

Contents

MRA-Based Wavelet Frames and Applications	
BIN DONG AND ZUOWEI SHEN	3
MRA-Based Wavelet Frames and Applications	5
Introduction	5
Lecture 1. Multiresolution analysis	9
1. Definitions and basics	9
2. Density of the union of V_n	11
3. Triviality of the intersections of V_n	13
4. Approximation	16
Lecture 2. MRA-based tight wavelet frames	23
1. Extension principles	25
2. Quasi-affine systems and associated algorithms	44
3. Higher dimension tight frame systems	54
Lecture 3. Pseudo-splines and tight frames	59
1. Definitions and basics	59
2. Wavelets From pseudo-splines	69
3. Regularity of pseudo-splines	77
4. Two lemmata	90
Lecture 4. Frame based image restorations	97
1. Modeling	98
2. Balanced approach	103
3. Analysis based approach	123
Lecture 5. Other applications of frames	131
1. Background and models	131
2. Frame based blind deconvolution	137
3. Frame based image segmentation	140
4. Scene reconstruction from range data	143
Bibliography	149

MRA-Based Wavelet Frames and Applications

Bin Dong and Zuowei Shen

MRA-Based Wavelet Frames and Applications

Bin Dong and Zuowei Shen

Introduction

This collection of lecture notes is based on a series of lectures presented by the authors at the Institute for Advanced Studies/Park City Mathematics 2010 Summer Program on “The Mathematics of Image Processing”. The aim is to provide a concise, systematic and self contained overview on wavelet frames derived from a multiresolution analysis (MRA-based wavelet frames) and its applications in image analysis and restoration.

The publication of the unitary extension principle [158] in 1997 was the starting point of the MRA-based wavelet frames, especially the MRA-based tight wavelet frames. The unitary extension principle makes constructing wavelet frame systems with multiresolution structure painless. Wavelet frames derived from a multiresolution provide sparse approximations for piecewise smooth functions and have fast decomposition and reconstruction algorithms associated with them. These two important properties of MRA-based wavelet frames lead to many applications of wavelet frames in image processing. These lecture notes provide a detailed and comprehensive description of MRA-based wavelet frame theory and its applications.

Frame theory and its applications, notably the Gabor frames (see e.g. [61, 99, 132]) and wavelet frames (without a multiresolution structure) (see e.g. [61, 132]), were developed long before the discovery of the multiresolution analysis of [131, 134] and the systematic construction of the MRA-based compactly supported orthonormal wavelets of [60]. The concept of frame can be traced back to [80]. The rich literature of Gabor and wavelet frames provides a wide range of applications including time frequency analysis for signal processing, coherent state in quantum mechanics, filter bank design in electrical engineering, edge and singularity detection in image processing, and etc. It is not an intention of these notes to give a complete survey on frame theory and applications. Details on the earlier development and applications of Gabor and wavelet frames are provided in [61, 99, 132, 134, 135] and the references therein. Instead, the purpose here is to summarize some of contributions of the authors and their collaborators on the MRA-based wavelet frames and their applications. Special attention is given to applications of MRA-based tight wavelet frames in image restorations and analysis; a theme that is easily followed. We hope that these notes can be used for a course on this subject in future. A short note of such is given in [167].

Department of Mathematics, The University of Arizona, 617 N. Santa Rita Ave., Tucson, AZ, 85721-0089

E-mail address: dongbin@math.arizona.edu

Department of Mathematics, National University of Singapore, Singapore 119076

E-mail address: matzuows@nus.edu.sg

©2011 American Mathematical Society

MRA-based wavelet frames, especially, MRA-based wavelet tight frames, can be viewed as a generalization of the MRA-based orthonormal wavelet of [60, 131]. Since the publication of [60, 131], especially after the compactly supported MRA-based orthonormal wavelets being constructed by [60], wavelet analysis and its applications have been one of the most active research areas in applied mathematics over the past two decades. Algorithms and methods based on wavelet analysis have become powerful tools in image/signal processing and analysis. One well known application of MRA-based wavelets is image compression (see e.g. [132]) using the orthonormal or bi-orthogonal MRA-based wavelet bases of [56, 60]. Tight wavelet frames derived from over sampled orthonormal wavelet basis are already used in noise removal by [57, 77]. The publication of the unitary extension principle of [158] initiated a new wave of theoretical development, as well as exploration of new applications of MRA-based tight wavelet frames. Further theoretical developments on MRA-based wavelet frames can be found in e.g. [52, 65, 106] and the references therein. In order to be useful in applications, tight wavelet frames that have a multiresolution structure are preferred because this guarantees the existence of fast decomposition and reconstruction algorithms. Recently, MRA-based tight wavelet frames have been used in image inpainting, image denoising, image deblurring, blind deblurring, and image decompositions (see e.g. [14, 15, 16, 22, 23, 25, 32, 34, 36, 38]). More recently in [71], the MRA-based tight wavelet frame is used to develop algorithms for image segmentation. In some of applications mentioned above, the unitary extension principle is used to design a tight wavelet frame system adapted to the real life problems in hand, see, e.g. [14, 25, 32, 36, 38]. In other applications, simple tight wavelet frames derived from the unitary extension principle are used. Frame based algorithms for 3D surface reconstruction from scattered points and medical imaging (e.g. tomography) are currently being explored.

The community's effort to develop redundant wavelet systems that have sparse approximations for various classes of functions has led to the development of the MRA-based wavelet frames. The properties of redundancy, sparse approximation and the existence of fast decomposition and reconstruction algorithms are the key factors for the wide usage of wavelet frame systems. There are a few other redundant wavelet systems that have been developed and widely used in image/signal processing and analysis. Such redundant systems include, for example, bi-frames of [52, 65, 101, 157], bandlets of [132], ridgelets of [78], curvelets of [26, 27], and shearlets of [102, 123]. We forgo discussing any other redundant wavelet systems in order to have a clear focus in these lecture notes. Interested readers should consult the relevant references for further details.

The lecture notes are organized as follows. In Lecture 1, we present the multiresolution analysis, which lays the foundation of MRA-based wavelets and tight wavelet frames and provides a platform for multi-level analysis of images/signals. In Lecture 2, based on the MRA structure, we introduce the unitary extension principle of [158], which provides a general and convenient framework to construct MRA-based tight frames (framelets) of $L_2(\mathbb{R}^d)$. The MRA-based tight frames derived from the unitary extension principle have fast decomposition and reconstruction algorithms with the same complexity as convolutions. Details of fast algorithms for framelet decomposition and reconstruction are also given. In Lecture 3, we introduce a general class of refinable functions, called pseudo-splines [65, 165, 75],

which includes B-splines, Daubechies orthogonal refinable function and interpolatory refinable functions as special cases. Comprehensive regularity analysis of pseudo-splines, as well as the construction of (anti)symmetric pseudo-spline tight frames is provided. In Lecture 4, we present tight frame based models namely, balanced, synthesis based and analysis based approach, as well as the corresponding fast algorithms for image restoration problems, e.g. image deblurring, denoising and inpainting. In Lecture 5, we consider some other interesting and important applications of tight frames. We discuss the model proposed in [18] on blind deblurring (motion deblurring to be specific) problems. The major difficulty of blind deblurring problems is that the convolution kernel is unknown and needs to be solved simultaneously with the underlying images. Then, we present a frame based image segmentation model with a fast algorithm for the general image segmentation problems of [71]. At the end of this lecture, we discuss the model proposed by [112] on reconstruction of scenes (visible surfaces) from scattered, noisy and possibly sparse range data (point clouds). High-quality 3D scene modeling has long been an important research topic in computer vision, robotic navigation, computer graphics and animation.

Multiresolution analysis

The concept of multiresolution analysis (MRA) was first introduced by Mallat [131] and Meyer [134]. It is a general framework that makes constructing orthonormal wavelet bases for $L_2(\mathbb{R})$ very easy. The MRA-based compactly supported orthonormal wavelet systems were constructed by Daubechies [60]. Furthermore, the MRA structure grants fast implementation of wavelet decomposition and reconstruction which makes wavelets a very practical tool for image/signal processing and analysis.

1. Definitions and basics

Here, we adopt a more general MRA structure as proposed in [68], rather than the structure originally proposed by Mallat and Meyer.

For a given function $\phi \in L_2(\mathbb{R})$, we define the *shift-invariant subspace* $V(\phi) \subset L_2(\mathbb{R})$ generated by ϕ as

$$V(\phi) := \overline{\text{span}\{\phi(\cdot - k), k \in \mathbb{Z}\}},$$

and denote V_n as the 2^n -dilate of $V(\phi)$, i.e.

$$(1.1) \quad V_n = \overline{\text{span}\{\phi(2^n \cdot - k), k \in \mathbb{Z}\}}, n \in \mathbb{Z}.$$

We have $V = V_0$. A subspace $S \subset L_2(\mathbb{R})$ is called *translation-invariant* if for any $t \in \mathbb{R}$ and $f \in S$, we have $f(\cdot - t) \in S$. The subspace S is called *s-shift-invariant* if for any $k \in \mathbb{Z}$ and $f \in S$, we have $f(\cdot - sk) \in S$, and in particular if $s = 1$, we call S a *shift-invariant* subspace.

Now for a given sequence of subspaces $\{V_n\}_{n \in \mathbb{Z}}$, we say that $\{V_n\}$ forms a *multiresolution analysis (MRA)* for $L_2(\mathbb{R})$, if the following conditions are satisfied:

$$(1.2) \quad V_n \subset V_{n+1}, n \in \mathbb{Z};$$

$$(1.3) \quad \overline{\cup_n V_n} = L_2(\mathbb{R});$$

$$(1.4) \quad \cap_n V_n = \{0\}.$$

Here we call ϕ the *generator* of the MRA. The major tasks of Lecture 1 are to investigate which conditions ϕ must satisfy in order for the corresponding sequence of subspaces $\{V_n\}_{n \in \mathbb{Z}}$ to form an MRA, and to study the approximation properties of an MRA.

To make (1.2) hold, the generator $\phi \in L_2(\mathbb{R})$ must be in V_1 . Once $\phi \in V_1$, the shifts of ϕ are in V_1 as well since V_1 is $\frac{1}{2}$ -shift invariant. Therefore, to have $V_0 \subset V_1$, it is natural to require ϕ to be refinable. A function $\phi \in L_2(\mathbb{R})$ is said to be *refinable* if

$$(1.5) \quad \phi(x) = 2 \sum_{k \in \mathbb{Z}} h_0[k] \phi(2x - k)$$

for some $h_0 \in \ell_2(\mathbb{Z})$. The sequence h_0 is called the *refinement mask* of ϕ . In Fourier domain, the definition of refinability of ϕ can be rewritten as

$$(1.6) \quad \widehat{\phi}(2\cdot) = \widehat{h}_0 \widehat{\phi}.$$

Here $\widehat{\phi}$ denotes the Fourier transform of ϕ , and \widehat{h}_0 denotes the Fourier series of sequence h_0 . Recall that the *Fourier transform* of a function $f \in L_1(\mathbb{R})$ is defined by

$$\widehat{f}(\xi) := \int_{\mathbb{R}} f(t) e^{-i\xi t} dt, \quad \xi \in \mathbb{R},$$

which can be extended to more general function spaces (e.g. $L_2(\mathbb{R})$ or space of tempered distributions) in a natural manner. The *Fourier series* of a sequence $h_0 \in \ell_2(\mathbb{Z})$ is defined as

$$\widehat{h}_0(\xi) := \sum_{j \in \mathbb{Z}} h_0[j] e^{-ij\xi}, \quad \xi \in \mathbb{R}.$$

It follows from [69, Theorem 2.14] (also see [68]) that $\phi \in V_1$ whenever (1.6) holds with \widehat{h}_0 being a 2π periodic measurable function. Therefore, we generally do not need to require $h_0 \in \ell_2(\mathbb{Z})$. However, we still assume that $h_0 \in \ell_2(\mathbb{Z})$ for convenience. Indeed, when the unitary extension principle is applied to construct tight wavelet frames, we normally assume that ϕ and its shifts form a *Bessel sequence*, i.e. there exist $C < \infty$, such that for any sequence $b \in \ell_2(\mathbb{Z})$,

$$(1.7) \quad \left\| \sum_{k \in \mathbb{Z}} b[k] \phi(\cdot - k) \right\|_{L_2(\mathbb{R})} \leq C \|b\|_{\ell_2(\mathbb{Z})}.$$

When ϕ and its shifts form a Bessel sequence, it is easy to prove that $\phi \in V_1$ under the assumption that ϕ is refinable with its mask $h_0 \in \ell_2(\mathbb{Z})$. Finally, we remark that when $\phi \in L_2(\mathbb{R})$ is compactly supported, ϕ automatically satisfies (1.7) (see Lemma 3.1 in Lecture 3).

Examples of refinable functions include B-splines or more generally pseudo-splines. Here, we give examples of refinable B-splines, and postpone detailed discussion on pseudo-splines until Lecture 3. A (centered) B-spline with order m , denoted as B_m , is defined in the Fourier domain by

$$(1.8) \quad \widehat{B}_m(\xi) = e^{-ij\frac{\xi}{2}} \left(\frac{\sin(\xi/2)}{\xi/2} \right)^m,$$

where $j = 0$ when m is even, $j = 1$ when m is odd. The corresponding refinement mask of B_m is

$$\widehat{a}(\xi) = e^{-ij\frac{\xi}{2}} \cos^m(\xi/2).$$

The B-spline B_m is a compactly supported function in $C^{m-2}(\mathbb{R})$ with m as the length of its support. For detailed discussions about B-splines, one may refer to [67].

It turns out that we only need to place very mild requirements on $\phi \in L_2(\mathbb{R})$ for conditions (1.3) and (1.4) be satisfied. For example, any compactly supported refinable function $\phi \in L_2(\mathbb{R})$ with $\widehat{\phi}(0) \neq 0$ will generate an MRA. In particular, any B-spline of (1.8) generates an MRA. Details will be discussed in the next two Sections.

2. Density of the union of V_n

In the following discussions we denote $\phi_n := 2^{n/2}\phi(2^n \cdot)$. Then V_n , defined in (1.1), is the 2^{-n} -shift-invariant subspace generated by ϕ_n . We know from [69, Theorem 2.14] that a function s belongs to V_n if and only if there exists some $2^{n+1}\pi$ -period measurable function τ such that

$$(1.9) \quad \hat{s} = \tau \widehat{\phi_n},$$

and $\tau \widehat{\phi_n} \in L_2(\mathbb{R})$.

Before we give a complete characterization of the density property (1.3), we first recall two classical results of [163].

Proposition 1.1. ([163, Theorem 9.5]) *For any function $f \in L_2(\mathbb{R})$ and every $t \in \mathbb{R}$, let f_t be the translate of f defined by*

$$f_t(x) = f(x - t) \quad x \in \mathbb{R}.$$

Then the mapping

$$t \mapsto f_t$$

is a uniformly continuous mapping of \mathbb{R} into $L_2(\mathbb{R})$.

Proposition 1.2. ([163, Theorem 9.17] or [181] page 100) *Let X be a closed translation-invariant subspace of $L_2(\mathbb{R})$ with its Fourier transforms defined as*

$$\widehat{X} := \{\hat{f} : f \in X\}.$$

Then we have $\widehat{X} = L_2(\Omega)$ for some measurable set $\Omega \in \mathbb{R}$. Let X_1 and X_2 are two closed translation-invariant subspaces of $L_2(\mathbb{R})$ with $\widehat{X}_1 = L_2(\Omega_1)$ and $\widehat{X}_2 = L_2(\Omega_2)$. Then $X_1 = X_2$ if and only if $\Omega_1 = \Omega_2$ modulo a null-set.

Remark 1.1. In particular, when one takes the Ω_2 in Proposition 1.2 to be \mathbb{R} , then we have that $\widehat{X}_1 = L_2(\Omega_1)$ and $\widehat{X}_2 = L_2(\mathbb{R})$. Proposition 1.2 tells us that $X_1 = X_2$ if and only if $\Omega_1 = \mathbb{R}$ modulo a null-set, where we note that $X_2 = L_2(\mathbb{R})$.

Now, we begin with the following lemma of [68]

Lemma 1.1. *Let $\{V_n\}_{n \in \mathbb{Z}}$ be a nested sequence, i.e. (1.2) is satisfied. Then $\overline{\cup_n V_n}$ is a closed translation-invariant subspace of $L_2(\mathbb{R})$.*

PROOF. Let $X := \cup V_n$. Then \overline{X} is certainly closed. We first show that if $f \in X$, and then for any $t_0 \in \mathbb{R}$, $f_{t_0} \in \overline{X}$. We take $f \in X$. Since V_n is a nested sequence, i.e. $V_n \subset V_{n+1}$, then for n sufficiently large we have $f \in V_n$ (i.e. $\exists K > 0$, s.t. for all $n > K$, $f \in V_n$). Since V_n is 2^{-n} -shift-invariant, $f_t = f(\cdot + t)$ is in X for any $t = 2^{-n}k$, $n, k \in \mathbb{Z}$, which means that f_t is in X for all dyadic $t = 2^{-n}k$, $k, n \in \mathbb{Z}$. Proposition 1.1 tells us that translation is a continuous operation in $L_2(\mathbb{R})$, i.e. for any $\varepsilon > 0$, $\exists \delta > 0$ and $|s - t| < \delta$, such that $\|f_s - f_t\|_{L_2(\mathbb{R})} < \varepsilon$. It is also well known that the dyadic set $A := \{2^k j\}_{k, n \in \mathbb{Z}}$ is dense in \mathbb{R} , which means for any $t_0 \in \mathbb{R}$ we can choose $t \in A$ s.t. $|t - t_0| < \delta$ for any given $\delta > 0$. Then by the continuity of translation of L_2 functions, for any $t_0 \in \mathbb{R}$, we can always find $t \in A$ such that for any $\varepsilon > 0$, $\|f_t - f_{t_0}\|_{L_2(\mathbb{R})} < \varepsilon$. Since \overline{X} is closed, we have $f_{t_0} \in \overline{X}$.

Now, we show that for $g \in \overline{X}$ and $t_0 \in \mathbb{R}$, $g_{t_0} \in \overline{X}$. We take $g \in \overline{X}$. Then there always exists a function $f \in X$, s.t. $\|g - f\|_{L_2(\mathbb{R})} < \varepsilon$, $\forall \varepsilon > 0$. We take $t_0 \in \mathbb{R}$, and then the equality $\|g_{t_0} - f_{t_0}\|_{L_2(\mathbb{R})} = \|g - f\|_{L_2(\mathbb{R})}$ and the closeness of \overline{X} show that

$g_{t_0} \in \overline{X}$, since we have already proved that $f_{t_0} \in \overline{X}$. This concludes that $\overline{X} = \overline{\cup V_n}$ is a closed translation-invariant subspace of $L_2(\mathbb{R})$. ■

Lemma 1.1 gives us the translation-invariance of $\overline{\cup V_n}$ which allows us to use Proposition 1.2 to prove the important theorem below which is first proven in [68] (also see [115]).

Theorem 1.1. *Let $\{V_n\}_{n \in \mathbb{Z}}$ be a nested sequence, i.e. (1.2) is satisfied. Then $\overline{\cup V_n} = L_2(\mathbb{R})$ if and only if $\Omega_0 := \cup_n \text{supp } \widehat{\phi}_n = \mathbb{R}$ (modulo a null-set).*

PROOF. Let $X := \overline{\cup V_n}$. From Proposition 1.2 we have $\widehat{X} = L_2(\Omega)$ for some measurable set $\Omega \subset \mathbb{R}$. Moreover, Proposition 1.2 tells us that $X = L_2(\mathbb{R})$ if and only if $\Omega = \mathbb{R}$ modulo a null-set. We verify that $\Omega = \Omega_0$ modulo a null-set which will complete the proof. Since each ϕ_n is in X , we must have $\text{supp } \widehat{\phi}_n \subset \Omega$ modulo a null-set, and so, $\Omega_0 \subset \Omega$ modulo a null-set. Now we suppose that $\Omega \setminus \Omega_0$ contains a set Ω_1 of positive measure. From (1.9), we have $\text{supp } \widehat{s} \subset \text{supp } \widehat{\phi}_n$ for any $s \in V_n$. Since $\widehat{\phi}_n$ vanishes on Ω_1 for all $n \in \mathbb{Z}$, so does the Fourier transform of each element in V_n , $n \in \mathbb{Z}$. Hence, each element in $\cup V_n$ has Fourier transform which vanishes on Ω_1 . Lastly, if we show that by taking the closure, each element in X has Fourier transform which vanishes on Ω_1 , then it contradicts with the fact that $\widehat{X} = L_2(\Omega)$ contains $L_2(\Omega_1)$. Thus we will have $\Omega \setminus \Omega_0$ is of measure zero, which then implies that $\Omega = \Omega_0$ modulo a null-set.

Now take $f \in X$, there exists $g \in \cup V_n$ with \widehat{g} vanishes on Ω_1 , such that for any $\varepsilon > 0$,

$$\|f - g\|_{L_2(\mathbb{R})} < \varepsilon.$$

Applying Parseval's identity we have

$$\|f - g\|_{L_2(\mathbb{R})} = \frac{1}{\sqrt{2\pi}} \|\widehat{f} - \widehat{g}\|_{L_2(\mathbb{R})}.$$

By the fact that \widehat{g} vanishes on Ω_1 we have

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \|\widehat{f} - \widehat{g}\|_{L_2(\mathbb{R})} &\geq \frac{1}{\sqrt{2\pi}} \|\widehat{f} - \widehat{g}\|_{L_2(\Omega_1)} \geq \frac{1}{\sqrt{2\pi}} \|\widehat{f}\|_{L_2(\Omega_1)} - \frac{1}{\sqrt{2\pi}} \|\widehat{g}\|_{L_2(\Omega_1)} \\ &= \frac{1}{\sqrt{2\pi}} \|\widehat{f}\|_{L_2(\Omega_1)}. \end{aligned}$$

Thus we have

$$\frac{1}{\sqrt{2\pi}} \|\widehat{f}\|_{L_2(\Omega_1)} < \varepsilon.$$

Therefore \widehat{f} also vanishes on Ω_1 and this concludes the proof of this theorem. ■

Theorem 1.2. [115] *Let $\phi \in L_2(\mathbb{R})$ and V_n defined by (1.1). Assume that ϕ is refinable. Then,*

$$\overline{\cup V_n} = L_2(\mathbb{R})$$

if and only if

$$\cap_n 2^n Z(\widehat{\phi})$$

is a set of measure zero, where $Z(\widehat{\phi}) := \{\xi \in \mathbb{R} : \widehat{\phi}(\xi) = 0\}$.

PROOF. We first note that (1.2) holds because ϕ is refinable. Indeed, from the refinement equation (1.6), we have

$$\widehat{\phi}(2^{-n+1}\xi) = \widehat{h}_0(2^{-n}\xi)\widehat{\phi}(2^{-n}\xi).$$

Since \widehat{h}_0 is 2π -periodic, $\widehat{h}_0(2^{-n}\cdot)$ is $2^{n+1}\pi$ -periodic and hence $\phi(2^{n-1}\cdot) \in V_n$ by (1.9). Since V_n is 2^{-n} -shift-invariant which is obviously also 2^{-n+1} -shift-invariant, we have $\{\phi(2^{n-1}\cdot -k) : k \in \mathbb{Z}\} \subset V_n$ which implies $V_{n-1} \subset V_n$.

From Theorem 1.1 we know that $\overline{\cup V_n} = L_2(\mathbb{R})$ if and only if $\Omega_0 := \cup \text{supp } \widehat{\phi}_n = \mathbb{R}$ modulo a null-set. We will show that $\Omega_0 = \mathbb{R}$ modulo a null-set if and only if $m(\cap 2^n Z(\widehat{\phi})) = 0$ which will complete the proof. The fact that $\cup \text{supp } \widehat{\phi}_n = \mathbb{R}$ modulo a null-set is equivalent to

$$m\left(\left(\cup \text{supp } \widehat{\phi}_n\right)^c\right) = 0.$$

Moreover we have

$$\left(\cup \text{supp } \widehat{\phi}_n\right)^c = \cap (\text{supp } \widehat{\phi}_n)^c = \cap Z(\widehat{\phi}_n),$$

modulo a null-set. Last, we show that $Z(\widehat{\phi}_n) = 2^n Z(\widehat{\phi})$. We first note that $\widehat{\phi}_n = 0$ is equivalent to $\widehat{\phi}(2^{-n}\cdot) = 0$. Now, we take any $\xi_0 \in Z(\widehat{\phi}_n) = \{\xi \in \mathbb{R} : \widehat{\phi}(2^{-n}\xi) = 0\}$. We have $2^{-n}\xi_0 \in Z(\widehat{\phi})$ and that means $\xi_0 \in 2^n Z(\widehat{\phi})$. We take $\xi_0 \in 2^n Z(\widehat{\phi})$ which means that $\xi_0 = 2^n \xi_1$, where $\widehat{\phi}(\xi_1) = 0$. Since $\widehat{\phi}(2^{-n}\xi_0) = \widehat{\phi}(\xi_1) = 0$, we obtain $\xi_0 \in Z(\widehat{\phi}_n)$. This shows that $Z(\widehat{\phi}_n) = 2^n Z(\widehat{\phi})$. Now we conclude that

$$m\left(\left(\cup \text{supp } \widehat{\phi}_n\right)^c\right) = m\left(\cap 2^n Z(\widehat{\phi})\right).$$

Therefore, $\cup \text{supp } \widehat{\phi}_n = \mathbb{R}$ modulo a null-set if and only if $m\left(\cap 2^n Z(\widehat{\phi})\right) = 0$. This concludes the proof of the theorem. ■

Corollary 1.1. *Let $\phi \in L_2(\mathbb{R})$ and V_n defined by (1.1). Assume that $\{V_n\}_n$ is nested. If $\widehat{\phi}$ is non-zero almost everywhere (a.e.) in some neighborhood of the origin, we have $\overline{\cup V_n} = L_2(\mathbb{R})$. In particular, assume that $\widehat{\phi}$ is continuous at the origin and $\widehat{\phi}(0) \neq 0$. Then we have $\overline{\cup V_n} = L_2(\mathbb{R})$.*

PROOF. Since $\widehat{\phi}_n = 2^{-n/2}\widehat{\phi}(2^{-n}\cdot)$, $\widehat{\phi}$ being non-zero a.e. on Ω implies $\widehat{\phi}_n$ is non-zero on $2^n\Omega$. Now if Ω is some neighborhood of the origin, we obtain that $\cup \text{supp } \widehat{\phi}_n = \mathbb{R}$, since $\cup 2^n\Omega = \mathbb{R}$. By Theorem 1.1 we have $\overline{\cup V_n} = L_2(\mathbb{R})$. ■

Corollary 1.2. *Let $\phi \in L_2(\mathbb{R})$ with $\widehat{\phi}(0) = 1$ be a compactly supported refinable function, and V_n is defined as in (1.1). Then $\overline{\cup V_n} = L_2(\mathbb{R})$.*

PROOF. Since ϕ is compactly supported, the set of zero points of $\widehat{\phi}$, i.e. $Z(\widehat{\phi})$ is of measure zero. Then this corollary follows directly from Theorem 1.2. ■

3. Triviality of the intersections of V_n

In this section, we study which requirements placed on ϕ guarantee that (1.4) is satisfied. Let us first introduce some basic concepts and notations that are used in the last section of this lecture as well.

We define the *bracket product* of two functions $f, g \in L_2(\mathbb{R})$, denoted as $[f, g]$, as follows:

$$[f, g](\xi) := \sum_{k \in \mathbb{Z}} f(\xi + 2\pi k) \overline{g(\xi + 2\pi k)}.$$

It is easy to verify that if $f, g \in L_2(\mathbb{R})$, then $[f, g] \in L_1([-\pi, \pi])$, and the Cauchy-Schwartz inequality gives that

$$(1.10) \quad |[f, g]|^2 \leq [f, f][g, g],$$

where the right hand side is finite a.e. (see e.g. [114]). Inequality (1.10) shows that bracket products for functions in $L_2(\mathbb{R})$ are well defined.

Using the notation of bracket products, one can easily show that (see Lemma 3.1 in Lecture 3) the shifts of ϕ forming an orthonormal set, i.e.

$$\langle \phi, \phi(\cdot - k) \rangle = \delta_{k,0},$$

with $\delta_{k,0} = 0$ if $k \neq 0$ and $\delta_{0,0} = 1$, is equivalent to

$$[\widehat{\phi}, \widehat{\phi}] = 1 \quad \text{a.e.}$$

We need the following lemma to prove the main theorem of this section.

Lemma 1.2. *For a given $\phi \in L_2(\mathbb{R})$, one can always find $\psi \in L_2(\mathbb{R})$ such that $V(\phi) \subset V(\psi)$ and the shifts of ψ are orthonormal.*

PROOF. For given $\phi \in L_2(\mathbb{R})$, we define ψ as follows:

$$\widehat{\psi}(\xi) := \begin{cases} \frac{\widehat{\phi}(\xi)}{\sqrt{[\widehat{\phi}, \widehat{\phi}](\xi)}}, & \text{if } [\widehat{\phi}, \widehat{\phi}](\xi) > 0; \\ 1, & \text{if } [\widehat{\phi}, \widehat{\phi}](\xi) = 0 \text{ and } \xi \in [-\pi, \pi]; \\ 0, & \text{elsewhere.} \end{cases}$$

Since $[\widehat{\phi}, \widehat{\phi}] \in L_1(\mathbb{R})$, it is finite almost everywhere, and thus

$$[\widehat{\psi}, \widehat{\psi}](\xi) = 1 \quad \text{for almost all } \xi \in \mathbb{R}.$$

Hence $\psi \in L_2(\mathbb{R})$ and the shifts of ψ are orthonormal. Moreover,

$$\widehat{\phi} = \sqrt{[\widehat{\phi}, \widehat{\phi}]} \widehat{\psi}.$$

Then (1.9) implies that $\phi \in V(\psi)$, and hence $V(\phi) \subset V(\psi)$. \blacksquare

The following result was first proven in [68] and the proof given here is from [115]).

Theorem 1.3. *Given any $\phi \in L_2(\mathbb{R})$, we have $\bigcap_n V_n = \{0\}$.*

PROOF. By lemma 1.2, it suffices to prove the theorem for the case when $\{\phi(\cdot - k), k \in \mathbb{Z}\}$ is an orthonormal set. We note that whenever $\{\phi(\cdot - k), k \in \mathbb{Z}\}$ forms an orthonormal set, so does $\{2^{n/2}\phi(2^n \cdot - k), k \in \mathbb{Z}\}$. Here we denote $\phi_{n,k} := 2^{n/2}\phi(2^n \cdot - k)$.

Let P_n be the orthogonal projector from $L_2(\mathbb{R})$ onto V_n . The theorem will be established if we can prove

$$(1.11) \quad P_n f \rightarrow 0 \quad \text{as } n \rightarrow -\infty$$

for every $f \in L_2(\mathbb{R})$. Note that compactly supported functions are dense in $L_2(\mathbb{R})$, which means there always exists $\tilde{f} \in L_2(\mathbb{R})$ which is compactly supported and

$$\|f - \tilde{f}\|_{L_2(\mathbb{R})} < \varepsilon.$$

Hence suppose we have $P_n \tilde{f} \rightarrow 0$, $n \rightarrow -\infty$, then

$$\begin{aligned} \|P_n f\|_{L_2(\mathbb{R})} - \|P_n \tilde{f}\|_{L_2(\mathbb{R})} &\leq \|P_n f - P_n \tilde{f}\|_{L_2(\mathbb{R})} \\ &= \|P_n(f - \tilde{f})\|_{L_2(\mathbb{R})} \\ &\leq \|f - \tilde{f}\|_{L_2(\mathbb{R})} \leq \varepsilon \end{aligned}$$

which means that $P_n \tilde{f} \rightarrow 0$ implies $P_n f \rightarrow 0$. So it is sufficient to show that (1.11) is true for any compactly supported function $f \in L_2(\mathbb{R})$. Let f be such a function. Then $P_n f$ can be expressed as

$$P_n f = \sum_{k \in \mathbb{Z}} c_n[k] \phi_{n,k},$$

where

$$c_n[k] = \int_{\mathbb{R}} f(x) \overline{\phi_{n,k}(x)} dx$$

are sequences in $\ell_2(\mathbb{Z})$. Since the shifts of ϕ are orthonormal, we deduce from the equation of $P_n f$ that

$$\begin{aligned} \|P_n f\|_{L_2(\mathbb{R})}^2 &= \|c_n\|_{\ell_2(\mathbb{Z})}^2 \\ &= \sum_{k \in \mathbb{Z}} \left| \int_{\mathbb{R}} f(x) \overline{\phi_{n,k}(x)} dx \right|^2 \\ &\leq \sum_{k \in \mathbb{Z}} \left(\int_{\mathbb{R}} |f(x) \phi_{n,k}(x)| dx \right)^2 \\ &= \sum_{k \in \mathbb{Z}} \left(\int_{\mathbb{R}} |f(x) 2^{n/2} \phi(2^n x - k)| dx \right)^2. \end{aligned}$$

The first identity follows by Plancherel formula for the orthonormal basis $\{\phi_{n,k}\}_{k \in \mathbb{Z}}$. If f is supported in the interval $[-R, R]$, then by the Cauchy-Schwartz inequality we have,

$$\|P_n f\|_{L_2(\mathbb{R})}^2 \leq \|f\|_{L_2(\mathbb{R})}^2 \sum_{k \in \mathbb{Z}} \int_{\{2^n[-R, R] - k\}} |\phi(x)|^2 dx.$$

Now if $n < 0$ and $|n|$ is sufficiently large, we have that

$$\{2^n[-R, R] - k_1\} \cap \{2^n[-R, R] - k_2\} = \emptyset, \quad k_1 \neq k_2.$$

Therefore,

$$(1.12) \quad \|P_n f\|_{L_2(\mathbb{R})}^2 \leq \|f\|_{L_2(\mathbb{R})}^2 \int_{E_n} |\phi(x)|^2 dx = \|f\|_{L_2(\mathbb{R})}^2 \int_{\mathbb{R}} \chi_{E_n} |\phi(x)|^2 dx,$$

where

$$E_n = \cup_{k \in \mathbb{Z}} (2^n[-R, R] - k).$$

Since when $n \rightarrow -\infty$, $\chi_{E_n} |\phi(x)|^2 \rightarrow 0$ for almost every $x \in \mathbb{R}$, and $\chi_{E_n} |\phi(x)|^2 \leq |\phi(x)|^2 \in L_1(\mathbb{R})$, then by the dominated convergence theorem we have

$$\lim_{n \rightarrow -\infty} \|f\|_{L_2(\mathbb{R})}^2 \int_{\mathbb{R}} \chi_{E_n} |\phi(x)|^2 dx = \|f\|_{L_2(\mathbb{R})}^2 \int_{\mathbb{R}} \lim_{n \rightarrow -\infty} \chi_{E_n} |\phi(x)|^2 dx = 0.$$

Thus $\lim_{n \rightarrow -\infty} \|P_n f\|_{L_2(\mathbb{R})}^2 = 0$. \blacksquare

Now we summarize the results we get in this section and the previous one in the following theorem.

Theorem 1.4. *Let $\phi \in L_2(\mathbb{R})$, and V_n defined as in (1.1). Then,*

- (1) $\cap_n V_n = \{0\}$;
- (2) *Assume, in addition, that ϕ is refinable. Then,*

$$\overline{\cup_n V_n} = L_2(\mathbb{R})$$

if and only if

$$\cap_n 2^n Z(\hat{\phi})$$

is a set of measure zero, where $Z(\hat{\phi}) := \{\xi \in \mathbb{R} : \hat{\phi}(\xi) = 0\}$.

In particular, the sequence of subspaces $\{V_n\}_{n \in \mathbb{Z}}$ generated by $\phi \in L_2(\mathbb{R})$ forms an MRA if either of the following condition is satisfied:

- (1) *ϕ is a compactly supported refinable function with $\hat{\phi}(0) \neq 0$;*
- (2) *ϕ is refinable and $\hat{\phi}$ is continuous at 0 with $\hat{\phi}(0) \neq 0$.*

4. Approximation

In order to obtain a function in V_n that provides a good approximation to a given function f , one needs an approximation scheme. One of the commonly used approximation schemes is the quasi-interpolation scheme. There are many quasi-interpolatory schemes that give good approximations. We start by defining the quasi-interpolatory scheme that is used in MRA-based tight wavelet frame approximation.

For any function $f \in L_2(\mathbb{R})$, the dyadic dilation operator \mathcal{D} is defined by $\mathcal{D}f(x) := \sqrt{2}f(2x)$ and the translation operator T is defined by $T_a f(x) := f(x - a)$ for $a \in \mathbb{R}$. Given $j \in \mathbb{Z}$, we have $T_a \mathcal{D}^j = \mathcal{D}^j T_{2^j a}$. Define $\phi_{n,k} := \mathcal{D}^n T_k \phi$.

For a given $\phi \in L_2(\mathbb{R})$ that generates an MRA $\{V_n\}_n$, the *quasi-interpolatory operator* used here is defined as

$$(1.13) \quad \mathcal{P}_n : f \mapsto \sum_{k \in \mathbb{Z}} \langle f, \phi_{n,k} \rangle \phi_{n,k},$$

for an arbitrary $f \in L_2(\mathbb{R})$.

We first provide the following proposition which will become handy for our analysis.

Proposition 1.3. *Let $f, \phi \in L_2(\mathbb{R})$. Assume that $[\hat{\phi}, \hat{\phi}]$ is essentially bounded. Then*

- (1) *The shifts of ϕ form a Bessel system, i.e. (1.7) is satisfied. Furthermore, the Bessel bound, C in (1.7), is equal to $\|[\hat{\phi}, \hat{\phi}]\|_{L_\infty([-\pi, \pi])}^{\frac{1}{2}}$.*
- (2) *The 2π -periodic function $[\hat{f}, \hat{\phi}] \in L_1([-\pi, \pi])$ is the Fourier series of the sequence $\{\langle f, \phi(\cdot - k) \rangle : k \in \mathbb{Z}\}$. As a consequence,*

$$\sum_{k \in \mathbb{Z}} \langle f, \phi(\cdot - k) \rangle e^{-ik\xi} = [\hat{f}, \hat{\phi}](\xi)$$

almost everywhere.

(3) The operator \mathcal{P}_n is bounded with bound $\|\widehat{[\widehat{f}, \widehat{\phi}]}\|_{L_\infty([-\pi, \pi])}$. Furthermore, the Fourier transform of $\mathcal{P}_n f$ satisfies

$$(1.14) \quad \widehat{\mathcal{P}_n f} = [\widehat{f}(2^n \cdot), \widehat{\phi}](2^{-n} \cdot) \widehat{\phi}(2^{-n} \cdot)$$

PROOF. For (1), we denote $\tilde{\phi} := [\widehat{f}, \widehat{\phi}]^{\frac{1}{2}}$. The Plancherel theorem says that for all finite sequences $b \in \ell_2(\mathbb{Z})$ we have

$$2\pi \left\| \sum_{k \in \mathbb{Z}} b[k] \phi(\cdot - k) \right\|_{L_2(\mathbb{R})}^2 = \|\widehat{b\tilde{\phi}}\|_{L_2(\mathbb{R})}^2.$$

By breaking the real line into $\mathbb{R} = \cup_{k \in \mathbb{Z}} \{[-\pi, \pi] + 2k\pi\}$ and noticing that \widehat{b} is 2π -periodic, we have

$$\begin{aligned} \|\widehat{b\tilde{\phi}}\|_{L_2(\mathbb{R})}^2 &= \int_{\mathbb{R}} |\widehat{b}(\xi)|^2 |\widehat{\phi}(\xi)|^2 d\xi \\ &= \sum_{k \in \mathbb{Z}} \int_{[-\pi, \pi] + 2k\pi} |\widehat{b}(\xi)|^2 |\widehat{\phi}(\xi)|^2 d\xi \\ &= \sum_{k \in \mathbb{Z}} \int_{-\pi}^{\pi} |\widehat{b}(\xi)|^2 |\widehat{\phi}(\xi + 2k\pi)|^2 d\xi \\ &= \|\widehat{b\tilde{\phi}}\|_{L_2([-\pi, \pi])}^2. \end{aligned}$$

Therefore, we obtain the following identity

$$(1.15) \quad 2\pi \left\| \sum_{k \in \mathbb{Z}} b[k] \phi(\cdot - k) \right\|_{L_2(\mathbb{R})}^2 = \|\widehat{b\tilde{\phi}}\|_{L_2([-\pi, \pi])}^2.$$

Suppose $\tilde{\phi}$ is essentially bounded. Then (1.15) is satisfied for all $b \in \ell_2(\mathbb{Z})$. Therefore,

$$\|\widehat{b\tilde{\phi}}\|_{L_2([-\pi, \pi])} \leq C \|\widehat{b}\|_{L_2([-\pi, \pi])},$$

with $C = \|\tilde{\phi}\|_{L_\infty([-\pi, \pi])}$ and thus (1.7) follows.

For (2), using Fubini's theorem and the fact that $e^{ik\xi}$ is 2π -periodic, we have,

$$\begin{aligned} \int_{-\pi}^{\pi} [\widehat{f}, \widehat{\phi}](\xi) e^{ik\xi} d\xi &= \int_{-\pi}^{\pi} \sum_{k \in \mathbb{Z}} \widehat{f}(\xi + 2k\pi) \overline{\widehat{\phi}(\xi + 2k\pi)} e^{ik\xi} d\xi \\ &= \sum_{k \in \mathbb{Z}} \int_{-\pi}^{\pi} \widehat{f}(\xi + 2k\pi) \overline{\widehat{\phi}(\xi + 2k\pi)} e^{ik\xi} d\xi \\ &= \sum_{k \in \mathbb{Z}} \int_{[-\pi, \pi] + 2k\pi} \widehat{f}(\xi) \overline{\widehat{\phi}(\xi)} e^{ik\xi} d\xi \\ &= \int_{\mathbb{R}} \widehat{f}(\xi) \overline{\widehat{\phi}(\xi)} e^{ik\xi} d\xi, \end{aligned}$$

for any $k \in \mathbb{Z}$. Now by the Plancherel theorem, we have

$$\int_{\mathbb{R}} \widehat{f}(\xi) \overline{\widehat{\phi}(\xi)} e^{ik\xi} d\xi = \int_{\mathbb{R}} \widehat{f}(\xi) \widehat{\phi}(\xi) e^{-ik\xi} d\xi = 2\pi \int_{\mathbb{R}} f(x) \overline{\phi(x - k)} dx.$$

Therefore, we have

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} [\widehat{f}, \widehat{\phi}](\xi) e^{ik\xi} d\xi = \langle f, \phi(\cdot - k) \rangle.$$

This means that $[\widehat{f}, \widehat{\phi}](\xi)$ is the Fourier series of $\{\langle f, \phi(\cdot - k) \rangle : k \in \mathbb{Z}\}$.

For (3), the boundedness of \mathcal{P}_n follows directly from the facts that $\|\mathcal{P}_n\| = \|\mathcal{P}_0\|$ for all $n \in \mathbb{Z}$ and the shifts of ϕ form a Bessel system. Now, we show the identity (1.14) in part (3). Notice that $\mathcal{P}_n = \mathcal{D}^n \mathcal{P}_0 \mathcal{D}^{-n}$. Therefore, all we need to show is that

$$(1.16) \quad \widehat{\mathcal{P}_0 f} = [\widehat{f}, \widehat{\phi}] \widehat{\phi}.$$

Indeed, we have

$$\begin{aligned} \widehat{\mathcal{P}_0 f}(\xi) &= \int_{\mathbb{R}} \sum_{k \in \mathbb{Z}} \langle f, \phi(\cdot - k) \rangle \phi(x - k) e^{-i\xi x} dx \\ &= \sum_{k \in \mathbb{Z}} \langle f, \phi(\cdot - k) \rangle e^{-ik\xi} \int_{\mathbb{R}} \phi(x - k) e^{-i\xi(x-k)} dx \\ &= \sum_{k \in \mathbb{Z}} \langle f, \phi(\cdot - k) \rangle e^{-ik\xi} \widehat{\phi}(\xi) \\ &= [\widehat{f}, \widehat{\phi}](\xi) \widehat{\phi}(\xi), \end{aligned}$$

where the switching of orders of summation and integration is justified by the fact that the summation converges in the sense of $L_2(\mathbb{R})$ and Fourier transform is continuous on $L_2(\mathbb{R})$. ■

We say that \mathcal{P}_n provides approximation order m if, for all $f \in \mathcal{W}_2^m(\mathbb{R})$,

$$\|f - \mathcal{P}_n f\|_{L_2(\mathbb{R})} = O(2^{-nm}).$$

Here $\mathcal{W}_2^m(\mathbb{R})$, $m \in \mathbb{R}_+$, are Sobolev spaces defined by

$$\mathcal{W}_2^m(\mathbb{R}) := \{f \in L_2(\mathbb{R}) : \|f\|_{\mathcal{W}_2^m(\mathbb{R})} := \sqrt{2\pi} \|(1 + |\cdot|)^m \widehat{f}\|_{L_2(\mathbb{R})} < \infty\}.$$

The rest of this section is devoted to the analysis of approximation orders of \mathcal{P}_n . As one will see, the approximation order of \mathcal{P}_n is closely related to the property of $1 - |\widehat{\phi}|^2$ and $[\widehat{\phi}, \widehat{\phi}] - |\widehat{\phi}|^2$ at the origin.

In the following discussions, we denote $|f| = O(|\cdot|^m)$ whenever $\lim_{x \rightarrow 0} \frac{|f(x)|}{|x|^m} \leq C$ with $C > 0$. We say that a function has zero of order m if $f^{(j)}(0) = 0$ for $j = 0, 1, \dots, m-1$. It is clear that whenever $f \in C^m$ and it has zero of order m , then $|f| = O(|\cdot|^m)$. For convenience of arguments, whenever we say that f has zero of order m , we implicitly assume that $f \in C^m$ (although we only require the existence of $f^{(m-1)}$ in the original definition).

The following theorem, whose proof follows similar ideas of [69], shows the approximation order of the operator \mathcal{P}_n depends on the behavior of the function $[\widehat{\phi}, \widehat{\phi}] - |\widehat{\phi}|^2$ at the origin.

Theorem 1.5. *Let \mathcal{P}_n be defined by (1.13) for a given function $\phi \in L_2(\mathbb{R})$ satisfying*

- (1) $[\widehat{\phi}, \widehat{\phi}]$ is essentially bounded;
- (2) $[\widehat{\phi}, \widehat{\phi}] - |\widehat{\phi}|^2 = O(|\cdot|^{2m})$;
- (3) $1 - |\widehat{\phi}|^2 = O(|\cdot|^{2m_0})$.

Then, \mathcal{P}_n provides approximation order $m_1 = \min\{m, 2m_0\}$.

PROOF. Part (3) of Proposition 1.3 gives us

$$\begin{aligned}\widehat{\mathcal{P}_n f} &= [\widehat{f}(2^n \cdot), \widehat{\phi}](2^{-n} \cdot) \widehat{\phi}(2^{-n} \cdot) \\ &= \sum_{k \in \mathbb{Z}} \widehat{f}(\xi + 2^{n+1} k \pi) \overline{\widehat{\phi}(2^{-n} \xi + 2k\pi)} \widehat{\phi}(2^{-n} \xi).\end{aligned}$$

Then we have

$$\begin{aligned}\|\widehat{\mathcal{P}_n f} - \widehat{f}\|_{L_2(\mathbb{R})}^2 &= \int_{\mathbb{R}} \left| [\widehat{f}(2^n \cdot), \widehat{\phi}](2^{-n} \xi) \widehat{\phi}(2^{-n} \xi) - \widehat{f}(\xi) \right|^2 d\xi \\ &= \int_{\mathbb{R}} \left| [\widehat{f}(2^n \cdot), \widehat{\phi}](2^{-n} \xi) \right|^2 \left| \widehat{\phi}(2^{-n} \xi) \right|^2 d\xi \\ &\quad - \int_{\mathbb{R}} [\widehat{f}(2^n \cdot), \widehat{\phi}](2^{-n} \xi) \widehat{\phi}(2^{-n} \xi) \overline{\widehat{f}(\xi)} d\xi \\ &\quad - \int_{\mathbb{R}} \overline{[\widehat{f}(2^n \cdot), \widehat{\phi}](2^{-n} \xi) \widehat{\phi}(2^{-n} \xi)} \widehat{f}(\xi) d\xi + \|\widehat{f}\|_{L_2(\mathbb{R})}^2\end{aligned}$$

Since $[\widehat{f}(2^n \cdot), \widehat{\phi}](2^{-n} \xi)$ is $2^{n+1}\pi$ -periodic, we have

$$\begin{aligned}\int_{\mathbb{R}} \left| [\widehat{f}(2^n \cdot), \widehat{\phi}](2^{-n} \xi) \right|^2 \left| \widehat{\phi}(2^{-n} \xi) \right|^2 d\xi \\ &= \sum_{l \in \mathbb{Z}} \int_{2^n([- \pi, \pi] + 2l\pi)} \left| [\widehat{f}(2^n \cdot), \widehat{\phi}](2^{-n} \xi) \right|^2 \left| \widehat{\phi}(2^{-n} \xi) \right|^2 d\xi \\ &= \int_{2^n[- \pi, \pi]} \left| [\widehat{f}(2^n \cdot), \widehat{\phi}](2^{-n} \xi) \right|^2 [\widehat{\phi}, \widehat{\phi}](2^{-n} \xi) d\xi.\end{aligned}$$

Similarly, we have

$$\begin{aligned}\int_{\mathbb{R}} [\widehat{f}(2^n \cdot), \widehat{\phi}](2^{-n} \xi) \widehat{\phi}(2^{-n} \xi) \overline{\widehat{f}(\xi)} d\xi &= \int_{2^n[- \pi, \pi]} \left| [\widehat{f}(2^n \cdot), \widehat{\phi}](2^{-n} \xi) \right|^2 d\xi \\ &= \int_{\mathbb{R}} \overline{[\widehat{f}(2^n \cdot), \widehat{\phi}](2^{-n} \xi) \widehat{\phi}(2^{-n} \xi)} \widehat{f}(\xi) d\xi.\end{aligned}$$

Now altogether, we have

$$\begin{aligned}\|\widehat{\mathcal{P}_n f} - \widehat{f}\|_{L_2(\mathbb{R})}^2 &= \int_{2^n[- \pi, \pi]} \left| [\widehat{f}(2^n \cdot), \widehat{\phi}](2^{-n} \xi) \right|^2 \left([\widehat{\phi}, \widehat{\phi}](2^{-n} \xi) - 1 \right) d\xi \\ &\quad + \|\widehat{f}\|_{L_2(\mathbb{R})}^2 - \int_{2^n[- \pi, \pi]} \left| [\widehat{f}(2^n \cdot), \widehat{\phi}](2^{-n} \xi) \right|^2 d\xi.\end{aligned}$$

Now suppose $\text{supp } \widehat{f} \subset 2^n[- \pi, \pi]$. Then

$$\sum_{k \in \mathbb{Z}} \widehat{f}(\xi + 2^{n+1} k \pi) \overline{\widehat{\phi}(2^{-n} \xi + 2k\pi)} = \widehat{f}(\xi) \overline{\widehat{\phi}(2^{-n} \xi)}$$

on $2^n[-\pi, \pi]$. Thus,

$$\begin{aligned} \|\widehat{\mathcal{P}_n f} - \widehat{f}\|_{L_2(\mathbb{R})}^2 &= \int_{2^n[-\pi, \pi]} \left| \widehat{f}(\xi) \widehat{\phi}(2^{-n}\xi) \right|^2 \left([\widehat{\phi}, \widehat{\phi}](2^{-n}\xi) - 1 \right) d\xi \\ &\quad + \int_{2^n[-\pi, \pi]} |\widehat{f}(\xi)|^2 (1 - |\widehat{\phi}(2^{-n}\xi)|^2) d\xi \\ &= \int_{2^n[-\pi, \pi]} \left| \widehat{f}(\xi) \widehat{\phi}(2^{-n}\xi) \right|^2 \left([\widehat{\phi}, \widehat{\phi}](2^{-n}\xi) - |\widehat{\phi}(2^{-n}\xi)|^2 \right) d\xi \\ &\quad + \int_{2^n[-\pi, \pi]} |\widehat{f}(\xi)|^2 \left(|\widehat{\phi}(2^{-n}\xi)|^2 - 1 \right)^2 d\xi \end{aligned}$$

Since $[\widehat{\phi}, \widehat{\phi}]$ is essentially bounded, then when n is large enough, we have

$$\begin{aligned} \|\widehat{\mathcal{P}_n f} - \widehat{f}\|_{L_2(\mathbb{R})}^2 &\leq C \int_{2^n[-\pi, \pi]} |\widehat{f}(\xi)|^2 \left| [\widehat{\phi}, \widehat{\phi}](2^{-n}\xi) - |\widehat{\phi}(2^{-n}\xi)|^2 \right| d\xi \\ &\quad + \int_{2^n[-\pi, \pi]} |\widehat{f}(\xi)|^2 \left(|\widehat{\phi}(2^{-n}\xi)|^2 - 1 \right)^2 d\xi \\ &= C \int_{2^n[-\pi, \pi]} (1 + |\xi|)^{2\tilde{m}} |\widehat{f}(\xi)|^2 \frac{\left| [\widehat{\phi}, \widehat{\phi}](2^{-n}\xi) - |\widehat{\phi}(2^{-n}\xi)|^2 \right|}{(1 + |\xi|)^{2\tilde{m}}} d\xi \\ &\quad + \int_{2^n[-\pi, \pi]} (1 + |\xi|)^{2\tilde{m}} |\widehat{f}(\xi)|^2 \frac{\left(|\widehat{\phi}(2^{-n}\xi)|^2 - 1 \right)^2}{(1 + |\xi|)^{2\tilde{m}}} d\xi. \end{aligned}$$

Therefore, $\mathcal{P}_n f$ provides approximation order \tilde{m} for all $f \in W_2^{\tilde{m}}(\mathbb{R})$ with $\text{supp } \widehat{f} \subset 2^n[-\pi, \pi]$ whenever we have

$$\left\| \frac{[\widehat{\phi}, \widehat{\phi}](2^{-n}\xi) - |\widehat{\phi}(2^{-n}\xi)|^2}{(1 + |\xi|)^{2\tilde{m}}} \right\|_{L_\infty(2^n[-\pi, \pi])} \leq C 2^{-2n\tilde{m}},$$

and

$$\left\| \frac{\left(|\widehat{\phi}(2^{-n}\xi)|^2 - 1 \right)^2}{(1 + |\xi|)^{2\tilde{m}}} \right\|_{L_\infty(2^n[-\pi, \pi])} \leq C 2^{-2n\tilde{m}},$$

for n large enough. The above two inequalities are equivalent to

$$\left\| \frac{[\widehat{\phi}, \widehat{\phi}](\xi) - |\widehat{\phi}(\xi)|^2}{(2^{-n} + |\xi|)^{2\tilde{m}}} \right\|_{L_\infty([-\pi, \pi])} \leq C,$$

and

$$\left\| \frac{\left(|\widehat{\phi}(\xi)|^2 - 1 \right)^2}{(2^{-n} + |\xi|)^{2\tilde{m}}} \right\|_{L_\infty([-\pi, \pi])} \leq C,$$

which are all satisfied when $\tilde{m} = m_1 = \min\{m, 2m_0\}$.

So far we have established that for $f \in W_2^{m_1}(\mathbb{R})$ with $\text{supp } \widehat{f} \subset 2^n[-\pi, \pi]$, we have

$$\|\mathcal{P}_n f - f\|_{L_2(\mathbb{R})} \leq C 2^{-nm_1}.$$

Now consider general $f \in W_2^{m_1}(\mathbb{R})$ and define $\widehat{g} := \chi_{2^n[-\pi, \pi]} \widehat{f}$. Then we have

$$\begin{aligned} \|\mathcal{P}_n f - f\|_{L_2(\mathbb{R})} - \|\mathcal{P}_n g - g\|_{L_2(\mathbb{R})} &\leq \|\mathcal{P}_n f - f - \mathcal{P}_n g + g\|_{L_2(\mathbb{R})} \\ &\leq \|\mathcal{P}_n(f - g)\|_{L_2(\mathbb{R})} + \|f - g\|_{L_2(\mathbb{R})} \\ &\leq C \|\widehat{f} - \widehat{g}\|_{L_2(\mathbb{R})}, \end{aligned}$$

where the last inequality follows from the boundedness of operator \mathcal{P}_n because $[\widehat{\phi}, \widehat{\phi}]$ is essentially bounded (part (1) of Proposition 1.3). Since we have already established that $\|\mathcal{P}_n g - g\|_{L_2(\mathbb{R})} \leq C 2^{-nm_1}$, all we need to show is that $\|\widehat{f} - \widehat{g}\|_{L_2(\mathbb{R})} \leq C 2^{-nm_1}$ for every n . Indeed, we have

$$\begin{aligned} \|\widehat{f} - \widehat{g}\|_{L_2(\mathbb{R})}^2 &= \int_{\mathbb{R}} (1 - \chi_{2^n[-\pi, \pi]}(\xi)) |\widehat{f}(\xi)|^2 d\xi \\ &= 2^n \int_{\mathbb{R}} (1 - \chi_{[-\pi, \pi]}(\xi)) |\widehat{f}(2^n \xi)|^2 d\xi \\ &= 2^n \int_{\mathbb{R} \setminus [-\pi, \pi]} |\widehat{f}(2^n \xi)|^2 d\xi \\ &= 2^n \int_{\mathbb{R} \setminus [-\pi, \pi]} \frac{(1 + |2^n \xi|)^{2m_1} |\widehat{f}(2^n \xi)|^2}{(1 + |2^n \xi|)^{2m_1}} d\xi. \end{aligned}$$

Letting $\nu := (1 + |\cdot|)^{m_1} \widehat{f}$, we then have

$$\begin{aligned} \|\widehat{f} - \widehat{g}\|_{L_2(\mathbb{R})}^2 &= 2^{n(1-2m_1)} \int_{\mathbb{R} \setminus [-\pi, \pi]} \frac{|\nu(2^n \xi)|^2}{(2^{-n} + |\xi|)^{2m_1}} d\xi \\ &\leq C_1 2^{n(1-2m_1)} \int_{\mathbb{R} \setminus [-\pi, \pi]} |\nu(2^n \xi)|^2 d\xi \\ &= C_1 2^{-2nm_1} \int_{2^n(\mathbb{R} \setminus [-\pi, \pi])} |\nu(\xi)|^2 d\xi \\ &\leq C_2 2^{-2nm_1}. \end{aligned}$$

This concludes the proof of the theorem. \blacksquare

The approximation order of the quasi-interpolatory operator \mathcal{P}_n may not reach the maximal approximation order that space V_n provides. Recall that for a given shift-invariant subspace V of $L_2(\mathbb{R})$ and any function $f \in L_2(\mathbb{R})$, the *approximation error* of V to function f is defined as

$$E(f, V) := \min\{\|f - g\|_{L_2(\mathbb{R})} : g \in V\}.$$

We say that the shift-invariant subspace $V(\phi)$ *provides approximation order m* (or we say ϕ *provides approximation order m*) if, for every $f \in \mathcal{W}_2^m(\mathbb{R})$,

$$(1.17) \quad E(f, V_n) = O(2^{-nm}).$$

If $\{V_n\}_n$ generates an MRA, we say that the MRA provides approximation order m if $V(\phi)$ (or ϕ) provides an approximation order m .

When the shifts of ϕ form a Bessel system, we have $\mathcal{P}_n f \in V_n$. Therefore, the approximation order of $\mathcal{P}_n f$ can never exceed that of $V(\phi)$. The two approximation orders coincide when, for example, \mathcal{P}_n is an orthogonal projector.

It was shown in [69] that the approximation order of $V(\phi)$ is completely determined by the following function

$$(1.18) \quad \Lambda_\phi := \left(1 - \frac{|\widehat{\phi}|^2}{[\widehat{\phi}, \widehat{\phi}]}\right)^{1/2}, \quad \text{on } [-\pi, \pi].$$

As shown in the following theorem, the behavior of the function Λ_ϕ at the origin is crucial. The proof of the theorem is omitted here and we refer interested readers to the original paper [69] for details.

Theorem 1.6. [69] *The shift-invariant subspace $V(\phi)$ (or ϕ) provides approximation order $m > 0$ if and only if the function Λ_ϕ defined on the set $([-\pi, \pi])$ as*

$$(1.19) \quad \Lambda_\phi := \left(1 - \frac{|\widehat{\phi}|^2}{[\widehat{\phi}, \widehat{\phi}]}\right)^{1/2},$$

satisfies $|\cdot|^m \Lambda_\phi \in L_\infty([-\pi, \pi])$.

The condition (1.19) is closely related to the Strang-Fix conditions of ϕ . Recall that a function ϕ satisfies the Strang-Fix (SF) conditions of order $m \in \mathbb{N}_+$ if

$$(1.20) \quad \widehat{\phi}(0) \neq 0, \quad \widehat{\phi}^{(j)}(2\pi k) = 0, \quad j = 0, 1, 2, \dots, m-1, \quad k \in \mathbb{Z} \setminus \{0\}.$$

It is shown in [69] that under certain conditions on ϕ (e.g. if ϕ is compactly supported and $\widehat{\phi}(0) \neq 0$), (1.19) is equivalent to the SF conditions, meaning $V(\phi)$ (or ϕ) provides approximation order m if and only if ϕ satisfies SF conditions of order m .

When $\widehat{\phi}$ is smooth enough, e.g. when ϕ is compactly supported and $\widehat{\phi}(0) \neq 0$, assumption (2) of Theorem 1.5 is equivalent to the Strang-Fix (SF) conditions. In this case, we can easily see that assumption (2) of Theorem 1.5 ensures that the approximation order of $V(\phi)$ is m . If, in addition, m_0 in assumption (3) of Theorem 1.5 satisfies the inequality $2m_0 \geq m$, then Theorem 1.5 asserts that the operator \mathcal{P}_n attains the approximation order of the space V_n .

If ϕ is a compactly supported refinable function with a finitely supported mask, h_0 , satisfying $\widehat{h}_0(0) = 1$, it satisfies the SF conditions of order m . Hence ϕ provides an approximation order m whenever \widehat{h}_0 has a zero of order m at π . On the other hand, Proposition 2.1 in the next lecture will show that assumption (3) of Theorem 1.5, i.e. $1 - |\widehat{\phi}|^2 = O(|\cdot|^{2m_0})$, is equivalent to the condition $1 - |\widehat{h}_0|^2 = O(|\cdot|^{2m_0})$. Therefore, the approximation order of $V(\phi)$, with ϕ satisfying some additional mild conditions, is entirely determined by the refinement mask h_0 of ϕ . In fact, as we will see in the following two lectures, not only the approximation order of $V(\phi)$, but also the regularity of ϕ can be easily characterized by h_0 . Furthermore, thanks to the MRA structure and the unitary extension principle [158], the constructions of tight frames and the approximation properties of the tight frames can also be easily obtained from corresponding properties of h_0 .

MRA-based tight wavelet frames

This lecture is devoted to MRA-based tight wavelet frames. We first introduce the general notion of tight wavelet frames in space $L_2(\mathbb{R})$ with some other basic concepts and notations.

For any function $f \in L_2(\mathbb{R})$, the dyadic dilation operator \mathcal{D} is defined by $\mathcal{D}f(x) := \sqrt{2}f(2x)$ and the translation operator T is defined by $T_t f(x) := f(x - t)$ for $t \in \mathbb{R}$. Given $j \in \mathbb{Z}$, we have $T_t \mathcal{D}^j = \mathcal{D}^j T_{2^j t}$.

