

# COMPACTLY SUPPORTED FRAME WAVELETS AND APPLICATIONS

---

A Dissertation Presented to  
the Faculty of the Department of Mathematics  
University of Houston

---

In Partial Fulfillment  
of the Requirements for the Degree  
Doctor of Philosophy

---

By  
Nikolaos Karantzas  
August 2019

# COMPACTLY SUPPORTED FRAME WAVELETS AND APPLICATIONS

---

Nikolaos Karantzas

APPROVED:

---

Prof. Emanuel Papadakis, Chair  
Dept. of Mathematics, UH

---

Prof. Demetrio Labate  
Dept. of Mathematics, UH

---

Prof. Bernhard G. Bodmann  
Dept. of Mathematics, UH

---

Prof. Saurabh Prasad  
Dept. of Electrical Engineering, UH

---

Dean, College of Natural Sciences and Mathematics

# Acknowledgements

There are many people without whom this work would not have been possible. I would like to thank my advisor Dr. Emanuel Papadakis and my co-advisor Dr. Demetrio Labate for their valuable help and guidance. Working with you has given me a new world to explore. Special thanks to the committee members Dr. Bernhard Bodmann and Dr. Saurabh Prasad for their insightful comments and advice. I also need to thank Dr. Nicholas Alikakos and Dimitris Valsamis for getting it all started. Special thanks to Dr. Nikolaos Atreas and Dr. Theodoros Stavropoulos for the fruitful collaboration.

Loukia, Leonidas, and Charis, thank you for your  $\infty$  support and for setting the bar so high. Many thanks to Vasso, Axilleas, Akis, Costas,  $x_0$ , Despina, George, Stefanos, Neil, Duong, Kayla, Daewa, Zach, Robert, Photis, Stratos and the Oropians. I am also very grateful to Mo, Kazem, and Soren for their friendship and effort that led to the completion of this dissertation. Special thanks to the Graduate Tuition Fellowship (GTF) program for supporting me financially throughout the whole duration of my studies.

This document is dedicated to my grandparents Charalampos and Maria.

# COMPACTLY SUPPORTED FRAME WAVELETS AND APPLICATIONS

---

An Abstract of a Dissertation  
Presented to  
the Faculty of the Department of Mathematics  
University of Houston

---

In Partial Fulfillment  
of the Requirements for the Degree  
Doctor of Philosophy

---

By  
Nikolaos Karantzas  
August 2019

# Abstract

Signal processing has been at the forefront of modern information technology as the need for storing, analyzing, and interpreting data gathered all around us is ever growing. Multi-dimensional sparse signal representations occupy a significant part of the literature on multi-scale decompositions. The interest in such representations arises from their ability to analyze, synthesize, and modify signals carrying information about the behavior of specific phenomena.

This work is devoted to the development and design of application-targeted tools for the multi-variable analysis of image data. Our main interests revolve around both the theoretical and practical aspects of signal processing, machine learning, and deep neural networks. In Chapter 1 we present the necessary mathematical background this work is based on. In Chapter 2 we develop a theoretical base for the construction of a specific class of compactly supported Parseval Framelets with directional characteristics. The framelets we construct arise from readily available refinable functions and their filters have few non-zero coefficients, custom-selected orientations and can act as finite-difference operators. We present explicit examples related to well-known directional representations (directional filter banks). Finally, in Chapter 3 we explore the capabilities of our construction in the growing field of deep convolutional neural networks.

# Contents

<b>1</b>	<b>Previous Mathematical Results</b>	<b>1</b>
1.1	The Fourier Transform . . . . .	1
1.2	The Short-Time Fourier Transform . . . . .	7
1.3	The Sampling Theorem . . . . .	8
1.4	Wavelet Analysis . . . . .	11
1.5	Multiresolution Analysis . . . . .	13
1.6	Vanishing Moments and the Wavelet Construction of Daubechies . . . . .	18
1.7	Frames in Hilbert Spaces . . . . .	22
<b>2</b>	<b>Compactly Supported Parseval Wavelet Frames for <math>L_2(\mathbb{R}^s)</math>.</b>	<b>26</b>
2.1	Dyadic Wavelet Frames in $L_2(\mathbb{R}^s)$ . . . . .	27
2.2	The Geometry of the Proposed Construction . . . . .	33
2.3	Wavelets with Directional Vanishing Moments and Customizable Filters. . . . .	43
2.4	Adaptive MRA and Compactly Supported Parseval Frames. . . . .	54
2.5	Algorithmic Construction of Parseval Framelets . . . . .	59
2.6	Filter Bank Examples . . . . .	62
<b>3</b>	<b>A Neural Network Application to the Kaggle Quick, Draw! Data Set</b>	<b>76</b>
3.1	Project Description . . . . .	76
3.2	The Data Set . . . . .	77
3.3	Scoring . . . . .	79
3.4	Preprocessing . . . . .	80

3.5	Models . . . . .	81
3.5.1	ResNet . . . . .	81
3.5.2	ResNetRF . . . . .	83
3.5.3	Model Architectures . . . . .	83
3.6	Training Strategy . . . . .	86
3.6.1	Feed-forward Scheme . . . . .	86
3.6.2	Model Compilation . . . . .	87
3.7	Results . . . . .	88
3.7.1	A Qualitative Error Examination . . . . .	91
3.7.2	Test Results . . . . .	95
<b>4</b>	<b>Conclusions</b>	<b>96</b>
<b>A</b>	<b>Filter Banks</b>	<b>98</b>
A.1	$3 \times 3$ Filters for ResNetRF-One . . . . .	98
A.1.1	Low-pass Filter . . . . .	98
A.1.2	High-pass Filters . . . . .	98
A.2	$7 \times 7$ Filters for ResNetRF-One . . . . .	99
A.2.1	Low-pass Filter . . . . .	99
A.2.2	High-pass Filters . . . . .	99
A.3	$3 \times 3$ Filters for ResNetRF-All . . . . .	104
A.3.1	Low-pass Filter . . . . .	104
A.3.2	High-pass Filters . . . . .	104
A.4	$7 \times 7$ Filters for ResNetRF-All . . . . .	105
A.4.1	Low-pass Filter . . . . .	105
A.4.2	High-pass Filters . . . . .	106
	<b>Bibliography</b>	<b>114</b>

# List of Figures

1.1	(Left) Audio signal changing over 4 seconds. (Right) Magnitude of the Fourier Transform of the audio signal. . . . .	4
1.2	(Left) Windowed signal over $[0, 1]$ . (Right) Magnitude of the Fourier Transform. . . . .	8
2.1	Left: Cubic spline interpolation of binary image containing a single horizontal non-zero band. Values in this image are constant in the direction of $(0, 1)$ . Right: $2D$ convolution with $h$ defining a wavelet with 4 directional vanishing moments in the direction of $(0, 1)$ . As expected, the lower polynomial degree of the intensity profile of the left panel relative to the number of DVM of $h$ parallel to $(0, 1)$ practically flattens the cubic spline bump in the middle of the left panel. . . . .	51
2.2	This is a $256 \times 256$ image freely available with Matlab 2017. We use it to demonstrate the interaction of the designed filters with singularities in various directions. . . . .	68
2.3	Application of $h_i, i = 3, 4, 5, 6$ constructed in Example 2.6.3 as discrete 2D-convolution kernels at native resolution. The first two filters act as first-order directional central-difference filters oriented at $135^\circ$ and $0^\circ$ , respectively. The last two act as second-order central-difference filters oriented at $0^\circ$ and $135^\circ$ , respectively. Note that singularity detection strength increases as edges are oriented closer to being perpendicular to the orientation of each filter. In Fig 2.5 we see that this effect may also be related to the anisotropy of the filter and its size. . . . .	69
2.4	The dashed lines show four successive positions of central bands defining this predesigned filter set. Once the central band has been set, we choose its nearest diametrically opposite bands to create all first and second-order finite difference filters allowed by this process. . . . .	72

2.5	Application of $h_i$ , $i = 5, 6, 17, 18$ at native resolution. The first two convolutions correspond to filters with orientations at $135^\circ$ and $153.43^\circ$ , respectively. The last two convolutions correspond to filters with orientations at $135^\circ$ and $153.43^\circ$ , respectively. . . . .	72
3.1	Strokes belonging to 2 classes drawn by linear interpolation. . . . .	79
3.2	Drawing images from the sampled strokes for a variety of classes. Note that the images are not binary. There is a difference in intensity values from stroke to stroke. . . . .	81
3.3	A residual block used for propagation of information over multiple layers. The image is taken from [39]. . . . .	82
3.4	ResNet18 MAP@3 validation precision (left) and cross entropy loss (right). . . . .	89
3.5	ResNet34 MAP@3 validation precision (left) and cross entropy loss (right). . . . .	90
3.6	ResNet50 MAP@3 validation precision (left) and cross entropy loss (right). . . . .	90
3.7	ResNet18 validation confusion matrices for 200 samples per class. . .	92
3.8	ResNet34 validation confusion matrices for 200 samples per class. . .	93
3.9	ResNet50 validation confusion matrices for 200 samples per class. . .	94

