^k)$ is a critical point of $L(C, W, Q, P; \bar{\Lambda}^k, \rho^k)$ if

$$0 \in \partial L(\bar{C}^k, \bar{W}^k, \bar{Q}^k, \bar{P}^k; \bar{\Lambda}^k, \rho^k). \tag{4.1.8}$$

Thus, there are two questions to answer when executing Step 1 of Algorithm I:

(1) Is Step 1 of Algorithm I well posed? In other words, is the existence of the points (C^k, W^k, Q^k, P^k) satisfying (4.1.5) guaranteed?
(2) How can we efficiently compute such points with arbitrarily given accuracy, i.e., the perturbation ϵ^k can be arbitrarily small?

In the next subsection, we will first describe the method for solving (4.1.5) that answers Question 2. Then, we will show in Proposition 4.2.6 that this method will answer Question 1 positively.

4.1.2. Algorithm II for step 1 of Algorithm I.

It can be seen that the constraint (4.1.5) is actually an ϵ^k-perturbation of the so-called critical point property (4.1.8). Thus, we need a method that can evaluate the corresponding critical points (C^k, W^k, Q^k, P^k) of the functional $\partial L(C, W, Q, P; \bar{\Lambda}^k, \rho^k)$ with arbitrary accuracy.

Based on the PAM algorithm developed in [**7**], we propose a coordinate-descent method with proximal regularization. The PAM method [**7**] was proposed for solving a class of non-smooth and non-convex optimization problems. Under certain conditions on the objective function, it was shown in [**7**, Theorem 6.2] that the PAM method has global convergence, i.e., the whole sequence generated by the method converges to some critical point. As we will show in section 4.3, the function $L(C, W, Q, P; \bar{\Lambda}^k, \rho^k)$ indeed satisfies the sufficient conditions for the global convergence of the PAM method, provided that the penalty parameters $\{\rho^k\}_{k\in\mathbb{N}}$ satisfies a mild condition. In other words, Step 1 is well defined provided the parameters in Algorithm I are appropriately chosen when the PAM method is used to solve Step 1. Algorithm II provides the outline of the method for solving (4.1.5).

The PAM method can be applied as follows. At the k-th outer iteration, the problem (4.1.8) can be solved with arbitrary accuracy using the set of inner iterations below, which can be viewed as a proximal regularization of a four block Gauss-Seidel method:

$$\begin{cases} C^{k,j} \in \operatorname{argmin}_C L(C, W^{k,j-1}, Q^{k,j-1}, P^{k,j-1}; \bar{\Lambda}^k, \rho^k) + \frac{c_1^{k,j-1}}{2}\|C - C^{k,j-1}\|_F^2; \\ W^{k,j} \in \operatorname{argmin}_W L(C^{k,j}, W, Q^{k,j-1}, P^{k,j-1}; \bar{\Lambda}^k, \rho^k) + \frac{c_2^{k,j-1}}{2}\|W - W^{k,j-1}\|_F^2; \\ Q^{k,j} \in \operatorname{argmin}_Q L(C^{k,j}, W^{k,j}, Q, P^{k,j-1}; \bar{\Lambda}^k, \rho^k) + \frac{c_3^{k,j-1}}{2}\|Q - Q^{k,j-1}\|_F^2; \\ P^{k,j} \in \operatorname{argmin}_P L(C^{k,j}, W^{k,j}, Q^{k,j}, P; \bar{\Lambda}^k, \rho^k) + \frac{c_4^{k,j-1}}{2}\|P - P^{k,j-1}\|_F^2, \end{cases} \tag{4.1.9}$$

where the proximal parameters $\{c_i^{k,j}\}_{k,j}$, can be arbitrarily chosen as long as they are bounded above and below respectively by predetermined positive finite constants $\underline{c}$ and $\bar{c}$, i.e.,

$$\underline{c} \leq c_i^{k,j} \leq \bar{c}, \quad k, j \in \mathbb{N},\, i = 1, 2, 3, 4.$$

It turns out that all subproblems in (4.1.9) have analytic solutions. The solutions to the first and second subproblems are the respective least squares solutions, the third can be obtained by soft-thresholding, and the last can be obtained by the singular value decomposition (SVD). We terminate the algorithm (4.1.9) when there exists

$$\Theta^{k,j} \in \partial L(C^{k,j}, W^{k,j}, Q^{k,j}, P^{k,j}; \bar{\Lambda}^k, \rho^k)$$

satisfying

$$\|\Theta^{k,j}\|_\infty \leq \epsilon^k, \quad (P^{k,j})^\top P^{k,j} = I_m.$$

As we shall see from the proof of Proposition 4.3.3, a direct extension of [**6**, Lemma 5(iii)] shows that we can obtain a smooth (alternative) representation of such a sub-gradient element $\Theta^{k,j} := (\Theta_1^{k,j}, \Theta_2^{k,j}, \Theta_3^{k,j}, \Theta_4^{k,j})$, and it is explicitly given by

$$\begin{cases} \Theta_1^{k,j} := \rho^k[(Q^{k,j-1} - C^{k,j}(W^{k,j-1})^\top + \frac{1}{\rho^k}\bar{\Lambda}_1^k)^\top W^{k,j-1} - (Q^{k,j} - C^{k,j}(W^{k,j})^\top + \frac{1}{\rho^k}\bar{\Lambda}_1^k)^\top W^{k,j}] \\ \qquad\quad + c_1^{k,j-1}(C^{k,j-1} - C^{k,j}); \\ \Theta_2^{k,j} := \rho^k[(Q^{k,j-1} - Q^{k,j})^\top C^{k,j} + P^{k,j-1} - P^{k,j}] + c_2^{k,j-1}(W^{k,j-1} - W^{k,j}); \\ \Theta_3^{k,j} := c_3^{k,j-1}(Q^{k,j-1} - Q^{k,j}); \\ \Theta_4^{k,j} := c_4^{k,j-1}(P^{k,j-1} - P^{k,j}). \end{cases} \tag{4.1.10}$$

Algorithm II below provides a detailed description of the method proposed for solving (4.1.5), which completes Algorithm I.

Algorithm II: Method for solving (4.1.5).

(1) Let $(C^{1,0}, W^{1,0}, Q^{1,0}, P^{1,0})$ be any initialization. For $k \geq 2$, set $(C^{k,0}, W^{k,0}, Q^{k,0}, P^{k,0}) := (C^{k-1}, W^{k-1}, Q^{k-1}, P^{k-1})$.

(2) Re-iterate on j until $\|\Theta^{k,j}\|_\infty \leq \epsilon^k$, where $\Theta^{k,j}$ is defined by (4.1.10).

$$\begin{cases} 1.\, C^{k,j} = (A^{k,j-1})^{-1}[B^\top F + (\bar{\Lambda}_1^k + \rho^k Q^{k,j-1})^\top W^{k,j-1} + c_1^{k,j-1} C^{k,j-1}], \\ 2.\, W^{k,j} = [(\bar{\Lambda}_1^k + \rho^k Q^{k,j-1}) C^{k,j} + \rho^k P^{k,j-1} + \bar{\Lambda}_2^k + c_2^{k,j-1} W^{k,j-1}](D^{k,j-1})^{-1}, \\ 3.\, Q^{k,j} = T_\eta^1\left(\frac{\rho^k W^{k,j}(C^{k,j})^\top - \bar{\Lambda}_1^k + c_3^{k,j-1} Q^{k,j-1}}{\rho^k + c_3^{k,j-1}}\right), \\ 4.\, P^{k,j} = U I_{n\times m} V^\top, \end{cases}$$

where $n \geq m$, $I_{n\times m} := [I_{m\times m} | 0_{m\times(n-m)}]^\top$,

$$A^{k,j-1} := B^\top B + \lambda L^\top L + (\rho^k + c_1^{k,j-1}) I;$$

$$D^{k,j-1} := \rho^k((C^{k,j})^\top C^{k,j} + I) + c_2^{k,j-1} I;$$

$\eta := \eta^{k,j-1} := \mu \cdot (\rho^k + c_3^{k,j-1})^{-1}$, T_η^1 is the soft-thresholding operator defined in (1.1.12), and the matrices $U := U^{k,j}$, $V := V^{k,j}$ are obtained from the singular value decomposition (SVD) of the following matrix

$$\frac{\rho^k W^{k,j} + c_4^{k,j-1} P^{k,j-1} - \bar{\Lambda}_2^k}{\rho^k + c_4^{k,j-1}} =: U \Sigma V^\top.$$

(3) Set $(C^k, W^k, Q^k, P^k) := (C^{k,j}, W^{k,j}, Q^{k,j}, P^{k,j})$ and

$$\Theta^k := \Theta^{k,j}. \tag{4.1.11}$$

4.1.3. Numerical results.

While the focus in this chapter is the convergence analysis of the optimization scheme for adaptive HARDI spatial regularization and denoising, we nevertheless perform experiments to illustrate how optimization on Stiefel manifolds can improve denoising performances. Here, as a simple experiment, we set $b = 2000$, and use 81 sampling gradients on the top slice of the HARDI testing data given in Chapter 3. The numerical experiments are carried out with results averaged over 50 trials under the SNR values of 6, 8 and 10.

Using the initialization $W^0 = W$ defined in (3.4.3), $C^0 = [B^\top B + \lambda L^\top L]^{-1} B^\top F$, $Q^0 = W(C^0)^\top$, $P^0 = C^0$, $\Lambda_1, \Lambda_2 = 0$, we compare the computational performances of the non-adaptive method (SH-based CSA QBI with non-adaptive spatial regularization) and the adaptive method (SH-based CSA QBI with adaptive spatial regularization) in Tables 4.1.1, 4.1.2 and 4.1.3. From Tables 4.1.1, 4.1.2 and 4.1.3, we can deduce that as the SNR becomes lower, denoising generally becomes more effective with the adaptive method.

SNR	NMSE (Non-adaptive)	NMSE (Adaptive)
10	0.0593	**0.0571**
8	0.0694	**0.0677**
6	0.0873	**0.0855**

TABLE 4.1.1. NMSE values of non-adaptive and adaptive SH-based methods under SNR levels 6, 8 and 10.

SNR	Pd (Non-adaptive)	Pd (Adaptive)
10	22.7	**21.8**
8	24.1	**23.3**
6	28.8	**27.7**

TABLE 4.1.2. Pd values of non-adaptive and adaptive SH-based methods under SNR levels 6, 8 and 10.

SNR	AD (Non-adaptive)	AD (Adaptive)
10	5.60	**5.36**
8	5.70	**5.42**
6	7.22	**6.69**

TABLE 4.1.3. AD values of non-adaptive and adaptive SH-based methods under SNR levels 6, 8 and 10.

4.2. Notations and Preliminaries

In this section, we introduce notations and preliminaries on non-smooth analysis. For any $v \in \mathbb{R}^n$, let $[v]_i$ denote its i-th component and let $\mathrm{diag}(v) \in \mathbb{R}^{n\times n}$ denote the diagonal matrix with diagonal entries $\{[v]_i\}_{i=1}^n$. For an index sequence $\mathcal{K} = \{k_0, k_1, k_2, \ldots\}$ that satisfies

$k_{j+1} > k_j$ for any $j \geq 0$, we denote

$$\lim_{k \in \mathcal{K}} x_k := \lim_{j \to \infty} x_{k_j}.$$

DEFINITION 4.2.1. ([**74**]) Let $S \subseteq \mathbb{R}^n$ and $\bar{x} \in S$. A vector v is normal to S at $\bar{x}$ in the regular sense, expressed as $v \in \widehat{N}_S(\bar{x})$, if

$$\langle v, x - \bar{x} \rangle \leq o(\|x - \bar{x}\|) \text{ for } x \in S,$$

where $o(\|y\|)$ is defined by $\lim_{\|y\| \to 0} \frac{o(\|y\|)}{\|y\|} = 0$. A vector is normal to S at $\bar{x}$ in the general sense, expressed as $v \in N_S(\bar{x})$, if there exist sequences $\{x^k\}_k \subset S$, $\{v^k\}_k$ such that $x^k \longrightarrow \bar{x}$ and $v^k \longrightarrow v$ with $v^k \in \widehat{N}_S(x^k)$. The cone $N_S(\bar{x})$ is called the normal cone to S at $\bar{x}$.

For a proper and lower semi-continuous function, denoted by $\sigma : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, the domain of σ is defined as

$$\operatorname{dom} \sigma = \{x \in \mathbb{R}^n : \sigma(x) < +\infty\}.$$

DEFINITION 4.2.2. ([**74**]) Consider a proper and lower semi-continuous function $\sigma : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ and a point $\bar{x}$ with finite $\sigma(\bar{x})$. Let $v \in \mathbb{R}^n$.

(1) The vector v is said to be a regular subgradient of σ at $\bar{x}$, expressed as $v \in \widehat{\partial}\sigma(\bar{x})$, if

$$\sigma(x) \geq \sigma(\bar{x}) + \langle v, x - \bar{x} \rangle + o(\|x - \bar{x}\|).$$

(2) The vector v is said to be a (general or limiting) subgradient of σ at $\bar{x}$, expressed as $v \in \partial\sigma(\bar{x})$, if there exist sequences $\left\{x^k\right\}_k$, $\left\{v^k\right\}_k$ such that $x^k \longrightarrow \bar{x}$, $\sigma(x^k) \longrightarrow \sigma(\bar{x})$ and $v^k \in \widehat{\partial}\sigma(x^k)$ with $v^k \longrightarrow v$.

(3) For each $x \in \operatorname{dom} \sigma$, x is called a (limiting)-critical point of σ if $0 \in \partial\sigma(x)$.

We now describe the formal definition of the Kurdyeka-Łojasiewcz (K-Ł) property and classes of functions that satisfy the K-Ł property.

DEFINITION 4.2.3. A function ψ satisfies the Kurdyeka-Łojasiewcz (K-Ł) property at a point $\bar{x} \in \operatorname{dom}(\partial\psi)$ if there exists $\theta \in [0, 1)$ such that

$$\frac{|\psi(x) - \psi(\bar{x})|^\theta}{\operatorname{dist}(0, \partial\psi(x))} \tag{4.2.1}$$

is bounded around $\bar{x}$ under the following notational conventions: $0^0 = 1$, $\infty/\infty = 0/0 = 0$. In other words, in a certain neighbourhood $\mathcal{U}$ of $\bar{x}$, there exists $\phi(s) = cs^{1-\theta}$ for some $c > 0$ and $\theta \in [0, 1)$ such that the K-Ł inequality holds:

$$\phi'(|\psi(x) - \psi(\bar{x})|)\operatorname{dist}(0, \partial\psi(x)) \geq 1, \quad \forall x \in \mathcal{U} \cap \operatorname{dom}(\partial\psi) \text{ and } \psi(x) \neq \psi(\bar{x}),$$

where $\mathrm{dom}(\partial\psi) := \{x : \partial\psi(x) \neq \emptyset\}$ and $\mathrm{dist}(0, \partial\psi(x)) := \min\{\|y\|_F : y \in \partial\psi(x)\}$.

This property was introduced by Łojasiewicz [**56**] on real analytic functions, for which (4.2.1) is bounded around any critical point $\bar{x}$ for $\theta \in [0, 1)$. Kurdyka extended this property to functions on the o-minimal structure in [**51**]. Recently, the K-Ł inequality was extended to non-smooth subanalytic functions [**16**]. While it is not trivial to check the conditions in the definition, we summarize a few large classes of functions that satisfy the K-Ł-property.

Real analytic functions. A smooth function φ on $\mathbb{R}$ is analytic if $(\frac{\varphi^{(k)}(\cdot)}{k!})^{\frac{1}{k}}$ is bounded for all $k \in \mathbb{N}$ and on any compact set $\mathcal{D} \subset \mathbb{R}$. One can verify whether a real function ψ on $\mathbb{R}^n$ is analytic by checking the analyticity of $\varphi(\cdot) = \psi(x + \cdot y)$ for any x, $y \in \mathbb{R}^n$. For example, any polynomial function is real analytic, such as $\|Ax - b\|_F^2$, $\mathrm{Tr}(x^\top Hx) + b^\top x$.

Semialgebraic functions. A set $\mathcal{D} \subseteq \mathbb{R}^n$ is called semialgebraic [**15**] if it can be represented as

$$\mathcal{D} = \bigcup_{i=1}^{s} \bigcap_{j=1}^{t} \{x \in \mathbb{R}^n : p_{ij}(x) = 0, q_{ij}(x) > 0\},$$

where p_{ij}, q_{ij} are real polynomial functions for $1 \leq i \leq s$, $1 \leq j \leq t$. A function ψ is called semialgebraic if its graph $\mathrm{Gr}(\psi) := \{(x, \psi(x)) : x \in \mathrm{dom}(\psi)\}$ is a semialgebraic set. Semialgebraic functions are subanalytic, so they satisfy the K-Ł inequality according to [**16**,**17**]. We list some known elementary properties of semialgebraic sets and functions below, as they help identify semialgebraic functions for our purposes.

1. If a set $\mathcal{D}$ is semialgebraic, so is its closure $\mathrm{cl}(\mathcal{D})$.
2. If $\mathcal{D}_1$ and $\mathcal{D}_2$ are both semialgebraic, so are $\mathcal{D}_1 \cup \mathcal{D}_2$, $\mathcal{D}_1 \cap \mathcal{D}_2$, and $\mathbb{R}^n \backslash \mathcal{D}_1$.
3. Indicator functions of semialgebraic sets are semialgebraic.
4. Finite sums and products of semialgebraic functions are semialgebraic.
5. The composition of semialgebraic functions is semialgebraic.

From items 1 and 2,

$$\mathcal{S} := \{P \in \mathbb{R}^{n \times m} : P^\top P = I_m\} = \bigcap_{j=1}^{m} \bigcap_{k=1}^{m} \{P : \sum_{i=1}^{n} p_{ki} p_{ji} = \delta_{j,k}\}$$

is a semialgebraic set. Hence, the indicator function $\delta_{\mathcal{S}}(P)$ is a semialgebraic function. The absolute value function $\phi(t) = |t|$ is also semialgebraic since its graph is $\mathrm{cl}(D)$, where

$$\mathcal{D} = \{(t, s) : t + s = 0, -t > 0\} \cup \{(t, s) : t - s = 0, t > 0\}.$$

Hence, the ℓ_1-norm $\|x\|_1$ is semialgebraic since it is the finite sum of absolute functions. Furthermore, the Euclidean norm $\|x\|_2$ is shown to be semialgebraic in [**15**]. According to item 5, $\|Ax-b\|_1$, and $\|Ax-b\|_2$, $\|X\|_{1,2}$ are all semialgebraic functions.

In the end of this section, we list some results that will be used in the subsequent discussion.

REMARK 4.2.4. [**74**, Example 6.7] If S is a closed non-empty subset of $\mathbb{R}^n$, then

$$\partial\delta_S(\bar{x}) = N_S(\bar{x}), \quad \bar{x} \in S.$$

Furthermore, for a smooth mapping $G : \mathbb{R}^n \to \mathbb{R}^m$, i.e., $G(x) := (g_1(x), \ldots, g_m(x))^\top$, define $S = G^{-1}(0) \subset \mathbb{R}^n$. Set

$$\nabla G(x) := [\frac{\partial g_j}{\partial x_i}(x)]_{i,j=1}^{n,m} \in \mathbb{R}^{n\times m}.$$

If $\nabla G(\bar{x})$ has full rank m at a point $\bar{x} \in S$, with $G(\bar{x}) = 0$, then its normal cone to S can be explicitly written as

$$N_S(\bar{x}) = \{\nabla G(\bar{x})y \,|\, y \in \mathbb{R}^m\}. \tag{4.2.2}$$

REMARK 4.2.5. [**12**, Proposition B.24(b)] Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is a real-valued function of the form $f = C + D$, where C is convex and D is continuously differentiable. Then $\bigcup_{x\in\mathcal{M}} \partial f(x)$ is a bounded set whenever the set $\mathcal{M}$ is bounded.