For given $\Psi := \{\psi_1, \dots, \psi_r\} \subset L_2(\mathbb{R})$, define the *wavelet system* (or *affine system*) as

$$(2.1) \quad X(\Psi) := \{\psi_{\ell, n, k} : 1 \leq \ell \leq r; n, k \in \mathbb{Z}\},$$

where $\psi_{\ell, n, k} = \mathcal{D}^n T_k \psi_\ell = 2^{n/2} \psi_\ell(2^n \cdot -k)$. The system $X(\Psi) \subset L_2(\mathbb{R})$ is called a *tight wavelet frame* of $L_2(\mathbb{R})$ if

$$\|f\|_{L_2(\mathbb{R})}^2 = \sum_{g \in X(\Psi)} |\langle f, g \rangle|^2,$$

holds for all $f \in L_2(\mathbb{R})$, where $\langle \cdot, \cdot \rangle$ is the inner product in $L_2(\mathbb{R})$ and $\|\cdot\|_{L_2(\mathbb{R})} = \sqrt{\langle \cdot, \cdot \rangle}$. This is equivalent to

$$f = \sum_{g \in X(\Psi)} \langle f, g \rangle g,$$

for all $f \in L_2(\mathbb{R})$. Here $\{\langle f, g \rangle\}_{g \in X(\Psi)}$ is called *canonical tight frame coefficients*.

It is clear that an orthonormal basis is a tight frame. When $X(\Psi)$ forms an orthonormal basis of $L_2(\mathbb{R})$, then $X(\Psi)$ is called an orthonormal wavelet basis. When $X(\Psi)$ forms a tight frame of $L_2(\mathbb{R})$, then $X(\Psi)$ is called a tight wavelet (affine) frame. We note that in some literature, the definition of tight frame we present is called the tight frame with bound one or Parseval frame.

More generally, the system $X(\Psi) \subset L_2(\mathbb{R})$ is called a *wavelet frame* of $L_2(\mathbb{R})$ if there are constants $0 < C_1 \leq C_2 < \infty$ such that

$$(2.2) \quad C_1 \|f\|_{L_2(\mathbb{R})}^2 \leq \sum_{g \in X(\Psi)} |\langle f, g \rangle|^2 \leq C_2 \|f\|_{L_2(\mathbb{R})}^2,$$

holds for all $f \in L_2(\mathbb{R})$. We call $X(\Psi)$ a Bessel system if only the right inequality holds.

Recall that for a given system $X(\Psi)$, the corresponding *frame operator* is defined by

$$Sf = \sum_{g \in X(\Psi)} \langle f, g \rangle g, \quad f \in L_2(\mathbb{R}).$$

It is clear that $X(\Psi)$ is a frame of $L_2(\mathbb{R})$ if and only if S is bounded and has a bounded inverse; $X(\Psi)$ is a tight frame of $L_2(\mathbb{R})$ if and only if S is the identity. For

comprehensive characterizations of frames in terms of various operators associated with them, the papers [156, 162] and the book [50] are good references.

The first interesting question is how to choose Ψ such that $X(\Psi)$ is a tight frame, or more generally a frame for $L_2(\mathbb{R})$. A general characterization of all wavelet frames whose dilation matrix is an integral (via dual Gramian analysis of [156, 162]) provided in [158] and derived from it a special characterization of tight wavelet frames in [158] and bi-frame in [157].

The dual Gramian analysis identifies the frame operator corresponding to the wavelet system $X(\Psi)$ as the dual Gramian matrix with each entry being written in terms of the Fourier transform of the generators Ψ . The dual Gramian analysis decomposes the operator S into a collection of simpler operators which are called fibers in [156, 162] in Fourier domain. The operator S is bounded and has a bounded inverse if and only if almost every fiber operator is uniformly bounded and has a uniformly bounded inverse. S is the identity operator if and only if almost every fiber operator is the identity operator. This leads to the conclusion that the wavelet system $X(\Psi)$ forms a tight frame of $L_2(\mathbb{R})$ if and only if the dual Gramian corresponding to the wavelet system $X(\Psi)$ is the identity operator almost everywhere. By writing each entry of the dual Gramian explicitly, one obtains the following theorem (see, e.g. Corollary 1.3 of [158]). For general results and dual Gramian analysis, interested readers should consult [158, 156, 162] for details. Note that the dual Gramian analysis was also applied to the Gabor frame analysis in [159] to derive the duality principle for the Gabor frames.

Theorem 2.1. [158] *The wavelet system $X(\Psi)$ is a tight frame of $L_2(\mathbb{R})$ if and only if the identities*

$$(2.3) \quad \sum_{\psi \in \Psi} \sum_{k \in \mathbb{Z}} |\widehat{\psi}(2^k \xi)|^2 = 1; \quad \sum_{\psi \in \Psi} \sum_{k=0}^{\infty} \widehat{\psi}(2^k \xi) \overline{\widehat{\psi}(2^k(\xi + (2j+1)2\pi))} = 0 \quad j \in \mathbb{Z}$$

hold for a.e. $\xi \in \mathbb{R}$. Furthermore, $X(\Psi)$ is an orthonormal basis of $L_2(\mathbb{R})$ if and only if (2.3) holds and $\|\psi\| = 1$ for all $\psi \in \Psi$.

There were many contributions over the last two decades to the study of the Bessel frame and other related properties of wavelet systems. Examples of univariate wavelet frames can be found in [63]; necessary and sufficient conditions for mother wavelets to generate frames were discussed (implicitly) in [61, 134]. Characterizations of univariate orthonormal bases associated with integer dilation were established independently in [98] and [108]. The multivariate counterparts of these results for the dyadic dilation appeared in [93]. Characterization of bi-frames (tight frame is a special case) in multivariate case for an integer dilation matrix was given in [101]. Independent of all these, as discussed before Theorem 2.1, a general characterization of all wavelet frames whose dilation matrix is an integral (via dual Gramian analysis) were provided in [158] and derived from it a special characterization of tight wavelet frames in [158] and bi-frame in [157].

Although Theorem 2.1 gives a complete characterization of the wavelet system $X(\Psi)$ being a tight frame of $L_2(\mathbb{R})$ and it helps to obtain tight wavelet frame systems with band-limited generators, directly applying this theorem provides little help in the construction of wavelet systems with compactly supported generators. On the other hand, because fast decomposition and reconstruction algorithms exist for compactly supported tight wavelet frames constructed from an MRA, these frames

are very handy to use in practice. This motivates the study of MRA-based tight wavelet frames in [158]. The key is the introduction of extension principles, e.g. the unitary extension principle. The MRA-based band-limited tight wavelet frames are also constructed in [6] without invoking extension principles.

1. Extension principles

In this section, we introduce the unitary extension principle (UEP), which leads to explicit constructions of tight wavelet frames based on the multiresolution analysis generated by a refinable function ϕ . In particular, examples of spline tight wavelet frame will be given. The approximation order of tight wavelet frames is studied and the fast algorithm for tight wavelet frame derived from the unitary extension principle is discussed as well. Also, the decomposition and reconstruction algorithms for the MRA-based tight wavelet frames are provided. We briefly introduce other extension principles as well. Although the main interest here is to construct compactly supported tight frames with finitely supported masks, we state the UEP in a more general setting. An even more general form of the UEP (e.g. multivariate case) can be found in [65, 158]).

1.1. Unitary extension principle

Let $\{V_n\}_{n \in \mathbb{Z}}$ be the MRA generated by the refinable function ϕ with refinement mask h_0 . The construction of tight frame systems starts with the construction of $\Psi \subset L_2(\mathbb{R})$. The objective of MRA-based construction of tight wavelet frames is to find $\Psi = \{\psi_1, \dots, \psi_r\} \subset V_1$ such that $X(\Psi)$ forms a 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 h_ℓ such that

$$(2.4) \quad \psi_\ell(x) = 2 \sum_{k \in \mathbb{Z}} h_\ell[k] \phi(2x - k).$$

The sequences h_1, \dots, h_r are called wavelet masks, or the high pass filters of the system; the refinement mask h_0 is also known as the low pass filter. In the Fourier domain, (2.4) can be written as

$$(2.5) \quad \widehat{\psi}_\ell(2\cdot) = \widehat{h}_\ell \widehat{\phi}, \quad \ell = 1, \dots, r,$$

where $\widehat{h}_1, \dots, \widehat{h}_r$ are 2π periodic functions. Throughout this lecture, ϕ and masks $\{h_\ell, \ell = 0, 1, \dots, r\}$ are assumed to satisfy the following mild assumptions:

Assumptions 2.1. All MRA-based constructions that are considered in this article are assumed to satisfy the following assumptions:

- (1) Each mask of $\{h_\ell : \ell = 0, 1, \dots, r\}$ is a sequence in $\ell_2(\mathbb{Z})$ and its Fourier series \widehat{h}_ℓ is measurable and (essentially) bounded.
- (2) The refinement mask h_0 of refinable function $\phi \in L_2(\mathbb{R})$ satisfies

$$|\widehat{h}_0(\xi) - 1| \leq C|\xi|.$$

- (3) The function $[\widehat{\phi}, \widehat{\phi}] := \sum_{k \in 2\pi\mathbb{Z}} |\widehat{\phi}(\cdot + k)|^2$ is essentially bounded, which is equivalent to (1.7) (see [70]), i.e. the shifts of ϕ form a Bessel system.

Under the assumption in item 2, the proof of Theorem 3.1 reveals the following property of ϕ

$$\lim_{\xi \rightarrow 0} \widehat{\phi}(\xi) = 1.$$

Our analysis in the previous lecture shows that the sequence of subspaces $\{V_n\}_n$ defined in (1.1) forms an MRA. When ϕ is a compactly supported refinable function generated by a finitely supported refinement mask with $\widehat{\phi}(0) = 1$, all conditions in Assumption 2.1 are satisfied. We note that item 2 in Assumption 2.1 is in fact not necessary and were not required in the original proof of UEP [158]. We include them here for the simplicity of the proof.

For the UEP in the most general setting with mild assumptions, interested readers should consult [65, 158] for the details. The univariate version of the UEP of [158] can be stated as following.

Theorem 2.2 (Unitary Extension Principle, (UEP) [158]). *Let $\phi \in L_2(\mathbb{R})$ be a refinable function with refinement mask h_0 and $\{h_1, \dots, h_r\}$ be a set of sequences. Assume that the refinable function ϕ and masks $\{h_0, h_1, \dots, h_r\}$ satisfy Assumption 2.1. Then the system $X(\Psi)$ where $\Psi = \{\psi_1, \dots, \psi_r\}$ defined in (2.4) forms a tight frame in $L_2(\mathbb{R})$ provided the equalities*

$$(2.6) \quad \sum_{\ell=0}^r |\widehat{h}_\ell(\xi)|^2 = 1 \quad \text{and} \quad \sum_{\ell=0}^r \widehat{h}_\ell(\xi) \overline{\widehat{h}_\ell(\xi + \pi)} = 0$$

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

$$(2.7) \quad \sigma(V_0) := \{\xi \in \mathbb{R} : [\widehat{\phi}, \widehat{\phi}](\xi) \neq 0\}.$$

Furthermore, assuming $r = 1$ and $\|\phi\|_{L_2(\mathbb{R})} = 1$, then $X(\Psi)$ is an orthonormal wavelet bases of $L_2(\mathbb{R})$.

First, we note that in order to use the UEP, the mask h_0 must necessarily satisfy

$$(2.8) \quad |\widehat{h}_0(\xi)|^2 + |\widehat{h}_0(\xi + \pi)|^2 \leq 1.$$

Indeed, the UEP condition (2.6) means that the following matrix has the two rows orthonormal to each other

$$\begin{pmatrix} \widehat{h}_0(\xi), & \widehat{h}_1(\xi), & \cdots, & \widehat{h}_r(\xi) \\ \widehat{h}_0(\xi + \pi), & \widehat{h}_1(\xi + \pi), & \cdots, & \widehat{h}_r(\xi + \pi), \end{pmatrix} \quad \text{i.e.} \quad \xi \in \mathbb{R}.$$

Then one can extend the above matrix to a $(r+1) \times (r+1)$ unitary matrix. The fact that the first column of the extended matrix has norm one implies the inequality (2.8).

Conditions in (2.6) can be written in terms of sequences $\{h_0, h_1, \dots, h_r\}$. The first condition becomes

$$(2.9) \quad \sum_{\ell=0}^r \sum_{k \in \mathbb{Z}} \overline{h_\ell[k]} h_\ell[k-p] = \delta_{p,0}, \quad p \in \mathbb{Z},$$

where $\delta_{p,0} = 1$ when $p = 0$ and 0 otherwise, and the second condition can be written as

$$(2.10) \quad \sum_{\ell=0}^r \sum_{k \in \mathbb{Z}} (-1)^{k-p} \overline{h_\ell[k]} h_\ell[k-p] = 0, \quad p \in \mathbb{Z}.$$

Before proving Theorem (2.2), we first prove the following Lemma 2.1. We will prove Lemma 2.1 in the Fourier domain which is similar to the proof of [65, Lemma 2.4]. Also, we demonstrate another proof of Lemma 2.1 in the time domain with some additional assumptions. The reason that we are showing the proof of

identity (2.12) in the time domain is because both the identity itself and the proof of it are the foundation of the fast decomposition and reconstruction algorithms for MRA-based wavelet frame systems.

For a given ϕ and its corresponding MRA $\{V_n\}_n$, the quasi-interpolatory operator is defined as (same as (1.13))

$$(2.11) \quad \mathcal{P}_n : f \mapsto \sum_{k \in \mathbb{Z}} \langle f, \phi_{n,k} \rangle \phi_{n,k},$$

for an arbitrary $f \in L_2(\mathbb{R})$.

Lemma 2.1. *Let $\phi \in L_2(\mathbb{R})$ be a refinable function with mask h_0 . Let*

$$\{h_0, h_1, \dots, h_r\}$$

be a set of sequences satisfying (2.6) (or equivalently (2.9) and (2.10)) and Assumption 2.1. Then

$$(2.12) \quad \mathcal{P}_n f = \mathcal{P}_{n-1} f + \sum_{\ell=1}^r \sum_{k \in \mathbb{Z}} \langle f, \psi_{\ell, n-1, k} \rangle \psi_{\ell, n-1, k}.$$

PROOF. To simplify notations, let ϕ be defined as ψ_0 , and

$$\mathcal{P}_{n,\ell} f := \sum_{k \in \mathbb{Z}} \langle f, \psi_{\ell, n, k} \rangle \psi_{\ell, n, k},$$

for $\ell = 0, 1, \dots, r$. Note that $\mathcal{P}_{n,0} = \mathcal{P}_n$. By Assumption 2.1, the system $\{\psi_{\ell, n, k}; k \in \mathbb{Z}\}$ for each ℓ and n , forms a Bessel system. Hence the operator $\mathcal{P}_{n,\ell} : L_2(\mathbb{R}) \mapsto L_2(\mathbb{R})$ is well-defined. Therefore, (2.12) is equivalent to

$$(2.13) \quad \widehat{\mathcal{P}_n f} = \sum_{\ell=0}^r \widehat{\mathcal{P}_{n-1,\ell} f}.$$

Since $\mathcal{P}_{n,\ell} = \mathcal{D}^n \mathcal{P}_{0,\ell} \mathcal{D}^{-n}$, we only need to show (2.13) for $n = 1$, i.e.

$$(2.14) \quad \widehat{\mathcal{P}_1 f} = \sum_{\ell=0}^r \widehat{\mathcal{P}_{0,\ell} f}.$$

By Proposition 1.3 item (3), (2.14) is equivalent to

$$(2.15) \quad [\widehat{f}(2\cdot), \widehat{\phi}(\frac{\cdot}{2})] \widehat{\phi}(\frac{\cdot}{2}) = \sum_{\ell=0}^r [\widehat{f}, \widehat{\psi}_\ell] \widehat{\psi}_\ell.$$

Indeed, since $\widehat{\psi}_\ell = \widehat{h}_\ell(\frac{\cdot}{2})\widehat{\phi}(\frac{\cdot}{2})$, we thus have

$$\begin{aligned}
\sum_{\ell=0}^r [\widehat{f}, \widehat{\psi}_\ell] \widehat{\psi}_\ell &= \sum_{\ell=0}^r [\widehat{f}, \widehat{h}_\ell(\frac{\cdot}{2})\widehat{\phi}(\frac{\cdot}{2})] \widehat{h}_\ell(\frac{\cdot}{2})\widehat{\phi}(\frac{\cdot}{2}) \\
&= \sum_{\ell=0}^r \left(\overline{\widehat{h}_\ell(\frac{\cdot}{2})} [\widehat{f}(2\cdot), \widehat{\phi}(\frac{\cdot}{2})] + \overline{\widehat{h}_\ell(\frac{\cdot}{2} + \pi)} [\widehat{f}(2\cdot), \widehat{\phi}(\frac{\cdot}{2} + \pi)] \right) \widehat{h}_\ell(\frac{\cdot}{2})\widehat{\phi}(\frac{\cdot}{2}) \\
&= \sum_{\ell=0}^r |\widehat{h}_\ell(\frac{\cdot}{2})|^2 [\widehat{f}(2\cdot), \widehat{\phi}(\frac{\cdot}{2})] \widehat{\phi}(\frac{\cdot}{2}) \\
&\quad + \sum_{\ell=0}^r \widehat{h}_\ell(\frac{\cdot}{2}) \overline{\widehat{h}_\ell(\frac{\cdot}{2} + \pi)} [\widehat{f}(2\cdot), \widehat{\phi}(\frac{\cdot}{2} + \pi)] \widehat{\phi}(\frac{\cdot}{2}) \\
&= [\widehat{f}(2\cdot), \widehat{\phi}(\frac{\cdot}{2})] \widehat{\phi}(\frac{\cdot}{2}).
\end{aligned}$$

The last identity above follows from (2.6) for $\frac{\xi}{2} \in \sigma(V_0)$ and the fact that $\widehat{\phi}(\frac{\xi}{2}) = 0$ when $\frac{\xi}{2} \notin \sigma(V_0)$. This concludes the proof of the lemma. \blacksquare

Remark 2.2. When Lemma 2.1 is proven in spacial domain, it reveals some details of the fast framelet decomposition and reconstruction algorithms given in Section 1.4. For simplicity, we assume that the refinable function and all framelets are compactly supported, and that their corresponding masks are finitely supported, i.e. we assume that ψ_ℓ , for $\ell = 0, 1, \dots, r$, are compactly supported and the masks h_ℓ are finite sequences; these assumptions are commonly satisfied in applications.

Since the system $\{\psi_{\ell,n,k}; k \in \mathbb{Z}\}$ for any $n \in \mathbb{Z}$, forms a Bessel system, the operators $\mathcal{P}_{n,\ell} : L_2(\mathbb{R}) \rightarrow L_2(\mathbb{R})$, for each ℓ , is bounded. Therefore, we only need to prove (2.12) for all f that are compactly supported.

By the definition of refinability of ϕ and (2.4), we have for $\ell = 0, 1, \dots, r$,

$$\psi_\ell(x) = 2 \sum_{k \in \mathbb{Z}} h_\ell[k] \phi(2x - k) = 2 \sum_{k \in \mathbb{Z}} h_\ell[k] \psi_0(2x - k).$$

Thus we have

$$\begin{aligned}
\psi_{\ell,n-1,k} &= 2^{(n-1)/2} \psi_\ell(2^{n-1} \cdot -k) \\
&= 2^{(n-1)/2+1} \sum_{k' \in \mathbb{Z}} h_\ell[k'] \psi_0(2^n \cdot -2k - k') \\
&= 2^{1/2} \sum_{k' \in \mathbb{Z}} h_\ell[k'] 2^{n/2} \psi_0(2^n \cdot -2k - k') \\
&= 2^{1/2} \sum_{k' \in \mathbb{Z}} h_\ell[k'] \psi_{0,n,2k+k'}
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
& \sum_{\ell=0}^r \sum_{k \in \mathbb{Z}} \langle f, \psi_{\ell, n-1, k} \rangle \psi_{\ell, n-1, k} \\
&= 2 \sum_{\ell=0}^r \sum_{k \in \mathbb{Z}} \left(\sum_{k' \in \mathbb{Z}} \overline{h_{\ell}[k']} \langle f, \psi_{0, n, 2k+k'} \rangle \right) \left(\sum_{k'' \in \mathbb{Z}} h_{\ell}[k''] \psi_{0, n, 2k+k''} \right) \\
&= 2 \sum_{\ell=0}^r \sum_{k \in \mathbb{Z}} \left(\sum_{j \in \mathbb{Z}} \overline{h_{\ell}[j-2k]} \langle f, \psi_{0, n, j} \rangle \right) \left(\sum_{j' \in \mathbb{Z}} h_{\ell}[j'-2k] \psi_{0, n, j'} \right) \\
&= 2 \sum_{j \in \mathbb{Z}} \sum_{j' \in \mathbb{Z}} \left(\sum_{\ell=0}^r \sum_{k \in \mathbb{Z}} \overline{h_{\ell}[j-2k]} h_{\ell}[j'-2k] \right) \langle f, \psi_{0, n, j} \rangle \psi_{0, n, j'},
\end{aligned}$$

where the switching of summations order is valid since all summations involved are finite sum.

Now we focus on proving that

$$(2.16) \quad \sum_{\ell=0}^r \sum_{k \in \mathbb{Z}} \overline{h_{\ell}[j-2k]} h_{\ell}[j'-2k] = \frac{1}{2} \delta_{j, j'}.$$

Suppose $j' - j$ is even, i.e. there exists $p \in \mathbb{Z}$ such that $j' = j + 2p$. Then we have

$$\begin{aligned}
\sum_{\ell=0}^r \sum_{k \in \mathbb{Z}} \overline{h_{\ell}[j-2k]} h_{\ell}[j'-2k] &= \sum_{\ell=0}^r \sum_{k \in \mathbb{Z}} \overline{h_{\ell}[2k+j]} h_{\ell}[2k+j'] \\
&= \sum_{\ell=0}^r \sum_{k \in \mathbb{Z}} \overline{h_{\ell}[2k+j]} h_{\ell}[2k+j+2p] \\
&= \begin{cases} \sum_{\ell=0}^r \sum_{k \in \mathbb{Z}} \overline{h_{\ell}[2k]} h_{\ell}[2k+2p], & j \text{ even} \\ \sum_{\ell=0}^r \sum_{k \in \mathbb{Z}} \overline{h_{\ell}[2k+1]} h_{\ell}[2k+2p+1], & j \text{ odd.} \end{cases} \\
&=: \begin{cases} c_1[p], & j \text{ even} \\ c_2[p], & j \text{ odd.} \end{cases}
\end{aligned}$$

Now replacing p by $-2p$ and breaking the summations (2.9) and (2.10) into even and odd parts, we have

$$\begin{aligned}
c_1[p] + c_2[p] &= \delta_{-2p, 0} = \delta_{p, 0} \\
c_1[p] - c_2[p] &= 0,
\end{aligned}$$

which leads to

$$c_1[p] = c_2[p] = \frac{1}{2} \delta_{p, 0}.$$

Since $j' - j = 2p$, we have $\delta_{p, 0} = \delta_{j, j'}$ which concludes that (2.16) is satisfied when $j' - j$ is even.

Suppose $j' - j$ is odd, which means there exists $p \in \mathbb{Z}$ such that $j' = j + 2p + 1$. Then we have

$$\begin{aligned} \sum_{\ell=0}^r \sum_{k \in \mathbb{Z}} \overline{h_\ell[j - 2k]} h_\ell[j' - 2k] &= \sum_{\ell=0}^r \sum_{k \in \mathbb{Z}} \overline{h_\ell[2k + j]} h_\ell[2k + j + 2p + 1] \\ &= \begin{cases} \sum_{\ell=0}^r \sum_{k \in \mathbb{Z}} \overline{h_\ell[2k]} h_\ell[2k + 2p + 1], & j \text{ even} \\ \sum_{\ell=0}^r \sum_{k \in \mathbb{Z}} \overline{h_\ell[2k + 1]} h_\ell[2k + 2p + 2], & j \text{ odd.} \end{cases} \\ &=: \begin{cases} c_1[p], & j \text{ even} \\ c_2[p], & j \text{ odd.} \end{cases} \end{aligned}$$

Now replacing p by $-2p - 1$ and breaking the summations (2.9) and (2.10) into even and odd parts, we have

$$\begin{aligned} c_1[p] + c_2[p] &= 0 \\ -c_1[p] + c_2[p] &= 0, \end{aligned}$$

which implies that $c_1[p] = c_2[p] = 0$ for all $p \in \mathbb{Z}$. Therefore, we have

$$\sum_{\ell=0}^r \sum_{k \in \mathbb{Z}} \overline{h_\ell[j - 2k]} h_\ell[j' - 2k] = 0 = \frac{1}{2} \delta_{j, j'}.$$

This concludes the proof of (2.16) and hence the identity (2.12) is proved.

Note that identity (2.12) essentially provides a decomposition and reconstruction formula for any function $f \in L_2(\mathbb{R})$ under the system $X(\Psi)$. This formula and the MRA structure generated by the underlying refinable function ϕ are used to prove that $X(\Psi)$ is a tight frame for $L_2(\mathbb{R})$.

The following two lemmata are necessary in order to prove the UEP. Lemma 2.2 below is in fact a direct consequence of $\cap_n V_n = \{0\}$ and the weakly compactness of the unit ball in $L_2(\mathbb{R})$. However, we provide a direct proof which is similar to the proof of $\cap_n V_n = \{0\}$ (Theorem 1.3).

Lemma 2.2. *Let refinable function $\phi \in L_2(\mathbb{R})$ satisfying Assumption 2.1 and operator \mathcal{P}_n be defined by (2.11). Then, for all $f \in L_2(\mathbb{R})$,*

$$\lim_{n \rightarrow -\infty} \mathcal{P}_n f = 0.$$

PROOF. Part (1) of Proposition 1.3 gives us

$$\begin{aligned} \|\mathcal{P}_n f\|_{L_2(\mathbb{R})}^2 &\leq C \sum_{k \in \mathbb{Z}} |\langle f, \phi_{n, k} \rangle|^2 \\ &= C \sum_{k \in \mathbb{Z}} \left(\int_{\mathbb{R}} |f(x) 2^{n/2} \phi(2^n x - k)| dx \right)^2. \end{aligned}$$

If f is supported in the interval $[-R, R]$, then by the Cauchy-Schwartz inequality we have,

$$\|\mathcal{P}_n f\|_{L_2(\mathbb{R})}^2 \leq C \|f\|_{L_2(\mathbb{R})}^2 \sum_{k \in \mathbb{Z}} \int_{\{2^n[-R, R] - k\}} |\phi(x)|^2 dx.$$

Now if $n < 0$ and $|n|$ is sufficiently large, we have that

$$\{2^n[-R, R] - k_1\} \cap \{2^n[-R, R] - k_2\} = \emptyset, \quad k_1 \neq k_2.$$

Therefore,

$$\|\mathcal{P}_n f\|_{L_2(\mathbb{R})}^2 \leq C \|f\|_{L_2(\mathbb{R})}^2 \int_{E_n} |\phi(x)|^2 dx = C \|f\|_{L_2(\mathbb{R})}^2 \int_{\mathbb{R}} \chi_{E_n} |\phi(x)|^2 dx,$$

where

$$E_n = \cup_{k \in \mathbb{Z}} (2^n[-R, R] - k).$$

Since when $n \rightarrow -\infty$, $\chi_{E_n} |\phi(x)|^2 \rightarrow 0$ for almost every $x \in \mathbb{R}$, and $\chi_{E_n} |\phi(x)|^2 \leq |\phi(x)|^2 \in L_1(\mathbb{R})$, then by the dominated convergence theorem we have

$$\lim_{n \rightarrow -\infty} \|f\|_{L_2(\mathbb{R})}^2 \int_{\mathbb{R}} \chi_{E_n} |\phi(x)|^2 dx = \|f\|_{L_2(\mathbb{R})}^2 \int_{\mathbb{R}} \lim_{n \rightarrow -\infty} \chi_{E_n} |\phi(x)|^2 dx = 0.$$

Thus

$$\lim_{n \rightarrow -\infty} \|\mathcal{P}_n f\|_{L_2(\mathbb{R})}^2 = 0,$$

for all compactly supported $f \in L_2(\mathbb{R})$. By a standard density argument and noting that \mathcal{P}_n is bounded for each n , we can conclude that $\lim_{n \rightarrow -\infty} \mathcal{P}_n f = 0$ for all $f \in L_2(\mathbb{R})$. ■

Lemma 2.3. *Let refinable function $\phi \in L_2(\mathbb{R})$ satisfy Assumption 2.1 and (2.8) and operator \mathcal{P}_n be defined by (2.11). Then for all $f \in L_2(\mathbb{R})$,*

$$\lim_{n \rightarrow \infty} \mathcal{P}_n f = f.$$

PROOF. We will prove this lemma by showing that for any $f \in L_2(\mathbb{R})$

$$(2.17) \quad \lim_{n \rightarrow \infty} \langle \mathcal{P}_n f, f \rangle = \|f\|_{L_2(\mathbb{R})}^2.$$

and for $n \in \mathbb{Z}$,

$$(2.18) \quad \|\mathcal{P}_n\|_{L_2(\mathbb{R})} \leq 1.$$

Once (2.17) and (2.18) are established, we will have

$$\begin{aligned} 0 &\leq \lim_{n \rightarrow \infty} \|\mathcal{P}_n f - f\|_{L_2(\mathbb{R})}^2 = \lim_{n \rightarrow \infty} \|\mathcal{P}_n f\|_{L_2(\mathbb{R})}^2 - 2\langle \mathcal{P}_n f, f \rangle + \|f\|_{L_2(\mathbb{R})}^2 \\ &= \lim_{n \rightarrow \infty} \|\mathcal{P}_n f\|_{L_2(\mathbb{R})}^2 - \|f\|_{L_2(\mathbb{R})}^2 \leq 0, \end{aligned}$$

which will conclude the proof of this lemma.

Now, we show (2.17), whose proof is similar to [50, Lemma 11.2.3]. Following the routine arguments in approximation theory, we only need to prove (2.17) for all $f \in L_2(\mathbb{R})$ with \widehat{f} continuous and compactly supported. By a similar proof of part (2) of Proposition 1.3, one can show that $[D^n \widehat{f}, \widehat{\phi}]$ is the Fourier series of $\{\langle f, \phi_{n,k} \rangle : k \in \mathbb{Z}\}$ for each $n \in \mathbb{Z}$. Since \widehat{f} is continuous and compactly supported, the infinite sum

$$[D^n \widehat{f}, \widehat{\phi}] = \sum_{k \in \mathbb{Z}} (D^n \widehat{f})(\xi + 2k\pi) \overline{\widehat{\phi}(\xi + 2k\pi)}$$

is in fact finite. Then, the boundedness of \widehat{f} implies that $[D^n \widehat{f}, \widehat{\phi}] \in L_2([-\pi, \pi])$. Now by Parseval's identity, we have

$$\langle \mathcal{P}_n f, f \rangle = \frac{1}{2\pi} \|[D^n \widehat{f}, \widehat{\phi}]\|_{L_2([-\pi, \pi])}^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \sum_{k \in \mathbb{Z}} (D^n \widehat{f})(\xi + 2k\pi) \overline{\widehat{\phi}(\xi + 2k\pi)} \right|^2 d\xi.$$

Since \widehat{f} is compactly supported, when n is large enough, we have

$$\langle \mathcal{P}_n f, f \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| (D^n \widehat{f})(\xi) \widehat{\phi}(\xi) \right|^2 d\xi.$$

Since item 2 of Assumption 2.1 implies that $\lim_{\xi \rightarrow 0} \widehat{\phi}(\xi) = 1$. Then, for any $\varepsilon > 0$, there exists $r \in (0, \pi)$ such that $1 - \varepsilon \leq |\widehat{\phi}(\xi)|^2 \leq 1 + \varepsilon$ for $\xi \in [-r, r]$. Furthermore, when n large enough, $D^n \widehat{f}$ has support in $[-r, r]$. Then we have,

$$\langle \mathcal{P}_n f, f \rangle = \frac{1}{2\pi} \int_{-r}^r \left| (D^n \widehat{f})(\xi) \widehat{\phi}(\xi) \right|^2 d\xi$$

and thus

$$\frac{1 - \varepsilon}{2\pi} \|D^n \widehat{f}\|_{L_2(\mathbb{R})}^2 \leq \langle \mathcal{P}_n f, f \rangle \leq \frac{1 + \varepsilon}{2\pi} \|D^n \widehat{f}\|_{L_2(\mathbb{R})}^2.$$

Now, (2.17) follows by noting that $\frac{1}{2\pi} \|D^n \widehat{f}\|_{L_2(\mathbb{R})}^2 = \|f\|_{L_2(\mathbb{R})}^2$.

From part (3) of Proposition 1.3, it is easy to see that it suffices to prove that $[\widehat{\phi}, \widehat{\phi}] \leq 1$ a.e. on $[-\pi, \pi]$. Define a sequence of functions $\{\phi_n\}_n$ as (known as the cascade algorithm)

$$\widehat{\phi}_n(\xi) := \widehat{h}_0\left(\frac{\xi}{2}\right) \widehat{\phi}_{n-1}\left(\frac{\xi}{2}\right) = \prod_{j=1}^n \widehat{h}_0(2^{-j}\xi) \widehat{\phi}_0(2^{-n}\xi),$$

with $\widehat{\phi}_0 := \chi_{[-\pi, \pi]}$. It is clear that $[\widehat{\phi}_0, \widehat{\phi}_0] \leq 1$, by induction, we have

$$\begin{aligned} [\widehat{\phi}_n, \widehat{\phi}_n](\xi) &= \sum_{k \in \mathbb{Z}} |\widehat{h}_0\left(\frac{\xi}{2} + k\pi\right)|^2 |\widehat{\phi}_{n-1}\left(\frac{\xi}{2} + k\pi\right)|^2 \\ &= \sum_{k \in \mathbb{Z}} |\widehat{h}_0\left(\frac{\xi}{2} + 2k\pi\right)|^2 |\widehat{\phi}_{n-1}\left(\frac{\xi}{2} + 2k\pi\right)|^2 \\ &\quad + \sum_{k \in \mathbb{Z}} |\widehat{h}_0\left(\frac{\xi}{2} + 2k\pi + \pi\right)|^2 |\widehat{\phi}_{n-1}\left(\frac{\xi}{2} + 2k\pi + \pi\right)|^2 \\ &= |\widehat{h}_0\left(\frac{\xi}{2}\right)|^2 \sum_{k \in \mathbb{Z}} |\widehat{\phi}_{n-1}\left(\frac{\xi}{2} + 2k\pi\right)|^2 \\ &\quad + |\widehat{h}_0\left(\frac{\xi}{2} + \pi\right)|^2 \sum_{k \in \mathbb{Z}} |\widehat{\phi}_{n-1}\left(\frac{\xi}{2} + 2k\pi + \pi\right)|^2 \\ &\leq |\widehat{h}_0(\xi/2)|^2 + |\widehat{h}_0(\xi/2 + \pi)|^2 \\ &\leq 1, \text{ a.e. } \xi \in \mathbb{R}. \end{aligned}$$

the last inequality follows from the fact that $|\widehat{h}_0(\xi/2)|^2 + |\widehat{h}_0(\xi/2 + \pi)|^2 \leq 1$. By item 2 of Assumption 2.1, we have that $\widehat{\phi}_n$ converges pointwise to $\widehat{\phi}$. (The proof of this argument is the same as that of Theorem 3.1. Although we assume in Theorem 3.1 that h_0 is finitely supported, only item 2 of Assumption 2.1 is necessary to show pointwise convergence of $\widehat{\phi}_n$ to $\widehat{\phi}$.) By Fatou's lemma, we have

$$[\widehat{\phi}, \widehat{\phi}] \leq \liminf_{n \rightarrow \infty} [\widehat{\phi}_n, \widehat{\phi}_n] \leq 1, \text{ a.e. } \xi \in \mathbb{R}$$

This concludes the proof of (2.18). \blacksquare

PROOF. (**Proof of Theorem 2.2**) By (2.12), we have

$$\mathcal{P}_n f = \mathcal{P}_{n-1} f + \sum_{\ell=1}^r \sum_{k \in \mathbb{Z}} \langle f, \psi_{\ell, n-1, k} \rangle \psi_{\ell, n-1, k}$$

Applying (2.12) inductively, one has that

$$\mathcal{P}_n f = \mathcal{P}_{n'} f + \sum_{\ell=1}^r \sum_{j=n'}^{n-1} \sum_{k \in \mathbb{Z}} \langle f, \psi_{\ell, j, k} \rangle \psi_{\ell, j, k}.$$

Letting $n' \rightarrow -\infty$ and employing Lemma 2.2, we have

$$(2.19) \quad \mathcal{P}_n f = \sum_{\ell=1}^r \sum_{j < n} \sum_{k \in \mathbb{Z}} \langle f, \psi_{\ell, j, k} \rangle \psi_{\ell, j, k}.$$

Then taking $n \rightarrow \infty$ at both sides of (2.19), noting that (2.8) holds by (2.6), and applying Lemma 2.3, we have, for all $f \in L_2(\mathbb{R})$,

$$f = \sum_{\ell=1}^r \sum_{n \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \langle f, \psi_{\ell, n, k} \rangle \psi_{\ell, n, k}.$$

This proves that $X(\Psi)$ is a tight frame for $L_2(\mathbb{R})$.

Finally, we show that if $r = 1$ and $\|\phi\|_{L_2(\mathbb{R})} = 1$ (or equivalently $\|\widehat{\phi}\|_{L_2(\mathbb{R})}^2 = 2\pi$), then $X(\Psi)$ forms an orthonormal basis for $L_2(\mathbb{R})$. Using (2.6), $\widehat{\phi}(2\cdot) = \widehat{h}_0 \widehat{\phi}$ and $\widehat{\psi}(2\cdot) = \widehat{h}_1 \widehat{\phi}$, we have

$$\begin{aligned} \|\widehat{\psi}\|_{L_2(\mathbb{R})}^2 &= 2 \int |\widehat{\psi}(2\xi)|^2 d\xi = 2 \int |\widehat{h}_1(\xi)|^2 |\widehat{\phi}(\xi)|^2 d\xi \\ &= 2 \int (1 - |\widehat{h}_0(\xi)|^2) |\widehat{\phi}(\xi)|^2 d\xi \\ &= 4\pi - 2 \int |\widehat{\phi}(2\xi)|^2 d\xi \\ &= 2\pi \end{aligned}$$

This proves that $\|\psi\|_{L_2(\mathbb{R})} = 1$ and hence $\|\psi_{n,k}\|_{L_2(\mathbb{R})} = 1$ for all $n, k \in \mathbb{Z}$. Since $X(\psi)$ is a tight frame for $L_2(\mathbb{R})$. Then for any atom $h \in X(\psi)$, we have

$$\|h\|_{L_2(\mathbb{R})}^2 = \sum_{g \in X(\psi)} |\langle h, g \rangle|^2 = \sum_{g \in X(\psi), g \neq h} |\langle h, g \rangle|^2 + \|h\|_{L_2(\mathbb{R})}^4,$$

which implies that

$$\sum_{g \in X(\psi), g \neq h} |\langle h, g \rangle|^2 = 0, \quad \text{for all } h \in X(\psi).$$

Therefore, $\langle h, g \rangle = 0$ whenever $h \neq g$ and hence $X(\psi)$ forms an orthonormal basis for $L_2(\mathbb{R})$. ■

In practice, it is common to decompose a given function down to a certain level instead of down to negative infinity. The following corollary states that if a function is decomposed to a finite level L , the system $\{\phi_{L,k}, \psi_{\ell,n,k} : 1 \leq \ell \leq r, n \geq L, k \in \mathbb{Z}\}$ forms a tight frame of $L_2(\mathbb{R})$. The proof of the corollary follows directly from (2.19) and that $X(\Psi)$ is a tight frame of $L_2(\mathbb{R})$.

Corollary 2.1. *Let $\Psi = \{\psi_\ell : 1 \leq \ell \leq r\}$ be the set of tight framelets constructed from the UEP with ϕ the corresponding refinable function. Then for any given $L \in \mathbb{Z}$, the system*

$$X(\phi, \Psi; L) := \{\phi_{L,k}, \psi_{\ell,n,k} : 1 \leq \ell \leq r, n \geq L, k \in \mathbb{Z}\}$$

forms a tight frame of $L_2(\mathbb{R})$, i.e. for any $f \in L_2(\mathbb{R})$,

$$f = \sum_{k \in \mathbb{Z}} \langle f, \phi_{L,k} \rangle \phi_{L,k} + \sum_{\ell=1}^r \sum_{n \geq L} \sum_{k \in \mathbb{Z}} \langle f, \psi_{\ell,n,k} \rangle \psi_{\ell,n,k}.$$

In general, we refer to the elements in Ψ as *mother wavelets*. If the generators in Ψ are constructed from a MRA, we call them *framelets*. For the special case $r = 1$, the above theorem is given in [124]. What makes constructing tight framelets painless is the ability to choose the number of generators r in the UEP. For example, one can easily construct tight framelets from splines. In fact, [158] gives a systematic construction of tight wavelet frame systems from B-splines by using the UEP. In the next section, we begin with a simple construction of tight frame from B-splines by using the UEP. We present more constructions of framelets from pseudo-splines using UEP in Lecture 3.

1.2. B-spline tight frames

Consider (centered) B-splines of order m . The corresponding refinement mask \widehat{h}_0 is given as $\widehat{h}_0(\xi) = e^{-ij\frac{\xi}{2}} \cos^m(\xi/2)$ with $j = 0$ when m is even and $j = 1$ when m is odd. We define m wavelet masks as

$$(2.20) \quad \widehat{h}_\ell(\xi) := -i^\ell e^{-ij\frac{\xi}{2}} \sqrt{\binom{m}{\ell}} \sin^\ell(\xi/2) \cos^{m-\ell}(\xi/2), \quad \ell = 1, 2, \dots, m.$$

Clearly, all assumptions in Assumption 2.1 are satisfied. Furthermore, we have

$$\sum_{\ell=0}^m |\widehat{h}_\ell(\xi)|^2 = (\cos^2(\xi/2) + \sin^2(\xi/2))^m = 1,$$

and

$$\sum_{\ell=0}^m \widehat{h}_\ell(\xi) \overline{\widehat{h}_\ell(\xi + \pi)} = e^{\frac{\pi}{2}ij} (\sin(\xi/2) \cos(\xi/2))^m (1 - 1)^m = 0.$$

Therefore, the m wavelets defined by

$$\widehat{\psi}_\ell := -i^\ell e^{-ij\frac{\xi}{2}} \sqrt{\binom{m}{\ell}} \frac{\cos^{m-\ell}(\xi/4) \sin^{m+\ell}(\xi/4)}{(\xi/4)^m}$$

generate a tight frame for $L_2(\mathbb{R})$. Note that each framelet ψ_ℓ is a real valued (anti)symmetric function supported in $[-(m+j)/2, (m+j)/2]$. The cases for $m = 2$ and $m = 4$ are given in the following two examples.

Example 2.1. Let $h_0 = [\frac{1}{4}, \frac{1}{2}, \frac{1}{4}]$ be the refinement mask of the piecewise linear B-spline $B_2(x) = \max(1 - |x|, 0)$. Define $h_1 = [-\frac{1}{4}, \frac{1}{2}, -\frac{1}{4}]$ and $h_2 = [\frac{\sqrt{2}}{4}, 0, -\frac{\sqrt{2}}{4}]$. Then h_0, h_1 and h_2 satisfy (2.9) and (2.10). Hence, the system $X(\Psi)$ where $\Psi = \{\psi_1, \psi_2\}$ defined in (2.4) by using h_1, h_2 and ϕ is a tight frame of $L_2(\mathbb{R})$ (see Figure 1).

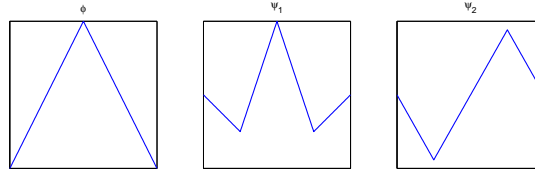


FIGURE 1. Piecewise linear refinable spline and corresponding framelets.

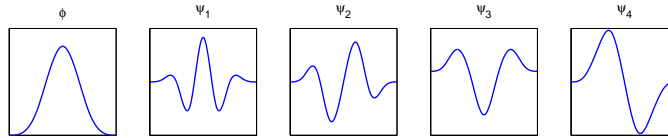


FIGURE 2. Piecewise cubic refinable B-spline and corresponding framelets.

Example 2.2. Let $h_0 = [\frac{1}{16}, \frac{1}{4}, \frac{3}{8}, \frac{1}{4}, \frac{1}{16}]$ be the refinement mask of piecewise cubic B-spline B_4 . Define h_1, h_2, h_3, h_4 as follows:

$$h_1 = [\frac{1}{16}, -\frac{1}{4}, \frac{3}{8}, -\frac{1}{4}, \frac{1}{16}], \quad h_2 = [-\frac{1}{8}, \frac{1}{4}, 0, -\frac{1}{4}, \frac{1}{8}],$$

$$h_3 = [\frac{\sqrt{6}}{16}, 0, -\frac{\sqrt{6}}{8}, 0, \frac{\sqrt{6}}{16}], \quad h_4 = [-\frac{1}{8}, -\frac{1}{4}, 0, \frac{1}{4}, \frac{1}{8}].$$

Then h_0, h_1, h_2, h_3, h_4 satisfy (2.9) and (2.6) and hence the system $X(\Psi)$ where $\Psi = \{\psi_1, \psi_2, \psi_3, \psi_4\}$ defined in (2.4) by h_1, h_2, h_3, h_4 and ϕ is a tight frame of $L_2(\mathbb{R})$ (see Figure 2).

An advantage of the tight wavelet frames derived from the UEP is that those systems have fast decomposition and reconstruction algorithms (see [65]), just as the orthonormal wavelet bases of [60]. Detailed discussions of decomposition and reconstruction algorithms are given in Section 1.4 for the univariate case and Section 3 for multivariate case.

1.3. Approximations

In this section, we focus on approximation properties of tight frame systems $X(\Psi)$ constructed from the UEP, its relations with approximation order of the underlying MRA, and vanishing moments of Ψ .

In Lecture 1 we introduced the concept of approximation order of MRA. Here, we introduce the concepts of vanishing moments and approximation orders of tight frame systems $X(\Psi)$ as follows.

Definition 2.1. Let $X(\Psi)$ be a tight frame system constructed from the UEP with the underlying MRA generated by the refinable function ϕ .

- (1) We say that *the tight frame system* $X(\Psi)$ *has vanishing moments of order* m_0 if $\widehat{\psi}_\ell$ has a zero of order m_0 at the origin, i.e. $|\widehat{\psi}_\ell| = O(|\cdot|^{m_0})$, for every $\ell = 1, 2, \dots, r$.
- (2) Define the *truncated operator* of the tight frame system $X(\Psi)$ as

$$(2.21) \quad \mathcal{Q}_n : f \mapsto \sum_{\psi \in \Psi, k \in \mathbb{Z}, j < n} \langle f, \psi_{j,k} \rangle \psi_{j,k}.$$

We say that *the tight frame system* $X(\psi)$ *provides approximation order* m_1 if, for all $f \in W_2^{m_1}(\mathbb{R})$,

$$\|f - \mathcal{Q}_n f\|_{L_2(\mathbb{R})} = O(2^{-nm_1}).$$

Since Ψ is constructed from the UEP with the underlying MRA generated by a refinable function ϕ , the order of vanishing moments can be characterized by the masks $\{\widehat{h}_\ell : \ell = 0, 1, \dots, r\}$, or $\widehat{\phi}$. This provides a very convenient way of analyzing the vanishing moments of Ψ because usually we do not have an explicit formula for Ψ . Now, we summarize this characterization in the following proposition.

Proposition 2.1. *Suppose Ψ is constructed from the UEP. Then it has vanishing moments of order m_0 if and only if one of the following three conditions is satisfied:*

- (1) $\sum_{\ell=1}^r |\widehat{h}_\ell|^2 = O(|\cdot|^{2m_0})$;
 (2) $1 - |\widehat{h}_0|^2 = O(|\cdot|^{2m_0})$;
 (3) $1 - |\widehat{\phi}|^2 = O(|\cdot|^{2m_0})$.

PROOF. Since Ψ is constructed from UEP, then by (2.6) we have

$$\sum_{\ell=1}^r |\widehat{h}_\ell(\cdot)|^2 = 1 - |\widehat{h}_0(\cdot)|^2.$$

Furthermore, since $\widehat{\psi}_\ell = \left(\widehat{h}_\ell \widehat{\phi}\right)(\cdot/2)$ and $\widehat{\phi}(0) = 1$, it is clear that $\sum_{\ell=1}^r |\widehat{h}_\ell(\cdot)|^2 = O(|\cdot|^{2m_0})$ if and only if $|\widehat{\psi}_\ell| = O(|\cdot|^{m_0})$ for all $\ell = 1, 2, \dots, r$, which establishes (1) and (2).

Let us consider (3). By the refinability of ϕ , we have

$$|\widehat{h}_0|^2 |\widehat{\phi}|^2 = |\widehat{\phi}(2\cdot)|^2.$$

Then

$$(2.22) \quad (1 - |\widehat{h}_0|^2) |\widehat{\phi}|^2 = |\widehat{\phi}|^2 - |\widehat{\phi}(2\cdot)|^2.$$

Suppose $1 - |\widehat{\phi}|^2 = O(|\cdot|^{\tilde{m}})$, i.e. $1 - |\widehat{\phi}|^2 = q + o(|\cdot|^{\tilde{m}})$ near the origin, with q some homogeneous polynomial of total degree \tilde{m} (here $\tilde{m} > 0$ because $\widehat{\phi}(0) = 1$). Then near the origin, we have

$$|\widehat{\phi}|^2 - |\widehat{\phi}(2\cdot)|^2 = q(2\cdot) - q(\cdot) + o(|\cdot|^{\tilde{m}}).$$

Since $q(2\cdot) - q(\cdot)$ is a nonzero homogeneous polynomial of total degree \tilde{m} , we see that $|\widehat{\phi}|^2 - |\widehat{\phi}(2\cdot)|^2 = O(|\cdot|^{\tilde{m}})$. Since $\widehat{\phi}(0) = 1$ and (2) gives us $1 - |\widehat{h}_0|^2 = O(|\cdot|^{2m_0})$, we have $\tilde{m} = 2m_0$. ■

Notice from (2.19) in the proof of Theorem 2.2 that if $X(\Psi)$ is constructed from the UEP with corresponding MRA generated by ϕ , we have

$$(2.23) \quad \mathcal{P}_n f = \mathcal{Q}_n f, \quad \text{for all } f \in L_2(\mathbb{R}),$$

where \mathcal{P}_n is the quasi-interpolatory operator defined by ϕ in (2.11). This means that the approximation order of tight frame system $X(\Psi)$ is the same as that of \mathcal{P}_n .

We note that when $X(\Psi)$ forms an orthonormal basis for $L_2(\mathbb{R})$, the set $\{\phi_{n,k} : k \in \mathbb{Z}\}$ forms an orthonormal basis of V_n , and hence \mathcal{Q}_n is an orthogonal projection onto V_n . Therefore, the approximation order of \mathcal{Q}_n , coincides with the approximation order of the underlying MRA. The same may not be true when $X(\Psi)$ is a tight frame for $L_2(\mathbb{R})$. In general, approximation order of \mathcal{Q}_n cannot exceed the approximation order provided by the underlying MRA.

The following theorem of [65] shows how the approximation order of the tight frame system $X(\Psi)$ depends on the refinable functions and framelets. The proof of it is a direct consequence of Theorem 1.5 and Proposition 2.1.

Theorem 2.3. [65] *Let $X(\Psi)$ be a tight frame system constructed from the UEP with underlying MRA generated by ϕ . Assume that Ψ has vanishing moments of order m_0 (equivalently, $1 - |\widehat{\phi}|^2 = O(|\cdot|^{2m_0})$ by Proposition 2.1) and that the refinable function ϕ satisfies $|\widehat{\phi, \phi} - |\widehat{\phi}|^2 = O(|\cdot|^{2m})$. Then the approximation order of the tight frame system is $m_1 = \min\{m, 2m_0\}$.*

It is easy to verify that for an arbitrary B-spline B_m , the order of the zero of $1 - |\widehat{B}_m|^2$ at the origin cannot exceed 2. Then Theorem 2.3 implies that the tight frame system $X(\Psi)$ constructed from B_m via the UEP (the corresponding masks are given by (2.20)) cannot have approximation order more than 2 (in fact, it is exactly 2 for all $m \geq 2$). Furthermore, Proposition 2.1 tells us that there is at least one wavelet among Ψ that has a vanishing moment of order 1. High order of approximation of a tight frame system $X(\Psi)$ gives good approximations to smooth functions. Furthermore, the high orders of vanishing moments of the framelets provide good sparse approximations to piecewise smooth functions. Hence, in order to have a good tight wavelet system, we need to have refinable functions whose Fourier transform are very flat at the origin. This leads to the introduction of pseudo-splines in [65, 75]. We revisit this issue with full details in Lecture 3.

Next, we briefly discuss how the order of vanishing moments plays a role in sparse approximation of piecewise smooth functions. The theory of the sparse approximation is based on nonlinear approximation and the characterization of various function spaces by a weighted norm of framelet coefficients. Here, we only provide some hubristic discussions. Interested readers should consult [9, 10, 106, 134] for comprehensive studies of this subject.

We start with the following proposition. It roughly says that when a wavelet function ψ has a high order vanishing moment, the majority of the corresponding wavelet coefficients are small. Hence, if all mother wavelets in Ψ have high vanishing moments, then the system $X(\Psi)$ can have a good sparse approximation to piecewise smooth functions. We note that a compactly supported wavelet ψ having order of vanishing moments m is equivalent to

$$\int x^k \psi(x) dx = 0, \quad \text{for all } 0 \leq k \leq m - 1.$$

Proposition 2.2. *If $\psi \in L_2(\mathbb{R})$ is a compactly supported function with order of vanishing moments m , then*

$$\langle f, \psi_{n,k} \rangle = O\left(\frac{1}{m!} 2^{-n(m+1/2)}\right)$$

for any $f \in L_2(\mathbb{R}) \cap C^m(\mathbb{R})$, when n is large enough.

PROOF. Without loss of generality, we assume $k = 0$. Then $\text{supp}\psi_{n,0} = 2^{-n}\text{supp}\psi \subseteq (-\varepsilon, \varepsilon)$ for some small $\varepsilon > 0$ when n is large enough. So, by $f \in C^m(\mathbb{R})$,

$$f(x) = f(0) + f'(0)x + \cdots + f^{(m-1)}(0)x^{m-1}/(m-1)! + f^{(m)}(\zeta_x)x^m/m!, \quad x \in (-\varepsilon, \varepsilon).$$

So, we have

$$\begin{aligned} \left| \int_{\mathbb{R}} f(x)\psi_{n,0}(x) dx \right| &= 2^{n/2} \left| \int_{\mathbb{R}} f(0)\psi(2^n x) dx + \int_{\mathbb{R}} f'(0)x\psi(2^n x) dx + \cdots \right. \\ &\quad \left. + \int_{\mathbb{R}} f^{(m-1)}(0) \frac{x^{m-1}}{(m-1)!} \psi(2^n x) dx + \int_{\mathbb{R}} f^{(m)}(\zeta_x) \frac{x^m}{m!} \psi(2^n x) dx \right| \\ &\leq \frac{2^{n/2}}{m!} \|f^{(m)}\|_{\infty} \int_{\mathbb{R}} |x|^m |\psi(2^n x)| dx \\ &= \frac{2^{-n(m+1/2)}}{m!} \|f^{(m)}\|_{\infty} \int_{\mathbb{R}} |x|^m |\psi(x)| dx \end{aligned}$$

which completes the proof. \blacksquare

Since ψ is compactly supported, the above proof can be modified to show that the majority of wavelet coefficients not near singularities can be small when f is a piecewise smooth function. This indicates that if all mother wavelets in Ψ have high orders of vanishing moments and short supports, then the system $X(\Psi)$ can have an ideally sparse approximation for piecewise smooth functions. However, higher orders of vanishing moments usually imply larger supports of Ψ , and shorter supports usually implies lower orders of vanishing moments of Ψ . Therefore, in practice, one should balance accordingly between the two properties as well as other properties like the regularity of Ψ , for each type of application.

1.4. Decomposition and reconstruction algorithms

The decomposition and reconstruction algorithms for MRA-based tight wavelet frames derived from the UEP is essentially the same as those of MRA-based orthonormal wavelets. Here, we assume that all masks used are finitely supported.

Since

$$\mathcal{P}_L f = D^L \mathcal{P}_0 D^{-L} f,$$

without loss of generality, one may use $\mathcal{P}_0 f \in V_0$ to approximate f . If it is necessary, we can always consider the function $f(2^{-L}\cdot)$ instead of f , since the approximation of a function f in space V_L is the same as that of the function $f(2^{-L}\cdot)$ in space V_0 .

When a tight wavelet frame is used, the given data is considered to be sampled as a local average $v[k] = \langle f, \phi(\cdot - k) \rangle$, which means that we can use

$$\mathcal{P}_0 f = \sum_{k \in \mathbb{Z}} v[k] \phi(\cdot - k)$$

to approximate the underlying function f . The accuracy of the approximation $\mathcal{P}_0 f$ to f is discussed in Theorem 1.5 which depends on the sampling function ϕ .

Given the sequence $h_\ell = \{h_\ell[k]\}_{k \in \mathbb{Z}}$ for any $\ell = 0, 1, \dots, r$, we define an infinite matrix H_ℓ which corresponds to h_ℓ as

$$H_\ell := (H_\ell[l, k]) := (\sqrt{2} \cdot \overline{h_\ell[k - 2l]}),$$

where the (l, k) th entry in H_ℓ is fully determined by the $(k - 2l)$ th entry in h_ℓ . Then for any $v \in \ell_2(\mathbb{Z})$, we have

$$(H_\ell v)[l] = \sqrt{2} \sum_{k \in \mathbb{Z}} \overline{h_\ell[k - 2l]} v[k].$$

Similarly, we can define H_ℓ^* , the adjoint of H_ℓ , as

$$(H_\ell^* v)[k] = \sqrt{2} \sum_{l \in \mathbb{Z}} h_\ell[k - 2l] v[l].$$

Let $\downarrow(\cdot)$ denote the downsampling operator which is defined as

$$\downarrow(v)[k] = v[2k], \quad k \in \mathbb{Z};$$

and $\uparrow(\cdot)$ denote the upsampling operator which is defined as

$$(\uparrow v)[k] = \begin{cases} v[k/2], & k \text{ even;} \\ 0, & k \text{ odd.} \end{cases}$$

Then we have

$$H_\ell v = \downarrow(\sqrt{2} \cdot \overline{h_\ell[\cdot]} * v) \quad \text{and} \quad H_\ell^* v = \sqrt{2} h_\ell * (\uparrow v).$$

In the literature of wavelets, it is traditional to use notations based on convolution with upsampling and downsampling.

We collect the coefficients in each level $n < 0$ to form an infinite column vector

$$v_{\ell, n} := [\dots, \langle f, \psi_{\ell, n, k} \rangle, \dots]^\top, \quad \ell = 0, 1, \dots, r,$$

where $\psi_0 := \phi$. From the proof of Lemma 2.1, we have shown that (identity (2.16))

$$\sum_{\ell=0}^r \sum_{k \in \mathbb{Z}} \overline{h_\ell[j - 2k]} h_\ell[j' - 2k] = \frac{1}{2} \delta_{j, j'},$$

which is equivalent to

$$(2.24) \quad \sum_{\ell=0}^r H_\ell^* H_\ell = I.$$

Thus, the decomposition and reconstruction process (2.12) can be written in the matrix form as

$$v_{0, n+1} = \sum_{\ell=0}^r H_\ell^* H_\ell v_{0, n+1}.$$

For a multiple level decomposition operator, we define \mathcal{A}_L , $L < 0$, as a (rectangular) block matrix given as:

$$(2.25) \quad \mathcal{A}_L := [H_0^{-L}; H_1 H_0^{-L-1}; \dots; H_r H_0^{-L-1}; \dots; H_1; \dots; H_r]^\top.$$

Then we have the reconstruction operator \mathcal{A}_L^* , the adjoint operator of \mathcal{A}_L , defined as

$$(2.26) \quad \mathcal{A}_L^* = [H_0^{*-L}; H_0^{*-L-1} H_1^*; \dots; H_0^{*-L-1} H_r^*; \dots; H_1^*; \dots; H_r^*]^\top.$$

Similar to (2.24), we have a multi-level perfect reconstruction formula $\mathcal{A}_L^* \mathcal{A}_L = I$, which is summarized in the following proposition; the proof is just straightforward calculation and is omitted.

Proposition 2.3. *The decomposition operator \mathcal{A}_L , as defined in (2.25) satisfies $\mathcal{A}_L^* \mathcal{A}_L = I$, where I is the identity operator.*

In practice, signals are of finite lengths. Therefore, we need a proper definition of operations $H_\ell v$ when v is of finite length. Now suppose $v = \{v[k]\}_{k=0}^{N-1}$. We define the periodic and symmetric extensions of v as

$$(2.27) \quad \tilde{v}[k] := \begin{cases} v[k \bmod N], & \text{periodic} \\ v_f[k \bmod 2N], & \text{symmetric,} \end{cases}$$

where v_f is defined as

$$v_f[k] := \begin{cases} v[k], & k = 0, 1, \dots, N-1 \\ v[2N-1-k], & k = N, N+1, \dots, 2N-1. \end{cases}$$

Therefore, whenever v is of finite length, we define

$$H_\ell v := H_\ell \tilde{v},$$

with \tilde{v} being one of the extensions in (2.27).

Note that when v is of finite length, the decomposition and reconstruction process can be realized as matrix multiplications (see e.g. [15, 35]). When \tilde{v} is a periodic or symmetric extension of v , the operation $H_\ell v$ can be understood as convolutions of v (followed by downsampling) with periodic or Neumann boundary conditions; where, the former is also known as circular convolution. In practice, one normally uses convolutions to implement the operation $H_\ell v$ instead of matrix multiplications due to efficiency considerations. A similar argument applies to $H_\ell^* v$ as well.

Algorithm 2.1 is the univariate fast framelet decomposition and reconstruction algorithm; Figure 3 is a diagram of the algorithm. For simplicity of notation, we denote $\tilde{h}_\ell := \sqrt{2} \cdot \overline{h_\ell[-\cdot]}$ and $\tilde{h}_\ell^* := \sqrt{2}h_\ell$. For a given finite signal $v \in \mathbb{R}^N$, we denote the finite length convolution operation as $\tilde{h}_\ell \otimes v$ (similarly for $\tilde{h}_\ell^* \otimes v$) which is understood as a restriction of $\tilde{h}_\ell * \tilde{v}$ on the set $\{0, 1, \dots, N-1\}$:

$$\tilde{h}_\ell \otimes v := (\tilde{h}_\ell * \tilde{v})|_{\{0, 1, \dots, N-1\}},$$

where \tilde{v} is either periodic or symmetric extension of v . Under this definition, $\tilde{h}_\ell \otimes v$ is a convolution of \tilde{h}_ℓ with the finite signal v using either periodic or Neumann boundary conditions.

Algorithm 2.1. *Given signal $v \in \mathbb{R}^N$ with N assumed to be an integer multiple of 2^L , $L \in \mathbb{N}_+$. Denote $v_{0,0} = v$. Then the L -level fast framelet decomposition and reconstruction are given as follows:*

- (1) **Decomposition:** For each $j = 1, 2, \dots, L$
 - (a) Obtain low frequency approximation to v at level j :

$$v_{0,j} = \downarrow (\tilde{h}_0 \otimes v_{0,j-1});$$

- (b) Obtain framelet coefficients of v at level j :

$$v_{\ell,j} = \downarrow (\tilde{h}_\ell \otimes v_{0,j-1}), \quad \ell = 1, 2, \dots, r.$$

- (2) **Reconstruction:** For each $j = L, L-1, \dots, 1$,

$$v_{0,j-1} = \sum_{\ell=0}^r \tilde{h}_\ell^* \otimes (\uparrow v_{\ell,j}).$$

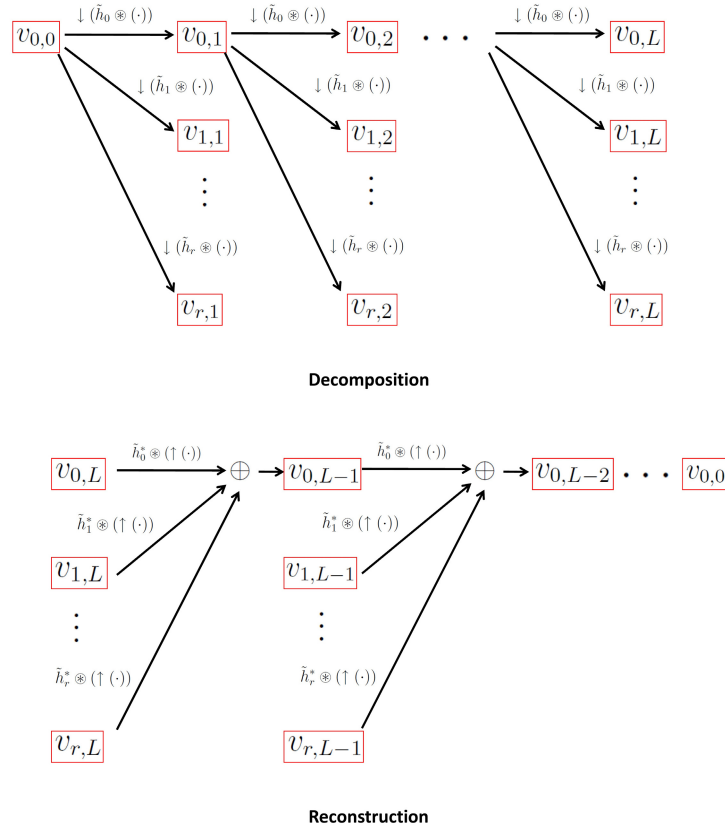


FIGURE 3. The above diagrams illustrate the fast framelet decomposition and reconstruction procedure given by Algorithm 2.1.