# List of Tables

3.1	(Left) Basic residual block. (Right) Bottleneck residual block. . . . .	84
3.2	Network architectures. We note that there are shortcut identity connections from block to block in all three architectures. . . . .	85
3.3	Number of trainable parameters. . . . .	88
3.4	Test accuracy provided by Kaggle . . . . .	95

# Chapter 1

## Previous Mathematical Results

In this chapter we give a short introduction leading up to the results presented in this dissertation. We wish to make this document as self-contained as possible. The proofs of the statements made here are omitted, but the inquiring reader may find a more detailed discussion in the books of Ole Christensen [15, Chapters 7, 8, 9][14, Chapters 3, 4, 5] as well as in the classical treatise of Weiss and Hernandez [40, Chapters 2, 3, 7, 8]. We begin with the Fourier Transform and the insights leading to the Short-Time Fourier Transform (STFT). We continue with Shannon's Sampling Theorem and close with the basics of one-dimensional wavelet bases and the concept of a multiresolution analysis (MRA), the wavelets of Daubechies, and frames in Hilbert spaces.

### 1.1 The Fourier Transform

The Fourier Transform in  $L_2(\mathbb{R})$  is a very important tool used both in theory and applications to extract the frequency content of a given function (signal). Recall that for a periodic signal  $f \in L_2[0, 1)$  its Fourier Series allows us to express  $f$  as a

superposition of oscillatory functions in the form  $\sum_{k \in \mathbb{Z}} c_k e^{2\pi i k x}$ . However, expansions of non-periodic functions are just as desired, but not equally simple. In this case, all frequencies can appear in the signal. Loosely speaking, the Fourier Series is replaced by an integral over the real numbers giving rise to the Fourier Transform in  $L_1(\mathbb{R})$ .

**Definition 1.1.1.** [15, Definition 7.1.1] Let  $f \in L_1(\mathbb{R})$ . The Fourier Transform associates to  $f$  a new function  $\widehat{f} : \mathbb{R} \rightarrow \mathbb{C}$  given by

$$\widehat{f}(\gamma) = \int_{\mathbb{R}} f(x) e^{-2\pi i \gamma x} dx, \quad \gamma \in \mathbb{R}.$$

The Fourier Transform in  $L_1(\mathbb{R})$  can be seen as an operator, denoted by  $\mathcal{F}$ , mapping the function  $f$  to  $\widehat{f}$ . Furthermore, the triangle inequality for integrals implies

$$|\mathcal{F}f(\gamma)| = |\widehat{f}(\gamma)| \leq \int_{\mathbb{R}} |f(x) e^{-2\pi i \gamma x}| dx = \int_{\mathbb{R}} |f(x)| dx < \infty,$$

so  $\mathcal{F}$  is a well-defined integral operator. We also observe that  $\mathcal{F}$  is linear, i.e., for  $f, g \in L_1(\mathbb{R})$  we have

$$\mathcal{F}(af + bg)(\gamma) = a\mathcal{F}(f)(\gamma) + b\mathcal{F}(g)(\gamma).$$

Moreover, the following theorem helps us gain insights into the range space of  $\mathcal{F}$  as well as establish that  $\mathcal{F}$  is also bounded.

**Theorem 1.1.2.** [15, Theorem 7.1.5] (Riemann-Lebesgue) Let  $f \in L_1(\mathbb{R})$ . Then  $\widehat{f} : L_1(\mathbb{R}) \rightarrow C_0(\mathbb{R})$ .

Theorem 1.1.2 implies that  $\mathcal{F}$  is a bounded linear operator since  $|\mathcal{F}f(\gamma)| \leq \|f\|_1$ , which in turn implies that

$$\|\mathcal{F}f\|_{\infty} \leq \|f\|_1, \quad f \in L_1(\mathbb{R}).$$

More importantly, we will later see that the Fourier Transform in  $L_1$  can be extended to a unitary bounded operator mapping  $L_2(\mathbb{R})$  onto  $L_2(\mathbb{R})$ . In the relevant literature,

several extensions of the Fourier Transform to other spaces have been established, e.g., to  $L_p(\mathbb{R})$ . The extension to  $L_2(\mathbb{R})$  is particularly useful since  $L_2(\mathbb{R})$  is a Hilbert space yielding some very special properties.

Next, under certain assumptions, knowledge of the Fourier Transform of a signal  $f$  can help us fully recover the signal in its original form. This result is known as the inversion formula of the Fourier Transform.

**Theorem 1.1.3.** [15, Theorem 7.1.7] (*Inversion of the Fourier Transform in  $L_1(\mathbb{R})$* )

*Assume that both  $f$  and  $\widehat{f}$  are  $L_1(\mathbb{R})$  functions. Then*

$$f(x) = \int_{\mathbb{R}} \widehat{f}(\gamma) e^{2\pi i x \gamma} d\gamma$$

*for almost every  $x \in \mathbb{R}$ . If  $f$  is continuous, the above formula holds pointwise for all  $x \in \mathbb{R}$ .*

As mentioned at the beginning of this section, the Fourier Transform can be used to reveal the frequency content in a given signal. For example, assume  $f$  is a continuous function representing a piece of music starting at time  $t = 0$  sec and running until time  $t = 4$  sec. This signal is therefore compactly supported over this time interval. The frequencies present are not immediately accessible and this is where the Fourier Transform comes into play. Figure 1.1 below shows how such a signal changes over time as well as the amplitude of its Fourier Transform.

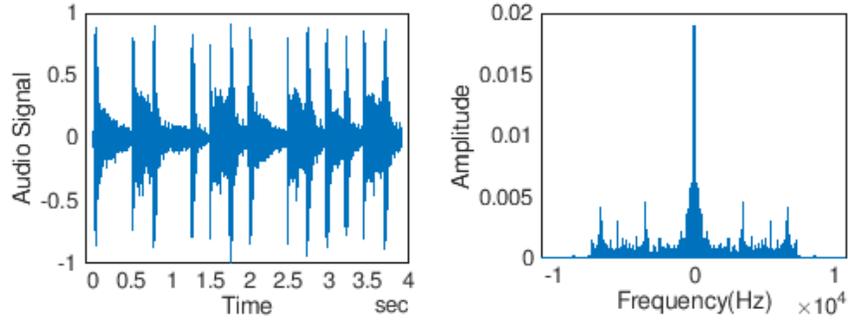


Figure 1.1: (Left) Audio signal changing over 4 seconds. (Right) Magnitude of the Fourier Transform of the audio signal.

By looking at the magnitude of  $\hat{f}$ , we are able to obtain the frequency information appearing in  $f$ . We can then expect the drums and bass in the lower frequency range, the vocals as mid range frequencies, and any possible noise in high frequencies. However, the Fourier transform is non-local in the sense that the frequency content in our signal does not yield information about how these frequencies line up in the time interval the signal was generated. This information is contained within the phase of  $\hat{f}$  and is essentially what allows us to recover the piece of music as it was originally played. In other words, if we write  $\hat{f}(\gamma) = |A(\gamma)|e^{i\phi(\gamma)}$ , then  $|A(\gamma)|$  determines the relative presence of a sinusoid  $e^{2\pi i\gamma t}$  in  $f(t)$  while  $\phi(\gamma)$  determines how the sinusoids line up relative to one another to form  $f(t)$ .

In real life applications, just as described above, signals are compactly supported. For this reason, this motivates examining the space  $C_c(\mathbb{R})$  of continuous compactly supported functions as a stepping stone to extend the Fourier Transform to the Hilbert space  $L_2(\mathbb{R})$ . The following lemma shows that if we equip  $C_c(\mathbb{R})$  with the  $L_2(\mathbb{R})$  norm, the Fourier Transform is an isometry from  $C_c(\mathbb{R})$  into  $L_2(\mathbb{R})$ .

**Lemma 1.1.4.** [15, Lemma 7.2.1] For any  $f \in C_c(\mathbb{R})$  we have

$$\int_{\mathbb{R}} |\mathcal{F}f(\gamma)|^2 d\gamma = \int_{\mathbb{R}} |f(x)|^2 dx.$$

Finally, the next theorem helps us arrive at the extension of the Fourier Transform in  $L_2(\mathbb{R})$ .

**Theorem 1.1.5.** [15, Theorem 3.3.2] (*Extension by uniform continuity*) Let  $X$  and  $Y$  be Banach spaces. Let  $K$  be a dense subspace of  $X$  and  $T : K \rightarrow Y$  a bounded linear operator. Then there exists a unique bounded linear operator  $S : X \rightarrow Y$  for which  $Sv = Tv$  for all  $v \in K$ . The operator  $S$  satisfies that  $\|S\| = \|T\|$ .