PROPOSITION 4.2.6. ([**6**]) *Suppose that L is a proper and lower semi-continuous function of the form*

$$L(x,y) = f(x) + g(y) + Q(x,y), \quad x \in \mathbb{R}^n,\, y \in \mathbb{R}^m,$$

where f, g are proper lower semi-continuous functions and $Q \in C^1$. Then for all $(x,y) \in \mathbb{R}^n \times \mathbb{R}^m$,

$$\begin{aligned} \partial L(x,y) &= (\nabla_x Q(x,y) + \partial f(x),\, \nabla_y Q(x,y) + \partial g(y)) \\ &= (\partial_x L(x,y),\, \partial_y L(x,y)). \end{aligned} \tag{4.2.3}$$

4.3. Well-posedness of Algorithms I and II

In this subsection, we will show that Step 1 of Algorithm I is well defined by using Algorithm II, provided that some mild condition is satisfied. By denoting

$$X := (C, W, Q, P), \tag{4.3.1}$$

we get $L(C, W, Q, P; \Lambda, \rho) = L(X; \Lambda, \rho)$. In other words, we will show that there are solutions to (4.1.5) and Algorithm II can always find one such solution, under some mild condition.

For Step 1 to be well defined, it needs an important property of Algorithm II, i.e., for each $k \in \mathbb{N}$,

$$(C^{k,j}, W^{k,j}, Q^{k,j}, P^{k,j}) \to (\bar{C}^k, \bar{W}^k, \bar{Q}^k, \bar{P}^k), \quad j \to \infty, \tag{4.3.2}$$

where $(\bar{C}^k, \bar{W}^k, \bar{Q}^k, \bar{P}^k)$ is a critical point of $L(C, W, Q, P; \bar{\Lambda}^k, \rho^k)$. The proof of the limiting property (4.3.2) is based on a result in [**7**]. We next briefly describe [**7**, Theorem 6.2], which considers the minimization of a function $f : \mathbb{R}^{n_1} \times \ldots \times \mathbb{R}^{n_p} \to \mathbb{R} \cup \{+\infty\}$ of the form

$$f(x) = g(x_1, x_2, \ldots, x_p) + \sum_{i=1}^{p} f_i(x_i), \tag{4.3.3}$$

where the terms f_i and g satisfy the following assumptions:

(1) $f_i : \mathbb{R}^{n_i} \to \mathbb{R} \cup \{+\infty\}$ is a proper lower semi-continuous function, $i = 1, 2, \ldots, p$;
(2) g is a C^1 function with locally Lipschitz continuous gradient;
(3) f is a K-Ł (Kurdyka-Łojasiewicz) function (see Remark 4.3.2 for more details), and $\inf_{\mathbb{R}^{n_1} \times \ldots \times \mathbb{R}^{n_p}} f > -\infty$.

For a function f that satisfies all assumptions listed above, it is shown in [**7**] that the PAM scheme generates a critical point of f.

The PAM method [**7**, Algorithm 4] is given as follows. Given initialization $(x_1^0, x_2^0, \ldots, x_p^0)$, for $j = 1, 2, \ldots$,

$$\begin{cases} x_1^j \in \arg\min_{x_1} f_1(x_1) + g(x_1, x_2^{j-1}, \ldots, x_p^{j-1}) + \frac{c_1^{j-1}}{2}\|x_1 - x_1^{j-1}\|_F^2; \\ x_2^j \in \arg\min_{x_2} f_2(x_2) + g(x_1^j, x_2, \ldots, x_p^{j-1}) + \frac{c_2^{j-1}}{2}\|x_2 - x_2^{j-1}\|_F^2; \\ \vdots \\ x_p^j \in \arg\min_{x_p} f_p(x_p) + g(x_1^j, x_2^j, \ldots, x_p) + \frac{c_p^{j-1}}{2}\|x_p - x_p^{j-1}\|_F^2, \end{cases} \tag{4.3.4}$$

where the proximal parameters $\{c_i^j\}_j$, can be arbitrarily chosen as long as they are bounded above and below respectively by predetermined positive finite constants $\underline{c}$ and $\bar{c}$, i.e.,

$$\underline{c} \le c_i^j \le \bar{c}, \quad k, j \in \mathbb{N},\ i = 1, 2, \ldots, p.$$

THEOREM 4.3.1. [**7**, Theorem 6.2] *Suppose that f is a K-Ł function of the form* (4.3.3). *Let $\{x^k\}_{k\in\mathbb{N}}$ be a sequence generated by the PAM scheme in* (4.3.4). *If the sequence $\{x^k\}_{k\in\mathbb{N}}$ is bounded, then the following assertions hold:*

(i) *The sequence $\{x^k\}_{k\in\mathbb{N}}$ has finite length, i.e., $\sum_{k=1}^{\infty} \|x^{k+1} - x^k\|_F < \infty$.*
(ii) *The sequence $\{x^k\}_{k\in\mathbb{N}}$ converges to a critical point $\bar{x}$ of f.*

Indeed, Algorithm II is a specific case of the PAM method presented in [**7**].

REMARK 4.3.2. [K-Ł Property] The global convergence of the PAM method established in [**7**, Theorem 6.2] requires the objective function f to satisfy the so-called Kurdyka-Łojasiewicz (K-Ł) property in its effective domain; see [**7**, Definition 3] for more details on the K-Ł property. It is shown in [**18**, Definition 5] that the so-called semialgebraic functions satisfy the K-Ł property. Indeed, all terms involved in (4.1.4) are semialgebraic functions, which include the ℓ_1-norm $\|WC^\top\|_1$, the linear and quadratic functions (e.g., the trace terms and $\|BC-F\|_F^2$), and $\delta_{\mathcal{S}}$, the indicator function of the Stiefel manifold $\mathcal{S}$. Since a finite sum of semialgebraic functions are also semialgebraic, the objective function (4.1.4) also satisfies the K-Ł property.

For the k-th iteration, define $X := (C, W, P, Q)$. Then the functional defined in (4.1.4) can be expressed as

$$L_k(X) = L(C, W, P, Q; \bar{\Lambda}^k, \rho^k) = f_1(C) + f_2(W) + f_3(Q) + f_4(P) + g_k(C, W, Q, P), \quad (4.3.5)$$

where

$$\begin{cases} f_1(C) := \frac{1}{2}\|BC-F\|_F^2 + \frac{\lambda}{2}\|LC\|_F^2, \quad f_2(W) := 0, \quad f_3(Q) := \frac{1}{\mu}\|Q\|_1, \quad f_4(P) := \delta_{\mathcal{S}}(P), \\ g_k(C, W, Q, P) := \langle \bar{\Lambda}_1^k, Q - WC^\top\rangle + \frac{\rho^k}{2}\|Q - WC^\top\|_F^2 \\ \qquad\qquad\qquad + \langle \bar{\Lambda}_2^k, P - W\rangle + \frac{\rho^k}{2}\|P - W\|_F^2. \end{cases}$$

PROPOSITION 4.3.3. *For each $k \in \mathbb{N}$, denote the functional given by* (4.3.5) *by L_k, and denote the sequence generated by Algorithm II by $\{(C^{k,j}, W^{k,j}, Q^{k,j}, P^{k,j})\}_{j\in\mathbb{N}}$. Then, $\Theta^{k,j}$ defined in* (4.1.10) *satisfies*

$$\Theta^{k,j} \in \partial L(C^{k,j}, W^{k,j}, Q^{k,j}, P^{k,j}; \bar{\Lambda}^k, \rho^k), \quad \forall\, j \in \mathbb{N}.$$

Suppose that the parameters γ, ρ^1 in Algorithm I are chosen such that

$$\gamma > 1, \quad \rho^1 > 0. \quad (4.3.6)$$

Then for each $k \in \mathbb{N}$,

$$\|\Theta^{k,j}\|_\infty \to 0 \quad \text{as } j \to \infty.$$

PROOF. To establish the first part of this proposition, recall the functions g, f_1, f_2, f_3, as defined by L_k in (4.3.5). Then, a direct calculation shows that $\Theta^{k,j} = (\Theta_1^{k,j}, \Theta_2^{k,j}, \Theta_3^{k,j}, \Theta_4^{k,j})$

defined by (4.1.10) can be expressed in terms of partial derivatives of $g := g_k$ as

$$\begin{aligned}
\Theta_1^{k,j} &= -\nabla_C g(C^{k,j}, W^{k,j-1}, Q^{k,j-1}, P^{k,j-1}) - c_1^{k,j-1}(C^{k,j} - C^{k,j-1}) + \nabla_C g(X^{k,j});\\
\Theta_2^{k,j} &= -\nabla_W g(C^{k,j}, W^{k,j}Q^{k,j-1}, P^{k,j-1}) - c_2^{k,j-1}(W^{k,j} - W^{k,j-1}) + \nabla_W g(X^{k,j});\\
\Theta_3^{k,j} &= -\nabla_Q g(C^{k,j}, W^{k,j}, Q^{k,j}, P^{k,j-1}) - c_3^{k,j-1}(Q^{k,j} - Q^{k,j-1}) + \nabla_Q g(X^{k,j});\\
\Theta_4^{k,j} &= -\nabla_P g(C^{k,j}, W^{k,j}, Q^{k,j}, P^{k,j}) - c_3^{k,j-1}(P^{k,j} - P^{k,j-1}) + \nabla_P g(X^{k,j}),
\end{aligned} \tag{4.3.7}$$

where $X^{k,j} = (C^{k,j}, W^{k,j}, Q^{k,j}, P^{k,j})$ is as defined in (4.3.1). On the other hand, given $(C^{k,j-1}, W^{k,j-1}, Q^{k,j-1}, P^{k,j-1})$, the PAM scheme (4.1.9) yields the following necessary first order optimality condition:

$$\begin{cases}
\nabla f_1(C^{k,j}) + \nabla_C g(C^{k,j}, W^{k,j-1}Q^{k,j-1}, P^{k,j-1}) + c_1^{k,j-1}(C^{k,j} - C^{k,j-1}) = 0;\\
\nabla f_2(W^{k,j}) + \nabla_W g(C^{k,j}, W^{k,j}Q^{k,j-1}, P^{k,j-1}) + c_2^{k,j-1}(W^{k,j} - W^{k,j-1}) = 0;\\
\nu^{k,j} + \nabla_Q g(X^{k,j}, Q^{k,j}, P^{k,j-1}) + c_3^{k,j-1}(Q^{k,j} - Q^{k,j-1}) = 0;\\
\omega^{k,j} + \nabla_P g(X^{k,j}, Q^{k,j}, P^{k,j}) + c_4^{k,j-1}(P^{k,j} - P^{k,j-1}) = 0,
\end{cases} \tag{4.3.8}$$

where $\nu^{k,j} \in \partial f_3(Q^{k,j})$ and $\omega^{k,j} \in \partial f_4(P^{k,j})$. Replacing the corresponding terms in (4.3.7) by (4.3.8) gives

$$\begin{cases}
\Theta_1^{k,j} = \nabla f_1(C^{k,j}) + \nabla_C g(C^{k,j}, W^{k,j}, Q^{k,j}, P^{k,j}) \in \partial_C L_k(X^{k,j}),\\
\Theta_2^{k,j} = \nabla f_2(W^{k,j}) + \nabla_W g(C^{k,j}, W^{k,j}, Q^{k,j}, P^{k,j}) \in \partial_W L_k(X^{k,j}),\\
\Theta_3^{k,j} = \nu^{k,j} + \nabla_Q g(C^{k,j}, W^{k,j}, Q^{k,j}, P^{k,j}) \in \partial_Q L_k(X^{k,j}),\\
\Theta_4^{k,j} = \omega^{k,j} + \nabla_P g(C^{k,j}, W^{k,j}, Q^{k,j}, P^{k,j}) \in \partial_P L_k(X^{k,j}).
\end{cases}$$

Thus, for each $k \in \mathbb{N}$,

$$\Theta^{k,j} \in \partial L(C^{k,j}, W^{k,j}, Q^{k,j}, P^{k,j}; \bar{\Lambda}^k, \rho^k), \quad \forall\, j \in \mathbb{N}.$$

Note that from (4.3.7), $\Theta^{k,j}$ is represented by differences of continuous partial derivatives of g and bounded multiples of partial derivatives of smooth proximal terms. Thus, in order to prove that for each $k \in \mathbb{N}$, $\|\Theta^{k,j}\|_\infty \to 0$, as $j \to \infty$ for the second part of the proposition, it suffices to show that for each $k \in \mathbb{N}$, the sequence $\{(C^{k,j}, W^{k,j}, Q^{k,j}, P^{k,j})\}_{j\in\mathbb{N}}$ is convergent. Then it remains to verify that the functionals $L_k(X)$ satisfy the conditions and assumptions made in Theorem 4.3.1. From its definition (4.3.5), it can be seen that the function L_k satisfies the assumptions (i) and (ii) of the function given by (4.3.3), and $L_k(X)$ is also a K-Ł function according to Remark 4.3.2. Thus, we only need to verify that for each $k \in \mathbb{N}$, L_k is bounded below and the sequence $\{X^{k,j}\}_{j\in\mathbb{N}}$ is bounded. For each $k \in \mathbb{N}$, the lower bound of L_k is proved by showing that L_k is a coercive function (i.e., $\lim\limits_{\|X\|_F\to\infty} L_k(X) = +\infty$), provided

that the parameters γ, ρ^1 satisfy (4.3.6). Clearly, f_1, f_2, f_3, f_4 of L_k in (4.3.5) are coercive. For the remaining term, $g_k(C, W, Q, P)$, we may rewrite it as

$$g_k(C, W, Q, P) := g_{1,k}(C, W, Q) + g_{2,k}(W, P),$$

where

$$\begin{cases} g_{1,k}(C, W, Q) := \frac{\rho^k}{2}\|Q - WC^\top + \frac{1}{\rho^k}\bar{\Lambda}_1^k\|_F^2 - \frac{1}{\rho^k}\|\bar{\Lambda}_1^k\|_F^2, \\ g_{2,k}(W, P) := \frac{\rho^k}{2}\|P - W + \frac{1}{\rho^k}\bar{\Lambda}_2^k\|_F^2 - \frac{1}{\rho^k}\|\bar{\Lambda}_2^k\|_F^2. \end{cases}$$

Thus, it can be seen that g_k is a coercive function.

The boundedness of the sequence $\{X^{k,j}\}_{j\in\mathbb{N}}$ is proved by contradiction. Suppose on the contrary that the sequence $X^{k_0,j}$ is not bounded so that $\lim_{j\to\infty} \|X^{k_0,j}\|_F = \infty$. As $L_{k_0}(X)$ is a coercive function, we have then $\lim_{j\to\infty} L_{k_0}(X^{k_0,j}) = +\infty$. However, by setting $x := X$, $f := L_{k_0}$, $\underline{\lambda} := \underline{c}$ in the last inequality of [**7**, page 31], we have that

$$L_{k_0}(X^{k_0,j+1}) + \underline{c}\|X^{k_0,j+1} - X^{k_0,j}\|_F^2 \le L_{k_0}(X^{k_0,j}), \quad j \in \mathbb{N},$$

which imply that $\{L_{k_0}(X^{k_0,j})\}_{j\in\mathbb{N}}$ is a non-increasing sequence. This leads to a contradiction and completes the proof. □

Due to the smooth (alternative) representation of the subgradient element $\Theta^{k,j}$ given by (4.3.7), for each fixed k, the convergence of $\{\Theta^{k,j}\}_{j\in\mathbb{N}}$ to zero implies that given any ϵ^k, one can find a positive integer $J := J(k)$ such that $\|\Theta^{k,j}\|_\infty \le \epsilon^k$, whenever $j \ge J$. Thus, Step 1 of Algorithm I always has a solution, and so Algorithm I is well posed.

4.4. Subsequence Convergence Analysis

For the convenience of notation and discussion, we rewrite the problem (4.1.3) using the notation of vectors. Recall that $C \in \mathbb{R}^{R\times N}$, $W \in \mathbb{R}^{m\times N}$, $Q \in \mathbb{R}^{m\times R}$, $P \in \mathbb{R}^{m\times N}$, with $N := N_x \cdot N_y$ and $m \ge N$. Denote $m_* := RN + mN + mR + mN$. Let $x \in \mathbb{R}^{m_*}$ denote the column vector formed by concatenating the columns of C, W, Q, P:

$$x := \begin{bmatrix} \mathrm{Vec}(C) \\ \hline \mathrm{Vec}(W) \\ \hline \mathrm{Vec}(Q) \\ \hline \mathrm{Vec}(P) \end{bmatrix} \tag{4.4.1}$$

where for a given matrix $X := [X_1|\dots|X_q] \in \mathbb{R}^{p\times q}$, $\mathrm{Vec}(X)$ is defined as the column vector formed by the concatenation of the column vectors $X_1, \dots, X_q$, i.e.,

$$\mathrm{Vec}(X) := \begin{bmatrix} X_1 \\ \hline \vdots \\ \hline X_q \end{bmatrix} \in \mathbb{R}^{pq}.$$

Then, the problem (4.1.3) can be rewritten as the following:

$$\min_{x\in\mathbb{R}^{m_*}} f(x), \quad \text{subject to} \quad h_1(x) = 0 \text{ and } h_2(x) = 0, \tag{4.4.2}$$

where

$$f(x) := \sum_{j=1}^{N} \left\{ \frac{1}{2}\|BC_j - F_j\|_2^2 + \frac{\lambda}{2}\|LC_j\|_2^2 \right\} + \mu \sum_{i=1}^{R} \|Q_i\|_1, \tag{4.4.3}$$

$$h_1(x) := \begin{bmatrix} \mathrm{Vec}(Q) - \mathrm{Vec}(WC^\top) \\ \hline \mathrm{Vec}(P) - \mathrm{Vec}(W) \end{bmatrix} \in \mathbb{R}^{m_1}, \quad h_2(x) := \begin{bmatrix} P_1^\top P_1 - 1 \\ P_2^\top P_1 \\ P_3^\top P_1 \\ \vdots \\ P_N^\top P_1 \\ \hline P_2^\top P_2 - 1 \\ P_3^\top P_2 \\ \vdots \\ P_N^\top P_2 \\ \hline \vdots \\ \hline P_{N-1}^\top P_{N-1} - 1 \\ P_N^\top P_{N-1} \\ \hline P_N^\top P_N - 1 \end{bmatrix} \in \mathbb{R}^{m_2}, \tag{4.4.4}$$

where

$$m_1 := mR + mN, \quad m_2 := N(N+1)/2. \tag{4.4.5}$$

Note that the column vector h_1 is formed by a concatenation of the column vectors of the matrices $Q - WC^\top$ and $P - W$. On the other hand, the column vector h_2 is formed by concatenating the columns in the *lower triangular half* of $P^\top P - I_{N\times N}$ (this is done to remove the redundant equations formed by the strictly upper diagonal entries of the symmetric matrix $P^\top P - I_{N\times N}$).