1.5. Other extension principles

the unitary extension principle of [158] was stated in a more general multivariate setting. In fact, the unitary extension principle is a consequence of a more general theorem on MRA-based tight wavelet frames. The details can be found in [158]. Since the publication of [158] in 1997, there are many generalizations of the unitary extension principle. Here, we briefly review some of them. The interested reader should consult the references mentioned below for further details.

The first generalization of the unitary extension principle is the oblique extension principle of [52, 65]. One of the major motivations to generalize the unitary extension principle is to obtain a spline tight wavelet system with better approximation power. As mentioned before, when the unitary extension principle is applied to construct tight wavelet frames from refinable spline functions, the approximation order of the corresponding truncated wavelet system cannot exceed 2; and there is at least one framelet that has vanishing moments of order 1.

The oblique extension principle can be used to obtain spline tight frame system whose truncated wavelet system has high approximation order and whose generators

have high order vanishing moments. The oblique extension principle is stated as following:

Proposition 2.4. (*The Oblique Extension Principle (OEP)*) *Suppose that the refinable function ϕ and the masks h_0, h_1, \dots, h_r satisfy Assumptions 2.1; and there exists a 2π -periodic function Θ that is non-negative, essentially bounded, continuous at the origin with $\Theta(0) = 1$. Assume that for $\xi \in \sigma(V_0)$, and $\xi + \pi \in \sigma(V_0)$, the following equalities*

$$(2.28) \quad \begin{aligned} & |\widehat{h}_0(\xi)|^2 \Theta(2\xi) + \sum_{\ell=1}^r |\widehat{h}_\ell(\xi)|^2 = \Theta(\xi); \\ & \widehat{h}_0(\xi) \widehat{h}_0(\xi + \pi) \Theta(2\xi) + \sum_{\ell=1}^r \widehat{h}_\ell(\xi) \widehat{h}_\ell(\xi + \pi) = 0. \end{aligned}$$

hold. Then the wavelet system $X(\Psi)$ defined by h_1, \dots, h_r is a tight wavelet frame.

PROOF. Setting $\theta := \Theta^{1/2}$, we define a function φ via $\widehat{\varphi} := \theta \widehat{\phi}$. Since θ is bounded, φ lies in $L_2(\mathbb{R})$. Consider the masks

$$\widehat{h}_0 := \frac{\theta(2\cdot) \widehat{h}_0}{\theta}, \quad \widehat{h}_i := \frac{\widehat{h}_i}{\theta}, \quad i = 1, \dots, r.$$

Applying (2.28), we obtain that $\sum_{i=0}^r |\widehat{h}_i(\xi)|^2 = 1, a.e.$ on $\sigma(V_0)$, hence $\widehat{h}_i, i = 0, 1, \dots, r$ is well-defined and bounded, and \widehat{h}_0 is the refinement mask of φ . Moreover, since Θ is continuous at 0 with $\Theta(0) = 1$, we obtain that $\widehat{\varphi}$ is continuous at 0 and $\widehat{\varphi}(0) = 1$. Now, apply the UEP (Theorem 2.2 and remarks after it) to \widehat{h}_i , and observe that the wavelets obtained from the masks \widehat{h}_i and φ are the same as the wavelets induced by \widehat{h}_i and ϕ . This concludes the proof of the theorem. ■

We note that the unitary extension principle can be viewed as a special case of the oblique extension principle by taking Θ to be 1.

Theorem 2.3 says that when the Fourier transform of the refinable function is not flat at the origin, one can chose a proper Θ which is flat at the origin. The vanishing moments of the resulting framelets have high order of the vanishing moment and the truncated tight wavelet system has a high approximation order. Detailed discussions can be found in [65]. This leads to many nice examples of spline tight wavelet frames with high order of vanishing moment and high approximation power in [52, 64, 65].

Example 2.3. Take $\widehat{h}_0(\xi) = (1 + e^{-i\xi})^4/16$ and

$$\Theta(\xi) = 2452/945 - 1657/840 \cos(\xi) + 44/105 \cos(2\xi) - 311/7560 \cos(3\xi).$$

Let

$$(2.29) \quad \begin{aligned} \widehat{h}_1(\xi) &= t_1 (1 - e^{-i\xi})^4 \left[1 + 8e^{-i\xi} + e^{-i2\xi} \right], \\ \widehat{h}_2(\xi) &= t_2 (1 - e^{-i\xi})^4 \left[1 + 8e^{-i\xi} + \left(\frac{7775}{4396} t - \frac{53854}{1099} \right) e^{-i2\xi} + 8e^{-i3\xi} + e^{-i4\xi} \right], \\ \widehat{h}_3(\xi) &= t_3 (1 - e^{-i\xi})^4 \left[1 + 8e^{-i\xi} + \left(21 + \frac{t}{8} \right) (e^{-i2\xi} + e^{-i4\xi}) \right. \\ &\quad \left. + t e^{-i3\xi} + 8e^{-i5\xi} + e^{-i6\xi} \right], \end{aligned}$$

where $t_3 = \sqrt{32655}/20160$, $t = 317784/7775 + 56\sqrt{16323699891}/2418025$, and

$$t_1 = \frac{\sqrt{11113747578360 - 245493856965 t}}{62697600}, \quad t_2 = \sqrt{1543080 - 32655 t}/40320.$$

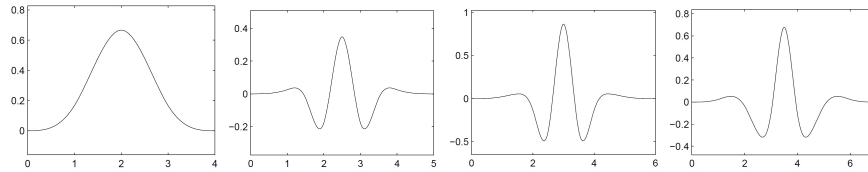


FIGURE 4. Figures from left to right are: refinable function ϕ , and framelets ψ_1 , ψ_2 and ψ_3 .

The above masks satisfy the OEP conditions, hence they lead to a tight wavelet frame. Here, all the wavelets have vanishing moments of order 4, i.e. $m_0 = 4$. Furthermore, the mother wavelets ψ_1, ψ_2, ψ_3 are symmetric. Note that for the above given ϕ the approximation order of the MRA is $m = 4$. Hence, the approximation order of the tight wavelet system is $4 = \min(m, 2m_0)$. The three filters $\{h_1, h_2, h_3\}$ are of size 7, 9, 11. The refinable function ϕ and the three framelets ψ_1, ψ_2 and ψ_3 are shown in Figure 4

Since it is impossible to obtain a tight wavelet frame that provides an arbitrary high approximation order from either the unitary extension principle or the oblique extension principle, a new approach is developed in [104]. This new approach of [104] starts with a non-stationary multiresolution analysis that has different refinable functions and refinement masks at different levels. A non-stationary version of the unitary extension principle is then established and the corresponding wavelet masks are obtained. For a different level, the corresponding set of wavelet masks is different, since the refinement mask is different. Hence, the wavelets and its masks vary by levels. By a proper choice of pseudo-spline masks, symmetric C^∞ real-valued tight wavelet frames in $L_2(\mathbb{R})$ with compact support and the spectral frame approximation order are obtained in [104]. Furthermore, these tight wavelet frames are used to characterize Sobolev spaces with arbitrary regularity in [105]. We omit the detailed discussions here and interested readers should consult [104] and [105] for details.

More recently, a concept of adaptive MRA (AMRA) was introduced in [102] to derive a fast and flexible decomposition strategy which adapts to data and provides a sparse approximation to the underlying function. This new AMRA is a variant of the classical MRA. For the general AMRA wavelet system, a unitary extension principle for filter design is derived. Then, it is applied to the directional representation system of shearlets, which, in turn, leads to the unitary extension principle for shearlets. This leads to a comprehensive theory for fast decomposition and reconstruction algorithms associated with 2D as well as 3D-shearlet systems encompassing tight shearlet frames with spatially compactly supported generators within such an AMRA structure. Furthermore, within the framework of [102] shearlet-like systems associated with parabolic scaling and unimodular matrices optimally close to rotation are studied.

Finally, both the unitary extension principle and the oblique extension principle can be generalized to a bi-frame setting which is called the mixed extension principle. Interested reader should consult [52, 65, 157] where the mixed extension principle is given in the multivariate setting with an arbitrary integer dilation matrix.

Furthermore, the mixed extension principle for $L_2(\mathbb{R}^d)$ of [157] is generalized for a pair of dual Sobolev spaces $H^s(\mathbb{R}^d)$ and $H^{-s}(\mathbb{R}^d)$ in [106]. Here, we briefly discuss the univariate case and encourage the reader to consult [106] for details of the multivariate case. The mixed extension principle is given to ensure that a pair of systems $X^s(\phi; \psi_1, \dots, \psi_r)$ and $X^{-s}(\tilde{\phi}; \tilde{\psi}_1, \dots, \tilde{\psi}_r)$ form a dual wavelet frame pair in the corresponding dual Sobolev spaces $H^s(\mathbb{R})$ and $H^{-s}(\mathbb{R})$. Recall that the system $X^s(\phi, \Psi) := X^s(\phi; \psi_1, \dots, \psi_r)$ is called a homogenous wavelet system generated by ϕ and $\Psi := \{\psi_1, \dots, \psi_r\}$, i.e.,

$$X^s(\phi, \Psi) := \{\phi(\cdot - k) : k \in \mathbb{Z}^d\} \cup \{2^{n(d/2-s)}\psi_\ell(2^n \cdot -k) : n \in \mathbb{N}_0, k \in \mathbb{Z}^d, 1 \leq \ell \leq r\}.$$

The key new feature in this general mixed extension principle is that the regularity and vanishing moment are shared by two different systems in the dual pair. In this case, if $\phi, \psi_1, \dots, \psi_r$ are in $H^s(\mathbb{R})$ for $s > 0$, then the regularity of $\phi, \psi_1, \dots, \psi_r$ is required, but $\tilde{\psi}_1, \dots, \tilde{\psi}_r$ is not required to have any order of vanishing moment. At the same time, the vanishing moments of $\tilde{\psi}_1, \dots, \tilde{\psi}_r$ are required, while $\tilde{\phi}, \tilde{\psi}_1, \dots, \tilde{\psi}_r$ are in $H^{-s}(\mathbb{R})$ can be tempered distributions instead of in $L_2(\mathbb{R})$. (Note that in order to have a dual pair in $L_2(\mathbb{R})$, both systems in the dual pair must have a certain order of regularity and vanishing moment at the same time.) This implies that it is not necessary for the systems $X^s(\phi; \psi_1, \dots, \psi_r)$ and $X^{-s}(\tilde{\phi}; \tilde{\psi}_1, \dots, \tilde{\psi}_r)$ to be normalized into a frame of $L_2(\mathbb{R})$. This leads to many simple constructions of frames in an arbitrarily given Sobolev space. For example, let B_m be the B -spline of order m . Then by [106], $\{2^{j(1/2-s)}B_m(2^j \cdot -k) : j \in \mathbb{N}_0, k \in \mathbb{Z}\}$ is a wavelet frame in $H^s(\mathbb{R})$ for any $0 < s < m - 1/2$. This construction is applied to multivariate box splines to obtain wavelet frames with small supports, which is known to be difficult if the traditional approach is taken. Recall that in the traditional approach, frames in Sobolev spaces are obtained by normalizing a frame in $L_2(\mathbb{R})$ to a frame in Sobolev space. Since it is hard to construct small support wavelet frames in $L_2(\mathbb{R}^d)$ in general, it is difficult to obtain small support frames in Sobolev spaces by this traditional method. Also, this general mixed extension principle also naturally leads to a characterization of the Sobolev norm of a function in terms of the weighted norm of its wavelet coefficient sequence (decomposition sequence) without requiring that dual wavelet frames be in $L_2(\mathbb{R})$. This approach is quite different from others in the literature (see e.g. [10, 9, 108, 134]). Furthermore, applying this general mixed extension principle obtains a characterization for a pair of systems $X^s(\phi; \psi_1, \dots, \psi_r)$ and $X^{-s}(\tilde{\phi}; \tilde{\psi}_1, \dots, \tilde{\psi}_r)$ in Sobolev spaces $H^s(\mathbb{R})$ and $H^{-s}(\mathbb{R})$ that form a pair of dual Riesz bases. This characterization, for example, leads to a proof of the fact that all interpolatory wavelet systems defined in [76] generated by an interpolatory refinable function $\phi \in H^s(\mathbb{R})$ with $s > 1/2$ are Riesz bases of the Sobolev space $H^s(\mathbb{R})$.

2. Quasi-affine systems and associated algorithms

In the context of signal and image processing, it is usually preferred to use wavelet systems that are shift-invariant. Recall that a system X , that contains countably many elements, is τ -shift-invariant with $\tau \in \mathbb{R}$, if for any $k \in \mathbb{Z}$ and $g \in X$, we have $g(\cdot - \tau k) \in X$. In particular, if a system is 1-shift-invariant, we simply call it *shift-invariant*. Notice that the wavelet system (affine system) $X(\Psi)$ defined by (2.1) is not shift-invariant. In order to achieve shift-invariance, we need to over-sample the affine system $X(\Psi)$ below level 0. This over-sampled affine

system is called a quasi-affine system, which was first introduced in [158]. One implicitly uses quasi-affine systems when the wavelet decomposition algorithm is applied without downsampling [57]. This type of wavelet transform is also known as the undecimated wavelet transform. As a consequence, the theory of quasi-affine systems established in [158] also provides a theoretic foundation of using the undecimated wavelet transform which was not available in the literature before. The definition of quasi-affine system is given as follows. Note that most part of this section is taken from [32].

2.1. Quasi-affine systems

The notion of quasi-affine systems was first introduced and extensively studied in [158].

Definition 2.2. Let $\Psi = \{\psi_1, \dots, \psi_r\}$ be a set of functions. A *quasi-affine system* from level L is defined as

$$X_L^q(\Psi) = \{\psi_{\ell,n,k}^q : 1 \leq \ell \leq r; n, k \in \mathbb{Z}\},$$

where $\psi_{\ell,n,k}^q$ is defined by

$$\psi_{\ell,n,k}^q := \begin{cases} \mathcal{D}^n T_k \psi_\ell, & n \geq L; \\ 2^{\frac{n-L}{2}} T_{2^{-L}k} \mathcal{D}^n \psi_\ell, & n < L. \end{cases}$$

The quasi-affine system is obtained by over sampling the wavelet frame system starting from level $L - 1$ and downward. Hence, the whole quasi-affine system is a 2^{-L} -shift-invariant system. The quasi-affine system from level 0 was first introduced in [158] to convert a non-shift invariant affine system to a shift invariant system. Further, it was shown in [158, Theorem 5.5] that a wavelet system $X(\Psi)$ (not necessarily an MRA-based wavelet system) is a tight frame of $L_2(\mathbb{R})$ if and only if the corresponding quasi-affine counterpart $X_L^q(\Psi)$ is a tight frame of $L_2(\mathbb{R})$. In fact, this result plays important roles in the proof of Theorem 2.1 by using the dual Gramian analysis. This result enables us to convert a wavelet system into a shift invariant system which makes it possible to use the powerful tool of the dual Gramian analysis of [156, 162]. Note that dual Gramian analysis of [156] is designed for shift invariant systems.

Here we provide a direct proof of a special case of [158, Theorem 5.5], the case for the MRA-based wavelet system, in Theorem 2.4. Precisely speaking, we prove that the quasi-affine system $X_L^q(\Psi)$ is a tight frame for $L_2(\mathbb{R})$ if Ψ is constructed from UEP (Theorem 2.2).

When we consider the MRA-based quasi-affine system $X_L^q(\Psi)$ generated by Ψ , the spaces V_n , $n < L$ in the MRA for the affine system are replaced by $V_n^{q,L}$, $n < L$, for the quasi-affine system. Note that the space V_n is spanned by the functions $\phi_{n,k}$, while the space $V_n^{q,L}$ is spanned by functions $\phi_{n,k}^q$, where $\phi_{n,k}^q$ is defined by

$$\phi_{n,k}^q := \begin{cases} \mathcal{D}^n T_k \phi, & n \geq L; \\ 2^{\frac{n-L}{2}} T_{2^{-L}k} \mathcal{D}^n \phi, & n < L. \end{cases}$$

The spaces $V_n^{q,L}$ for all $n \in \mathbb{Z}$ are 2^{-L} -shift invariant.

Similar to an affine system, we can define the quasi-interpolatory operator $\mathcal{P}_n^{q,L}$ and the truncated operator $\mathcal{Q}_n^{q,L}$, $n \in \mathbb{Z}$, for the quasi-affine system by

$$(2.30) \quad \mathcal{P}_n^{q,L} : f \mapsto \sum_{k \in \mathbb{Z}} \langle f, \phi_{n,k}^q \rangle \phi_{n,k}^q$$

and

$$(2.31) \quad \mathcal{Q}_n^{q,L} : f \mapsto \sum_{\ell=1}^r \sum_{j < n, k \in \mathbb{Z}} \langle f, \psi_{\ell,j,k}^q \rangle \psi_{\ell,j,k}^q.$$

Whenever $\{\phi_{n,k}^q : k \in \mathbb{Z}\}$ is a Bessel system, the quasi-interpolatory operator $\mathcal{P}_n^{q,L}$ maps $f \in L_2(\mathbb{R})$ into $V_n^{q,L}$. From the definition of $\phi_{n,k}^q$, we can see that $\mathcal{P}_n^{q,L} = \mathcal{P}_n$ when $n \geq L$, and that these two operators are different only when $n < L$. Moreover, since for an arbitrary $f \in L_2(\mathbb{R})$ and $n < L$,

$$\begin{aligned} \mathcal{P}_n^{q,L} f &= \sum_{k \in \mathbb{Z}} \langle f, \phi_{n,k}^q \rangle \phi_{n,k}^q = \mathcal{D}^L \sum_{k \in \mathbb{Z}} \langle \mathcal{D}^{-L} f, 2^{\frac{n-L-0}{2}} T_k \mathcal{D}^{n-L} \phi \rangle 2^{\frac{n-L-0}{2}} T_k \mathcal{D}^{n-L} \phi \\ &= \mathcal{D}^L \mathcal{P}_{n-L}^{q,0} \mathcal{D}^{-L} f, \end{aligned}$$

one only needs to understand the case $L = 0$. In this case we simplify our notation by setting

$$(2.32) \quad \mathcal{P}_n^q := \mathcal{P}_n^{q,0}, \quad \mathcal{Q}_n^q := \mathcal{Q}_n^{q,0}$$

for the quasi-interpolatory operators and $V_n^q := V_n^{q,0}$, for $n \in \mathbb{Z}$. From now on, we only give the properties for \mathcal{P}_n^q , the corresponding spaces V_n^q , and the associated quasi-affine system $X^q(\Psi) := X_0^q(\Psi)$. The corresponding results for the over sampling rate of $2^{-L}\mathbb{Z}$ can be obtained in a similar manner.

Also, for operator \mathcal{P}_n^q , $n \in \mathbb{Z}$, we have a decomposition and reconstruction formula similar to (2.12). Similar to the presentation of affine systems, we present proof of Lemma 2.4 in the Fourier domain under Assumption 2.1, and in the time domain with some additional assumptions.

Lemma 2.4. *Let $X(\Psi)$, where the framelets $\Psi = \{\psi_1, \dots, \psi_r\}$, be the affine tight frame system obtained from the UEP with corresponding refinable function ϕ and the set of refinement masks $\{h_0, h_1, \dots, h_r\}$. Let $X^q(\Psi)$ be the quasi-affine system derived from $X(\Psi)$. Assume that Assumption 2.1 is satisfied. Then we have, for all $f \in L_2(\mathbb{R})$,*

$$(2.33) \quad \mathcal{P}_n^q f = \mathcal{P}_{n-1}^q f + \sum_{\ell=1}^r \sum_{k \in \mathbb{Z}} \langle f, \psi_{\ell,n-1,k}^q \rangle \psi_{\ell,n-1,k}^q.$$

PROOF. When $n \geq 0$, we have $\phi_{n,k}^q = \phi_{n,k}$ and $\psi_{n,k}^q = \psi_{n,k}$, which imply that

$$\mathcal{P}_n f = \sum_{k \in \mathbb{Z}} \langle f, \phi_{n,k} \rangle \phi_{n,k} = \sum_{k \in \mathbb{Z}} \langle f, \phi_{n,k}^q \rangle \phi_{n,k}^q = \mathcal{P}_n^q f,$$

and

$$\sum_{\ell=1}^r \sum_{k \in \mathbb{Z}} \langle f, \psi_{\ell,n,k}^q \rangle \psi_{\ell,n,k}^q = \sum_{\ell=1}^r \sum_{k \in \mathbb{Z}} \langle f, \psi_{\ell,n,k} \rangle \psi_{\ell,n,k}.$$

Then (2.12) implies that

$$\begin{aligned} \mathcal{P}_n^q f &= \mathcal{P}_n f = \mathcal{P}_{n-1} f + \sum_{\ell=1}^r \sum_{k \in \mathbb{Z}} \langle f, \psi_{\ell,n-1,k} \rangle \psi_{\ell,j,k} \\ &= \mathcal{P}_{n-1}^q f + \sum_{\ell=1}^r \sum_{k \in \mathbb{Z}} \langle f, \psi_{\ell,n-1,k}^q \rangle \psi_{\ell,n-1,k}^q, \end{aligned}$$

i.e. the identity (2.33) holds when $n \geq 1$. Next, we show (2.33) holds for $n \leq 0$. We denote ϕ as ψ_0 .

First, we note that the system $\{\psi_{\ell,n,k}^q : k \in \mathbb{Z}\}$ for any given $\ell = 0, 1, \dots, r$ and $n \leq 0$ is a Bessel system. Indeed, by Assumption 2.1, we know that $\{\psi_{\ell,n,k} : k \in \mathbb{Z}\}$ is a Bessel system, i.e. there exists constant $0 < C < \infty$ such that for all $f \in L_2(\mathbb{R})$,

$$\sum_{k \in \mathbb{Z}} |\langle f, \psi_{\ell,n,k} \rangle|^2 \leq C \|f\|_{L_2(\mathbb{R})}^2.$$

Notice that

$$\begin{aligned} \{\psi_{\ell,n,k}^q : k \in \mathbb{Z}\} &= \{2^n \psi_{\ell}(2^n x - 2^n k) : k \in \mathbb{Z}\} \\ &= \bigcup_{l \in \mathbb{Z}/2^{-n}\mathbb{Z}} \{2^n \psi_{\ell}(2^n(\cdot - l) - k) : k \in \mathbb{Z}\}. \end{aligned}$$

By a simple change of variables, one can easily check that the system $\{2^n \psi_{\ell}(2^n(\cdot - l) - k) : k \in \mathbb{Z}\}$ is also a Bessel system with bound $2^{n/2} \sqrt{C}$ for every $l \in \mathbb{Z}/2^{-n}\mathbb{Z}$. Therefore, we have

$$\sum_{k \in \mathbb{Z}} |\langle f, \psi_{\ell,n,k}^q \rangle|^2 \leq C \|f\|_{L_2(\mathbb{R})}^2.$$

Denote

$$\mathcal{P}_{n,\ell}^q f := \sum_{k \in \mathbb{Z}} \langle f, \psi_{\ell,n,k}^q \rangle \psi_{\ell,n,k}^q,$$

for $\ell = 0, 1, \dots, r$. Note that $\mathcal{P}_{n,0}^q = \mathcal{P}_n^q$. Since, the system $\{\psi_{\ell,n,k}^q; k \in \mathbb{Z}\}$ for each ℓ and n , forms a Bessel system, the operator $\mathcal{P}_{n,\ell}^q : L_2(\mathbb{R}) \rightarrow L_2(\mathbb{R})$ is well-defined. Therefore, to prove (2.33) for $n \leq 0$, it is equivalent to show

$$(2.34) \quad \widehat{\mathcal{P}_n^q f} = \sum_{\ell=0}^r \widehat{\mathcal{P}_{n-1,\ell}^q f}.$$

Letting $I_n := 2^n(\mathbb{Z}/2^{-n}\mathbb{Z})$, we observe that

$$\begin{aligned} \mathcal{P}_{n,\ell}^q f &= \sum_{k \in \mathbb{Z}} \langle f, \psi_{\ell,n,k}^q \rangle \psi_{\ell,n,k}^q \\ &= 2^n \sum_{\alpha \in I_n} \sum_{k \in \mathbb{Z}} \langle f, (T_{\alpha} \psi_{\ell})_{n,k} \rangle (T_{\alpha} \psi_{\ell})_{n,k}. \end{aligned}$$

Then by Proposition 1.3 item (3), we have

$$\begin{aligned} \widehat{\mathcal{P}_{n,\ell}^q f}(\xi) &= 2^n \sum_{\alpha \in I_n} [\widehat{f}(2^n \cdot), \widehat{T_{\alpha} \psi_{\ell}}](2^{-n} \xi) \widehat{T_{\alpha} \psi_{\ell}}(2^{-n} \xi) \\ &= 2^n \sum_{\alpha \in I_n} \sum_{k \in \mathbb{Z}} \widehat{f}(\xi + 2^{n+1} k \pi) \overline{\widehat{T_{\alpha} \psi_{\ell}}(\frac{\xi}{2^n} + 2k\pi)} \widehat{T_{\alpha} \psi_{\ell}}(2^{-n} \xi) \\ &= 2^n \sum_{\alpha \in I_n} \sum_{k \in \mathbb{Z}} \widehat{f}(\xi + 2^{n+1} k \pi) e^{i\alpha(2^{-n} \xi + 2k\pi)} \widehat{\psi_{\ell}}(\frac{\xi}{2^n} + 2k\pi) e^{-i\alpha 2^{-n} \xi} \widehat{\psi_{\ell}}(2^{-n} \xi) \\ &= 2^n \sum_{\alpha \in I_n} \sum_{k \in \mathbb{Z}} \widehat{f}(\xi + 2^{n+1} k \pi) \widehat{\psi_{\ell}}(\frac{\xi}{2^n} + 2k\pi) e^{i\alpha 2k\pi} \widehat{\psi_{\ell}}(2^{-n} \xi) \\ &= \sum_{k \in \mathbb{Z}} \widehat{f}(\xi + 2^{n+1} k \pi) \widehat{\psi_{\ell}}(\frac{\xi}{2^n} + 2k\pi) \left(2^n \sum_{\alpha \in I_n} e^{i\alpha 2k\pi} \right) \widehat{\psi_{\ell}}(2^{-n} \xi). \end{aligned}$$

Notice that

$$2^n \sum_{\alpha \in I_n} e^{i\alpha 2k\pi} = 2^n \sum_{l \in \mathbb{Z}/2^{-n}\mathbb{Z}} \left(e^{i2^{n+1}k\pi} \right)^l = \begin{cases} 1, & k \in 2^{-n}\mathbb{Z} \\ 0, & k \notin 2^{-n}\mathbb{Z} \end{cases}.$$

Therefore

$$\begin{aligned} \widehat{\mathcal{P}_{n,\ell}^q f} &= \sum_{k \in 2^{-n}\mathbb{Z}} \widehat{f}(\xi + 2^{n+1}k\pi) \overline{\widehat{\psi}_\ell\left(\frac{\xi}{2^n} + 2k\pi\right)} \widehat{\psi}_\ell(2^{-n}\xi) \\ &= [\widehat{f}, \widehat{\psi}_\ell(\frac{\cdot}{2^n})] \widehat{\psi}_\ell(\frac{\cdot}{2^n}). \end{aligned}$$

Now, identity (2.34) is equivalent to

$$(2.35) \quad [\widehat{f}, \widehat{\phi}(\frac{\cdot}{2^n})] \widehat{\phi}(\frac{\cdot}{2^n}) = \sum_{\ell=0}^r [\widehat{f}, \widehat{\psi}_\ell(\frac{\cdot}{2^{n-1}})] \widehat{\psi}_\ell(\frac{\cdot}{2^{n-1}}),$$

for $n \leq 0$. Using $\widehat{\psi}_\ell = \widehat{h}_\ell(\frac{\cdot}{2}) \widehat{\phi}(\frac{\cdot}{2})$ and the first identity of the UEP condition (2.6), we have

$$\sum_{\ell=0}^r [\widehat{f}, \widehat{\psi}_\ell(\frac{\cdot}{2^{n-1}})] \widehat{\psi}_\ell(\frac{\cdot}{2^{n-1}}) = \sum_{\ell=0}^r |\widehat{h}_\ell(\frac{\cdot}{2^n})|^2 [\widehat{f}, \widehat{\phi}(\frac{\cdot}{2^n})] \widehat{\phi}(\frac{\cdot}{2^n}) = [\widehat{f}, \widehat{\phi}(\frac{\cdot}{2^n})] \widehat{\phi}(\frac{\cdot}{2^n}).$$

This concludes the proof of (2.35) and hence the lemma follows. \blacksquare

Remark 2.3. Lemma 2.4 can be proven in the time domain. We prove a simple case here because the proof will reveal some details of the fast framelet decomposition and reconstruction algorithm for quasi-affine systems. For simplicity, we assume that ψ_ℓ , for $\ell = 0, 1, \dots, r$, are compactly supported and the masks h_ℓ are finite sequences.

We only need to prove (2.33) for $n \leq 0$. Since system $\{\psi_{\ell,n,k}^q : k \in \mathbb{Z}\}$ for any given $\ell = 0, 1, \dots, r$ and $n \leq 0$ is a Bessel system, the operators $\mathcal{P}_{n,\ell}^q : L_2(\mathbb{R}) \mapsto L_2(\mathbb{R})$, for each ℓ is bounded. Therefore, we only need to show (2.33) for all compactly supported f .

By the definitions of a refinable equation (1.5) and a framelet (2.4), one obtains for $\ell = 0, 1, \dots, r$,

$$\psi_\ell = 2 \sum_{k \in \mathbb{Z}} h_\ell[k] \phi(2 \cdot -k).$$

This leads to

$$\begin{aligned} \psi_{\ell,n-1,k}^q &= 2^{n-1} T_k \psi_\ell(2^{n-1} \cdot) = 2^n T_k \left(\sum_{k' \in \mathbb{Z}} h_\ell[k'] \psi_0(2^n \cdot -k') \right) \\ &= \sum_{k' \in \mathbb{Z}} h_\ell[k'] 2^n \psi_0(2^n(\cdot - k - 2^{-n}k')) \\ &= \sum_{k' \in 2^{-n}\mathbb{Z}} h_\ell[2^n k'] 2^n \psi_0(2^n(\cdot - k - k')). \end{aligned}$$

We define the dilated sequence $h_{\ell,n}$ by

$$(2.36) \quad h_{\ell,n}[k] = \begin{cases} h_\ell[2^n k], & k \in 2^{-n}\mathbb{Z}; \\ 0, & k \notin 2^{-n}\mathbb{Z}. \end{cases}$$

The sequence $h_{\ell,n}$ is obtained inductively by inserting 0 between every two entries in $h_{\ell,n+1}$, and $h_{\ell,0} = h_{\ell}$. With the dilated sequence, we have

$$\psi_{\ell,n-1,k}^q = \sum_{k' \in \mathbb{Z}} h_{\ell,n}[k'] \psi_{0,n,k+k'}^q.$$

Therefore, we have

$$\begin{aligned} & \sum_{\ell=0}^r \sum_{k \in \mathbb{Z}} \langle f, \psi_{\ell,n-1,k}^q \rangle \psi_{\ell,n-1,k}^q \\ &= \sum_{\ell=0}^r \sum_{k \in \mathbb{Z}} \left(\sum_{k' \in \mathbb{Z}} \overline{h_{\ell,n}[k']} \langle f, \psi_{0,n,k'+k}^q \rangle \right) \left(\sum_{k'' \in \mathbb{Z}} h_{\ell,n}[k''] \psi_{0,n,k''+k}^q \right) \\ &= \sum_{k' \in \mathbb{Z}} \sum_{k'' \in \mathbb{Z}} \left(\sum_{\ell=0}^r \sum_{k \in \mathbb{Z}} \overline{h_{\ell,n}[k]} h_{\ell,n}[k+k''-k'] \right) \langle f, \psi_{0,n,k'}^q \rangle \psi_{0,n,k''}^q, \end{aligned}$$

where the switching of summation order is valid since all summations involved are finite sums.

Now, we focus on showing

$$\sum_{\ell=0}^r \sum_{k \in \mathbb{Z}} \overline{h_{\ell,n}[k]} h_{\ell,n}[k+k''-k'] = \delta_{k'-k'',0}.$$

When $k' - k'' \in 2^{-n}\mathbb{Z}$, there exists $p \in \mathbb{Z}$ such that $k' - k'' = 2^{-n}p$ and we have

$$\begin{aligned} \sum_{\ell=0}^r \sum_{k \in \mathbb{Z}} \overline{h_{\ell,n}[k]} h_{\ell,n}[k+k''-k'] &= \sum_{\ell=0}^r \sum_{k \in \mathbb{Z}} \overline{h_{\ell,n}[k]} h_{\ell,n}[k-2^{-n}p] \\ &= \sum_{\ell=0}^r \sum_{k \in 2^{-n}\mathbb{Z}} \overline{h_{\ell,n}[k]} h_{\ell,n}[k-2^{-n}p] \\ &= \sum_{\ell=0}^r \sum_{k \in \mathbb{Z}} \overline{h_{\ell}[k]} h_{\ell}[k-p] = \delta_{p,0}. \end{aligned}$$

The last identity follows by (2.9). The sum is nonzero if and only if $p = 0$, which is equivalent to $k' = k''$. When $k' - k'' \notin 2^{-n}\mathbb{Z}$, there exist $p_1, p_2 \in \mathbb{Z}$ and $p_2 \notin 2^{-n}\mathbb{Z}$ such that $k' - k'' = 2^{-n}p_1 + p_2$. Then we have

$$\begin{aligned} \sum_{\ell=0}^r \sum_{k \in \mathbb{Z}} \overline{h_{\ell,n}[k]} h_{\ell,n}[k+k''-k'] &= \sum_{\ell=0}^r \sum_{k \in \mathbb{Z}} \overline{h_{\ell,n}[k]} h_{\ell,n}[k-2^{-n}p_1-p_2] \\ &= \sum_{\ell=0}^r \sum_{k \in 2^{-n}\mathbb{Z}} \overline{h_{\ell,n}[k]} h_{\ell,n}[k-2^{-n}p_1-p_2]. \end{aligned}$$

Since $k - 2^{-n}p_1 - p_2 \notin 2^{-n}\mathbb{Z}$ when $k \in 2^{-n}\mathbb{Z}$, we have $h_{\ell,n}[k - 2^{-n}p_1 - p_2] = 0$ for any $k \in 2^{-n}\mathbb{Z}$ and the last identity is equal to 0. In conclusion, for the dilated filters $h_{0,n}, h_{1,n}, \dots, h_{r,n}$, we still have a result similar to (2.9)

$$(2.37) \quad \sum_{\ell=0}^r \sum_{k \in \mathbb{Z}} \overline{h_{\ell,n}[k]} h_{\ell,n}[k-p] = \delta_{p,0}, \quad n \leq 0, \quad p \in \mathbb{Z}.$$

Thus we have proven (2.33), which leads to

$$\sum_{\ell=0}^r \sum_{k \in \mathbb{Z}} \langle f, \psi_{\ell, n-1, k}^q \rangle \psi_{\ell, n-1, k}^q = \sum_{k \in \mathbb{Z}} \langle f, \psi_{0, n, k}^q \rangle \psi_{0, n, k}^q = \mathcal{P}_n^q f.$$

This is the identity we need to prove the case when $n \leq 0$. In all, identity (2.33) holds for any $n \in \mathbb{Z}$.

Remark 2.4. We note that in the proof of identity (2.12) for the affine system, one needs both identities of (2.6); while in the proof of identity (2.33) for $n \leq 0$, when the quasi-affine system is used, one only needs the first identity of (2.6).

The next theorem shows that a result similar to (2.23) also holds for quasi-affine systems, and that $X^q(\Psi)$ is a tight frame for $L_2(\mathbb{R})$.

Theorem 2.4. *Let $X(\Psi)$, where the framelets $\Psi = \{\psi_1, \dots, \psi_r\}$, be the affine tight frame system obtained from the UEP. Then the quasi-affine system $X^q(\Psi)$ forms a tight frame for $L_2(\mathbb{R})$, and $\mathcal{P}_n^q f = \mathcal{Q}_n^q f$ for all $f \in L_2(\mathbb{R})$.*

PROOF. We begin by showing that $\mathcal{P}_n^q f = \mathcal{Q}_n^q f$. Telescoping on (2.33) we have, for any $f \in L_2(\mathbb{R})$ and $n > n'$,

$$(2.38) \quad \mathcal{P}_n^q f = \mathcal{P}_{n'}^q f + \sum_{\ell=1}^r \sum_{j=n'}^{n-1} \sum_{k \in \mathbb{Z}} \langle f, \psi_{\ell, j, k}^q \rangle \psi_{\ell, j, k}^q.$$

Thus the proof of $\mathcal{P}_n^q f = \mathcal{Q}_n^q f$ is transferred to the proof of $\mathcal{P}_{n'}^q f \rightarrow 0$ as $n' \rightarrow -\infty$. Since we are interested in the behavior of $\mathcal{P}_{n'}^q$ as $n' \rightarrow -\infty$, we now assume that $n' < 0$.

As shown in the proof of Lemma 2.4, $\{\phi_{n', k}^q : k \in \mathbb{Z}\}$ is a Bessel system for any given $n' \in \mathbb{Z}$. Thus, we have

$$\sum_{k \in \mathbb{Z}} |\langle f, \phi_{n', k}^q \rangle|^2 \leq C \|f\|_{L_2(\mathbb{R})}^2.$$

Using the above inequality and

$$\mathcal{P}_{n'}^q f = \sum_{k \in \mathbb{Z}} \langle f, \phi_{n', k}^q \rangle \phi_{n', k}^q,$$

we have the norm of $\mathcal{P}_{n'}^q f$ satisfying

$$(2.39) \quad \|\mathcal{P}_{n'}^q f\|_{L_2(\mathbb{R})}^2 \leq C \sum_{k \in \mathbb{Z}} |\langle f, \phi_{n', k}^q \rangle|^2.$$

Based on routine arguments in approximation theory, we only need to check the value of $\|\mathcal{P}_{n'}^q f\|_{L_2(\mathbb{R})}$ when f is supported on an interval $[-R, R]$ for an arbitrary given $R > 0$. Applying the Cauchy-Schwartz inequality to (2.39), we have,

$$(2.40) \quad \begin{aligned} \|\mathcal{P}_{n'}^q f\|_{L_2(\mathbb{R})}^2 &\leq C \|f\|_{L_2(\mathbb{R})}^2 \sum_{k \in \mathbb{Z}} \int_{[-R, R]} \left| 2^{n'} \phi(2^{n'}(x-k)) \right|^2 dx \\ &= 2^{n'} C \|f\|_{L_2(\mathbb{R})}^2 \sum_{k \in \mathbb{Z}} \int_{E_{n', k}} |\phi(x)|^2 dx, \end{aligned}$$

where $E_{n', k} = 2^{n'}([-R, R] - k)$. Now, we prove that

$$(2.41) \quad 2^{n'} \sum_{k \in \mathbb{Z}} \int_{E_{n', k}} |\phi(x)|^2 dx \rightarrow 0, \quad \text{as } n' \rightarrow -\infty.$$

By the monotone convergence theorem, we have

$$2^{n'} \sum_{k \in \mathbb{Z}} \int_{E_{n',k}} |\phi(x)|^2 dx = 2^{n'} \int \sum_{k \in \mathbb{Z}} \chi_{E_{n',k}} |\phi(x)|^2 dx.$$

Observe that $\sum_{k \in \mathbb{Z}} \chi_{E_{n',k}} \leq C_R$ with the constant C_R only depending on R . Therefore, we have

$$2^{n'} \sum_{k \in \mathbb{Z}} \chi_{E_{n',k}} |\phi(x)|^2 \rightarrow 0 \text{ a.e. and } 2^{n'} \sum_{k \in \mathbb{Z}} \chi_{E_{n',k}} |\phi(x)|^2 \leq C_R |\phi(x)|^2.$$

Then, the dominated convergence theorem implies (2.41), and hence (2.38) becomes

$$\mathcal{P}_n^q f = \sum_{\ell=1}^r \sum_{j < n} \sum_{k \in \mathbb{Z}} \langle f, \psi_{\ell,j,k}^q \rangle \psi_{\ell,j,k}^q = \mathcal{Q}_n^q f.$$

This completes the proof of $\mathcal{P}_n^q f = \mathcal{Q}_n^q f$.

Since $\mathcal{P}_n^q f = \mathcal{P}_n f$ for $n \geq 0$, we have $\mathcal{P}_n^q f = \mathcal{P}_n f \rightarrow f$ as $n \rightarrow \infty$ (by Lemma 2.3). This implies that $X^q(\Psi)$ forms a tight frame of $L_2(\mathbb{R})$ and thus concludes the proof of this theorem. ■

From the above theorem, the following result on approximation orders of quasi-affine tight frame systems $X^q(\Psi)$ is obvious. Note that the approximation order of $X^q(\Psi)$ is defined similarly as $X(\Psi)$ in Definition 2.1.

Proposition 2.5. *Let $X^q(\Psi)$ be a quasi-affine tight frame system constructed from UEP with the underlying MRA generated by a refinable function ϕ . Assume that Ψ has vanishing moments of order m_0 and ϕ provides approximation order m . Then the approximation order of the quasi-affine tight frame system is $m_1 = \min\{m, 2m_0\}$.*

Similar to the affine counterpart, we normally do not perform full levels of decomposition and reconstruction as given by Theorem 2.4. We normally stop at a certain level instead. The following corollary states that by doing so, we still have a tight frame system which includes the shifts of ϕ at a certain scale.

Corollary 2.2. *Let $\Psi = \{\psi_\ell : 1 \leq \ell \leq r\}$ be the set of tight framelets constructed from the UEP with ϕ the corresponding refinable function. Then for any given $L \in \mathbb{Z}$, the system*

$$X^q(\phi, \Psi; L) := \{\phi_{L,k}^q, \psi_{\ell,n,k}^q : 1 \leq \ell \leq r, n \geq L, k \in \mathbb{Z}\}$$

forms a tight frame of $L_2(\mathbb{R})$, i.e. for any $f \in L_2(\mathbb{R})$,

$$f = \sum_{k \in \mathbb{Z}} \langle f, \phi_{L,k}^q \rangle \phi_{L,k}^q + \sum_{\ell=1}^r \sum_{n \geq L} \sum_{k \in \mathbb{Z}} \langle f, \psi_{\ell,n,k}^q \rangle \psi_{\ell,n,k}^q.$$

In practical calculations, we always assume that the function we have is $\mathcal{P}_0^q f \in V_0^q$, and perform decomposition down to V_{-L}^q for some $L \in \mathbb{Z}^+$. Then the decomposition and reconstruction formula can be written as

$$\mathcal{P}_0^q f = \mathcal{P}_{-L}^q f + \sum_{\ell=1}^r \sum_{j=-L}^{-1} \sum_{k \in \mathbb{Z}} \langle f, \psi_{\ell,j,k}^q \rangle \psi_{\ell,j,k}^q.$$

This process corresponds to the so-called undecimated wavelet decomposition and reconstruction. Details of this process are described in the next section.

2.2. Algorithms for quasi-affine tight frames

In this subsection we provide the decomposition and reconstruction algorithms for quasi-affine tight frame systems following a similar discussion as in Section 1.4.

Since

$$\mathcal{P}_L^{q,L} f = \mathcal{P}_L f = D^L \mathcal{P}_0 D^{-L} f = D^L \mathcal{P}_0^q D^{-L} f,$$

one may use $\mathcal{P}_0^q f \in V_0^q$ to approximate f without loss of generality. If it is necessary, we can always consider function $f(2^{-L}\cdot)$ instead of f , since the approximation of a function f in space $V_L^{q,L}$ is the same as that of the function $f(2^{-L}\cdot)$ in space V_0^q .

In level $n \leq 0$, the dilated filter is denoted by $h_{\ell,n}$, which is defined by (same as (2.36))

$$(2.42) \quad h_{\ell,n}[k] = \begin{cases} h_{\ell}[2^n k], & k \in 2^{-n}\mathbb{Z}; \\ 0, & k \in \mathbb{Z} \setminus 2^{-n}\mathbb{Z}. \end{cases}$$

Given a sequence $h_{\ell,n} = \{h_{\ell,n}[k]\}_{k \in \mathbb{Z}}$ for any $\ell = 0, 1, \dots, r$, and $n \leq 0$, we define an infinite matrix $H_{\ell,n}^q$ corresponding to $h_{\ell,n}$ as

$$H_{\ell,n}^q := (H_{\ell,n}^q[l, k]) := \overline{h_{\ell,n}[k-l]},$$

where the (l, k) th entry in $H_{\ell,n}^q$ is fully determined by the $(k-l)$ th entry in $h_{\ell,n}$. Then for any vector $v \in \ell_2(\mathbb{Z})$, we have

$$(H_{\ell,n}^q v)[l] = \sum_{k \in \mathbb{Z}} \overline{h_{\ell,n}[k-l]} v[k].$$

Note that the matrix multiplication defined above is in fact the convolution of v with filter $\overline{h_{\ell,n}[\cdot]}$, i.e.

$$H_{\ell,n}^q v = \overline{h_{\ell,n}[\cdot]} * v.$$

We denote $H_{\ell,n}^{q*}$ as the adjoint of $H_{\ell,n}^q$ with

$$(H_{\ell,n}^{q*} v)[k] = \sum_{l \in \mathbb{Z}} h_{\ell,n}[k-l] v[l] = (h_{\ell,n} * v)[k].$$

We collect the coefficients in each level $n < 0$ to form an infinite column vector

$$v_{\ell,n} := [\dots, \langle f, \psi_{\ell,n,k}^q \rangle, \dots]^{\top}, \quad \ell = 0, 1, \dots, r,$$

where $\psi_0^q := \phi^q$. From the proof of Lemma 2.4, we have shown that (identity (2.37))

$$\sum_{\ell=0}^r \sum_{k \in \mathbb{Z}} \overline{h_{\ell,n}[k]} h_{\ell,n}[k-p] = \delta_{p,0}, \quad p \in \mathbb{Z}.$$

which is equivalent to

$$(2.43) \quad \sum_{\ell=0}^r H_{\ell,n}^{q*} H_{\ell,n}^q = I.$$

Thus, the decomposition and reconstruction process (2.33) can be written in matrix form as

$$v_{0,n} = \sum_{\ell=0}^r H_{\ell,n}^{q*} H_{\ell,n}^q v_{0,n} \quad n \leq 0.$$

For a multiple level decomposition operator, we define \mathcal{A}_L^q , for $L < 0$, as a (rectangular) block matrix given as:

$$(2.44) \quad \mathcal{A}_L^q := \left[\left(\prod_{n=L}^0 H_{0,n}^q \right); \left(H_{1,L}^q \prod_{n=L+1}^0 H_{0,n}^q \right); \dots; \left(H_{r,L}^q \prod_{n=L+1}^0 H_{0,n}^q \right); \dots; H_{1,0}^q; \dots; H_{r,0}^q \right]^\top,$$

where the product of operators is defined as,

$$\prod_{n=L_1}^{L_2} T_n := T_{L_1} T_{L_1+1} \cdots T_{L_2}.$$

Then, we have the reconstruction operator \mathcal{A}_L^{q*} , the adjoint operator of \mathcal{A}_L^q , defined as

$$(2.45) \quad \mathcal{A}_L^{q*} = \left[\left(\prod_{n=0}^L H_{0,n}^{q*} \right); \left(\prod_{n=0}^{L+1} H_{0,n}^{q*} H_{1,L}^{q*} \right); \dots; \left(\prod_{n=0}^{L+1} H_{0,n}^{q*} H_{r,L}^{q*} \right); \dots; H_{1,0}^{q*}; \dots; H_{r,0}^{q*} \right].$$

For the case $L = 0$, we denote

$$\mathcal{A}_0^q := [H_{0,0}^q; H_{1,0}^q; \dots; H_{r,0}^q]^\top; \quad \mathcal{A}_0^{q*} := [H_{0,0}^{q*}; H_{1,0}^{q*}; \dots; H_{r,0}^{q*}].$$

The next proposition shows that the decomposition and reconstruction process is perfect, i.e. $\mathcal{A}_L^{q*} \mathcal{A}_L^q = I$.

Proposition 2.6. *The decomposition operator \mathcal{A}_L^q , as defined in (2.44) satisfies $\mathcal{A}_L^{q*} \mathcal{A}_L^q = I$, where I is the identity operator.*

Here, we provide the univariate fast framelet decomposition and reconstruction algorithm for quasi-affine tight frame systems in Algorithm 2.2. For simplicity of notation, we denote $\tilde{h}_{\ell,n} := \overline{h_{\ell,n}[-\cdot]}$. For a given finite signal $v \in \mathbb{R}^N$, we denote the finite length convolution operation as $\tilde{h}_{\ell,n} \otimes v$, and understand it to be the restriction of $H_{\ell,n} \tilde{v} = \tilde{h}_{\ell,n} * \tilde{v}$ on the set $\{0, 1, \dots, N-1\}$:

$$\tilde{h}_{\ell,n} \otimes v := (H_{\ell,n} \tilde{v})|_{\{0,1,\dots,N-1\}} = (\tilde{h}_{\ell,n} * \tilde{v})|_{\{0,1,\dots,N-1\}},$$

where \tilde{v} is either periodic or a symmetric extension of v given by (2.27).

Algorithm 2.2. *Given signal $v \in \mathbb{R}^N$ with $N \in \mathbb{N}_+$. Denote $v_{0,0} = v$. Then the L -level fast quasi-affine framelet decomposition and reconstruction is given as follows:*

- (1) **Decomposition:** For each $j = 1, 2, \dots, L$
 - (a) Obtain low frequency approximation to v at level j :

$$v_{0,j} = \tilde{h}_{0,1-j} \otimes v_{0,j-1};$$

- (b) Obtain framelet coefficients of v at level j :

$$v_{\ell,j} = \tilde{h}_{\ell,1-j} \otimes v_{0,j-1}, \quad \ell = 1, 2, \dots, r.$$

- (2) **Reconstruction:** For each $j = L, L-1, \dots, 1$,

$$v_{0,j-1} = \sum_{\ell=0}^r h_{\ell,1-j} \otimes v_{\ell,j}.$$

3. Higher dimension tight frame systems

So far we have focused on univariate tight frame systems. However, almost all theorems and propositions mentioned so far were established in a multivariate setting. For example, the unitary extension principle, the mixed extension principle and the oblique extension principle are stated for a multivariate setting in [65, 158]. In fact, examples of multivariate spline tight frames from Box splines by using the unitary extension principle are provided in [160]. Here, we present simplified versions of them.

3.1. MRA and tight frame systems for $L_2(\mathbb{R}^d)$

For a given function $\phi \in L_2(\mathbb{R}^d)$, we define the *shift-invariant* subspace $V(\phi) \subset L_2(\mathbb{R}^d)$ generated by ϕ as

$$V(\phi) := \overline{\text{span}\{\phi(\cdot - k), k \in \mathbb{Z}^d\}},$$

and denote V_n as the 2^n -dilate of $V(\phi)$, i.e.

$$(2.46) \quad V_n = \overline{\text{span}\{\phi(2^n \cdot - k), k \in \mathbb{Z}^d\}}, \quad n \in \mathbb{Z}.$$

We say that $\{V_n\}_n$ forms a *multiresolution analysis (MRA)* for $L_2(\mathbb{R}^d)$ if the following conditions are satisfied:

$$V_n \subset V_{n+1}; \quad \overline{\bigcup_n V_n} = L_2(\mathbb{R}^d); \quad \text{and} \quad \bigcap_n V_n = \{0\}.$$

The function ϕ is refinable if

$$(2.47) \quad \phi(x) = 2^d \sum_{k \in \mathbb{Z}^d} h_0[k] \phi(2x - k)$$

for some $h_0 \in \ell_2(\mathbb{Z}^d)$. In frequency domain, the above equation can be rewritten as

$$\widehat{\phi}(2 \cdot) = \widehat{h}_0 \widehat{\phi}.$$

The Fourier transform of a function $f \in L_1(\mathbb{R}^d)$ is defined by

$$\widehat{f}(\xi) := \int_{\mathbb{R}^d} f(t) e^{-i\xi \cdot t} dt, \quad \xi \in \mathbb{R}^d,$$

which can be extended to more general function spaces (e.g. $L_2(\mathbb{R}^d)$) in a natural manner. The Fourier series of a sequence $h_0 \in \ell_2(\mathbb{Z}^d)$ is defined as

$$\widehat{h}_0(\xi) := \sum_{j \in \mathbb{Z}^d} h_0[j] e^{-ij \cdot \xi}, \quad \xi \in \mathbb{R}^d.$$

The following theorem is the corresponding \mathbb{R}^d version of Theorem 1.4.

Theorem 2.5. [115] *Let $\phi \in L_2(\mathbb{R}^d)$, and V_n defined as (2.46). Then,*

- (1) $\bigcap_n V_n = \{0\}$;
- (2) *Assume, in addition, that ϕ is refinable. Then,*

$$\overline{\bigcup_n V_n} = L_2(\mathbb{R}^d)$$

if and only if

$$\bigcap_n 2^n Z(\widehat{\phi})$$

is a set of measure zero, where $Z(\widehat{\phi}) := \{\xi \in \mathbb{R}^d : \widehat{\phi}(\xi) = 0\}$.

In particular, the sequence of subspaces $\{V_n\}_{n \in \mathbb{Z}}$ generated by a refinable function $\phi \in L_2(\mathbb{R}^d)$ forms an MRA if either of the following conditions is satisfied:

- (1) ϕ is a compactly supported refinable function;

(2) ϕ is refinable and $\widehat{\phi}$ is continuous at 0 with $\widehat{\phi}(0) \neq 0$.

PROOF. The proof is a direct analog to that of Theorem 1.4. ■

With a refinable function ϕ generating an MRA for $L_2(\mathbb{R}^d)$, we can construct tight frames systems for $L_2(\mathbb{R}^d)$.

Definition 2.3. Let $\Psi = \{\psi_1, \dots, \psi_r\}$ be a set of functions. A wavelet/affine system generated by Ψ is defined as

$$X(\Psi) = \{\psi_{\ell,n,k} : 1 \leq \ell \leq r; n \in \mathbb{Z}; k \in \mathbb{Z}^d\}$$

where $\psi_{\ell,n,k}$ is defined by

$$\psi_{\ell,n,k} := \mathcal{D}^n T_k \psi_\ell := 2^{nd/2} \psi_\ell(2^n \cdot -k).$$

A quasi-affine system from level 0 is defined as

$$X^q(\Psi) = \{\psi_{\ell,n,k}^q : 1 \leq \ell \leq r; n \in \mathbb{Z}; k \in \mathbb{Z}^d\},$$

where $\psi_{\ell,n,k}^q$ is defined by

$$\psi_{\ell,n,k}^q := \begin{cases} \mathcal{D}^n T_k \psi_\ell, & n \geq 0; \\ 2^{\frac{nd}{2}} T_{2^{-n}k} \mathcal{D}^n \psi_\ell, & n < 0. \end{cases}$$

The following is the unitary extension principle for $L_2(\mathbb{R}^d)$ which was originally proved in [158]. For simplicity, we assume the underlying refinable function ϕ is compactly supported with $\widehat{\phi}(0) = 1$ and the associated masks are finely supported. Under these assumptions, items 1-3 in Assumption 2.1 are satisfied.

Theorem 2.6 (Unitary Extension Principle for $L_2(\mathbb{R}^d)$ [158]). *Let $\phi \in L_2(\mathbb{R}^d)$ be the compactly supported refinable function with finitely supported refinement mask h_0 satisfying $\widehat{\phi}(0) = 1$. Let $\{h_1, \dots, h_r\}$ be a set of finitely supported sequences. Then the wavelet system $X(\Psi)$ and its quasi-affine counterpart $X^q(\Psi)$, with $\Psi = \{\psi_1, \dots, \psi_r\}$ defined by $\widehat{\psi}_\ell(2 \cdot) = \widehat{h}_\ell \widehat{\phi}$, forms a tight frame in $L_2(\mathbb{R}^d)$ provided the equalities*

$$(2.48) \quad \sum_{\ell=0}^r |\widehat{h}_\ell(\xi)|^2 = 1 \quad \text{and} \quad \sum_{\ell=0}^r \widehat{h}_\ell(\xi) \overline{\widehat{h}_\ell(\xi + \nu)} = 0,$$

hold for all $\nu \in \{0, \pi\}^d \setminus \{0\}$ and $\xi \in [-\pi, \pi]^d$. Furthermore, assuming $r = 2^d - 1$ and $\|\phi\|_{L_2(\mathbb{R}^d)} = 1$, then $X(\Psi)$ is an orthonormal wavelet bases of $L_2(\mathbb{R}^d)$.

PROOF. The proof is analogous to that of Theorem 2.2 and Theorem 2.4; hence we only sketch it here.

First, one can establish the following equalities using (2.48),

$$\begin{aligned} \mathcal{P}_n f &= \mathcal{P}_{n-1} f + \sum_{\ell=1}^r \sum_{k \in \mathbb{Z}^d} \langle f, \psi_{\ell,n-1,k} \rangle \psi_{\ell,n-1,k} \\ &= \mathcal{P}_{n'} f + \sum_{\ell=1}^r \sum_{j=n'}^{n-1} \sum_{k \in \mathbb{Z}^d} \langle f, \psi_{\ell,j,k} \rangle \psi_{\ell,j,k}, \end{aligned}$$

and

$$\begin{aligned}
(2.49) \quad \mathcal{P}_n^q f &= \mathcal{P}_{n-1}^q f + \sum_{\ell=1}^r \sum_{k \in \mathbb{Z}^d} \langle f, \psi_{\ell, n-1, k}^q \rangle \psi_{\ell, n-1, k}^q \\
&= \mathcal{P}_{n'}^q f + \sum_{\ell=1}^r \sum_{j=n'}^{n-1} \sum_{k \in \mathbb{Z}^d} \langle f, \psi_{\ell, j, k}^q \rangle \psi_{\ell, j, k}^q.
\end{aligned}$$

Finally one can show that

$$\mathcal{P}_{n'} f \rightarrow 0 \quad \text{and} \quad \mathcal{P}_{n'}^q f \rightarrow 0, \quad n' \rightarrow -\infty$$

and

$$\mathcal{P}_n f \rightarrow f \quad \text{and} \quad \mathcal{P}_n^q f \rightarrow f, \quad n \rightarrow \infty$$

which establishes that $X(\Psi)$ and $X^q(\Psi)$ form tight frames for $L_2(\mathbb{R}^d)$. \blacksquare

One way to construct tight frames for $L_2(\mathbb{R}^d)$ is by taking *tensor products* of univariate tight frames. For simplicity of notation, we consider 2D case, i.e. $d = 2$. Arguments for higher dimensions are similar. Given a set of univariate masks $\{h_\ell : \ell = 0, 1, \dots, r\}$, define the 2D masks $h_{i,j}[k_1, k_2]$ as

$$(2.50) \quad h_{i,j}[k_1, k_2] := h_i[k_1]h_j[k_2], \quad 0 \leq i, j \leq r; (k_1, k_2) \in \mathbb{Z}^2,$$

or equivalently in Fourier domain

$$\widehat{h}_{i,j}[\xi_1, \xi_2] := \widehat{h}_i[\xi_1]\widehat{h}_j[\xi_2], \quad 0 \leq i, j \leq r; (\xi_1, \xi_2) \in \mathbb{R}^2.$$

Then the corresponding 2D refinable function and mother wavelets are defined by

$$\psi_{i,j}(x, y) = \psi_i(x)\psi_j(y), \quad 0 \leq i, j \leq r; (x, y) \in \mathbb{R}^2.$$

Denote the collection of mother wavelets as

$$\Psi_2 := \{\psi_{i,j}; 0 \leq i, j \leq r; (i, j) \neq (0, 0)\},$$

which contains $(r+1)^2 - 1$ elements. (Here, ψ_0 is defined to be the underlying univariate refinable function.) If the univariate masks $\{h_\ell\}$ are constructed from UEP, then it is easy to verify that $\{h_{i,j}\}$ satisfies (2.48) and thus $X(\Psi_2)$ and $X^q(\Psi_2)$ are tight frames for $L_2(\mathbb{R}^2)$.

3.2. Algorithms for tensor product tight frame systems

In this section, we only discuss the fast framelet decomposition and reconstruction algorithm for 2D finite signals. Since we are using tensor product tight frame systems, the 2D fast algorithms can be easily derived based on 1D fast algorithms for both wavelet and quasi-affine systems.

Now, we introduce some notation. Let the set of masks $\{h_\ell : \ell = 0, 1, \dots, r\}$ be constructed from univariate UEP (Theorem 2.2). For a given signal $v \in \mathbb{R}^{N_1} \times \mathbb{R}^{N_2}$ with $N_1, N_2 \in \mathbb{N}_+$, let $D_{\ell,j}^x$ and $D_{\ell,j}^y$ denote the one level of decomposition with respect to the x -index and y -index of v , which are defined as follows:

$$\begin{aligned}
(D_{\ell,j}^x v)[\cdot, k_2] &= \begin{cases} \downarrow (\tilde{h}_\ell \otimes v[\cdot, k_2]), & \text{wavelet system,} \\ \tilde{h}_{\ell,j} \otimes v[\cdot, k_2], & \text{quasi-affine system;} \end{cases} \\
(D_{\ell,j}^y v)[k_1, \cdot] &= \begin{cases} \downarrow (\tilde{h}_\ell \otimes v[k_1, \cdot]), & \text{wavelet system,} \\ \tilde{h}_{\ell,j} \otimes v[k_1, \cdot], & \text{quasi-affine system,} \end{cases}
\end{aligned}$$

where $\tilde{h}_\ell = \sqrt{2} \cdot \overline{h_\ell[-\cdot]}$ and $\tilde{h}_{\ell,j} = \overline{h_{\ell,j}[-\cdot]}$ with $h_{\ell,j}$ defined as in (2.42). Similarly, we can define one level of reconstruction $D_{\ell,j}^{x*}$ and $D_{\ell,j}^{y*}$ as

$$(D_{\ell,j}^{x*}v)[\cdot, k_2] = \begin{cases} \tilde{h}_\ell^* \otimes \uparrow (v[\cdot, k_2]), & \text{wavelet system,} \\ h_{\ell,j} \otimes v[\cdot, k_2], & \text{quasi-affine system;} \end{cases}$$

$$(D_{\ell,j}^{y*}v)[k_1, \cdot] = \begin{cases} \tilde{h}_\ell^* \otimes \uparrow (v[k_1, \cdot]), & \text{wavelet system,} \\ h_{\ell,j} \otimes v[k_1, \cdot], & \text{quasi-affine system,} \end{cases}$$

where $\tilde{h}_\ell^* = \sqrt{2}h_\ell$.