Using the fact that  $C_c(\mathbb{R})$  is dense in  $L_2(\mathbb{R})$  and Theorem 1.1.5 we have the extension theorem of the Fourier Transform in  $L_2(\mathbb{R})$ .

**Theorem 1.1.6.** [15, Theorem 7.2.2] The Fourier Transform can be extended to a unitary mapping of  $L_2(\mathbb{R})$  onto  $L_2(\mathbb{R})$  satisfying the following:

(i) (*Plancherel*) For all  $f \in L_2(\mathbb{R})$ , we have  $\|\mathcal{F}f\|_2 = \|f\|_2$ .

(ii) (*Parseval*) For all  $f, g \in L_2(\mathbb{R})$ , we have  $\langle \mathcal{F}f, \mathcal{F}g \rangle = \langle f, g \rangle$ .

Recall that unitary operators are invertible and so Theorem 1.1.6 helps us define the inverse Fourier Transform as a mapping from  $L_2(\mathbb{R})$  onto  $L_2(\mathbb{R})$ .

**Theorem 1.1.7.** [15, Theorem 7.2.3] Assume that  $f \in L_2(\mathbb{R})$  and that  $\widehat{f} \in L_1(\mathbb{R})$ . Then

$$f(x) = \int_{\mathbb{R}} \widehat{f}(\gamma) e^{2\pi i x \gamma} d\gamma$$

for almost every  $x \in \mathbb{R}$ . If  $f$  is continuous, the above formula holds pointwise for all  $x \in \mathbb{R}$ .

We have seen how the Fourier Transform is defined as a bounded unitary operator in  $L_2(\mathbb{R})$ . We also mentioned the non-local nature of the Fourier Transform. To address this issue we will define The Short-Time Fourier Transform, but before doing so we introduce various useful linear operators on  $L_2(\mathbb{R})$ .

**Definition 1.1.8.** Let  $f \in L_2(\mathbb{R})$ . For  $a, b \in \mathbb{R}$  and  $c \in \mathbb{R}^+$ , we define:

- (i) The *translation* operator  $\tau_a f(x) := f(x - a)$ ,  $x \in \mathbb{R}$ ,
- (ii) the *modulation* operator  $M_b f(x) := e^{2\pi i b x} f(x)$ ,  $x \in \mathbb{R}$ , and
- (iii) the *dilation* operator  $D_c f(x) := \sqrt{c} f(cx)$ ,  $x \in \mathbb{R}$ .

**Lemma 1.1.9.** *The translation, modulation, and dilation operators are unitary linear operators mapping  $L_2(\mathbb{R})$  onto  $L_2(\mathbb{R})$  satisfying the following:*

- (i) For  $a \in \mathbb{R}$ , we have  $\tau_a^{-1} = \tau_{-a} = \tau_a^*$ .
- (ii) For  $b \in \mathbb{R}$ , we have  $M_b^{-1} = M_{-b} = M_b^*$ .
- (iii) For  $c \in \mathbb{R}^+$ , we have  $D_c^{-1} = D_{1/c} = D_c^*$ .

Compositions of these operators have proven to play an important role in various branches of mathematics. For example, the results presented in this document revolve around wavelet systems, which consist of dilated and translated versions of fixed functions. In terms of the Fourier Transform, the following theorem shows the behavior of  $\mathcal{F}$  when composed with the translation, modulation, and dilation operators. We note that the translation and modulation operators have dual roles.

**Theorem 1.1.10.** *Let  $f \in L_2(\mathbb{R})$ . Then*

(a) For  $a \in \mathbb{R}$ , the Fourier Transform of  $\tau_a f$  satisfies

$$\mathcal{F}\tau_a f(\gamma) = \widehat{f}(\gamma)e^{-2\pi ia\gamma} = M_{-a}\widehat{f}(\gamma).$$

(b) For  $b \in \mathbb{R}$ , the Fourier Transform of  $M_b f$  satisfies

$$\mathcal{F}M_b f(\gamma) = \widehat{f}(\gamma - b) = \tau_b \widehat{f}(\gamma).$$

(c) For each  $c \in \mathbb{R}^+$ , the Fourier Transform of  $D_c f$  satisfies

$$\mathcal{F}D_c f(\gamma) = D_{1/c}\widehat{f}(\gamma).$$

## 1.2 The Short-Time Fourier Transform

The Short-Time Fourier Transform serves as a remedy for the non-local nature of the Fourier Transform. It is used to analyze the frequency and phase content of local sections of a signal. In practical applications, the process of computing the Short-Time Fourier Transform is to partition a longer time signal into shorter equal length intervals and then compute the Fourier transform separately on each short interval. This reveals the spectrum over each interval.

For example, let us consider  $f(t) = \sin(2\pi 15t) + \sin(2\pi 20t)$ ,  $t \in [0, 1]$ . Suppose we are interested in obtaining the frequency content of this function over the interval  $[1/4, 3/4]$ . We consider the function  $f\chi_{[1/4, 3/4]}$  and compute its Fourier Transform by

$$\int_{\mathbb{R}} f(t)\chi_{[1/4, 3/4]}(t)e^{-2\pi i\gamma t} dt = \int_{1/4}^{3/4} f(t)e^{-2\pi i\gamma t} dt.$$

Figure 1.2 below shows the windowed signal and the magnitude of its Short-Time Fourier Transform.

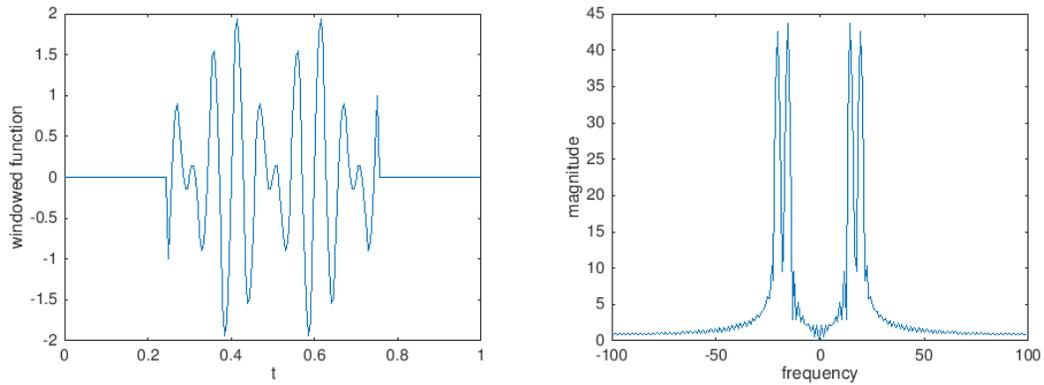


Figure 1.2: (Left) Windowed signal over  $[0, 1]$ . (Right) Magnitude of the Fourier Transform.

One thing to keep in mind is that calculating the Fourier Transform of a function  $f\chi_{\text{window}}$  will produce unwanted oscillations in the frequency domain due to the discontinuity cut-offs imposed by the characteristic function. These oscillations can be seen in the figure above, where instead of having just two spikes corresponding to frequencies  $\gamma = 15$  and  $\gamma = 20$ , we get visible oscillations around these frequencies as well. However, we do gain some understanding of the time information contained in our signal  $f$ . We can then use the translation operator to slide the window accordingly and cover the whole time interval over which our signal exists.

### 1.3 The Sampling Theorem

We saw how knowledge of the frequency content in a given signal can be sufficient to completely reconstruct the signal. In this Section we present Shannon's Sampling Theorem, which not only is the key tool that leads to a different view on the same problem, but also leads to a notion of efficient signal compression.

The task at hand is the following: How can we fully recover a signal  $f : \mathbb{R} \rightarrow \mathbb{C}$  if we only know a countable set of its values  $\{f(x_k)\}_{k \in I}$ ? Formulated this way the problem is ill-posed: there are infinitely many functions having the same values on a given countable set. So we need appropriate conditions to impose on the function  $f$  for such a problem to make sense. This is usually done by requiring  $f$  to belong to a certain function space. A classical example is to consider the space of functions whose Fourier transform is compactly supported or, in other words, the space of *band-limited* functions.

**Definition 1.3.1.** [15, Definition 7.4.1] The Paley-Wiener space  $PW$  is the subspace of  $L_2(\mathbb{R})$  defined by

$$PW := \left\{ f \in L_2(\mathbb{R}) : \text{supp } \hat{f} \subset \left[ -\frac{1}{2}, \frac{1}{2} \right] \right\}$$

The Paley-Wiener space, just as any other subspace of  $L_2(\mathbb{R})$ , consists of equivalence classes of functions. Due to the fact that the Fourier transform of functions in  $PW$  has compact support, each of these equivalence classes contains a continuous representative.