We now provide an alternative expression for $\mathrm{Vec}(WC^\top)$ in the first block of h_1 by using a classical result involving the Vec notation and the Kronecker product [**54**, equation (13.6)]:

$$\begin{aligned}\mathrm{Vec}(WC^\top) &= \mathrm{Vec}(WI_{N\times N}C^\top)\\ &= (C\otimes I_{N\times N})\mathrm{Vec}(W)\\ &= \begin{bmatrix} c_{11}W_1 + c_{12}W_2 + \ldots + c_{1N}W_N\\ c_{21}W_1 + c_{22}W_2 + \ldots + c_{2N}W_N\\ \vdots\\ c_{R1}W_1 + c_{R2}W_2 + \ldots + c_{RN}W_N\end{bmatrix},\end{aligned} \tag{4.4.6}$$

where the values $[c_{ij}]_{i,j}$ are the entries of the matrix C.

Let λ denote the concatenation of the two Lagrange multiplier vectors of Λ_1 and Λ_2 given by

$$\lambda := \begin{bmatrix}\mathrm{Vec}(\Lambda_1)\\ \mathrm{Vec}(\Lambda_2)\end{bmatrix} \in \mathbb{R}^{m_1+m_2}.$$

Then, the corresponding augmented Lagrangian of (4.4.2) can be expressed as

$$L(x,\lambda;\rho) := f(x) + \sum_{i=1}^{m_1}[\lambda]_i[h_1(x)]_i + \frac{\rho}{2}\sum_{i=1}^{m_1}[h_1(x)]_i^2, \quad \text{subject to} \quad x\in\Gamma,$$

where

$$\Gamma = \{x \,:\, h_2(x) = 0\}.$$

Therefore, a point (C^*, W^*, Q^*, P^*) is a KKT point for (4.1.3) if and only if the vector x^* defined by (4.4.1) is a KKT point for (4.4.2), i.e., there exist $w^*\in\partial f(x^*)$, $\lambda^*\in\mathbb{R}^{m_1}$, $v^*\in\mathbb{R}^{m_2}$ such that

$$w^* + \sum_{i=1}^{m_1}[\lambda^*]_i\nabla[h_1(x^*)]_i + \sum_{i=1}^{m_2}[v^*]_i\nabla[h_2(x^*)]_i = 0;\ h_1(x^*) = 0;\ \text{and } h_2(x^*) = 0, \tag{4.4.7}$$

where λ^*, v^* are column vectors with components $\{[\lambda^*]_i\}_{i=1}^{m_1}$ and $\{[v^*]_i\}_{i=1}^{m_2}$ respectively.

In this section, we establish the subsequence convergence property of Algorithm I, i.e., there exists at least one convergent subsequence of the sequence generated by Algorithm I and it converges to a KKT point of (4.4.2).

THEOREM 4.4.1. *Suppose that $\gamma > 1$ and the sequences $\{\rho^k\}_{k\in\mathbb{N}}$, $\{(C^k, W^k, Q^k, P^k)\}_{k\in\mathbb{N}}$ generated in Algorithm I satisfy the following conditions: firstly, there exists a positive constant M such that*

$$\|C^k\|_2 \le M\rho^k, \quad k\in\mathbb{N}, \tag{4.4.8}$$

and secondly, there exists a positive constant ϵ *such that*

$$|\lambda_i(Z^k) - \lambda_j(Y^k)| \geq \varepsilon, \quad \forall i, j, \quad k \in \mathbb{N}, \tag{4.4.9}$$

where $\lambda_i(Z^k)$, $\lambda_j(Y^k)$ *denote respectively the* i*-th eigenvalue of* Z^k*, the* j*-th eigenvalue of* Y^k *with the matrices* $\{Z^k\}_{k\in\mathbb{N}}$, $\{Y^k\}_{k\in\mathbb{N}}$ *defined by*

$$Z^k := B^\top B + \lambda L^\top L + \rho^k I, \quad Y^k := \rho^k W^k (W^k)^\top, \quad k \in \mathbb{N}. \tag{4.4.10}$$

Then, the limit point set of $\{(C^k, W^k, Q^k, P^k)\}_{k\in\mathbb{N}}$ *is non-empty, and every limit point is a KKT point of the original problem* (4.1.3).

Proof. [Sketch of the proof] The proof of the subssequence convergence property of Algorithm I in this section is organized as follows. Firstly, in subsection 4.4.1, we establish a crucial ingredient needed by the convergence analysis, namely the linear independence of the gradient vectors $\{\nabla[h_1(x)]_i\}_{i=1}^{m_1} \cup \{\nabla[h_2(x)]_i\}_{i=1}^{m_2}$ when $x \in \Gamma$. Consequently, any locally optimal solution to (4.4.2) is necessarily a KKT point of (4.4.2). Secondly, in subsection 4.4.2, we show that any limit point of a sequence generated by Algorithm I is also a KKT point of (4.1.3). Lastly, in subsection 4.4.3, we show that for (4.4.2), the sequence $\{(C^k, W^k, Q^k, P^k)\}_{k\in\mathbb{N}}$ generated by Algorithm I must be bounded. These results together establish the subsequence convergence property of Algorithm I. □

Remark 4.4.2. We remark that (4.4.8) says that $\{\|C^k\|_2\}_{k\in\mathbb{N}}$ should grow no faster than $\{\rho^k\}_{k\in\mathbb{N}}$, up to a multiplicative constant, which is not that difficult to achieve practically. The inequality (4.4.9) is not a particularly strict condition due to the expressions of Z^k and Y^k in (4.4.10) where we expect $\|W^k(W^k)^\top\|_2 \approx 1$ and the parameter λ can be chosen so that $B^\top B + \lambda L^\top L$ is sufficiently (symmetric) positive definite.

Before we commence on the subsequence convergence analysis, we first establish the existence of Lagrange multipliers of the problem (4.1.3) which justifies why an augmented Lagrangian scheme on (4.1.3) is valid.

4.4.1. Linear independence and KKT first order necessary conditions.

It is noted that the objective function f in (4.4.2) is only Lipschitz continuous on bounded sets, which is equivalent to the notion of strict continuity (see [**74**, Definition 9.1]). In order to establish that a locally optimal solution satisfies the KKT first order necessary conditions in the non-smooth case, we need to invoke Theorem 4.4.3 below. However, note that since f is a sum of a convex function and a continuously differentiable function, it follows from

[**74**, Example 9.14] that f is locally Lipschitz continuous on $\mathbb{R}^n$, i.e., f is Lipschitz continuous on any compact subset of $\mathbb{R}^n$. This turns out to be equivalent to the notion of strict continuity (see [**74**, Definition 9.1]) that is a key condition we need to apply Theorem 4.4.3.

THEOREM 4.4.3. [**74**, Exercise 10.52] (*nonsmooth Lagrange multiplier rule*). *For a nonempty, closed set $\mathcal{X} \in \mathbb{R}^n$ and strictly continuous functions $f_0 : \mathbb{R}^n \to \mathbb{R}$ and $F : \mathbb{R}^n \to \mathbb{R}^m$ with $F = (f_1, \ldots, f_m)$, consider the problem*

$$\min_{x \in \mathcal{X}} f_0(x) + \theta\left(F(x)\right),$$

where $\theta : \mathbb{R}^m \to \mathbb{R}$ is proper, lower semi-continuous and convex with effective domain D. Suppose that $\bar{x}$ is a locally optimal solution at which the following constraint qualification is satisfied:

$$0 \in \partial(yF)(\bar{x}) + N_{\mathcal{X}}(\bar{x}),\ y \in N_D(F(\bar{x})) \ \Rightarrow\ y = 0, \tag{4.4.11}$$

where $y \in \mathbb{R}^n$ and $yF := \sum_{i=1}^m y_i f_i$. Then there exists a vector $\bar{y}$ such that

$$0 \in \partial(f_0 + \bar{y}F)(\bar{x}) + N_{\mathcal{X}}(\bar{x}),\ \ \bar{y} \in \partial\theta(F(\bar{x})). \tag{4.4.12}$$

Moreover, the set of such vectors $\bar{y}$ is compact.

Before applying the above result, we first show that the gradient vectors $\{\nabla[h_1(x)]_i\}_{i=1}^{m_1} \cup \{\nabla[h_2(x)]_i\}_{i=1}^{m_2}$ of the equality constraints satisfy a linear independence constraint qualification whenever $x \in \Gamma$, i.e., it satisfies the orthogonality constraints. This leads to the KKT first order necessary conditions.

LEMMA 4.4.4. *Suppose that $x \in \Gamma$. Then, the gradient vectors $\{\nabla[h_1(x)]_i\}_{i=1}^{m_1} \cup \{\nabla[h_2(x)]_i\}_{i=1}^{m_2}$ of the equality constraints in* (4.4.2) *are linearly independent. Consequently, if $\bar{x}$ is a locally optimal solution of the problem* (4.4.2)*, then $\bar{x}$ is a KKT point for* (4.4.2).

PROOF. From the definition (4.4.1) of x, (4.4.4), (4.4.5) and (4.4.6), it can be seen that

$$\nabla h_1(x) = \begin{bmatrix} M_1(x) & 0_{RN\times mN} \\ C^\top \otimes I_{m\times m} & -I_{mN\times mN} \\ -I_{mR\times mR} & 0_{mR\times mN} \\ 0_{mN\times mR} & I_{mN\times mN} \end{bmatrix} \quad \text{and} \quad \nabla h_2(x) = \begin{bmatrix} 0_{RN\times m_2} \\ 0_{mN\times m_2} \\ 0_{mR\times m_2} \\ M_2(x) \end{bmatrix}, \tag{4.4.13}$$

where $M_1(x) \in \mathbb{R}^{RN\times mR}$, $M_2(x) \in \mathbb{R}^{mN\times m_2}$ are given by

$$M_1(x) := \Big[T_1|T_2|\dots|T_N\Big]^\top, \quad T_i := \begin{bmatrix} W_i & & & \\ & W_i & & \\ & & \ddots & \\ & & & W_i \end{bmatrix} \in \mathbb{R}^{mR\times R},\ i = 1, 2, \dots, N,$$

$$M_2(x) = \begin{bmatrix} 2P_1 & P_2 & P_3 & \dots & P_N & 0 & 0 & \dots & 0 & & 0 & 0 & 0 \\ 0 & P_1 & 0 & \dots & 0 & 2P_2 & P_3 & \dots & P_N & & \vdots & \vdots & \vdots \\ 0 & 0 & P_1 & \ddots & \vdots & 0 & P_2 & \dots & 0 & \cdots & 0 & 0 & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 & \vdots & \ddots & \ddots & \vdots & & 2P_{N-1} & P_N & 0 \\ 0 & 0 & \dots & 0 & P_1 & 0 & \dots & 0 & P_2 & & 0 & P_{N-1} & 2P_N \end{bmatrix}.$$

Since $x \in \Gamma$, the column vectors $\{P_i\}_{i=1}^N$ are orthogonal to each other, and thus the columns of $M_2(x)$ are orthogonal to each other. Furthermore, the first three blocks of $\nabla h_2(x)$ form a zero matrix. Thus, from the configurations of the zero and identity block matrices in $\nabla h_1(x)$, $\nabla h_2(x)$ as seen in (4.4.13), one can easily show that $\{\nabla[h_1(x)]_i\}_{i=1}^{m_1} \cup \{\nabla[h_2(x)]_i\}_{i=1}^{m_2}$ are linearly independent for any $x \in \Gamma$.

Secondly, suppose that $\bar{x}$ is a locally optimal solution of the problem (4.4.2). Then $\bar{x} \in \Gamma$. It can be seen from the arguments above that $\nabla h_2(\bar{x})$ is of full column rank. Furthermore, applying (4.2.2) in Remark 4.2.4 on the smooth function h_2 leads to

$$N_\Gamma(\bar{x}) = \{\nabla h_2(\bar{x})z \,|z \in \mathbb{R}^{m_2}\} = \left\{ \sum_{i=1}^{m_2} [z]_i \nabla[h_2(\bar{x})]_i \;\Big|\; z \in \mathbb{R}^{m_2} \right\}. \tag{4.4.14}$$

A direct calculation shows that the constraint qualification in (4.4.11) amounts to verifying

$$0 \in \nabla h_1(\bar{x})y + \nabla h_2(\bar{x})z,\ y \in \mathbb{R}^n \implies y = 0,$$

which holds true by the linear independence of $\{\nabla[h_1(\bar{x})]_i\}_{i=1}^{m_1} \cup \{\nabla[h_2(\bar{x})]_i\}_{i=1}^{m_2}$ as $\bar{x} \in \Gamma$. Notice that δ_0 is proper, lower semi-continuous and convex with effective domain $D = \{0\}$. Then, by applying Theorem 4.4.3 on the setting: $f_0 := f$, $\theta := \delta_0$, $F := h_1$ and $\mathcal{X} := \Gamma$, we established (4.4.12). Together with (4.4.14), we have shown the existence of vectors $\bar{w} \in \partial f(\bar{x}), \bar{y} \in \mathbb{R}^{m_1}, \bar{z} \in \mathbb{R}^{m_2}$ such that

$$\bar{w} + \sum_{i=1}^{m_1} [\bar{y}]_i \nabla[h_1(\bar{x})]_i + \sum_{i=1}^{m_2} [\bar{z}]_i \nabla[h_2(\bar{x})]_i = 0,$$

In other words, the locally optimal point $\bar{x}$ is also a KKT point of (4.4.2). □

4.4.2. Limit points as KKT points.

In this subsection, we show that any limit point generated by Algorithm I is also a KKT point of (4.1.3), i.e., any limit point x^* of the corresponding sequence $\{x^k\}_{k\in\mathbb{N}}$ w.r.t. (C^k, W^k, Q^k, P^k) is a KKT point for (4.4.2). Recall that the normal cone $\partial\delta_{\mathcal{S}}(C, W, Q, P) = N_{\mathcal{S}}(C, W, Q, P)$ in vector notation is given by (4.4.14). Thus, in vector notation, finding the solution satisfying the constraint (4.1.5) at Step 1 of Algorithm I is equivalent to calculating a solution x^k such that there exist vectors $\omega^k \in \partial f(x^k)$ and v^k which satisfy

$$\|\omega^k + \sum_{i=1}^{m_1}([\bar{\lambda}^k]_i + \rho^k[h_1(x^k)]_i)\nabla[h_1(x^k)]_i + \sum_{i=1}^{m_2}[v^k]_i\nabla[h_2(x^k)]_i\|_2 \leq \epsilon^k \tag{4.4.15}$$

with $h_2(x^k) = 0$, $k \in \mathbb{N}$.

Remark 4.4.5. Algorithm I can be recast as an equality constrained version of [**4**, Algorithm 3.1] in vector notation. However, we cannot directly apply the results provided in [**4**, Theorems 4.1-4.2] to our problem, as our objective function f defined by (4.4.3) is not in C^1.

In vector notation, the main result is as follows.

Theorem 4.4.6. *Suppose that $\{x^k\}_{k\in\mathbb{N}}$ is a sequence generated by Algorithm I. Let x^* be a limit point of this sequence, i.e., there exists a subsequence $\mathcal{K} \subseteq \mathbb{N}$ such that $\lim_{k\in\mathcal{K}} x^k = x^*$. Then x^* is also a KKT point of* (4.4.2).

Proof. The proof consists of two main parts. The first part shows that x^* is a feasible point of (4.4.2), i.e., $h_1(x^*) = 0$ and $h_2(x^*) = 0$. The second part shows that x^* satisfies the remaining KKT property in (4.4.7).

We start with the proof of the feasibility of x^* for h_2. After running Step 1 of Algorithm I, we obtain $h_2(x^k) = 0$ for all $k \in \mathcal{K}$, therefore, $h_2(x^*) = 0$, i.e., $x^* \in \Gamma$. The next step is to show $h_1(x^*) = 0$, which is discussed in two cases. We now prove the case where the sequence $\{\rho^k\}_{k\in\mathbb{N}}$ is bounded. Recall that in Algorithm I, $\gamma > 1$. Thus, the update rule (4.1.7) on ρ^k in Step 3 suggests that from some iteration k_0 onwards, the penalty parameter ρ^k will keep the same, which implies that $\|h_1(x^{k+1})\|_\infty \leq \tau\|h_1(x^k)\|_\infty, k \geq k_0$, for some constant $\tau \in [0, 1)$. The feasibility $h_1(x^*) = 0$ is then proved.

In the other case where the sequence $\{\rho^k\}_{k\in\mathbb{N}}$ is not bounded, for each $k \in \mathcal{K}$, there exist vectors δ^k with $\|\delta^k\| \leq \epsilon^k$ and $\epsilon^k \downarrow 0$ such that

$$w^k + \sum_{i=1}^{m_1}([\bar{\lambda}^k]_i + \rho^k[h_1(x^k)]_i)\nabla[h_1(x^k)]_i + \sum_{i=1}^{m_2}[v^k]_i\nabla[h_2(x^k)]_i = \delta^k, \tag{4.4.16}$$

for some $w^k \in \partial f(x^k)$. Dividing both sides of (4.4.16), we have

$$\sum_{i=1}^{m_1} \left([\bar{\lambda}^k/\rho^k]_i + [h_1(x^k)]_i \right) \nabla [h_1(x^k)]_i + \sum_{i=1}^{m_2} [\hat{v}^k]_i \nabla [h_2(x^k)]_i = \frac{\delta^k - w^k}{\rho^k}, \tag{4.4.17}$$

where $\hat{v}^k := (\rho^k)^{-1} v^k$. Define

$$\Xi(x)^\top := [\nabla h_1(x) \,|\, \nabla h_2(x)],$$

and

$$\eta^k := \left([\bar{\lambda}^k/\rho^k]_1 + [h_1(x^k)]_1, \ldots, [\bar{\lambda}^k/\rho^k]_{m_1} + [h_1(x^k)]_{m_1}, [\hat{v}^k]_1, \ldots, [\hat{v}^k]_{m_2} \right)^\top.$$

Then the equality (4.4.17) can be re-written as

$$\Xi(x^k)^\top \eta^k = (\delta^k - w^k)/\rho^k.$$

By Lemma 4.4.4, $\{\nabla [h_1(x^*)]_i\}_{i=1}^{m_1} \cup \{\nabla [h_2(x^*)]_i\}_{i=1}^{m_2}$ are linearly independent as $x^* \in \Gamma$. Moreover, the gradient vectors ∇h_1, ∇h_2 are continuous and $h_2(x^k) = 0$ for all $k \in \mathcal{K}$. Note that by the continuity of the gradient vectors ∇h_1 and ∇h_2, $\Xi(x^k) \to \Xi(x^*)$, which has full rank as $x^* \in \Gamma$.

Therefore, the matrix $\Xi(x^k)$ has full row rank with sufficiently large k by a standard continuity argument in optimization. Thus, $\Xi(x^k)\Xi(x^k)^\top$ is nonsingular, which leads to

$$\eta^k = [\Xi(x^k)\Xi(x^k)^\top]^{-1} \Xi(x^k)(\delta^k - w^k)/\rho^k.$$

Since f is the summation of a convex function and a continuously differentiable function, $\bigcup_{x \in \mathcal{M}} \partial f(x)$ is a bounded set whenever $\mathcal{M}$ is bounded using a simple modification of (4.2.3). Here, we set $\mathcal{M} = \{x^k\}_{k \in \mathcal{K}}$ which is clearly a bounded set. Thus, we have $\{w^k\}_{k \in \mathcal{K}}$ is bounded. Together with $\|\delta^k\| \leq \epsilon^k \downarrow 0$, taking limits as $k \in \mathcal{K}$ goes to infinity gives $\eta^k \to 0$. The boundedness of the safeguard Lagrange multipliers $\{\bar{\lambda}^k\}_k$ implies that $[h_1(x^*)]_i = 0 = [\hat{v}]_j$ for all i, j. Thus, $h_1(x^*) = 0$ and this ends the first part of the proof.