Now, we present the following 2D fast framelet decomposition and reconstruction algorithm, which unifies the algorithm for wavelet systems and quasi-affine systems. A diagram illustrating one level of decomposition and reconstruction is given in Figure 5.

Algorithm 2.3. *Given signal $v \in \mathbb{R}^{N_1} \times \mathbb{R}^{N_2}$ with $N_1, N_2 \in \mathbb{N}_+$. If the underlying system is affine, then N_1 and N_2 are assumed to be integer multiples of 2^L with L the total level of decompositions. Denote $v_{0,0,0} = v$. Then the L -level fast framelet decomposition and reconstruction is given as follows:*

- (1) **Decomposition:** For each $j = 1, 2, \dots, L$
 (a) Obtain low frequency approximation to v at level j :

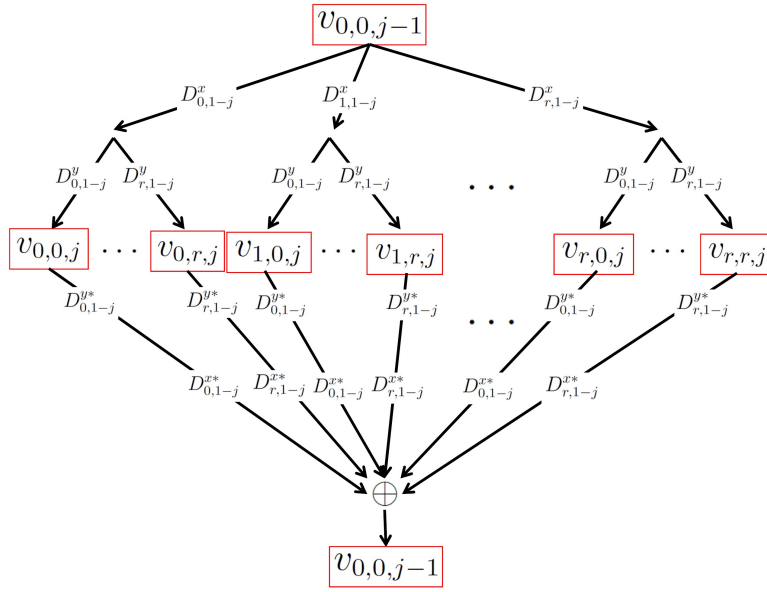
$$v_{0,0,j} = D_{0,1-j}^y(D_{0,1-j}^x v_{0,0,j-1});$$

- (b) Obtain framelet coefficients of v at level j : for each $\ell_1 = 0, 1, \dots, r$ and $\ell_2 = 0, 1, \dots, r$ with $(\ell_1, \ell_2) \neq (0, 0)$,

$$v_{\ell_1, \ell_2, j} = D_{\ell_2, 1-j}^y(D_{\ell_1, 1-j}^x v_{0,0,j-1}).$$

- (2) **Reconstruction:** For each $j = L, L-1, \dots, 1$,

$$v_{0,0,j-1} = \sum_{\ell_1=0}^r \sum_{\ell_2=0}^r D_{\ell_1, 1-j}^{x*}(D_{\ell_2, 1-j}^{y*} v_{\ell_1, \ell_2, j}).$$



1-Level Decomposition & Reconstruction

FIGURE 5. The above diagram illustrates one level of the fast framelet decomposition and reconstruction procedure given by Algorithm 2.3.

Pseudo-splines and tight frames

A class of refinable and compactly supported functions known as pseudo-splines encompasses a wide variety of refinable functions which provide much flexibility in the construction of wavelets and framelets. Functions such as B-splines, interpolatory, or Daubechies' orthogonal refinable functions (see [61, 60]) are special cases of pseudo-splines. Pseudo-splines were first introduced in [65, 165] in order to construct tight framelets with satisfactory approximation order for the truncated frame series of the tight wavelet system generated by the unitary extension principle.

The organization of this lecture is as follows. After a general introduction of pseudo-splines in Section 1, a construction of tight framelets is presented in Section 2. As a special example, we can obtain orthonormal wavelet bases for $L_2(\mathbb{R})$ which were originally constructed by Daubechies in [60]. Furthermore, the technique used to estimate the regularity of pseudo-splines can be applied to discover that the tight frame systems constructed in [65] and [75] always have one framelet whose dilations and shifts already form a Riesz basis for $L_2(\mathbb{R})$. This leads to a new understanding about the structure of pseudo-spline tight frame systems.

An optimal regularity analysis of pseudo-splines does not come easily, as has already been illustrated in a regularity estimate of the orthonormal refinable functions, which is a special case of pseudo-splines (see [54] and [61]). In [75], a systematic regularity analysis of both types of pseudo-splines was given. In Section 3, we provide regularity analysis for pseudo-splines. In particular, the exact decay of the Fourier transforms of the pseudo-splines is given which leads to an estimate of their Hölder regularity. The last section contains two technical lemmata which are needed in the proofs of the properties of pseudo-splines. Finally, we note that some key part of this lecture (e.g. regularity analysis) follows from [75].

1. Definitions and basics

This section is devoted to the definitions of various pseudo-splines and their basic properties.

1.1. Definitions

Pseudo-splines are defined in terms of their refinement masks. It starts with the simple trigonometric identity $1 = \cos^2(\xi) + \sin^2(\xi)$. For given nonnegative integers l and m with $l \leq m - 1$, we have

$$(3.1) \quad 1 = (\cos^2(\xi/2) + \sin^2(\xi/2))^{m+l}.$$

The refinement masks of pseudo-splines are defined by the summation of the first $l + 1$ terms of the binomial expansion of (3.1). In particular, the refinement mask

of a *pseudo-spline of Type I with order (m, l)* is given by, for $\xi \in [-\pi, \pi]$,

$$(3.2) \quad |_1\widehat{a}(\xi)|^2 := |_1\widehat{a}_{(m,l)}(\xi)|^2 := \cos^{2m}(\xi/2) \sum_{j=0}^l \binom{m+l}{j} \sin^{2j}(\xi/2) \cos^{2(l-j)}(\xi/2)$$

and the refinement mask of a *pseudo-spline of Type II with order (m, l)* is given by, for $\xi \in [-\pi, \pi]$,

$$(3.3) \quad {}_2\widehat{a}(\xi) := {}_2\widehat{a}_{(m,l)}(\xi) := \cos^{2m}(\xi/2) \sum_{j=0}^l \binom{m+l}{j} \sin^{2j}(\xi/2) \cos^{2(l-j)}(\xi/2).$$

Define

$$(3.4) \quad P_{m,l}(y) := \sum_{j=0}^l \binom{m+l}{j} y^j (1-y)^{l-j}$$

and

$$(3.5) \quad R_{m,l}(y) := (1-y)^m P_{m,l}(y),$$

where m, l are nonnegative integers with $l \leq m-1$. Then, it is obvious that

$${}_2\widehat{a}(\xi) = R_{m,l}(\sin^2(\xi/2)).$$

Except for some special circumstances, we simply notation by dropping the subscript “ (m, l) ” in $_1\widehat{a}_{(m,l)}(\xi)$ and $_2\widehat{a}_{(m,l)}(\xi)$ for simplicity. We note that the mask of Type I is obtained by taking the square root of the mask of Type II using the Fejér-Riesz lemma (see e.g. [61] and [152]), i.e. ${}_2\widehat{a}(\xi) = |_1\widehat{a}(\xi)|^2$. Type I was introduced and used by [65] to construct tight framelets.

Finally, a family of *dual pseudo-splines* was first discovered by [81] and the general refinement mask of a dual pseudo-spline of order (m, l) with $0 \leq l \leq m-1$ is

$$(3.6) \quad \widehat{b}_{m,l}(\xi) = e^{i\xi/2} \cos^{2m+1}(\xi/2) \sum_{j=0}^l \binom{m+1/2+l}{j} \sin^{2j}(\xi/2) \cos^{2(l-j)}(\xi/2).$$

Notice that dual pseudo-spline of order $(m, 0)$ is the B-spline of degree $2m$ and dual pseudo-spline of order $(m, m-1)$ is the limit function of the dual $2m$ -point scheme. The masks of dual pseudo-splines are very similar in format to those of pseudo-splines. In fact, if we let $r = m + \frac{1}{2}$, then (3.6) can be rewritten as

$$\widehat{b}_{r,l}(\xi) = e^{i\xi/2} \cos^{2r}(\xi/2) \sum_{j=0}^l \binom{r+l}{j} \sin^{2j}(\xi/2) \cos^{2(l-j)}(\xi/2).$$

Therefore, the analysis of dual pseudo-splines is similar to that of pseudo-splines by using Newton’s generalized binomial theorem. Hence, we focus on the analysis of pseudo-splines because the analysis of dual pseudo-splines is analogous. Interested readers should work out the details as exercises.

1.2. Underlying refinable functions

Given refinement masks ${}_k\hat{a}$, the corresponding pseudo-splines can be formally defined in terms of their Fourier transforms ${}_k\hat{\phi}(\xi) := \prod_{j=1}^{\infty} {}_k\hat{a}(2^{-j}\xi)$ for $k = 1, 2$. To make sure that ${}_k\hat{\phi}$ is well defined, we need to consider the convergence of the right-hand-side which is given below. For finitely supported refinement masks, the following theorem is well known and one can find the proof in e.g. [61]. Here, we provide an alternative proof motivated from a proof of [116].

Proposition 3.1. [116] *Suppose that h_0 is a finitely supported refinement mask satisfying $\hat{h}_0(0) = 1$. Then, the product*

$$(3.7) \quad \hat{\phi}(\xi) := \prod_{j=1}^{\infty} \hat{h}_0(2^{-j}\xi)$$

converges uniformly on any compact subset of \mathbb{R} . Furthermore, $\hat{\phi}$ is the Fourier transform of a compactly supported tempered distribution ϕ that satisfies refinement equation (1.6).

PROOF. Since \hat{h}_0 is a trigonometric polynomial with $\hat{h}_0(0) = 1$, we have

$$(3.8) \quad |\hat{h}_0(\xi)| \leq 1 + C|\xi| \leq e^{C|\xi|},$$

for some $C > 0$. This leads to

$$\prod_{j=1}^{\infty} |\hat{h}_0(2^{-j}\xi)| \leq e^{C|\xi|}.$$

Therefore, the right-hand-side of (3.7) converges uniformly on any compact subset K of \mathbb{R} . It is clear that $\hat{\phi}$ satisfies refinement equation (1.6). Note that in the above proof, we only require that the refinement mask satisfy (3.8). The finiteness of the support of the refinement mask is not needed.

Let $c_0 = \max_{\xi} |\hat{h}_0(\xi)|$ and $c_1 = \max_{|\xi| \leq 2} |\hat{\phi}(\xi)|$. For any positive j , consider ξ , such that $2^j \leq |\xi| < 2^{j+1}$. Since

$$\hat{\phi}(\xi) = \prod_{k=1}^j \hat{h}_0(2^{-k}\xi) \hat{\phi}(2^{-j}\xi),$$

we have

$$(3.9) \quad |\hat{\phi}(\xi)| \leq c_1 c_0^j \leq c_1 c_0^{\log_2 |\xi|} = c_1 |\xi|^{\log_2 c_0}.$$

Therefore, $\hat{\phi}$ has at most polynomial growth at infinity. Hence, ϕ is a tempered distribution with $\hat{\phi}(0) = 1$.

Finally it remains to prove that ϕ is compactly supported. Let $\phi_0 = \delta$ be the Dirac function. Then it is a compactly supported tempered distribution whose Fourier transform is 1. Define the cascade algorithm

$$\hat{\phi}_{n+1}(\xi) = \hat{h}_0(\xi/2) \hat{\phi}_n(\xi/2).$$

Then, the infinity product in the right-hand-side of (3.7) converges uniformly on any compact subset \mathbb{R} if and only if ϕ_n does. Furthermore, (3.9) shows that $\hat{\phi}_n$ has at most polynomial growth at infinity. Assume that the support of the refinement mask is $[-q, q]$. Let K be any compact subset in \mathbb{R} containing $[-q, q]$. Then it can be shown inductively that the support of ϕ_n is in K for all $n \geq 0$. Now, let g be

an arbitrary test function whose support is disjoint with K . Then, applying the dominated converge theorem, one has

$$\langle \phi, g \rangle = (2\pi)^{1/2} \langle \widehat{\phi}, \widehat{g} \rangle = (2\pi)^{1/2} \langle \lim_n \widehat{\phi}_n, \widehat{g} \rangle = (2\pi)^{1/2} \lim_n \langle \widehat{\phi}_n, \widehat{g} \rangle = \lim_n \langle \phi_n, g \rangle = 0.$$

Hence, ϕ is compactly supported. \blacksquare

One of the fundamental questions for the study of refinable functions is that given a refinement mask h_0 , does there exist a corresponding refinable function $\phi \in L_2(\mathbb{R})$ such that the refinement equation (1.5) is satisfied. This question will be answered by the following result where part of the proof is taken from [32] (see [125] for a more general analysis). To establish these results, we need to introduce the cascade algorithm. For a given refinement mask h_0 , define the *cascade algorithm* as

$$(3.10) \quad \widehat{f}_n(\xi) = \widehat{h}_0\left(\frac{\xi}{2}\right) \widehat{f}_{n-1}\left(\frac{\xi}{2}\right) = \prod_{j=1}^n \widehat{h}_0(2^{-j}\xi) \widehat{f}_0(2^{-n}\xi), \quad n > 0,$$

with initial function f_0 satisfying $\widehat{f}_0(\xi) = \chi_{[-\pi, \pi]}(\xi)$. Proposition 3.1 shows that f_n converges uniformly on each compact set.

Theorem 3.1. [32] *Suppose h_0 is finitely supported and satisfies the following condition:*

$$(3.11) \quad \begin{cases} |\widehat{h}_0(\xi)|^2 + |\widehat{h}_0(\xi + \pi)|^2 \leq 1, & a.e. \quad \xi \in \mathbb{R}; \\ \widehat{h}_0(0) = 1. \end{cases}$$

Then, the limit ϕ of the cascade algorithm (3.10) is a compactly supported function in $L_2(\mathbb{R})$ satisfying $\widehat{\phi}(0) = 1$.

PROOF. Since h_0 is finitely supported and $\widehat{h}_0(0) = 1$, then by Proposition 3.1, there exists a unique compactly supported refinable distribution ϕ whose Fourier transform given by (3.7) satisfies $\widehat{\phi}(0) = 1$ (also see [61, 29]). Now, consider

$$\begin{aligned} \|\widehat{f}_n\|_{L_2(\mathbb{R})}^2 &= \int_{-2^n\pi}^{2^n\pi} \prod_{j=1}^n |\widehat{h}_0(2^{-j}\xi)|^2 d\xi \\ (\widehat{h}_0 \text{ is } 2\pi\text{-periodic}) &= \int_0^{2^{n+1}\pi} \prod_{j=1}^n |\widehat{h}_0(2^{-j}\xi)|^2 d\xi \\ &= \int_0^{2^n\pi} \prod_{j=1}^n |\widehat{h}_0(2^{-j}\xi)|^2 d\xi + \int_{2^n\pi}^{2^{n+1}\pi} \prod_{j=1}^n |\widehat{h}_0(2^{-j}\xi)|^2 d\xi \\ &= \int_0^{2^n\pi} \prod_{j=1}^n |\widehat{h}_0(2^{-j}\xi)|^2 d\xi + \int_0^{2^n\pi} \prod_{j=1}^n |\widehat{h}_0(2^{-j}(\xi + 2^n\pi))|^2 d\xi \\ &= \int_0^{2^n\pi} \prod_{j=1}^{n-1} |\widehat{h}_0(2^{-j}\xi)|^2 \left(|\widehat{h}_0(2^{-n}\xi)|^2 + |\widehat{h}_0(2^{-n}\xi + \pi)|^2 \right) d\xi. \end{aligned}$$

By (3.11), we have

$$\|\widehat{f}_n\|_{L_2(\mathbb{R})}^2 \leq \int_0^{2^n\pi} \prod_{j=1}^{n-1} |\widehat{h}_0(2^{-j}\xi)|^2 d\xi = \|\widehat{f}_{n-1}\|_{L_2(\mathbb{R})}^2.$$

Then by induction, we have

$$\|\widehat{f}_n\|_{L_2(\mathbb{R})}^2 \leq \|\widehat{f}_0\|_{L_2(\mathbb{R})}^2 = 2\pi, \quad \text{for all } n \geq 0.$$

Since the sequence $\{\widehat{f}_n\}_n$ converges to $\widehat{\phi}$ pointwise, then by Fatou's lemma

$$\|\widehat{\phi}\|_{L_2(\mathbb{R})}^2 \leq \liminf_{n \rightarrow \infty} \|\widehat{f}_n\|_{L_2(\mathbb{R})}^2 \leq 2\pi < \infty,$$

and hence $\phi \in L_2(\mathbb{R})$. ■

By the definition of the refinement masks of the pseudo-splines, the pseudo-splines of both types satisfy (3.11), and hence all pseudo-splines are compactly supported functions in $L_2(\mathbb{R})$. In fact, they are much smoother than merely being in $L_2(\mathbb{R})$ as is shown in Section 3. Furthermore, from Theorem 1.4 in Lecture 1 the sequence of subspaces $\{V_n\}_n$ (defined via (1.1)) generated by any pseudo-spline forms an MRA for $L_2(\mathbb{R})$.

Since most pseudo-splines do not have an analytical form (B-splines as the exception). It is important to know whether the infinite product (3.7) converges in $L_2(\mathbb{R})$, so that one can obtain the approximation of the corresponding refinable function.

Proposition 3.2. *The cascade algorithm defined in (3.10) by the refinement mask h_0 of an arbitrary given pseudo-spline with any type and order converges to the corresponding pseudo-spline ϕ in $L_2(\mathbb{R})$.*

PROOF. Since we already have pointwise convergence of the cascade algorithm by Proposition 3.1, all we need to show is that $\|\widehat{\phi}\|_{L_2(\mathbb{R})}^2 = \lim_{n \rightarrow \infty} \|\widehat{f}_n\|_{L_2(\mathbb{R})}^2$. Indeed, we have

$$0 \leq \|\widehat{f}_n - \widehat{\phi}\|_{L_2(\mathbb{R})}^2 = \langle \widehat{f}_n - \widehat{\phi}, \widehat{f}_n - \widehat{\phi} \rangle = \|\widehat{f}_n\|_{L_2(\mathbb{R})}^2 - 2\operatorname{Re}\langle \widehat{f}_n, \widehat{\phi} \rangle + \|\widehat{\phi}\|_{L_2(\mathbb{R})}^2.$$

Since \widehat{f}_n pointwise converges to $\widehat{\phi}$, one needs to show that $\lim_{n \rightarrow \infty} \langle \widehat{f}_n, \widehat{\phi} \rangle = \|\widehat{\phi}\|_{L_2(\mathbb{R})}^2$ and $\|\widehat{\phi}\|_{L_2(\mathbb{R})}^2 = \lim_{n \rightarrow \infty} \|\widehat{f}_n\|_{L_2(\mathbb{R})}^2$. This leads to

$$0 \leq \lim_{n \rightarrow \infty} \|\widehat{f}_n - \widehat{\phi}\|_{L_2(\mathbb{R})}^2 \leq \lim_{n \rightarrow \infty} \left(\|\widehat{f}_n\|_{L_2(\mathbb{R})}^2 - \|\widehat{\phi}\|_{L_2(\mathbb{R})}^2 \right).$$

By the dominated convergence theorem, it suffices to show that

$$(3.12) \quad |\widehat{f}_n|^2 \leq C|\widehat{\phi}|^2 \quad \text{and} \quad |\widehat{f}_n\widehat{\phi}| \leq C|\widehat{\phi}|^2$$

Since we have already shown that $\phi \in L_2(\mathbb{R})$, for any given $n \geq 0$, the existence of $C > 0$ satisfying (3.12) is obvious when $|\xi| > 2^n\pi$ since $\widehat{f}_n(\xi) = 0$. Now, for $|\xi| \leq 2^n\pi$, since ϕ is refinable, it follows that

$$|\widehat{\phi}(\xi)|^2 = \left| \prod_{j=1}^{\infty} \widehat{h}_0(2^{-j}\xi) \right|^2 = \left| \prod_{j=1}^n \widehat{h}_0(2^{-j}\xi) \prod_{j=n+1}^{\infty} \widehat{h}_0(2^{-j}\xi) \right|^2 = |\widehat{f}_n(\xi)|^2 |\widehat{\phi}(2^{-n}\xi)|^2.$$

By (3.16) one can easily check that the inequality

$$|\widehat{\phi}(\xi)|^2 \geq 1/C > 0, \quad \xi \in [-\pi, \pi]$$

holds for some $C > 0$ for all pseudo-splines. Hence, (3.12) holds for all pseudo-splines. ■

1.3. Approximation

Recall the definition of the refinement mask of a pseudo-spline of Type II with order (m, l)

$${}_2\hat{a}(\xi) = \cos^{2m}(\xi/2) \sum_{j=0}^l \binom{m+l}{j} \sin^{2j}(\xi/2) \cos^{2(l-j)}(\xi/2).$$

One can see that a pseudo-spline of Type II with order (m, l) is a convolution of a B-spline of order $2m$ with a distribution. The mask of the distribution is $P_{m,l}(\sin^2(\xi/2))$, where $P_{m,l}(y)$ is given by (3.4). Since Lemma 3.3 says that $P_{m,l}(y) = \sum_{j=0}^l \binom{m-1+j}{j} y^j$, one has $P_{m,l}(\sin^2(\xi/2)) \geq 1$ for all ξ . Hence, pseudo-spline of Type II with order (m, l) can be viewed as a convolution of a B-spline with a distribution whose Fourier transform is at least 1 at any point ξ . Therefore, the pseudo-spline ${}_2\phi$ of Type II with order (m, l) satisfies the SF condition of order $2m$. By a similar discussion with some special treatment of the case when m is odd, one can conclude that the pseudo-spline ${}_1\phi$ of Type I with order (m, l) satisfies the SF condition of order m .

Consider the masks of ${}_1\hat{a}$ of the pseudo-splines of type I of order (m, l) . Since $1 - |{}_1\hat{a}|^2$ has a factor $\sin^{2(l+1)}(\xi/2)$, we have $1 - |{}_1\hat{a}|^2 = O(|\cdot|^{2l+2})$. Then by Theorem 1.5 and Proposition 2.1, the operator $\mathcal{P}_n f$ defined by (1.13) corresponding to the pseudo-spline ${}_1\phi$ of order (m, l) provides approximation order $\min\{m, 2l + 2\}$ for sufficiently smooth functions. Hence, Theorem 2.3 implies that any tight frame system constructed by the UEP using pseudo-splines of type I of order (m, l) provides approximation order $\min\{m, 2l + 2\}$.

The following theorem shows that, if we start from pseudo-splines of Type II with order (m, l) in Construction 3.3, the corresponding tight frame system $X(\Psi)$ provides approximation order $2l + 2$.

Theorem 3.2. *Let m and l be nonnegative integers satisfying $l \leq m - 1$. Let ${}_k\phi$, $k = 1, 2$ be the pseudo-spline of Type I and II with order (m, l) and ${}_k\hat{a}$, $k = 1, 2$ be its refinement mask. Let \mathcal{P}_n be defined by (1.13) for a given refinable function ${}_k\phi$. Then, \mathcal{P}_n provides approximation order $\min\{m, 2l + 2\}$ for Type I and $2l + 2$ for type II. Consequently, the corresponding truncated operator \mathcal{Q}_n defined in (2.21) for any tight wavelet frame system constructed from a refinable function ${}_k\phi$, $k = 1, 2$ of order (m, l) via the unitary extension principle provides approximation order $\min\{m, 2l + 2\}$ for Type I and $2l + 2$ for Type II.*

PROOF. Notice that ${}_2\hat{a} = |{}_1\hat{a}|^2$ and $1 - {}_2\hat{a}^2 = (1 + {}_2\hat{a})(1 - {}_2\hat{a})$. Since $1 + {}_2\hat{a}(0) \neq 0$, the order of zeros of $1 - {}_2\hat{a}^2$ at the origin is the same as that of $1 - {}_2\hat{a}$, which is exactly $2l + 2$ based on the discussion above. This, together with Theorem 1.5 and Proposition 2.1, concludes the proof of the theorem. ■

For fixed m , since the value of the mask $|{}_k\hat{a}(\xi)|$, for $k = 1, 2$ and $\xi \in \mathbb{R}$, increases with l (by (1) of Lemma 3.3 in section 4), and the length of the mask ${}_ka$ also increases with l , we conclude that the decay rate of the Fourier transform of a pseudo-spline decreases with l (see Section 3 for details) and the support of the corresponding pseudo-spline increases with l . In particular, for fixed m , the pseudo-spline with order $(m, 0)$ has the highest order of smoothness with the shortest support, the pseudo-spline with order $(m, m - 1)$ has the lowest order of

smoothness with the largest support in the family. When we move from B-splines to orthogonal or interpolatory refinable functions, we sacrifice the smoothness and short support of the B-splines to gain some other desirable properties, such as orthogonality or interpolatory property. What do we get for the pseudo-splines of other orders? When we move from B-splines to pseudo-splines, we gain the approximation power of the truncated tight frame systems derived from them and the vanishing moments of wavelets.

As mentioned in Lecture 1, if ϕ satisfies the SF condition of order m , then ϕ (or $V(\phi)$, or the underlying MRA) provides approximation order m . Therefore, even though the $(V_n)_{n \in \mathbb{Z}}$ may be generated by a different pseudo-spline of order (m, l) with m fixed, the corresponding spaces $(V_n)_{n \in \mathbb{Z}}$ provide the same approximation order. However, in many applications we are more interested in the performance of the tight frame system constructed using UEP. Therefore, we normally use the truncated operator \mathcal{Q}_n defined in (2.21) or, equivalently, the quasi-interpolatory operator $\mathcal{P}_n : L_2(\mathbb{R}) \mapsto V_n$ defined in (2.11) to approximate f . Theorem 3.2 tells us that for B-splines, the approximation order of the corresponding \mathcal{Q}_n (hence \mathcal{P}_n) can never exceed 2 even if a high order B-spline is used. On the other hand, for pseudo-spline of either type with order (m, l) , $0 \leq l \leq m - 1$, the approximation order of \mathcal{Q}_n (hence \mathcal{P}_n) is $\min\{m, 2l + 2\}$ for Type I and $2l + 2$ for Type II. This indicates that the tight frame system derived from a pseudo-spline normally gives better approximation order when the truncated series is used to approximate the underlying functions than that derived from B-splines. For fixed m , the choice of l depends entirely on applications. According to the practical problems in hand, one must properly balance the approximation order, the length of support of the wavelet, and regularity.

1.4. Stability and orthogonality

Recall that ϕ is *stable*, if $\{\phi(\cdot - k)\}_{k \in \mathbb{Z}}$ forms a Riesz basis for $V(\phi)$, i.e. there exist $0 < C_1, C_2 < \infty$, such that for any sequence $b \in \ell_2(\mathbb{Z})$,

$$(3.13) \quad C_1 \|b\|_{\ell_2(\mathbb{Z})} \leq \left\| \sum_{k \in \mathbb{Z}} b[k] \phi(\cdot - k) \right\|_{L_2(\mathbb{R})} \leq C_2 \|b\|_{\ell_2(\mathbb{Z})}.$$

In particular if only the upper bound of (3.13) is satisfied, we say that $\{\phi(\cdot - k)\}_{k \in \mathbb{Z}}$ forms a *Bessel* sequence (as we have already discussed in Lecture 1). On the other hand, if $C_1 = C_2 = 1$, then we say that ϕ is *orthonormal*, i.e. $\{\phi(\cdot - k)\}_{k \in \mathbb{Z}}$ forms an orthonormal basis for $V(\phi)$. Recall that ϕ is orthonormal if and only if $\langle \phi, \phi(\cdot - k) \rangle = \delta_{k,0}$ for all $k \in \mathbb{Z}$.

The stability of function $\phi \in L_2(\mathbb{R})$ can also be characterized by its bracket product which is well known in the literature (see e.g. [70, 61, 114, 156]). Now, we summarize and prove this characterization in the following lemma.

Lemma 3.1. *For any given $\phi \in L_2(\mathbb{R})$, then ϕ is stable, i.e. (3.13) is satisfied, if and only if $[\widehat{\phi}, \widehat{\phi}]$ satisfies*

$$(3.14) \quad 0 < C_1 \leq [\widehat{\phi}, \widehat{\phi}](\xi) \leq C_2;$$

holds for almost every $\xi \in \mathbb{R}$ with

$$\text{ess inf } [\widehat{\phi}, \widehat{\phi}] = C_1 \quad \text{and} \quad \text{ess sup } [\widehat{\phi}, \widehat{\phi}] = C_2.$$

As a consequence, ϕ and its shifts form an orthonormal basis of $V(\phi)$ if and only if $[\widehat{\phi}, \widehat{\phi}] = 1$ holds almost everywhere. Furthermore, assume that ϕ is compactly supported, and then the upper bound of (3.14) or equivalently (3.13) is satisfied.

PROOF. For simplicity, we denote $\tilde{\phi} := [\widehat{\phi}, \widehat{\phi}]^{\frac{1}{2}}$. Then for all finite sequences $b \in \ell_2(\mathbb{Z})$, we have the following identity (see proof of part (1) of Proposition 1.3)

$$(3.15) \quad 2\pi \left\| \sum_{k \in \mathbb{Z}} b[k] \phi(\cdot - k) \right\|_{L_2(\mathbb{R})}^2 = \|\widehat{b\tilde{\phi}}\|_{L_2([- \pi, \pi])}^2.$$

Suppose $\tilde{\phi}$ satisfies (3.14). Then the upper bound of (3.14) implies that (3.15) is satisfied for all $b \in \ell_2(\mathbb{Z})$. Therefore,

$$C_1 \|\widehat{b}\|_{L_2([- \pi, \pi])}^2 \leq \|\widehat{b\tilde{\phi}}\|_{L_2([- \pi, \pi])}^2 \leq C_2 \|\widehat{b}\|_{L_2([- \pi, \pi])}^2$$

with

$$C_1 = \text{ess inf } [\widehat{\phi}, \widehat{\phi}] \quad \text{and} \quad C_2 = \text{ess sup } [\widehat{\phi}, \widehat{\phi}],$$

and hence (3.13) follows.

Now suppose the upper bound of (3.13) is satisfied. If $\tilde{\phi}$ is not essentially bounded, then there exists a sequence of domains $\Omega_n \subset [-\pi, \pi]$ with positive measure such that $\tilde{\phi} > n$ on Ω_n . Let c^n be a sequence in $\ell_2(\mathbb{Z})$ satisfying $\widehat{c^n} = \chi_{\Omega_n}$ as $n \rightarrow \infty$. Then by (3.15), we have

$$2\pi \left\| \sum_{k \in \mathbb{Z}} c^n[k] \phi(\cdot - k) \right\|_{L_2(\mathbb{R})}^2 = \|\widehat{c^n \tilde{\phi}}\|_{L_2([- \pi, \pi])}^2 \geq n \|\widehat{c^n}\|_{L_2([- \pi, \pi])}^2 = 2\pi n \|c^n\|_{\ell_2(\mathbb{Z})}^2.$$

This contradicts with the upper bound of (3.13). Therefore, $\tilde{\phi}$ must be essentially bounded and the upper bound of (3.14) follows. Notice that $\tilde{\phi}$ is essentially bounded away from 0 if and only if $1/\tilde{\phi}$ is essentially bounded. Then using a similar argument we can show that the lower bound of (3.14) follows from that of (3.13).

Now, we show that whenever ϕ is compactly supported, the upper bound of (3.14), or equivalently (3.13), is satisfied automatically. Indeed, part (2) of Proposition 1.3 implies that $\tilde{\phi}$ is the Fourier series of $\{\langle \phi, \phi(\cdot - k) \rangle : k \in \mathbb{Z}\}$ which is a finite sequence since ϕ is compactly supported. Then $\tilde{\phi}$ is in fact a trigonometric polynomial which is continuous and hence bounded on $[-\pi, \pi]$. ■

Consider the compactly supported centered B-splines

$$\widehat{B}_m(\xi) = e^{-ij\frac{\xi}{2}} \left(\frac{\sin(\xi/2)}{\xi/2} \right)^m.$$

Since it is obvious that there is $C_1 > 0$, such that $|\widehat{B}_m(\xi)| > \sqrt{C_1}$ for all $\xi \in [-\pi, \pi]$, we have

$$[\widehat{B}_m, \widehat{B}_m](\xi) = |\widehat{B}_m(\xi)|^2 + \sum_{k \in \mathbb{Z} \setminus 0} |\widehat{B}_m(\xi + 2k\pi)|^2 \geq |\widehat{B}_m(\xi)|^2 \geq C_1.$$

Hence all B-splines are stable. Next, we prove that all pseudo-splines are stable as given by [74].

Proposition 3.3. [74] *All pseudo-splines for both types with arbitrary orders are stable.*

PROOF. Here, we only prove stability for pseudo-splines of type II. The arguments for pseudo-splines of type I are similar.

Let $\phi_{(m,l)}$ be pseudo-spline of type II with order (m, l) and $\widehat{a}_{(m,l)}$ be its refinement mask. Since pseudo-splines are compactly supported functions in $L_2(\mathbb{R})$, Lemma 3.1 tells us that we only need to show that there is a constant $C_1 > 0$ such that $C_1 \leq [\widehat{\phi}_{(m,l)}, \widehat{\phi}_{(m,l)}](\xi)$ holds for every $\xi \in \mathbb{R}$.

Since for each fixed $m \geq 1$ and for every $0 \leq l \leq m - 1$, we have

$$P_{m,l}(\sin^2(\xi/2)) \geq 1,$$

and the following inequality

$$\cos^{2m}(\xi/2) \leq \widehat{a}_{(m,l)}(\xi)$$

holds for all $\xi \in \mathbb{R}$. Therefore, by (3.7), we have for all $\xi \in \mathbb{R}$,

$$(3.16) \quad |\widehat{B}_{2m}(\xi)| \leq |\widehat{\phi}_{(m,l)}(\xi)|.$$

Since B_{2m} is stable, there is a constant $C_1 > 0$ such that $C_1 \leq [\widehat{\phi}_{(m,l)}, \widehat{\phi}_{(m,l)}](\xi)$ holds for every $\xi \in \mathbb{R}$. Altogether, we have that $\phi_{(m,l)}$ is stable. ■

Let ϕ_m be the pseudo-splines of type I with order $(m, m - 1)$. Next, we show that ϕ_m is orthonormal, which means ϕ_m satisfies (3.13) with $C_1 = C_2 = 1$. By Lemma 3.1, we know that ϕ_m is orthonormal if and only if

$$(3.17) \quad [\widehat{\phi}_m, \widehat{\phi}_m] = 1$$

almost everywhere. We now have the following proposition.

Proposition 3.4. *Pseudo-splines of type I with order $(m, m - 1)$ are orthonormal refinable functions.*

PROOF. All we need to show is that $[\widehat{\phi}_m, \widehat{\phi}_m] = 1$ a.e. We first note that for any $k \in \mathbb{Z}$,

$$\int_{-\pi}^{\pi} [\widehat{\phi}_m, \widehat{\phi}_m](\xi) e^{ik\xi} d\xi = \int_{\mathbb{R}} |\widehat{\phi}_m(\xi)|^2 e^{ik\xi} d\xi.$$

Define f_0 such that $\widehat{f}_0(\xi) = \chi_{[-\pi, \pi]}(\xi)$. Using the identity (3.27) and following the same arguments as in the proof of Theorem 3.4, we have

$$\int_{\mathbb{R}} |\widehat{f}_n(\xi)|^2 e^{ik\xi} d\xi = \int_{\mathbb{R}} |\widehat{f}_0(\xi)|^2 e^{ik\xi} d\xi = \int_{-\pi}^{\pi} e^{ik\xi} d\xi = 2\pi\delta_{k,0}.$$

Then by the dominated convergence theorem (similarly as the proof of Theorem 3.4), we have

$$\int_{\mathbb{R}} |\widehat{\phi}_m(\xi)|^2 e^{ik\xi} d\xi = 2\pi\delta_{k,0},$$

which means

$$\int_{-\pi}^{\pi} [\widehat{\phi}_m, \widehat{\phi}_m](\xi) e^{ik\xi} d\xi = 2\pi\delta_{k,0}.$$

This proves that $[\widehat{\phi}_m, \widehat{\phi}_m] = 1$ a.e. ■

Following arguments similar to those used in proving the previous proposition, we can show that all pseudo-splines of type II with order $(m, m - 1)$ are indeed interpolatory as given by the following proposition. Recall that a continuous function $\phi \in L_2(\mathbb{R})$ is said to be interpolatory if $\phi(k) = \delta_{k,0}$ for $k \in \mathbb{Z}$ (see e.g. [79]).

One of the key steps in the proof of the following proposition is to show that $|\widehat{\phi}(\xi)| \leq C(1 + |\xi|)^{-1-\varepsilon}$. In fact, the Fourier transform of pseudo-splines of type II with order $(m, m-1)$ has a faster decay than $O((1 + |\xi|)^{-1-\varepsilon})$, which means they are actually smoother than merely continuous. We provide an optimal decay analysis of all pseudo-splines in Theorem 3.8.

Proposition 3.5. *All pseudo-splines of type II with order $(m, m-1)$ are interpolatory.*

PROOF. Let ϕ be pseudo-spline of type II with order $(m, m-1)$ and \widehat{a} be its refinement mask. First, we show that $|\widehat{\phi}(\xi)| \leq C(1 + |\xi|)^{-1-\varepsilon}$ for some $\varepsilon > 0$. Such decay makes sure that $\widehat{\phi} \in L_1(\mathbb{R})$, which ensures the continuity of ϕ and thus $\phi(k)$ is well-defined for all $k \in \mathbb{Z}$.

Now, we show the decay of $\widehat{\phi}$. Recall from (3.3) that

$$\widehat{a}(\xi) = \cos^{2m}(\xi/2) \sum_{j=0}^{m-1} \binom{2m-1}{j} \sin^{2j}(\xi/2) \cos^{2(m-j-1)}(\xi/2).$$

Define

$$\mathcal{L}(\xi/2) := \sum_{j=0}^{m-1} \binom{2m-1}{j} \sin^{2j}(\xi/2) \cos^{2(m-j-1)}(\xi/2).$$

Then

$$\begin{aligned} |\widehat{\phi}(\xi)| &= \prod_{j=1}^{\infty} \left| \left(\frac{1 + e^{-i(2^{-j}\xi)}}{2} \right) \right|^{2m} \prod_{j=1}^{\infty} \mathcal{L}(2^{-j}\xi) \\ (3.18) \quad &= \left| \left(\frac{1 - e^{-i\xi}}{i\xi} \right) \right|^{2m} \prod_{j=1}^{\infty} \mathcal{L}(2^{-j}\xi) \\ &\leq C(1 + |\xi|)^{-2m} \prod_{j=1}^{\infty} \mathcal{L}(2^{-j}\xi). \end{aligned}$$

By item 1 of Lemma 3.3, it is easy to see that $\sup_{\xi} \mathcal{L}(\xi) = 2^{2m-2}$. We now provide a decay estimate of $\prod_{j=1}^{\infty} \mathcal{L}(2^{-j}\xi)$. Since \mathcal{L} is a trigonometric polynomial with $\mathcal{L}(0) = 1$, we have $\mathcal{L}(\xi) \leq 1 + C|\xi|$. Therefore,

$$\sup_{|\xi| \leq 1} \prod_{j=1}^{\infty} \mathcal{L}(2^{-j}\xi) \leq \sup_{|\xi| \leq 1} \prod_{j=1}^{\infty} e^{C2^{-j}|\xi|} \leq e^C.$$

For a given $|\xi| \geq 1$, there exists $J \geq 1$ such that $2^{J-1} \leq |\xi| \leq 2^J$. Therefore,

$$\begin{aligned} \prod_{j=1}^{\infty} \mathcal{L}(2^{-j}\xi) &= \prod_{j=1}^J \mathcal{L}(2^{-j}\xi) \prod_{j=1}^{\infty} \mathcal{L}(2^{-j-J}\xi) \\ (3.19) \quad &\leq 2^{J(2m-2)} e^C \\ &\leq C(1 + |\xi|)^{2m-2}. \end{aligned}$$

Combining inequalities (3.18) and (3.19), we have the desired decay estimate of $\widehat{\phi}$.

Now, proving ϕ is interpolatory is equivalent to showing that

$$(3.20) \quad \sum_{k \in \mathbb{Z}} \phi(k) e^{-ik\xi} = 1$$

almost everywhere. Define $\phi^o(\xi) := \sum_{k \in \mathbb{Z}} \widehat{\phi}(\xi + 2\pi k)$. Since $\widehat{\phi} \in L_1(\mathbb{R})$, it is easy to show that $\phi^o \in L_1([-\pi, \pi])$. Considering the Fourier coefficients of ϕ^o , we have

$$\widehat{\phi^o}[k] = \int_{-\pi}^{\pi} \phi^o(\xi) e^{ik\xi} d\xi = \int_{\mathbb{R}} \widehat{\phi}(\xi) e^{ik\xi} d\xi = 2\pi\phi(k).$$

Since ϕ is compactly supported, we know that $\{\phi[k] : k \in \mathbb{Z}\}$ is a finite sequence. Then $\sum_{k \in \mathbb{Z}} |\widehat{\phi^o}[k]| < \infty$, which means ϕ^o has a Fourier series with absolute convergence. Therefore,

$$(3.21) \quad \sum_{k \in \mathbb{Z}} \widehat{\phi}(\xi + 2\pi k) = \phi^o(\xi) = \sum_{k \in \mathbb{Z}} \phi(k) e^{-ik\xi}.$$

Identity (3.21) is known as the *Poisson summation formula*. Combining (3.21) with (3.20) we have that ϕ is interpolatory if and only if

$$\sum_{k \in \mathbb{Z}} \widehat{\phi}(\xi + 2\pi k) = 1$$

almost everywhere. Noting that $\widehat{a} + \widehat{a}(\cdot + \pi) = 1$, the rest of the proof follows similar to that of Proposition 3.4 and we leave the details to the readers as an exercise. ■

Note that pseudo-splines with order $(m, 0)$ for both types are in fact B-splines. Pseudo-splines of Type I with order $(m, m - 1)$ are the refinable functions with orthonormal shifts (called orthogonal refinable functions) given in [60]. The pseudo-splines of Type II with order $(m, m - 1)$ are the interpolatory refinable functions (which were first introduced in [79] and a systematic construction was given in [60]). The other pseudo-splines fill the gap between the B-splines and orthogonal or interpolatory refinable functions.

Similarly, we have the stability of all dual pseudo-splines.

Proposition 3.6. [72] *All dual pseudo-splines are stable.*

2. Wavelets From pseudo-splines

For a given pseudo-spline, there are many ways to construct pseudo-spline tight wavelet frames from it. Here, we provide a general construction of pseudo-spline tight wavelet frames.

2.1. Framelets

In this Section, we give a construction of a tight wavelet frame system from pseudo-splines. The construction is from [75] which is motivated from [51] and one of the constructions of [65]. The construction can be applied to any refinable function whose mask is a trigonometric polynomial and satisfies (3.11).

Construction 3.3. *Let $\phi \in L_2(\mathbb{R})$ be a compactly supported refinable function with its trigonometric polynomial refinement mask \widehat{h}_0 satisfying $\widehat{h}_0(0) = 1$ and (3.11). Let*

$$T = 1 - |\widehat{h}_0|^2 - |\widehat{h}_0(\cdot + \pi)|^2 \quad \text{and} \quad \mathcal{A} := \frac{\sqrt{T}}{2},$$

where \sqrt{T} is obtained via the Fejér-Riesz lemma. Define

$$\widehat{h}_1(\xi) := e^{-i\xi} \overline{\widehat{h}_0(\xi + \pi)}, \quad \widehat{h}_2(\xi) := \mathcal{A}(\xi) + e^{-i\xi} \mathcal{A}(-\xi)$$

and

$$\widehat{h}_3(\xi) := e^{-i\xi} \overline{\widehat{h}_2(\xi + \pi)}.$$

Let $\Psi := \{\psi_1, \psi_2, \psi_3\}$, where

$$\widehat{\psi}_\ell(\xi) = \widehat{h}_\ell(\xi/2) \widehat{\phi}(\xi/2), \quad \ell = 1, 2, 3.$$

Then $X(\Psi)$ is a tight frame for $L_2(\mathbb{R})$. Moreover, assume that ϕ , hence its refinement mask \widehat{h}_0 , is symmetric about the origin. Then, ψ_1 is symmetric about $\frac{1}{2}$, ψ_2 is symmetric about $\frac{1}{4}$ and ψ_3 is antisymmetric about $\frac{1}{4}$.

PROOF. By Theorem 2.2, in order to verify that $X(\Psi)$ is a tight frame for $L_2(\mathbb{R})$, one needs to show that the masks $\{\widehat{h}_0, \widehat{h}_1, \widehat{h}_2, \widehat{h}_3\}$ satisfy (2.6). Using the definitions of \widehat{h}_1 and \widehat{h}_3

$$\widehat{h}_1 = e^{-i\xi} \overline{\widehat{h}_0(\cdot + \pi)} \quad \text{and} \quad \widehat{h}_3 = e^{-i\xi} \overline{\widehat{h}_2(\cdot + \pi)},$$

we have

$$\sum_{\ell=0}^3 \widehat{h}_\ell \overline{\widehat{h}_\ell(\cdot + \pi)} = \widehat{h}_0 \overline{\widehat{h}_0(\cdot + \pi)} - \widehat{h}_0 \overline{\widehat{h}_0(\cdot + \pi)} + \widehat{h}_2 \overline{\widehat{h}_2(\cdot + \pi)} - \widehat{h}_2 \overline{\widehat{h}_2(\cdot + \pi)} = 0.$$

Next, we show that

$$(3.22) \quad \sum_{\ell=0}^3 |\widehat{h}_\ell|^2 = 1.$$

Since

$$|\widehat{h}_0|^2 + |\widehat{h}_1|^2 = |\widehat{h}_0|^2 + |\widehat{h}_0(\cdot + \pi)|^2,$$

it remains to show that

$$|\widehat{h}_2|^2 + |\widehat{h}_3|^2 = 1 - |\widehat{h}_0|^2 - |\widehat{h}_0(\cdot + \pi)|^2 = T.$$

Since h_0 has real coefficients, we have $\widehat{h}_0(-\xi) = \overline{\widehat{h}_0(\xi)}$, and thus $T(\xi) = T(-\xi)$. Furthermore, $T(\xi)$ is nonnegative and π -periodic. Then by the Fejér-Riesz lemma (and spectral factorization [61, Lemma 6.1.3]), the function $\mathcal{A}(\xi)$ is also π -periodic and has real Fourier coefficients. In summary, we have

$$(3.23) \quad \mathcal{A}(\xi) = \mathcal{A}(\xi + \pi), \quad \text{and} \quad |\mathcal{A}(\xi)|^2 = |\mathcal{A}(-\xi)|^2, \quad \text{for all } \xi \in \mathbb{R}.$$

Since

$$\widehat{h}_2(\xi) = \mathcal{A}(\xi) + e^{-i\xi} \mathcal{A}(-\xi) \quad \text{and} \quad \widehat{h}_3(\xi) = e^{-i\xi} \overline{\widehat{h}_2(\cdot + \pi)} = e^{-i\xi} \mathcal{A}(-\xi) - \mathcal{A}(\xi),$$

applying (3.23), one obtains

$$\begin{aligned} |\widehat{h}_2(\xi)|^2 &= \left(\mathcal{A}(\xi) + e^{-i\xi} \mathcal{A}(-\xi) \right) \left(\overline{\mathcal{A}(\xi)} + e^{i\xi} \overline{\mathcal{A}(-\xi)} \right) \\ &= |\mathcal{A}(\xi)|^2 + |\mathcal{A}(-\xi)|^2 + e^{i\xi} \mathcal{A}(\xi) \overline{\mathcal{A}(-\xi)} + e^{-i\xi} \mathcal{A}(-\xi) \overline{\mathcal{A}(\xi)} \\ &= 2|\mathcal{A}(\xi)|^2 + e^{i\xi} \mathcal{A}(\xi) \overline{\mathcal{A}(-\xi)} + e^{-i\xi} \mathcal{A}(-\xi) \overline{\mathcal{A}(\xi)} \end{aligned}$$

and

$$\begin{aligned} |\widehat{h}_3(\xi)|^2 &= \left(e^{-i\xi} \mathcal{A}(-\xi) - \mathcal{A}(\xi) \right) \left(e^{i\xi} \overline{\mathcal{A}(-\xi)} - \overline{\mathcal{A}(\xi)} \right) \\ &= |\mathcal{A}(\xi)|^2 + |\mathcal{A}(-\xi)|^2 - e^{i\xi} \mathcal{A}(\xi) \overline{\mathcal{A}(-\xi)} - e^{-i\xi} \mathcal{A}(-\xi) \overline{\mathcal{A}(\xi)} \\ &= 2|\mathcal{A}(\xi)|^2 - e^{i\xi} \mathcal{A}(\xi) \overline{\mathcal{A}(-\xi)} - e^{-i\xi} \mathcal{A}(-\xi) \overline{\mathcal{A}(\xi)}. \end{aligned}$$

Hence,

$$|\widehat{h}_2(\xi)|^2 + |\widehat{h}_3(\xi)|^2 = 4|\mathcal{A}(\xi)|^2 = T(\xi),$$

which gives (3.22) and thus concludes that the masks $\{\widehat{h}_0, \widehat{h}_1, \widehat{h}_2, \widehat{h}_3\}$ satisfy (2.6). Therefore, $X(\Psi)$ is indeed a tight frame for $L_2(\mathbb{R})$ by the unitary extension principle.

Now we show that ψ_1 is symmetric about $\frac{1}{2}$ while ψ_2 is symmetric about $\frac{1}{4}$ and ψ_3 is antisymmetric about $\frac{1}{4}$. It is well known that a function $f \in L_2(\mathbb{R})$, is symmetric about the point $\gamma_1 \in \mathbb{R}$ if and only if

$$f(x) = f(2\gamma_1 - x) \quad \text{a.e.},$$

which is equivalent to

$$(3.24) \quad \widehat{f}(\xi) = e^{-i2\gamma_1\xi} \widehat{f}(-\xi) \quad \text{a.e..}$$

Similarly, a function $f \in L_2(\mathbb{R})$ is antisymmetric about the point $\gamma_2 \in \mathbb{R}$ if and only if

$$f(x) = -f(2\gamma_2 - x) \quad \text{a.e.},$$

which is equivalent to

$$(3.25) \quad \widehat{f}(\xi) = -e^{-i2\gamma_2\xi} \widehat{f}(-\xi) \quad \text{a.e..}$$

By the definition of \widehat{h}_1 and the fact that \widehat{h}_0 is symmetric about the origin and 2π -periodic, one obtains

$$\widehat{h}_1(\xi) = e^{-i\xi} \overline{\widehat{h}_0(\xi + \pi)} = e^{-2i\xi} (e^{i\xi} \overline{\widehat{h}_0(-\xi + \pi)}) = e^{-2i\xi} \widehat{h}_1(-\xi).$$

Since ϕ is symmetric about the origin, then by (3.24) one obtains

$$(3.26) \quad \widehat{\phi}(\xi) = \widehat{\phi}(-\xi), \quad \text{for all } \xi \in \mathbb{R}.$$

Therefore,

$$\widehat{\psi}_1(\xi) = \widehat{h}_1(\xi/2) \widehat{\phi}(\xi/2) = e^{-i\xi} \widehat{h}_1(-\xi/2) \widehat{\phi}(-\xi/2) = e^{-i\xi} \widehat{\psi}_1(-\xi),$$

which, by (3.24), means that ψ_1 is symmetric about $\frac{1}{2}$. Similarly by the definition of \widehat{h}_2 , one obtains

$$\widehat{h}_2(\xi) = \mathcal{A}(\xi) + e^{-i\xi} \mathcal{A}(-\xi) = e^{-i\xi} (\mathcal{A}(-\xi) + e^{i\xi} \mathcal{A}(\xi)) = e^{-i\xi} \widehat{h}_2(-\xi).$$

Applying (3.26) and the definition of $\widehat{\psi}_2$, one obtains,

$$\widehat{\psi}_2(\xi) = \widehat{h}_2(\xi/2) \widehat{\phi}(\xi/2) = e^{-i\frac{\xi}{2}} \widehat{h}_2(-\xi/2) \widehat{\phi}(-\xi/2) = e^{-i\frac{\xi}{2}} \widehat{\psi}_2(-\xi),$$

which, by (3.24), means that ψ_2 is symmetric about $\frac{1}{4}$. Similarly, we can show that ψ_3 is antisymmetric about $\frac{1}{4}$. ■

Remark 3.5. Theorem 3.6 proves that the first framelet ψ_1 constructed from dual pseudo-splines generates a Riesz basis for $L_2(\mathbb{R})$. Furthermore, when pseudo-splines of type I with order $(m, m-1)$ are used in Construction 3.3, the set Ψ contains only one element ψ which coincides with Daubechies' orthonormal wavelets of order m [61, 60] (see the next subsection for more details).

Next, we give one example of (anti)symmetric tight framelets constructed from Construction 3.3 using pseudo-splines of Type II with order $(3, 1)$.

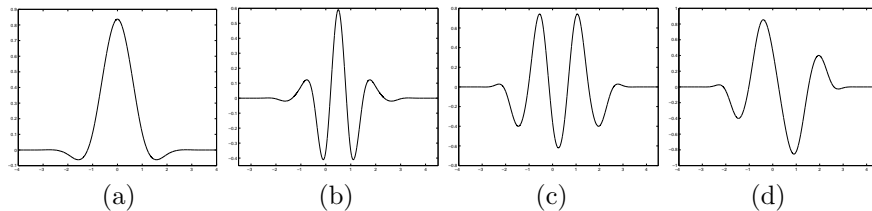


FIGURE 1. (a) is the pseudo-spline of Type II with order (3, 1) and (b)-(d) are the corresponding (anti)symmetric tight framelets.

Example 3.1. Let \widehat{h}_0 to be the mask of the pseudo-spline of Type II with order (3, 1) i.e.

$$\widehat{h}_0(\xi) = \cos^6(\xi/2)(1 + 3\sin^2(\xi/2)).$$

We define

$$\widehat{h}_1(\xi) := e^{-i\xi} \overline{\widehat{h}_0(\xi + \pi)} = e^{-i\xi} \sin^6(\xi/2)(1 + 3\cos^2(\xi/2)),$$

$$\widehat{h}_2(\xi) := \mathcal{A}(\xi) + e^{-i\xi}\mathcal{A}(-\xi) \quad \text{and} \quad \widehat{h}_3(\xi) := e^{-i\xi}\mathcal{A}(-\xi) - \mathcal{A}(\xi),$$

where

$$\begin{aligned} \mathcal{A} = \frac{1}{2} & \left(0.00123930398199e^{-4i\xi} + 0.00139868605052e^{-2i\xi} - 0.22813823298962 \right. \\ & \left. + 0.44712319189971e^{2i\xi} - 0.22162294894260e^{4i\xi} \right). \end{aligned}$$

The graphs of Ψ are given by (b)-(d) in Figure 1. The tight frame system has approximation order 4.

Since Construction 3.3 is generic and applies on any refinement mask, we can apply this construction on dual pseudo-splines. Notice that dual pseudo-splines are symmetric about $-1/2$. Then, we have the following simple results on the symmetry of the framelets constructed via Construction 3.3. Furthermore, arguments similar to those used in proving Theorem 3.6 can be used to prove that the first framelet ψ_1 constructed from dual pseudo-splines generates a Riesz basis for $L_2(\mathbb{R})$.

Proposition 3.7. *Let $\Psi = \{\psi_1, \psi_2, \psi_3\}$ be the tight framelets constructed from Construction 3.3 from dual pseudo-splines. Then, ψ_1 is antisymmetric about $\frac{1}{2}$, ψ_2 is symmetric about 0 and ψ_3 is antisymmetric about 0.*

2.2. Orthonormal wavelets

In this subsection, we show that the tight frame systems constructed from Construction 3.3 using pseudo-splines of type I with order $(m, m-1)$ are orthonormal bases for $L_2(\mathbb{R})$. In fact, they are the orthonormal bases originally constructed by Daubechies in [60]. For simplicity, we denote pseudo-splines of type I with order $(m, m-1)$ simply as ϕ_m , and its mask as a_m . In fact, one can see easily that the masks, a_m , for all m are exactly the same as the refinement masks of the orthonormal refinable functions in [60] by item 1 of Lemma 3.3.

By definition of pseudo-splines, we have

$$(3.27) \quad |\widehat{a}_m(\xi)|^2 + |\widehat{a}_m(\xi + \pi)|^2 = 1, \quad \text{and} \quad \widehat{a}_m(0) = 1.$$

Under the condition (3.27), Construction 3.3 produces only one framelet ψ with mask

$$\widehat{h}_1(\xi) = e^{-i\xi} \overline{\widehat{a}_m(\xi + \pi)}.$$

Now, we show that the tight frame system $X(\psi)$ is in fact an orthonormal basis for $L_2(\mathbb{R})$. Based on the unitary extension principle (Theorem 2.2), all we need to show is that $\|\phi_m\|_2 = 1$ for any given $m \geq 1$.

Theorem 3.4. *Given ϕ_m , pseudo-spline of type I with order $(m, m-1)$, let ψ be the single framelet constructed from Construction 3.3 using ϕ_m . Then $X(\psi)$ forms a compactly supported orthonormal wavelet basis for $L_2(\mathbb{R})$.*

PROOF. As we discussed above, all we need to show is that $\|\phi_m\|_2 = 1$ or equivalently, $\|\widehat{\phi}_m\|_2^2 = 2\pi$.

Consider the cascade algorithm defined by (same as (3.10))

$$\widehat{f}_n(\xi) = \widehat{a}_m\left(\frac{\xi}{2}\right) \widehat{f}_{n-1}\left(\frac{\xi}{2}\right) = \prod_{j=1}^n \widehat{a}_m(2^{-j}\xi) \widehat{f}_0(2^{-n}\xi), \quad n > 0,$$

with initial function f_0 satisfying $\widehat{f}_0(\xi) = \chi_{[-\pi, \pi]}(\xi)$. By Proposition 3.1, $\widehat{f}_n \rightarrow \widehat{\phi}_m$ pointwise as $n \rightarrow \infty$.

Following a proof similar to that of Theorem 3.1, we have

$$\begin{aligned} \|\widehat{f}_n\|_{L_2(\mathbb{R})}^2 &= \int_{-2^n\pi}^{2^n\pi} \prod_{j=1}^n |\widehat{a}_m(2^{-j}\xi)|^2 d\xi \\ &= \int_0^{2^n\pi} \prod_{j=1}^{n-1} |\widehat{a}_m(2^{-j}\xi)|^2 (|\widehat{a}_m(2^{-n}\xi)|^2 + |\widehat{a}_m(2^{-n}\xi + \pi)|^2) d\xi \\ (\text{by (3.27)}) &= \int_0^{2^n\pi} \prod_{j=1}^{n-1} |\widehat{a}_m(2^{-j}\xi)|^2 d\xi \\ &= \|\widehat{f}_{n-1}\|_{L_2(\mathbb{R})}^2 = \cdots = \|\widehat{f}_0\|_{L_2(\mathbb{R})}^2 = 2\pi. \end{aligned}$$

Note that

$$|\widehat{\phi}_m(\xi)|^2 \geq |B_{2m}(\xi)| = \left(\frac{\sin(\xi/2)}{\xi/2} \right)^{2m} \geq C > 0, \quad \xi \in [-\pi, \pi].$$

Then arguments similar to those in Proposition 3.2 imply that

$$\|\widehat{\phi}_m\|_{L_2(\mathbb{R})}^2 = \lim_{n \rightarrow \infty} \|\widehat{f}_n\|_{L_2(\mathbb{R})}^2 = 2\pi,$$

which concludes the proof of this theorem. \blacksquare

2.3. Riesz wavelets from pseudo-splines

In this section, we focus on the structure of the tight frame systems constructed from pseudo-splines by applying the unitary extension principle [158]. We show that in all pseudo-spline tight frame systems constructed using Construction 3.3, there is one wavelet whose dilations and shifts already form a Riesz basis for $L_2(\mathbb{R})$.

A system $X(\psi)$ is a *Riesz basis* if there exist $1 < C_1 \leq C_2 < \infty$ such that, for all sequences $c \in \ell_2(\mathbb{Z}^2)$,

$$C_1 \|c\|_{\ell_2(\mathbb{Z}^2)} \leq \left\| \sum_{(n,k) \in \mathbb{Z}^2} c[n,k] \psi_{n,k} \right\|_{L_2(\mathbb{R})} \leq C_2 \|c\|_{\ell_2(\mathbb{Z}^2)}$$

holds and the span of $\{\psi_{n,k} : n, k \in \mathbb{Z}\}$ is dense in $L_2(\mathbb{R})$. The function ψ is called *Riesz wavelet* if $X(\psi)$ forms a Riesz basis for $L_2(\mathbb{R})$ which is also called the *Riesz wavelet system*. When only the right hand side of the above inequalities holds, the wavelet system $X(\psi)$ is a Bessel system. Note that the definition of a Bessel system used here is equivalent to the earlier definition which uses the right hand side of the inequality in the definition of the frame in (2.2) (see e.g. [50, 156]).

For a given stable refinable function $\phi \in L_2(\mathbb{R})$, the key step in the MRA-based construction of the Riesz wavelet ψ is to select some desirable wavelet mask h_1 with

$$\widehat{\psi}(2\cdot) = \widehat{h}_1 \widehat{\phi}.$$

When $\{\phi(\cdot - k) : k \in \mathbb{Z}\}$ forms an orthonormal basis for $V_0(\phi)$, e.g. ϕ is a pseudo-spline of Type I with order $(m, m-1)$, and if we choose h_1 as

$$h_1[k] = (-1)^{k-1} \overline{h_0[1-k]}, \quad k \in \mathbb{Z},$$

or equivalently

$$(3.28) \quad \widehat{h}_1(\xi) = e^{-i\xi} \overline{\widehat{h}_0(\xi + \pi)}.$$

then Theorem 2.2 implies that the corresponding wavelet system $X(\psi)$ is an orthonormal basis for $L_2(\mathbb{R})$. It was shown in [103] that if ϕ is a B-spline, then the wavelet system $X(\psi)$ with the corresponding wavelet mask h_1 given by (3.28) is a Riesz basis for $L_2(\mathbb{R})$. The first goal of this section is to show that if ϕ is any pseudo-spline with mask h_0 , then the wavelet system $X(\psi)$ with the corresponding wavelet mask h_1 given by (3.28) is a Riesz basis for $L_2(\mathbb{R})$. To prove this, we use the following theorem which is a special case of [103, Theorem 2.1]. Note that when both refinement masks are finitely supported the result is similar to those already obtained in [55, 56, 53].

Theorem 3.5. *Let h_0 be a finitely supported refinement mask of a refinable function $\phi \in L_2(\mathbb{R})$ with $\widehat{h}_0(0) = 1$ and $\widehat{h}_0(\pi) = 0$, such that \widehat{h}_0 can be factorized into the form*

$$(3.29) \quad |\widehat{h}_0(\xi)| = \left| \left(\frac{1 + e^{-i\xi}}{2} \right)^n \mathcal{L}(\xi) \right| = \cos^n(\xi/2) |\mathcal{L}(\xi)|, \quad \xi \in [-\pi, \pi],$$

where \mathcal{L} is the Fourier series of a finitely supported sequence with $\mathcal{L}(\pi) \neq 0$. Suppose that

$$|\widehat{h}_0(\xi)|^2 + |\widehat{h}_0(\xi + \pi)|^2 \neq 0, \quad \xi \in [-\pi, \pi].$$

Define

$$\widehat{\psi}(2\xi) := e^{-i\xi} \overline{\widehat{h}_0(\xi + \pi)} \widehat{\phi}(\xi)$$

and

$$(3.30) \quad \tilde{\mathcal{L}}(\xi) := \frac{\mathcal{L}(\xi)}{|\widehat{h}_0(\xi)|^2 + |\widehat{h}_0(\xi + \pi)|^2}.$$

Assume that

$$(3.31) \quad \rho_{\mathcal{L}} := \|\mathcal{L}(\xi)\|_{L_\infty(\mathbb{R})} < 2^{n-\frac{1}{2}} \quad \text{and} \quad \rho_{\tilde{\mathcal{L}}} := \|\tilde{\mathcal{L}}(\xi)\|_{L_\infty(\mathbb{R})} < 2^{n-\frac{1}{2}},$$

Then $X(\psi)$ is a Riesz basis for $L_2(\mathbb{R})$.