**Theorem 1.3.2.** [15, Theorem 7.4.3] (*Continuity of functions in  $PW$* ) Assume that  $f \in PW$ . Then  $f$  is equivalent to a continuous function. Moreover,  $\hat{f} \in L_1(\mathbb{R})$ .

In order to get a representation for functions in  $PW$  we first need to establish an orthonormal basis for  $PW$ . Shannon's theorem shows that the Paley-Wiener space has an orthonormal basis consisting of translates of the sinc function

$$\text{sinc}(x) := \begin{cases} \frac{\sin(\pi x)}{\pi x}, & x \neq 0 \\ 1, & x = 0. \end{cases}$$

Moreover, the theorem states that any continuous function  $f$  in the Paley-Wiener space can be fully recovered from its samples on the integer grid.

**Theorem 1.3.3.** [15, Theorem 7.4.5] (Shannon's sampling theorem) *The functions  $\{\text{sinc}(\cdot - k)\}_{k \in \mathbb{Z}}$  form an orthonormal basis for  $PW$ . If  $f \in PW$  is continuous, then*

$$f(x) = \sum_{k \in \mathbb{Z}} f(k) \text{sinc}(x - k),$$

*with two convergence interpretations of the infinite series:*

(i) *The symmetric partial sums converge pointwise, i.e.,*

$$\lim_{N \rightarrow \infty} \sum_{k=-N}^N f(k) \text{sinc}(x - k) = f(x),$$

*for all  $x \in \mathbb{R}$ .*

(ii) *The symmetric partial sums converge in  $L_2(\mathbb{R})$ , i.e.,*

$$\lim_{N \rightarrow \infty} \left\| f - \sum_{k=-N}^N f(k) \text{sinc}(\cdot - k) \right\|_{L_2(\mathbb{R})} = 0.$$

With the appropriate scaling, Shannon's sampling theorem can be extended to functions whose Fourier transform is supported over any arbitrary interval. In fact, if  $\text{supp } \hat{f} \subset [-a/2, a/2]$ , the result takes the form

$$f(x) = \sum_{k \in \mathbb{Z}} f\left(\frac{k}{a}\right) \text{sinc}(ax - k), \quad x \in \mathbb{R}. \quad (1.1)$$

The underlying principle in Shannon's theorem was the basis for modern communication technology. Most signals in real life applications depend on continuous variables and their processing is facilitated efficiently if they can be handled in terms of sequences of samples.

Returning to our example of the recorded piece of music, all frequencies might appear in the signal. But humans can only hear frequencies within a certain range (at most up to 20,000 Hz). Thus, we can discard the high frequencies and consider the piece of music  $f$  as band-limited, e.g., with  $\text{supp } \widehat{f} \subset [-20000, 20000]$ . Equation (1.1) shows that this signal can be reconstructed from its samples at the points  $k/40000$ ,  $k \in \mathbb{Z}$ . Thus, all information about the signal is contained in a discrete sequence of numbers. This principle was used in CD players and in other applications where an analog-to-digital conversion is needed.

However, sinc has a very slow decay rate. Thus, in trying to have a good finite-sum approximation of a given signal, we need to store a large number of its samples  $f(k)$ . This problem gave rise to the analysis of wavelets: functions with faster decay and the ability to achieve more efficient representations of signals (compression)!

## 1.4 Wavelet Analysis

Although the first wavelets appeared a century ago, their systematic analysis began in the 80s. The concept of a Multiresolution Analysis (MRA) was introduced in 1987, and shortly thereafter Daubechies used it to construct a special class of wavelet orthonormal bases with very useful properties in the context of data compression.

Wavelet theory gives us a way of constructing orthonormal bases for  $L_2(\mathbb{R})$  from dyadic dilations and translations of a fixed set of functions. This unique structure was the key idea that led to the success of wavelets in various signal processing applications. In fact, if  $\{e_k\}_{k \in \mathbb{Z}}$  is an orthonormal basis for  $L_2(\mathbb{R})$ , we know that all functions  $f \in L_2(\mathbb{R})$  have an expansion  $f = \sum_{k \in \mathbb{Z}} c_k e_k$  for suitable coefficients  $\{c_k\}_{k \in \mathbb{Z}}$ . However, in order for this representation to be of practical use, one hopes

that the relevant signals  $f$  can be well approximated by finite partial sums with only a few nonzero coefficients. In other words, we want the representation of  $f$  to be sparse.

**Definition 1.4.1.** [15, Definition 8.1.1] Let  $\psi \in L_2(\mathbb{R})$ . For  $j, k \in \mathbb{Z}$ , define the functions  $\psi_{j,k}$  by

$$\psi_{j,k}(x) := D_2^j \tau_k \psi(x) = 2^{j/2} \psi(2^j x - k), \quad x \in \mathbb{R}.$$

The function  $\psi$  is called an orthonormal wavelet if the functions  $\{\psi_{j,k}\}_{j,k \in \mathbb{Z}}$  form an orthonormal basis for  $L_2(\mathbb{R})$ .

If  $\psi$  is an orthonormal wavelet, then for any function  $f \in L_2(\mathbb{R})$ , we have a representation of the form

$$f = \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \langle f, \psi_{j,k} \rangle \psi_{j,k}.$$

This is a very special property which suggests that only very special functions can be orthonormal wavelets. The first orthonormal wavelet construction was achieved by Haar who in 1910 proved that with

$$\psi(x) = \begin{cases} 1, & 0 \leq x < 1/2, \\ -1, & 1/2 \leq x < 1, \\ 0, & \text{otherwise,} \end{cases} \quad (1.2)$$

the functions  $\{\psi_{j,k}\}_{j,k \in \mathbb{Z}}$  form an orthonormal basis for  $L_2(\mathbb{R})$ . The proof of the basis property is quite technical and exceeds the purpose of this introductory presentation. However, the reader can refer to [29] for more details. Another famous construction was the *Shannon wavelet* for which  $\psi = \chi_S$ , where  $S = [-1, -1/2) \cup [1/2, 1)$ . In this case, Theorem 1.1.10 suggests that for a function  $\psi \in L_2(\mathbb{R})$ , we have

$$\widehat{\psi}_{j,k}(\gamma) = 2^{-j/2} e^{-2\pi i k 2^{-j} \gamma} \widehat{\psi}(2^{-j} \gamma) \quad (1.3)$$

for  $j, k \in \mathbb{Z}$ . A closer examination of the sets  $\{2^j S : j \in \mathbb{Z}\}$  reveals that they form a mutually disjoint cover of  $\mathbb{R} \setminus \{0\}$ . Since

$$\{e^{-2\pi i k \cdot} \chi_S : k \in \mathbb{Z}\}$$

forms an orthonormal basis for  $L_2(S)$ , the functions  $2^{-j/2} e^{-2\pi i k 2^{-j} \cdot}$  in Equation (1.3), restricted on  $2^j S$ , form an orthonormal basis for  $L_2(2^j S)$  for each  $j \in \mathbb{Z}$ . Hence, the collection of functions  $\{\psi_{j,k} : j, k \in \mathbb{Z}\}$  forms an orthonormal basis for  $L_2(\mathbb{R})$ , i.e.,  $\psi$  is an orthonormal wavelet.

As one suspects, the construction of orthonormal wavelets is not a straightforward task. The following section is based on the theoretical foundations developed for achieving this goal.

## 1.5 Multiresolution Analysis

In 1987, Mallat and Meyer introduced *Multiresolution Analysis* (MRA) as a general tool to construct wavelet orthonormal bases [51, 52].

**Definition 1.5.1.** [15, Definition 8.2.1] A MRA for  $L_2(\mathbb{R})$  consists of a sequence of closed subspaces  $\{V_j\}_{j \in \mathbb{Z}}$  of  $L_2(\mathbb{R})$  and a function  $\phi \in V_0$  such that the following conditions hold:

- (i) The spaces  $V_j$  are nested, i.e.,

$$\cdots V_{-1} \subset V_0 \subset V_1 \subset \cdots .$$

- (ii)  $\overline{\cup_{j \in \mathbb{Z}} V_j} = L_2(\mathbb{R})$  and  $\cap_{j \in \mathbb{Z}} V_j = \{0\}$ .

- (iii) For all  $j \in \mathbb{Z}$ , we have  $V_{j+1} = D_2 V_j$ .

(iv)  $f \in V_0$  implies  $\tau_k f \in V_0$  for all  $k \in \mathbb{Z}$ .

(v)  $\{\tau_k \phi\}_{k \in \mathbb{Z}}$  is an orthonormal basis for  $V_0$ .