Next, we will show that x^* is a KKT point of the problem (4.4.2). Since $\{w^k\}_{k \in \mathcal{K}}$ is bounded, there exists a subsequence $\mathcal{K}_2 \subseteq \mathcal{K}$ such that $\lim_{k \in \mathcal{K}_2} w^k = w^*$. Recall that $\lim_{k \in \mathcal{K}_2} x^k = x^*$ and $w^k \in \partial f(x^k)$. Thus,

$$w^* \in \partial f(x^*),$$

due to the closedness property of the limiting Fréchet subdifferential. Together with the fact that $[\lambda^{k+1}]_i = [\bar{\lambda}^k]_i + \rho^k [h_1(x^k)]_i$, it can be seen from Algorithm I that for $k \in \mathcal{K}_2$,

$$w^k + \sum_{i=1}^{m_1} [\lambda^{k+1}]_i \nabla [h_1(x^k)]_i + \sum_{i=1}^{m_2} [v^k]_i \nabla [h_2(x^k)]_i = \delta^k, \tag{4.4.18}$$

for some vectors δ^k with $\|\delta^k\| \leq \epsilon^k \downarrow 0$ and $w^k \in \partial f(x^k)$. Define

$$\pi^k := \Big([\lambda^{k+1}]_1, \ldots, [\lambda^{k+1}]_{m_1}, [v^k]_1, \ldots, [v^k]_{m_2} \Big)^\top. \tag{4.4.19}$$

Then (4.4.18) can be re-written as

$$\Xi(x^k)^\top \pi^k = \delta^k - w^k.$$

By the same arguments in the first part, we have the matrix $\Xi(x^k)\Xi(x^k)^\top$ is nonsingular for sufficiently large $k \in \mathcal{K}_2$ and

$$\pi^k = [\Xi(x^k)\Xi(x^k)^\top]^{-1}\Xi(x^k)(\delta^k - w^k). \tag{4.4.20}$$

Hence, by taking the limit on (4.4.20) as $k \in \mathcal{K}_2$ goes to infinity, we have

$$\pi^k \to \pi^* = -[\Xi(x^*)\Xi(x^*)^\top]^{-1}\Xi(x^*)w^*.$$

By the definition (4.4.19) of π^k, taking limit as $k \in \mathcal{K}_2$ approaches infinity on both sides of (4.4.18) leads to

$$w^* + \sum_{i=1}^{m_1} [\lambda^*]_i \nabla [h_1(x^*)]_i + \sum_{i=1}^{m_2} [v^*]_i \nabla [h_2(x^*)]_i = 0,$$

where λ^*, v^* are obtained from π^* similar to (4.4.19). Thus x^* is a KKT point of (4.4.2) and this completes the second part of the proof. □

4.4.3. Existence of limit points.

The results presented in the previous subsections assume the existence of a limit point of the sequence $\{x_k\}_{k\in\mathbb{N}}$, i.e., the sequence generated by Algorithm I contains at least one convergent subsequence. In this subsection, we prove the existence of such a subsequence by showing that the sequence is bounded.

PROPOSITION 4.4.7. *Let $\{(C^k, W^k, Q^k, P^k)\}_{k\in\mathbb{N}}$ be the sequence generated by Algorithm I. Suppose that $\gamma > 1$ and the sequences $\{\rho^k\}_{k\in\mathbb{N}}$, $\{(C^k, W^k, Q^k, P^k)\}_{k\in\mathbb{N}}$ generated in Algorithm I satisfy the conditions* (4.4.8) *and* (4.4.9). *Then, $\{(C^k, W^k, Q^k, P^k)\}_{k\in\mathbb{N}}$ is bounded and thus contains at least one convergent subsequence.*

PROOF. The boundedness of $\{P^k\}_{k\in\mathbb{N}}$ is easy to see from Step 1 of Algorithm I. It remains to show that $\{(C^k, W^k, Q^k)\}_{k\in\mathbb{N}}$ is bounded. Using a direct extension of the result [**18**,

Proposition 3], the first three partial subdifferentials of L in (4.1.5) yield the following: there exist $\phi^k \in \mu\partial\|Q^k\|_1$ and matrices

$$\begin{cases} \zeta_1^k = (B^\top B + \lambda L^\top L + \rho^k I)C^k - [B^\top F + \rho^k(Q^k + \bar{\Lambda}_1^k)^\top W^k]; \\ \zeta_2^k = \rho^k W^k((C^k)^\top C^k + I) - [(\bar{\Lambda}_1^k + \rho^k Q^k)C^k + \rho^k P^k + \bar{\Lambda}_2^k]; \\ \zeta_3^k = \phi^k + \rho^k[Q^k - (W^k C^{k\top} - \frac{1}{\rho^k}\bar{\Lambda}_1^k)], \end{cases} \tag{4.4.21}$$

where $\|\zeta_i^k\|_\infty \leq \epsilon^k$, $i = 1, 2, 3$. Post-multiplying C^k to the terms in the third equation of (4.4.21) and then summing its resultant with the first equation of (4.4.21), we obtain

$$W^k = P^k + \frac{1}{\rho^k}(\zeta_2^k - \bar{\Lambda}_2^k) + \frac{1}{\rho^k}(\zeta_3^k - \phi^k)C^k.$$

Now $\left\{(P^k, \zeta_2^k, \bar{\Lambda}_2^k, \zeta_3^k, \phi^k)\right\}_{k\in\mathbb{N}}$ is bounded and $\left\{\rho^k\right\}_{k\in\mathbb{N}}$ is a non-decreasing positive sequence. Together with the first assumption (4.4.8), we deduce that $\left\{W^k\right\}_{k\in\mathbb{N}}$ is bounded.

We now establish that with the second assumption (4.4.9), $\left\{C^k\right\}_{k\in\mathbb{N}}$ and $\left\{Q^k\right\}_{k\in\mathbb{N}}$ are bounded sequences. Substituting the third equation of (4.4.21) into the first equation of (4.4.21) gives

$$Z^k C^k - C^k Y^k = E^k,$$

where

$$E^k := \zeta_1^k + B^\top F + (\zeta_3^k - \phi^k)^\top W^k$$

is bounded as all the terms in E^k are bounded. Invoking again the classical result involving the Vec notation and the Kronecker product yields

$$[Z^k \otimes I - I \otimes Y^k]\mathrm{Vec}(C^k) = \mathrm{Vec}(E^k).$$

By [**54**, Theorem 13.16], the eigenvalues of $[Z^k \otimes I - I \otimes Y^k]$ are given by $\lambda_i(Z^k) - \lambda_j(Y^k)$, where $\lambda_i(Z^k)$, $\lambda_j(Y^k)$ denote respectively the i-th eigenvalue of Z^k, the j-th eigenvalue of Y^k. Thus the condition (4.4.9) ensures that $\left\{[Z^k \otimes I - I \otimes Y^k]\right\}_{k\in\mathbb{N}}$ is a sequence of invertible matrices with bounded inverses. Therefore,

$$\left\{\mathrm{Vec}(C^k)\right\}_{k\in\mathbb{N}} = \left\{[Z^k \otimes I - I \otimes Y^k]^{-1}\mathrm{Vec}(E^k)\right\}_{k\in\mathbb{N}}$$

is bounded. By the third equation of (4.4.21),

$$Q^k = \frac{1}{\rho^k}(\zeta_3^k - \phi^k) + (W^k(C^k)^\top - \frac{1}{\rho^k}\bar{\Lambda}_1^k).$$

So $\left\{Q^k\right\}_{k\in\mathbb{N}}$ is bounded since $\left\{\rho^k\right\}_{k\in\mathbb{N}}$ is a non-decreasing positive sequence and

$$\left\{(\zeta_3^k, \phi^k, W^k, C^k, \bar{\Lambda}_1^k)\right\}_{k\in\mathbb{N}}$$

is bounded. This completes the proof. □

CHAPTER 5

An Augmented Lagrangian Method for ℓ_1-Regularized Optimization on Stiefel Manifolds

As it turns out, the PAMAL method in Chapter 4 can be applied on a class of ℓ_1-regularized optimization problems defined on Stiefel manifolds which are commonly used in applications. This final chapter is based on a paper submitted for publication [**33**] and is organized as follows. In section 5.1, we give an introduction to ℓ_1-regularized optimization problems with orthogonality constraints and review related work and optimization methods, including the SOC method [**53**]. Section 5.2 describes in detail the PAMAL method (by Algorithms 1 and 2) when applied to ℓ_1-regularized optimization problems with orthogonality constraints. In section 5.3, we provide the convergence analysis of Algorithms 1 and 2. Numerical results in section 5.4 illustrate that for the compressed modes problem [**69**], the PAMAL method is noticeably faster than the SOC method in producing compressed modes with comparable quality.

5.1. ℓ_1-Regularized Optimization Problems with Orthogonality Constraints

In the last few decades, the concept of sparsity has been extensively exploited in a wide range of applications in imaging and information science. Most of these methods focus on the sparsity of the coefficients used for representing the corresponding vector with a set of atoms. The majority of sparsity-driven applications use the ℓ_1-norm as the convex relaxation of the sparsity-prompting function in their variational formulations. Such applications include compressed sensing [**28**, **31**, **40**], model selection and learning [**62**, **81**], and image recovery [**32**, **76**, **77**]. Most of the optimization problems arising from these applications are convex problems. In the last ten years, there has been a huge growth in literature on efficient numerical solvers for these problems; see e.g., [**24**, **49**, **67**].

Nevertheless, there are also many applications in which the data must satisfy non-convex constraints. One commonly seen non-convex constraint is the orthogonality constraint, i.e.,

the data for estimation can be expressed as an orthogonal matrix. Examples of such applications include sparse principal component analysis [**57**], eigenvalue problems in subspace tracking [**87**] and mathematical physics [**69**], and orthogonal Procrustes problem in shape analysis [**41**]. Because orthogonality constraints are non-convex, such problems can be difficult, except in a few simple cases. In recent years, the idea of sparsity is also exploited for data with orthogonal constraints, and ℓ_1-regularization is introduced in the resulting variational model to regularize the sparsity of the data. We briefly describe two representative applications that involve ℓ_1-regularized optimization problems with orthogonality constraints.

(a) *Compressed modes (waves) in physics* [**69**, **70**]. Compressed modes are spatially localized solutions to the eigenvalue problem of the Schrödinger's equation. By considering the independent-particle Schrödinger's equation for a finite system of electrons, the corresponding eigenvalue problem can be reformulated as follows:

$$\min_{X\in\mathbb{R}^{n\times m}} \frac{1}{\mu}\|X\|_1 + \mathrm{Tr}(X^\top HX) \quad \text{s.t.} \quad X^\top X = I_m, \tag{5.1.1}$$

where $\|X\|_1 := \sum_{i=1}^{n}\sum_{j=1}^{m}|X_{i,j}|$, μ is a pre-defined positive parameter that balances the sparsity and the accuracy of the solution, H denotes the discretized Hamiltonian and the columns of X denote the eigenvectors with local support, the so-called compressed modes.

(b) *Feature selection* [**80**, **88**]. Feature selection seeks to choose a smaller subset of features with most information from high dimensional feature sets. It is used in computer vision [**80**] and social media data [**88**], etc. The models for feature selection in [**80**, **88**] adhere to the following ℓ_1-regularized optimization problem with (weighted) orthogonality constraints:

$$\min_{X\in\mathbb{R}^{n\times m}} \frac{1}{\mu}\|X\|_{2,1} + \mathrm{Tr}(X^\top HX), \quad \text{s.t.} \quad X^\top MX = I_m, \tag{5.1.2}$$

where $\|X\|_{2,1} := \sum_{i=1}^{n}(\sum_{j=1}^{m}X_{i,j}^2)^{1/2}$, H is a symmetric matrix and M is a symmetric positive definite matrix of the form $M = R^\top R$, for some $R \in \mathbb{R}^{n\times n}$.

This chapter aims at developing a numerical method to solve (5.1.1), as well as (5.1.2) with minor modifications. The proposed PAMAL method can be viewed as a method that hybridizes the augmented Lagrangian method [**4**] and the proximal alternating minimization (PAM) techniques proposed in [**7**]. The convergence analysis established in this chapter shows that under very mild assumptions on the associated penalty parameters, the sequence generated by the proposed method has the *subsequence convergence property*, i.e., there exists

at least one convergent subsequence and any convergent subsequence converges to a Karush-Kuhn Tucker (KKT) point of (5.2.1) (see (5.3.3) for details).

We now give a brief review on the existing methods that can be applied to solve problems with orthogonality constraints:

$$\min_{X\in\mathbb{R}^{n\times m}} J(X) \quad \text{s.t.} \quad X^\top X = I_m, \tag{5.1.3}$$

where J might be non-convex and non-differentiable. Existing numerical methods that are applicable to (5.1.3) can be classified under two categories: feasible and infeasible approaches.

The feasible approaches satisfy the constraints during each iteration, i.e., each point of the sequence generated by the approach satisfies the orthogonality constraints in (5.1.3). In fact, various optimization methods such as Newton's method, the conjugate gradient method, and the method of steepest descent have been used to solve (5.1.3) as feasible approaches. Most of the existing methods are based on the study of the manifold structures of the orthogonality constraints (see e.g., [**1**, **42**, **48**, **60**, **86**]). These methods require the objective function J to be differentiable, which is not applicable to the problem (5.1.1) studied in this chapter. Furthermore, it is not trivial to satisfy the orthogonality constraints in (5.1.1) during each iteration, as suggested in [**53**]. Therefore, the feasible approach might not be ideal to solve (5.1.3) as its objective function is often non-differentiable.

The PAMAL method proposed in this chapter is an infeasible approach. The infeasible approaches simplify the constrained problem (5.1.3) by relaxing the constraints and iteratively diminish the degree of infeasibility. As a result, intermediate points of the generated sequence may not satisfy the orthogonality constraints. The penalty method (e.g., [**13**, **66**]) approximates the problem (5.1.3) by penalizing the deviations from the constraints:

$$\min_{X\in\mathbb{R}^{n\times m}} J(X) + \frac{1}{2\kappa}\|X^\top X - I_m\|_F^2,$$

where κ denotes some penalty parameter decreasing to zero. If $J(X) = \mathrm{Tr}(X^\top HX)$, then the quadratic penalty model can be viewed as an exact penalty method with a finite penalty parameter κ; see e.g., [**85**]. While the penalty method is simple, it suffers from ill-conditioning issues, especially when the penalty parameter κ decreases to zero. Thus, the standard augmented Lagrangian method [**44**, **45**] is often preferred as it does not require the parameter κ to decrease to zero. When being applied to solve (5.1.3), the standard augmented Lagrangian method yields the following scheme:

$$\begin{cases} X^{k+1} & \in & \mathrm{argmin}_X J(X) + \frac{\rho^k}{2}\|X^\top X - I_m\|_F^2 + \mathrm{Tr}((\Lambda^k)^\top (X^\top X - I_m)), \\ \Lambda^{k+1} & = & \Lambda^k + \rho^k((X^{k+1})^\top X^{k+1} - I_m). \end{cases}$$

The first subproblem of the above augmented Lagrangian scheme is rather complex and generally has no analytic solution. Indeed, it is not trivial to design an efficient solver for the first subproblem.

Aiming at a more computationally efficient method for solving (5.1.3), the splitting of orthogonality constraints (SOC) method [**53**] introduces auxiliary variables to split the orthogonality constraints, which leads to another formulation of (5.1.3):

$$\min_{X,P\in\mathbb{R}^{n\times m}} J(X) \quad \text{s.t.} \quad X = P, \quad P^\top P = I_m. \tag{5.1.4}$$

Using the ideas of alternating direction method of multipliers (ADMM) and the split Bregman method, the SOC method solves (5.1.4) by alternately updating three variables $\{X, P, B\}$:

$$\begin{cases} X^{k+1} \in \operatorname{argmin}_X J(X) + \frac{\rho}{2}\|X - P^k + B^k\|_F^2; \\ P^{k+1} \in \operatorname{argmin}_P \frac{\rho}{2}\|P - (X^{k+1} + B^k)\|_F^2, \text{ s.t. } P^\top P = I; \\ B^{k+1} = B^k + X^{k+1} - P^{k+1}. \end{cases} \tag{5.1.5}$$

In contrast with the standard augmented Lagrangian method, each subproblem in the iterations of the SOC method has an analytic solution. However, the trade-off is its challenging convergence analysis. To the best of our knowledge, it remains an open question whether the SOC method (5.1.5) has the subsequence convergence property.

5.2. An Augmented Lagrangian Method with Proximal Alternating Minimization

In this section, we apply the PAMAL method on the following selected class of ℓ_1-regularized optimization problems with orthogonality constraints:

$$\min_{X,Q,P\in\mathbb{R}^{n\times m}} \frac{1}{\mu}\|Q\|_1 + \operatorname{Tr}(X^\top H X) + \delta_{\mathcal{S}}(P) \quad \text{s.t. } Q - X = 0,\ P - W = 0. \tag{5.2.1}$$

Denoting the Lagrange multipliers as Λ_1, Λ_2 , $\Lambda := (\Lambda_1, \Lambda_2)$ and ρ as a given positive parameter, the corresponding augmented Lagrangian of (5.2.1) is then given by

$$\begin{aligned} L(X, Q, P; \Lambda, \rho) :=& \frac{1}{\mu}\|Q\|_1 + \operatorname{Tr}(X^\top H X) + \delta_{\mathcal{S}}(P) \\ &+ \langle \Lambda_1, Q - X\rangle + \frac{\rho}{2}\|Q - X\|_F^2 \\ &+ \langle \Lambda_2, P - X\rangle + \frac{\rho}{2}\|P - X\|_F^2. \end{aligned} \tag{5.2.2}$$

We now describe Algorithm 1 which is analogous to Algorithm I in the previous chapter.

5.2.1. Algorithm 1.

Here is an outline of Algorithm 1.

Algorithm 1: Method for solving (5.2.2)

Given pre-defined parameters $\{\epsilon^k\}_{k\in\mathbb{N}}$, $\bar{\Lambda}^1 = (\bar{\Lambda}_1^1, \bar{\Lambda}_2^1)$, ρ^1, $\bar{\Lambda}_{p,\min}$, $\bar{\Lambda}_{p,\max}$, τ, γ that satisfy the conditions in Remark 5.2.1, for $k = 1, 2, \ldots$,

(1) Compute (X^k, Q^k, P^k) such that there exists $\Theta^k \in \partial L(X^k, Q^k, P^k; \bar{\Lambda}^k, \rho^k)$ satisfying

$$\|\Theta^k\|_\infty \leq \epsilon^k, \quad (P^k)^\top P^k = I, \tag{5.2.3}$$

where $\{\epsilon^k\}_{k\in\mathbb{N}}$ is a sequence of positive tolerance parameters.

(2) Estimate the multipliers Λ_1^{k+1}, Λ_2^{k+1} by

$$\Lambda_1^{k+1} = \bar{\Lambda}_1^k + \rho^k(Q^k - X^k), \quad \Lambda_2^{k+1} = \bar{\Lambda}_2^k + \rho^k(P^k - X^k),$$

where $\bar{\Lambda}_p^k$ is the projection of Λ_p^k onto the set $\{\Lambda_p : \bar{\Lambda}_{p,\min} \leq \Lambda_p \leq \bar{\Lambda}_{p,\max}\}$, $p = 1, 2$, given by (4.1.6).