Recall that the refinement masks of pseudo-splines of Type I and II are

$$|_1\hat{a}(\xi)| := \cos^m(\xi/2) \left(\sum_{j=0}^l \binom{m+l}{j} \sin^{2j}(\xi/2) \cos^{2(l-j)}(\xi/2) \right)^{\frac{1}{2}}$$

and

$${}_2\hat{a}(\xi) := \cos^{2m}(\xi/2) \sum_{j=0}^l \binom{m+l}{j} \sin^{2j}(\xi/2) \cos^{2(l-j)}(\xi/2).$$

Using the polynomial $P_{m,l}(y)$ given by (3.4) with $y = \sin^2(\xi/2)$, we can rewrite $_1\hat{a}$ and $_2\hat{a}$ as

$$(3.32) \quad |_1\hat{a}| = \left((1-y)^m P_{m,l}(y) \right)^{\frac{1}{2}}, \quad {}_2\hat{a} = (1-y)^m P_{m,l}(y).$$

Hence, the corresponding \mathcal{L} function in (3.29) for pseudo-splines of Type I and II are

$$(3.33) \quad |_1\mathcal{L}| = (P_{m,l}(y))^{\frac{1}{2}}, \quad |_2\mathcal{L}| = P_{m,l}(y).$$

Furthermore, using $R_{m,l}(y)$ defined by (3.5), we have

$$|_1\hat{a}(\xi)|^2 + |_1\hat{a}(\xi + \pi)|^2 = R_{m,l}(y) + R_{m,l}(1-y)$$

and

$$|_2\hat{a}(\xi)|^2 + |_2\hat{a}(\xi + \pi)|^2 = R_{m,l}^2(y) + R_{m,l}^2(1-y),$$

with $y = \sin^2(\xi/2)$. Hence,

$$(3.34) \quad |_1\tilde{\mathcal{L}}| = \frac{(P_{m,l}(y))^{\frac{1}{2}}}{R_{m,l}(y) + R_{m,l}(1-y)} \quad \text{and} \quad |_2\tilde{\mathcal{L}}| = \frac{P_{m,l}(y)}{R_{m,l}^2(y) + R_{m,l}^2(1-y)}.$$

Theorem 3.5 tells us that the key step is to estimate the upper bounds of $|\mathcal{L}(\xi)|$ and $|\tilde{\mathcal{L}}(\xi)|$. The estimation of $\|_1\tilde{\mathcal{L}}\|_{L_\infty(\mathbb{R})}$ and $\|_2\tilde{\mathcal{L}}\|_{L_\infty(\mathbb{R})}$ are based on the following result:

Proposition 3.8. *Let m and l be given nonnegative integers with $l \leq m-1$ and $|_1\tilde{\mathcal{L}}|$ and $|_2\tilde{\mathcal{L}}|$ be defined in (3.34). Then,*

- (1) $\|_1\tilde{\mathcal{L}}\|_{L_\infty(\mathbb{R})} = \sup_{y \in [0,1]} \frac{(P_{m,l}(y))^{\frac{1}{2}}}{R_{m,l}(y) + R_{m,l}(1-y)} < 2^{m-\frac{1}{2}}$.
- (2) $\|_2\tilde{\mathcal{L}}\|_{L_\infty(\mathbb{R})} = \sup_{y \in [0,1]} \frac{P_{m,l}(y)}{R_{m,l}^2(y) + R_{m,l}^2(1-y)} < 2^{2m-\frac{1}{2}}$.

PROOF. Item 1 of Lemma 3.3 gives

$$(3.35) \quad P_{m,l}(y) = \sum_{j=0}^l \binom{m+l}{j} y^j (1-y)^{l-j} = \sum_{j=0}^l \binom{m-1+j}{j} y^j,$$

with $y \in [0, 1]$. Hence both $(P_{m,l}(y))^{\frac{1}{2}}$ and $P_{m,l}(y)$ attain their maximum at $y = 1$ and the maximum values are:

$$(P_{m,l}(1))^{\frac{1}{2}} = \binom{m+l}{l}^{\frac{1}{2}} \quad \text{and} \quad P_{m,l}(1) = \binom{m+l}{l}.$$

By item 3 of Lemma 3.3, one obtains

$$\begin{aligned} \|_1\tilde{\mathcal{L}}\|_{L_\infty(\mathbb{R})} &= \sup_{y \in [0,1]} \frac{(P_{m,l}(y))^{\frac{1}{2}}}{R_{m,l}(y) + R_{m,l}(1-y)} \\ &\leq \binom{m+l}{l}^{\frac{1}{2}} \max_{y \in [0,1]} \frac{1}{R_{m,l}(y) + R_{m,l}(1-y)} \\ &\leq \frac{2^{m+l-1} \binom{m+l}{l}^{\frac{1}{2}}}{\sum_{j=0}^l \binom{m+l}{j}}. \end{aligned}$$

Applying item 3 of Lemma 3.2, i.e.

$$(3.36) \quad \frac{2^l \binom{m+l}{l}^{\frac{1}{2}}}{\sum_{j=0}^l \binom{m+l}{j}} \leq 1,$$

one obtains

$$\|_1\tilde{\mathcal{L}}\|_{L_\infty(\mathbb{R})} \leq 2^{m-1} < 2^{m-\frac{1}{2}}.$$

The proof of (2) is similar to that of (1). Indeed, by item 4 of Lemma 3.3 and applying (3.36) again, we have

$$\begin{aligned} \|_2\tilde{\mathcal{L}}\|_{L_\infty(\mathbb{R})} &= \sup_{y \in [0,1]} \frac{P_{m,l}(y)}{R_{m,l}^2(y) + R_{m,l}^2(1-y)} \\ &\leq \binom{m+l}{l} \max_{y \in [0,1]} \frac{1}{R_{m,l}^2(y) + R_{m,l}^2(1-y)} \\ &= \frac{2^{2m+2l-1} \binom{m+l}{l}}{(\sum_{j=0}^l \binom{m+l}{j})^2} \leq 2^{2m-1} < 2^{2m-\frac{1}{2}}. \end{aligned}$$

■

Theorem 3.6. *Let ${}_k\phi$, $k = 1, 2$ be the pseudo-spline of Type I and II with order (m, l) . The refinement masks ${}_ka$, $k = 1, 2$, are given in (3.2) and (3.3). Define*

$$(3.37) \quad {}_k\hat{\psi}(2\xi) := e^{-i\xi} \overline{{}_k\hat{a}(\xi + \pi)} {}_k\hat{\phi}(\xi), \quad k = 1, 2,$$

then $X({}_k\psi)$ forms a Riesz basis for $L_2(\mathbb{R})$.

PROOF. To apply Theorem 3.5, we first note that by items 3 and 4 of Lemma 3.3,

$$|_1\hat{a}(\xi)|^2 + |_1\hat{a}(\xi + \pi)|^2 = R_{m,l}(\sin^2(\xi/2)) + R_{m,l}(\cos^2(\xi/2)) \neq 0$$

and

$$|_2\hat{a}(\xi)|^2 + |_2\hat{a}(\xi + \pi)|^2 = R_{m,l}^2(\sin^2(\xi/2)) + R_{m,l}^2(\cos^2(\xi/2)) \neq 0,$$

for all $\xi \in [-\pi, \pi]$.

Since Proposition 3.8 showed that

$$\rho_{1\tilde{\mathcal{L}}} = \|_1\tilde{\mathcal{L}}\|_{L_\infty(\mathbb{R})} < 2^{m-\frac{1}{2}} \quad \text{and} \quad \rho_{2\tilde{\mathcal{L}}} = \|_2\tilde{\mathcal{L}}\|_{L_\infty(\mathbb{R})} < 2^{2m-\frac{1}{2}},$$

we only need to check whether

$$\rho_{1\mathcal{L}} = \|_1\mathcal{L}\|_{L_\infty(\mathbb{R})} < 2^{m-\frac{1}{2}}, \quad \rho_{2\mathcal{L}} = \|_2\mathcal{L}\|_{L_\infty(\mathbb{R})} < 2^{2m-\frac{1}{2}}.$$

Indeed, we have

$$|_k\hat{a}(\xi)|^2 + |_k\hat{a}(\xi + \pi)|^2 \leq 1 \quad \text{for all } \xi \in \mathbb{R}.$$

Hence,

$$|{}_k\mathcal{L}(\xi)| \leq |{}_k\tilde{\mathcal{L}}(\xi)| \quad \text{for all } \xi \in \mathbb{R}.$$

This concludes the proof. \blacksquare

Remark 3.6. The Riesz wavelet constructed in the above theorem has the same length of support and at least the same order of smoothness as that of the corresponding pseudo-spline. The order of its vanishing moments is the same as the order of the B-spline factor of the pseudo-spline. But, in general, its dual Riesz wavelet system is not compactly supported. However, this is not a problem for some applications. In applications like image compression, the short Riesz wavelet system can be applied to obtain a fast reconstruction algorithm, while decomposition is obtained by solving a linear system of equations (see [121]).

In Construction 3.3 as well as in the three constructions of [65] for pseudo-splines of Type I where the number of wavelets is either two or three (see [65] Section 3.1 for details), we observe that one of the wavelets ψ_1 is defined by

$$\widehat{\psi}_1 := e^{-i\xi} \overline{\widehat{h}_0(\xi + \pi)} \widehat{\phi}(\xi/2),$$

where \widehat{h}_0 is the refinement mask of a pseudo-spline. Then, from Theorem 3.6, $X(\psi_1)$ forms a Riesz basis for $L_2(\mathbb{R})$. This means that each pseudo-spline tight frame system constructed in [65] or by Construction 3.3, already has a subsystem form a Riesz basis for $L_2(\mathbb{R})$. Finally, we note that if dual pseudo-splines are used in Construction 3.3 or any of the constructions in [65], we will have the same conclusion by a similar argument as for pseudo-splines.

3. Regularity of pseudo-splines

This section is on theoretical analysis of pseudo-splines. Those who are interested in the basics of MRA-base wavelet frames and applications may skip this section. Also, we note that the material of this section is mainly taken from [75].

For $\alpha = n + \beta$, $n \in \mathbb{N}$, $0 \leq \beta < 1$, the Hölder space C^α (see e.g. [61]) is defined to be the set of functions which are n times continuously differentiable and whose n^{th} derivative $f^{(n)}$ satisfies the condition,

$$|f^{(n)}(x+h) - f^{(n)}(x)| \leq C|h|^\beta, \quad \forall x, h.$$

It is well known (see [61]) that if

$$\int_{\mathbb{R}} |\widehat{f}(\xi)| (1 + |\xi|)^\alpha < \infty,$$

then $f \in C^\alpha$. In particular, if $|\widehat{f}(\xi)| \leq C(1 + |\xi|)^{-1-\alpha-\varepsilon}$, then $f \in C^\alpha$.

The main idea here is to obtain a lower bound of the regularity of pseudo-splines with order (m, l) by estimating the decay of the Fourier transform of them. It turns out that this lower bound coincides with the upper bound when m goes to infinity, as will be shown in Theorem 3.9. It is well known that the exact Sobolev regularity of a given refinable function can be obtained via its mask by applying the transfer operator (see e.g. [61, 161] and references in there). Although the Sobolev exponent of a given refinable function can be computed exactly through finding the spectrum of the transfer operator derived from the corresponding refinement mask, this approach does not lead itself to a systematic computation of Sobolev exponents for a class of refinable functions, such as pseudo-splines. This is simply

because different refinable functions lead to different transfer operators. This is the main reason we give a systematic estimate of the decay of the Fourier transform of pseudo-splines instead.

Since for any compactly supported refinable function ϕ in $L_2(\mathbb{R})$ with $\widehat{\phi}(0) = 1$, the refinement mask h_0 must satisfy $\widehat{h}_0(0) = 1$ and $\widehat{h}_0(\pi) = 0$ (see e.g. [61] or [116]), then $\widehat{h}_0(\xi)$ can be factored as

$$\widehat{h}_0(\xi) = \left(\frac{1 + e^{-i\xi}}{2} \right)^n \mathcal{L}(\xi),$$

where n is the maximum multiplicity of zeros of \widehat{h}_0 at π and $\mathcal{L}(\xi)$ is a trigonometric polynomial with $\mathcal{L}(0) = 1$. Hence, we have

$$\widehat{\phi}(\xi) = \prod_{j=1}^{\infty} \widehat{a}(2^{-j}\xi) = \prod_{j=1}^{\infty} \left(\frac{1 + e^{-i2^{-j}\xi}}{2} \right)^n \prod_{j=1}^{\infty} \mathcal{L}(2^{-j}\xi) = \left(\frac{1 - e^{-i\xi}}{i\xi} \right)^n \prod_{j=1}^{\infty} \mathcal{L}(2^{-j}\xi).$$

This shows that any compactly supported refinable function in $L_2(\mathbb{R})$ is the convolution of a B-spline of some order, say n , with a distribution (see [155]). Indeed, a B-spline of order n can be rewritten as

$$\widehat{B}_n = \left(\frac{1 - e^{-i\xi}}{i\xi} \right)^n.$$

It is a piecewise polynomial of degree $n - 1$ in $C^{n-1-\varepsilon}(\mathbb{R})$, supported on $[0, n]$, and has refinement mask $\left(\frac{1+e^{-i\xi}}{2} \right)^n$. Since $\mathcal{L}(\xi)$ is bounded, $\mathcal{L}(\xi)$ is actually the refinement mask of a refinable distribution (see e.g. [61]). Therefore, ϕ is the convolution of the B-spline B_n with the distribution. The regularity of ϕ comes from the B-spline factor while the distribution factor takes away the regularity. But the distribution component provides some desirable properties for ϕ , such as interpolatory properties, orthogonality of its shifts and approximation order of certain quasi-interpolants.

The decay of $|\widehat{\phi}|$ can be characterized by $|\widehat{h}_0|$ as stated in the following theorem. The proof of this theorem can be found in [61]. Note that in the following theorem, we write $|\widehat{h}_0|$ in the form of

$$|\widehat{h}_0(\xi)| = \left| \left(\frac{1 + e^{-i\xi}}{2} \right)^n \mathcal{L}(\xi) \right| = \cos^n(\xi/2) |\mathcal{L}(\xi)|, \quad \xi \in [-\pi, \pi].$$

Theorem 3.7. [61] *Let h_0 be the refinement mask of the refinable function ϕ of the form*

$$|\widehat{h}_0(\xi)| = \cos^n(\xi/2) |\mathcal{L}(\xi)|, \quad \xi \in [-\pi, \pi].$$

Suppose that

$$(3.38) \quad \begin{aligned} |\mathcal{L}(\xi)| &\leq |\mathcal{L}(\frac{2\pi}{3})| && \text{for } |\xi| \leq \frac{2\pi}{3}, \\ |\mathcal{L}(\xi)\mathcal{L}(2\xi)| &\leq |\mathcal{L}(\frac{2\pi}{3})|^2 && \text{for } \frac{2\pi}{3} \leq |\xi| \leq \pi. \end{aligned}$$

Then $|\widehat{\phi}(\xi)| \leq C(1 + |\xi|)^{-n+\kappa}$, with $\kappa = \log(|\mathcal{L}(\frac{2\pi}{3})|)/\log 2$, and this decay is optimal.

This theorem allows us to estimate the decay of the Fourier transform of a refinable function via its refinement mask. Since $|\widehat{\phi}|^2 = |\widehat{\phi}|^2$, the decay rate of $|\widehat{\phi}|^2$ is half of that of $|\widehat{\phi}|$. Thus we can focus on the analysis of the decay of the Fourier

transforms of pseudo-splines of Type II. Based on (1) of Lemma 3.3, we show that $P_{m,l}(y)$, defined in (3.4), satisfies (3.38). This directly leads to the estimate of the regularity of pseudo-splines. Note that the corresponding result for $l = m - 1$ was proven in [54] which led to the optimal estimates for the decay of the Fourier transforms of the orthogonal and interpolatory refinable functions. Here, the more general result for pseudo-splines is obtained by a simpler proof than the original one of [54] and [61].

Proposition 3.9. [75] *Let $P_{m,l}(y)$ be defined as in (3.4), where l, m are nonnegative integers with $l \leq m - 1$. Then*

$$(3.39) \quad P_{m,l}(y) \leq P_{m,l}\left(\frac{3}{4}\right), \quad \text{for } y \in \left[0, \frac{3}{4}\right],$$

$$(3.40) \quad P_{m,l}(y)P_{m,l}(4y(1-y)) \leq \left(P_{m,l}\left(\frac{3}{4}\right)\right)^2, \quad \text{for } y \in \left[\frac{3}{4}, 1\right].$$

PROOF. Since $P_{m,l}(y)$ is monotonically increasing (using item 1 of Lemma 3.3), (3.39) is obviously true. Hence, we focus on the proof of (3.40).

Throughout this proof, we let m be fixed. Let

$$W_{m,l}(y) := P_{m,l}(y)P_{m,l}(4y(1-y)) - \left(P_{m,l}\left(\frac{3}{4}\right)\right)^2.$$

Then, the inequality (3.40) is equivalent to

$$(3.41) \quad W_{m,l}(y) \leq 0 \quad \text{for all } y \in \left[\frac{3}{4}, 1\right].$$

Note that when $l = 0$, $P_{m,0}(y) = 1$ for all $y \in [0, 1]$; then (3.41) is obviously true for $l = 0$. Therefore, if we can show that

$$(3.42) \quad W_{m,l+1}(y) - W_{m,l}(y) \leq 0, \quad \text{for all } y \in \left[\frac{3}{4}, 1\right], \quad l = 0, 1, \dots, m-2,$$

then (3.41) follows from (3.42) which is equivalent to (3.40). We now focus on the inequality (3.42).

Using item 1 of Lemma 3.3, we have

$$\begin{aligned} W_{m,l+1}(y) - W_{m,l}(y) &= \left(\sum_{j=0}^{l+1} \binom{m-1+j}{j} y^j \right) \left(\sum_{j=0}^{l+1} \binom{m-1+j}{j} (4y(1-y))^j \right) \\ &\quad - \left(\sum_{j=0}^l \binom{m-1+j}{j} y^j \right) \left(\sum_{j=0}^l \binom{m-1+j}{j} (4y(1-y))^j \right) \\ &\quad + P_{m,l}^2\left(\frac{3}{4}\right) - P_{m,l+1}^2\left(\frac{3}{4}\right). \end{aligned}$$

Splitting the sum $\sum_{j=0}^{l+1} \binom{m-1+j}{j} y^j$, one obtains

$$\begin{aligned}
W_{m,l+1}(y) - W_{m,l}(y) &= \left(\sum_{j=0}^l \binom{m-1+j}{j} y^j \right) \left(\sum_{j=0}^{l+1} \binom{m-1+j}{j} (4y(1-y))^j \right) \\
&\quad + \binom{m+l}{l+1} y^{l+1} \sum_{j=0}^{l+1} \binom{m-1+j}{j} (4y(1-y))^j \\
&\quad - \left(\sum_{j=0}^l \binom{m-1+j}{j} y^j \right) \left(\sum_{j=0}^l \binom{m-1+j}{j} (4y(1-y))^j \right) \\
&\quad + P_{m,l}^2\left(\frac{3}{4}\right) - P_{m,l+1}^2\left(\frac{3}{4}\right).
\end{aligned}$$

Combining the first and the third term, one obtains

$$\begin{aligned}
W_{m,l+1}(y) - W_{m,l}(y) &= \binom{m+l}{l+1} (4y(1-y))^{l+1} \sum_{j=0}^l \binom{m-1+j}{j} y^j \\
&\quad + \binom{m+l}{l+1} y^{l+1} \sum_{j=0}^{l+1} \binom{m-1+j}{j} (4y(1-y))^j \\
&\quad + P_{m,l}^2\left(\frac{3}{4}\right) - P_{m,l+1}^2\left(\frac{3}{4}\right) \\
&= \binom{m+l}{l+1} \left((4y(1-y))^{l+1} \sum_{j=0}^l \binom{m-1+j}{j} y^j \right. \\
(3.43) \quad &\quad \left. + y^{l+1} \sum_{j=0}^{l+1} \binom{m-1+j}{j} (4y(1-y))^j \right) \\
&\quad + P_{m,l}^2\left(\frac{3}{4}\right) - P_{m,l+1}^2\left(\frac{3}{4}\right).
\end{aligned}$$

Since $W_{m,l+1}\left(\frac{3}{4}\right) - W_{m,l}\left(\frac{3}{4}\right) = 0 - 0 = 0$, it suffices to show that $W_{m,l+1}(y) - W_{m,l}(y)$ monotonically decreases on $[\frac{3}{4}, 1]$, which is equivalent to showing that

$$G(y) := (4y(1-y))^{l+1} \sum_{j=0}^l \binom{m-1+j}{j} y^j + y^{l+1} \sum_{j=0}^{l+1} \binom{m-1+j}{j} (4y(1-y))^j$$

monotonically decreases on $[\frac{3}{4}, 1]$. We now compute G' as follows:

$$\begin{aligned}
G'(y) &= (l+1)(4-8y)(4y(1-y))^l \sum_{j=0}^l \binom{m-1+j}{j} y^j \\
&\quad + (4y(1-y))^{l+1} \sum_{j=1}^l \binom{m-1+j}{j} j y^{j-1} \\
&\quad + (l+1)y^l \sum_{j=0}^{l+1} \binom{m-1+j}{j} (4y(1-y))^j \\
&\quad + y^{l+1}(4-8y) \sum_{j=1}^{l+1} \binom{m-1+j}{j} j (4y(1-y))^{j-1} \\
&= (l+1)(4-8y)(4y(1-y))^l \sum_{j=0}^l \binom{m-1+j}{j} y^j \\
&\quad + (4y(1-y))^{l+1} \sum_{j=0}^{l-1} \binom{m+j}{j+1} (j+1) y^j \\
&\quad + (l+1)y^l \sum_{j=0}^{l+1} \binom{m-1+j}{j} (4y(1-y))^j \\
&\quad + y^{l+1}(4-8y) \sum_{j=0}^l \binom{m+j}{j+1} (j+1) (4y(1-y))^j.
\end{aligned}$$

Applying item 1 of Lemma 3.2 to the second and the fourth term above, one obtains

$$\begin{aligned}
G'(y) &= (l+1)(4-8y)(4y(1-y))^l \sum_{j=0}^l \binom{m-1+j}{j} y^j \\
&\quad + (4y(1-y))^{l+1} \sum_{j=0}^l \binom{m-1+j}{j} (m+j) y^j \\
&\quad - (m+l) \binom{m-1+l}{l} y^l (4y(1-y))^{l+1} \\
&\quad + (l+1)y^l \sum_{j=0}^l \binom{m-1+j}{j} (4y(1-y))^j \\
&\quad + (l+1) \binom{m+l}{l+1} y^l (4y(1-y))^{l+1} \\
&\quad + y^{l+1}(4-8y) \sum_{j=0}^l \binom{m-1+j}{j} (m+j) (4y(1-y))^j.
\end{aligned}$$

By $(l+1) \binom{m+l}{l+1} = (m+l) \binom{m-1+l}{l}$ (item 1 of Lemma 3.2), we have

$$(l+1) \binom{m+l}{l+1} y^l (4y(1-y))^{l+1} - (m+l) \binom{m-1+l}{l} y^l (4y(1-y))^{l+1} = 0.$$

Hence,

$$G'(y) = \sum_{j=0}^l \binom{m-1+j}{j} \left((l+1)(4-8y)(4y(1-y))^l y^j + (m+j)(4y(1-y))^{l+1} y^j \right. \\ \left. + (l+1)y^l (4y(1-y))^j + (m+j)(4-8y)y^{l+1}(4y(1-y))^j \right).$$

Noting the common factor $y^j(4y(1-y))^j$ in the right hand side of the above equation, one obtains

$$G'(y) = \sum_{j=0}^l \binom{m-1+j}{j} y^j (4y(1-y))^j \left((l+1)(4-8y)(4y(1-y))^{l-j} \right. \\ \left. + (m+j)(4y(1-y))^{l+1-j} + (l+1)y^{l-j} + (m+j)(4-8y)y^{l+1-j} \right).$$

For $0 \leq j \leq l \leq m-2$, consider

$$g_{l,j}(y) := (l+1)(4-8y)(4y(1-y))^{l-j} + (m+j)(4y(1-y))^{l+1-j} \\ + (l+1)y^{l-j} + (m+j)(4-8y)y^{l+1-j} \\ = (l+1)(4y(1-y))^{l-j} (4y(1-y) - (8y-4)) \\ + (l+1)y^{l-j} (1 - (8y-4)y) \\ + (m+j-l-1) \left((4y(1-y))^{l+1-j} - (8y-4)y^{l+1-j} \right).$$

The inequality $4y(1-y) \leq y$ and $8y-4 \geq 2$ for $y \in [3/4, 1]$ show that $g_{l,j}(y) \leq 0$ and $G'(y) \leq 0$ on this interval. ■

Remark 3.7. It is clear that $W_{m,0}(y) = 0$, $y \in [\frac{3}{4}, 1]$, because $P_{m,0} = 1$. It was also proven by [54] that $W_{m,m-1}(y) \leq 0$, $y \in [\frac{3}{4}, 1]$, which is equivalent to (3.40). The decreasing of $W_{m,l}(y)$, for $y \in [\frac{3}{4}, 1]$, as l increases shown above indicates some difficulties in directly proving (3.40) for an arbitrary l , $0 < l < m-1$, because it has a smaller margin than the case when $l = m-1$. In fact, to some extent, the proof of (3.40) for the case when $l = m-1$ relies on a numerical check for $m \leq 12$ (see [61]). Inequality (3.40) for the case $l = m-1$ as proven in [54] (also see [61]) is one of the cornerstones of the wavelet theory, because it immediately leads to the optimal estimate of the decay of the Fourier transforms (hence, an estimate of the regularity) of both interpolatory and orthogonal refinable functions. Here, we take a different approach by proving that $W_{m,l}(y)$, $y \in [\frac{3}{4}, 1]$, decreases as l increases. As a result, we obtain (3.40) for all $0 \leq l \leq m-1$ by the fact that $W_{m,0}(y) = 0$, $y \in [\frac{3}{4}, 1]$. This shows that introducing the concepts of the pseudo-splines gives a better understanding and a more complete picture of the proof of (3.40) and also, we hope, enriches the theory of wavelets. Note that the proof of (3.40) for all $0 \leq l \leq m-1$ given here does not rely on any numerical computation and is simpler than the original proof of [54] and [61]. More recently, the authors in [136] provided a simpler proof of Proposition 3.9.

With this proposition, one obtains the regularity of pseudo-splines by applying Theorem 3.7.

Theorem 3.8. [75] Let ${}_2\phi$ be the pseudo-spline of Type II with order (m, l) . Then

$$|{}_2\widehat{\phi}(\xi)| \leq C(1 + |\xi|)^{-2m+\kappa},$$

where $\kappa = \log(P_{m,l}(\frac{3}{4}))/\log 2$. Consequently, ${}_2\phi \in C^{\alpha_2-\varepsilon}$ with $\alpha_2 = 2m - \kappa - 1$. Furthermore, let ${}_1\phi$ be the pseudo-spline of Type I with order (m, l) . Then

$$|{}_1\widehat{\phi}(\xi)| \leq C(1 + |\xi|)^{-m+\frac{\kappa}{2}}.$$

Consequently, ${}_1\phi \in C^{\alpha_1-\varepsilon}$ with $\alpha_1 = m - \frac{\kappa}{2} - 1$.

PROOF. Since

$$\begin{aligned} {}_2\widehat{a}(\xi) &= \cos^{2m}(\xi/2) \sum_{j=0}^l \binom{m+l}{j} \sin^{2j}(\xi/2) \cos^{2(l-j)}(\xi/2) \\ &= (\cos(\xi/2))^{2m} P_{m,l}(\sin^2(\xi/2)), \end{aligned}$$

the function $|\mathcal{L}(\xi)|$ in Theorem 3.7 is exactly $P_{m,l}(\sin^2(\xi/2))$. Applying (3.39) of Proposition 3.9

$$P_{m,l}(y) \leq P_{m,l}(\frac{3}{4}), \quad y \in [0, \frac{3}{4}],$$

and using $y = \sin^2(\xi/2)$, we have

$$\begin{aligned} |\mathcal{L}(\xi)| &= P_{m,l}(\sin^2(\xi/2)) \\ &= P_{m,l}(y) \leq P_{m,l}(\frac{3}{4}) = P_{m,l}(\sin^2(\frac{\pi}{3})) \quad \text{for } |\xi| \leq \frac{2\pi}{3}. \end{aligned}$$

Note that

$$|\mathcal{L}(2\xi)| = P_{m,l}(\sin^2(\xi)) = P_{m,l}(4\sin^2(\xi/2)(1 - \sin^2(\xi/2))) = P_{m,l}(4y(1 - y)).$$

Applying (3.40) of Proposition 3.9

$$P_{m,l}(y)P_{m,l}(4y(1 - y)) \leq (P_{m,l}(\frac{3}{4}))^2, \quad y \in [\frac{3}{4}, 1],$$

we have

$$\begin{aligned} |\mathcal{L}(\xi)\mathcal{L}(2\xi)| &= P_{m,l}(\sin^2(\xi/2))P_{m,l}(4\sin^2(\xi/2)(1 - \sin^2(\xi/2))) \\ &= P_{m,l}(y)P_{m,l}(4y(1 - y)) \\ &\leq (P_{m,l}(\frac{3}{4}))^2 = (P_{m,l}(\sin^2(\frac{\pi}{3})))^2, \quad \text{for } \frac{2\pi}{3} \leq |\xi| \leq \pi. \end{aligned}$$

Hence, by Theorem 3.7, ${}_2\widehat{\phi}$ satisfies

$$|{}_2\widehat{\phi}(\xi)| \leq C(1 + |\xi|)^{-2m+\kappa},$$

where $\kappa = \log(P_{m,l}(\frac{3}{4}))/\log 2$. This leads to ${}_2\phi \in C^{\alpha_2-\varepsilon}$, where $\alpha_2 = 2m - \kappa - 1$.

Since the decay of $|{}_1\widehat{\phi}|$ is exactly half of $|{}_2\widehat{\phi}|$, we have

$$|{}_1\widehat{\phi}(\xi)| \leq C(1 + |\xi|)^{-m+\frac{\kappa}{2}},$$

consequently ${}_1\phi \in C^{\alpha_1-\varepsilon}$, where $\alpha_1 = m - \frac{\kappa}{2} - 1$. ■

Table 1 gives the decay rates $\beta_{m,l}$ of the Fourier transform of pseudo-splines of Type II with order (m, l) , for $2 \leq m \leq 8$ and $1 \leq l \leq m - 1$. The regularity exponent of the corresponding pseudo-spline is, at least, $\alpha_2 = \beta_{m,l} - 1 - \varepsilon$. The decay rate of the Fourier transform of the pseudo-spline of Type I with the same

order is $\frac{\beta_{m,l}}{2}$ and its regularity exponent α_1 is $\frac{\alpha_2-1}{2}$. Therefore, the table shows that for either type of the pseudo-splines and fixed order m , the decay rate of their Fourier transform decreases as l increases, while for fixed l , it increases as m increases. This is true indeed as shown in the following proposition.

TABLE 1. Decay rates $\beta_{m,l} = 2m - \kappa$ of pseudo-splines of Type II with order (m, l) , for $2 \leq m \leq 8$ and $1 \leq l \leq m - 1$.

(m, l)	$l = 1$	$l = 2$	$l = 3$	$l = 4$	$l = 5$	$l = 6$	$l = 7$
$m = 2$	2.67807						
$m = 3$	4.29956	3.27208					
$m = 4$	6.00000	4.73321	3.82507				
$m = 5$	7.75207	6.27890	5.19506	4.35316			
$m = 6$	9.54057	7.88626	6.64465	5.66363	4.86449		
$m = 7$	11.35614	9.54057	8.15608	7.04717	6.13261	5.36349	
$m = 8$	13.19265	11.23182	9.71691	8.48992	7.46770	6.59988	5.85310

Proposition 3.10. [75] *Let $\beta_{m,l} = 2m - \kappa$ with $\kappa = \log P_{m,l}(\frac{3}{4}) / \log 2$ as given in Theorem 3.8 and $0 \leq l \leq m - 1$. Then:*

- (1) *For fixed m , $\beta_{m,l}$ decreases as l increases.*
- (2) *For fixed l , $\beta_{m,l}$ increases as m increases.*
- (3) *When $l = m - 1$, $\beta_{m,l}$ increases as m increases.*

Consequently, the decay rate $\beta_{2,1} = 2.67807$ is the smallest among all $\beta_{m,l}$, with $m \geq 2$ and $0 \leq l \leq m - 1$.

PROOF. Part (1) follows directly from item 1 of Lemma 3.3, which shows that $P_{m,l}(\frac{3}{4})$ increases with l for fixed m .

For part (2), note that

$$\beta_{m,l} = 2m - \frac{\log P_{m,l}(\frac{3}{4})}{\log 2}.$$

Consider

$$2^{\beta_{m,l}} = 2^{2m - \frac{\log P_{m,l}(\frac{3}{4})}{\log 2}} = \frac{4^m}{P_{m,l}(\frac{3}{4})} = \frac{1}{4^{-m} P_{m,l}(\frac{3}{4})}.$$

Hence, part (2) is equivalent to the fact that

$$I_m := 4^{-m} P_{m,l}(\frac{3}{4})$$

decreases as m increases for fixed l , which is equivalent to that for fixed $0 \leq l \leq m - 1$,

$$(3.44) \quad I_{m+1} - I_m < 0.$$

Note that

$$\begin{aligned} I_{m+1} - I_m &= 4^{-m-1} P_{m+1,l}(\frac{3}{4}) - 4^{-m} P_{m,l}(\frac{3}{4}) \\ &= 4^{-m-1} \sum_{j=0}^l \left(\binom{m+j}{j} - 4 \binom{m-1+j}{j} \right) \left(\frac{3}{4} \right)^j. \end{aligned}$$

Inequality (3.44) follows from the fact that for $0 \leq j \leq m-1$,

$$(3.45) \quad \binom{m+j}{j} = \frac{m+j}{m} \binom{m-1+j}{j} = \left(1 + \frac{j}{m}\right) \binom{m-1+j}{j} < 4 \binom{m-1+j}{j}.$$

This concludes the proof of part (2).

Following an argument similar to that of part (2), one can show that part (3) is equivalent to having that

$$J_m := 4^{-m} P_{m,m-1} \left(\frac{3}{4}\right)$$

decreases as m increases, which is equivalent to

$$(3.46) \quad J_{m+1} - J_m < 0 \quad \text{for } m \geq 1.$$

Following a direct computation, we have

$$J_{m+1} - J_m = 4^{-m-1} \left(\sum_{j=0}^m \binom{m+j}{j} \left(\frac{3}{4}\right)^j - 4 \sum_{j=0}^{m-1} \binom{m-1+j}{j} \left(\frac{3}{4}\right)^j \right).$$

Let

$$M := \sum_{j=0}^m \binom{m+j}{j} \left(\frac{3}{4}\right)^j - 4 \sum_{j=0}^{m-1} \binom{m-1+j}{j} \left(\frac{3}{4}\right)^j.$$

Then, (3.46) is equivalent to $M < 0$ for $m \geq 1$. It is easy to see that $M < 0$, when $m = 1$. For the case $m \geq 2$, we first note that

$$\begin{aligned} M &= \sum_{j=0}^{m-1} \binom{m+j}{j} \left(\frac{3}{4}\right)^j - 4 \sum_{j=0}^{m-1} \binom{m-1+j}{j} \left(\frac{3}{4}\right)^j + \binom{2m}{m} \left(\frac{3}{4}\right)^m \\ &= \sum_{j=1}^{m-1} \binom{m-1+j}{j-1} \left(\frac{3}{4}\right)^j - 3 \sum_{j=0}^{m-1} \binom{m-1+j}{j} \left(\frac{3}{4}\right)^j + \binom{2m}{m} \left(\frac{3}{4}\right)^m, \end{aligned}$$

where the last identity follows from item 1 of Lemma 3.2. Substituting j for $j-1$ in the first term, one obtains that

$$(3.47) \quad M = \frac{3}{4} \sum_{j=0}^{m-2} \binom{m+j}{j} \left(\frac{3}{4}\right)^j - 3 \sum_{j=0}^{m-1} \binom{m-1+j}{j} \left(\frac{3}{4}\right)^j + \binom{2m}{m} \left(\frac{3}{4}\right)^m,$$

Splitting the second term in (3.47), one obtains

$$(3.48) \quad \begin{aligned} M &= \frac{3}{4} \sum_{j=0}^{m-2} \binom{m+j}{j} \left(\frac{3}{4}\right)^j - 3 \sum_{j=0}^{m-2} \binom{m-1+j}{j} \left(\frac{3}{4}\right)^j \\ &\quad + \binom{2m}{m} \left(\frac{3}{4}\right)^m - 3 \binom{2m-2}{m-1} \left(\frac{3}{4}\right)^{m-1} \end{aligned}$$

For the last two terms of (3.48), we have

$$\begin{aligned} \binom{2m}{m} \left(\frac{3}{4}\right)^m - 3 \binom{2m-2}{m-1} \left(\frac{3}{4}\right)^{m-1} &= \left(\frac{3}{4}\right)^m \left(\binom{2m}{m} - 4 \binom{2m-2}{m-1} \right) \\ &= \left(\frac{3}{4}\right)^m \left(\left(4 - \frac{2}{m}\right) \binom{2m-2}{m-1} - 4 \binom{2m-2}{m-1} \right) \\ &< 0. \end{aligned}$$

Therefore,

$$\begin{aligned}
M &< \frac{3}{4} \sum_{j=0}^{m-2} \binom{m+j}{j} \left(\frac{3}{4}\right)^j - 3 \sum_{j=0}^{m-2} \binom{m-1+j}{j} \left(\frac{3}{4}\right)^j \\
&< \sum_{j=0}^{m-2} \binom{m+j}{j} \left(\frac{3}{4}\right)^j - 3 \sum_{j=0}^{m-2} \binom{m-1+j}{j} \left(\frac{3}{4}\right)^j \\
&= \sum_{j=0}^{m-2} \left(\binom{m+j}{j} - 3 \binom{m-1+j}{j} \right) \left(\frac{3}{4}\right)^j.
\end{aligned}$$

Applying (3.45), one obtains, for $0 \leq j \leq m-2$,

$$\binom{m+j}{j} = \left(1 + \frac{j}{m}\right) \binom{m-1+j}{j} < 3 \binom{m-1+j}{j}.$$

Therefore, we conclude that $M < 0$ and part (3) follows.

Finally, note that the decay rate of the Fourier transform of the pseudo-spline of Type I with order $(2, 1)$ is $\frac{\beta_{m,l}}{2} \approx 1.33903$. Hence, it follows from parts (1)-(3) that the decay rate of an arbitrary pseudo-spline of either type with order (m, l) , $m > 2$, $0 \leq l \leq m-1$ is higher than 1.33903. ■

Proposition 3.10 reveals that the decay rates of the Fourier transforms of either type of pseudo-splines increase as m increases for fixed l and decrease as l increases for fixed m . Here, we give an asymptotical analysis of the decay rate which, in turn, gives an asymptotical analysis of the regularity of ${}_1\phi$ and ${}_2\phi$ as the order $(m, l) \rightarrow \infty$.

Theorem 3.9. [75] *Let ${}_1\phi$ and ${}_2\phi$ be the pseudo-splines of Type I and II respectively with order (m, l) . Fix $l = \lfloor \lambda m \rfloor$, $0 \leq \lambda \leq 1$, where $\lfloor \lambda m \rfloor$ denotes the largest integer which is smaller than or equal to λm . Then, we have*

$$|\widehat{{}_1\phi}(\xi)| \leq C(1 + |\xi|)^{-\frac{\mu}{2}m} \quad \text{and} \quad |\widehat{{}_2\phi}(\xi)| \leq C(1 + |\xi|)^{-\mu m},$$

where $\mu = \frac{\log\left(\frac{4}{1+\lambda}\right)^{\lambda+1} \left(\frac{\lambda}{3}\right)^{\lambda}}{\log 2}$, asymptotically for large m . This means that the asymptotic rates of the pseudo-spline of Type I and Type II are $\frac{\mu}{2}$ and μ respectively.

PROOF. We will only analyze pseudo-splines of type II, and the analysis for pseudo-splines of type I is similar. We first show the following identity:

$$(3.49) \quad x^{-l} P_{m,l}(x) \geq y^{-l} P_{m,l}(y), \quad \text{for } 0 < x \leq y \leq 1.$$

Indeed, item 1 of Lemma 3.3 implies that

$$x^{-l} P_{m,l}(x) = \sum_{j=0}^l \binom{m-1+j}{j} x^{j-l} \geq \sum_{j=0}^l \binom{m-1+j}{j} y^{j-l} = y^{-l} P_{m,l}(y),$$

for $0 < x \leq y \leq 1$.

In order to compute the asymptotic rate, we need to estimate the upper and lower bound of $P_{m,l}\left(\frac{3}{4}\right)$ in terms of m and l . For this, we let $x = \frac{3}{4}$ and $y = 1$ in (3.49) and obtain

$$(3.50) \quad P_{m,l}\left(\frac{3}{4}\right) \geq \left(\frac{3}{4}\right)^l P_{m,l}(1) = \left(\frac{3}{4}\right)^l \binom{m+l}{l}.$$

Next, let $x = \frac{1}{2}$ and $y = \frac{3}{4}$ in (3.49), we obtain

$$P_{m,l}\left(\frac{3}{4}\right) \leq \left(\frac{3}{2}\right)^l P_{m,l}\left(\frac{1}{2}\right).$$

Since

$$P_{m,l}\left(\frac{1}{2}\right) = \sum_{j=0}^l \binom{m+l}{j} 2^{-j} 2^{j-l} = 2^{-l} \sum_{j=0}^l \binom{m+l}{j},$$

one obtains

$$(3.51) \quad P_{m,l}\left(\frac{3}{4}\right) \leq \left(\frac{3}{4}\right)^l \sum_{j=0}^l \binom{m+l}{j}.$$

Putting (3.50) and (3.51) together, we obtain the following estimates of $P_{m,l}\left(\frac{3}{4}\right)$,

$$\left(\frac{3}{4}\right)^l \binom{m+l}{l} \leq P_{m,l}\left(\frac{3}{4}\right) \leq \left(\frac{3}{4}\right)^l \sum_{j=0}^l \binom{m+l}{j}.$$

For $l \leq m-1$, we have

$$\sum_{j=0}^l \binom{m+l}{j} \leq m \binom{m+l}{l}.$$

Hence,

$$(3.52) \quad \left(\frac{3}{4}\right)^l \binom{m+l}{l} \leq P_{m,l}\left(\frac{3}{4}\right) \leq m \left(\frac{3}{4}\right)^l \binom{m+l}{l}.$$

Next, we use this estimate to analyze the decay of $|\widehat{\phi}_2|$ as m goes to infinity. The upper bound of $P_{m,l}\left(\frac{3}{4}\right)$ in (3.52) implies that

$$2m - \frac{\log P_{m,l}\left(\frac{3}{4}\right)}{\log 2} \geq 2m - \frac{\log \left(m \left(\frac{3}{4}\right)^l \binom{m+l}{l}\right)}{\log 2}.$$

We estimate the right hand side of the above inequality asymptotically for large (m, l) to obtain the asymptotical lower bound of $2m - \frac{\log P_{m,l}\left(\frac{3}{4}\right)}{\log 2}$. For this, we first recall the Stirling approximation, i.e. $m! \sim \sqrt{2\pi} e^{(m+\frac{1}{2}) \log m - m}$ (see e.g. [88]), where $a_m \sim b_m$ means that $\frac{a_m}{b_m} \rightarrow 1$, $m \rightarrow \infty$. By the Stirling approximation, we have

$$(3.53) \quad \begin{aligned} \log m! &\sim \log \sqrt{2\pi} e^{(m+\frac{1}{2}) \log m - m} \\ &\sim m \log m - m. \end{aligned}$$

Applying (3.53), one obtains

$$\begin{aligned} \log \binom{m+l}{l} &= \log(m+l)! - \log m! - \log l! \\ &\sim (m+l) \log(m+l) - (m+l) - (m \log m - m) - (l \log l - l) \\ &\sim (m+l) \log(m+l) - m \log m - l \log l. \end{aligned}$$

Thus,

$$\begin{aligned} 2m - \frac{\log(m \frac{3^l}{4} \binom{m+l}{l})}{\log 2} &= 2m - \frac{\log m + l \log \frac{3}{4} + \log \binom{m+l}{l}}{\log 2} \\ &\sim m \left(2 - \frac{\frac{l}{m} \log \frac{3}{4} + (1 + \frac{l}{m}) \log(m+l) - \log m - \frac{l}{m} \log l}{\log 2} \right). \end{aligned}$$

By the assumption, $l = \lfloor \lambda m \rfloor$, $0 \leq \lambda \leq 1$. Hence, when m is sufficiently large, $\frac{l}{m} \sim \lambda$ and therefore,

$$\begin{aligned} 2m - \frac{\log(m \frac{3^l}{4} \binom{m+l}{l})}{\log 2} &\sim m \left(2 - \frac{\log(1+\lambda) \left(\frac{3+3\lambda}{4\lambda} \right)^\lambda}{\log 2} \right) \\ &= m \left(\frac{\log \left(\frac{4}{1+\lambda} \right)^{\lambda+1} \left(\frac{\lambda}{3} \right)^\lambda}{\log 2} \right). \end{aligned}$$

Now we obtain the asymptotical lower bound of $2m - \frac{\log P_{m,l}(\frac{3}{4})}{\log 2}$, i.e. asymptotically, for large m with $l = \lfloor \lambda m \rfloor$,

$$(3.54) \quad 2m - \frac{\log |P_{m,l}(\frac{3}{4})|}{\log 2} \geq m \left(\frac{\log \left(\frac{4}{1+\lambda} \right)^{\lambda+1} \left(\frac{\lambda}{3} \right)^\lambda}{\log 2} \right).$$

Next, we use the left hand side of (3.52) to obtain the asymptotical upper bound of $2m - \frac{\log P_{m,l}(\frac{3}{4})}{\log 2}$. First note that (3.52) gives

$$2m - \frac{\log P_{m,l}(\frac{3}{4})}{\log 2} \leq 2m - \frac{l \log \frac{3}{4} + \log \binom{m+l}{l}}{\log 2}.$$

Applying arguments similar to the estimate of the lower bound by using (3.53), we obtain the following

$$\begin{aligned} 2m - \frac{l \log \frac{3}{4} + \log \binom{m+l}{l}}{\log 2} &\sim m \left(2 - \frac{\frac{l}{m} \log \frac{3}{4} + (1 + \frac{l}{m}) \log(m+l) - \log m - \frac{l}{m} \log l}{\log 2} \right) \\ &\sim m \left(\frac{\log \left(\frac{4}{1+\lambda} \right)^{\lambda+1} \left(\frac{\lambda}{3} \right)^\lambda}{\log 2} \right). \end{aligned}$$

This leads to the asymptotical lower bound of $2m - \frac{\log P_{m,l}(\frac{3}{4})}{\log 2}$, i.e. asymptotically, for large m with $l = \lfloor \lambda m \rfloor$,

$$(3.55) \quad 2m - \frac{\log P_{m,l}(\frac{3}{4})}{\log 2} \leq m \left(\frac{\log \left(\frac{4}{1+\lambda} \right)^{\lambda+1} \left(\frac{\lambda}{3} \right)^\lambda}{\log 2} \right).$$

Combining (3.54) and (3.55), we conclude that for large m , the asymptotical upper and lower bounds coincide and equal to

$$(3.56) \quad 2m - \frac{\log P_{m,l}(\frac{3}{4})}{\log 2} \sim m \left(\frac{\log \left(\frac{4}{1+\lambda} \right)^{\lambda+1} \left(\frac{\lambda}{3} \right)^\lambda}{\log 2} \right) = \mu m.$$

Therefore the equation (3.56) gives that, fixing $l = \lfloor \lambda m \rfloor$ and asymptotically, for large m , we have

$$|{}_2\widehat{\phi}(\xi)| \leq C(1 + |\xi|)^{-\mu m} \quad \text{and} \quad |{}_1\widehat{\phi}(\xi)| \leq C(1 + |\xi|)^{-\frac{\mu}{2}m},$$

where $\mu = \frac{\log(\frac{4}{1+\lambda})^{\lambda+1}(\frac{4}{3})^\lambda}{\log 2}$. ■

Remark 3.8. The above theorem shows that, asymptotically for large m , the smoothness of the pseudo-splines of type I and II increases at a rate $\mu/2$ and μ respectively. The proof of Theorem 3.9 also leads to the following two observations:

- (1) Consider pseudo-splines of Type II with order $(m, m - p)$, where p is a fixed positive integer independent of m . The asymptotic rate is $2 - \frac{\log 3}{\log 2} \approx 0.4150$. Indeed, when $l = m - p$, $\lambda \sim \frac{l}{m} = \frac{m-p}{m} \sim 1$ for sufficiently large m . Similarly, for pseudo-splines of Type I with order $(m, m - p)$, the corresponding asymptotic rate is $1 - \frac{\log 3}{2 \log 2} \approx 0.2075$.
- (2) Assume that l is fixed for all m . The asymptotic rates of pseudo-splines of Type I and II with order (m, l) are 1 and 2 respectively. This is simply because, for the fixed integer l , $\lambda \sim \frac{l}{m} \sim 0$ for sufficiently large m .

TABLE 2. Asymptotically for large m , the smoothness of ${}_2\phi$ increases at rate μ , which is given in the following table with some choices of l .

$m \rightarrow \infty$	$l = 0$	$l = \frac{m}{10}$	$l = \frac{m}{8}$	$l = \frac{m}{6}$	$l = \frac{m}{4}$	$l = \frac{m}{2}$	$l = m - 1$
$\mu \approx$	2.0000	1.5581	1.4857	1.3789	1.2013	0.8301	0.4150

Example 3.2. In Table 2, we give μ , the asymptotical rate of pseudo-splines of Type II with order $(m, \lfloor \lambda m \rfloor)$, as m goes to infinity and the parameter $\lambda = \frac{1}{10}, \frac{1}{8}, \frac{1}{6}, \frac{1}{4}, \frac{1}{2}, 1$. The asymptotic rate μ_0 for pseudo-splines of Type I with the same order is just $\mu_0 = \frac{\mu}{2}$.

A similar discussion can lead the regularity analysis for the dual pseudo-splines as it was done in [72].

Theorem 3.10. [72] *Let $\tilde{\phi}_{m,l}$ be the dual pseudo-spline of order (m, l) with $0 \leq l \leq m - 1$. Then*

$$|\widehat{\tilde{\phi}}_{m,l}| \leq C(1 + |\xi|)^{-\gamma_{m,l}}$$

with

$$\gamma_{m,l} = 2m + 1 - \log(P_{m-\frac{1}{2},l}(\frac{3}{4})) / \log 2,$$

and the decay rate $\gamma_{m,l}$ is optimal. Consequently, $\tilde{\phi}_{m,l} \in C^\alpha$ with $\alpha = \gamma_{m,l} - 1 - \varepsilon$.

4. Two lemmata

This section gives the two key technical lemmata that were used in the previous sections.

Lemma 3.2. *For given nonnegative integers m, j, l , we have:*

- (1) $\binom{m+1}{j} = \binom{m}{j} + \binom{m}{j-1}$ for $j \geq 1$ and $(j+1)\binom{m+j}{j+1} = (m+j)\binom{m-1+j}{j}$.
- (2) $2(m+1)\sum_{j=0}^{l-1}\binom{m+l}{j} - l\sum_{j=0}^l\binom{m+l}{j} \geq 0$, for $m \geq 1$ and $1 \leq l \leq m-1$.
- (3) $\frac{2^l\binom{m+l}{l}^{\frac{1}{2}}}{\sum_{j=0}^l\binom{m+l}{j}} \leq 1$, for all $m \geq 1$ and $0 \leq l \leq m-1$.

PROOF. The identities in (1) are well known and can be proven directly by the definition of the binomial coefficients.

For (2), since $m > l$, we have

$$(m+1)\sum_{j=0}^{l-1}\binom{m+l}{j} - l\sum_{j=0}^{l-1}\binom{m+l}{j} \geq 0.$$

Subtracting this inequality from (2), we conclude that it remains to check if

$$(m+1)\sum_{j=0}^{l-1}\binom{m+l}{j} - l\binom{m+l}{l} \geq 0$$

holds, in order to verify (2). Since $(m+1)\binom{m+l}{l-1} = l\binom{m+l}{l}$, we have

$$(m+1)\sum_{j=0}^{l-1}\binom{m+l}{j} > (m+1)\binom{m+l}{l-1} = l\binom{m+l}{l}.$$

This gives (2) immediately.

Finally, we prove (3) by induction with respect to m . Since (3) is obviously true for $l = 0$, we now focus on $1 \leq l \leq m-1$. When $m = 1$, the inequality trivially holds. Assume (3) holds when $m = m_0$, i.e.

$$2^{2l}\binom{m_0+l}{l} \leq \left(\sum_{j=0}^l\binom{m_0+l}{j}\right)^2,$$

for all $1 \leq l \leq m_0 - 1$. Consider the case $m = m_0 + 1$. We first show that (3) holds for all l , where $1 \leq l \leq m_0 - 1$. For $1 \leq l \leq m_0 - 1$, we have

$$\begin{aligned}
2^{2l} \binom{m_0 + l + 1}{l} &= \frac{m_0 + l + 1}{m_0 + 1} 2^{2l} \binom{m_0 + l}{l} \\
&\leq \frac{m_0 + l + 1}{m_0 + 1} \left(\sum_{j=0}^l \binom{m_0 + l}{j} \right)^2 \quad (\text{by induction hypothesis}) \\
&= \left(\sum_{j=0}^l \binom{m_0 + l}{j} + \left(\sqrt{\frac{m_0 + l + 1}{m_0 + 1}} - 1 \right) \sum_{j=0}^l \binom{m_0 + l}{j} \right)^2 \\
&= \left(\sum_{j=0}^l \binom{m_0 + l}{j} \right. \\
&\quad \left. + \frac{l}{m_0 + 1 + \sqrt{(m_0 + l + 1)(m_0 + 1)}} \sum_{j=0}^l \binom{m_0 + l}{j} \right)^2 \\
&< \left(\sum_{j=0}^l \binom{m_0 + l}{j} + \frac{l}{2m_0 + 2} \sum_{j=0}^l \binom{m_0 + l}{j} \right)^2 \\
&\leq \left(\sum_{j=0}^l \binom{m_0 + l}{j} + \sum_{j=0}^{l-1} \binom{m_0 + l}{j} \right)^2 \quad (\text{from (2)}) \\
&= \left(1 + \sum_{j=1}^l \binom{m_0 + l}{j} + \sum_{j=1}^l \binom{m_0 + l}{j-1} \right)^2 \\
&= \left(\sum_{j=0}^l \binom{m_0 + l + 1}{j} \right)^2 \quad (\text{from (1)}).
\end{aligned}$$

This shows that (3) holds for all $1 \leq l \leq m_0 - 1$, it remains to show (3) holds for $l = m_0$, i.e. to show

$$(3.57) \quad 2^{2m_0} \binom{2m_0 + 1}{m_0} \leq \left(\sum_{j=0}^{m_0} \binom{2m_0 + 1}{j} \right)^2.$$

Applying $\binom{n}{j} = \binom{n}{n-j}$, we have

$$\begin{aligned}
\sum_{j=0}^{m_0} \binom{2m_0 + 1}{j} &= \frac{1}{2} \left(\sum_{j=0}^{m_0} \binom{2m_0 + 1}{j} + \sum_{j=0}^{m_0} \binom{2m_0 + 1}{j} \right) \\
&= \frac{1}{2} \left(\sum_{j=0}^{m_0} \binom{2m_0 + 1}{j} + \sum_{j=m_0+1}^{2m_0+1} \binom{2m_0 + 1}{j} \right) \\
&= \frac{1}{2} \sum_{j=0}^{2m_0+1} \binom{2m_0 + 1}{j} = 2^{2m_0}.
\end{aligned}$$

Then (3.57) is equivalent to $\binom{2m_0+1}{m_0} \leq \sum_{j=0}^{m_0} \binom{2m_0+1}{j}$, which is obviously true. This concludes the proof of (3). \blacksquare

Next, we give several basic properties of the polynomials $P_{m,l}(y)$ and $R_{m,l}(y)$ defined by (3.4) and (3.5). Part (2)-(4) of the following lemma were mainly used in the proof of Theorem 3.2.

Lemma 3.3. *For nonnegative integers m and l with $l \leq m - 1$, let $P_{m,l}(y)$ and $R_{m,l}(y)$ be the polynomials defined in (3.4) and (3.5). Then:*

- (1) $P_{m,l}(y) = \sum_{j=0}^l \binom{m-1+j}{j} y^j$.
- (2) $R'_{m,l}(y) = -(m+l) \binom{m+l-1}{l} y^l (1-y)^{m-1}$.
- (3) Define $Q(y) := R_{m,l}(y) + R_{m,l}(1-y)$. Then,

$$\min_{y \in [0,1]} Q(y) = Q\left(\frac{1}{2}\right) = 2^{1-m-l} \sum_{j=0}^l \binom{m+l}{j}.$$

- (4) Define $S(y) := R_{m,l}^2(y) + R_{m,l}^2(1-y)$. Then,

$$\min_{y \in [0,1]} S(y) = S\left(\frac{1}{2}\right) = 2^{1-2m-2l} \left(\sum_{j=0}^l \binom{m+l}{j} \right)^2.$$

PROOF. For fixed m , we prove (1) by induction with respect to l . It is obviously true for $l = 0$. Now suppose (1) holds for l_0 . Consider $l = l_0 + 1$,

$$\begin{aligned} P_{m,l}(y) &= \sum_{j=0}^{l_0+1} \binom{m+l_0+1}{j} y^j (1-y)^{l_0-j+1} \\ &= (1-y)^{l_0+1} + \sum_{j=1}^{l_0+1} \binom{m+l_0+1}{j} y^j (1-y)^{l_0-j+1}. \end{aligned}$$

Applying the first identity in (1) of Lemma 3.2, we have,

$$\begin{aligned} P_{m,l}(y) &= (1-y)^{l_0+1} + \sum_{j=1}^{l_0+1} \binom{m+l_0}{j} y^j (1-y)^{l_0-j+1} \\ &\quad + \sum_{j=1}^{l_0+1} \binom{m+l_0}{j-1} y^j (1-y)^{l_0-j+1} \\ &= \sum_{j=0}^{l_0+1} \binom{m+l_0}{j} y^j (1-y)^{l_0-j+1} + \sum_{j=1}^{l_0+1} \binom{m+l_0}{j-1} y^j (1-y)^{l_0-j+1} \\ &= \sum_{j=0}^{l_0} \binom{m+l_0}{j} y^j (1-y)^{l_0-j+1} + \binom{m+l_0}{l_0+1} y^{l_0+1} \\ &\quad + \sum_{j=0}^{l_0} \binom{m+l_0}{j} y^{j+1} (1-y)^{l_0-j} \\ &= (1-y)P_{m,l_0}(y) + \binom{m+l_0}{l_0+1} y^{l_0+1} + yP_{m,l_0}(y) \\ &= P_{m,l_0}(y) + \binom{m+l_0}{l_0+1} y^{l_0+1}. \end{aligned}$$

Then, by the inductive hypothesis,

$$\begin{aligned} P_{m,l}(y) &= \sum_{j=0}^{l_0} \binom{m-1+j}{j} y^j + \binom{m+l_0}{l_0+1} y^{l_0+1} \\ &= \sum_{j=0}^{l_0+1} \binom{m-1+j}{j} y^j. \end{aligned}$$

We prove (2) by induction with respect to l for given m . It is obviously true when $l = 0$. Suppose (2) holds for l_0 , i.e. $R'_{m,l_0}(y) = -(m+l_0) \binom{m+l_0-1}{l_0} y^{l_0} (1-y)^{m-1}$, and consider the case $l = l_0 + 1 \leq m-1$. Using (1) and definition of $R_{m,l}(y)$ in (3.5), we have

$$\begin{aligned} R_{m,l_0+1}(y) &= (1-y)^m P_{m,l_0+1}(y) \\ &= (1-y)^m \left(P_{m,l_0}(y) + \binom{m+l_0}{l_0+1} y^{l_0+1} \right). \end{aligned}$$

Since $R_{m,l_0}(y) = (1-y)^m P_{m,l_0}(y)$, we have

$$R_{m,l_0+1}(y) = \binom{m+l_0}{l_0+1} y^{l_0+1} (1-y)^m + R_{m,l_0}(y).$$

Then,

$$\begin{aligned} R'_{m,l_0+1}(y) &= (l_0+1) \binom{m+l_0}{l_0+1} y^{l_0} (1-y)^m - m \binom{m+l_0}{l_0+1} y^{l_0+1} (1-y)^{m-1} \\ &\quad + R'_{m,l_0}(y) \\ &= (l_0+1) \binom{m+l_0}{l_0+1} y^{l_0} (1-y)^m - m \binom{m+l_0}{l_0+1} y^{l_0+1} (1-y)^{m-1} \\ &\quad - (m+l_0) \binom{m+l_0-1}{l_0} y^{l_0} (1-y)^{m-1}. \end{aligned}$$

Pulling the common factor $y^{l_0} (1-y)^{m-1}$ out, one obtains

$$\begin{aligned} R'_{m,l_0+1}(y) &= y^{l_0} (1-y)^{m-1} \left((l_0+1) \binom{m+l_0}{l_0+1} (1-y) - m \binom{m+l_0}{l_0+1} y \right. \\ &\quad \left. - (m+l_0) \binom{m+l_0-1}{l_0} \right) \\ &= y^{l_0} (1-y)^{m-1} \left((l_0+1) \binom{m+l_0}{l_0+1} - (l_0+1) \binom{m+l_0}{l_0+1} y \right. \\ &\quad \left. - m \binom{m+l_0}{l_0+1} y - (m+l_0) \binom{m+l_0-1}{l_0} \right) \end{aligned}$$

Combining the second and the third term, one obtains

$$\begin{aligned} R'_{m,l_0+1}(y) &= y^{l_0} (1-y)^{m-1} \left((l_0+1) \binom{m+l_0}{l_0+1} - (m+l_0+1) \binom{m+l_0}{l_0+1} y \right. \\ &\quad \left. - (m+l_0) \binom{m+l_0-1}{l_0} \right). \end{aligned}$$

By the second identity in (1) of Lemma 3.2, one obtains $(l_0 + 1) \binom{m+l_0}{l_0+1} = (m + l_0) \binom{m+l_0-1}{l_0}$. Hence, $R'_{m,l_0+1}(y) = -(m + l_0 + 1) \binom{m+l_0}{l_0+1} y^{l_0+1} (1-y)^{m-1}$. This concludes the proof of (2).

For (3), we compute $Q'(y)$, i.e.

$$Q'(y) = R'_{m,l}(y) + (R_{m,l}(1-y))' = R'_{m,l}(y) - R'_{m,l}(1-y).$$

Applying (2), one obtains

$$Q'(y) = (m+l) \binom{m+l-1}{l} \left(y^{m-1}(1-y)^l - (1-y)^{m-1}y^l \right).$$

Now, we show that $Q'(y) \leq 0$ on $[0, \frac{1}{2}]$, $Q'(y) \geq 0$ on $[\frac{1}{2}, 1]$. Note that

$$y^{m-l-1} \leq (1-y)^{m-l-1}, \quad \text{for all } y \in [0, \frac{1}{2}].$$

Multiplying both sides by $y^l(1-y)^l$,

$$y^{m-1}(1-y)^l \leq (1-y)^{m-1}y^l, \quad \text{for all } y \in [0, \frac{1}{2}].$$

Similarly we have

$$y^{m-1}(1-y)^l \geq (1-y)^{m-1}y^l, \quad \text{for all } y \in [\frac{1}{2}, 1].$$

We conclude that

$$Q'(y) \begin{cases} \leq 0, & y \in [0, \frac{1}{2}] \\ \geq 0, & y \in [\frac{1}{2}, 1]. \end{cases}$$

This means that $Q(y)$ reaches its minimum value at point $y = \frac{1}{2}$. Now, we compute $Q(\frac{1}{2})$. Note that $Q(\frac{1}{2}) = 2R_{m,l}(\frac{1}{2}) = 2^{1-m}P_{m,l}(\frac{1}{2})$. Recall that $P_{m,l}(y)$ is defined in (3.4), i.e. $P_{m,l}(y) = \sum_{j=0}^l \binom{m+l}{j} y^j (1-y)^{l-j}$. Then $Q(\frac{1}{2}) = 2^{1-m} 2^{-l} \sum_{j=0}^l \binom{m+l}{j} = 2^{1-m-l} \sum_{j=0}^l \binom{m+l}{j}$.