In Definition 1.5.1, condition (ii) means that  $\cup_{j \in \mathbb{Z}} V_j$  is dense in  $L_2(\mathbb{R})$ , i.e., for any  $f \in L_2(\mathbb{R})$  and any  $\epsilon > 0$ , there exists a function  $g \in \cup_{j \in \mathbb{Z}} V_j$  such that  $\|f - g\|_2 \leq \epsilon$ . The function  $g$  will belong to  $V_J$  for some  $J \in \mathbb{Z}$ , and so by condition (i), to all spaces  $V_j$  with  $j \geq J$ . Lastly, a closer look reveals that the choice of the function  $\phi$  in a MRA actually determines the spaces  $V_j$  uniquely. For this reason, we say that  $\phi$  generates the MRA, but we have to emphasize the fact that only very special functions  $\phi$  can generate multiresolution analyses.  $\phi$  is called an orthonormal scaling function.

**Lemma 1.5.2.** [15, Theorem 8.2.2] *Assume that conditions (iii - iv - v) in Definition 1.5.1 hold. Then*

(i)  $V_j = D_2^j V_0$  for all  $j \in \mathbb{Z}$ .

(ii)  $V_j = \overline{\text{span}}\{D_2^j \tau_k \phi\}_{k \in \mathbb{Z}}$  for all  $j \in \mathbb{Z}$ .

For example, the *Haar* MRA is defined by the function  $\phi = \chi_{[0,1]}$  and the relevant spaces  $\{V_j\}_{j \in \mathbb{Z}}$  are given by

$$V_j = \left\{ f \in L_2(\mathbb{R}) : f \text{ is constant on } 2^{-j}[k, k+1) \text{ for all } k \in \mathbb{Z} \right\}.$$

The orthonormal wavelet associated with this MRA is the Haar wavelet  $\psi$  defined in Equation (1.2). Note that  $\psi$  is a finite linear combination of dilated and translated versions of  $\phi$ . It turns out that a similar result is true for all orthonormal wavelets generated from a multiresolution analysis.

We will now see how we can use a multiresolution analysis to construct an orthonormal basis for  $L_2(\mathbb{R})$ , but we first need to consider a class of vector spaces related to the spaces  $\{V_j\}_{j \in \mathbb{Z}}$ .

**Definition 1.5.3.** [15, Definition 8.2.4] Assume that  $V_j$  is a sequence of closed subspaces of  $L_2(\mathbb{R})$  and that condition (i) in Definition 1.5.1 holds. For any  $j \in \mathbb{Z}$ , let  $W_j$  denote the orthogonal complement of  $V_j$  with respect to  $V_{j+1}$ , i.e.,

$$W_j = \{f \in V_{j+1} : \langle f, g \rangle = 0 \text{ for all } g \in V_j\}.$$

Out of the spaces  $W_j$ ,  $j \in \mathbb{Z}$ , it turns out that the role of  $W_0$  is crucial. In fact, the following result states that the collection of functions  $\{D_2^j \tau_k \psi\}_{k,j \in \mathbb{Z}}$  forms an orthonormal basis for  $L_2(\mathbb{R})$  if there exists a function  $\psi \in W_0$  such that  $\{\tau_k \psi\}_{k \in \mathbb{Z}}$  is an orthonormal basis for  $W_0$ .

**Proposition 1.5.4.** [15, Proposition 8.2.5] Assume that  $\phi \in L_2(\mathbb{R})$  generates a MRA. Let  $\psi \in L_2(\mathbb{R})$  and suppose that  $\{\tau_k \psi\}_{k \in \mathbb{Z}}$  is an orthonormal basis for  $W_0$ . Then the following hold:

- (i) For each  $j \in \mathbb{Z}$ , the functions  $\{D_2^j \tau_k \psi\}_{k \in \mathbb{Z}}$  form an orthonormal basis for  $W_j$ .
- (ii) The functions  $\{D_2^j \tau_k \psi\}_{j,k \in \mathbb{Z}}$  form an orthonormal basis for  $L_2(\mathbb{R})$ , i.e.,  $\psi$  is an orthonormal wavelet.
- (iii) The functions  $\{\tau_k \phi\}_{k \in \mathbb{Z}} \cup \{D_2^j \tau_k \psi\}_{j \in \mathbb{N}, k \in \mathbb{Z}}$  form an orthonormal basis for  $L_2(\mathbb{R})$ .

Conceptually, Proposition 1.5.4 offers a simplification in the construction of orthonormal wavelets: the wavelet system  $\{D_2^j \tau_k \psi\}_{j,k \in \mathbb{Z}}$  involves the operations of scaling and translation, while  $\{\tau_k \psi\}_{k \in \mathbb{Z}}$  just consists of translations. The following result is a key step in the construction of an appropriate function  $\psi$ .

**Proposition 1.5.5.** [15, Proposition 8.2.6] Assume that  $\phi \in L_2(\mathbb{R})$  generates a MRA. Then there exists a 1-periodic function  $H_0 \in L_2(0,1)$  such that

$$\widehat{\phi}(2\gamma) = H_0(\gamma) \widehat{\phi}(\gamma), \quad \gamma \in \mathbb{R}. \tag{1.4}$$

The above equation is called a *refinement equation*. A function  $\phi$  that satisfies a refinement equation is said to be *refinable*. The function  $H_0$  is called a *low-pass filter* associated with the refinable function  $\phi$ . We are now ready to present the main construction of orthonormal wavelets via a MRA. We note that the result is only based on knowledge of the low-pass filter  $H_0$ .

**Theorem 1.5.6.** [15, Theorem 8.2.7] Assume that  $\phi \in L_2(\mathbb{R})$  generates a MRA and let  $H_0 \in L_2(0, 1)$  be a 1-periodic function satisfying the scaling Equation (1.4). Define the 1-periodic function  $H_1$  by

$$H_1(\gamma) := \overline{H_0(\gamma + 1/2)} e^{-2\pi i \gamma}. \quad (1.5)$$

Also define  $\psi$  via

$$\widehat{\psi}(2\gamma) = H_1(\gamma) \widehat{\phi}(\gamma). \quad (1.6)$$

Then the following hold:

- (i)  $\{\tau_k \psi\}_{k \in \mathbb{Z}}$  is an orthonormal basis for  $W_0$ .
- (ii)  $\{D_2^j \tau_k \psi\}_{j, k \in \mathbb{Z}}$  is an orthonormal basis for  $L_2(\mathbb{R})$ , i.e.,  $\psi$  is an orthonormal wavelet.

Equation (1.6) implies that  $\psi$  is a refinable function. The function  $H_1$  is called a *high-pass filter*. We also notice that  $\psi$  is defined in terms of its Fourier transform, which means that we have to apply the inverse Fourier transform in order to obtain an expression for  $\psi$ .

**Proposition 1.5.7.** [15, Proposition 8.2.8] Assume Equation (1.6) holds for some 1-periodic trigonometric polynomial  $H_1 \in L_2(0, 1)$ , i.e.,  $H_1(\gamma) = \sum_{k \in \mathbb{Z}} d_k e^{2\pi i k \gamma}$ . Then

$$\psi(x) = \sqrt{2} \sum_{k \in \mathbb{Z}} d_k \phi(2x + k), \quad x \in \mathbb{R}. \quad (1.7)$$

The results in the previous proposition imply that we can find the orthonormal wavelet  $\psi$  whenever  $H_0$  in the refinement equation is known. Moreover, in most cases of practical interest, the low-pass filter  $H_0$  is actually a trigonometric polynomial of the form

$$H_0(\gamma) = \sum_{k=-N}^N c_k e^{2\pi i k \gamma}.$$

A closer look at Equation (1.7) leads naturally to a criterion for obtaining a compactly supported orthonormal wavelet. Indeed, the function  $\psi$  is given in terms of the refinable function  $\phi$ . Thus, if  $\phi \in L_2(\mathbb{R})$  is compactly supported and  $H_0$  is a trigonometric polynomial, then the orthonormal wavelet  $\psi$  is compactly supported.

We have seen how the choice of a scaling function  $\phi$  characterizes a multiresolution analysis completely. It therefore makes sense to seek a formulation of a MRA in terms of the properties of the function  $\phi$ . Such a formulation is presented below.

**Theorem 1.5.8.** *[15, Theorem 8.2.11] Let  $\phi \in L_2(\mathbb{R})$ . Define the spaces  $V_j$  as in Definition 1.5.1. Then  $\phi$  generates a MRA if the following conditions hold:*

(i)  $\inf_{\gamma \in (-\epsilon, \epsilon)} \left| \widehat{\phi}(\gamma) \right| > 0$  for some  $\epsilon > 0$ .

(ii) The scaling Equation (1.4) is satisfied for a bounded 1-periodic function  $H_0$ .

(iii)  $\{\tau_k \phi\}_{k \in \mathbb{Z}}$  is an orthonormal system.