(3) Update the penalty parameter ρ^{k+1} by

$$\rho^{k+1} = \begin{cases} \rho^k, & \text{if } \|R_i^k\|_\infty \leq \tau\|R_i^{k-1}\|_\infty,\ i = 1, 2, \\ \gamma\rho^k, & \text{otherwise,} \end{cases}$$

where $R_1^k := Q^k - X^k$, $R_2^k := P^k - X^k$, $k \in \mathbb{N}$.

Similar to the steps in Algorithm I, Step 1 of Algorithm 1 seeks the updates of primal variables such that there is an associated sub-gradient element of L, which satisfies a specified level of tolerance. Step 2 of Algorithm 1 updates the multiplier estimates by first computing the first-order approximations of the multipliers, which are then projected on a suitable box to ensure compactness. Step 3 of Algorithm 1 updates the penalty parameter ρ^k according to the degree of infeasibility. Here, a significant difference between Algorithms 1 and I lies in the choice of the parameters ϵ^k, $\bar{\Lambda}_{p,\min}$, $\bar{\Lambda}_{p,\max}$, ρ^k, γ, τ to ensure that Step 1 is well defined. Remark 5.2.1 discusses the setting of the parameters for Algorithm 1.

REMARK 5.2.1. (Parameter setting) The parameters in Algorithm 1 are set as follows. The sequence of positive tolerance parameters $\{\epsilon^k\}_{k\in\mathbb{N}}$ in (5.2.3) is chosen such that

$$\lim_{k\to\infty} \epsilon^k = 0.$$

The safeguard matrices $\bar{\Lambda}_{p,\min}$, $\bar{\Lambda}_{p,\max}$ are finite-valued matrices satisfying

$$-\infty < [\bar{\Lambda}_{p,\min}]_{i,j} < [\bar{\Lambda}_{p,\max}]_{i,j} < +\infty, \forall i, j, \quad p = 1, 2.$$

As we shall see, for Step 1 to be well defined, it suffices to have $\gamma > 1$, $\tau \in [0, 1)$ and ρ^1 to be a positive penalty parameter such that

$$2H + \rho^1 I \succ 0,$$

where H is the (discrete) Hamiltonian given in (5.1.1).

REMARK 5.2.2. (Relation with the SOC method) Both the PAMAL and the SOC methods [**69**] use the same splitting technique as described in (5.2.1). The main difference between the PAMAL method and the SOC method lies in how (X^k, Q^k, P^k) is updated. In the PAMAL method, the update (Step 1) is done by calling Algorithm 2 which runs several inner iterations to obtain an approximate solution to a critical point (X^k, Q^k, P^k) for L_k with a pre-defined tolerance ϵ_k, i.e.,

$$\Theta^k \in \partial L(X^k, Q^k, P^k; \bar{\Lambda}^k, \rho^k), \text{ s.t.} \quad \|\Theta^k\| \leq \epsilon^k. \tag{5.2.4}$$

The tolerance parameter sequence $\{\epsilon_k\}_{k\in\mathbb{N}}$ can be set to decrease to zero. In contrast, the SOC method only uses a single inner iteration in every outer iteration to solve the problem in Step 1. Thus, there is no guarantee that the corresponding tolerance sequence will converge to zero, which makes the convergence analysis of the SOC method a very challenging task. Despite the fact that multiple inner iterations might be used in Algorithm 1, the flexibility on the accuracy of the solution in Step 1, which is controlled by the setting of the parameters, makes it actually more computationally efficient. For example, when being applied to solving compressed modes problems, the PAMAL method uses much less outer iterations to meet the stopping criteria, and the number of the inner iterations in most outer iterations is only 1 or 2. As a result, the total number of inner iterations of the PAMAL method is less than that of the SOC method (see Tables 5.4.1 and 5.4.2).

Step 1 is the most crucial and difficult step of Algorithm 1. The constraints (5.2.3) can also be viewed as the relaxed KKT conditions for minimizing the augmented Lagrangian L given by (5.2.2).

Thus, there are two questions to answer when executing Step 1 of Algorithm 1:

(1) Is Step 1 of Algorithm 1 well posed? In other words, is the existence of the points (X^k, Q^k, P^k) satisfying (5.2.3) guaranteed?

(2) How can we efficiently compute such points with arbitrarily given accuracy, i.e., can the perturbation ϵ^k be arbitrarily small?

In the next subsection, we will first describe the method for solving (5.2.3) that answers Question 2. Then, we will also show that this method will answer Question 1 positively.

5.2.2. Algorithm 2 for step 1 of Algorithm 1.

It can be seen that the constraint (5.2.3) is actually an ϵ^k-perturbation of the so-called critical point property

$$0 \in \partial L(X, Q, P; \bar{\Lambda}^k, \rho^k). \tag{5.2.5}$$

Thus, we need a method that can evaluate the corresponding critical points (X^k, Q^k, P^k) of the functional $L(X, Q, P; \bar{\Lambda}^k, \rho^k)$ with arbitrary accuracy.

Based on the PAM algorithm [**7**], we propose a coordinate-descent method with proximal regularization. The PAM method [**7**] is proposed for solving a class of non-smooth and non-convex optimization problems. Under certain conditions on the objective function, it is shown in [**7**, Theorem 6.2] that the PAM method has global convergence, i.e., the sequence generated by the method converges to some critical point. As we will show later, the functional $L(X, Q, P; \bar{\Lambda}^k, \rho^k)$ indeed satisfies the sufficient conditions for the global convergence of the PAM method, provided that the penalty parameters $\{\rho^k\}_{k\in\mathbb{N}}$ satisfy a mild condition. In other words, Step 1 is well defined provided that the parameters in Algorithm 1 are appropriately chosen when the PAM method is employed. Algorithm 2 gives the outline of the method for solving (5.2.3).

The PAM method can be applied to solve Step 1 of Algorithm 1. Indeed, at the k-th outer iteration, the problem (5.2.5) can be solved with arbitrary accuracy using the following set of inner iterations, which can be viewed as a proximal regularization of a three block Gauss-Seidel method:

$$\begin{cases} X^{k,j} \in \arg\min_X L(X, Q^{k,j-1}, P^{k,j-1}; \bar{\Lambda}^k, \rho^k) + \frac{c_1^{k,j-1}}{2}\|X - X^{k,j-1}\|_F^2; \\ Q^{k,j} \in \arg\min_Q L(X^{k,j}, Q, P^{k,j-1}; \bar{\Lambda}^k, \rho^k) + \frac{c_2^{k,j-1}}{2}\|Q - Q^{k,j-1}\|_F^2; \\ P^{k,j} \in \arg\min_P L(X^{k,j}, Q^{k,j}, P; \bar{\Lambda}^k, \rho^k) + \frac{c_3^{k,j-1}}{2}\|P - P^{k,j-1}\|_F^2, \end{cases} \tag{5.2.6}$$

where the proximal parameters $\{c_i^{k,j}\}_{k,j}$ can be arbitrarily chosen as long as they satisfy

$$\underline{c} \le c_i^{k,j} \le \bar{c}, \quad k, j \in \mathbb{N}, \quad i = 1, 2, 3,$$

for some pre-determined positive values $\underline{c}$ and $\bar{c}$.

It turns out that all subproblems in (5.2.6) have analytic solutions. The solution to the first subproblem is the respective least squares solution, the solution to the second subproblem can be obtained by soft-thresholding, and the solution to the last subproblem can be obtained by the singular value decomposition (SVD). We terminate the algorithm (5.2.6) when there exists $\Theta^{k,j} \in \partial L(X^{k,j}, Q^{k,j}, P^{k,j}; \bar{\Lambda}^k, \rho^k)$ satisfying

$$\|\Theta^{k,j}\|_\infty \leq \epsilon^k, \quad (P^{k,j})^\top P^{k,j} = I_m.$$

An explicit expression of the term $\Theta^{k,j} := (\Theta_1^{k,j}, \Theta_2^{k,j}, \Theta_3^{k,j})$ is given as follows:

$$\begin{cases} \Theta_1^{k,j} := \rho^k(Q^{k,j-1} - Q^{k,j} + P^{k,j-1} - P^{k,j}) + c_1^{k,j-1}(X^{k,j-1} - X^{k,j}); \\ \Theta_2^{k,j} := c_2^{k,j-1}(Q^{k,j-1} - Q^{k,j}); \\ \Theta_3^{k,j} := c_3^{k,j-1}(P^{k,j-1} - P^{k,j}). \end{cases} \tag{5.2.7}$$

Algorithm 2 describes the method proposed for solving (5.2.3), which completes Algorithm 1.

Algorithm 2: Proposed method for solving (5.2.3).

---

(1) Let $(X^{1,0}, Q^{1,0}, P^{1,0})$ be any initialization. For $k \geq 2$, set $(X^{k,0}, Q^{k,0}, P^{k,0}) := (X^{k-1}, Q^{k-1}, P^{k-1})$.

(2) Re-iterate on j until $\|\Theta^{k,j}\|_\infty \leq \epsilon^k$, where $\Theta^{k,j}$ is defined by (5.2.7).

$$\begin{cases} 1.\, X^{k,j} = (Z^{k,j-1})^{-1}[\bar{\Lambda}_1^k + \bar{\Lambda}_2^k + \rho^k Q^{k,j-1} + \rho^k P^{k,j-1} + c_1^{k,j-1} X^{k,j-1}], \\ 2.\, Q^{k,j} = T_\eta^1(\frac{\rho^k X^{k,j} - \bar{\Lambda}_1^k + c_2^{k,j-1} Q^{k,j-1}}{\rho^k + c_2^{k,j-1}}), \\ 3.\, P^{k,j} = U I_{n\times m} V^\top, \end{cases}$$

where

$$Z^{k,j-1} := 2H + (2\rho^k + c_1^{k,j-1}) I_n,$$

$\eta := \eta^{k,j-1} := \mu \cdot (\rho^k + c_2^{k,j-1})^{-1}$, T_η^1 is the soft-thresholding operator defined in (1.1.12), and the matrices $U := U^{k,j}$, $V := V^{k,j}$ are obtained from the SVD of the following matrix

$$\frac{\rho^k X^{k,j} + c_3^{k,j-1} P^{k,j-1} - \bar{\Lambda}_2^k}{\rho^k + c_3^{k,j-1}} =: U\Sigma V^\top.$$

(3) Set $(X^k, Q^k, P^k) := (X^{k,j}, Q^{k,j}, P^{k,j})$ and

$$\Theta^k := \Theta^{k,j}.$$

5.2.3. Well-posedness of Algorithms 1 and 2.

In this subsection, we will show that Step 1 of Algorithm 1 is well defined by using Algorithm 2, provided that some mild condition is satisfied. By denoting

$$W := (X, Q, P),$$

we get $L(X, Q, P; \Lambda, \rho) = L(W; \Lambda, \rho)$. In other words, we will show that the solutions for (5.2.3) are non-empty and Algorithm 2 can always find a solution, under some mild condition. For Step 1 to be well defined, it needs an important property of Algorithm 2, i.e., for each $k \in \mathbb{N}$,

$$(X^{k,j}, Q^{k,j}, P^{k,j}) \to (\bar{X}^k, \bar{Q}^k, \bar{P}^k), \quad j \to \infty, \tag{5.2.8}$$

where $(\bar{X}^k, \bar{Q}^k, \bar{P}^k)$ is a critical point of $L(X, Q, P; \bar{\Lambda}^k, \rho^k)$. The proof of the limiting property (5.2.8) is based on the PAM result in [**7**]. For the k-th iteration, recalling that $W = (X, P, Q)$, then the functional defined in (5.2.2) can be expressed as

$$L_k(W) := L(X, Q, P; \bar{\Lambda}^k, \rho^k) = f_1(X) + f_2(Q) + f_3(P) + g_k(X, Q, P), \tag{5.2.9}$$

where

$$\begin{cases} f_1(X) := \mathrm{Tr}(X^\top H X), \quad f_2(Q) := \frac{1}{\mu}\|Q\|_1, \quad f_3(P) := \delta_{\mathcal{S}}(P), \\ g_k(X, Q, P) := \langle \bar{\Lambda}_1^k, Q - X\rangle + \frac{\rho^k}{2}\|Q - X\|_F^2 \\ \qquad\qquad +\langle \bar{\Lambda}_2^k, P - X\rangle + \frac{\rho^k}{2}\|P - X\|_F^2. \end{cases}$$

Then, we have

PROPOSITION 5.2.3. *For each $k \in \mathbb{N}$, denote the functional given by (5.2.9) by L_k, and denote the sequence generated by Algorithm 2 by $\{(X^{k,j}, Q^{k,j}, P^{k,j})\}_{j\in\mathbb{N}}$. Then, $\Theta^{k,j}$ defined in (5.2.7) satisfies*

$$\Theta^{k,j} \in \partial L(X^{k,j}, Q^{k,j}, P^{k,j}; \bar{\Lambda}^k, \rho^k), \quad \forall\, j \in \mathbb{N}.$$

If the parameters γ, ρ^1 in Algorithm 1 are chosen such that

$$\gamma > 1, \quad \rho^1 > 0, \quad \rho^1 I_n + 2H \succ 0, \tag{5.2.10}$$

then for each $k \in \mathbb{N}$,

$$\|\Theta^{k,j}\|_\infty \to 0 \quad \text{as } j \to \infty.$$

PROOF. To establish the first part of this proposition, recall the functions g_k, f_1, f_2, f_3, as defined by L_k in (5.2.9). Then, a direct calculation shows that $\Theta^{k,j} = (\Theta_1^{k,j}, \Theta_2^{k,j}, \Theta_3^{k,j})$

defined by (5.2.7) can be expressed in terms of partial derivatives of $g := g_k$ as

$$\begin{aligned}
\Theta_1^{k,j} &= -\nabla_X g(X^{k,j}, Q^{k,j-1}, P^{k,j-1}) - c_1^{k,j-1}(X^{k,j} - X^{k,j-1}) + \nabla_X g(X^{k,j}, Q^{k,j}, P^{k,j});\\
\Theta_2^{k,j} &= -\nabla_Q g(X^{k,j}, Q^{k,j}, P^{k,j-1}) - c_2^{k,j-1}(Q^{k,j} - Q^{k,j-1}) + \nabla_Q g(X^{k,j}, Q^{k,j}, P^{k,j});\\
\Theta_3^{k,j} &= -\nabla_P g(X^{k,j}, Q^{k,j}, P^{k,j}) - c_3^{k,j-1}(P^{k,j} - P^{k,j-1}) + \nabla_P g(X^{k,j}, Q^{k,j}, P^{k,j}).
\end{aligned} \tag{5.2.11}$$

On the other hand, given $(X^{k,j-1}, Q^{k,j-1}, P^{k,j-1})$, the PAM scheme (5.2.6) yields the following necessary first order optimality condition:

$$\begin{cases}
\nabla f_1(X^{k,j}) + \nabla_X g(X^{k,j}, Q^{k,j-1}, P^{k,j-1}) + c_1^{k,j-1}(X^{k,j} - X^{k,j-1}) = 0;\\
\nu^{k,j} + \nabla_Q g(X^{k,j}, Q^{k,j}, P^{k,j-1}) + c_2^{k,j-1}(Q^{k,j} - Q^{k,j-1}) = 0;\\
\omega^{k,j} + \nabla_P g(X^{k,j}, Q^{k,j}, P^{k,j}) + c_3^{k,j-1}(P^{k,j} - P^{k,j-1}) = 0,
\end{cases} \tag{5.2.12}$$

where $\nu^{k,j} \in \partial f_2(Q^{k,j})$ and $\omega^{k,j} \in \partial f_3(P^{k,j})$. Replacing the corresponding terms in (5.2.11) by (5.2.12) gives

$$\begin{cases}
\Theta_1^{k,j} = \nabla f_1(X^{k,j}) + \nabla_X g(X^{k,j}, Q^{k,j}, P^{k,j}) \in \partial_X L_k(W^{k,j}),\\
\Theta_2^{k,j} = \nu^{k,j} + \nabla_Q g(X^{k,j}, Q^{k,j}, P^{k,j}) \in \partial_Q L_k(W^{k,j}),\\
\Theta_3^{k,j} = \omega^{k,j} + \nabla_P g(X^{k,j}, Q^{k,j}, P^{k,j}) \in \partial_P L_k(W^{k,j}).
\end{cases}$$

Thus, for each $k \in \mathbb{N}$,

$$\Theta^{k,j} \in \partial L(X^{k,j}, Q^{k,j}, P^{k,j}; \bar{\Lambda}^k, \rho^k), \quad \forall\, j \in \mathbb{N}.$$

As for the second part of the proposition, to prove that for each $k \in \mathbb{N}$, $\|\Theta^{k,j}\|_\infty \to 0$, as $j \to \infty$, it suffices to show that for each $k \in \mathbb{N}$, the sequence $\{(X^{k,j}, Q^{k,j}, P^{k,j})\}_{j\in\mathbb{N}}$ is convergent. Then it remains to verify that the functionals $L_k(W)$ satisfy the conditions and assumptions made in Theorem 4.3.1. From its definition (5.2.9), it can be seen that the function L_k satisfies the assumptions (i) and (ii) of the function given by (4.3.3), and $L_k(W)$ is also a K-Ł function according to Remark 4.3.2. Thus, we only need to verify that for each $k \in \mathbb{N}$, L_k is bounded below and the sequence $\{W^{k,j}\}_{j\in\mathbb{N}}$ is bounded. For each $k \in \mathbb{N}$, the lower bound of L_k is proved by showing that L_k is a coercive function provided that the parameters γ, ρ^1 satisfy (5.2.10). Clearly, the two terms f_2 and f_3 of L_k in (5.2.9) are coercive. For the remaining term, $f_1(X) + g_k(X, Q, P)$, we may rewrite it as the following:

$$f_1(X) + g_k(X, Q, P) := g_{1,k}(X, P) + g_{2,k}(X, Q),$$

where

$$\begin{cases} g_{1,k}(X,P) := \frac{1}{2}\mathrm{Tr}(X^\top(2H+\rho^k I_n)X) - \langle \rho^k P + \bar{\Lambda}_2^k, X\rangle + \langle \bar{\Lambda}_2^k, P\rangle + \frac{\rho^k}{2}\|P\|_F^2, \\ g_{2,k}(X,Q) := \frac{\rho^k}{2}\|Q - X + \bar{\Lambda}_1^k/\rho^k\|_F^2 - \|\bar{\Lambda}_1^k/\rho^k\|_F^2. \end{cases}$$

It can be seen that $g_{2,k}$ is bounded below. Since $P \in \mathcal{S}$ (i.e., $P^\top P = I_m$), so $\|P\|_\infty = 1$ and $\|P\|_F = \sqrt{m}$. Thus we have

$$g_{1,k}(X,P) \geq \frac{1}{2}\mathrm{Tr}(X^\top(2H+\rho^k I_n)X) - \|X\|_1 - \langle \bar{\Lambda}_2^k, X\rangle - \|\bar{\Lambda}_2^k\|_1 + \frac{\rho^k m}{2}.$$

Therefore, $g_{1,k}$ is coercive as long as $2H + \rho^k I_n$ is symmetric positive definite and $P \in \mathcal{S}$. Notice that the sequence $\{\rho^k\}_{k\in\mathbb{N}}$ is set in Step 3 of Algorithm 1 such that it is non-decreasing when $\gamma > 1$ which implies that $\rho^k \geq \rho^1$ for any $k > 1$. If the initial parameter ρ^1 is set sufficiently large such that $\rho^1 I_n + 2H \succ 0$, we have the positive definiteness of $2H + \rho^k I_n$ for any $k \geq 1$ and thus the term $f_1 + g_k$ is also coercive. In short, the functions $\{L_k\}_{k\in\mathbb{N}}$ defined as (5.2.9) are all coercive.