With (3), the proof of (4) is simpler. Since

$$S'(y) = 2R_{m,l}(y)R'_{m,l}(y) + 2R_{m,l}(1-y)(R_{m,l}(1-y))',$$

using the identities $R_{m,l}(y) = (1-y)^m P_{m,l}(y)$, $R'_{m,l}(y) = -(m+l) \binom{m+l-1}{l} y^l (1-y)^{m-1}$ and $(R_{m,l}(1-y))' = (m+l) \binom{m+l-1}{l} y^{m-1} (1-y)^l$, we obtain

$$\begin{aligned} \frac{S'(y)}{2(m+l) \binom{m+l-1}{l}} &= y^m P_{m,l}(1-y) y^{m-1} (1-y)^l - (1-y)^m P_{m,l}(y) y^l (1-y)^{m-1} \\ &= y^{2m-1} \sum_{j=0}^l \binom{m-1+j}{j} (1-y)^{l+j} \\ &\quad - (1-y)^{2m-1} \sum_{j=0}^l \binom{m-1+j}{j} y^{l+j} \\ &= \sum_{j=0}^l \binom{m-1+j}{j} \left((1-y)^{l+j} y^{2m-1} - y^{l+j} (1-y)^{2m-1} \right). \end{aligned}$$

Since, for each $0 \leq j \leq l$, when $y \in [0, \frac{1}{2}]$, $y^{2m-l-j-1} \leq (1-y)^{2m-l-j-1}$, and when $y \in [\frac{1}{2}, 1]$, $y^{2m-l-j-1} \geq (1-y)^{2m-l-j-1}$, then by similar arguments in (2) we conclude,

$$S'(y) \begin{cases} \leq 0, & y \in [0, \frac{1}{2}] \\ \geq 0, & y \in [\frac{1}{2}, 1]. \end{cases}$$

Thus $\min_{y \in [0,1]} S(y) = S(\frac{1}{2})$. Since $R_{m,l}(\frac{1}{2}) = 2^{-m-l} \sum_{j=0}^l \binom{m+l}{j}$, we have

$$S(\frac{1}{2}) = 2R_{m,l}^2(\frac{1}{2}) = 2^{1-2m-2l} \left(\sum_{j=0}^l \binom{m+l}{j} \right)^2.$$

■

Remark 3.9. From (1) of Lemma 3.3 we know that the refinement mask of the pseudo-spline of Type I in (3.2) can be written as

$$|_1\hat{a}(\xi)|^2 = \cos^{2m}(\xi/2) \sum_{j=0}^l \binom{m-1+j}{j} \sin^{2j}(\xi/2).$$

Hence, the pseudo-spline of Type I with order $(m, m-1)$ is indeed the refinable function whose shifts form an orthonormal system constructed in [60] and the pseudo-spline of Type II with order $(m, m-1)$ is indeed the autocorrelation of the orthogonal refinable function, which is interpolatory.

Frame based image restorations

Image restoration is often formulated as an inverse problem. For simplicity of notation, we denote images as vectors in \mathbb{R}^n with n equal to the total number of pixels. The objective is to find the unknown true image $u \in \mathbb{R}^n$ from an observed image (or measurements) $f \in \mathbb{R}^l$ defined by

$$(4.1) \quad f = Au + \eta,$$

where η is a white Gaussian noise with variance σ^2 , and $A \in \mathbb{R}^{l \times n}$ is a linear operator, typically a convolution operator for image deconvolution problems, a projection operator for image inpainting and the identity for image denoising. This lecture is devoted to frame based image restorations.

As we discussed in Lecture 2, fast framelet decomposition and reconstruction algorithms are linear operators. When we consider signals living 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 . Then based on our analysis in Lecture 2, we have $W^\top W = I$, where I is the identity matrix. Thus, for every vector $u \in \mathbb{R}^n$,

$$(4.2) \quad u = W^\top(Wu).$$

The components of the vector Wu are called the canonical coefficients representing u . The matrix W is generated from the masks $\{h_\ell : \ell = 0, 1, \dots, r\}$ constructed from the univariate UEP (Theorem 2.6). We emphasize that we view the image as a column vector and the corresponding tight wavelet frame transform as the matrix W action just for simplicity of notation. In computations of image restorations, we use 2D wavelet decomposition and reconstruction algorithms given by Algorithm 2.3 instead of matrix multiplications.

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 images; namely, the analysis based approach, the synthesis based approach and the balanced approach. The analysis based approach was first proposed in [84, 170]. In this approach, we assume that the coefficient vector Wu can be sparsely approximated, and is usually formulated as a minimization problem involving a penalty on the term $\|Wu\|_1$. The synthesis based approach was first introduced in [66, 86, 87, 90, 91]. In that approach, the underlying image u is assumed to be synthesized from a sparse coefficient vector α with $u = W^\top \alpha$, and it is usually formulated as a minimization problem involving a penalty on the term $\|\alpha\|_1$. The balanced approach was first used in [34, 36] for high resolution image reconstruction. It was further developed for various image

restoration problems in [14, 15, 16, 17, 25, 32, 38]. In this approach, the underlying image u is assumed to be synthesized from some sparse coefficient vector α via $u = W^\top \alpha$ which is penalized by $\|\alpha\|_1$, and the distance of α to the range of W is minimized at the same time. Although the synthesis based, analysis based and balanced approaches are developed independently in the literature, the balanced approach can be viewed as a way to balance the analysis and synthesis based approach.

First, Section 1 gives a general introduction. Then Section 2 illustrates the ideas of the development of the balanced approach, as well as details of a fast algorithm, under the scenario of image inpainting [16, 168]. Then an accelerated algorithm of [168] for both balanced and synthesis based approach for general image restorations (4.1) is discussed. Within this section we discuss the linearized Bregman algorithm and its connection to the Uzawa's algorithm as an alternative to either the balanced or the synthesis based approaches. In Section 3, we present the analysis based approach [24], and its applications in image deblurring.

1. Modeling

This section establishes the models for frame based image restorations, especially, image inpainting, denoising, and deblurring.

1.1. Preliminaries

We start with some basic notations. For any $x \in \mathbb{R}^n$,

$$\|x\|_p := \left(\sum_{j=1}^n |x_j|^p \right)^{1/p},$$

$1 \leq p < \infty$. We denote the inner product $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{R}$ as

$$\langle x, y \rangle = x^\top y = \sum_{j=1}^n x_j y_j.$$

Then $\|x\|_2^2 = \langle x, x \rangle$. Let $\|x\|_D$ denote the D -norm, where D is a symmetric positive definite matrix, defined by $\|x\|_D = \sqrt{x^\top D x}$. For any real symmetric matrix H , $\rho_{\max}(H)$ denotes the maximum eigenvalue of H in magnitude. For any $m \times n$ real matrices A , $\|A\|_2 = \sqrt{\rho_{\max}(A^\top A)}$.

Now we recall some concepts and facts from general convex analysis, which are used throughout this lecture. A function $F : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ is convex if the domain of F is convex and for all u in the domain of F ,

$$F(tu + (1-t)v) \leq tF(u) + (1-t)F(v), \quad \text{for all } t \in [0, 1],$$

and it is said to be strictly convex if

$$F(tu + (1-t)v) < tF(u) + (1-t)F(v), \quad \text{for all } t \in (0, 1) \text{ and } u \neq v.$$

Here $\bar{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$.

A function F is said to be subdifferentiable at u if there exists $u^* \in \mathbb{R}^n$ such that

$$(4.3) \quad F(v) \geq F(u) + \langle v - u, u^* \rangle, \quad \text{for all } v \in \mathbb{R}^n,$$

and u^* is called a subgradient of F at u . The set of subgradients of F at u is denoted as $\partial F(u)$, which, in general, is not a singleton. By definition of subgradients, it is

easy to see that for $u^* \in \partial F(u)$ and $v^* \in \partial F(v)$ with $F(u)$ and $F(v)$ being finite, we have

$$(4.4) \quad \langle u^* - v^*, u - v \rangle \geq 0.$$

The same as derivatives, we have the following chain rule for subgradients, where the equality is the set equality,

$$\partial(F(Tu)) = T^\top(\partial F)(Tu),$$

where T is some linear operator on \mathbb{R}^n , whenever both sides of equality are well defined. If F is convex and differentiable at u , then we have $\partial F(u) = \{\nabla F(u)\}$, i.e. $\partial F(u)$ contains $\nabla F(u)$ as its unique element. If F is 2nd order differentiable, then F is convex if and only if the Hessian of F , $\nabla^2 F$ is semi-positive definite; and F is strictly convex if and only if $\nabla^2 F$ is positive definite.

It is well known (see e.g. [82]) that if a convex function F is bounded in a neighborhood of some point $v \in \{u \in \mathbb{R}^n : F(u) < \infty\}$, then it is continuous and subdifferentiable everywhere in the interior of $\{u \in \mathbb{R}^n : F(u) < \infty\}$. In particular if the convex function F nowhere assumes the values $\pm\infty$, then it is continuous and subdifferentiable everywhere in \mathbb{R}^n (this statement is not true if F is defined on a Banach space, instead of a finite dimensional space).

The general optimality condition for a convex subdifferentiable functional $F(u)$ can be characterized by subgradients which is known as the Fermat's rule [154, Theorem 10.1]. The statement is as follows:

$$(4.5) \quad u^* = \arg \min_u F(u) \quad \text{if and only if } 0 \in \partial F(u^*).$$

In other words, u^* minimizes $F(u)$ if and only if 0 is a subgradient of F at u^* .

A convex function $F : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ is said to be proper, if it nowhere takes the value $-\infty$ and is not identically equal to $+\infty$. We call F lower semi-continuous at v if

$$\liminf_{u \rightarrow v} F(u) \geq F(v).$$

Note that if a convex function F nowhere assumes $\pm\infty$, then it is proper and lower semi-continuous (in fact it is continuous based on the above discussions). The reason to introduce the concept of proper and lower semi-continuity is to include functions F that actually assumes value ∞ . For example, given a set $\Omega \in \mathbb{R}^n$, we can define the corresponding indicator function \mathcal{C}_Ω as

$$\mathcal{C}_\Omega(u) = \begin{cases} 0, & u \in \Omega \\ \infty, & u \in \Omega^c \end{cases}.$$

The function \mathcal{C}_Ω is proper, but it is lower semi-continuous if and only if Ω is closed. If the set Ω is convex, then \mathcal{C}_Ω is a convex function. Note that we can use indicator functions to move convex constraints for an optimization problem into the corresponding objective function. For example, minimizing $F(u)$ subject to $u \in \Omega$ is equivalent to minimizing $F(u) + \mathcal{C}_\Omega(u)$, which is still a convex optimization problem whenever Ω is convex.

The function F is said to be coercive, if

$$F(u) \rightarrow \infty \quad \text{when } \|u\|_2 \rightarrow \infty.$$

Then the general existence theorem of convex optimization problems states as follows (see e.g. [82]): the problem

$$\min_{u \in \mathbb{R}^n} F(u)$$

has at least one solution if F is proper, convex, lower semi-continuous and coercive; if, in addition, F is strictly convex, then the solution is unique.

A function F is said to have an L -Lipschitz continuous gradient if

$$(4.6) \quad \|\nabla F(u) - \nabla F(v)\| \leq L\|u - v\|, \quad \forall u, v$$

for some $L > 0$. We call L the Lipschitz constant of F .

Finally, we recall the following result on convergence of minimizing sequences, which is a special case of one of the results of [130]. The proof of this result is taken from [24].

Proposition 4.1. *Let $F(u)$ be a convex function defined on \mathbb{R}^n and nowhere assumes the values $\pm\infty$. Suppose F has a unique minimizer $u^* \in \mathbb{R}^n$. Then any minimizing sequence $\{u_k\}_{k=0}^\infty$, i.e. $F(u_k) \rightarrow F(u^*)$ as $k \rightarrow \infty$, converges to u^* in any Euclidean norm of \mathbb{R}^n .*

PROOF. Since all norms of \mathbb{R}^n are equivalent, we shall focus on the ℓ_2 -norm and prove that

$$(4.7) \quad \lim_{k \rightarrow \infty} \|u_k - u^*\|_2 = 0.$$

Since $u^* \in \mathbb{R}^n$ is the unique minimizer of $F(u)$, we have $F(u) > F(u^*)$ for all $u \neq u^*$. Now, we suppose that (4.7) does not hold, so there exists a subsequence u_{k_i} such that $\|u_{k_i} - u^*\|_2 > \epsilon$ for some $\epsilon > 0$ and for all i . Then, $F(u_{k_i}) > \min\{F(u) : \|u - u^*\|_2 = \epsilon\}$. Indeed, let v be the intersection of the sphere $\{u : \|u - u^*\|_2 = \epsilon\}$ and the line segment from u^* to u_{k_i} ; then there exists a positive number $t \in (0, 1)$ such that $v = tu^* + (1-t)u_{k_i}$. By the convexity of F and the definition of u^* , we have

$$\begin{aligned} F(u_{k_i}) &> tF(u^*) + (1-t)F(u_{k_i}) \geq F(tu^* + (1-t)u_{k_i}) = F(v) \\ &\geq \min\{F(u) : \|u - u^*\|_2 = \epsilon\}. \end{aligned}$$

Denote $\tilde{u} = \arg \min\{F(u) : \|u - u^*\|_2 = \epsilon\}$, which exists because spheres in \mathbb{R}^n are compact and F is continuous. By noting that $\tilde{u} \neq u^*$, we have

$$F(u^*) = \lim_{i \rightarrow +\infty} F(u_{k_i}) \geq F(\tilde{u}) > F(u^*),$$

which is a contradiction. \blacksquare

Remark 4.10. The conclusion of Proposition 4.1 is no longer true if F is a functional defined on a general normed space. The reason is that $\inf_u \{F(u) : \|u - u^*\| = \epsilon\}$ may not be attainable for any $u \in \{\|u - u^*\| = \epsilon\}$, and $\inf_u \{F(u) : \|u - u^*\| = \epsilon\}$ could be equal to $\inf_u F(u)$ (see [130] for a counterexample, as well as general theories on the convergence of minimizing sequences).

1.2. Image restoration models

The balanced approach, synthesis based approach and analysis based approach can be integrated into one minimization problem as follows:

$$(4.8) \quad \min_{\alpha \in \mathbb{R}^m} \frac{1}{2} \|AW^\top \alpha - f\|_D^2 + \frac{\kappa}{2} \|(I - WW^\top)\alpha\|_2^2 + \|\text{diag}(\lambda)\alpha\|_1,$$

where $0 \leq \kappa \leq \infty$ and λ is a given vector defined as

$$(4.9) \quad \lambda = (\lambda_1, \lambda_2, \dots, \lambda_m)^\top,$$

with $\lambda_i > 0$ for $i = 1, \dots, m$. We denote $\lambda_{\min} := \min_i \lambda_i > 0$. The matrix D is some symmetric positive definite matrix. It is usually chosen as a preconditioner to the equation $AW^\top \alpha = f$. Throughout this lecture, we take $D = (AA^\top + \theta I)^{-1}$ for some properly chosen $\theta > 0$. We refer to the model (4.8) as the (*single-system*) *balanced approach*.

When $0 < \kappa < \infty$, the problem (4.8) is called the balanced approach. When $\kappa = 0$, the problem (4.8) is reduced to the *synthesis based approach*:

$$(4.10) \quad \min_{\alpha \in \mathbb{R}^m} \frac{1}{2} \|AW^\top \alpha - f\|_D^2 + \|\text{diag}(\lambda)\alpha\|_1.$$

When $\kappa = \infty$, the problem (4.8) is reduced to the analysis based approach. To see this, we note that the distance $\|(I - WW^\top)\alpha\|$ must be 0 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 (4.8) as

$$\min_{\alpha \in \text{Range}(W)} \frac{1}{2} \|AW^\top \alpha - f\|_D^2 + \|\text{diag}(\lambda)\alpha\|_1 = \min_{u \in \mathbb{R}^n} \frac{1}{2} \|Au - f\|_D^2 + \|\text{diag}(\lambda)Wu\|_1.$$

Now, we refer to the problem

$$(4.11) \quad \min_{u \in \mathbb{R}^n} \frac{1}{2} \|Au - f\|_D^2 + \|\text{diag}(\lambda)Wu\|_1,$$

as the *analysis based approach*. We remark that the analysis based approach does not require using the inverse of W . Hence, it can be generalized to any linear transform W .

The objective functions in (4.8) with $0 \leq \kappa \leq \infty$ and in (4.11) are convex (but not strictly convex in general), proper and continuous on \mathbb{R}^m . Furthermore, as Theorem 4.2 shows that they are also coercive. Therefore, there exists at least one solution for both problems.

It is clear that when $0 < \kappa < \infty$, (4.8) balances between (4.10) and (4.11); hence, is called the balanced approach. Since the algorithms and convergence analysis on the balanced approach are still valid when $\kappa = 0$, we include discussion of the synthesis based approach into our presentation of the balanced approach. We even intend to call the model (4.10) the general balanced approach for $0 \leq \kappa < \infty$ and include the synthesis based approach as a special case.

We note that when the rows of W form an orthonormal basis, instead of being a redundant tight frame, the above three approaches are exactly the same, since in this case, $WW^\top = I$. However, for redundant tight frame system W , the analysis based, synthesis based and balanced approach cannot be derived from one another. In fact, it was observed in [47, 83] for example, that there is a gap between the analysis based and synthesis based approach. Both of them have their own favorable data sets and applications. In general, it is hard to draw definitive conclusions on which approach is better without specifying the applications and data sets. We further note that the ℓ_1 -minimization problems in compressed sensing is akin to the synthesis based approach in nature. On the other hand, the TV-norm minimization problem in imaging restoration is, for many cases, an analysis based approach. For frame based image restoration, numerical simulations in [24] show that the analysis based approach tends to generate smoother images than the synthesis

based approach. This is because the coefficient Wu is quite often linked to the smoothness of the underlying image [10, 9, 97, 106, 110]. However, the synthesis based approach tends to exploit more of the sparse representation of the underlying solution in terms of the given frame system by utilizing the redundancy. Therefore, the synthesis based approach usually enhances and sharpens edges, although it may introduce some artifacts as shown in [16]. The balanced approach bridges the analysis based and synthesis based approach and it balances the smoothness and the sparsity provided by frames as shown in [14, 15, 16, 25, 32, 34, 36, 38].

To solve the synthesis based approach (4.10), the proximal forward and backward splitting algorithm was used in [62, 66, 86, 87, 90, 91]. Also, the accelerated proximal gradient (APG) algorithms of [168], designed for the general balanced approach, can be applied to get a fast algorithm for the synthesis based approach.

For the analysis based approach (4.11), the coordinate descent method was used in [84, 170]. More recently, the split Bregman algorithm was used to develop a fast algorithm for the analysis based approach in frame based image restoration in [24], where numerical simulations showed that the split Bregman algorithm is efficient for image deblurring, decomposition, denoise, and inpainting. The split Bregman algorithm was first proposed in [96] and was shown to be powerful in [96, 185] when it was applied to various PDE based image restoration approaches, e.g., ROF and nonlocal variational models. Convergence analysis of the split Bregman was given in [24].

For the balanced approach in frame based image restoration, the model and algorithm were first developed in [32, 34, 36, 38]. The balanced approach was reformulated as the proximal forward-backward splitting algorithm in [14, 15, 16, 25]. The balanced approach gives satisfactory simulation results, as shown in [14, 15, 16, 25, 32, 34, 36, 38]. Recently, fast algorithms for the balanced approach in frame based image restoration whose convergence speeds are much faster than those of the proximal forward-backward splitting algorithm were developed in [168]. The accelerated proximal gradient algorithms proposed in [168] are based on and extended from several variants of accelerated proximal gradient algorithms that were studied in [4, 140, 141, 142, 143, 175]. These accelerated proximal gradient algorithms have an attractive iteration complexity of $O(1/\sqrt{\epsilon})$ for achieving ϵ -optimality. Also these accelerated proximal gradient algorithms are simple and only use the soft-thresholding operator, just like algorithms such as the linearized Bregman algorithm, the split Bregman algorithm and the proximal forward-backward splitting algorithm.

Finally, we note that the formulation of (4.8) can be extended to image restoration of two-layered images [24, 84, 170]. Real images usually have two layers, referring to cartoons (the piecewise smooth part of the image) and textures (the oscillating pattern part of the image). Usually, different layers have sparse approximations under different tight frame systems. Therefore, these two different layers should be considered separately. One natural idea is to use two tight frame systems that can sparsely represent cartoons and textures respectively. The corresponding image restoration problem can be formulated as the following ℓ_1 -minimization problem:

$$(4.12) \quad \min_{\alpha^1, \alpha^2 \in \mathbb{R}^m} \frac{1}{2} \|A(\sum_{i=1}^2 W_i^\top \alpha^i) - f\|_D^2 + \sum_{i=1}^2 \frac{\kappa_i}{2} \|(I - W_i W_i^\top) \alpha^i\|_2^2 + \sum_{i=1}^2 \|\text{diag}(\lambda^i) \alpha^i\|_1,$$

where, for $i = 1, 2$, $W_i^T W_i = I$, $\kappa_i > 0$, λ_i is a given positive weight vector, and D is a given symmetric positive definite matrix. We refer to model (4.12) as the *two-system balanced approach*.

2. Balanced approach

The balanced approach for frame based image restorations was first developed in [34, 36, 38] for high resolution image reconstruction from a few low resolution images. The problem of high resolution image constructions is converted to the problem of filling in the missing wavelet frame coefficients, i.e. inpainting in wavelet frame transformation domain, by designing a proper wavelet tight frame [34, 36, 38]. The ideas of [34, 36] are used in [37] to develop the balanced approach for frame based image inpainting (in pixel domain) whose complete analysis of convergence and optimal properties of the solution are given in [16]. Analysis of the convergence and optimal properties of the solutions of algorithms in [34, 36, 38] is given in [14, 25, 32].

In this section, we use image inpainting to illustrate how the ideas of the balanced approach are formed and developed. Also, we prove the convergence of the image inpainting algorithm. Then we present an accelerated algorithm of [168] for the balanced approach (which includes the synthesis based approach) and the proof of convergence given by [168] is included. After that, we recall the linearized Bregman algorithm as an alternative to the balanced and synthesis approaches. We prove the convergence of the linearized Bregman algorithm by connecting it to the Uzawa's algorithm. Numerical experiments on both image inpainting and deblurring are presented at the end of this section.

2.1. Balanced approach for image inpainting

The mathematical model for image inpainting can be stated as follows. Let the original image $u \in \mathbb{R}^n$ be defined on the domain $\Omega = \{1, 2, \dots, n\}$ and the nonempty set $\Lambda \subsetneq \Omega$ be the given observed region. Then the observed (incomplete) image f is

$$(4.13) \quad f(i) = \begin{cases} u(i) + \eta(i), & i \in \Lambda, \\ \text{arbitrary}, & i \in \Omega \setminus \Lambda, \end{cases}$$

where $\eta(i)$ is the noise. The goal is to find u from f . When $\eta(i) = 0$ for all $i \in \Lambda$, we require that $u(i) = f(i)$ and u is just the solution of an interpolation problem. Otherwise, we seek a smooth solution u that satisfies $|u(i) - f(i)| \leq \eta(i)$ for all $i \in \Lambda$. In both cases, variational approaches will penalize some cost functionals (which normally are weighted function norms of the underlying solution) to control the roughness of the solution, see for instance [7, 42].

Image inpainting is to recover data by interpolation. There are many interpolation schemes available, e.g., spline interpolation, but the majority of them are only good for smooth functions. Images are either piecewise smooth function or formed by textures which do not have the global smoothness required by most interpolation schemes. The major challenge in image inpainting is to keep the features, e.g. edges, of images which cannot be well preserved by many of the available interpolation algorithms. Furthermore, since images are usually contaminated by noise, a good inpainting algorithm should have a built-in denoising component.

The simple idea of the balanced approach for frame based image inpainting comes as follows: one may use any simple interpolation scheme to interpolate the given data that leads to an inpainted image. Edges might be blurred and noises are still present in this inpainted image. One of the simplest ways to sharpen the image and remove noises at the same time is to set small coefficients under a tight wavelet frame transform to zero. When we reconstruct the image using the modified frame coefficients, it will no longer interpolate the data, and the simplest way to make it interpolate the given data is to put the given data back. One may iterate this procedure till convergence.

To be precise, let \mathcal{P}_Λ be the diagonal matrix with diagonal entries 1 for the indices in Λ and 0 otherwise. Starting with the initial guess u_0 , the iterative procedure mentioned above can be expressed as

$$(4.14) \quad u_{k+1} = \mathcal{P}_\Lambda f + (I - \mathcal{P}_\Lambda)W^\top \mathcal{T}_\lambda(Wu_k).$$

Here

$$(4.15) \quad \mathcal{T}_\lambda([\beta_1, \beta_2, \dots, \beta_m]^\top) := [t_{\lambda_1}(\beta_1), t_{\lambda_2}(\beta_2), \dots, t_{\lambda_m}(\beta_m)]^\top$$

is the soft-thresholding function [77]:

$$(4.16) \quad t_{\lambda_i}(\beta_i) := \begin{cases} \text{sgn}(\beta_i)(|\beta_i| - \lambda_i), & \text{if } |\beta_i| > \lambda_i, \\ 0, & \text{if } |\beta_i| \leq \lambda_i, \end{cases}$$

Note that by using soft-thresholding instead of hard-thresholding which is traditionally used to sharpen edges, we reduce artifacts and obtain the desired minimization property in each iteration. The thresholding operator \mathcal{T}_λ also plays two other important roles, namely, removing noises in the image and perturbing the frame coefficients Wu_k so that information contained in the given region Λ can permeate into the missing region $\Omega \setminus \Lambda$.

Now, define $\alpha_k = \mathcal{T}_\lambda(Wu_k)$. Then (4.14) can be rewritten as an iteration in terms of α_k :

$$\alpha_{k+1} = \mathcal{T}_\lambda\left(W\mathcal{P}_\Lambda f + W(I - \mathcal{P}_\Lambda)W^\top \alpha_k\right).$$

Then we have the following frame based image inpainting algorithm:

Algorithm 4.1.

- (i) Set an initial guess α_0 .
- (ii) Iterate on k until convergence:

$$(4.17) \quad \alpha_{k+1} = \mathcal{T}_\lambda\left(W\mathcal{P}_\Lambda f + W(I - \mathcal{P}_\Lambda)W^\top \alpha_k\right)$$

- (iii) Let α^* to the output of Step (ii). If $\eta(i) = 0$ for all $i \in \Lambda$ in (4.13), we set $u^* := \mathcal{P}_\Lambda f + (I - \mathcal{P}_\Lambda)W^\top \alpha^*$ to be the solution (to the inpainting problem); otherwise, since \mathcal{T}_λ can remove noise, we set $u^\diamond = W^\top \alpha^*$ to be the solution (to the inpainting-plus-denoising problem).

Algorithm 4.1 was first proposed in [37], and a complete analysis of its convergence was given in [16]. It was proved in [16] that α_k generated by Algorithm 4.1 is a converging sequence that minimizes

$$(4.18) \quad \min_{\alpha \in \mathbb{R}^m} \frac{1}{2} \|\mathcal{P}_\Lambda W^\top \alpha - \mathcal{P}_\Lambda f\|_2^2 + \frac{1}{2} \|(I - WW^\top)\alpha\|_2^2 + \|\text{diag}(\lambda)\alpha\|_1.$$

Note that the problem (4.18) is a special case of (4.8) with $D = I$, $A = \mathcal{P}_\Lambda$ and $\kappa = 1$. In fact, (4.18) was discovered first and was later generalized to (4.8)

and (4.12). Here, we consider a slightly more general problem than (4.18), with $\kappa \in [0, \infty)$,

$$(4.19) \quad \min_{\alpha \in \mathbb{R}^m} \frac{1}{2} \|\mathcal{P}_\Lambda W^\top \alpha - \mathcal{P}_\Lambda f\|_2^2 + \frac{\kappa}{2} \|(I - WW^\top)\alpha\|_2^2 + \|\text{diag}(\lambda)\alpha\|_1.$$

Now, we follow arguments that is similar to those in [16] to show that Algorithm 4.1 converges to a minimizer of (4.18) (i.e. (4.19) with $\kappa = 1$). Define

$$(4.20) \quad F_1(\alpha) := \|\text{diag}(\lambda)\alpha\|_1 \quad \text{and} \quad F_2^\kappa(\alpha) := \frac{1}{2} \|\mathcal{P}_\Lambda(W^\top \alpha) - \mathcal{P}_\Lambda f\|_2^2 + \frac{\kappa}{2} \|(I - WW^\top)\alpha\|_2^2.$$

Here for the case $\kappa = 1$ we denote $F_2^1(\alpha)$ simply as $F_2(\alpha)$. Then the minimization problem (4.18) can be written as

$$\min_{\alpha \in \mathbb{R}^m} F_1(\alpha) + F_2(\alpha).$$

For any proper, convex, lower semi-continuous function φ which takes values in $(-\infty, +\infty]$, its *proximal operator* (see e.g. [109, 137, 138]) is defined by

$$(4.21) \quad \text{prox}_\varphi(\mathbf{x}) := \arg \min_{\mathbf{y}} \left\{ \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2^2 + \varphi(\mathbf{y}) \right\}.$$

It is easy to show that (see e.g. [58])

$$(4.22) \quad \text{prox}_{F_1}(\beta) = \arg \min_{\alpha \in \mathbb{R}^m} \left\{ \frac{1}{2} \|\beta - \alpha\|_2^2 + \|\text{diag}(\lambda)\alpha\|_1 \right\} = \mathcal{T}_\lambda(\beta).$$

Now, we can rewrite the iterations of α_k in (4.17) in terms of proximal operators defined in (4.21).

Lemma 4.1. *Iteration of α_k given by (4.17) is equivalent to*

$$(4.23) \quad \alpha_{k+1} = \text{prox}_{F_1}(\alpha_k - \nabla F_2(\alpha_k)).$$

PROOF. Indeed, by (4.22), we have

$$\begin{aligned} \alpha_{k+1} &= \text{prox}_{F_1}[W\mathcal{P}_\Lambda f + W(I - \mathcal{P}_\Lambda)W^\top \alpha_k] \\ &= \text{prox}_{F_1}[\alpha_k - \alpha_k + W\mathcal{P}_\Lambda f + WW^\top \alpha_k - W\mathcal{P}_\Lambda W^\top \alpha_k] \\ &= \text{prox}_{F_1}[\alpha_k - ((I - WW^\top)\alpha_k + W\mathcal{P}_\Lambda(\mathcal{P}_\Lambda W^\top \alpha_k - \mathcal{P}_\Lambda f))]. \end{aligned}$$

Since $(I - WW^\top)^2 = I - WW^\top$, we have

$$\begin{aligned} \alpha_{k+1} &= \text{prox}_{F_1}[\alpha_k - ((I - WW^\top)^2 \alpha_k + W\mathcal{P}_\Lambda(\mathcal{P}_\Lambda W^\top \alpha_k - \mathcal{P}_\Lambda f))] \\ &= \text{prox}_{F_1}[\alpha_k - \nabla \left(\frac{1}{2} \|\mathcal{P}_\Lambda W^\top \alpha_k - \mathcal{P}_\Lambda f\|_2^2 + \frac{1}{2} \|(I - WW^\top)\alpha_k\|_2^2 \right)] \\ &= \text{prox}_{F_1}(\alpha_k - \nabla F_2(\alpha_k)). \end{aligned}$$

This concludes the proof of the lemma. \blacksquare

The new form of iteration (4.23) of α_k is known as the proximal forward-backward splitting. A rather comprehensive convergence analysis of this type of iterations was given in [58]. Let us now recall the main convergence theorem in [58] for the finite dimensional case.

Theorem 4.1. *Consider the minimization problem*

$$(4.24) \quad \min_{x \in \mathbb{R}^m} F_1(x) + F_2(x),$$

where $F_1 : \mathbb{R}^m \mapsto \bar{\mathbb{R}}$ is a proper, convex, lower semi-continuous function, and $F_2 : \mathbb{R}^m \mapsto \mathbb{R}$ is a convex, differentiable function with an L -Lipschitz continuous gradient. Assume a minimizer of (4.24) exists. Then for any initial guess x_0 , the iteration (called the proximal forward-backward splitting):

$$x_{k+1} = \text{prox}_{F_1/L}(x_k - \nabla F_2(x_k)/L)$$

converges to a minimizer of $F_1(x) + F_2(x)$.

We do not provide a proof for Theorem 4.1 because it does not reveal any information on the speed of convergence of the sequence α_k . In the next section, we prove another version of Theorem 4.1 (Theorem 4.4) for a general balanced approach (4.8).

By Lemma 4.1, proving convergence of Algorithm 4.1 is now reduced to proving that F_1 and F_2 defined in (4.20) satisfy all the assumptions in Theorem 4.1.

Theorem 4.2. *The sequence α_k defined by (4.17) converges to a minimizer of the minimization problem (4.18) for any initial guess α_0 .*

PROOF. Obviously F_1 is a proper, convex, lower semi-continuous (in fact continuous) function, and F_2 is a convex and differentiable function. Therefore, all we need to prove are:

- (1) F_2 is 1-Lipschitz.
- (2) $F_1(\alpha) + F_2(\alpha)$ is coercive, i.e. $F_1(\alpha) + F_2(\alpha) \rightarrow \infty$ as $\|\alpha\|_2 \rightarrow \infty$, which implies that $\min_{\alpha} \{F_1(\alpha) + F_2(\alpha)\}$ has at least one solution.

For (1), noting that $W(I - \mathcal{P}_{\Lambda})W^{\top} \succeq 0$, we have

$$\begin{aligned} \nabla^2 F_2 &= (I - WW^{\top}) + W\mathcal{P}_{\Lambda}W^{\top} \\ &\preceq (I - WW^{\top}) + WW^{\top} \\ &= I \end{aligned}$$

Therefore, F_2 is 1-Lipschitz.

For (2), we have

$$F_1(\alpha) + F_2(\alpha) \geq F_1(\alpha) = \sum_{j=1}^m \lambda_j |\alpha_j| \geq \lambda_{\min} \|\alpha\|_1 \geq \lambda_{\min} \|\alpha\|_2,$$

which proves coercivity of $F_1(\alpha) + F_2(\alpha)$. ■

Theorem 4.2 proves convergence of Algorithm 4.1 to (4.19) for the case $\kappa = 1$. In fact, Theorem 4.1 suggests the following Algorithm 4.2 for the general inpainting model (4.19) with $0 \leq \kappa < \infty$.

Algorithm 4.2.

- (i) Set an initial guess α_0 .
- (ii) Iterate on k until convergence:

$$(4.25) \quad \alpha_{k+1} = \mathcal{T}_{\lambda/L} \left(\alpha_k - \nabla F_2^{\kappa}(\alpha_k)/L \right),$$

where L is the Lipschitz constant of F_2^{κ} defined by (4.6).

- (iii) Let α^* to the output of Step (ii). If $\eta(i) = 0$ for all $i \in \Lambda$ in (4.13), we set $u^* := \mathcal{P}_{\Lambda}f + (I - \mathcal{P}_{\Lambda})W^{\top}\alpha^*$ to be the solution (to the inpainting problem); otherwise, since $\mathcal{T}_{\lambda/L}$ can remove noise, we set $u^{\diamond} = W^{\top}\alpha^*$ to be the solution (to the inpainting-plus-denoising problem).

Then we have the following convergence theorem.

Theorem 4.3. *The sequence α_k defined by (4.25) converges to a minimizer of the minimization problem (4.19) with $0 \leq \kappa < \infty$ for any initial guess α_0 .*

PROOF. All we need to show is that L , the Lipschitz constant of F_2^κ , is bounded. In fact, we will show that $L \leq \max\{1, \kappa\}$. Indeed, we have

$$\begin{aligned} \nabla^2 F_2^\kappa &= \kappa(I - WW^\top) + W\mathcal{P}_\Lambda W^\top \\ &\leq \kappa(I - WW^\top) + WW^\top \\ &\leq \kappa I + (1 - \kappa)WW^\top \\ &\leq \begin{cases} I, & \text{for } 0 \leq \kappa \leq 1 \\ \kappa I, & \text{for } \kappa \geq 1 \end{cases} \end{aligned}$$

which implies $L \leq \max\{1, \kappa\}$. ■

2.2. Proximal forward-backward splitting for general balanced approach

This section presents a proximal forward-backward splitting algorithm for the general balanced approach (4.8) with $0 \leq \kappa < \infty$, as well as some convergence analysis of the algorithm. The corresponding algorithm for the balanced approach involving multiple frame systems (4.12) can be obtained similarly. Let

$$F_1(\alpha) = \|\text{diag}(\lambda)\alpha\|_1 \text{ and } F_2(\alpha) = \frac{1}{2}\|AW^\top\alpha - f\|_D^2 + \frac{\kappa}{2}\|(I - WW^\top)\alpha\|_2^2.$$

Then, the proximal forward-backward splitting algorithm can be written in the following short form (see [14, 15, 16, 17, 25, 32, 34, 36, 38, 58])

$$\alpha_{k+1} = \text{prox}_{F_1/L}(\alpha_k - \nabla F_2(\alpha_k)/L).$$

The detailed form of the above algorithm is given as follows:

Algorithm 4.3. *For a given vector $\lambda \in \mathbb{R}_+^m$, choose any $\alpha_0 \in \mathbb{R}^m$. For $k = 0, 1, 2, \dots$, generate α_{k+1} from α_k according to the following iterations:*

- (i) Set $g_k = \alpha_k - \nabla F_2(\alpha_k)/L$;
- (ii) Set $\alpha_{k+1} = \mathcal{T}_{\lambda/L}(g_k)$.

Here, the gradient of F_2 is given by

$$\nabla F_2(\alpha) = WA^T D(AW^\top\alpha - f) + \kappa(I - WW^\top)\alpha,$$

and the Hessian matrix of F_2 is

$$\nabla^2 F_2 = WA^T DAW^\top + \kappa(I - WW^\top).$$

It is obvious that F_2 is L -Lipschitz, and $L \leq \rho_{\max}(A^\top DA) + \kappa$. This generalizes the inpainting algorithms given in the previous section to algorithms for various image restoration problems. Although the original development of algorithms took a different path, this idea is used in the proof of the convergence of the balanced approach frame based algorithms given in [14, 15, 16, 17, 25, 32, 34, 36, 38].

First, we prove the following lemma ([168, Lemma 2.1]) which characterizes the optimal solutions of (4.8).

Lemma 4.2. *For each $\lambda > 0$, the optimal solution α^* of (4.8) satisfies*

$$\|\alpha^*\|_1 \leq C,$$

where

$$C = \begin{cases} \min\{\|f\|_D^2/2, \|\text{diag}(\lambda)\alpha_{\text{LS}}\|_1\}/\lambda_{\min} & \text{if } A \text{ is surjective} \\ \|f\|_D^2/(2\lambda_{\min}) & \text{otherwise,} \end{cases}$$

with $\alpha_{\text{LS}} := WA^\top(AA^\top)^{-1}f$.

PROOF. Let α^* be an optimal solution of (4.8). Then

$$\lambda_{\min}\|\alpha^*\|_1 \leq F_1(\alpha^*) + F_2(\alpha^*) \leq F_1(0) + F_2(0) = \frac{1}{2}\|f\|_D^2$$

Hence $\|\alpha^*\|_1 \leq \|f\|_D^2/(2\lambda_{\min})$. In addition, if A is surjective then we have

$$\lambda_{\min}\|\alpha^*\|_1 \leq F_1(\alpha^*) + F_2(\alpha^*) \leq F_1(\alpha_{\text{LS}}) + F_2(\alpha_{\text{LS}}) = \|\text{diag}(\lambda)\alpha_{\text{LS}}\|_1.$$

This concludes the proof of the lemma. \blacksquare

Now, we prove that the sequence $\{\alpha_k\}_{k \geq 0}$ generated by Algorithm 4.3 converges to a minimizer of:

$$\min_{\alpha \in \mathbb{R}^m} F_1(\alpha) + F_2(\alpha),$$

under suitable conditions. A similar proof can be found in e.g. [4]. For notational convenience we denote $F(\alpha) := F_1(\alpha) + F_2(\alpha)$ and

$$\ell_F(\alpha; \beta) := F_2(\beta) + \langle \nabla F_2(\beta), \alpha - \beta \rangle + \|\text{diag}(\lambda)\alpha\|_1,$$

where the sum of the first two terms is the linear approximation of F_2 at β . Since F_2 has an L -Lipschitz continuous gradient and is convex, we have the following inequality

$$(4.26) \quad F(\alpha) - \frac{L}{2}\|\alpha - \beta\|_2^2 \leq \ell_F(\alpha; \beta), \quad \text{for all } \alpha, \beta \in \mathbb{R}^m.$$

Theorem 4.4. *Consider the minimization problem*

$$(4.27) \quad \min_{\alpha \in \mathbb{R}^m} F_1(\alpha) + F_2(\alpha),$$

where $F_1 = \|\text{diag}(\lambda)\alpha\|_1$ and $F_2 : \mathbb{R}^m \mapsto \mathbb{R}$ is a convex, differentiable function with an L -Lipschitz continuous gradient. Let $F := F_1 + F_2$ and denote α^* be a solution of (4.27). Then the sequence $\{\alpha_k\}_{k \geq 0}$ generated by Algorithm 4.3 satisfies

$$(4.28) \quad F(\alpha_k) - F(\alpha^*) \leq \frac{L\|\alpha^* - \alpha_0\|_2^2}{2k}.$$

As a consequence, for given $\epsilon > 0$, we have

$$(4.29) \quad F(\alpha_k) - F(\alpha^*) \leq \epsilon, \quad \text{whenever } k \geq \frac{L(C + \|\alpha_0\|_2)^2}{2\epsilon}.$$

When (4.27) has a unique solution α^* , we have

$$\lim_{k \rightarrow \infty} \|\alpha_k - \alpha^*\|_2 = 0.$$

PROOF. For a given $k \geq 1$ and any optimal solution α^* , let

$$\alpha_{k+1} := \mathcal{T}_{\lambda/L}(\alpha_k - \nabla F_2(\alpha_k)/L).$$

First, we show that

$$(4.30) \quad \alpha_{k+1} \in \arg \min_{\alpha} \{\ell_F(\alpha; \alpha_k) + L\langle \alpha_{k+1} - \alpha_k, \alpha \rangle\}.$$

By letting $g_k = \alpha_k - \nabla F_2(\alpha_k)/L$ and removing constant terms in (4.30), it is easy to see that (4.30) is equivalent to

$$(4.31) \quad \alpha_{k+1} \in \arg \min_{\alpha} \{ \langle \alpha_{k+1} - g_k, \alpha \rangle + \|\text{diag}(\lambda)\alpha\|_1/L \}.$$

Now by the definition of α_{k+1} , we have

$$g_k[j] - \alpha_{k+1}[j] = \begin{cases} \text{sgn}(g_k[j])\lambda_j/L, & \text{when } |g_k[j]| > \lambda_j/L \\ g_k[j], & \text{when } |g_k[j]| \leq \lambda_j/L. \end{cases}$$

This implies that

$$g_k - \alpha_{k+1} \in \partial \|\text{diag}(\lambda)\alpha_{k+1}\|_1/L,$$

or equivalently,

$$0 \in \alpha_{k+1} - g_k + \partial \|\text{diag}(\lambda)\alpha_{k+1}\|_1/L,$$

which implies (4.31) and hence (4.30). By (4.30), we now have

$$(4.32) \quad \ell_F(\alpha_{k+1}; \alpha_k) + L \langle \alpha_{k+1} - \alpha_k, \alpha_{k+1} \rangle \leq \ell_F(\alpha^*; \alpha_k) + L \langle \alpha_{k+1} - \alpha_k, \alpha^* \rangle.$$

Letting $\alpha = \alpha_{k+1}$ and $\beta = \alpha_k$ in (4.26), we have

$$(4.33) \quad F(\alpha_{k+1}) \leq \ell_F(\alpha_{k+1}; \alpha_k) + \frac{L}{2} \|\alpha_{k+1} - \alpha_k\|_2^2.$$

Applying (4.32) to (4.33), we have

$$\begin{aligned} F(\alpha_{k+1}) &\leq \ell_F(\alpha^*; \alpha_k) + L \langle \alpha_{k+1} - \alpha_k, \alpha^* - \alpha_{k+1} \rangle + \frac{L}{2} \|\alpha_{k+1} - \alpha_k\|_2^2 \\ &= \ell_F(\alpha^*; \alpha_k) + \frac{L}{2} \|\alpha^* - \alpha_k\|_2^2 - \frac{L}{2} \|\alpha^* - \alpha_{k+1}\|_2^2, \end{aligned}$$

where the last equality follows from straightforward calculations. Now, we have

$$(4.34) \quad \begin{aligned} F(\alpha_{k+1}) &\leq \ell_F(\alpha^*; \alpha_k) + \frac{L}{2} \|\alpha^* - \alpha_k\|_2^2 - \frac{L}{2} \|\alpha^* - \alpha_{k+1}\|_2^2, \\ &\leq F(\alpha^*) + \frac{L}{2} \|\alpha^* - \alpha_k\|_2^2 - \frac{L}{2} \|\alpha^* - \alpha_{k+1}\|_2^2. \end{aligned}$$

Here, the last inequality follows from the definition of ℓ_F and the convexity of F_2 .

Subtracting $F(\alpha^*)$ from both sides of the last inequality of (4.34) gives

$$F(\alpha_{k+1}) - F(\alpha^*) \leq \frac{L}{2} \|\alpha^* - \alpha_k\|_2^2 - \frac{L}{2} \|\alpha^* - \alpha_{k+1}\|_2^2.$$

Telescoping on the above inequality, we have

$$(4.35) \quad \sum_{j=1}^{k+1} F(\alpha_j) - (k+1)F(\alpha^*) \leq \frac{L}{2} \|\alpha^* - \alpha_0\|_2^2.$$

Now, by using (4.30) with $\alpha = \alpha_k$, we have

$$\begin{aligned} \ell_F(\alpha_{k+1}; \alpha_k) + L \langle \alpha_{k+1} - \alpha_k, \alpha_{k+1} \rangle &\leq \ell_F(\alpha_k; \alpha_k) + L \langle \alpha_{k+1} - \alpha_k, \alpha_k \rangle \\ &= F(\alpha_k) + L \langle \alpha_{k+1} - \alpha_k, \alpha_k \rangle. \end{aligned}$$

By applying the above inequality to (4.33) we obtain

$$F(\alpha_{k+1}) - F(\alpha_k) \leq -\frac{L}{2} \|\alpha_{k+1} - \alpha_k\|_2^2.$$

Combining the above inequality with (4.35), we have

$$(k+1)(F(\alpha_{k+1}) - F(\alpha^*)) \leq \frac{L}{2} \|\alpha^* - \alpha_0\|_2^2,$$

and thus (4.28) follows. In addition, by applying the following inequalities

$$\|\alpha^* - \alpha_0\|_2 \leq \|\alpha^*\|_2 + \|\alpha_0\|_2 \leq \|\alpha^*\|_1 + \|\alpha_0\|_2,$$

and Lemma 4.2, we obtain (4.29).

The conclusion that $\alpha_k \rightarrow \alpha^*$ whenever α^* is the unique minimizer of F follows directly from Proposition 4.1. ■

Remark 4.11. Theorem 4.1 whose proof is given in [58] says that the sequence α_k itself converges to an optimal solution α^* whenever there exists one as given in Theorem 4.1, which is stronger than what we have shown above. The interested reader should consult [58] for details. However, the speed of convergence, which is more important in practice, is not clear from the proof of [58]. Here, we focused on showing that α_k is a minimizing sequence that reaches to an ϵ -optimal solution after $O(L/\epsilon)$ iterations. Furthermore, in image restorations we are more interested in finding a solution at which the cost functional has a value that is close to the minimal value, rather than knowing whether the solution is close to a minimizer or not. In this regard, the proof here is more informative.

2.3. Accelerated algorithm

As shown in Theorem 4.4, the proximal forward-backward splitting algorithm generates an ϵ -optimal solution of (4.8) in $O(L/\epsilon)$ iterations, which is reasonably efficient. However, in practice, faster algorithms are always desired. For example when the matrix A is huge (e.g. Radon transform for tomography in 3D), each iteration of Algorithm 4.3 can be expensive. Therefore, one always wishes to reduce the total number of iterations. As we will see from Theorem 4.5 the accelerated proximal gradient (APG) algorithm of [168] (see also the FISTA algorithm of [4]) is much more efficient than the proximal forward-backward splitting algorithm because it only requires $O(\sqrt{L}/\epsilon)$ iterations to obtain an ϵ -optimal solution.

The APG algorithm of [168] is obtained by adjusting the step $g_k = \alpha_k - \nabla F_2(\alpha_k)/L$ in the proximal forward-backward splitting algorithm. This idea has already appeared in [4, 173]. Next, we describe the APG algorithm for solving (4.8).

Algorithm 4.4. (APG for (4.8)) *For a given vector $\lambda \in \mathbb{R}_+^m$, choose $\alpha_0 = \alpha_{-1} \in \mathbb{R}^m$, $t_0 = 1$, and $t_{-1} = 0$. For $k = 0, 1, 2, \dots$, generate α_{k+1} from α_k according to the following iteration:*

- (i) Set $\beta_k = \alpha_k + \frac{t_{k-1}-1}{t_k}(\alpha_k - \alpha_{k-1})$.
- (ii) Set $g_k = \beta_k - \nabla F_2(\beta_k)/L$.
- (iii) Set $\alpha_{k+1} = \mathcal{T}_{\lambda/L}(g_k)$.
- (iv) Compute $t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}$.

Note that, when $\kappa = 0$, i.e. $F_2(\alpha) = \frac{1}{2}\|AW^\top \alpha - f\|_D^2$, Algorithm 4.4 leads to an efficient algorithm for the synthesis based approach. When we take $t_k = 1$ for all k in the APG algorithm, it is reduced to the proximal forward-backward splitting algorithm for the balanced approach in frame based image restorations, and it is the popular iterative shrinkage/thresholding algorithms proposed in [62, 66, 90, 91].

The iterative shrinkage/thresholding algorithms and the proximal forward-backward splitting algorithms have been developed and analyzed independently

by many researchers. These algorithms only require gradient evaluations and soft-thresholding operations, so the computation cost for each iteration is very cheap. But, for any $\epsilon > 0$, these algorithms terminate in $O(L/\epsilon)$ iterations with an ϵ -optimal solution as proven in Theorem 4.4. The APG algorithm proposed here, however, reaches an ϵ -optimal solution in $O(\sqrt{L}/\epsilon)$ iterations. Thus APG algorithm of [168] greatly accelerates the proximal forward-backward splitting algorithms used in [14, 15, 16, 17, 25, 32, 34, 36, 38] for the balanced approach in frame based image restorations.

Here, we recall the proof of convergence of Algorithm 4.4 given by [168, Theorem 2.1], which gives an upper bound on the number of iterations for the APG algorithm for solving (4.8) to achieve ϵ -optimality. The proof follows a similar idea to that of [4, 173] and Theorem 4.4.

We note that for the sequence t_k defined in Algorithm 4.4, we have the following inequality which can be easily verified by induction

$$(4.36) \quad t_k \geq \frac{k+1}{2}.$$

Theorem 4.5. *Let $\{\alpha_k\}$, $\{\beta_k\}$, and $\{t_k\}$ be the sequences given by Algorithm 4.4. Then for any $k \geq 1$ and any optimal solution α^* to the minimization problem (4.8) with $0 \leq \kappa < \infty$, we have*

$$(4.37) \quad F(\alpha_k) - F(\alpha^*) \leq \frac{2L\|\alpha^* - \alpha_0\|_2^2}{k^2}.$$

Hence

$$(4.38) \quad F(\alpha_k) - F(\alpha^*) \leq \epsilon, \quad \text{whenever } k \geq \sqrt{\frac{2L}{\epsilon}} (\|\alpha_0\|_2 + C),$$

where C is given by Lemma 4.2. Furthermore, if α^* is the unique minimizer of $F(\alpha)$, then $\alpha_k \rightarrow \alpha^*$ as $k \rightarrow \infty$.

PROOF. For a given $k \geq 1$ and any optimal solution α^* , let $s_k := \mathcal{T}_{\lambda/L}(g_k)$ and $\tilde{\alpha} := ((t_k - 1)\alpha_k + \alpha^*)/t_k$. We first show that

$$(4.39) \quad s_k \in \arg \min_{\alpha} \{\ell_F(\alpha; \beta_k) + L\langle s_k - \beta_k, \alpha \rangle\}.$$

By definition of g_k in Algorithm 4.4 and removing constant terms in (4.39), it is easy to see that (4.39) is equivalent to

$$(4.40) \quad s_k \in \arg \min_{\alpha} \{\langle s_k - g_k, \alpha \rangle + \|\text{diag}(\lambda)\alpha\|_1/L\}.$$

Now by the definition of s_k , we have

$$g_k[j] - s_k[j] = \begin{cases} \text{sgn}(g_k[j])\lambda_j/L, & \text{when } |g_k[j]| > \lambda_j/L \\ g_k[j], & \text{when } |g_k[j]| \leq \lambda_j/L. \end{cases}$$

This implies that

$$g_k - s_k \in \partial\|\text{diag}(\lambda)s_k\|_1/L,$$

or equivalently

$$0 \in s_k - g_k + \partial\|\text{diag}(\lambda)s_k\|_1/L,$$

which implies (4.40) and hence (4.39). By (4.39), we now have

$$(4.41) \quad \ell_F(s_k; \beta_k) + L\langle s_k - \beta_k, s_k \rangle \leq \ell_F(\tilde{\alpha}; \beta_k) + L\langle s_k - \beta_k, \tilde{\alpha} \rangle.$$

Letting $\alpha = s_k$ and $\beta = \beta_k$ in (4.26) and noting that $\alpha_{k+1} = s_k$, we have

$$F(\alpha_{k+1}) = F(s_k) \leq \ell_F(s_k; \beta_k) + \frac{L}{2} \|s_k - \beta_k\|_2^2.$$

Applying (4.41) to the above inequality, we have

$$\begin{aligned} F(\alpha_{k+1}) &\leq \ell_F(\tilde{\alpha}; \beta_k) + L \langle s_k - \beta_k, \tilde{\alpha} - s_k \rangle + \frac{L}{2} \|s_k - \beta_k\|_2^2 \\ &= \ell_F(\tilde{\alpha}; \beta_k) + \frac{L}{2} \|\tilde{\alpha} - \beta_k\|_2^2 - \frac{L}{2} \|\tilde{\alpha} - s_k\|_2^2 \\ &= \ell_F(\tilde{\alpha}; \beta_k) + \frac{L}{2} \|\tilde{\alpha} - \beta_k\|_2^2 - \frac{L}{2} \|\tilde{\alpha} - \alpha_{k+1}\|_2^2, \end{aligned}$$

where the second to the last equality follows from straightforward calculations. Plug the definition of $\tilde{\alpha}$ into the above inequality and denoting $\gamma_k := (t_{k-1} - 1)\alpha_{k-1} - t_{k-1}\alpha_k$, one obtains

(4.42)

$$\begin{aligned} F(\alpha_{k+1}) &\leq \frac{t_k - 1}{t_k} \ell_F(\alpha_k; \beta_k) + \frac{1}{t_k} \ell_F(\alpha^*; \beta_k) \\ &\quad + \frac{L}{2t_k^2} \|(t_k - 1)\alpha_k + \alpha^* - t_k\beta_k\|_2^2 - \frac{L}{2t_k^2} \|(t_k - 1)\alpha_k + \alpha^* - t_k\alpha_{k+1}\|_2^2, \\ &= \frac{t_k - 1}{t_k} \ell_F(\alpha_k; \beta_k) + \frac{1}{t_k} \ell_F(\alpha^*; \beta_k) + \frac{L}{2t_k^2} \|\alpha^* - \gamma_k\|_2^2 - \frac{L}{2t_k^2} \|\alpha^* - \gamma_{k+1}\|_2^2 \\ &\leq \frac{t_k - 1}{t_k} F(\alpha_k) + \frac{1}{t_k} F(\alpha^*) + \frac{L}{2t_k^2} \|\alpha^* - \gamma_k\|_2^2 - \frac{L}{2t_k^2} \|\alpha^* - \gamma_{k+1}\|_2^2. \end{aligned}$$

Here, the first inequality follows from the fact that ℓ_F is convex and $t_k \geq 1$ by (4.36). The last inequality follows from the convexity of F_2 .

Subtracting $F(\alpha^*)$ from both sides of the last inequality of (4.42), multiplying t_k^2 at both sides and noticing that $t_{k-1}^2 = t_k(t_k - 1)$, we have

$$t_k^2 (F(\alpha_{k+1}) - F(\alpha^*)) \leq t_{k-1}^2 (F(\alpha_k) - F(\alpha^*)) + \frac{L}{2} \|\alpha^* - \gamma_k\|_2^2 - \frac{L}{2} \|\alpha^* - \gamma_{k+1}\|_2^2.$$

Telescoping on the above inequality, and using $t_{-1} = 0$ and $\gamma_0 = \alpha_0$, we have

$$t_k^2 (F(\alpha_{k+1}) - F(\alpha^*)) \leq \frac{L}{2} \|\alpha^* - \alpha_0\|_2^2.$$

By (4.36), the inequality (4.37) follows; and using Lemma 4.2, we obtain (4.38).

The conclusion that $\alpha_k \rightarrow \alpha^*$ whenever α^* is the unique minimizer of F follows directly from Proposition 4.1. \blacksquare

Remark 4.12.

- (1) Theorem 4.5 shows that the number of iterations to get an ϵ -optimal solution is of order $O(\sqrt{L/\epsilon})$. Therefore, less iterations are required if we have a smaller Lipschitz constant L . For the general balanced approach, we have $L \leq \rho_{\max}(A^\top DA) + \kappa$. Then for given A and D , the smaller κ is, the faster α_k from the APG algorithm converges.
- (2) For the inpainting problem considered in the previous section, we have $A = \mathcal{P}_\Lambda$ and $D = I$. We have shown in the proof of Theorem 4.3 that $L = \max\{1, \kappa\}$. Thus APG converges fastest whenever $\kappa \leq 1$ and $\kappa = 1$

is the case where we have fast convergence while the reconstructed image has good regularity.

- (3) It is clear that the Algorithm (4.4) can be used for both the balanced approach and the synthesis based approach (the combination of the two is called the general balanced approach here). In general, the balanced approach and the synthesis based approach can have the same algorithm by using different cost functionals, while algorithms for the analysis based approach need to be developed separately.

We finally present the APG algorithm for the two-system balanced approach (4.12) whose convergence can be proved in a manner similarly to Theorem 4.5. We denote

$$F_2(\alpha^1, \alpha^2) := \frac{1}{2} \|A(\sum_{i=1}^2 W_i^\top \alpha^i) - f\|_D^2 + \sum_{i=1}^2 \frac{\kappa_i}{2} \|(I - W_i W_i^\top) \alpha^i\|_2^2$$

and denote its Lipschitz constant as L (see [168] for an estimate of L). We note that the two-system balanced approach generally works better than (4.8) when the given image has two different layers, e.g. cartoon and textures that can be sparsely approximated by two different tight frame systems.

Algorithm 4.5. (APG for (4.12)) *Let $i = 1, 2$. For given vectors $\lambda^i \in \mathbb{R}_+^m$, choose $\alpha_0^i = \alpha_{-1}^i \in \mathbb{R}^m$, $t_0 = 1$, and $t_{-1} = 0$. For $k = 0, 1, 2, \dots$, generate α_{k+1}^i from α_k^i according to the following iteration:*

(i) *Set*

$$\beta_k^1 = \alpha_k^1 + \frac{t_{k-1} - 1}{t_k} (\alpha_k^1 - \alpha_{k-1}^1) \quad \text{and} \quad \beta_k^2 = \alpha_k^2 + \frac{t_{k-1} - 1}{t_k} (\alpha_k^2 - \alpha_{k-1}^2).$$

(ii) *Set*

$$g_k^1 = \beta_k^1 - \nabla_{\alpha^1} F_2(\beta_k^1, \beta_k^2) / L \quad \text{and} \quad g_k^2 = \beta_k^2 - \nabla_{\alpha^2} F_2(\beta_k^1, \beta_k^2) / L.$$

(iii) *Set*

$$\alpha_{k+1}^1 = \mathcal{T}_{\lambda^1/L}(g_k^1) \quad \text{and} \quad \alpha_{k+1}^2 = \mathcal{T}_{\lambda^2/L}(g_k^2).$$

(iv) *Compute $t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}$.*

2.4. Linearized bregman algorithm

Recently, the linearized Bregman algorithm has been applied to develop a fast algorithm for frame based image deblurring in [23], which converges to the minimizer of the following minimization problem:

$$(4.43) \quad \min_{\alpha \in \mathbb{R}^m} \left\{ \frac{\kappa}{2} \|\alpha\|_2^2 + \|\text{diag}(\lambda)\alpha\|_1 : AW^\top \alpha = f \right\},$$

when A is surjective. Furthermore, the optimal solution u_λ^* converges to the minimizer of $\min_{\alpha \in \mathbb{R}^m} \{\|\alpha\|_1 : AW^\top \alpha = f\}$ as $\lambda \rightarrow \infty$, where, for simplicity, each entry of the vector λ is set to be the same (see [23]). Hence, the linearized Bregman algorithm is used to solve a variation of the synthesis based approach when λ is sufficiently large. The linearized Bregman algorithm was first proposed to solve ℓ_1 -minimization problems in compressed sensing by [184] and was made efficient in [147]. The convergence analysis of the linearized Bregman algorithm was given in [21, 22]. Then, it was used in the nuclear norm minimization for matrix completion

by [13], where the authors realized that the linearized Bregman can be reformulated as the Uzawa's algorithm (see e.g. [82] for more details about the Uzawa algorithm).

A simple computation of [23] shows that when A is invertible, (4.43) is equivalent to

$$\min_{\alpha \in \mathbb{R}^m} \left\{ \frac{\kappa}{2} \|(I - WW^\top)\alpha\|_2^2 + \|\text{diag}(\lambda)\alpha\|_1 : AW^\top\alpha = f \right\}.$$

This looks like a variation of the balanced approach. However, when the large parameter vector λ is chosen, which happens when one applies linearized Bregman, it is closer to a variation of the synthesis based approach. A detailed discussion of the case when A is not invertible is provided in [23].

The advantage of Bregman iterations (either linearized Bregman or split Bregman iterations) in frame based image restorations is that big coefficients come back at first after few iterations and stay. This, in particular, is important to image deblurring, since big wavelet frame coefficients contain information of edges and features of images. The main goal of deblurring is to restore blurred edges and features.

Iterative algorithms involving Bregman distance were introduced to image and signal processing by many authors including [30, 31]. See [145] for an overview. In [145], a Bregman iteration was proposed for the nondifferentiable TV energy for image restoration. Then, in [184], it was shown to be remarkably successful for ℓ_1 -norm minimization problems in compressive sensing. To further improve the performance of the Bregman iteration, a linearized Bregman algorithm was invented in [59]; see also [184]. More details and an improvement called "kicking" of the linearized Bregman algorithm were described in [147], and a rigorous theory was given in [21, 22]. Wavelet-based denoising using the Bregman iteration was introduced in [182], and was further extended by using translation invariant wavelets in [129]. Here, we focus on the linearized Bregman algorithm for wavelet frame based image restorations of [23].

The linearized Bregman algorithm is given as follows.