Condition (iii) in Theorem 1.5.8 is still a tricky one to satisfy for a function  $\phi \in L_2(\mathbb{R})$ . So it would be convenient to obtain an alternative characterization of the functions  $\phi \in L_2(\mathbb{R})$  for which  $\{\tau_k \phi\}_{k \in \mathbb{Z}}$  is an orthonormal system. Such a characterization is given below:

**Theorem 1.5.9.** [15, Theorem 8.2.12] Let  $\phi \in L_2(\mathbb{R})$ . Then  $\{\tau_k\phi\}_{k \in \mathbb{Z}}$  is an orthonormal system if and only if

$$\sum_{k \in \mathbb{Z}} \left| \widehat{\phi}(\gamma + k) \right|^2 = 1, \quad a.e. \gamma \in \mathbb{R}.$$

## 1.6 Vanishing Moments and the Wavelet Construction of Daubechies

In this section we explain why orthonormal wavelets proved to be very useful in signal processing applications, as well as present Daubechies' construction of compactly supported orthonormal wavelets. Let us assume that the orthonormal wavelet  $\psi \in L_2(\mathbb{R})$  comes from a MRA generated by the function  $\phi \in L_2(\mathbb{R})$ . Then we already mentioned in Proposition 1.5.4 that the collection of functions

$$\{\tau_k\phi\}_{k \in \mathbb{Z}} \cup \{D_2^j\tau_k\psi\}_{j \in \mathbb{N}, k \in \mathbb{Z}}$$

forms an orthonormal basis for  $L_2(\mathbb{R})$ . We therefore know that any  $f \in L_2(\mathbb{R})$  has a representation

$$f = \sum_{k \in \mathbb{Z}} \langle f, \tau_k\phi \rangle \tau_k\phi + \sum_{j=1}^{\infty} \sum_{k \in \mathbb{Z}} \langle f, \psi_{j,k} \rangle \psi_{j,k}. \quad (1.8)$$

Such a representation allows us to reconstruct the signal  $f$  based on the coefficients

$$\left\{ \langle f, \tau_k\phi \rangle \right\}_{k \in \mathbb{Z}} \cup \left\{ \langle f, \psi_{j,k} \rangle \right\}_{j \in \mathbb{N}, k \in \mathbb{Z}}.$$

However, in practical applications one cannot store an infinite sequence of nonzero numbers. So one has to select a finite number of these coefficients to represent a given signal  $f$ . Thus, the goal is to deal with the infinite sequence of coefficients in a way that only large enough coefficients are kept and the rest can be ignored. Then the function  $f$  will be hopefully well-approximated by a finite sum of only a few terms.

This is usually done by thresholding: that is, we specify an appropriate  $\epsilon > 0$  and keep only the coefficients for which  $|\langle f, \psi_{j,k} \rangle| \geq \epsilon$ . Wavelet analysis has been very successful in signal processing since we can often choose orthonormal wavelets for which a large number of the coefficients in the sequence  $\{\langle f, \psi_{j,k} \rangle\}_{j \in \mathbb{N}, k \in \mathbb{Z}}$  are small for a given signal  $f$ .

For example, recall the Haar orthonormal wavelet given by Equation (1.2) and consider the terms in the second infinite sum in Equation (1.8). We then have that

$$2^{j/2} \langle f, \psi_{j,k} \rangle = 2^j \int_{2^{-j}[k, k+1/2)} f(x) dx - 2^j \int_{2^{-j}[k+1/2, k+1)} f(x) dx.$$

The above expression implies that the coefficients in the representation of  $f$  are dependent on the behavior of  $f$  on the interval  $2^{-j}[k, k+1)$ . If the signal is continuous, then many coefficients are expected to be small, at least for larger values of  $j$ . Assume that  $f$  is discontinuous at a point  $x_0$ . Then if  $j, k$  are chosen so that  $x_0 \notin 2^{-j}[k, k+1)$ , the above argument still works. On the other hand, if  $j, k$  are chosen so that  $x_0 \in 2^{-j}[k, k+1)$ , then the coefficients can be expected to be approximately half the size of the jump

$$f(x_0^+) - f(x_0^-),$$

at least whenever  $j$  is large. Thus, large values of the coefficients that persist through the different scales  $j$  indicate discontinuities in the function  $f$ . In other words, knowledge of the coefficients allows us to detect discontinuities in a signal.

There exist orthonormal wavelets that perform much better than the Haar wavelet. The key feature turns out to be what is called their *vanishing moments*:

**Definition 1.6.1.** [15, Definition 8.3.2] Let  $N \in \mathbb{N}$ . A function  $\psi$  has  $N$  vanishing moments if

$$\int_{\mathbb{R}} x^l \psi(x) dx = 0$$

for  $l = 0, 1, \dots, N - 1$ .

It can be shown that the Haar orthonormal wavelet has only one vanishing moment. The next theorem shows that if an orthonormal wavelet has a large number of vanishing moments, then only a few coefficients will be large.

**Theorem 1.6.2.** [15, Theorem 8.3.3] (*Decay of wavelet coefficients*) Assume that the function  $\psi \in L_2(\mathbb{R})$  is compactly supported and has  $N$  vanishing moments. Then for any  $N$  times differentiable function  $f \in L_2(\mathbb{R})$  for which the  $N$ th derivative is bounded, there exists a constant  $C > 0$  such that

$$|\langle f, \psi_{j,k} \rangle| \leq C 2^{-jN} 2^{-j/2}$$

for all  $j \geq 1$  and  $k \in \mathbb{Z}$ .

The estimate above states that a high number of vanishing moments implies that the terms in the sequence  $\{\langle f, \psi_{j,k} \rangle\}$  have a fast decay rate as  $j \rightarrow \infty$ . It turns out that the problem of obtaining orthonormal wavelets with a certain number of vanishing moments can be formulated in terms of the low-pass filter  $H_0$ .

**Theorem 1.6.3.** [15, Theorem 8.3.4] Let  $\phi$  be a compactly supported scaling function associated with a MRA and let  $\psi$  be the corresponding orthonormal wavelet. Then the following are equivalent:

- (i)  $\psi$  has  $N$  vanishing moments.
- (ii) The function  $H_0$  can be factorized as

$$H_0(\gamma) = \left( \frac{1 + e^{-2\pi i \gamma}}{2} \right)^N L(\gamma)$$

for some 1-periodic trigonometric polynomial  $L$ .

Additionally, knowledge of the low-pass filter  $H_0$  determines (up to scalar multiplication) the associated refinable function  $\phi$ . Indeed, for any  $K \in \mathbb{N}$ , the scaling equation shows that

$$\widehat{\phi}(\gamma) = H_0(\gamma/2)H_0(\gamma/4) \cdots H_0(\gamma/2^K)\widehat{\phi}(\gamma/2^K).$$

One can then prove that  $\widehat{\phi}$  is continuous at zero and that

$$\widehat{\phi}(\gamma) = \lim_{K \rightarrow \infty} \left( \widehat{\phi}(\gamma/2^K) \prod_{j=1}^K H_0(\gamma/2^j) \right) = \widehat{\phi}(0) \prod_{j=1}^{\infty} H_0(\gamma/2^j).$$

The orthonormal wavelets introduced by Daubechies are the best known constructions based on the above idea. Except for  $N = 1$  (Haar), the orthonormal wavelets of Daubechies are not given by an explicit formula, but it is known that their smoothness increases with  $N$ . This construction is based on a family of polynomials given by

$$P_{N-1}(y) = \sum_{k=0}^{N-1} \binom{2N-1}{k} y^k (1-y)^{N-1-k}, \quad N \in \mathbb{N}.$$

**Theorem 1.6.4.** [15, Theorem 8.3.5] (Daubechies) *For any  $N \in \mathbb{N}$ , there exists a trigonometric polynomial  $L$  such that*

$$|L(y)|^2 = P_{N-1}(\sin^2(\pi\gamma)).$$

*With such a choice for  $L$ , the following statements hold:*

- (i) *The function  $H_0$  is associated with a MRA.*
- (ii) *With  $H_0$  as in (i), the orthonormal wavelet  $\psi$  has  $N$  vanishing moments and is supported on  $[0, 2N - 1]$ .*

The orthonormal wavelets of Daubechies were used in a variety of important applications, including signal compression (FBI fingerprint database), noise reduction, self-similarity problems, etc.

In the following section we wish to expand our universe of representation systems from bases to *frames* in Hilbert spaces. There are multiple reasons to justify such a generalization due to the inflexible character of bases. For example, it is often impossible to construct application-specific bases. Moreover, slight modifications of bases might lose the basis property completely.