The boundedness of the sequence $\{W^{k,j}\}_{j\in\mathbb{N}}$ is proved by contradiction. Suppose on the contrary that the sequence $\{W^{k_0,j}\}_{j\in\mathbb{N}}$ is not bounded. This means that $\lim_{j\to\infty} \|W^{k_0,j}\|_F = \infty$. As $L_{k_0}(W)$ is a coercive function, we have then $\lim_{j\to\infty} L_{k_0}(W^{k_0,j}) = +\infty$. However, by setting $x := W$, $f := L_{k_0}$, $\underline{\lambda} := \underline{c}$ in the last inequality of [**18**, page 31], we have that

$$L_{k_0}(W^{k_0,j+1}) + \underline{c}\|W^{k_0,j+1} - W^{k_0,j}\|_F^2 \leq L_{k_0}(W^{k_0,j}), \quad j \in \mathbb{N},$$

which implies that $\{L_{k_0}(W^{k_0,j})\}_{j\in\mathbb{N}}$ is a non-increasing sequence, which leads to a contradiction. This completes the proof. □

REMARK 5.2.4. The condition $2H + \rho^1 I_n \succ 0$ is a mild condition. Take the compressed modes problem [**69**] for example. In the case of the free-electron (FE) model, the discretized Hamiltonian matrix $H \succ 0$ and thus ρ^1 can be taken to be any positive number. In the case of the Kronig-Penney (KP) model, the magnitudes of the negative eigenvalues of the corresponding matrix H are generally less than 1. Thus, we may set $\rho^1 > 2$.

5.3. Convergence Analysis

For the convenience of notation and discussion, we rewrite the problem (5.2.1) using the notation of vectors. Let $x \in \mathbb{R}^{3mn}$ denote the column vector formed by concatenating the columns of X, Q, P:

$$x := \mathrm{Vec}([X|Q|P]). \tag{5.3.1}$$

Then, the problem (5.2.1) can be rewritten as the following:

$$\min_{x\in\mathbb{R}^{3mn}} f(x), \quad \text{subject to} \quad h_1(x)=0; \text{ and } h_2(x)=0; \tag{5.3.2}$$

where $h_1(x) \in \mathbb{R}^{2mn}$ denotes $\mathrm{Vec}([Q-X|P-X])$, $h_2(x)$ denote the $fracm(m+1)2 \times 1$ vector obtained by vectorizing only the the lower triangular entries of the symmetric matrix $P^\top P - I_m$ given in (4.4.4), and

$$f(x) = \sum_{j=1}^{m} \left(\mu^{-1}\|Q_j\|_1 + X_j^\top H X_j\right).$$

Let λ denote the concatenation of the two Lagrange multiplier vectors of Λ_1 and Λ_2 given by

$$\lambda := \mathrm{Vec}([\Lambda_1|\Lambda_2]).$$

Then, the corresponding augmented Lagrangian of (5.3.2) can be expressed as

$$L(x,\lambda;\rho) := f(x) + \sum_{i=1}^{m_1}[\lambda]_i[h_1(x)]_i + \frac{\rho}{2}\sum_{i=1}^{m_1}[h_1(x)]_i^2, \quad \text{subject to} \quad x\in\Gamma,$$

where $m_1 := 2mn, m_2 := m(m+1)/2$, and

$$\Gamma = \{x \,:\, h_2(x)=0\}.$$

Therefore, a point (X^*, Q^*P^*) is a KKT point for (5.2.1) if and only if the vector x^* defined by (5.3.1) is a KKT point for (5.3.2), i.e., there exists $w^* \in \partial f(x^*)$, $\lambda^* \in \mathbb{R}^{m_1}$, $v^* \in \mathbb{R}^{m_2}$ such that

$$w^* + \sum_{i=1}^{m_1}[\lambda^*]_i \nabla[h_1(x^*)]_i + \sum_{i=1}^{m_2}[v^*]_i\nabla[h_2(x^*)]_i = 0; h_1(x^*)=0; \text{ and } h_2(x^*)=0. \tag{5.3.3}$$

where λ^*, v^* are column vectors with components $\{[\lambda]_i\}_{i=1}^{m_1}$ and $\{[v]_i\}_{i=1}^{m_2}$ respectively.

In this section, we establish the subsequence convergence property of Algorithm 1, i.e., there exists at least one convergent subsequence of the sequence generated by Algorithm 1 and it converges to a KKT point of (5.3.2).

THEOREM 5.3.1. *Suppose that the positive parameters γ, ρ^1 in Algorithm 1 are chosen so that $\gamma > 1$, $2H + \rho^1 I_n \succ 0$. Let $\{(X^k, Q^k, P^k)\}_{k\in\mathbb{N}}$ be the sequence generated by Algorithm 1. Then, the limit point set of $\{(X^k, Q^k, P^k)\}_{k\in\mathbb{N}}$ is non-empty, and every limit point is a KKT point of the original problem* (5.3.2).

PROOF. [Sketch of the proof] The proof of the subsequence convergence property of Algorithm 1 is organized as follows. Firstly, in section 5.3.1, we establish a crucial ingredient needed for the convergence analysis, namely the linear independence of the gradient vectors $\{\nabla[h_1(x)]_i\}_{i=1}^{m_1} \cup \{\nabla[h_2(x)]_i\}_{i=1}^{m_2}$ when $x \in \Gamma$. Consequently, any locally optimal solution to (5.3.2) is necessarily a KKT point of (5.3.2). Secondly, in section 5.3.2, we show that any limit point of a sequence generated by Algorithm 1 is also a KKT point of (5.2.1). Lastly, in section 5.3.3, we show that for (5.3.2), the sequence $\left\{(X^k, Q^k, P^k)\right\}_{k\in\mathbb{N}}$ generated by Algorithm 1 must be bounded. These results together establish the subsequence convergence property of Algorithm 1. □

Before we commence on the subsequence convergence analysis, we first establish the existence of Lagrange multipliers of the problem (5.2.1) which justifies why an augmented Lagrangian scheme on (5.2.1) is valid.

5.3.1. Linear independence and KKT first order necessary conditions.

Analogous to the previous chapter, we have the corresponding linear independence and KKT first order necessary conditions in this subsection.

LEMMA 5.3.2. *Suppose that $x \in \Gamma$. Then the gradient vectors $\{\nabla[h_1(x)]_i\}_{i=1}^{m_1} \cup \{\nabla[h_2(x)]_i\}_{i=1}^{m_2}$ of the equality constraints in* (5.3.2) *are linearly independent. Consequently, if $\bar{x}$ is a locally optimal solution of the problem* (5.3.2)*, then $\bar{x}$ is a KKT point for* (5.3.2).

PROOF. From the definition (5.3.1) of x, by setting $m_3 := mn$, it can be seen that

$$\nabla h_1(x) = \begin{bmatrix} -I_{m_3\times m_3} & -I_{m_3\times m_3} \\ I_{m_3\times m_3} & 0_{m_3\times m_3} \\ 0_{m_3\times m_3} & I_{m_3\times m_3} \end{bmatrix} \quad \text{and} \quad \nabla h_2(x) = \begin{bmatrix} 0_{m_3\times m_2} \\ 0_{m_3\times m_2} \\ M(x) \end{bmatrix}, \tag{5.3.4}$$

where $M(x) \in \mathbb{R}^{m_3\times m_2}$ is given by

$$M(x) = \begin{bmatrix} 2P_1 & P_2 & P_3 & \dots & P_m & 0 & 0 & \dots & 0 & & 0 & 0 & 0 \\ 0 & P_1 & 0 & \dots & 0 & 2P_2 & P_3 & \dots & P_m & & \vdots & \vdots & \vdots \\ 0 & 0 & P_1 & \ddots & \vdots & 0 & P_2 & \dots & 0 & \cdots & 0 & 0 & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 & \vdots & \ddots & \ddots & \vdots & & 2P_{m-1} & P_m & 0 \\ 0 & 0 & \dots & 0 & P_1 & 0 & \dots & 0 & P_2 & & 0 & P_{m-1} & 2P_m \end{bmatrix}.$$

Since $x \in \Gamma$, the column vectors $\{P_i\}_{i=1}^m$ are orthogonal to each other, and thus the columns of $M(x)$ are orthogonal to each other. Furthermore, the first $2m_3$ rows of $\nabla h_2(x)$ form a zero matrix. Thus, from the matrix structures of $\nabla h_1(x)$, $\nabla h_2(x)$ in (5.3.4), one can easily show that $\{\nabla[h_1(x)]_i\}_{i=1}^{m_1} \cup \{\nabla[h_2(x)]_i\}_{i=1}^{m_2}$ are linearly independent for any $x \in \Gamma$.

Secondly, if $\bar{x}$ is a locally optimal solution of the problem (5.3.2), then $\bar{x} \in \Gamma$. We shall not repeat the argument as it follows verbatim to the proof of Lemma 4.4.4. Thus, the locally optimal point $\bar{x}$ is also a KKT point of (5.3.2). □

5.3.2. Limit points as KKT points.

In this subsection, we show that any limit point generated by Algorithm 1 is also a KKT point of (5.2.1), i.e., any limit point x^* of the corresponding sequence $\{x^k\}_{k\in\mathbb{N}}$ w.r.t. (X^k, Q^k, P^k) is a KKT point for (5.3.2). In vector notation, the main result is stated as follows.

THEOREM 5.3.3. *Suppose that $\{x^k\}_{k\in\mathbb{N}}$ is a sequence generated by Algorithm 1. Let x^* be a limit point of this sequence, i.e., there exists a subsequence $\mathcal{K} \subseteq \mathbb{N}$ such that $\lim_{k\in\mathcal{K}} x^k = x^*$. Then x^* is also a KKT point of* (5.3.2).

PROOF. The proof follows verbatim of that of Theorem 4.4.6. □

5.3.3. Existence of limit points.

The results presented in the previous subsections assume the existence of a limit point of the sequence $\{x_k\}_{k\in\mathbb{N}}$, i.e., the sequence generated by Algorithm 1 contains at least one convergent subsequence. In this subsection, we prove the existence of the sequence by showing that it is bounded.

PROPOSITION 5.3.4. *Let $\{(X^k, Q^k, P^k)\}_{k\in\mathbb{N}}$ be the sequence generated by Algorithm 1. Suppose that the parameters γ, ρ^1 in Algorithm 1 are chosen so that $\gamma > 1$ and $2H + \rho^1 I_n \succ 0$. Then, $\{(X^k, Q^k, P^k)\}_{k\in\mathbb{N}}$ is bounded and thus contains at least one convergent subsequence.*

PROOF. The boundedness of $\{P^k\}_{k\in\mathbb{N}}$ is easy to see from Step 1 of Algorithm 1. It remains to show that $\{(X^k, Q^k)\}_{k\in\mathbb{N}}$ is bounded. Using a direct extension of the result [**18**, Proposition 3], the first two partial subdifferentials of L in (5.2.3) yield the following: there exist $\nu^k \in \frac{1}{\mu}\partial\|Q^k\|_1$ and $\zeta^k = (\zeta_1^k, \zeta_2^k) \in \mathbb{R}^{n\times m} \times \mathbb{R}^{n\times m}$ such that

$$\begin{cases} \zeta_1^k = 2HX^k - \bar{\Lambda}_1^k + \rho^k(X^k - Q^k) - \bar{\Lambda}_2^k + \rho^k(X^k - P^k); \\ \zeta_2^k = \nu^k + \bar{\Lambda}_1^k + \rho^k(-X^k + Q^k), \end{cases} \tag{5.3.5}$$

where $\|\zeta^k\|_\infty \leq \epsilon^k$. Summing the above two equations gives

$$(2H + \rho^k)X^k = \zeta_1^k + \zeta_2^k + \bar{\Lambda}_2^k + \rho^k P^k - \nu^k.$$

Together with $2H + \rho^k I_n \succ 0$, we have

$$X^k = (2H + \rho^k I_n)^{-1}(\zeta_1^k + \zeta_2^k + \bar{\Lambda}_2^k + \rho^k P^k - \nu^k).$$

Let $H = V\mathrm{diag}(\lambda_1, \ldots, \lambda_n)V^\top$ denote the SVD of the symmetric matrix H. Then,

$$\begin{aligned} X^k &= V\mathrm{diag}(1/(2\lambda_1 + \rho^k), \ldots, 1/(2\lambda_n + \rho^k))V^\top(\zeta_1^k + \zeta_2^k + \bar{\Lambda}_2^k - \nu^k) \\ &\quad + V\mathrm{diag}(\rho^k/(2\lambda_1 + \rho^k), \ldots, \rho^k/(2\lambda_n + \rho^k))V^\top P^k. \end{aligned}$$

Recall that $\{\rho^k\}_{k\in\mathbb{N}}$ is non-decreasing and $2H + \rho^1 I_n \succ 0$. We have then, for $k \in \mathbb{N}$, $2H + \rho^k I_n \succ 0$, which gives $2\lambda_i + \rho^k > 0$, $i = 1, 2, \ldots, n$. Thus, for all $k \in \mathbb{N}$,

$$0 < 1/(2\lambda_i + \rho^k) \leq 1/(2\lambda_i + \rho^1) < +\infty, \ i = 1, 2, \ldots, n, \tag{5.3.6}$$

$$0 < \rho^k/(2\lambda_i + \rho^k) \leq \max(\rho^1/(2\lambda_i + \rho^1), 1), \ i = 1, 2, \ldots, n. \tag{5.3.7}$$

Together with the fact that $\{\zeta^k\}_{k\in\mathbb{N}}$ and $\{\bar{\Lambda}^k\}_{k\in\mathbb{N}}$ are both bounded, combining the two inequalities (5.3.6) and (5.3.7) shows that the sequence $\{X^k\}_{k\in\mathbb{N}}$ is bounded. Then, the boundedness of the sequence $\{Q^k\}_{k\in\mathbb{N}}$ can also be derived from (5.3.5). □

It is noted that the result still holds if the ℓ_1-term $\frac{1}{\mu}\|Q\|_1$ in (5.2.1) is replaced by any convex function with bounded subgradients on its domain, e.g., $\frac{1}{\mu}\|Q\|_{2,1}$.

5.4. The Compressed Modes for Variational Problems in Physics

This section is organized as follows. In subsection 5.4.1, we present some background information on compressed modes. In subsection 5.4.2, we review some existing methods to obtain compressed modes, which include the SOC method introduced by Lai and Osher in [**53**]. Finally, we compare the performance of the PAMAL method against that of the SOC method in subsection 5.4.3 on the compressed modes problem.

5.4.1. Background on compressed modes.

Motivated by the localized Wannier functions [**61**] used in solid state physics and quantum chemistry, a variational approach is developed in [**69**] to produce the so-called *compressed modes*, which are spatially localized solutions to the time-independent Schrödinger's equation:

$$\hat{H}\phi(x) = \lambda\phi(x), \quad x \in \Omega. \tag{5.4.1}$$

In this equation, Ω is a bounded subset of $\mathbb{R}^d$ and $\hat{H}$ denotes the corresponding Hamiltonian defined by

$$\hat{H} = -\frac{1}{2}\Delta + V,$$

where Δ denotes the Laplacian operator and V denotes the potential energy function, represented by a multiplication operator with a bounded measurable function. Spatially localized solutions to the eigenvalue problem (5.4.1) not only enable efficient computations related to the general Schrödinger's equation, but also fit certain observations in physics. For example, the screened correlations in condensed matter are typically short-ranged [**73**].

In [**69**], the authors considered the independent-particle Schrödinger's equation for a finite system of N electrons, with the electron spin neglected for simplicity. The ground state energy of these electrons, denoted by E_0, can be formulated as a variational problem, which minimizes the total energy subject to orthonormality conditions for the stationary states:

$$E_0 = \min_{\Phi_N} \sum_{j=1}^{N} \langle \phi_j, \hat{H}\phi_j \rangle \quad \text{s.t.} \quad \langle \phi_j, \phi_k \rangle = \delta_{jk}, \tag{5.4.2}$$

where $\langle \phi_j, \phi_k \rangle := \int_\Omega \phi_j(x)\overline{\phi_k(x)}\, dx$. The solutions $\Phi_N = \{\phi_i\}_{i=1}^N$ form a set of orthonormal eigenfunctions which are usually not spatially localized. Therefore, an ℓ_1-regularized model is proposed in [**69**] to obtain the solutions of (5.4.2) with better spatial localization:

$$E = \min_{\Psi_N} \sum_{j=1}^{N} \frac{1}{\mu}|\psi_j|_1 + \langle \psi_j, \hat{H}\psi_j \rangle \quad \text{s.t.} \quad \langle \psi_j, \psi_k \rangle = \delta_{jk}, \tag{5.4.3}$$

where $|\psi_j|_1 := \int_\Omega |\psi_j(x)|\, dx$ and the constant μ is a pre-defined parameter that balances the sparsity and the accuracy of the solution. It is shown in [**8**, **68**] that with fixed N, the approximation error of the energy E calculated by (5.4.3) to the ground state energy E_0 is decreasing as $\mu \to \infty$. By considering $\Omega = [0, L]^d$ with periodic boundary conditions and equally spaced nodes in each direction, the discretized version of (5.4.2) is expressed as

$$\Psi_N = \text{argmin}_{\Psi \in \mathbb{R}^{n\times N}} \frac{1}{\mu}\|\Psi\|_1 + \text{Tr}(\Psi^\top H \Psi) \quad \text{s.t.} \quad \Psi^\top \Psi = I, \tag{5.4.4}$$

where $\|\Psi\|_1 := \sum_{i,j} |\Psi_{i,j}|$, and H is a symmetric matrix formed by the discretization on Hamiltonian $\hat{H}$. The solution $\Psi_N = \{\psi_j\}_{j=1}^N$ is called the compressed modes (CMs) for the eigenvector problem (5.4.1).

5.4.2. Existing methods for compressed modes.

The presence of non-convex orthogonality constraints in (5.4.4) makes it a challenging problem to solve. By re-formulating the constrained optimization problem (5.4.4) as the

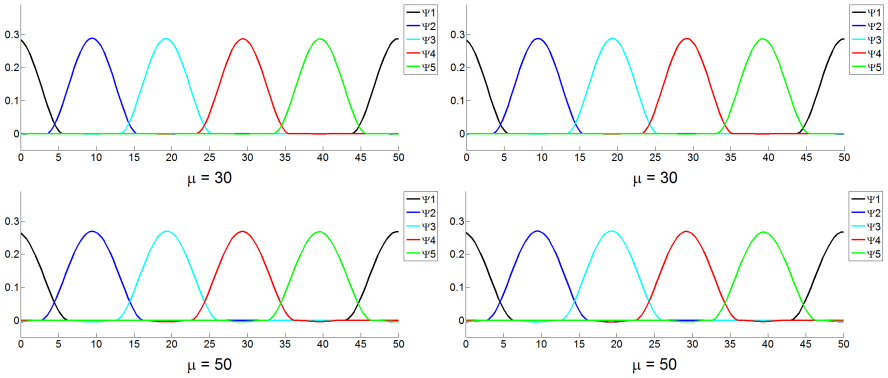

FIGURE 5.4.1. The comparison of the first five modes obtained for the 1D FE model with different values of μ. The first column shows the results computed by the SOC method [**69**]; and the second column shows the results computed by the PAMAL method.