Algorithm 4.6. (*Linearized Bregman*)

- (i) Set initial guess $\beta_0 \in \mathbb{R}^m$.
- (ii) For $k = 0, 1, \dots$, perform the following two-line-iteration until convergence:

$$(4.44) \quad \begin{aligned} \alpha_{k+1} &= \frac{1}{\kappa} \mathcal{T}_\lambda(\beta_k) \\ \beta_{k+1} &= \beta_k + WA^\top(f - AW^\top\alpha_{k+1}). \end{aligned}$$

The linearized Bregman algorithm looks similar to proximal forward-backward splitting algorithm. For simplicity, we suppose $\kappa = 1$. If we let $F_1(\alpha) = \|\text{diag}(\lambda)\alpha\|_1$ and $F_2(\alpha) = \frac{1}{2}\|AW^\top\alpha - f\|_2^2$, then from (4.44) we have

$$\begin{aligned} \alpha_{k+1} &= \mathcal{T}_\lambda(\beta_k) = \text{prox}_{F_1}(\beta_{k-1} - \nabla F_2(\alpha_k)) \\ &= \text{prox}_{F_1}(\beta_0 - \sum_{j=1}^k \nabla F_2(\alpha_j)). \end{aligned}$$

Therefore, one essential difference between the linearized Bregman algorithm (4.44) and the proximal forward-backward splitting algorithm is that the linearized Bregman uses all $\{\alpha_j : j = 1, 2, \dots, k\}$ to update α_{k+1} , while the proximal forward-backward splitting algorithm only uses the approximation of the previous iteration α_k . On the other hand, the linearized Bregman algorithm is in fact a primal-dual algorithm, where the variable β_k links to the dual variable of α_k . We further explain this in the following discussions.

It was proved in [22] that if the sequence $\{\alpha_k\}_{k \in \mathbb{N}}$ from (4.44) converges, its limit is the unique solution of the following optimization problem

$$(4.45) \quad \min_{u \in \mathbb{R}^n} \left\{ \|\text{diag}(\lambda)\alpha\|_1 + \frac{\kappa}{2} \|\alpha\|_2^2 : AW^\top \alpha = f \right\},$$

where $0 < \kappa < \infty$. Later, in [21], it was proved that the sequence $\{u_k\}_{k \in \mathbb{N}}$ converges. A rather crucial discovery that makes it possible to link the linearized Bregman algorithm to the Uzawa's algorithm is the discovery of the optimization problem (4.45) corresponding to the linearized Bregman algorithm [22].

The linearized Bregman algorithm coincides with the Uzawa's algorithm (see e.g. [13, 82, 151] for details on Uzawa's algorithm) applied to the problem (4.45). In this section, we first show the equivalence of the linearized Bregman algorithm with Uzawa's algorithm when applied to (4.45). Then we prove convergence of Uzawa's method in a more general setting, which in turn gives a convergence proof of linearized Bregman algorithm.

For simplicity, we denote $B = AW^\top$ and denote the objective function in (4.45) as $F_\lambda(\alpha)$, i.e.

$$F_\lambda(\alpha) := \|\text{diag}(\lambda)\alpha\|_1 + \frac{\kappa}{2} \|\alpha\|_2^2.$$

The Lagrangian for the problem (4.45) is given by

$$\mathcal{L}(\alpha, p) = F_\lambda(\alpha) + \langle p, f - B\alpha \rangle.$$

If strong duality holds (which is indeed the case for (4.45)), then α^* and p^* are called primal-dual optimal if the pair (α^*, p^*) is a saddle point of the Lagrangian $\mathcal{L}(\alpha, p)$, i.e. the pair satisfies

$$(4.46) \quad \sup_p \inf_\alpha \mathcal{L}(\alpha, p) = \mathcal{L}(\alpha^*, p^*) = \inf_\alpha \sup_p \mathcal{L}(\alpha, p).$$

The Uzawa's algorithm is an iterative procedure that solves the above saddle point problem as follows

$$(4.47) \quad \begin{aligned} \alpha_{k+1} &= \arg \min_\alpha \mathcal{L}(\alpha, p_k) \\ p_{k+1} &= p_k + \delta_k (f - B\alpha_{k+1}), \end{aligned}$$

where $\{\delta_k\}_{k \in \mathbb{N}}$ are positive step sizes. Note that the second step in (4.47) is one step of gradient ascend solving the maximization problem $\sup_p \mathcal{L}(\alpha_{k+1}, p)$, because $\nabla \mathcal{L}(\alpha_{k+1}, p) = f - Bu_{k+1}$. Therefore, Uzawa's algorithm is an alternative minimization and maximization method solving the saddle point problem (4.46).

By adding some terms independent of variable α , we have

$$\begin{aligned} \arg \min_\alpha \mathcal{L}(\alpha, p_k) &= \arg \min_\alpha F_\lambda(\alpha) + \langle p_k, f - B\alpha \rangle \\ &= \arg \min_\alpha \|\text{diag}(\lambda)\alpha\|_1 + \frac{\kappa}{2} \|\alpha - \frac{1}{\kappa} B^\top p_k\|_2^2 \\ &= \frac{1}{\kappa} \mathcal{T}_\lambda(B^\top p_k). \end{aligned}$$

Applying B^\top on both sides of the second equation in (4.47), letting $\beta_k = B^\top p_k$ and $\delta_k = 1$, we obtain the linearized Bregman algorithm (4.44). In fact, the sequence α_k for the linearized Bregman algorithm is exactly the same as Uzawa's method. Therefore, we focus on analyzing the convergence of Uzawa's algorithm (4.47).

First, we establish the strong convexity of $F_\lambda(\alpha)$ in the following lemma.

Lemma 4.3. *Let $v_1 \in \partial F_\lambda(\alpha_1)$ and $v_2 \in \partial F_\lambda(\alpha_2)$. Then*

$$(4.48) \quad \langle v_1 - v_2, \alpha_1 - \alpha_2 \rangle \geq \kappa \|\alpha_1 - \alpha_2\|_2^2.$$

PROOF. An element in $v_i \in \partial F_\lambda(\alpha_i)$, $i = 1, 2$, takes the form $v_i = w_i + \kappa \alpha_i$, with $w_i \in \partial \|\text{diag}(\lambda)\alpha_i\|_1$. Therefore, we have

$$\langle v_1 - v_2, \alpha_1 - \alpha_2 \rangle = \langle w_1 - w_2, \alpha_1 - \alpha_2 \rangle + \kappa \|\alpha_1 - \alpha_2\|_2^2.$$

Thus we only need to show that the first term of the right-hand side is nonnegative. By the definition of subgradient, it is easy to see that $\langle w_i, \alpha_i \rangle = \|\text{diag}(\lambda)\alpha_i\|_1$. Furthermore, we have

$$\begin{aligned} \langle w_1, \alpha_2 \rangle &= \langle \text{diag}(\lambda)^{-1}w_1, \text{diag}(\lambda)\alpha_2 \rangle \\ &\leq \|\text{diag}(\lambda)^{-1}w_1\|_\infty \|\text{diag}(\lambda)\alpha_2\|_1 \leq \|\text{diag}(\lambda)\alpha_2\|_1. \end{aligned}$$

Similarly we have $\langle w_2, \alpha_1 \rangle \leq \|\text{diag}(\lambda)\alpha_1\|_1$. Therefore, we have

$$\begin{aligned} \langle w_1 - w_2, \alpha_1 - \alpha_2 \rangle &= \langle w_1, \alpha_1 \rangle + \langle w_2, \alpha_2 \rangle - \langle w_1, \alpha_2 \rangle - \langle w_2, \alpha_1 \rangle \\ &= \|\text{diag}(\lambda)\alpha_1\|_1 + \|\text{diag}(\lambda)\alpha_2\|_1 - \langle w_1, \alpha_2 \rangle - \langle w_2, \alpha_1 \rangle \\ &\geq 0 \end{aligned}$$

This concludes the proof of the lemma. \blacksquare

Theorem 4.6. *Suppose the step sizes obey $0 < \inf_k \delta_k \leq \sup_k \delta_k < 2\kappa/\|B\|_2^2$. Then the sequence α_k obtained from Uzawa's algorithm (4.47) converges to the unique solution of (4.45).*

PROOF. Let (α^*, p^*) be the optimal primal-dual pair for the problem (4.45). Then the optimality conditions in (4.46) and (4.47) give

$$\begin{aligned} v_{k+1} - B^\top p_k &= 0, \\ v^* - B^\top p^* &= 0, \end{aligned}$$

where $v_{k+1} \in \partial F_\lambda(\alpha_{k+1})$, and $v^* \in \partial F_\lambda(\alpha^*)$. Then we have,

$$(v_{k+1} - v^*) - B^\top(p_k - p^*) = 0,$$

and it follows from Lemma 4.3 that

$$\langle \alpha_{k+1} - \alpha^*, B^\top(p_k - p^*) \rangle = \langle \alpha_{k+1} - \alpha^*, v_{k+1} - v^* \rangle \geq \kappa \|\alpha_{k+1} - \alpha^*\|_2^2.$$

Thus,

$$\langle B(\alpha_{k+1} - \alpha^*), p_k - p^* \rangle \geq \kappa \|\alpha_{k+1} - \alpha^*\|_2^2.$$

Noting that $B\alpha^* = f$ and using the second equation in (4.47), we have

$$\begin{aligned} \|p_{k+1} - p^*\|_2^2 &= \|p_k - p^* + \delta_k(f - B\alpha_{k+1})\|_2^2 \\ &= \|p_k - p^* + \delta_k B(\alpha^* - \alpha_{k+1})\|_2^2 \\ &= \|p_k - p^* - \delta_k B(\alpha_{k+1} - \alpha^*)\|_2^2. \end{aligned}$$

If we let $r_k := \|p_k - p^*\|_2$, we then have

$$\begin{aligned} r_{k+1}^2 &= r_k^2 - 2\delta_k \langle p_k - p^*, B(\alpha_{k+1} - \alpha^*) \rangle + \delta_k^2 \|B(\alpha_{k+1} - \alpha^*)\|_2^2 \\ &\leq r_k^2 - 2\kappa\delta_k \|\alpha_{k+1} - \alpha^*\|_2^2 + \delta_k^2 \|B\|_2^2 \|\alpha_{k+1} - \alpha^*\|_2^2. \end{aligned}$$

From the assumptions of δ_k , we have $2\kappa\delta_k - \delta_k^2 \|B\|_2^2 \geq C$, for some $C > 0$. Therefore,

$$r_{k+1}^2 \leq r_k^2 - C \|\alpha_{k+1} - \alpha^*\|_2^2.$$

The above inequality shows that r_k is a monotonically decreasing sequence and thus converges. Applying limits to both sides of the above inequality gives $\lim_{k \rightarrow \infty} \|\alpha_{k+1} - \alpha^*\|_2 = 0$ and this concludes the proof of the theorem. ■

Now, we have the following convergence theorem for the linearized Bregman algorithm (4.44), which is a direct consequence of Theorem 4.6.

Theorem 4.7. *The sequence α_k obtained from Algorithm 4.6 converges to the unique solution of (4.45), provided that $\kappa > \|B\|_2^2/2$.*

PROOF. Convergence follows from Theorem 4.6, and the uniqueness of the solution of (4.45) follows from the fact that

$$\|\text{diag}(\lambda)\alpha\|_1 + \frac{\kappa}{2}\|\alpha\|_2^2$$

is strictly convex for any $\kappa > 0$. ■

Finally, we note that in order to speed-up Algorithm 4.6, i.e. to reduce the total number of iterations required for a given stopping criterion, a preconditioned version of Algorithm 4.6 was proposed in [23]. The idea is that instead of considering equation $AW^\top\alpha = f$ in (4.45), we consider the equivalent equation

$$D^{\frac{1}{2}}AW^\top\alpha = D^{\frac{1}{2}}f,$$

where D is a symmetric positive definite matrix that approximates the pseudoinverse of A . Then by replacing A with $D^{\frac{1}{2}}A$ and f with $D^{\frac{1}{2}}f$ in Algorithm 4.6, one obtains the preconditioned linearized Bregman algorithm. Here, we use the same notation D as in (4.8) for example, because they serve the same purpose. Therefore, just as in (4.8), we choose $D = (AA^\top + \theta I)^{-1}$.

2.5. Role of the redundancy

Tight frames are different from orthonormal systems because tight frames are redundant. What does the redundancy bring us here? We start with a philosophical point about the algorithm and then give some quantitative analysis on the error reduction with each iteration.

Assume that some blocks of pixels are missing in a given image and we want to solve the inpainting problem in the wavelet frame domain as mentioned before. Since the framelets used are compactly supported, the coefficients of those framelets whose supports fall in the missing blocks are missing and the coefficients of those framelets whose supports overlap with the missing blocks are inaccurate. The main step of Algorithm 4.1 perturbs the frame coefficients Wu_n via thresholding so that information contained in the available coefficients will permeate into the missing frame coefficients. Here, the redundancy is very important, since the available coefficients and its associated atoms in the system contain information of the missing

coefficients and missing atoms only if the system is redundant, as the atoms in an orthonormal basis are orthogonal to each other and do not contain information of other atoms in L_2 -sense.

While applying the thresholding operator on the frame coefficients is a very important step in Algorithm 4.1, to remove noise, perturb the coefficients and sharpen the edges, it also introduces new errors and artifacts. We use image inpainting as an example to show how the numerical errors and artifacts introduced by thresholding can be reduced by the redundancy of the system W . We take the computed solution u^* for the case $\eta = 0$ as an example. Similar analysis holds for the computation of each iteration. Our computed solution u^* that interpolates the given data satisfies $\alpha^* = \mathcal{T}_\lambda(Wu^*)$ and

$$u^* = \mathcal{P}_\Lambda f + (I - \mathcal{P}_\Lambda)W^\top \mathcal{T}_\lambda Wu^*.$$

That is, on Λ , $W^\top \mathcal{T}_\lambda Wu^*$ is replaced by f . But since $\mathcal{P}_\Lambda f = \mathcal{P}_\Lambda u^* = \mathcal{P}_\Lambda W^\top Wu^*$, we are actually replacing $\mathcal{P}_\Lambda W^\top \mathcal{T}_\lambda Wu^*$ by $\mathcal{P}_\Lambda W^\top Wu^*$, which generates artifacts. Hence to reduce the artifacts, we require that the norm of

$$\mathcal{P}_\Lambda W^\top Wu^* - \mathcal{P}_\Lambda W^\top \mathcal{T}_\lambda Wu^* = \mathcal{P}_\Lambda W^\top (Wu^* - \mathcal{T}_\lambda Wu^*)$$

is small.

Clearly, the smaller the norm of $W^\top \mathbf{e} := W^\top (Wu^* - \mathcal{T}_\lambda Wu^*)$ is, the smaller the artifact is. Note that the reconstruction operator W^\top can eliminate the error components sitting in the kernel of W^\top . In fact, since W^\top projects all sequences down to the orthogonal complement of the kernel of W^\top , which is the range of W , the component of \mathbf{e} in the kernel of W^\top does not contribute. The redundant system reduces the errors as long as the component of \mathbf{e} in the kernel of W^\top is not zero. Therefore, in general, the larger the kernel of W^\top is, i.e. the more redundant the frame system is, the greater the error reduction. To increase the redundancy, we use quasi-affine tight frame system (which is also called an undecimated tight frame system since no downsampling is required during decompositions). In contrast, if W is an orthonormal system (not redundant at all), then the kernel of W^\top is just $\{0\}$ and we have $\|W^\top \mathbf{e}\| = \|\mathbf{e}\|$. The point here is that redundancy reduces the norm of $(\alpha^* - Wf)$.

2.6. Some simulations

First, we provide a comparison of computation efficiency between the PFBS and the APG algorithm under the scenario of image inpainting. Qualities of inpainting and deblurring using both single- and two-system balanced approach (4.8) and (4.12) are presented as well. In particular, we present results using the synthesis based approach (i.e. (4.8) and (4.12) with $\kappa = 0$). Finally, we present deblurring results using the linearized Bregman algorithm given by Algorithm 4.6 with preconditioning.

For the balanced approach (4.8), we take W corresponding to the piecewise linear B-spline quasi-affine tight frame system. For the two-system balanced approach (4.12), we take W_1 the same as W for (4.8), while W_2 is generated by a local discrete cosine transform.

Throughout this section, we adopt the following stopping criterion for Algorithm 4.4 (we use a similar stopping criterion for Algorithm 4.5):

$$\min \left\{ \frac{\|\alpha_k - \alpha_{k-1}\|_2}{\max\{1, \|\alpha_k\|_2\}}, \gamma \frac{\|AW^\top \alpha_k - f\|_D - \|AW^\top \alpha_{k-1} - f\|_D}{\|AW^\top \alpha_k - f\|_D} \right\} < 5 \times 10^{-4},$$

where $\gamma = 1$ for image inpainting and $\gamma = 4$ for deblurring (in order to prevent the algorithm from stopping too early since for deblurring the residual $\|AW^\top \alpha_k - f\|_D$ usually changes slowly with k). To measure the quality of the restored image, we use the PSNR value defined by

$$\text{PSNR} := -20 \log_{10} \frac{\|u - \tilde{u}\|}{255n},$$

where u and \tilde{u} are the original and restored images respectively, and n is total number of pixels in u . Finally, we note that all algorithms are programmed in MATLAB and run on a 64-bits Windows laptop with a Quad Core 1.73GHz CPU and 8.0G RAM.

2.6.1. Balanced Approach for Image Inpainting. In Table 1, we compare the efficiency of the APG and the PFBS algorithms for image inpainting without noise, i.e. $A = \mathcal{P}_\Lambda$, $D = I$ and $\eta = 0$. Level of framelet decomposition is chosen to be 1 for optimal quality of restored images. One can see that the APG algorithm is more efficient than the PFBS algorithm while the quality of restored images is comparable. Some image results are shown in Figure 1.

Inpainting results using the two-system balanced approach (4.12) are also presented in Table 1. It is worth noticing that for images containing both cartoons and textures, e.g. “barbara512”, the two-system balanced approach (4.12) works better than the single-system balanced approach (4.8). Such improvement is shown in Figure 2, where one can see that textures are better restored using the two-system balanced approach (4.12), because in general, textures can be sparsely approximated by local cosine transforms.

Table 1: Comparisons of PFBS and APG algorithms: image inpainting without noise (i.e., $\eta = 0$ in (4.1)). Level of framelet decomposition is taken to be 1.

Inpainting	APG for (4.8)			PFBS for (4.8)			APG for (4.12)		
	$\kappa = 1, \lambda = 0.03$			$\kappa = 1, \lambda = 0.03$			$\kappa^i = 1, \lambda^i = 0.01$		
$\sigma = 0$	iter	psnr	time (s)	iter	psnr	time (s)	iter	psnr	time (s)
peppers256	19	39.38	3.07	59	38.09	9.17	29	39.09	6.11
goldhill256	21	35.17	3.34	60	34.79	10.51	30	35.34	6.47
boat256	22	33.79	3.42	64	33.64	11.30	29	33.78	6.34
camera256	20	32.95	3.23	62	32.62	9.67	29	33.23	7.04
bridge256	24	34.85	3.80	74	33.93	11.84	33	34.39	7.64
bowl256	19	36.93	2.97	59	36.45	9.11	35	38.54	7.90
barbara512	23	34.48	24.44	88	33.49	91.27	30	39.08	39.66
baboon512	24	31.68	25.28	73	30.98	76.28	32	31.65	41.61
fingerprint512	23	32.89	24.25	73	30.41	77.25	34	33.46	44.79
zebra512	23	34.38	24.42	67	33.42	69.18	33	34.91	42.44

Results of the synthesis based approach, i.e. (4.8) and (4.12) with $\kappa = 0$, are presented in Table 2. The level of framelet decomposition is chosen to be 2 for optimal quality of restored images. When comparing Table 2 with Table 1 one can see that the balanced approach usually produces better results than the synthesis based approach. This is due to the lack of smoothness for the synthesis based approach ($\kappa = 0$).

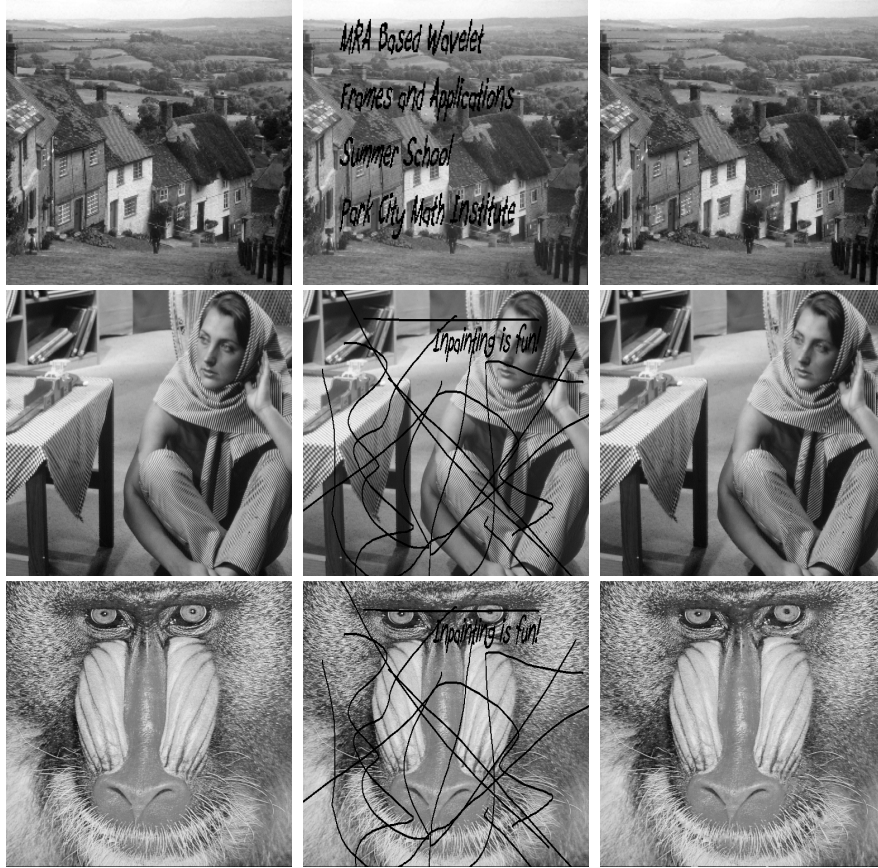


FIGURE 1. Inpainting results for “goldhill256” (first row), “barbara512” (second row) and “baboon512” (third row) using the single-system balanced approach (4.8) solved by APG algorithm (images taken from Table 1). Column 1-3 are the original image, observed image and restored image.

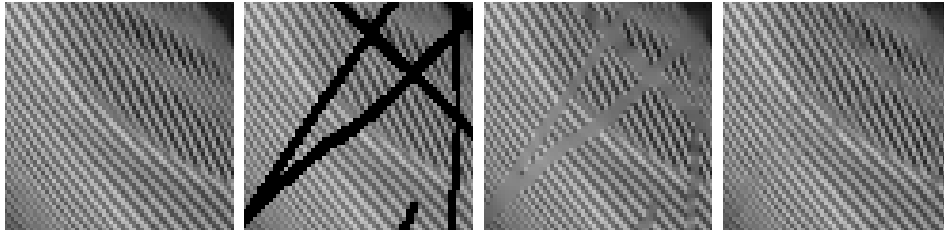


FIGURE 2. Comparison of image inpainting using the single-system balanced approach (4.8) and the two-system balanced approach (4.12). Images from left to right are: original image, observed image, and restored image using single- and two-system. The image presented is a zoom-in view of “barbara512” (taken from Table 1).

Table 2: Numerical results of APG algorithm in solving synthesis based method, i.e. (4.8) with $\kappa = 0$, for image inpainting without noise. Level of framelet decomposition is taken to be 2.

Inpainting	Single-System			Two-Systems		
	$\lambda = 0.03$			$\lambda_1 = \lambda_2 = 0.01$		
$\sigma = 0$	iter	psnr	time	iter	psnr	time
peppers256	51	39.00	13.51	83	39.36	26.02
goldhill256	62	33.45	16.71	78	34.40	24.42
boat256	61	32.93	16.65	29	33.52	9.22
camera256	67	31.27	17.82	29	32.53	9.39
bridge256	79	33.41	21.15	81	33.99	25.80
bowl256	59	35.78	15.98	29	37.35	9.22
barbara512	59	32.48	102.97	186	36.39	367.66
baboon512	81	29.86	143.44	29	30.36	57.35
fingerpint512	75	34.66	132.93	188	34.93	367.11
zebra512	80	33.90	141.98	84	34.95	165.97

2.6.2. *Image Deblurring.* Here, we present numerical results using the single- and two-system balanced approach (4.8) and (4.12) solved by the APG algorithm for image deblurring with noise. The matrix A in (4.1) is taken to be a convolution matrix with the corresponding kernel being a Gaussian function (generated in MATLAB by “`fspecial('gaussian',15,1.5);`” to be precise) and η is generated from a zero mean Gaussian distribution with variance $\sigma = 3$. Also, numerical results of image deblurring using the (preconditioned) linearized Bregman algorithm are presented.

For all the methods we used (i.e. the APG algorithm for the single- and two-system balanced approach, and the linearized Bregman algorithm), we choose $\theta = 0.01$ for the matrix $D = (AA^\top + \theta I)^{-1}$ which can be computed efficiently via fast Fourier transform (FFT). We choose the level of framelet decomposition to be 1 for the AGP algorithms and 4 for linearized Bregman algorithm for optimal quality of

the restored images. Finally, we note that for the linearized Bregman algorithm, we use the following stopping criterion:

$$\text{std}(AW^T \alpha_k - f) < \sigma.$$

In Table 3, we summarize the deblurring results for all of three algorithms, and the corresponding images are presented in Figure 3. Note that all three models produce comparable results in terms of PSNR values. However, the APG algorithm for both the single- and two-system balanced approaches are faster than the linearized Bregman algorithm. Also, we show a zoom-in view for “barbara” in Figure 4 where we can see that texture components of the image are best recovered by using the two-system balanced approach (4.12).

Table 3: Comparisons of APG algorithm for both single-system balanced approach (4.8) and two-system balanced approach (4.12), and linearized Bregman algorithm for (4.45).

deblurring	Single-System		Two-Systems		Linearized Bregman	
	$\kappa = 1, \lambda = 0.005$		$\kappa^2 = 1, \lambda^2 = 0.005$		$\kappa = 1, \lambda = 0.01$	
	psnr	time (s)	psnr	time (s)	psnr	time (s)
barbara	24.64	20.7	24.75	63.2	24.67	126.2
duck	31.15	17.5	31.04	41.8	30.87	100.1
wolf	30.83	25.9	30.6	72.6	30.66	186.5



FIGURE 3. Deblurring results for “barbara”, “duck” and “wolf”. Column 1-5 are: the original image, observed image, results of single-system, results of two-system and results of the linearized Bregman algorithm.

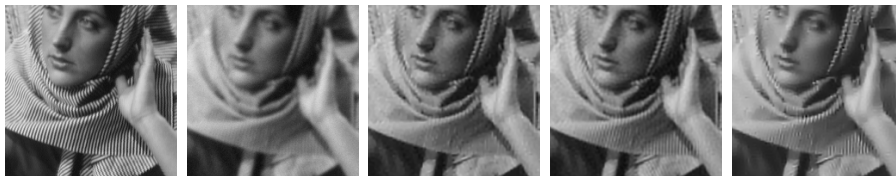


FIGURE 4. Zoom-in to the texture part of “barbara”. Images from left to right are: original image, observed image, results of single-system, results of two-system and results of linearized Bregman algorithm.

3. Analysis based approach

In this section, we present the split Bregman algorithm [96] for a more general analysis based approach:

$$(4.49) \quad \min_{u \in \mathbb{R}^n} H(u) + \|\text{diag}(\lambda)Wu\|_1,$$

where $H(\cdot)$ is some smooth convex function. In particular, when

$$H(u) = \frac{1}{2}\|Au - f\|_D^2,$$

(4.49) coincides with (4.11) given in the introduction. The optimization problem (4.49) can be solved by the split Bregman algorithm which was proposed by [96]. We provide a convergence analysis of the split Bregman algorithm following the proof given by [24].

3.1. Split bregman algorithm

The ℓ_1 term involved in (4.49) is nonsmooth and nonseparable. This prevents us from using optimization methods for smooth functions. A natural idea to solve (4.49) is to use a smoothed ℓ_1 norm to approximate the actual ℓ_1 norm and then apply optimization methods for smooth functions. This approach is commonly used in some early algorithms for PDE based, particularly for total variation (TV) based, variational models for image processing

$$\min_{u \in \mathbb{R}^n} H(u) + \lambda|\nabla u|,$$

(see e.g., [180, 40]). However, in these smoothing methods, the better the approximations (to the ℓ_1 norm) are, the slower the algorithms converge. Therefore, in order to make these algorithms converge faster, the smooth approximation to the ℓ_1 norm cannot be too accurate. Consequently, one cannot obtain a sparse solution by solving the minimization problem with a smoothed norm. However, sparsity is important in many cases of ℓ_1 regularization problems.

Another difficulty for the analysis based approach of (4.49) is that the term $\|\text{diag}(\lambda)Wu\|_1$ is not separable. Therefore, one cannot simply use soft thresholding as one normally does in the synthesis based approach, since it is impossible to keep any sequence in the range of W after applying a thresholding operator. An iterative method to find a solution in the range of W for a special case of (4.49), where $H(u) = \frac{1}{2}\|u - f\|^2$, was proposed by [33] by exploring duality.

As mentioned above, compared with the synthesis based approach, the ℓ_1 norm of $|Wu|$ involved in (4.49) is neither smooth nor separable. To overcome this, one

transfers (4.49) to a problem involving only separable nonsmooth terms. This is the main idea of split Bregman. In particular, one replaces the term Wu in (4.49) by a new variable d first and then adds a new constraint $d = Wu$ into (4.49). Hence, (4.49) becomes

$$(4.50) \quad \min_{u,d} H(u) + \|\text{diag}(\lambda)d\|_1 \quad \text{subject to} \quad d = Wu,$$

In order to solve (4.50), an iterative algorithm based on the Bregman distance with an inexact solver was proposed in [96]. This leads to the alternating split Bregman algorithm for (4.50). The split Bregman algorithm for (4.50) was first demonstrated in [96] to be an efficient tool for solving problems arising from TV norm minimization problems of PDE based models for image restoration, such as denoising.

The derivation of the split Bregman algorithm in [96, 24] is based on Bregman distance. The split Bregman method can also be understood as the augmented Lagrangian method (see e.g. [94]) applying to (4.50) (see e.g. [85, 171]). Here, we use the augmented Lagrangian method to derive the split Bregman algorithm, since this connects to the Uzawa's algorithm as we discussed for the linearized Bregman algorithm. Note that (4.50) is equivalent to the following optimization problem

$$(4.51) \quad \min_{u,d} H(u) + \|\text{diag}(\lambda)d\|_1 + \frac{\mu}{2}\|Wu - d\|_2^2 \quad \text{subject to} \quad d = Wu.$$

The Lagrangian for problem (4.51), which is called the augmented Lagrangian, is given as

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

The saddle points of $\mathcal{L}_\mu(u, d, p)$ can be obtained by the following iterative procedure

$$\begin{cases} (u_{k+1}, d_{k+1}) = \arg \min_{u,d} H(u) + \|\text{diag}(\lambda)d\|_1 + \langle p_k, d - Wu \rangle + \frac{\mu}{2}\|Wu - d\|_2^2, \\ p_{k+1} = p_k + \delta\mu(d_{k+1} - Wu_{k+1}), \end{cases}$$

which consists of one step of joint optimization of variables (u, d) followed by one step of gradient ascend of variable p . Now, letting $b_k = -p_k/\mu$, we have

$$\begin{cases} (u_{k+1}, d_{k+1}) = \arg \min_{u,d} H(u) + \|\text{diag}(\lambda)d\|_1 + \mu\langle b_k, Wu - d \rangle + \frac{\mu}{2}\|Wu - d\|_2^2, \\ b_{k+1} = b_k + \delta(Wu_{k+1} - d_{k+1}), \end{cases}$$

which is equivalent to

$$\begin{cases} (u_{k+1}, d_{k+1}) = \arg \min_{u,d} H(u) + \|\text{diag}(\lambda)d\|_1 + \frac{\mu}{2}\|Wu - d + b_k\|_2^2, \\ b_{k+1} = b_k + \delta(Wu_{k+1} - d_{k+1}). \end{cases}$$

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

$$(4.52) \quad \begin{cases} u_{k+1} = \arg \min_u H(u) + \frac{\mu}{2}\|Wu - d_k + b_k\|_2^2, \\ d_{k+1} = \arg \min_d \|\text{diag}(\lambda)d\|_1 + \frac{\mu}{2}\|d - Wu_{k+1} - b_k\|_2^2, \\ b_{k+1} = b_k + \delta(Wu_{k+1} - d_{k+1}). \end{cases}$$

Since $H(u)$ is convex and differentiable, the subproblem in the first line is easy to solve. Further, noting that the first term of the subproblem in the second line is the

ℓ_1 norm, the subproblem in the second line can be solved by a simple soft shrinkage. These features 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 has a relatively small memory footprint when compared to second order methods that require explicit representations of the Hessian matrix. Also, the method is easy to code. Both of these characteristics make this split Bregman method a practical algorithm for large scale problems.

For image restoration problems (4.1), we choose $H(u) = \frac{1}{2}\|Au - f\|_D^2$ and then (4.50) becomes

$$(4.53) \quad \min_u \frac{1}{2}\|Au - f\|_D^2 + \|\text{diag}(\lambda)d\|_1 \quad \text{subject to} \quad d = Wu,$$

which is equivalent to the analysis based approach (4.11). We summarize the split Bregman algorithm (4.52) for the analysis based approach in the following algorithm.

Algorithm 4.7. (*Split Bregman*)

- (i) Set initial guess d_0 and b_0 .
- (ii) For $k = 0, 1, \dots$, perform the following iterations until convergence

$$(4.54) \quad \begin{aligned} u_{k+1} &= (A^\top DA + \mu I)^{-1}(A^\top Df + \mu W^\top(d_k - b_k)), \\ d_{k+1} &= \mathcal{T}_{\lambda/\mu}(Wu_{k+1} + b_k), \\ b_{k+1} &= b_k + \delta(Wu_{k+1} - d_{k+1}). \end{aligned}$$

Note that when A is diagonalizable by discrete Fourier transform, e.g. a convolution matrix, the first equation of (4.54) can be solved efficiently by FFT.

3.2. Convergence analysis of split bregman algorithm

Now, we provide a convergence analysis of the split Bregman algorithm (4.52) for a general convex differentiable function $H(u)$. In particular, the convergence analysis applies to Algorithm 4.7 for image restoration problems. We note that, since all the subproblems involved in (4.52) are convex, the first order optimality condition gives

$$(4.55) \quad \begin{cases} 0 = \nabla H(u_{k+1}) + \mu W^\top(Wu_{k+1} - d_k + b_k), \\ 0 = p_{k+1} + \mu(d_{k+1} - Wu_{k+1} - b_k), \quad \text{with } p_{k+1} \in \partial\|\text{diag}(\lambda)d_{k+1}\|_1, \\ b_{k+1} = b_k + \delta(Wu_{k+1} - d_{k+1}). \end{cases}$$

This simple observation will be used in the proof of the convergence of the unconstrained split Bregman method (4.52). Note that we will not use the fact that $W^\top W = I$, because it is not required by the proof. Therefore, the following convergence analysis also applies when W is replaced by any other linear operator like ∇ (see [24] for details).

Theorem 4.8. *Assume that there exists at least one solution u^* of (4.49). Assume that $0 < \delta \leq 1$ and $\mu > 0$. Then, we have the following properties for the unconstrained split Bregman algorithm (4.52):*

$$(4.56) \quad \lim_{k \rightarrow +\infty} \|\text{diag}(\lambda)Wu_k\|_1 + H(u_k) = \|\text{diag}(\lambda)Wu^*\|_1 + H(u^*).$$

Furthermore,

$$(4.57) \quad \lim_{k \rightarrow +\infty} \|u_k - u^*\|_2 = 0$$

whenever (4.49) has a unique solution.

PROOF. Let u^* be an arbitrary solution of (4.49). By the first order optimality condition, u^* must satisfy

$$(4.58) \quad 0 = W^\top p^* + \nabla H(u^*),$$

where $p^* \in \partial \|\text{diag}(\lambda)d^*\|_1$ with $d^* = Wu^*$. Let

$$b^* = \frac{1}{\mu} p^*.$$

We obtain

$$(4.59) \quad \begin{cases} 0 = \nabla H(u^*) + \mu W^\top (Wu^* - d^* + b^*), \\ 0 = p^* + \mu(d^* - Wu^* - b^*), \quad \text{with } p^* \in \partial \|\text{diag}(\lambda)d^*\|_1, \\ b^* = b^* + \delta(Wu^* - d^*). \end{cases}$$

Therefore, (u^*, d^*, b^*) is a fixed point of (4.55). Consequently, if the iteration (4.52) converges, it converges to a solution of (4.49) when (4.56) is proven.

Denote the errors by

$$u_k^e = u_k - u^*, \quad d_k^e = d_k - d^*, \quad b_k^e = b_k - b^*.$$

Subtracting the first equation of (4.59) from the first equation of (4.55), we have

$$0 = \nabla H(u_{k+1}) - \nabla H(u^*) + \mu W^\top (Wu_{k+1}^e - d_k^e + b_k^e).$$

Taking the inner product of the left- and right- hand sides with respect to u_{k+1}^e , we have

$$(4.60) \quad 0 = \langle \nabla H(u_{k+1}) - \nabla H(u^*), u_{k+1}^e \rangle + \mu \|Wu_{k+1}^e\|_2^2 - \mu \langle W^\top d_k^e, u_{k+1}^e \rangle + \mu \langle W^\top b_k^e, u_{k+1}^e \rangle.$$

When similar manipulations are applied to the second equation of (4.55) and the second equation of (4.59), we obtain

$$(4.61) \quad 0 = \langle p_{k+1} - p^*, d_{k+1} - d^* \rangle + \mu \|d_{k+1}^e\|_2^2 - \mu \langle Wu_{k+1}^e, d_{k+1}^e \rangle - \mu \langle b_k^e, d_{k+1}^e \rangle,$$

where $p_{k+1} \in \partial \|\text{diag}(\lambda)d_{k+1}\|_1$ and $p^* = \mu b^* \in \partial \|\text{diag}(\lambda)d^*\|_1$. By summing (4.60) and (4.61), we get

$$(4.62) \quad \begin{aligned} 0 &= \langle \nabla H(u_{k+1}) - \nabla H(u^*), u_{k+1}^e \rangle + \langle p_{k+1} - p^*, d_{k+1} - d^* \rangle \\ &\quad + \mu (\|Wu_{k+1}^e\|_2^2 + \|d_{k+1}^e\|_2^2 - \langle Wu_{k+1}^e, d_{k+1}^e \rangle + \langle b_k^e, Wu_{k+1}^e - d_{k+1}^e \rangle). \end{aligned}$$

Now, subtracting the third equation of (4.59) from the third equation of (4.55) we obtain

$$b_{k+1}^e = b_k^e + \delta(Wu_{k+1}^e - d_{k+1}^e).$$

Taking the inner product of the above identity with itself and reorganizing the terms, we have

$$(4.63) \quad \langle b_k^e, Wu_{k+1}^e - d_{k+1}^e \rangle = \frac{1}{2\delta} (\|b_{k+1}^e\|_2^2 - \|b_k^e\|_2^2) - \frac{\delta}{2} \|Wu_{k+1}^e - d_{k+1}^e\|_2^2.$$

Substituting (4.63) into (4.62), we have

$$\begin{aligned}
(4.64) \quad & \frac{\mu}{2\delta} (\|b_k^e\|_2^2 - \|b_{k+1}^e\|_2^2) \\
& = \langle \nabla H(u_{k+1}) - \nabla H(u^*), u_{k+1}^e \rangle + \langle p_{k+1} - p^*, d_{k+1} - d^* \rangle \\
& \quad + \mu \left(\|Wu_{k+1}^e\|_2^2 + \|d_{k+1}^e\|_2^2 - \langle Wu_{k+1}^e, d_k^e + d_{k+1}^e \rangle - \frac{\delta}{2} \|Wu_{k+1}^e - d_{k+1}^e\|_2^2 \right) \\
& = \langle \nabla H(u_{k+1}) - \nabla H(u^*), u_{k+1}^e \rangle + \langle p_{k+1} - p^*, d_{k+1} - d^* \rangle \\
& \quad + \mu \left(\frac{1-\delta}{2} \|Wu_{k+1}^e - d_{k+1}^e\|_2^2 + \frac{1}{2} \|Wu_{k+1}^e\|_2^2 + \frac{1}{2} \|d_{k+1}^e\|_2^2 - \langle Wu_{k+1}^e, d_k^e \rangle \right) \\
& = \langle \nabla H(u_{k+1}) - \nabla H(u^*), u_{k+1}^e \rangle + \langle p_{k+1} - p^*, d_{k+1} - d^* \rangle \\
& \quad + \mu \left(\frac{1-\delta}{2} \|Wu_{k+1}^e - d_{k+1}^e\|_2^2 + \frac{1}{2} \|Wu_{k+1}^e - d_k^e\|_2^2 + \frac{1}{2} \|d_{k+1}^e\|_2^2 - \frac{1}{2} \|d_k^e\|_2^2 \right)
\end{aligned}$$

By summing the above equation from $k = 0$ to $k = K$, we get

$$\begin{aligned}
(4.65) \quad & \frac{\mu}{2\delta} (\|b_0^e\|_2^2 - \|b_{K+1}^e\|_2^2) + \frac{\mu}{2} \|d_0^e\|_2^2 \\
& = \sum_{k=0}^K \langle \nabla H(u_{k+1}) - \nabla H(u^*), u_{k+1} - u^* \rangle + \sum_{k=0}^K \langle p_{k+1} - p^*, d_{k+1} - d^* \rangle \\
& \quad + \mu \left(\frac{1-\delta}{2} \sum_{k=0}^K \|Wu_{k+1}^e - d_{k+1}^e\|_2^2 + \frac{1}{2} \sum_{k=0}^K \|Wu_{k+1}^e - d_k^e\|_2^2 + \frac{1}{2} \|d_{K+1}^e\|_2^2 \right).
\end{aligned}$$

By (4.4), and that $0 < \delta \leq 1$, all terms involved in the right-hand side of the above equation are nonnegative. This observation leads to the following inequality:

$$(4.66) \quad \frac{\mu}{2\delta} \|b_0^e\|_2^2 + \frac{\mu}{2} \|d_0^e\|_2^2 \geq \sum_{k=0}^K \langle \nabla H(u_{k+1}) - \nabla H(u^*), u_{k+1} - u^* \rangle.$$

By the assumption $\mu > 0$, we have $\sum_{k=0}^{+\infty} \langle \nabla H(u_{k+1}) - \nabla H(u^*), u_{k+1} - u^* \rangle < +\infty$, which leads to

$$(4.67) \quad \lim_{k \rightarrow +\infty} \langle \nabla H(u_k) - \nabla H(u^*), u_k - u^* \rangle = 0.$$

By (4.3), we have

$$H(u_k) - H(u^*) - \langle u_k - u^*, \nabla H(u^*) \rangle \geq 0,$$

and

$$H(u^*) - H(u_k) - \langle u^* - u_k, \nabla H(u_k) \rangle \geq 0,$$

which leads to

$$0 \leq H(u_k) - H(u^*) - \langle u_k - u^*, \nabla H(u^*) \rangle \leq \langle \nabla H(u_k) - \nabla H(u^*), u_k - u^* \rangle.$$

This, together with (4.67), leads to

$$(4.68) \quad \lim_{k \rightarrow +\infty} H(u_k) - H(u^*) - \langle u_k - u^*, \nabla H(u^*) \rangle = 0.$$

Similarly, (4.65) also provides us the following inequalities

$$\frac{\mu}{2\delta} \|b_0^e\|_2^2 + \frac{\mu}{2} \|d_0^e\|_2^2 \geq \sum_{k=0}^K \langle p_{k+1} - p^*, d_{k+1} - d^* \rangle,$$

and

$$\frac{\mu}{2\delta} \|b_0^e\|_2^2 + \frac{\mu}{2} \|d_0^e\|_2^2 \geq \frac{\mu}{2} \sum_{k=0}^K \|Wu_{k+1}^e - d_k^e\|_2^2.$$

Then by noticing that $Wu^* = d^*$ and following the same argument as above, we have

$$(4.69) \quad \lim_{k \rightarrow +\infty} \|\text{diag}(\lambda)d_k\|_1 - \|\text{diag}(\lambda)d^*\|_1 - \langle d_k - d^*, p^* \rangle = 0,$$

and

$$(4.70) \quad \lim_{k \rightarrow +\infty} \|Wu_{k+1} - d_k\|_2 = 0.$$

Since $\|\cdot\|_1$ is continuous, by (4.69) and (4.70), we obtain

$$(4.71) \quad \lim_{k \rightarrow +\infty} \|\text{diag}(\lambda)Wu_k\|_1 - \|\text{diag}(\lambda)Wu^*\|_1 - \langle Wu_k - Wu^*, p^* \rangle = 0.$$

Summing this and (4.68) yields

$$(4.72) \quad \lim_{k \rightarrow +\infty} \left((\|\text{diag}(\lambda)Wu_k\|_1 + H(u_k)) - (\|\text{diag}(\lambda)Wu^*\|_1 + H(u^*)) - \langle u_k - u^*, \nabla H(u^*) + W^T p^* \rangle \right) = 0.$$

This, together with (4.58), proves (4.56), and (4.57) follows from Proposition 4.1. \blacksquare

Remark 4.13. The proof of convergence still works if $H(u)$ is merely convex and continuous (then $H(u)$ is subdifferentiable). In that case, one can simply replace the first line of (4.55) by

$$(4.73) \quad 0 = q_{k+1} + \mu W^\top (Wu_{k+1} - d_k + b_k), \quad q_{k+1} \in \partial H(u_{k+1}),$$

and revise accordingly in the proof of Theorem 4.8. However, the equation (4.73) may be much harder to solve than the original one given in (4.55). For many applications, $H(u)$ corresponds to fidelity term and takes the form $H(u) = \frac{1}{2} \|Au - f\|_D^2$, which is differentiable. Then the first equation of (4.55) corresponds to a linear system of equations which can be solved efficiently whenever A is diagonalizable by discrete Fourier transform, e.g. when A is a convolution matrix.

3.3. Simulations

Now, we apply Algorithm 4.7 to image deblurring problems. The matrix A in (4.1) is taken to be a convolution matrix with the corresponding kernel being a Gaussian function (generated in MATLAB by “fspecial('gaussian',15,1.5);”) and η is generated from a zero mean Gaussian distribution with variance $\sigma = 3$. For the matrix $D = (AA^\top + \theta I)^{-1}$ we choose $\theta = 0.01$, and we pick the level of framelet decomposition to be 4. For stopping criterion, we use

$$\frac{\|Wu_{k+1} - d_{k+1}\|_2}{\|f\|_2} < 5 \times 10^{-4}.$$

Table 4 summarizes the deblurring results of Algorithm 4.7 and the corresponding images are shown in Figure 5. For convenience of comparison, we also present the results from Table 3 and Figure 3. As one can see, the single-system balanced approach outperform the analysis based approach in terms of both quality and speed. The two-system balanced approach is comparable to both Linearized Bregman and split Bregman algorithms, but it works better when images contain both cartoons and textures (see Figure 6).



FIGURE 5. Deblurring results for “barbara”, “duck” and “wolf”. Column 1-5 are: the original image, results of single-system, results of two-system, results of the linearized Bregman algorithm and results of the analysis based approach.

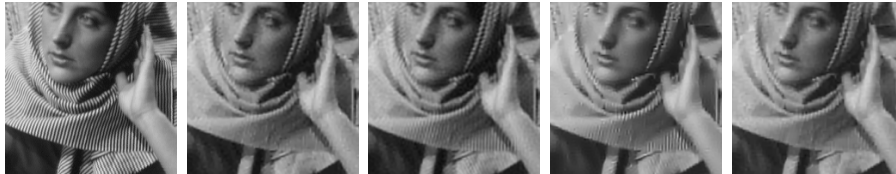


FIGURE 6. Zoom-in to texture component of “barbara”. Images from left to right are: the original image, results of single-system, results of two-system, results of linearized Bregman algorithm and results of the analysis based approach.

Table 4: Comparisons of APG algorithm for both single-system balanced approach (4.8) and two-system balanced approach (4.12), and linearized Bregman algorithm for (4.45).

Deblurring	Single-System		Two-Systems		Linearized Bregman		Analysis Approach	
$\sigma = 3$	$\kappa = 1, \lambda = 0.005$		$\kappa^2 = 1, \lambda^2 = 0.005$		$\kappa = 1, \lambda = 0.01$		$\mu = 0.1, \lambda = 0.05$	
	psnr	time (s)	psnr	time (s)	psnr	time (s)	psnr	time (s)
barbara	24.64	20.7	24.75	63.2	24.67	126.2	24.58	32.18
duck	31.15	17.5	31.04	41.8	30.87	100.1	31.00	26.57
wolf	30.83	25.9	30.6	72.6	30.66	186.5	30.75	81.40

Other applications of frames

1. Background and models

The previous lecture focused on using sparse approximations of wavelet frames to solve image restoration problems; however, these techniques have much broader applications. In this lecture, we consider some other interesting and important applications of tight frames. In Section 2, we discuss the model proposed in [18] on blind deblurring (motion deblurring to be specific) problems. Blind deblurring problems can still be modeled as (4.1), but the convolution kernel that determines A is unknown and needs to be solved simultaneously with u . In Section 3, we present a frame based image segmentation model of [71]. This model is no longer of the form (4.1) for image restoration problems. However, the concept of sparse approximation of tight frame systems still applies. More importantly, this model sheds some light on geometric interpretations of tight frames and grants links between tight frames and differential operators. In Section 4, we recall the model proposed by [112] on reconstruction of scenes (visible surfaces) from scattered, noisy and possibly sparse range data (point clouds).

Tight frames and the concept of sparse approximation of them have even more applications than described in this lecture. For example, researchers have recently applied frames to tomography (e.g. CT, PET and MRI). Medical images are similar to usual photographs (e.g. images of people, building or nature scenes) in the sense that they can be sparsely approximated by tight frames. Therefore, the basic idea of frame based tomography is to use the ℓ_1 -norm to penalize frame coefficients of the image that needs to be recovered with an appropriate fidelity term that models the specific signal acquisition technique (i.e. types of tomography) and type of noise. Also, applications of tight frames to geometric modeling have been developed.

1.1. Blind deconvolution

Motion blur caused by camera shake has been one of the prime causes of poor image quality in digital imaging, especially when using telephoto lenses or using long shutter speeds. In practice, using long shutter speeds is unavoidable when there is little light in the scene. Many researchers have worked on recovering clear images from motion-blurred images. The motion blur caused by camera shake is usually a spatially-invariant blurring process, which can be modeled as

$$(5.1) \quad f = u * p + \eta,$$

where $*$ is the convolution operator, u is the clear image to recover, f is the observed blurred image and p is the blur kernel (or *point spread function*), and η is the noise. If the blur kernel is given a priori, the model (5.1) is the same as (4.1), and the recovery of the clear image u is called a *non-blind deconvolution* problem; otherwise it is called a *blind deconvolution* problem.

It is known that the non-blind deconvolution problem is an ill-conditioned problem because of its high sensitivity to noise. Blind deconvolution is not only ill-conditioned but also ill-posed since both the blur kernel and the clear image are unknown. Thus the problem is highly under-determined as there are much more unknowns than available information. Since the relative motion between the camera and the scene can be arbitrary and usually hard to estimate a priori, motion deblurring is a typical blind deconvolution problem.

Early works on blind deblurring commonly use a single image and assume a parametric model of the blur kernel p , such that the blur kernel can be obtained by only estimating a few parameters (e.g. [149]). In practice, this is usually an oversimplification. In order to handle more complicated motion blurring, multi-image based approaches have been proposed to obtain more information about the blur kernel by either actively or passively capturing multiple images on the scene (e.g., [3, 5, 46, 153, 172]).

In recent years, there has been steady progress on removing motion blur from a single image for more general blur kernels. There are two typical approaches. One approach is to use some probabilistic priors on image's edge distribution to derive the blur kernel (e.g., [89, 119, 127]) or to manually select blurred edges to obtain the local blur kernel ([113]). The main issue with this type of method is that the assumed probabilistic priors do not always hold true for general images, especially for nature images of complex structures. The second approach is to formulate the blind deconvolution as a joint minimization problem with a certain regularization on both the blur kernel p and the clear image u :

$$(5.2) \quad E(u, p) = \min_{p, g} \Phi(u * p - f) + \lambda_1 \Psi_1(p) + \lambda_2 \Psi_2(u),$$

where $\Phi(p * u - f)$ is the fidelity term, $\Psi_1(p)$ and $\Psi_2(u)$ is the regularization term on the kernel and the clear image respectively.

In [2, 44, 49] TV-based approaches were proposed to solve general blind deconvolution problems, where the fidelity term in (5.2) is the usual ℓ_2 norm on image intensities; and both regularization terms in (5.2) are total variation measurement (ℓ_1 norm on function derivatives). In [166] a more sophisticated minimization model was presented where the fidelity term is a weighted ℓ_2 norm on both similarities in image intensity and similarities in the image's derivatives; and the regularization terms are the ℓ_1 norm of the blur kernel's derivatives and the ℓ_1 norm of the image intensities. This method demonstrated impressive performance on recovering motion-blurred images containing moderately complex structures. However, the minimization model proposed in [166] is quite complicated and requires a fairly accurate prior on the size of the blur kernel. Moreover, the recovered images from this method, as well as other TV-based methods, were lacking sufficient details and textures on the regions of complex structure, which limited their applications in practice.

In Section 2, we follow the tight frame based blind deconvolution model of [18]. This model takes into account both the sparsity of the image u and that of the kernel p . To be precise, both the image and the kernel are assumed to have sparse representation under two different quasi-affine tight frame systems. A fast algorithm is provided in [18] based on the linearized Bregman algorithm. Here, we recall the general model of [18] and leave the details to Section 2. The tight frame based blind deconvolution model can be found in [19, 20] as well.

Given a blurred image f , our goal is to recover the clear image u and the blur kernel p from (5.1):

$$f = u * p + \eta.$$

Let α denote the framelet coefficients of the clear image u with $u = W_1^\top \alpha$, and let β denote the framelet coefficients of the blur kernel p with $p = W_2^\top \beta$. Then,

$$f = (W_1^\top \alpha) * (W_2^\top \beta) + \eta.$$

Thus, we formulate the *frame based blind motion deblurring model* as:

$$\min_{\alpha, \beta} \|\alpha\|_1 + \lambda \|\beta\|_1$$

subject to

$$\|(W_1^\top \alpha) * (W_2^\top \beta) - f\|_2 \leq \sigma,$$

where W_i , $i = 1, 2$, are the decomposition operators of framelet systems, λ is the parameter which balances the sparseness between the kernel and the image, and σ is the noise parameter of the observed image f . The solution of this minimization is approximated by applying the linearized Bregman algorithm alternatively as shown in Section 2 with full details.

1.2. Segmentation

As we mentioned before, tight frames give redundant representations to signals and images, and the redundancy of tight frames usually leads to sparse approximation of images, which is known to be a desirable property for image restoration problems, like denoising, inpainting, deblurring, etc. Also, there is some research on texture classification and segmentation using wavelets or wavelet frames [178, 1]. However, utilizing the property of sparse approximation of tight frames for general image segmentation problems has not been considered in the literature. In Section 3 we recall a frame based segmentation model by [71].

The segmentation model of [71] is applicable to general images. However, we focus on medical images. Segmenting biological structures, e.g. cortical or subcortical structures, blood vessels, tumors etc., from various types of medical images (e.g. CT, MRI, ultrasound, etc.) is very important for detecting abnormalities, studying and tracking progress of diseases, and surgery planning. Medical image segmentation is a difficult problem due to medical images commonly having poor contrasts, different types of noise, and missing or diffusive boundaries. Numerous algorithms have been developed in the literature targeting either general segmentation problems or the segmentation of specific biological structures (see [48, 133, 126, 176, 183, 128, 179, 150, 174, 92, 139, 122] and the references therein).

Now, we give a brief preview of the frame based segmentation model of [71] and leave the details to Section 3. Since it is confusing to denote images as vectors in \mathbb{R}^n for segmentation problems, we shall understand image f as s -dimensional arrays, i.e. $f \in \mathbb{R}^{N_1 \times N_2 \times \dots \times N_s} =: \mathcal{S}_s$, with s typically 2 or 3. We still denote the fast framelet decomposition and reconstruction as W and W^\top , while they should be understood as linear operators, instead of matrices, defined precisely as in Algorithm 2.3. We note that we can use the same notation for image restoration problems because, mathematically speaking, it makes no difference.

For a given observed image $f \in \mathcal{S}_s$ (Figure 1(a)), a typical objective of segmentation is to divide the image domain into two sub-domains: one is the domain of the object of interest Ω (white region in Figure 1(b)) and the other is the background

Ω^c (black region in Figure 1(b)). The quality of segmentation is usually measured by: (1), how well the boundary of Ω (Figure 1(c)), denoted as $\partial\Omega$, fits the boundary of the object of interest in f ; (2), the smoothness of $\partial\Omega$. Here (2) is an important requirement when noise or other types of error are present in the image f .

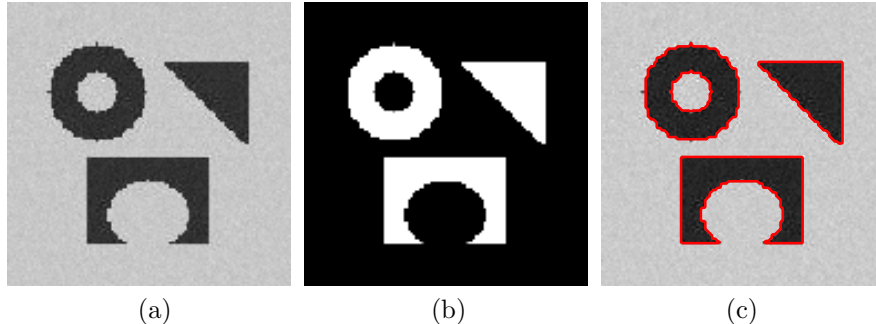


FIGURE 1. Figure (a) shows an observed image. Figure (b) shows the domain decomposition based on some segmentation where the white regions indicate the object of interest while the black regions indicate the background. Figure (c) shows the boundary of the segmented object (red curves).

One of the most popular ways of modeling segmentation problems is to use level set formulations. The key idea of the level set based segmentation models (see e.g. [41, 146, 148, 164]) is to find an appropriate level set function $u \in \mathcal{S}_s$ such that $\Omega := \{u \geq \alpha\}$ provides a desired segmentation of the given image f for some α ($\alpha = 0$ is commonly used). The beauty of the level set formulation is that it models general domains (not necessarily connected) with complicated boundaries, and the quality of segmentation is easily controlled by adding appropriate penalties to the level set function u .

The frame based segmentation model of [71] combines the property of sparse approximation of tight frame systems to piecewise smooth functions with the idea of the level set method. We now illustrate how the frame based segmentation model of [71] is motivated. The idea is motivated by that of [43, 45, 12] where TV-based segmentation models are considered.

Let χ_Ω denotes the characteristic function supporting on domain Ω . A desired segmentation of a given image $f \in \mathcal{S}_s$ should be a solution to the following non-convex optimization problem

$$(5.3) \quad \min_{\Omega; c^1, c^2 \in \mathbb{R}} \|\lambda \cdot W\chi_\Omega\|_1 + \mu \left(\|\chi_\Omega(c^1 - f)\|_2^2 + \|\chi_{\Omega^c}(c^2 - f)\|_2^2 \right).$$

Remark 5.14.

- (1) The first term of (5.3) controls the regularity of $\partial\Omega$. The frame transform operator W can be understood as a certain discretization of some differential operator. For example in 2D, when W corresponds to the piecewise linear wavelet, W contains difference operators that approximate $\partial_x u$ and $\partial_y u$ with the corresponding 2D filters $[1, 2, 1]^\top \cdot [1, 0, -1]$ and $[1, 0, -1]^\top \cdot [1, 2, 1]$; also it contains difference operators that approximate $\partial_{xx} u$ and $\partial_{yy} u$ with the corresponding 2D filters $[1, 2, 1]^\top \cdot [-1, 2, -1]$

and $[-1, 2, -1]^\top \cdot [1, 2, 1]$. This observation indicates that with different choices of parameter λ and tight frames, the frame transform operator W approximates different differential operators. Therefore, the first term of (5.3) provides rather rich geometric information of $\partial\Omega$.

- (2) The second term of (5.3) is the fidelity term that controls the accuracy of the segmentation. The two constants c^1 and c^2 approximate the average values of f inside and outside of Ω respectively. In general, one can consider c^1 and c^2 as arrays in \mathcal{S}_s that approximate f itself inside and outside of Ω . However, for simplicity, we will only consider c^1 and c^2 as constants here.
- (3) Note that when $W = \nabla$, the model (5.3) is known as the piecewise constant Mumford-Shah model studied by [43], where the authors employed the level set method and used gradient descent to solve the energy (5.3).

The major difficulty of solving (5.3) is that it is a nonconvex optimization problem which in general is hard to solve. Here, we describe a way that one can partially convexify the problem (5.3) following a similar idea as [45, 12].

If one replaces χ_Ω by u and χ_{Ω^c} by $1 - u$ in (5.3), it is easy to rewrite the problem (5.3) equivalently as

$$(5.4) \quad \min_{u \in \{0,1\}; c^1, c^2 \in \mathbb{R}} \|\lambda \cdot Wu\|_1 + \mu \langle \zeta(c^1, c^2), u \rangle,$$

where $u \in \{0, 1\}$ forces u to be a binary array, and $\zeta(c^1, c^2)$ is defined as

$$\zeta(c^1, c^2) := (c^1 - f)^2 - (c^2 - f)^2.$$

The problem (5.4) is equivalent to (5.3) in the sense that χ_Ω is a solution to (5.4) if and only if Ω is a solution to (5.3), for any given c^1 and c^2 . Note that the objective function of (5.4) is now convex with respect to u , while the non-convexity is moved to the constraints $u \in \{0, 1\}$, which are usually referred to as binary constraints.

Optimization problems with binary constraints are usually very hard to solve due to the existence of local minimum. One common way of relaxing binary constraints like $u \in \{0, 1\}$ is to use $0 \leq u \leq 1$ instead. This leads to the *frame based segmentation model* of [71] given as follows:

$$(5.5) \quad \min_{0 \leq u \leq 1; c^1, c^2 \in \mathbb{R}} \|\lambda \cdot Wu\|_1 + \mu \langle \zeta(c^1, c^2), u \rangle.$$

Note that (5.5) is still a non-convex problem; however, with fixed c^1 and c^2 it is a convex problem.

When c^1 and c^2 are fixed, an optimal solution u^* of (5.5) is generally not a solution of (5.4), unless u^* happens to be binary itself. When u^* is not binary, a segmentation of image f can be obtained by taking $\Omega_\alpha := \{u^* \geq \alpha\}$ for some $\alpha \in [0, 1]$. Now the question is, for which α , is the function χ_{Ω_α} a solution to (5.4)? For the case $W = \nabla$, this question is answered by [45, 12], where they showed that for almost every $\alpha \in [0, 1]$, χ_{Ω_α} is a solution to (5.4) (when $W = \nabla$, the model (5.4) is known as the piecewise Mumford-Shah model first considered by [43]). Although there is no corresponding theory for general tight frames where $W \neq \nabla$, our numerical simulations indicate that a similar result holds. Note that for different α , the set Ω_α is generally different. This raises the question of which choice of α results in Ω_α providing the best segmentation result. Recall from the theory of linear programming that an optimizer of a linear program must lie on one of the extreme points of the feasible set (see e.g. [11]). Therefore, when μ

is comparably large, the problem (5.5) is close to a linear program, and thus u^* is close to a binary array (which is confirmed by numerical experiments). As a consequence, choices of α are not crucial and we choose $\alpha = 0.5$.