## 1.7 Frames in Hilbert Spaces

Just as in the case of bases in a Hilbert space  $\mathcal{H}$ , a frame is also a collection of functions  $\{f_n\}_{n \in \mathbb{N}} \subset \mathcal{H}$  such that each  $f \in \mathcal{H}$  has a representation

$$f = \sum_{n=1}^{\infty} c_n(f) f_n$$

and the convergence of the series is with respect to the norm of  $\mathcal{H}$ . However, in the case of a basis representation of an element  $f \in \mathcal{H}$ , the coefficients  $c_n(f)$  are unique. This is not necessarily the case when it comes to frames, i.e., frames are not always bases. The concept of frames was first introduced by Duffin and Schaeffer in 1952 [24], and it took nearly 35 years before the value of frame theory was properly recognized. In 1985 Daubechies, Grossmann, and Meyer [20] proved they could use frames to find expansions of  $L_2(\mathbb{R})$  functions similar to those obtained by orthonormal bases.

**Definition 1.7.1.** [14, Definition 5.1.1] A sequence  $\{f_k\}_{k \in \mathbb{N}}$  of elements in a Hilbert space  $\mathcal{H}$  is a frame for  $\mathcal{H}$  if there exist constants  $A, B > 0$  such that every  $f \in \mathcal{H}$  satisfies

$$A\|f\|^2 \leq \sum_{k=1}^{\infty} |\langle f, f_k \rangle|^2 \leq B\|f\|^2. \quad (1.9)$$

The constants  $A, B$  are called *frame bounds*, and they are not unique since any  $A', B' > 0$  with  $A' < A$  and  $B' > B$  are also frame bounds. If  $A = B$ , then

$\{f_k\}_{k \in \mathbb{N}}$  is called a *Tight frame*, and if  $A = B = 1$ , then  $\{f_k\}_{k \in \mathbb{N}}$  is called a *Parseval frame*. A frame is called *redundant* if it is not a basis for  $\mathcal{H}$ .

It is important to understand the nuances involved in satisfying an inequality like (1.9). Specifically, for an upper bound analysis, every sequence  $\{f_k\}_{k \in \mathbb{N}} \subset \mathcal{H}$  satisfying

$$\sum_{k=1}^{\infty} |\langle f, f_k \rangle|^2 \leq B \|f\|^2 \quad (1.10)$$

for some constant  $B > 0$  is called a *Bessel family*. Thus, every frame is a Bessel family. Having the Bessel property leads to important results due to the following facts.

**Theorem 1.7.2.** [14, Lemma 3.2.1 & Theorem 3.2.3] *Let  $\{f_k\}_{k \in \mathbb{N}}$  be a sequence in  $\mathcal{H}$ .*

(a) *Suppose that  $\sum_{k=1}^{\infty} c_k f_k$  is convergent for all  $\{c_k\}_{k \in \mathbb{N}} \in \ell_2(\mathbb{N})$ . Then*

$$T : \ell_2(\mathbb{N}) \rightarrow \mathcal{H}, \quad T\{c_k\}_{k \in \mathbb{N}} := \sum_{k=1}^{\infty} c_k f_k$$

*defines a bounded linear operator. The adjoint operator is given by*

$$T^* : \mathcal{H} \rightarrow \ell_2(\mathbb{N}), \quad T^* f = \{\langle f, f_k \rangle\}_{k \in \mathbb{N}}.$$

*Furthermore,*

$$\sum_{k=1}^{\infty} |\langle f, f_k \rangle|^2 \leq \|T\|^2 \|f\|^2,$$

*for all  $f \in \mathcal{H}$ .*

(b) *The sequence  $\{f_k\}_{k \in \mathbb{N}}$  is a Bessel family if and only if*

$$T : \{c_k\}_{k \in \mathbb{N}} \rightarrow \sum_{k=1}^{\infty} c_k f_k$$

*is a well-defined bounded operator from  $\ell_2(\mathbb{N})$  into  $\mathcal{H}$  and  $\|T\| \leq \sqrt{B}$ .*

Theorem 1.7.2 shows that if one is only interested in the Bessel property and not in the value of the Bessel bound, then it suffices to establish that  $T$  is well-defined. Moreover, if  $\{f_k\}_{k \in \mathbb{N}}$  is a sequence in  $\mathcal{H}$  and  $\sum_{k=1}^{\infty} c_k f_k$  is convergent for all  $\{c_k\}_{k \in \mathbb{N}} \in \ell_2(\mathbb{N})$ , then  $\{f_k\}_{k \in \mathbb{N}}$  is a Bessel family. On the other hand, if  $\{f_k\}_{k \in \mathbb{N}}$  is a Bessel family in  $\mathcal{H}$ , then  $\sum_{k=1}^{\infty} c_k f_k$  converges unconditionally for all  $\{c_k\}_{k \in \mathbb{N}} \in \ell_2(\mathbb{N})$ .

For an interpretation of the lower frame bound, we note that Definition 1.7.1 implies that a frame for  $\mathcal{H}$  is *complete*, i.e.,

$$\overline{\text{span}}\{f_k\}_{k \in \mathbb{N}} = \mathcal{H}.$$

Indeed, assuming  $\{f_k\}_{k \in \mathbb{N}}$  is not complete, let  $\overline{\text{span}}\{f_k\}_{k \in \mathbb{N}} = W \subset \mathcal{H}$ . Then we know that  $\mathcal{H} = W \oplus W^\perp$ . So there exists a nonzero element  $w \in W^\perp$  satisfying

$$0 < A\|w\|^2 \leq \sum_{k=1}^{\infty} |\langle w, f_k \rangle|^2 = 0,$$

which is a contradiction. In general, obtaining a lower bound that satisfies the frame condition is not an easy task. Interestingly, we often have to consider sequences that are not complete in  $\mathcal{H}$  but form frames for the closed linear span of their elements.

**Definition 1.7.3.** [14, Definition 5.1.3] Let  $\{f_k\}_{k \in \mathbb{N}}$  be a sequence in  $\mathcal{H}$ . We say  $\{f_k\}_{k \in \mathbb{N}}$  is a *frame sequence* if it is a frame for  $\overline{\text{span}}\{f_k\}_{k \in \mathbb{N}}$ .

The operator  $T$  defined in Theorem 1.7.2 is called the *synthesis operator*, and its adjoint operator  $T^*$  is called the *analysis operator*. The composition operator

$$S : \mathcal{H} \rightarrow \mathcal{H}, \quad Sf = TT^*f = \sum_{k=1}^{\infty} \langle f, f_k \rangle f_k$$

is called the *frame operator*. We note that since  $\{f_k\}_{k \in \mathbb{N}}$  is a Bessel family, the series  $\sum_{k=1}^{\infty} \langle f, f_k \rangle f_k$  converges unconditionally for all  $f \in \mathcal{H}$ .

**Lemma 1.7.4.** [14, Lemma 5.1.5] Let  $\{f_k\}_{k \in \mathbb{N}}$  be a frame with frame operator  $S$  and frame bounds  $A, B$ . Then the following hold:

(a)  $S$  is bounded, invertible, self-adjoint, and positive.

(b)  $\{S^{-1}f_k\}_{k \in \mathbb{N}}$  is a frame with bounds  $B^{-1}, A^{-1}$ ; if  $A, B$  are the optimal bounds for  $\{f_k\}_{k \in \mathbb{N}}$ , then the bounds  $B^{-1}, A^{-1}$  are the optimal bounds for  $\{S^{-1}f_k\}_{k \in \mathbb{N}}$  and the frame operator for  $\{S^{-1}f_k\}_{k \in \mathbb{N}}$  is  $S^{-1}$ .

If  $\{f_k\}_{k \in \mathbb{N}}$  is a frame for  $\mathcal{H}$ , then the frame  $\{S^{-1}f_k\}_{k \in \mathbb{N}}$  is called the *dual frame* of  $\{f_k\}_{k \in \mathbb{N}}$ . The following theorem allows us to view frames as a generalization of bases and is therefore the tool that provides flexibility in the representation of functions in  $\mathcal{H}$ .

**Theorem 1.7.5.** [14, Theorem 5.1.6] Let  $\{f_k\}_{k \in \mathbb{N}}$  be a frame with frame operator  $S$ . Then

$$f = \sum_{k=1}^{\infty} \langle f, S^{-1}f_k \rangle f_k$$

for all  $f \in \mathcal{H}$ . The series converges unconditionally for all  $f \in \mathcal{H}$ .

Theorem 1.7.5 shows that frame theory is parallel to the main discussion of this chapter; For a given signal  $f \in L_2(\mathbb{R})$ , all information about  $f$  can be found in the sequence of frame coefficients  $\{\langle f, S^{-1}f_k \rangle\}_{k \in \mathbb{N}}$ . In what follows we relate the preliminary analysis of this chapter with the design of *multi-dimensional Parseval Wavelet frames* for the qualitative analysis of visual signals.