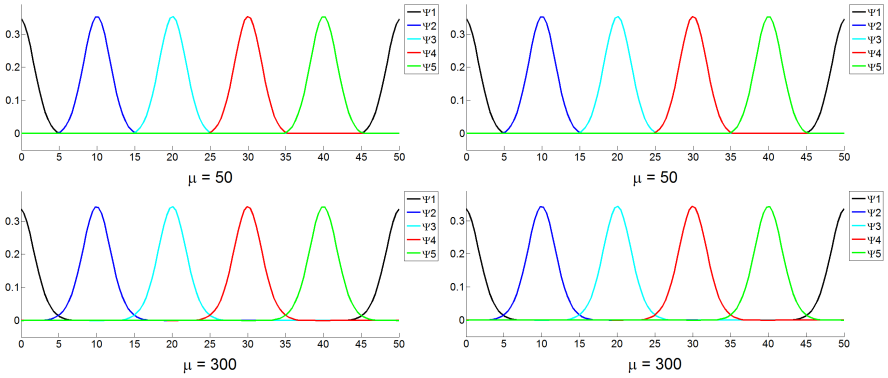

FIGURE 5.4.2. The comparison of the first five modes obtained for the 1D KP model with two different values of μ. The first column shows the results computed by the SOC method [**69**]; and the second column shows the results computed by the PAMAL method.

following,

$$\min_{\Psi,Q,P} \frac{1}{\mu}\|Q\|_1 + \mathrm{Tr}(\Psi^\top H\Psi) \quad \text{s.t.} \quad Q-\Psi=0,\ P-\Psi=0,\ P^\top P = I,$$

a split Bregman iteration based method was proposed in [**69**] to solve it. It is demonstrated in the numerical experiments conducted in [**69**] that the SOC method can produce compressed modes of good quality. Nevertheless, to the best of our knowledge, there is no analysis of its convergence property provided in the literature.

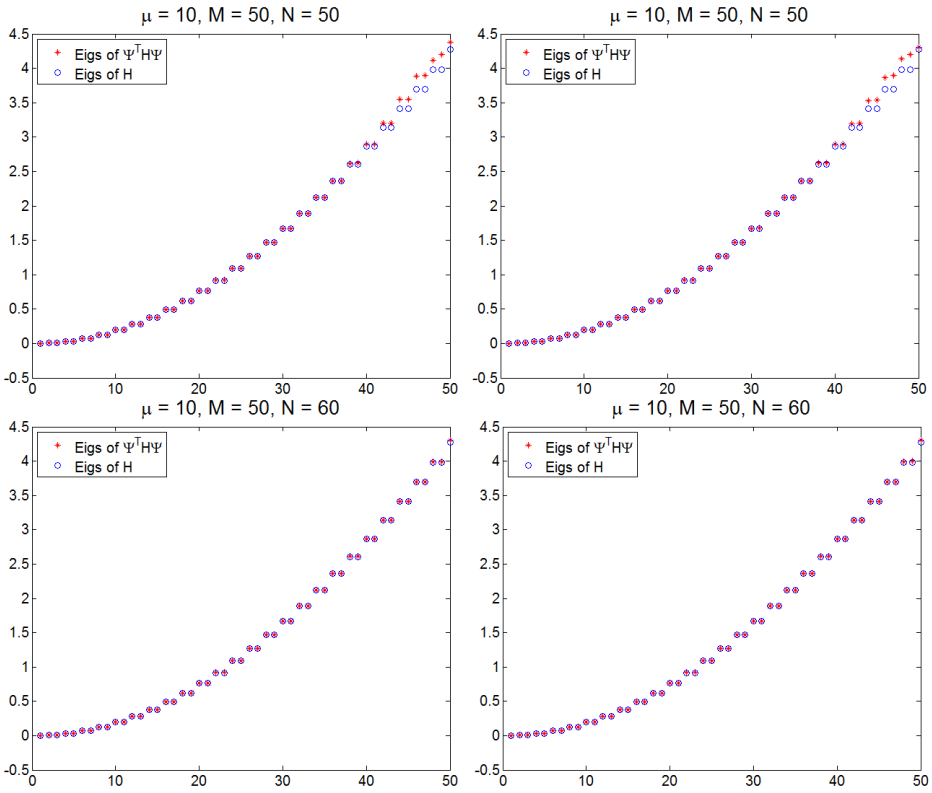

FIGURE 5.4.3. The comparison of the first 50 eigenvalues obtained for the 1D FE model with different values of N. The first column shows the results computed by the SOC method [**69**]; and the second column shows the results computed by the PAMAL method.

More recently, a convex relaxation approach is proposed in [**52**], which re-models the CMs problem into a density matrix minimization problem with ℓ_1-regularization:

$$\begin{aligned} &\min_{P\in\mathbb{R}^{n\times n}} \ \mathrm{Tr}(HP) + \frac{1}{\mu}\|P\|_1 \\ &\text{s.t. } P = P^\top, \quad \mathrm{Tr}(P) = N, \quad 0 \preceq P \preceq I. \end{aligned} \tag{5.4.5}$$

In [**52**], the convex model (5.4.5) is solved by the split Bregman method, with the convergence analysis provided.

5.4.3. Computations of CMs by the PAMAL method.

In this subsection, we applied the PAMAL method to solve the compressed modes problem under the same setting as [**69**], which includes both the free-electron (FE) and the Kronig-Penney (KP) models. Through the experiments, potential functions are approximated by

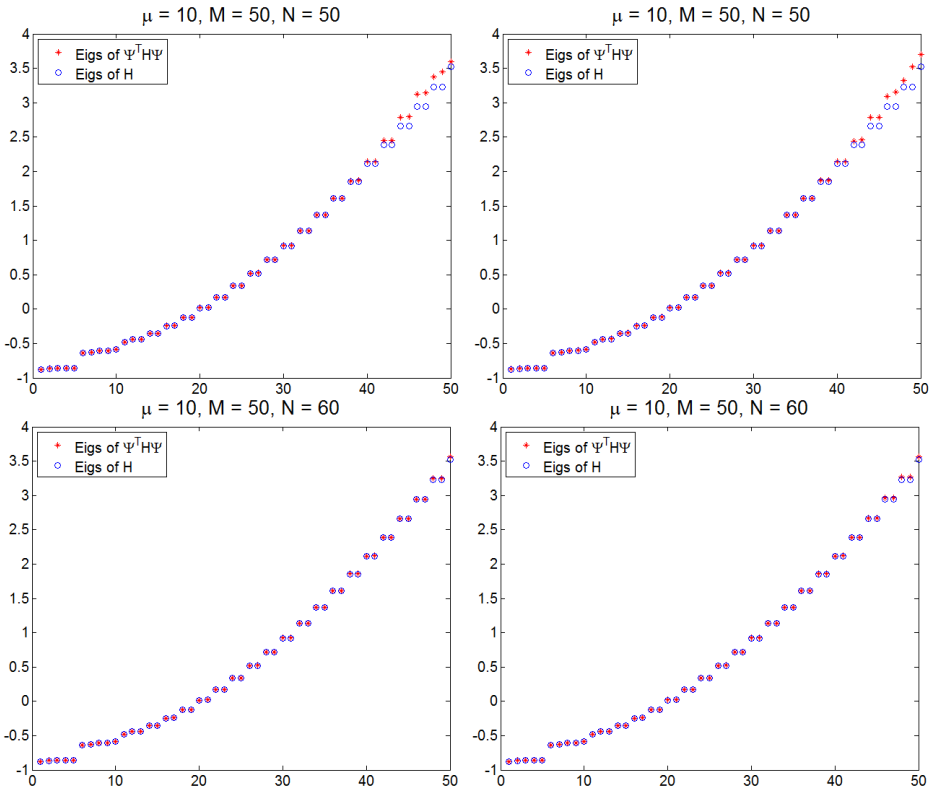

FIGURE 5.4.4. The comparison of the first 50 eigenvalues obtained for the 1D KP model with different values of N. The first column shows the results computed by the SOC method [**69**]; and the second column shows the results computed by the PAMAL method.

Gaussians. More specifically, we set

$$V \equiv 0, \quad \text{and} \quad V(x) = -V_0 \sum_{j=1}^{N_{el}} \exp[-\frac{(x-10j)^2}{2\delta^2}],$$

with $V_0 := 1$, $N_{el} := 5$, $\delta := 3$ respectively in the FE and KP models. Moreover, the domain $\Omega := [0, 50]$ is discretized with $n = 128$ equally spaced nodes.

The parameters of the PAMAL method are set as follows: $\tau = 0.99$, $\gamma = 1.01$, $\rho^1 = 2|\lambda_{\min}(H)| + N/2$, $\bar{\Lambda}_{p,\min} = -100$, $\bar{\Lambda}_{p,\max} = 100$, $p = 1, 2$, and

$$\epsilon^k = (0.999)^k, \quad k \in \mathbb{N}.$$

The parameters in Algorithm 2 are set as $\underline{c} = c_i^{k,j} = \bar{c} = 0.5$, for all k, j, i in both the FE and KP models. In the SOC method, as recommended by [**69**], we use the same penalty parameters ($\lambda = \mu N/20$, $r = \mu N/5$) in [**69**, equations 15–17]. In both the PAMAL method and the SOC

method, the same random orthonormal matrix initialization is used. In order to produce CMs of reasonable localization, we set the stopping criterion as $|J(P^k) - J(P^{k-1})| < 10^{-5}$, where J is the objective function given in (5.4.4), i.e., $J(\Psi) := \frac{1}{\mu}\|\Psi\|_1 + \mathrm{Tr}(\Psi^\top H\Psi)$.

Both methods are implemented in MATLAB and the experiments are done on a PC with a 1.70GHz CPU and 4G of RAM. The number of outer iterations, total number of inner iterations and CPU time, of the PAMAL and SOC methods are averaged over 50 experimental trials. See Table 5.4.1 and Table 5.4.2 for the comparison of the computational costs of the two methods. In general, with the same stopping criterion, the proposed PAMAL method is at least twice as fast as the SOC method. As discussed in Remark 5.2.2, the performance gain of the PAMAL methods comes from the flexibility on the accuracy of the solution for Step 1 in Algorithm 1.

Problems			No. of outer iterations		Total no. of inner iterations		CPU time (s)	
N	M	μ	PAMAL	SOC	PAMAL	SOC	PAMAL	SOC
5	5	30	77	237	82	237	0.07	0.15
5	5	50	87	499	92	499	0.07	0.27
50	50	10	512	3124	522	3124	1.35	7.25
60	50	10	484	4147	497	4147	1.54	11.02

TABLE 5.4.1. Computational costs of the PAMAL method and the SOC method for the FE model.

Problems			No. of outer iterations		Total no. of inner iterations		CPU time (s)	
N	M	μ	PAMAL	SOC	PAMAL	SOC	PAMAL	SOC
5	5	50	66	304	75	304	0.06	0.17
5	5	300	62	1826	71	1826	0.05	0.94
50	50	10	496	3179	507	3179	1.38	7.44
60	50	10	478	4118	491	4118	1.55	10.99

TABLE 5.4.2. Computational costs of the PAMAL method and the SOC method for the KP model.

The first five CMs of the 1D FE and KP models computed by the SOC/PAMAL methods are shown in the first/second columns of Figure 5.4.1 and Figure 5.4.2 respectively. It can be seen that the CMs computed by the PAMAL method are compactly supported functions and their localization degree is largely similar to that of the CMs obtained via the SOC method,

as shown in Figure 5.4.1 and Figure 5.4.2. We next examine the approximation behavior of the unitary transformations derived from the CMs to the eigenmodes of the Schrödinger operator. The approximation accuracy is demonstrated by comparing the first M eigenvalues $(\sigma_1, \ldots, \sigma_M)$ of the matrix $\mathrm{Tr}(\Psi_N^\top H \Psi_N)$ obtained by the M eigenvalues $(\lambda_1, \ldots, \lambda_M)$ of the corresponding Schrödinger operators. Figure 5.4.3 and Figure 5.4.4 reveal that the approximation accuracies of the SOC and PAMAL methods are similar for the FE and KP models respectively, where it can be seen that $\{\sigma_i\}_{i=1}^m$ converges to $\{\lambda_i\}_{i=1}^m$ with increasing number N of CMs.

To summarize this chapter, we proposed the PAMAL method, a numerical method for solving a class of ℓ_1-regularized optimization problems with orthogonality constraints. It is shown that the proposed method has the subsequence convergence property, which is not provided in the existing SOC method [**69**]. In addition, the experiments show that when being applied to solve the compressed modes problem, the proposed PAMAL method is noticeably faster than the SOC method in producing modes of comparable quality.

Remark 5.4.1. (Possible Future Research Work/Directions) It would be interesting to investigate the effectiveness of the PAMAL method (over the SOC method) on feature selection problems, which include computer vision [**80**] and social media data [**88**].

Bibliography

[1] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, 2009.

[2] R.A. Adams and J.J.F. Fournier, *Sobolev Spaces*, Academic Press, 2003.

[3] I. Aganj, C. Lenglet, G. Sapiro, E. Yacoub, K. Ugurbil, and N. Harel, *Reconstruction of the orientation distribution function in single-and multiple-shell q-ball imaging within constant solid angle*, Magn. Reson. Med. **64** (2010), no. 2, 554–566.

[4] R. Andreani, E.G. Birgin, J.M. Martínez, and M.L. Schuverdt, *On augmented Lagrangian methods with general lower-level constraints*, SIAM J. Optim. **18** (2007), no. 4, 1286–1309.

[5] K. Atkinson and W. Han, *Spherical Harmonics and Approximations on the Unit Sphere: An Introduction*, Springer, 2012.

[6] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, *Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality*, Math. Oper. Res. **35** (2010), no. 2, 438–457.

[7] H. Attouch, J. Bolte, and B.F. Svaiter, *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods*, Math. Program. **137** (2013), no. 1–2, 91–129.

[8] F. Barekat, *On the consistency of compressed modes for variational problems*, arXiv preprint arXiv:1310.4552 (2013).

[9] P.J. Basser, J. Mattiello, and D. LeBihan, *Estimation of the effective self-diffusion tensor from the NMR spin echo*, J. Magn. Reson., Ser B **103** (1994), no. 3, 247–254.

[10] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci. **2** (2009), no. 1, 183–202.

[11] J.J. Benedetto and S. Li, *The theory of multiresolution analysis frames and applications to filter banks*, Appl. Comput. Harmon. Anal. **5** (1998), no. 4, 389–427.

[12] D.P. Bertsekas, *Nonlinear Programming*, Athena Scientific, 1999.

[13] F. Bethuel, H. Brezis, and F. Hélein, *Asymptotics for the minimization of a Ginzburg-Landau functional*, Calc. Var. Partial Differential Equations **1** (1993), no. 2, 123–148.

[14] P.K. Bhattacharyya, *Distributions: Generalized Functions with Applications in Sobolev Spaces*, Walter de Gruyter, 2012.

[15] J. Bochnak, M. Coste, and M.-F. Roy, *Real Algebraic Geometry*, Springer, 1998.

[16] J. Bolte, A. Daniilidis, and A. Lewis, *The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems*, SIAM J. Optim. **17** (2007), no. 4, 1205–1223.

[17] J. Bolte, A. Daniilidis, A. Lewis, and M. Shiota, *Clarke subgradients of stratifiable functions*, SIAM J. Optim. **18** (2007), no. 2, 556–572.

[18] J. Bolte, S. Sabach, and M. Teboulle, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Math. Program. **146** (2014), no. 1-2, 459–494.

[19] C. de. Boor, *A Practical Guide to Splines*, Springer, 1978.

[20] C. de. Boor, R. DeVore, and A. Ron, *On the construction of multivariate (pre) wavelets*, Constr. Approx. **9** (1993), no. 2-3, 123–166.

[21] J. Cai, R. Chan, L. Shen, and Z. Shen, *Restoration of chopped and nodded images by framelets*, SIAM J. Sci. Comp. **30** (2008), no. 3, 1205–1227.

[22] J. Cai, R. Chan, and Z. Shen, *A framelet-based image inpainting algorithm*, Appl. Comput. Harmon. Anal. **24** (2008), no. 2, 131–149.

[23] J. Cai, B. Dong, S. Osher, and Z. Shen, *Image restoration: Total variation, wavelet frames, and beyond*, J. Amer. Math. Soc. **25** (2012), no. 4, 1033–1089.

[24] J. Cai, S. Osher, and Z. Shen, *Convergence of the linearized Bregman iteration for ℓ^1-norm minimization*, Math. Comput. **78** (2009), no. 268, 2127–2136.

[25] ———, *Linearized Bregman iterations for frame-based image deblurring*, SIAM J. Imaging Sci. **2** (2009), no. 1, 226–252.

[26] ———, *Split Bregman methods and frame based image restoration*, Multiscale Model. Simul. **8** (2009), no. 2, 337–369.

[27] F. Cakoni and D. Colton, *Qualitative Methods in Inverse Scattering Theory: An Introduction*, Springer, 2005.

[28] E.J. Candès and B. Recht, *Exact matrix completion via convex optimization*, Found. Comput. Math. **9** (2009), no. 6, 717–772.

[29] E.J. Candès and J. Romberg, *Quantitative robust uncertainty principles and optimally sparse decompositions*, Found. Comput. Math. **6** (2006), no. 2, 227–254.

[30] E.J. Candès, J. Romberg, and T. Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inf. Theory **52** (2006), no. 2, 489–509.

[31] ———, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math. **59** (2006), no. 8, 1207–1223.

[32] T. Chan, A. Marguina, and P. Mulet, *High-order total variation-based image restoration*, SIAM J. Sci. Comput. **22** (2000), no. 2, 503–516.

[33] W. Chen, H. Ji, and Y. You, *An augmented Lagrangian method for ℓ_1-regularized optimization problems with orthogonality constraints,*. SIAM J. Sci. Comput., submitted.

[34] O. Christensen, *An Introduction to Frames and Riesz Bases*, Birkhäuser, 2003.

[35] A. Daducci, E.J. Canales-Rodríguez, M. Descoteaux, E. Garyfallidis, Y. Gur, Y.-C. Lin, M. Mani, S. Merlet, M. Paquette, A. Ramirez-Manzanares, et al., *Quantitative comparison of reconstruction methods for intra-voxel fiber recovery from diffusion MRI*, IEEE Trans. Med. Imaging **33** (2014), no. 2, 384–399.

[36] I. Daubechies, *Orthonormal bases of compactly supported wavelets*, Comm. Pure Appl. Math. **41** (1988), no. 7, 909–996.

[37] L. Debnath and P. Mikusiński, *Hilbert Spaces with Applications*, Academic Press, 2005.

[38] M. Descoteaux, E. Angelino, S. Fitzgibbons, and R. Deriche, *Regularized, fast, and robust analytical Q-ball imaging*, Magnet. Reson. Med. **58** (2007), no. 3, 497–510.

[39] M. Descoteaux, R. Deriche, T. R. Knösche, and A. Anwander, *Deterministic and probabilistic tractography based on complex fibre orientation distributions*, IEEE Trans. Med. Imaging **28** (2009), no. 2, 269–286.

[40] D.L. Donoho, *Compressed sensing*, IEEE Trans. Inf. Theory **52** (2006), no. 4, 1289–1306.