1.3. Scene reconstruction from range data

High-quality 3D scene modeling has long been an important research topic in computer vision, robotic navigation, computer graphics and animation. The 3D geometrical model of a scene is usually reconstructed from pre-acquired range data sets of the scene, which are usually very noisy, corrupted with a lot of outliers, and sparsely sampled with large gaps. In the past, there have been extensive studies on reconstructing objects or scenes using range data from a single view or from multiple views. Interested readers are referred to a recent survey [100, Chapter 4] for more details. In Section 4, we focus on how to reconstruct the scene model using 3D range data of a single view. Since range data from a single view can only provide 3D information for the visible surfaces of the scene, a piecewise smooth explicit surface model is usually adequate to describe visible surfaces of a scene. It is noted that discontinuities on the reconstructed piecewise smooth surface provide very important information for many applications (e.g. robotics), because surface discontinuities represent either the boundaries of objects in the scene or sharp geometrical changes of individual objects. Thus, in contrast to what most traditional surface-fitting methods tend to do, a desired reconstruction algorithm should well preserve surface discontinuities.

Reconstructing a piecewise smooth surface using range data can be formulated as a function reconstruction problem. Assume that we are given a set of scattered data sites:

$$\Xi = \{x[1], x[2], \dots, x[n]\} \subset \mathbb{R}^2$$

and associated function values

$$f|_{\Xi} = \{f[1], f[2], \dots, f[n]\},$$

where f_i is the function value of an unknown data function $f(x)$ at $x[i]$ and possibly contains noise, i.e.,

$$f[i] = f(x[i]) + \epsilon[i].$$

Our goal is then to reconstruct the data function f under the assumption that f is a piecewise smooth function. It is emphasized that the discontinuities of f need to be well preserved in the reconstruction, because they encode important information about boundaries of objects on which many high-level tasks depend. On the other hand, the input data sites Ξ can be scattered such that they are non-uniformly sampled, with large gaps. Furthermore, the obtained function values $f|_{\Xi}$ are usually corrupted by different types of noise, e.g. Gaussian or impulsive noise. All of these challenges make reconstruction a rather difficult problem.

The frame based model of [112] is motivated by work in [107, 117, 169] which use a simple principal shift invariant space and its associated wavelet transform to fit scattered data, [118] proposed an efficient algorithm to approximate the solution of a smoothing spline in a principal shift invariant space.

Let $\Omega \subset \mathbb{R}^2$ be a bounded domain of interest that contains all data, and let ϕ be a carefully chosen compactly supported continuous function (e.g. uniform B-splines, box splines, nodal basis functions). We look for fitting functions in the

space spanned by 2^{-L} -dilates and 2^{-L} -shifts of ϕ whose support intersects Ω , i.e.,

$$S^L(\phi, \Omega) = \left\{ \sum_{k \in \mathbb{Z}^2} c(k) \phi(2^L \cdot -k) : c(k) = 0 \text{ whenever } \text{supp } \phi(2^L \cdot -k) \cap \Omega = \emptyset \right\},$$

where $L \in \mathbb{Z}^+$ is a scaling parameter that controls the refinement of the space. Then, any fitting function $s(x) \in S^L(\phi, \Omega)$ can be written as a finite expansion:

$$(5.6) \quad s(x) = \sum_{k \in \mathcal{I}} u[k] \phi(2^L x - k),$$

where $\mathcal{I} := \{k \in \mathbb{Z}^2 : \text{supp } \phi(2^L \cdot -k) \cap \Omega \neq \emptyset\}$. Let \mathbf{u} and \mathbf{f} denote the column vector $\{u[k]\}_{k \in \mathcal{I}}$ and $\{f[i]\}_{1 \leq i \leq n}$ respectively. Our goal here is to find \mathbf{u} such that $s(x)$ approximates the unknown function $f(x)$ based on the observed data \mathbf{f} . The function $f(x)$ that we are approximating is generally piecewise smooth and can be sparsely approximated by tight frame systems. Therefore, the key idea of [112] is to reconstruct $s(x)$ such that $s(x)$ can be sparsely approximated by certain tight frame systems, which naturally leads to the penalization of the ℓ_1 -norm of tight frame coefficients of $s(x)$ as shown in the previous lecture. Therefore, we will obtain a desired \mathbf{u} by solving the following problem

$$\min_{\mathbf{u}} \frac{1}{2} \|\mathbf{A}\mathbf{u} - \mathbf{f}\|_2^2 + \|\text{diag}(\lambda)\mathbf{W}\mathbf{u}\|_1,$$

where

$$A(i, k) = \phi(2^L x[i] - k), \quad i = 1, 2, \dots, n, \quad k \in \mathcal{I},$$

for some given $L \in \mathbb{Z}^+$. We shall now leave the details to Section 4.

2. Frame based blind deconvolution

As we see from (5.1), blind deblurring is an under-constrained problem with many possible solutions. Thus we need to enforce extra constraints on the image and the kernel to overcome the ambiguity. In past years, there have been extensive studies showing that most natural images usually have a sparse approximation under some tight frame systems, which include, for example, shift-invariant orthonormal wavelets, Gabor transform, framelet, curvelet, etc. This inspires the authors in [18] to use the sparseness of the image under the suitable tight frame system as the constraint on the image. The same sparseness constraint is also applicable to the kernel if we understand the kernel as a special image.

2.1. Problem formulation and analysis

Given a blurred image f , our goal is to recover the clear image u and the blur kernel p from (5.1):

$$f = u * p + \eta.$$

Let W_1 and W_2 denote the fast framelet decomposition corresponding to two possibly different tight frame systems. Denote α and β the framelet coefficients of the clear image u with $u = W_1^\top \alpha$ and the blur kernel p with $p = W_2^\top \beta$. Then,

$$f = (W_1^\top \alpha) * (W_2^\top \beta) + \eta.$$

Thus, we can formulate the blind motion deblurring problem as:

$$(5.7) \quad E(\alpha, \beta) = \min_{\alpha, \beta} \|\alpha\|_1 + \lambda \|\beta\|_1$$

subject to

$$(5.8) \quad \|(W_1^\top \alpha) * (W_2^\top \beta) - f\|_2 \leq \sigma,$$

where λ is the parameter which balances the sparseness between the kernel and the image, σ is the noise parameter of the observed image f .

It is known that the ℓ_1 norm is a fairly good measurement of the sparseness of vectors. Thus, the first and second term in (5.7) penalize the sparseness of framelet coefficients of the image u and kernel p under two tight frame systems. Since u and p are usually two different types of images, the tight frame systems that can provide an ideally sparse approximation to them should be different. The constraint in (5.8) is the fidelity constraint between the observed and computed image. The minimization problem (5.7-5.8) is a challenging one. In the next section, we present an alternating minimization approach to solve this minimization problem efficiently. The key idea is to adopt a modified version of the linearized Bregman algorithm.

2.2. Numerical algorithm and analysis

The minimization (5.7-5.8) is a joint minimization problem on α and β . It may not always yield a physically meaningful solution because (5.7-5.8) does not have a unique solution in general. In order to obtain a physically meaningful solution, we chose to impose the following physical conditions:

$$(5.9) \quad \begin{cases} p = W_2^\top \beta \geq 0, & \text{and } \sum p = 1; \\ u = W_1^\top \alpha \geq 0. \end{cases}$$

Even with the constraint (5.9), the problem (5.7)-(5.8) is not easy to solve. Here, we take the traditional heuristic alternative approach. The outline of the alternative algorithm for solving (5.7)-(5.8) is given as follows:

Algorithm 5.1. Let $\alpha_0 = W_1 f$ and $\beta_0 = W_2 \delta_c$ be initial guess, where f is the observed image and δ_c takes value one at the center pixel of the image and zero elsewhere. Iterate on k until convergence.

- (1) Fixing β_k , solve (5.7-5.8) with respect to α , i.e., set $\alpha_{k+\frac{1}{2}}$ be a solution of

$$(5.10) \quad \min_{\alpha} \|\alpha\|_1 \quad \text{subject to} \\ \|(W_2^\top \beta_k) * (W_1^\top \alpha) - f\|_2 \leq \sigma,$$

Then impose

$$\alpha_{k+1} = W_1 u_{k+1},$$

where

$$u_{k+1}[j] = \begin{cases} W_1^\top \alpha_{k+\frac{1}{2}}[j], & \text{if } W_1^\top \alpha_{k+\frac{1}{2}}[j] \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

- (2) Fixing α_{k+1} , solve (5.7-5.8) with respect to β , i.e., set $\beta_{k+\frac{1}{2}}$, be a solution of

$$(5.11) \quad \min_{\beta} \|\beta\|_1 \quad \text{subject to} \\ \|(W_1^\top \alpha_{k+1}) * (W_2^\top \beta) - f\|_2 \leq \sigma.$$

Then impose

$$\beta_{k+1} = W_2 h_{k+1},$$

where

$$h_{k+1}[j] = \begin{cases} W_2^\top \beta_{k+\frac{1}{2}}[j], & \text{if } W_2^\top \beta_{k+\frac{1}{2}}[j] \geq 0, \\ 0, & \text{otherwise,} \end{cases}$$

followed by normalization $h_{k+1} = \frac{h_{k+1}}{\|h_{k+1}\|_1}$.

In each step of the alternative algorithm, there are two minimization problems (5.10) and (5.11), which in fact take exactly the same form. In the following, we present an efficient algorithm to solve such minimization problems, which is derived from the linearized Bregman algorithm (Algorithm 4.6). For (5.10), let $[p]_*$ denote the matrix form of the convolution operator corresponding to the kernel p . Define

$$w_0 = x_0 = 0.$$

Then we generate a sequence of w_i and x_i as follows

$$(5.12) \quad \begin{cases} w_{i+1} = \nu \mathcal{T}_\mu(x_i), \\ x_{i+1} = x_i - Z_k W_1 [W_2^\top \beta_k]_*^\top ([W_2^\top \beta_k]_* W_1^\top w_{i+1} - f), \end{cases}$$

where \mathcal{T}_μ is the soft-thresholding operator defined by (4.15) and (4.16), and $Z^{(k)}$ is a preconditioning matrix to accelerate the convergence of the iteration, which is chosen in the implementation as

$$Z_k = ([W_2^\top \beta_k]_*^\top ([W_2^\top \beta_k]_* + \lambda_i \Delta))^{-1},$$

where Δ is the discrete Laplacian. The convergence of the sequence $w^{(i)}$ is guaranteed by the following proposition which is a direct consequence of Theorem 4.7. It is proved in [22] that the sequence $w^{(i)}$ generated by (5.12) actually converges to an approximated solution of (5.10) when $\nu\mu \rightarrow \infty$.

Proposition 5.1. *The sequence w_i generated via (5.12) with a proper ν converges to the unique solution of*

$$(5.13) \quad \begin{cases} \min_\alpha \|\alpha\|_1 + \frac{1}{2\nu\mu} \|\alpha\|_2^2, \\ \text{s.t. } (W_2^\top \beta_k) * (W_1^\top \alpha) = f, \end{cases}$$

if there exists at least one solution of $(W_2^\top \beta_k) * (W_1^\top \alpha) = f$.

PROOF. It is a direct consequence of Theorem 4.7 by noticing that the constraint $(W_2^\top \beta_k) * (W_1^\top \alpha) = f$ is equivalent to $Z_k^{\frac{1}{2}} (W_2^\top \beta_k) * (W_1^\top \alpha) = Z_k^{\frac{1}{2}} f$, and Z_k commutes with convolution operators. ■

The method (5.12) is extremely efficient. Usually it takes only a few iterations for (5.12) to get an approximate solution of (5.10) (cf. [22]), and the accuracy of the approximation is adequate during the alternative iterations in the main algorithm.

For (5.11), the same method can be applied to find the solution: start with $z_i = x_i = 0$, and generate the sequence

$$(5.14) \quad \begin{cases} z_{i+1} = \nu \mathcal{T}_\mu(x_i), \\ x_{i+1} = x_i - Z_k W_2 [W_1^\top \alpha_k]_*^\top ([W_1^\top \alpha_k]_* W_2^\top z_{i+1} - f). \end{cases}$$

Then after a few iterations, z_i becomes a good approximated solution to (5.11).

2.3. Numerical experiments

Here, we present deblurring result using Algorithm 5.1 for one image with synthetic motion blur and one real image that was taken by a digital camera using a long shutter speed under a weak light condition. The synthetic motion blur is generated by MATLAB function “fspecial(‘motion’,11,120)” and noise with $\sigma = 3$ is also added to the image. For simplicity, we choose $W_1 = W_2$ as the decomposition matrix corresponding to piecewise linear quasi-affine tight frame system. Numerical results of the two images are given in Figure 2 and 3. For more numerical experiments, one should consult [18].



FIGURE 2. Deblurring results for the synthetic motion blur. Column 1-3 are: the original image, observed blurry and noisy image, and deblurring result of Algorithm 5.1. Row 2 and 3 show two different close-ups of the corresponding images shown in row 1.

3. Frame based image segmentation

In this section, we review the frame based segmentation model proposed by [71]. First, we introduce a few notations that will be used throughout this section.

An observed s -dimensional image f is an array in $\mathcal{S}_s := \mathbb{R}^{N_1 \times \dots \times N_s}$. Framelet decomposition and reconstruction operators are denoted as W and W^\top which are given by Algorithm 2.3 and satisfy $W^\top W = I$ with I the identity operator. Here, we use the tensor product quasi-affine tight frame systems discussed in Lecture 2. Suppose there are r univariate tight framelets constructed from UEP. The total



FIGURE 3. Deblurring results for the real image. For each row, the left image is the observed blurry and noisy image, and the right image is the deblurred image using Algorithm 5.1. Row 2 shows a close-up of the corresponding images shown in row 1.

number of s -dimensional tight framelets obtained by tensor product is $r_s := (r + 1)^s - 1$.

For simplicity of notation, we suppose the level of framelet decomposition is 1. We denote W_0 as the operator that corresponds to the decomposition with respect to the refinable function; and denote W_i , with $1 \leq i \leq r_s$, as the operator that corresponds to the decomposition with respect to the i -th framelet. Under this notation, we denote $Wu \in \mathcal{S}_s \times \mathbb{R}^{r_s+1}$ as

$$Wu := [W_0u, W_1u, W_2u, \dots, W_{r_s}u].$$

3.1. Frame based segmentation model

For a given image $f \in \mathcal{S}_s$, we consider the following optimization problem

$$(5.15) \quad \min_{0 \leq u \leq 1; c^1, c^2 \in \mathbb{R}} \|\lambda \cdot Wu\|_1 + \mu \langle \zeta(c^1, c^2), u \rangle,$$

where $\zeta(c^1, c^2)$ is defined as

$$\zeta(c^1, c^2) := (c^1 - f)^2 - (c^2 - f)^2$$

with c^1 and c^2 real constants. Here $\lambda \in \mathcal{S}_s \times \mathbb{R}^{r_s+1}$ is a weight function that weights W_iu differently for different i in general. However, we choose the following λ for simplicity

$$\lambda[\cdot, i] = \begin{cases} 0, & i = 0; \\ \frac{1}{1 + \sigma \sum_{j=1}^{r_s} |W_j f|^2}, & i = 1, 2, \dots, r_s. \end{cases}$$

Notice that λ can be regarded as the so-called edge indicator function (see e.g. [28, 45, 12]) under the framelet transform W .

To solve (5.15), one can alternatively optimize variables u and c^i , $i = 1, 2$. Since when u is fixed, the optimal values c^i can be easily determined. Therefore, the key step is to optimize (5.15) with c^i fixed. Here, we adopt the idea of the split Bregman algorithm that was discussed in Section 3.

Let $d = Wu$ and write (5.15) equivalently as the following problem

$$\min_{d=Wu} \|\lambda \cdot d\|_1 + \mu \langle \zeta(c^1, c^2), u \rangle + \mathcal{C}_{0 \leq u \leq 1},$$

where $\mathcal{C}_{0 \leq u \leq 1}$ is the indicator function w.r.t. the set $\{0 \leq u \leq 1\}$. Then, from (4.52) in Section 3, we have the following iterative scheme

$$(5.16) \quad \begin{aligned} u_{k+1} &= \operatorname{argmin}_u \mu \langle \zeta(c^1, c^2), u \rangle + \mathcal{C}_{0 \leq u \leq 1} + \frac{\nu}{2} \|Wu - d_k + b_k\|_2^2 \\ d_{k+1} &= \operatorname{argmin}_d \|\lambda \cdot d\|_1 + \frac{\nu}{2} \|Wu_{k+1} - d + b_k\|_2^2 \\ b_{k+1} &= b_k + (Wu_{k+1} - d_{k+1}). \end{aligned}$$

This, together with the optimization of c^1 and c^2 , leads to the following algorithm for (5.15):

Algorithm 5.2. *Given image f .*

- (1) Start with $u_0 = 0$, $c^1 = \min f$, and $c^2 = \max f$.
- (2) For a given $\alpha \in [0, 1]$, iterate until convergence:

$$(5.17) \quad \begin{aligned} u_{k+\frac{1}{2}} &= W^\top (d_k - b_k) - \frac{\mu}{\nu} \zeta(c_k^1, c_k^2) \\ u_{k+1} &= \max\{\min\{u_{k+\frac{1}{2}}, 1\}, 0\} \\ d_{k+1} &= \mathcal{T}_{\lambda/\nu}(Wu_{k+1} + b_k) \\ b_{k+1} &= b_k + (Wu_{k+1} - d_{k+1}) \\ c_{k+1}^1 &= M(f, \Omega_{k+1}), \quad c_{k+1}^2 = M(f, (\Omega_{k+1})^c), \quad \Omega_{k+1} = \{u_{k+1} > \alpha\}, \end{aligned}$$

where $M(f, \Omega)$ returns the mean value of f within domain Ω .

- (3) Let u^* be the output of step (2). Then the segmentation of f is given by $\Omega^* := \{u^* \geq \alpha\}$.

Note that Algorithm 5.2 is very efficient. For each iteration k , the most time consuming operation is performing the fast framelet decomposition and reconstruction, which are of the same complexity as fast Fourier transform (FFT) (see Lecture 2, Section 1.4 and 3). Furthermore, numerical experiments show that we usually only need a few hundred iterations until the algorithm converges to a given tolerance.

3.2. Numerical results

In this section, we compare the frame based segmentation model (5.15) with the TV-based segmentation model [45, 12] using a 2D-slice of an MRI image and a 3D CT angiography (CTA) image. For more results of 3D CTA image segmentation, please see [71].

In our implementation, we adopt the stopping criterion: $\|b^{k+1}\| < 10^{-3}$. Based on this stopping criterion, the number of iterations for 2D and 3D cases varies from 100 to 500. Within each iteration, the comparably expensive operation is the framelet decomposition and reconstruction, i.e. W and W^\top . Although the complexity of applying W and W^\top is of the same order as FFT, by applying the fast algorithm of [65], in practice the constants really matter. We note, however, that

this constant is not big, and hence the framelet decomposition and reconstruction can be done rather efficiently. For example, for a 3D image of size $50 \times 50 \times 50$, the computational time for one level of framelet decomposition and reconstruction is approximately 5–6 times slower than the forward and inverse FFT. This comparison is done using MATLAB2007. Throughout this section, the parameter α in (5.17) is chosen to be 0.5, and the level of framelet decomposition L is chosen to be 1 for the 2D image and 2 for the 3D image for the best quality of segmentations.

We will solve the TV-based model using the split Bregman algorithm as given in [95]. Basically if one replace the update for $u^{k+\frac{1}{2}}$ in the first line of (5.17) by

$$\Delta u_{k+\frac{1}{2}} = \nabla \cdot (d_k - b_k) - \frac{\mu}{\nu} \zeta(c_k^1, c_k^2),$$

replace all W by ∇ and λ by g , then we obtain the split Bregman algorithm for TV-based segmentation model (see [95] for details). Note that the Laplace equation above is solved by FFT, instead of Gauss-Seidel relaxation as proposed in [95].

Since the decomposition operator W for tight frames corresponds to higher order difference operators, the frame based segmentation models should be more sensitive to low contrast boundaries and yet still robust to noise. To test this, we take a clean and high contrast image (first image in Figure 4) and then lower the contrast by 80%, add blurring and noise. We use the corrupted image (second image in Figure 4) as the input for both the TV-based and the frame based segmentation models.

In order to truly show the improvement of using tight frame systems, we pick the same set of parameters (μ, ν) and use the same λ for both the TV-based segmentation model and the frame based segmentation models (5.15). One can see from both Figure 4 and 5 that by using tight frame systems we can capture more features from the images and obtain better segmentations, especially for low contrast and blurry images.

4. Scene reconstruction from range data

4.1. Formulation and algorithm

For a given set of scattered sites $\{x[i]\}_{i=1}^n \subset \Omega \subset \mathbb{R}^2$ and the corresponding data $\{f[i]\}_{i=1}^n$, our task is to approximate the unknown surface $f(x)$ using $s(x) \in S^L(\phi, \Omega)$ by finding an appropriate vector \mathbf{u} such that

$$s(x) = \sum_{k \in \mathcal{I}} u[k] \phi(2^L x - k),$$

where ϕ is the 2D tensor product of some uniform B-spline function and

$$\mathcal{I} := \{k \in \mathbb{Z}^2 : \text{supp } \phi(2^L \cdot -k) \cap \Omega \neq \emptyset\}.$$

Since the function f that we want to recover can be sparsely approximated by tight frame systems, the desired \mathbf{u} can be obtained by solving the following problem

$$(5.18) \quad \min_{\mathbf{u}} \frac{1}{2} \|\mathbf{A}\mathbf{u} - \mathbf{f}\|_2^2 + \|\text{diag}(\lambda)W\mathbf{u}\|_1,$$

where

$$A(i, k) = \phi(2^L x[i] - k), \quad i = 1, 2, \dots, n, \quad k \in \mathcal{I}.$$

The matrix W is the fast wavelet frame transform associated with the bivariate B-spline tight frame. Note that $W\mathbf{u}$ is the canonical frame coefficient sequence of $s(x)$, hence the regularization term $\|\text{diag}(\lambda)W\mathbf{u}\|_1$ penalizes the roughness of the

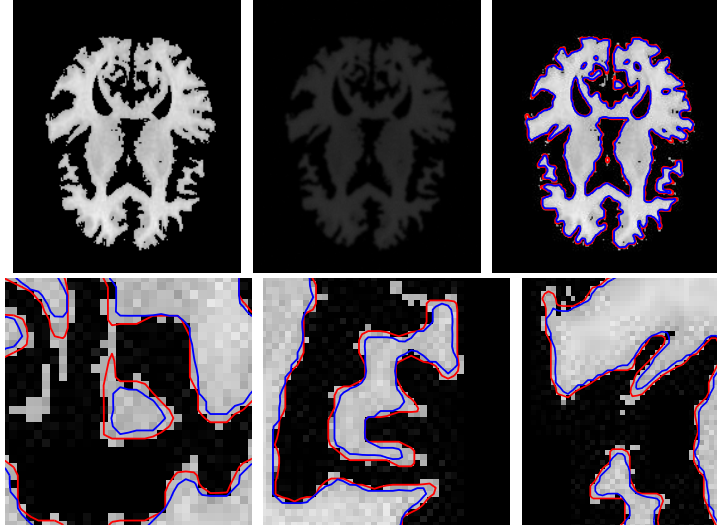


FIGURE 4. Comparisons of TV-based segmentation model with frame based segmentation model (5.15). First row: images from left to right are: original image; observed low contrast blurry image; segmentation results of the TV-based (blue) and the frame based (red) segmentations. Second row: zoom-in views of segmentation results. The set of parameters used is $(\mu, \nu) = (500, 5)$.

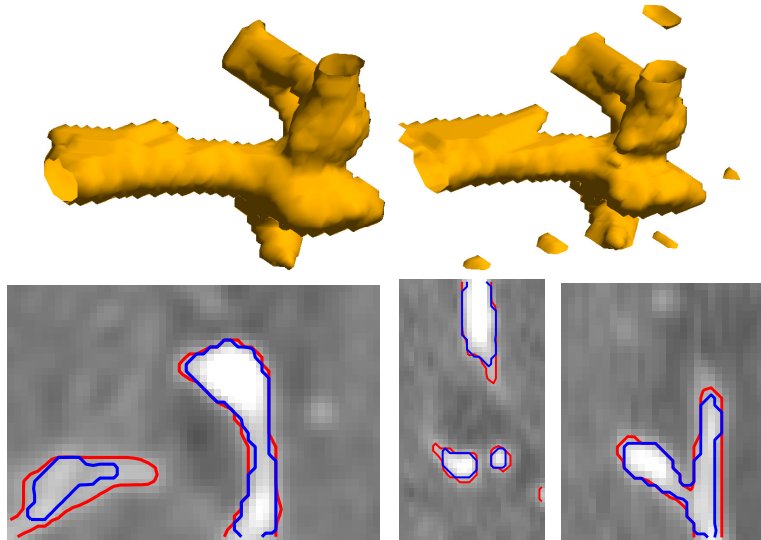


FIGURE 5. First row: comparisons of TV-based segmentation model (left) with frame based segmentation model (5.15) (right). Second row: axial, sagittal and coronal views of the results of the TV-based (blue) and the frame based (red) segmentation. The set of parameters used is $(\mu, \nu) = (100, 0.5)$.

fitting function $s(x)$, and also encourages preserving sharp edges as it is biased toward sparse solutions of $s(x)$. We note that the matrix W is constructed via the Neumann boundary condition, which implies that the symmetric boundary condition is used to extend \mathbf{u} across the boundaries of Ω .

The key part of the scene reconstruction model is to solve (5.18) efficiently, which can be achieved using split Bregman algorithm given in Section 3 of Lecture 4. For the readers' convenience, we recall the algorithm for problem (5.18). Given initializations $\mathbf{u}^0 = 0, \mathbf{d}^0 = \mathbf{b}^0 = 0$, iterate the following steps until convergence

$$(5.19) \quad \begin{aligned} \mathbf{u}_{k+1} &= (A^\top A + \mu I)^{-1} (A^\top \mathbf{f} + \mu W(\mathbf{d}_k - \mathbf{b}_k)), \\ \mathbf{d}_{k+1} &= \mathcal{T}_{\lambda/\mu}(W\mathbf{u}_{k+1} + \mathbf{b}_k), \\ \mathbf{b}_{k+1} &= \mathbf{b}_k + \delta(W\mathbf{u}_{k+1} - \mathbf{d}_{k+1}). \end{aligned}$$

The stopping criterion for the iteration (5.19) is chosen as $\|\mathbf{d}_k - W\mathbf{u}_k\| \leq \epsilon$ with ϵ being some given tolerance. One aspect of split Bregman iteration that contributes to its efficiency is that it is unnecessary to exactly solve the first equation of (5.19). Therefore, we approximate the solution \mathbf{u}_{k+1} of the first equation of (5.19) by a few steps of the conjugate gradient method.

In practice, there are usually a lot of outliers in the observed data (See e.g. [177]) and it is well-known that an ℓ_2 fidelity term is very sensitive to outliers. When there are outliers in the data, we use the ℓ_1 fidelity term instead:

$$(5.20) \quad \min_{\mathbf{u}} \|A\mathbf{u} - \mathbf{f}\|_1 + \|\text{diag}(\lambda)W\mathbf{u}\|_1,$$

The implications of the ℓ_1 data fidelity term have been well studied in the literature (e.g. [120, 111, 39, 144]). In particular, [144] justified the use of the ℓ_1 data fitting term for processing data corrupted with outliers in both theoretical analysis and numerical experiments.

In terms of Bayesian statistical estimation (see e.g. [8, 111]), the models (5.18) and (5.20) can be interpreted as the maximization of a *posteriori* estimator:

$$(5.21) \quad \max_{\mathbf{u}} \text{Exp} \left(-\frac{1}{p} \|A\mathbf{u} - \mathbf{f}\|_p^p \right) \text{Exp} (-\|\text{diag}(\lambda)W\mathbf{u}\|_1),$$

where the first factor is the likelihood that models the noise, and the second factor is the prior of \mathbf{u} . The prior is usually chosen based on the a priori knowledge we have for the unknown \mathbf{u} . For example, we know that the surface we need to recover from range image data is piecewise smooth and hence can be sparsely approximated by tight wavelet frames. Therefore, the prior given by (5.21) is reasonable. As for the likelihood, when the noise $\boldsymbol{\eta} := A\mathbf{u} - \mathbf{f}$ is i.i.d. Gaussian with zero mean, one should use $p = 2$ in order to obtain a statistically correct likelihood. When the noise $\boldsymbol{\eta}$ is impulsive noise, i.e. outliers, one may choose $p = 1$, because impulsive noise is sparse and can be well modeled by the ℓ_1 -norm (i.e. we believe that outliers are i.i.d. Laplace distribution). Notice that for $0 < p < 1$, the objective function of (5.21) is not concave. Therefore, although the ℓ_p -norm with $0 < p < 1$ may model impulsive noise better than the ℓ_1 -norm, we usually use the ℓ_1 -norm in practice for computational concerns.

In contrast to the ℓ_2 fitting term, the ℓ_1 fitting term is not smooth. In our numerical simulations, we approximate the non-differentiable ℓ_1 fitting term by a

smoothed version

$$(5.22) \quad \sum_{i=1}^n \sqrt{(A\mathbf{u}[i] - \mathbf{f}[i])^2 + s}$$

to approximate $\|A\mathbf{u} - \mathbf{f}\|_1$, where s is a small positive number.

To solve the ℓ_1 -norm fidelity model (5.20), we again use the split Bregman algorithm. Given initializations $\mathbf{u}^0 = 0, \mathbf{d}^0 = \mathbf{b}^0 = 0$, iterate the following steps until convergence

$$(5.23) \quad \begin{aligned} \mathbf{u}_{k+1} &= \arg \min_{\mathbf{u}} \sum_{i=1}^n \sqrt{(A\mathbf{u}[i] - \mathbf{f}[i])^2 + s} + \frac{\mu}{2} \|W\mathbf{u} - \mathbf{d}^k + \mathbf{b}^k\|_2^2, \\ \mathbf{d}_{k+1} &= \mathcal{T}_{\lambda/\mu}(W\mathbf{u}_{k+1} + \mathbf{b}_k), \\ \mathbf{b}_{k+1} &= \mathbf{b}_k + \delta(W\mathbf{u}_{k+1} - \mathbf{d}_{k+1}). \end{aligned}$$

For the first minimization subproblem of (5.23), since its objective functional is differentiable, it can be solved by the standard gradient descent method. One of the merits of the split Bregman algorithm that contributes to its efficiency is that it is unnecessary to solve the first subproblem of (5.23) to the full convergence. Instead, only a small number of inner iterations will be adequate at each split Bregman algorithm.

We remark that although we are using a smoothed version of the ℓ_1 data fidelity term in the above robust algorithm (5.23), the split Bregman algorithm actually can also be used to minimize an energy functional with both an ℓ_1 data fidelity term and an ℓ_1 regularization term without any smoothing (e.g. L1-TV minimization problem). Interested readers are referred to [73, 85] for more details.

4.2. Numerical experiments

The main purpose of this section is to briefly evaluate the performance of the tight frame based model (5.18) and (5.20) on reconstructing a piecewise smooth surface from sparse range data corrupted by Gaussian or impulse noise. More experiments and comparisons with some other existing methods can be found in [112]. Range image data are obtained from OSU (MSU/WSU) Range Image Database (<http://sampl.ece.ohio-state.edu/data/3DDB/RID/index.htm>). Figure 6 and 7 contain the reconstruction results of two scenes, where noise that is added to the original data is Gaussian and impulse plus small Gaussian noise respectively. Note that the input range data are scattered point sets, which are visualized as surfaces.

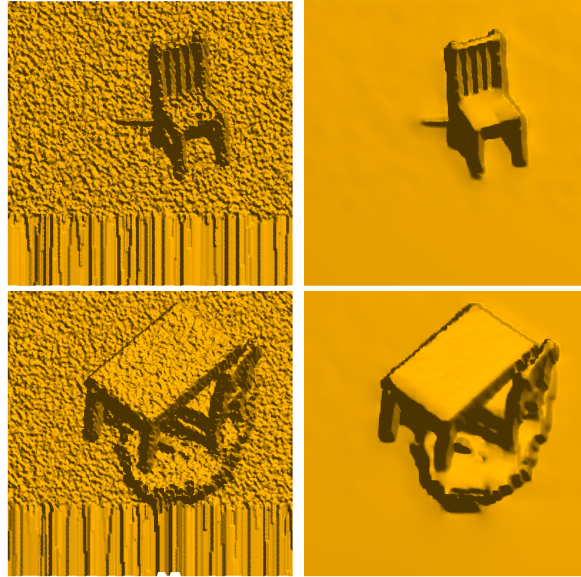


FIGURE 6. Reconstruction of scenes from range image data corrupted by Gaussian noise.

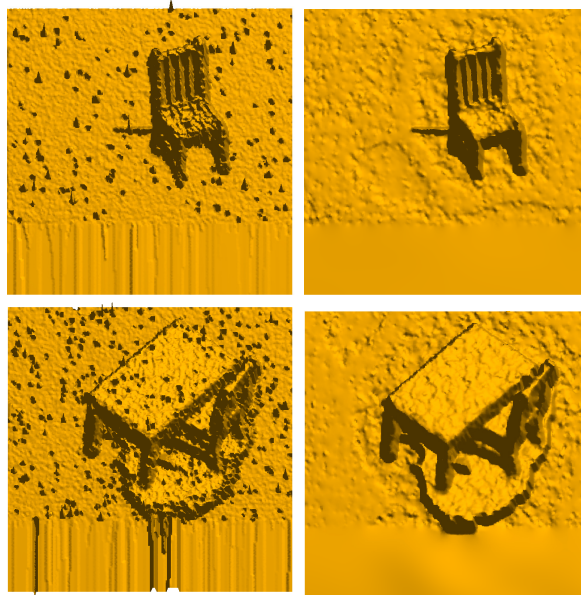


FIGURE 7. Reconstruction of scenes from range image data corrupted by impulse noise.

Bibliography

1. S. Arivazhagan and L. Ganesan, *Texture segmentation using wavelet transform*, Pattern Recognition Letters **24** (2003), no. 16, 3197–3203.
2. L. Bar, B. Berkels, M. Rumpf, and G. Sapiro, *A variational framework for simultaneous motion estimation and restoration of motion-blurred video*, International Conference on Computer Vision, 1–8, 2007.
3. B. Bascle, A. Blake, and A. Zisserman, *Motion deblurring and super-resolution from an image sequence*, Computer Vision/ECCV'96 (1996), 571–582.
4. A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM Journal on Imaging Sciences **2** (2009), no. 1, 183–202.
5. M. Ben-Ezra and S.K. Nayar, *Motion-based motion deblurring*, IEEE Transactions on Pattern Analysis and Machine Intelligence **26** (2004), no. 6, 689.
6. J.J. Benedetto and S. Li, *The Theory of Multiresolution Analysis Frames and Applications to Filter Banks*, Applied and Computational Harmonic Analysis **5** (1998), no. 4, 389–427.
7. M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, *Image inpainting*, Proceedings of the 27th annual conference on Computer graphics and interactive techniques, ACM Press/Addison-Wesley Publishing Co., 2000, pp. 417–424.
8. J. Besag, *Towards Bayesian image analysis*, Journal of Applied Statistics **20** (1993), no. 5, 107–119.
9. L. Borup, R. Gribonval, and M. Nielsen, *Bi-framelet systems with few vanishing moments characterize Besov spaces* 1*, Applied and Computational Harmonic Analysis **17** (2004), no. 1, 3–28.
10. ———, *Tight wavelet frames in lebesgue and sobolev spaces*, Journal of Function Spaces and Applications **2** (2004), no. 3, 227–252.
11. S.P. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge Univ Pr, 2004.
12. X. Bresson, S. Esedoglu, P. Vanderghenst, J.P. Thiran, and S. Osher, *Fast global minimization of the active contour/snake model*, Journal of Mathematical Imaging and Vision **28** (2007), no. 2, 151–167.
13. J.F. Cai, E.J. Candès, and Z. Shen, *A Singular Value Thresholding Algorithm for Matrix Completion*, SIAM Journal on Optimization **20**(4) (2010), 1956–1982.
14. J.F. Cai, R. Chan, L. Shen, and Z. Shen, *Restoration of chopped and noded images by framelets*, SIAM J. Sci. Comput **30** (2008), no. 3, 1205–1227.
15. J.F. Cai, R.H. Chan, L. Shen, and Z. Shen, *Convergence analysis of tight framelet approach for missing data recovery*, Advances in Computational Mathematics (2008), 1–27.
16. J.F. Cai, R.H. Chan, and Z. Shen, *A framelet-based image inpainting algorithm*, Applied and Computational Harmonic Analysis **24** (2008), no. 2, 131–149.
17. ———, *Simultaneous cartoon and texture inpainting*, Inverse Problems and Imaging (IPI) **4** (2010), no. 3, 379–395.
18. J.F. Cai, H. Ji, C. Liu, and Z. Shen, *Blind motion deblurring from a single image using sparse approximation*, Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE, 2009, pp. 104–111.
19. ———, *Blind motion deblurring using multiple images*, Journal of Computational Physics **228** (2009), no. 14, 5057–5071.
20. ———, *High-quality curvelet-based motion deblurring from an image pair*, Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE, 2009, pp. 1566–1573.
21. J.F. Cai, S. Osher, and Z. Shen, *Convergence of the linearized Bregman iteration for ℓ_1 -norm minimization*, Mathematics of Computation **78** (2009), 2127–2136.

22. ———, *Linearized Bregman iterations for compressed sensing*, Math. Comp **78** (2009), 1515–1536.
23. ———, *Linearized Bregman iterations for frame-based image deblurring*, SIAM J. Imaging Sci **2** (2009), no. 1, 226–252.
24. ———, *Split Bregman methods and frame based image restoration*, Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal **8** (2009), no. 2, 337–369.
25. J.F. Cai and Z. Shen, *Framelet based deconvolution*, J. Comp. Math **28** (2010), no. 3, 289–308.
26. E.J. Candes and D.L. Donoho, *New tight frames of curvelets and optimal representations of objects with C_2 singularities*, Comm. Pure Appl. Math **56** (2004), 219–266.
27. ———, *Continuous curvelet transform.: II. Discretization and frames*, Applied and Computational Harmonic Analysis **19** (2005), no. 2, 198–222.
28. Vicent Caselles, R Kimmel, and G Sapiro, *Geodesic active contours*, Int J Comput Vision **22** (1997), no. 1, 61–79 (English).
29. A.S. Cavaretta, W. Dahmen, and C.A. Micchelli, *Stationary subdivision*, Amer Mathematical Society, 1991.
30. A.E. Cetin, *Reconstruction of signals from Fourier transform samples*, Signal Processing **16** (1989), no. 2, 129–148.
31. AE Cetin, *An iterative algorithm for signal reconstruction from bispectrum*, IEEE Transactions on Signal Processing **39** (1991), no. 12, 2621–2628.
32. A. Chai and Z. Shen, *Deconvolution: A wavelet frame approach*, Numerische Mathematik **106** (2007), no. 4, 529–587.
33. A. Chambolle, *An algorithm for total variation minimization and applications*, Journal of Mathematical Imaging and Vision **20** (2004), no. 1, 89–97.
34. R.H. Chan, T.F. Chan, L. Shen, and Z. Shen, *Wavelet algorithms for high-resolution image reconstruction*, SIAM Journal on Scientific Computing **24** (2003), no. 4, 1408–1432.
35. R.H. Chan, S.D. Riemenschneider, L. Shen, and Z. Shen, *High-resolution image reconstruction with displacement errors: A framelet approach*, International Journal of Imaging Systems and Technology **14** (2004), no. 3, 91–104.
36. ———, *Tight frame: an efficient way for high-resolution image reconstruction*, Applied and Computational Harmonic Analysis **17** (2004), no. 1, 91–115.
37. R.H. Chan, L. Shen, and Z. Shen, *A framelet-based approach for image inpainting*, Research Report **4** (2005), 325.
38. R.H. Chan, Z. Shen, and T. Xia, *A framelet algorithm for enhancing video stills*, Applied and Computational Harmonic Analysis **23** (2007), no. 2, 153–170.
39. T.F. Chan and S. Esedoglu, *Aspects of Total Variation Regularized L' Function Approximation*, SIAM Journal on Applied Mathematics **65** (2005), no. 5, 1817.
40. T.F. Chan, G.H. Golub, and P. Mulet, *A nonlinear primal-dual method for total variation-based image restoration*, SIAM Journal on Scientific Computing **20** (1999), no. 6, 1964–1977.
41. T.F. Chan and J. Shen, *Image processing and analysis: variational, pde, wavelet, and stochastic methods*, Society for Industrial Mathematics, 2005.
42. ———, *Variational image inpainting*, Commun. Pure Appl. Math **58** (2005), 579–619.
43. T.F. Chan and L.A. Vese, *Active contours without edges*, IEEE Transactions on image processing **10** (2001), no. 2, 266–277.
44. T.F. Chan and C.K. Wong, *Total variation blind deconvolution*, IEEE Transactions on Image Processing **7** (1998), no. 3, 370–375.
45. Tony F Chan, Selim Esedoglu, and Mila Nikolova, *Algorithms for finding global minimizers of image segmentation and denoising models*, Siam J Appl Math **66** (2006), no. 5, 1632–1648 (English).
46. J. Chen, L. Yuan, C.K. Tang, and L. Quan, *Robust dual motion deblurring*, Proc. Conf. Comp. Vision and Pattern Recognition, 2008, pp. 1–8.
47. S.S. Chen, D.L. Donoho, and M.A. Saunders, *Atomic decomposition by basis pursuit*, SIAM review (2001), 129–159.
48. Y. Chen, H.D. Tagare, S. Thiruvankadam, F. Huang, D. Wilson, K.S. Gopinath, R.W. Briggs, and E.A. Geiser, *Using prior shapes in geometric active contours in a variational framework*, International Journal of Computer Vision **50** (2002), no. 3, 315–328.

49. S. Cho, Y. Matsushita, S. Lee, and P. POSTECH, *Removing non-uniform motion blur from images*, IEEE 11th International Conference on Computer Vision, 2007. ICCV 2007, 2007, pp. 1–8.
50. O. Christensen, *Frames and bases: An introductory course*, Birkhauser, 2008.
51. C.K. Chui and W. He, *Compactly supported tight frames associated with refinable functions*, Applied and Computational Harmonic Analysis **8** (2000), no. 3, 293–319.
52. C.K. Chui, W. He, and J. Stöckler, *Compactly supported tight and sibling frames with maximum vanishing moments*, Applied and computational harmonic analysis **13** (2002), no. 3, 224–262.
53. A. Cohen, *Biorthogonal wavelets*, Wavelets: A Tutorial in Theory and Applications, Academic Press Professional, Inc., 1993, pp. 123–152.
54. A. Cohen and J.P. Conze, *Régularité des bases ondelettes et mesures ergodiques*, Rev. Mat. Iberoamericana **8** (1992), no. 3, 351–365.
55. A. Cohen and I. Daubechies, *A stability criterion for biorthogonal wavelet bases and their related subband coding scheme*, Duke Math. J **68** (1992), no. 2, 313–335.
56. A. Cohen, I. Daubechies, and J.C. Feauveau, *Biorthogonal Bases of Compactly Supported Wavelets*, Communications on Pure and Applied Mathematics **45** (1992), no. 5, 485–560.
57. R.R. Coifman and D.L. Donoho, *Translation-invariant de-noising*, Lecture Notes in Statistics–New York–Springer Verlag (1995), 125–125.
58. P.L. Combettes and V.R. Wajs, *Signal recovery by proximal forward-backward splitting*, Multiscale Modeling and Simulation **4** (2006), no. 4, 1168–1200.
59. J. Darbon and S. Osher, *Fast discrete optimizations for sparse approximations and deconvolutions*, preprint 2007.
60. I. Daubechies, *Orthonormal bases of compactly supported wavelets*, Commun. Pure Appl. Math. **41** (1988), no. 7, 909–996.
61. ———, *Ten lectures on wavelets*, vol. CBMS-NSF Lecture Notes, SIAM, nr. 61, Society for Industrial Mathematics, 1992.
62. I. Daubechies, M. Defrise, and C. De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Communications on pure and applied mathematics **57** (2004), no. 11, 1413–1457.
63. I. Daubechies, A. Grossmann, and Y. Meyer, *Painless nonorthogonal expansions*, Journal of Mathematical Physics **27** (1986), 1271.
64. I. Daubechies and B. Han, *Pairs of dual wavelet frames from any two refinable functions*, Constructive approximation **20** (2004), no. 3, 325–352.
65. I. Daubechies, B. Han, A. Ron, and Z. Shen, *Framelets: Mra-based constructions of wavelet frames*, Applied and Computational Harmonic Analysis **14** (2003), no. 1, 1–46 (English).
66. I. Daubechies, G. Teschke, and L. Vese, *Iteratively solving linear inverse problems under general convex constraints*, Inverse Problems and Imaging **1** (2007), no. 1, 29.
67. C. De Boor, *A practical guide to splines*, Springer Verlag, 2001.
68. C. De Boor, R.A. DeVore, and A. Ron, *On the construction of multivariate (pre) wavelets*, Constructive approximation **9** (1993), no. 2, 123–166.
69. ———, *Approximation from shift-invariant subspaces of $L_2(\mathbb{R}^d)$* , Transactions of the American Mathematical Society **341** (1994), no. 2, 787–806.
70. ———, *The structure of finitely generated shift-invariant spaces in $L_2(\mathbb{R}^d)$* , J. Funct. Anal **119** (1994), no. 1, 37–78.
71. B. Dong, A. Chien, and Z. Shen, *Frame based segmentation for medical images*, Communications in Mathematical Sciences **9** (2011), no. 2, 551–559.
72. B. Dong, N. Dyn, and K. Hormann, *Properties of dual pseudo-splines*, Applied and Computational Harmonic Analysis **29** (2010), no. 1, 104–110.
73. B. Dong, E. Savitsky, and S. Osher, *A Novel Method for Enhanced Needle Localization Using Ultrasound-Guidance*, Advances in Visual Computing (2009), 914–923.
74. B. Dong and Z. Shen, *Linear independence of pseudo-splines*, Proceedings of the American Mathematical Society **134** (2006), no. 9, 2685–2694.
75. ———, *Pseudo-splines, wavelets and framelets*, Applied and Computational Harmonic Analysis **22** (2007), no. 1, 78–104.
76. D.L. Donoho, *Interpolating wavelet transforms*, Preprint, Department of Statistics, Stanford University (1992).

77. ———, *De-noising by soft-thresholding*, IEEE transactions on information theory **41** (1995), no. 3, 613–627.
78. ———, *Ridge Functions and Orthonormal Ridgelets*, Journal of Approximation Theory **111** (2001), no. 2, 143–179.
79. S. Dubuc, *Interpolation through an iterative scheme*, Journal of mathematical analysis and applications **114** (1986), no. 1, 185–204.
80. R.J. Duffin and A.C. Schaeffer, *A class of nonharmonic Fourier series*, Transactions of the American Mathematical Society **72** (1952), no. 2, 341–366.
81. N. Dyn, K. Hormann, M.A. Sabin, and Z. Shen, *Polynomial reproduction by symmetric subdivision schemes*, Journal of Approximation Theory **155** (2008), no. 1, 28–42.
82. I. Ekeland and R. Temam, *Convex analysis and variational problems*, Society for Industrial Mathematics, 1999.
83. M. Elad, P. Milanfar, and R. Rubinstein, *Analysis versus synthesis in signal priors*, Inverse Problems **23** (2007), 947.
84. M. Elad, J.L. Starck, P. Querre, and D.L. Donoho, *Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)*, Applied and Computational Harmonic Analysis **19** (2005), no. 3, 340–358.
85. E. Esser, *Applications of lagrangian-based alternating direction methods and connections to split bregman*, CAM report **9** (2009), 31.
86. M.J. Fadili and J.L. Starck, *Sparse representations and bayesian image inpainting*, Proc. SPARS **5** (2005).
87. M.J. Fadili, J.L. Starck, and F. Murtagh, *Inpaining and zooming using sparse representations*, The Computer Journal **52** (2009), no. 1, 64.
88. W. Feller, *An introduction to probability theory and its applications. Vol. 1*, Wiley New York, 1968.
89. R. Fergus, B. Singh, A. Hertzmann, S.T. Roweis, and W.T. Freeman, *Removing camera shake from a single photograph*, ACM Transactions on Graphics (TOG) **25** (2006), no. 3, 787–794.
90. M.A.T. Figueiredo and R.D. Nowak, *An EM algorithm for wavelet-based image restoration*, IEEE Transactions on Image Processing **12** (2003), no. 8, 906–916.
91. ———, *A bound optimization approach to wavelet-based image deconvolution*, Image Processing, 2005. ICIP 2005. IEEE International Conference on, vol. 2, IEEE, 2005, pp. II–782.
92. N. Flasque, M. Desvignes, J.M. Constans, and M. Revenu, *Acquisition, segmentation and tracking of the cerebral vascular tree on 3D magnetic resonance angiography images*, Medical Image Analysis **5** (2001), no. 3, 173–183.
93. M. Frazier, G. Garrigós, K. Wang, and G. Weiss, *A characterization of functions that generate wavelet and related expansion*, Journal of Fourier Analysis and Applications **3** (1997), 883–906.
94. R. Glowinski and P. Le Tallec, *Augmented Lagrangian and operator-splitting methods in nonlinear mechanics*, Society for Industrial Mathematics, 1989.
95. T. Goldstein, X. Bresson, and S. Osher, *Geometric Applications of the Split Bregman Method: Segmentation and Surface Reconstruction*, UCLA CAM Report (2009), 09–06.
96. T. Goldstein and S. Osher, *The split Bregman algorithm for L1 regularized problems*, SIAM Journal on Imaging Sciences **2** (2009), no. 2, 323–343.
97. R. Gribonval and M. Nielsen, *On approximation with spline generated framelets*, Constructive Approximation **20** (2004), no. 2, 207–232.
98. G. Gripenberg, *A necessary and sufficient condition for the existence of a father wavelet*, Studia Math **114** (1995), no. 3, 207–226.
99. K. Gröchenig, *Foundations of time-frequency analysis*, Birkhauser, 2001.
100. M. Gross and H. Pfister, *Point-based graphics*, Morgan Kaufmann Pub, 2007.
101. B. Han, *On Dual Wavelet Tight Frames*, Applied and Computational Harmonic Analysis **4** (1997), no. 4, 380–413.
102. B. Han, G. Kutyniok, and Z. Shen, *A Unitary Extension Principle for Shearlet Systems*, Arxiv preprint arXiv:0912.4529 (2009).
103. B. Han and Z. Shen, *Wavelets with short support*, SIAM Journal on Mathematical Analysis **38** (2007), no. 2, 530–556.
104. ———, *Compactly supported symmetric C wavelets with spectral approximation order*, SIAM J. Math. Anal **40** (2008), 905–938.

105. ———, *Characterization of Sobolev spaces of arbitrary smoothness using nonstationary tight wavelet frames*, Israel Journal of Mathematics **172** (2009), no. 1, 371–398.
106. ———, *Dual wavelet frames and Riesz bases in Sobolev spaces*, Constructive Approximation **29** (2009), no. 3, 369–406.
107. X. He, L. Shen, and Z. Shen, *A data-adaptive knot selection scheme for fitting splines*, IEEE Signal Processing Letters **8** (2001), no. 5, 137.
108. E. Hernández and G. Weiss, *A first course on wavelets*, Studies in Advanced Mathematics, CRC Press, 1996.
109. J.B. Hiriart-Urruty and C. Lemaréchal, *Convex analysis and minimization algorithms: Fundamentals*, Springer, 1993.
110. Y. Hur and A. Ron, *CAPlets: wavelet representations without wavelets*, preprint (2005).
111. A. Hyvärinen, J. Hurri, and P.O. Hoyer, *Natural Image Statistics: A Probabilistic Approach to Early Computational Vision*, Springer-Verlag New York Inc, 2009.
112. H. Ji, Z. Shen, and Y. Xu, *Wavelet frame based scene reconstruction from range data*, Journal of Computational Physics **229** (2010), no. 6, 2093–2108.
113. J. Jia, *Single image motion deblurring using transparency*, IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR'07, 2007, pp. 1–8.
114. R.Q. Jia and C.A. Micchelli, *Using the refinement equations for the construction of pre-wavelets II: Powers of two*, Curves and surfaces (1991), 209–246.
115. R.Q. Jia and Z. Shen, *Multiresolution and wavelets*, Proc. Edinb. Math. Soc., II. Ser. **37** (1994), no. 2, 271–300.
116. Q. Jiang and Z. Shen, *On existence and weak stability of matrix refinable functions*, Constructive Approximation **15** (1999), no. 3, 337–353.
117. M.J. Johnson, *Scattered Data Interpolation from Principal Shift-Invariant Spaces*, Journal of Approximation Theory **113** (2001), no. 2, 172–188.
118. M.J. Johnson, Z. Shen, and Y. Xu, *Scattered data reconstruction by regularization in B-spline and associated wavelet spaces*, Journal of Approximation Theory **159** (2009), no. 2, 197–223.
119. N. Joshi, R. Szeliski, and D.J. Kriegman, *PSF estimation using sharp edge prediction*, IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008, 2008, pp. 1–8.
120. S.A. Kassam and J.B. Thomas, *Signal detection in non-Gaussian noise*, Springer-Verlag New York, NY, 1988.
121. A. Khodakovsky, P. Schröder, and W. Sweldens, *Progressive geometry compression*, Proceedings of the 27th annual conference on Computer graphics and interactive techniques, Citeseer, 2000, pp. 271–278.
122. C. Kirbas and F. Quek, *A review of vessel extraction techniques and algorithms*, ACM Computing Surveys **36** (2004), no. 2, 81–121.
123. G. Kutyniok and D. Labate, *Resolution of the wavefront set using continuous shearlets*, AMERICAN MATHEMATICAL SOCIETY **361** (2009), no. 5, 2719–2754.
124. W.M. Lawton, *Tight frames of compactly supported affine wavelets*, Journal of Mathematical Physics **31** (1990), 1898.
125. W.M. Lawton, S.L. Lee, and Z. Shen, *Convergence of multidimensional cascade algorithm*, Numerische Mathematik **78** (1998), no. 3, 427–438.
126. M. Leventon, W.E.L. Grimson, and O. Faugeras, *Statistical shape influence in geodesic active contours*, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, IEEE Computer Society; 1999, 2000.
127. A. Levin, *Blind motion deblurring using image statistics*, Advances in Neural Information Processing Systems **19** (2007), 841.
128. C. Li, C.Y. Kao, J.C. Gore, and Z. Ding, *Minimization of region-scalable fitting energy for image segmentation*, IEEE Transactions on Image Processing **17** (2008), no. 10, 1940–1949.
129. M. Li, B. Hao, and X.C. Feng, *Iterative regularization and nonlinear inverse scale space based on translation invariant wavelet shrinkage*, INTERNATIONAL JOURNAL OF WAVELETS MULTIREOLUTION AND INFORMATION PROCESSING **6** (2008), no. 1, 83.
130. C.G. Looney, *Convergence of minimizing sequences*, Journal of Mathematical Analysis and Applications **61** (1977), no. 3, 835–840.
131. S.G. Mallat, *Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R})$* , Transactions of the American Mathematical Society **315** (1989), no. 1, 69–87.

132. ———, *A wavelet tour of signal processing*, vol. 2nd ed. New York: Academic, Academic press, 1999.
133. T. McInerney and D. Terzopoulos, *Deformable models in medical image analysis: a survey*, *Medical image analysis* **1** (1996), no. 2, 91–108.
134. Y. Meyer, *Wavelets and operators*. Translated by DH Salinger, Cambridge Studies in Advanced Mathematics, 1992.
135. ———, *Oscillating patterns in image processing and nonlinear evolution equations: the fifteenth Dean Jacqueline B. Lewis memorial lectures*, Amer Mathematical Society, 2001.
136. Q. Mo, Y. Shen, and S. Li, *A new proof of some polynomial inequalities related to pseudo-splines*, *Applied and Computational Harmonic Analysis* **23** (2007), no. 3, 415–418.
137. J.J. Moreau, *Fonctions convexes duales et points proximaux dans un espace hilbertien*, *CR Acad. Sci. Paris Sér. A Math* **255** (1962), 2897–2899.
138. ———, *Proximité et dualité dans un espace hilbertien*, *Bull. Soc. Math. France* **93** (1965), no. 2, 273–299.
139. D. Nain, A. Yezzi, and G. Turk, *Vessel segmentation using a shape driven flow*, *Lecture Notes in Computer Science* (2004), 51–59.
140. A.S. Nemirovskij and D.B. Yudin, *Problem complexity and method efficiency in optimization*. transl. from the russian by er dawson, Wiley-Interscience Series in Discrete Mathematics, 1983.
141. Y. Nesterov, *A method of solving a convex programming problem with convergence rate $o(1/k^2)$* , *Soviet Mathematics Doklady*, vol. 27, 1983, pp. 372–376.
142. ———, *On an approach to the construction of optimal methods for minimizing smooth convex functions*, *Ehkon. Mat. Metody* **24** (1988), no. 3, 509–517.
143. ———, *Smooth minimization of non-smooth functions*, *Mathematical Programming* **103** (2005), no. 1, 127–152.
144. M. Nikolova, *A variational approach to remove outliers and impulse noise*, *Journal of Mathematical Imaging and Vision* **20** (2004), no. 1, 99–120.
145. S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, *An iterative regularization method for total variation based image restoration*, *Multiscale Model. Simul* **4** (2005), no. 2, 460–489.
146. S. Osher and R.P. Fedkiw, *Level set methods and dynamic implicit surfaces*, Springer, 2003.
147. S. Osher, Y. Mao, B. Dong, and W. Yin, *Fast linearized bregman iteration for compressive sensing and sparse denoising*, *Communications in Mathematical Sciences* **8** (2010), no. 2, 93–111.
148. S. Osher and J. Sethian, *Fronts propagating with curvature-dependent speed- Algorithms based on Hamilton-Jacobi formulations*, *Journal of Computational Physics* **79** (1988), 12–49.
149. G. Pavlovic and AM Tekalp, *Maximum likelihood parametric blur identification based on acontinuous spatial domain model*, *IEEE Transactions on image processing* **1** (1992), no. 4, 496–504.
150. S.M. Pizer, P.T. Fletcher, S. Joshi, A. Thall, J.Z. Chen, Y. Fridman, D.S. Fritsch, A.G. Gash, J.M. Glotzer, M.R. Jiroutek, et al., *Deformable m-reps for 3d medical image segmentation*, *International Journal of Computer Vision* **55** (2003), no. 2, 85–106.
151. E. Polak, *Computational methods in optimization*, Academic Press New York, 1971.
152. G. Pólya and G. Szegő, *Aufgaben und Lehrsätze aus der Analysis*, Springer, 1971.
153. R. Raskar, A. Agrawal, and J. Tumblin, *Coded exposure photography: motion deblurring using fluttered shutter*, *ACM Transactions on Graphics (TOG)* **25** (2006), no. 3, 795–804.
154. R.T. Rockafellar and J.B.W. Roger, *Variational analysis*, Springer Verlag, 1997.
155. A. Ron, *Factorization theorems for univariate splines on regular grids*, *Israel Journal of Mathematics* **70** (1990), no. 1, 48–68.
156. A. Ron and Z. Shen, *Frames and Stable Bases for Shift-Invariant Subspaces of $L_2(\mathbb{R}^d)$* , *Canadian Journal of Mathematics* **47** (1995), no. 5, 1051–1094.
157. ———, *Affine systems in $L_2(\mathbb{R}^d)$ II: dual systems*, *Journal of Fourier Analysis and Applications* **3** (1997), no. 5, 617–638.
158. ———, *Affine Systems in $L_2(\mathbb{R}^d)$: The Analysis of the Analysis Operator*, *Journal of Functional Analysis* **148** (1997), no. 2, 408–447.
159. ———, *Weyl-Heisenberg frames and Riesz bases in $L_2(\mathbb{R}^d)$* , *Duke Mathematical Journal* **89** (1997), no. 2, 237–282.

160. ———, *Compactly supported tight affine spline frames in $L_2(\mathbb{R}^d)$* , Mathematics of Computation **67** (1998), no. 221, 191–207.
161. ———, *The Sobolev regularity of refinable functions*, Journal of Approximation Theory **106** (2000), 185–225.
162. ———, *Generalized shift-invariant systems*, Constructive Approximation **22** (2005), no. 1, 1–45.
163. W. Rudin, *Real and complex analysis*, Tata McGraw-Hill, 2006.
164. G. Sapiro, *Geometric partial differential equations and image analysis*, Cambridge Univ Pr, 2001.
165. I.W. Selesnick, *Smooth Wavelet Tight Frames with Zero Moments* 1*, Applied and Computational Harmonic Analysis **10** (2001), no. 2, 163–181.
166. Q. Shan, J. Jia, and A. Agarwala, *High-quality motion deblurring from a single image*, ACM Transactions on Graphics-TOG **27** (2008), no. 3, 73–73.
167. Z. Shen, *Wavelet frames and image restorations*, Proceedings of the International Congress of Mathematicians, vol. 4, 2010, pp. 2834–2863.
168. Z. Shen, K. C. Toh, and S. Yun, *An accelerated proximal gradient algorithm for frame based image restorations via the balanced approach*, SIAM Journal on Imaging Sciences **4**(2) (2011), 573–596.
169. Z. Shen and S. Waldron, *Scattered data interpolation by box splines*, Study in Advanced Mathematics, AMS/IP **42** (2008), no. 2, 749.
170. J.L. Starck, M. Elad, and D.L. Donoho, *Image decomposition via the combination of sparse representations and a variational approach*, IEEE transactions on image processing **14** (2005), no. 10, 1570–1582.
171. X.C. Tai and C. Wu, *Augmented Lagrangian method, dual methods and split Bregman iteration for ROF model*, Scale Space and Variational Methods in Computer Vision (2009), 502–513.
172. Y.W. Tai, H. Du, MS Brown, and S. Lin, *Image/video deblurring using a hybrid camera*, Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, IEEE, 2008, pp. 1–8.
173. K.C. Toh and S. Yun, *An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems*, Pacific Journal of Optimization **6** (2010), no. 20, 615–640.
174. A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, W.E. Grimson, and A. Willisky, *A shape-based approach to the segmentation of medical imagery using level sets*, IEEE Transactions on Medical Imaging **22** (2003), no. 2, 137–154.
175. P. Tseng, *On accelerated proximal gradient methods for convex-concave optimization*, submitted to SIAM Journal on Optimization (2008).
176. Z. Tu, K.L. Narr, P. Dollar, I. Dinov, P.M. Thompson, and A.W. Toga, *Brain anatomical structure segmentation by hybrid discriminative/generative models*, IEEE Transactions on Medical Imaging **27** (2008), no. 4, 495–508.
177. M. Umasuthan and AM Wallace, *Outlier removal and discontinuity preserving smoothing of rangedata*, IEE Proceedings-Vision, Image and Signal Processing **143** (1996), no. 3, 191–200.
178. M. Unser, *Texture classification and segmentation using wavelet frames*, IEEE Transactions on image processing **4** (1995), no. 11, 1549–1560.
179. L.A. Vese and T.F. Chan, *A multiphase level set framework for image segmentation using the Mumford and Shah model*, International Journal of Computer Vision **50** (2002), no. 3, 271–293.
180. C.R. Vogel, M.E. Oman, et al., *Fast, robust total variation-based reconstruction of noisy, blurred images*, IEEE Transactions on Image Processing **7** (1998), no. 6, 813–824.
181. N. Wiener, *The Fourier integral and certain of its applications*, Cambridge Univ Pr, 1988.
182. J. Xu and S. Osher, *Iterative regularization and nonlinear inverse scale space applied to wavelet-based denoising*, IEEE Transactions on Image Processing **16** (2007), no. 2, 534.
183. J. Yang, L.H. Staib, and J.S. Duncan, *Neighbor-constrained segmentation with 3d deformable models*, LECTURE NOTES IN COMPUTER SCIENCE (2003), 198–209.
184. W. Yin, S. Osher, D. Goldfarb, and J. Darbon, *Bregman iterative algorithms for l_1 -minimization with applications to compressed sensing*, SIAM J. Imaging Sci **1** (2008), no. 1, 143–168.

185. X. Zhang, M. Burger, X. Bresson, and S. Osher, *Bregmanized nonlocal regularization for deconvolution and sparse reconstruction*, SIAM Journal on Imaging Sciences **3** (2010), no. 3, 253–276.