[41] I.L. Dryden and K.V. Mardia, *Statistical Shape Analysis*, John Wiley & Sons, 1998.

[42] A. Edelman, T.A. Arias, and S.T. Smith, *The geometry of algorithms with orthogonality constraints*, SIAM J. Matrix Anal. **20** (1998), no. 2, 303–353.

[43] E. Esser, *Applications of Lagrangian-based alternating direction methods and connections to split Bregman*, UCLA CAM Reports **9** (2009).

[44] M. Fortin and R. Glowinski, *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-value Problems*, Elsevier, 2000.

[45] R. Glowinski and P. Le Tallec, *Augmented Lagrangian and Operator-splitting Methods in Nonlinear Mechanics*, SIAM, 1989.

[46] S.S. Goh, *Wavelet bases for Hilbert spaces of functions*, Complex Var. Elliptic Equ. **52** (2007), no. 2-3, 245–260.

[47] S.S. Goh and K.M. Teo, *Wavelet frames for Hilbert spaces of functions, in progress.*

[48] D. Goldfarb, Z. Wen, and W. Yin, *A curvilinear search method for p-harmonic flows on spheres*, SIAM J. Imaging Sci. **2** (2009), no. 1, 84–109.

[49] T. Goldstein and S. Osher, *The split Bregman method for L1-regularized problems*, SIAM J. Imaging Sci. **2** (2009), no. 2, 323–343.

[50] H. Jung, K. Sung, K.S. Nayak, E.Y. Kim, and J.C. Ye, *k-t FOCUSS: A general compressed sensing framework for high resolution dynamic MRI*, Magnet. Reson. Med. **61** (2009), no. 1, 103–116.

[51] K. Kurdyka, *On gradients of functions definable in o-minimal structures*, Annales de l'institut Fourier, 1998, pp. 769–783.

[52] R. Lai, J. Lu, and S. Osher, *Density matrix minimization with ℓ_1 regularization*, arXiv preprint arXiv:1403.1525 (2014).

[53] R. Lai and S. Osher, *A splitting method for orthogonality constrained problems*, J. Sci. Comput. **58** (2014), no. 2, 431–449.

[54] A.J. Laub, *Matrix Analysis for Scientists and Engineers*, SIAM, 2005.

[55] D. Liang, B. Liu, J. Wang, and L. Ying, *Accelerating SENSE using compressed sensing*, Magnet. Reson. Med. **62** (2009), no. 6, 1574–1584.

[56] S. Łojasiewicz, *Sur la géométrie semi-et sous-analytique*, Annales de l'institut Fourier, 1993, pp. 1575–1595.

[57] Z. Lu and Y. Zhang, *An augmented Lagrangian approach for sparse principal component analysis*, Math. Program. **135** (2012), no. 1, 149–193.

[58] M. Lustig, D. Donoho, and J.M. Pauly, *Sparse MRI: The application of compressed sensing for rapid MR imaging*, Magnet. Reson. Med. **58** (2007), no. 6, 1182–1195.

[59] S. Mallat, *Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R})$*, Trans. Amer. Math. Soc. **315** (1989), no. 1, 69–87.

[60] J. H. Manton, *Optimization algorithms exploiting unitary constraints*, IEEE Trans. Signal Process. **50** (2002), no. 3, 635–650.

[61] N. Marzari and D. Vanderbilt, *Maximally localized generalized Wannier functions for composite energy bands*, Phys. Rev. B **56** (1997), no. 20, 12847–12865.

[62] M. Meinshausen and B. Yu, *Lasso-type recovery of sparse representations for high-dimensional data*, Ann. Stat. **1** (2008), no. 37, 246–270.

[63] Y. Meyer, *Wavelets and Operators*, Vol. 1, Cambridge University Press, 1995.

[64] O. Michailovich and Y. Rathi, *On approximation of orientation distributions by means of spherical ridgelets*, IEEE Trans. Image Process. **19** (2010), no. 2, 461–477.

[65] O. Michailovich, Y. Rathi, and S. Dolui, *Spatially regularized compressed sensing for high angular resolution diffusion imaging*, IEEE Trans. Med. Imaging **30** (2011), no. 5, 1100–1115.

[66] J. Nocedal and S.J. Wright, *Numerical Optimization*, 2nd ed., Springer, 2006.

[67] S. Osher, Y. Mao, B. Dong, and W. Yin, *Fast linearized Bregman iteration for compressive sensing and sparse denoising*, Comm. Math. Sci. **8** (2010), no. 1, 93–111.

[68] S. Osher and K. Yin, *On the completeness of the compressed modes in the eigenspace*, UCLA CAM Reports: 13-62 (2013).

[69] V. Ozolins, R. Lai, R. Caflisch, and S. Osher, *Compressed modes for variational problems in mathematics and physics*, Proc. Nat. Acad. Sci. U.S.A. **110** (2013), no. 46, 18368–18373.

[70] ———, *Compressed plane waves - compactly supported multiresolution basis for the Laplace operator*, Proc. Nat. Acad. Sci. U.S.A. **111** (2014), no. 5, 1691–1696.

[71] I.C.S. Patarroyo, S. Dolui, O. Michailovich, and E. Vrscay, *Reconstruction of HARDI data using a split Bregman optimization approach*, Image Anal. Recognit., 2013, pp. 589–596.

[72] C. Pierpaoli, P. Jezzard, P.J. Basser, A. Barnett, and G. Di Chiro, *Diffusion tensor MR imaging of the human brain*, Radiology **201** (1996), no. 3, 637–648.

[73] E. Prodan and W. Kohn, *Nearsightedness of electronic matter*, Proc. Nat. Acad. Sci. U.S.A. **102** (2005), no. 33, 11635–11638.

[74] R.T. Rockafellar, R.J.-B. Wets, and M. Wets, *Variational Analysis*, Springer, 1998.

[75] A. Ron and Z. Shen, *Affine systems in $L_2(\mathbb{R}^d)$: The analysis of the analysis operator*, J. Funct. Anal. **148** (1997), no. 2, 408–447.

[76] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D **60** (1992), no. 1, 259–268.

[77] Z. Shen, *Wavelet Frames and Image Restorations*, Proc. ICM, 2010.

[78] E.O. Stejskal and J.E. Tanner, *Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient*, J. Chem. Phys. **42** (1965), no. 1, 288–292.

[79] J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis*, Springer, 2002.

[80] J. Tang and H. Liu, *Unsupervised feature selection for linked social media data*, Proc. 18th ACM SIGKDD International Conf. Knowl. Discov. Data Min., 2012, pp. 904–912.

[81] R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. Roy. Statist. Soc. Ser. B **1** (1996), no. 58, 267–288.

[82] D.S. Tuch, *Q-ball imaging*, Magnet. Reson. Med. **52** (2004), no. 6, 1358–1372.

[83] D.S. Tuch, R.M. Weisskoff, J.W. Belliveau, and V.J. Wedeen, *High angular resolution diffusion imaging of the human brain*, Proc. Ann. Meet. of ISMRM, 1999, pp. 321–321.

[84] V.J. Wedeen, T.G. Reese, D.S. Tuch, M.R. Weigel, J.G. Dou, R.M. Weiskoff, and D. Chessler, *Mapping fiber orientation spectra in cerebral white matter with Fourier-transform diffusion MRI*, Proc. Ann. Meet. ISMRM, 2000, pp. 82–82.

[85] Z. Wen, C. Yang, X. Liu, and Y. Zhang, *Trace-penalty minimization for large-scale eigenspace computation* (2013).

[86] Z. Wen and W. Yin, *A feasible method for optimization with orthogonality constraints*, Math. Program. **142** (2013), no. 1-2, 397–434.

[87] B. Yang, *Projection approximation subspace tracking*, IEEE Trans. Signal Proces. **43** (1995), no. 1, 95–107.

[88] Y. Yang, H.T. Shen, Z. Ma, Z. Huang, and X. Zhou, *$\ell_{2,1}$-norm regularized discriminative feature selection for unsupervised learning*, Proc. International Joint Conf. Art. Intell., 2011, pp. 1589–1594.

[89] X. Zhang, M. Burger, X. Bresson, and S. Osher, *Bregmanized nonlocal regularization for deconvolution and sparse reconstruction*, SIAM J. Imag. Sci. **3** (2010), no. 3, 253–276.

Appendix: Proofs of Theoretical Results in Chapter 3

(A) Proof of Corollary 3.1.2. We want to show that the modified spherical harmonics $\tilde{Y}_l^m$, $|m| \leq l$, $l \in 2\mathbb{N}_0$, are eigenfunctions of the FRT $\mathcal{R}$ with eigenvalues $2\pi P_l(0)$, i.e., for $l \in 2\mathbb{N}_0$, $|m| \leq l$,

$$\mathcal{R}[\tilde{Y}_l^m](\mathbf{u}) = 2\pi P_l(0)\tilde{Y}_l^m(\mathbf{u}), \quad u \in \mathbb{S}^2, \tag{1}$$

where $P_l(0)$ is the Legendre polynomial of degree l evaluated at 0, given explicitly in (3.1.11). Note that by (1.3.19), we have a relation between Y_l^m and $\tilde{Y}_l^m$: for $|m| \leq l$, $l \in 2\mathbb{N}_0$,

$$\tilde{Y}_l^m = \begin{cases} 2^{-1/2}[Y_l^m + (-1)^m Y_l^{-m}], & \text{if } m < 0, \\ Y_l^m, & \text{if } m = 0, \\ -2^{-1/2}\mathrm{i}[Y_l^m - (-1)^m Y_l^{-m}], & \text{if } m > 0, \end{cases}$$

which means that in order to establish (1), it suffices to show that the spherical harmonics Y_l^m, $|m| \leq l$, $l \in \mathbb{N}_0$, are eigenfunctions of the FRT $\mathcal{R}$, with eigenvalues $2\pi P_l(0)$, i.e., for $l \in \mathbb{N}_0$, $|m| \leq l$,

$$\mathcal{R}[Y_l^m](\mathbf{u}) = 2\pi P_l(0)Y_l^m(\mathbf{u}), \quad u \in \mathbb{S}^2, \tag{2}$$

We now show (2). Firstly, note that the FRT integral can be expressed as the following limit:

$$\mathcal{R}[Y_l^m](\mathbf{u}) = \lim_{n\to\infty} \int_{\mathbf{w}\in\mathbb{S}^2} \delta_n(\mathbf{u}^\top \mathbf{w}) Y_l^m(\mathbf{w}) d\mathbf{w},$$

where δ_n is a delta sequence defined by $\delta_n(x) := \frac{n}{\sqrt{\pi}} \exp(-n^2x^2)$ with

$$\lim_{n\to\infty} \int_{-1}^{1} \delta_n(x) f(x)\, dx = f(0), \quad f \in C[-1,1].$$

By the Funk-Hecke Theorem (Theorem 3.1.1), for $|m| \leq l$, $l \in \mathbb{N}_0$,

$$\begin{aligned} \mathcal{R}[Y_l^m](\mathbf{u}) &= \lim_{n\to\infty} \int_{\mathbf{w}\in\mathbb{S}^2} \delta_n(\mathbf{u}^\top \mathbf{w}) Y_l^m(\mathbf{w}) d\mathbf{w} \\ &= \lim_{n\to\infty} 2\pi \left[\int_{-1}^{1} \delta_n(t) P_l(t)\, dt \right] Y_l^m(\mathbf{u}) \\ &= 2\pi P_l(0) Y_l^m(\mathbf{u}). \end{aligned}$$

(B) Proof of Proposition 3.3.1. We want to prove (3.3.4): the sequence $\{P_{2N}(0)\}_{N\in\mathbb{N}}$ decays asymptotically at the rate of $N^{-1/2}$, i.e., there exist positive constants C_1 and C_2 such that

$$C_1 N^{-1/2} \leq |P_{2N}(0)| \leq C_2 N^{-1/2}, \quad N \in \mathbb{N}.$$

Note that from (3.1.11),

$$|P_{2N}(0)| = \prod_{n=1}^{N} \frac{(2n-1)}{2n}, \quad N \in \mathbb{N}.$$

Setting $a_n := \frac{1}{2n}$, $n \in \mathbb{N}$, and taking logarithms, we have

$$\log \prod_{n=1}^{N} \frac{(2n-1)}{2n} = \log \prod_{n=1}^{N} (1-a_n) = \sum_{n=1}^{N} \log(1-a_n).$$

With this choice of a_n, $n \in \mathbb{N}$, we have the following claim which will be shown later.

Claim (I): For any $N \in \mathbb{N}$,

$$\sum_{n=1}^{N}(-a_n) - \frac{1}{2} \leq \sum_{n=1}^{N} \log(1-a_n) \leq \sum_{n=1}^{N}(-a_n). \tag{3}$$

Note that (3) is equivalent to

$$\exp(-1/2) \cdot \exp\left\{\sum_{n=1}^{N}(-a_n)\right\} \leq \prod_{n=1}^{N}(1-a_n) \leq \exp\left\{\sum_{n=1}^{N}(-a_n)\right\}, \quad N \in \mathbb{N}. \tag{4}$$

Therefore, due to (4), in order to establish the desired result (3.3.4), we need to show that there exist positive constants C_1' and C_2' (independent of N) such that

$$C_1' N^{-1/2} \leq \exp\left\{\sum_{n=1}^{N}(-a_n)\right\} \leq C_2' N^{-1/2}, \quad N \in \mathbb{N}. \tag{5}$$

To show (5), we use the following result regarding the classical harmonic series:

$$\lim_{N\to\infty} \sum_{n=1}^{N} \left(\frac{1}{n}\right) - \log N = \gamma,$$

where γ is the Euler-Mascheroni constant. Therefore, there exist positive constants C_1'' and C_2'' such that

$$\log N + C_1'' \leq \sum_{n=1}^{N} \frac{1}{n} \leq \log N + C_2'', \quad N \in \mathbb{N}. \tag{6}$$

Since $a_n = \frac{1}{2n}$, $n \in \mathbb{N}$, using (6), it is not hard to see that (5) holds.

(C) Proof of Claim (I), i.e., (3). We begin the proof by recalling the Maclaurin's series

$$\log(1-x) = -x - \sum_{k=2}^{\infty} \frac{x^k}{k}, \quad -1 \leq x < 1. \tag{7}$$

Replacing x with $a_n = \frac{1}{2n}$, $n \in \mathbb{N}$, in (7) and summing (7) gives

$$\sum_{n=1}^{N} \log(1-a_n) = \sum_{n=1}^{N}(-a_n) - \sum_{n=1}^{N}\sum_{k=2}^{\infty} \frac{1}{k2^k} \cdot \frac{1}{n^k}, \quad N \in \mathbb{N}. \tag{8}$$

To prove (3) from (8), it suffices to show that

$$0 \leq \sum_{n=1}^{N} \sum_{k=2}^{\infty} \frac{1}{k2^k} \cdot \frac{1}{n^k} \leq \frac{1}{2}, \quad N \in \mathbb{N}. \tag{9}$$

The lower bound in (9) is obvious, so we concentrate on establishing the upper bound of (9).

Note that for $k \geq 2$, $f(x) := \frac{1}{x^k}$, $x > 0$, is a concave function satisfying

$$\sum_{n=2}^{N} f(n) \leq \int_{1}^{N} f(x)\, dx = \frac{1}{(k-1)}(1 - \frac{1}{N^{k-1}}) \leq 1, \tag{10}$$

for all $k \geq 2$ and $N \in \mathbb{N}$. Thus, from (10)

$$\sum_{n=1}^{N} \frac{1}{n^k} \leq 2,$$

which gives

$$\sum_{k=2}^{\infty} \frac{1}{k2^k} (\sum_{n=1}^{N} \frac{1}{n^k}) \leq 2 \sum_{k=2}^{\infty} \frac{1}{k2^k} \leq 2 \cdot (\frac{1}{2}) \sum_{k=2}^{\infty} \frac{1}{2^k} = \frac{1}{2}.$$

This completes the proof of Claim (I).

(D) Proof of Theorem 3.3.2. For $s \geq 0$, if $f \in H^s_{\mathrm{sym}}(\mathbb{S}^2)$, then

$$f = \sum_{l \in 2\mathbb{N}_0} \sum_{|m| \leq l} \tilde{f}_{m,l} \tilde{Y}_l^m,$$

where its modified spherical harmonic coefficients $\left\{\tilde{f}_{m,l}\right\}_{m,l}$ satisfy

$$\sum_{l \in 2\mathbb{N}_0} \sum_{|m| \leq l} (1 + l^2)^s \, |\tilde{f}_{m,l}|^2 < \infty. \tag{11}$$

It follows from Corollary 3.1.2 that

$$g := \mathcal{R}[f] = \sum_{l \in 2\mathbb{N}_0} \sum_{|m| \leq l} \tilde{g}_{m,l} \tilde{Y}_l^m,$$

where $\{\tilde{g}_{m,l}\}_{m,l}$ are the modified SH coefficients of g given as

$$\tilde{g}_{m,l} := 2\pi P_l(0)\, \tilde{f}_{m,l}, \quad |m| \leq l,\, l \in 2\mathbb{N}_0. \tag{12}$$

This series expansion is well defined because

$$\sum_{l \in 2\mathbb{N}_0} \sum_{|m| \leq l} (1 + l^2)^{s+1/2} |\tilde{g}_{m,l}|^2 < \infty,$$

which follows from (12), (11) and $\{P_{2N}(0)\}_{N \in \mathbb{N}} \sim O(N^{-1/2})$ from Proposition 3.3.1. Thus $g \in H^{s+1/2}_{\mathrm{sym}}(\mathbb{S}^2)$.

Next, consider an arbitrary $g \in H^{s+1/2}_{\mathrm{sym}}(\mathbb{S}^2)$, and define the mapping $\mathcal{R}^\dagger$ on g as

$$f^\dagger := \mathcal{R}^\dagger[g] = \sum_{l\in 2\mathbb{N}_0} \sum_{|m|\le l} \tilde{f}^\dagger_{m,l} \tilde{Y}^m_l,$$

where $\{\tilde{f}^\dagger_{m,l}\}_{m,l}$ are the modified SH coefficients of $f^\dagger$ given as

$$\tilde{f}^\dagger_{m,l} := [2\pi P_l(0)]^{-1}\, \tilde{g}_{m,l}$$

Since $g \in H^{s+1/2}_{\mathrm{sym}}(\mathbb{S}^2)$, i.e.,

$$\sum_{l\in 2\mathbb{N}_0} \sum_{|m|\le l} (1+l^2)^{s+1/2}\, |\tilde{g}_{m,l}|^2 < \infty.$$

and $\{[P_{2N}(0)]^{-1}\}_{N\in\mathbb{N}} \sim O(N^{1/2})$ from Proposition 3.3.1, we have

$$\sum_{l\in 2\mathbb{N}_0} \sum_{|m|\le l} (1+l^2)^{s} |\tilde{f}^\dagger_{m,l}|^2 < \infty.$$

Thus $f^\dagger \in H^s_{\mathrm{sym}}(\mathbb{S}^2)$.

Finally, combining what we have proved so far, we see that, given $s \ge 0$, $f \in H^s_{\mathrm{sym}}(\mathbb{S}^2)$ and $g \in H^{s+1/2}_{\mathrm{sym}}(\mathbb{S}^2)$, there holds $\mathcal{R}^\dagger\mathcal{R} f = f$ and $\mathcal{R}\mathcal{R}^\dagger g = g$. This shows that an inverse of $\mathcal{R}$ exists and $\mathcal{R}^{-1} = \mathcal{R}^\dagger$, completing the proof.