## Chapter 2

# Compactly Supported Parseval Wavelet Frames for $L_2(\mathbb{R}^s)$ .

From the early years of filter banks and wavelets, image decompositions for compression and analysis have been on the focus of many researchers. The vast majority of multi-dimensional designs was based on real-valued tensor product constructs of one-dimensional multi-scale decompositions. However, even in the early 90s, it was realized that such constructs do not seem to give optimal results, especially on curved boundaries [45, 57]. This motivated several researchers to explore non-separable (non-tensor product) designs [5, 6], or other dilation operators [45], which later led to the popular design of shearlets and curvelets ([11, 10, 22, 9, 13, 47]).

In this chapter we attempt to propose an alternative view on this problem. We develop a new method to design wavelet frames which combine the advantages of compactly supported wavelets, namely small support and vanishing moments, but also the directionality and anisotropy of curvelets and shearlets. One of the key novelties of this work is that we trade classical filter design, formulated as a problem of

solving trigonometric polynomial systems of equations in the frequency domain, for a much more computationally efficient method based on Singular Value Decomposition (SVD) (Theorems 2.2.6 and 2.3.2). This new method is the key contribution of this work.

## 2.1 Dyadic Wavelet Frames in $L_2(\mathbb{R}^s)$

We begin with a refinable function with compact support or, in other words, a function  $\phi \in L_2(\mathbb{R}^s)$  satisfying the following conditions:

- (a) The Fourier transform  $\widehat{\phi}$  is continuous in a neighborhood of the origin and  $\widehat{\phi}(0) = 1$ .
- (b) The  $\mathbb{Z}^s$ -periodic function  $\Phi = \sum_{k \in \mathbb{Z}^s} |\widehat{\phi}(\cdot + k)|^2$  is in  $L_\infty(\mathbb{T}^s)$ , the space of all measurable essentially bounded functions on  $\mathbb{T}^s$ . The *spectrum* of  $\Phi$  is denoted by  $\sigma_\phi = \{\gamma \in \mathbb{T}^s : \Phi(\gamma) \neq 0\}$ .
- (c) The function  $\phi$  is *refinable*, i.e.,  $\widehat{\phi}(2\gamma) = H_0(\gamma)\widehat{\phi}(\gamma)$  for almost every  $\gamma$  and for some  $\mathbb{Z}^s$ -periodic function  $H_0 \in L_2(\mathbb{T}^s)$  called a *low-pass filter*.
- (d) For  $v \in \mathbb{N}$  we consider a vector of refinable functions

$$\Psi = (\psi_1, \dots, \psi_v)_{i=1}^v \in L_2^{1 \times v}$$

called a *multi-wavelet* satisfying  $\widehat{\Psi}(2\gamma) = H_1(\gamma)\widehat{\Psi}(\gamma)$  for almost every  $\gamma \in \mathbb{R}^s$  and for another  $\mathbb{Z}^s$ -periodic vector-valued function  $H_1 \in L_2^{v \times 1}(\mathbb{T}^s)$  called a *high-pass filter*.

We define the dilation and translation operators on  $L_2(\mathbb{R}^s)$  by  $D_2 f = 2^{s/2} f(2\cdot)$  and  $\tau_k f = f(\cdot - k)$ ,  $k \in \mathbb{Z}^s$ , respectively. For the above selection of  $\Psi$  its corresponding

homogeneous wavelet family or affine Family  $X_\Psi$  is defined by

$$X_\Psi = \left\{ \psi_{i,j,k} = D_2^j \tau_k \psi_i : j \in \mathbb{Z}, k \in \mathbb{Z}^s, i = 1, \dots, v \right\}. \quad (2.1)$$

Additionally, for any  $j_0 \in \mathbb{Z}$  we define the *non-homogeneous Wavelet Family*  $X_{\phi, \Psi}^{(j_0)}$  by

$$X_{\phi, \Psi}^{(j_0)} = \left\{ D_2^j \tau_k \psi_i : j \geq j_0, k \in \mathbb{Z}^s, i = 1, \dots, v \right\} \cup \left\{ D_2^{j_0} \tau_k \phi : k \in \mathbb{Z}^s \right\}. \quad (2.2)$$

**Definition 2.1.1.** If there exist two positive constants  $C_1$  and  $C_2$  such that the inequality

$$C_1 \|f\|_2^2 \leq \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}^s} \sum_{i=1}^v |\langle f, \psi_{i,j,k} \rangle|^2 \leq C_2 \|f\|_2^2$$

holds for any  $f \in L_2$ , we say  $X_\Psi$  is an *affine frame* or a *homogeneous wavelet frame* for  $L_2(\mathbb{R}^s)$  and the elements of  $\Psi$  are often called *framelets*. If  $C_1 = C_2$ , then  $X_\Psi$  is called a *tight wavelet frame* and if  $C_1 = C_2 = 1$ , then  $X_\Psi$  is called a *Parseval wavelet frame* or a *Parseval framelet*.

Homogeneous wavelet frames have only theoretical interest. In applications we are more interested in non-homogeneous frames because they model an image decomposition into various fine scales and a coarse residual created by the integer translates of the refinable function.

As mentioned at the beginning of this chapter, the majority of multi-dimensional affine wavelets are orthonormal or Riesz wavelets defined as tensor products of one-dimensional multiresolution analysis wavelets. However, tensor product constructs tend to favor horizontal or vertical image characteristics and even introduce directional filtering variability depending on orientation. This fact was recognized by Kovacevic and Vetterli [45] who attempt to construct the first finite length filters for non-tensor product filter banks. Notably, different are the non-tensor product constructs of [6, 4, 25, 2, 31, 33, 30] which start from a single, compactly supported

refinable function whose integer shifts form a Riesz or an orthonormal basis [25]. General dilation matrices and properties such as compact support, decay, smoothness, symmetry and vanishing moments are explored in depth. We remark that all these constructs produce only real-valued wavelets.

An alternative way, which combines directionality and avoids the preferred horizontal and vertical filtering orientations of real-valued tensor products, is the introduction of complex-valued wavelets and frames pioneered by Kingsbury [43, 55] and more recently [37, 36].

The construction of refinable functions with stable integer shifts is all but an easy task, as the work of Cabrelli *et al.* [8] demonstrates. It is therefore quite easier to resort to plain refinable functions whose integer shifts form a Bessel family. We fully adopt this position, which breaks away from the MRA-orthodoxy. Ron and Shen [54] demonstrated that this can be done with the so-called Extension Principles with added benefits, the combination of small filter support with symmetry or anti-symmetry.

Our work is influenced by [54, 21, 32, 16]. We focus on the Unitary Extension Principle characterization of wavelet frames, which gives an elegant interpretation of framelet sets via systems of equations involving their low and high pass filters.

**Theorem 2.1.2.** (*Unitary Extension Principle*) *Let  $\phi \in L_2(\mathbb{R}^s)$  satisfy conditions (a)-(d) at the beginning of this section and let  $v \in \mathbb{N}$ . The family of functions  $X_\Psi$  given by Equation (2.1) is a Parseval Framelet for  $L_2(\mathbb{R}^s)$  if and only if there exists a complex-valued vector function  $H_1 \in L_2^{v \times 1}(\mathbb{T}^s)$  satisfying*

$$\overline{H_0(\gamma + q)}H_0(\gamma) + H_1^*(\gamma + q)H_1(\gamma) = \delta_{0,q} \quad (2.3)$$

*for all  $q \in \{0, 1/2\}^s$  and for almost every  $\gamma \in \mathbb{T}^s$ .*

**Remark 2.1.3.** We can see that the system of Equations (2.3) implies that if the first row of the *modulation matrix*

$$\begin{pmatrix} H_0(\gamma) & H_{1,1}(\gamma) & \cdots & H_{1,v}(\gamma) \\ H_0(\gamma + q_2) & H_{1,1}(\gamma + q_2) & \cdots & H_{1,v}(\gamma + q_2) \\ \vdots & \vdots & \ddots & \vdots \\ H_0(\gamma + q_{2^s-1}) & H_{1,1}(\gamma + q_{2^s-1}) & \cdots & H_{1,v}(\gamma + q_{2^s-1}) \end{pmatrix}$$

satisfies

$$|H_0(\gamma)|^2 + \sum_{k=1}^v |H_{1,k}(\gamma)|^2 = 1$$

for almost every  $\gamma \in \mathbb{T}^s$ , and if it is orthogonal to every other row, then  $X_\Psi$  forms a Parseval wavelet frame for  $L_2(\mathbb{R}^s)$  associated with  $\phi$ . Since the modulation matrix has  $2^s$  rows, we observe that we must have  $v \geq 2^s - 1$ .

Solving the Equations in (2.3) can be very challenging. Our goal is not to propose new filters and framelets but to provide a design framework through which one can create ensembles of Parseval framelets defined by sets of high pass finite-length filters. We want these filters to be a mix of well-known filters and of other custom-made ones as we wish to use them to capture edges, textures and surfaces of singularities with enough sensitivity to preselected orientations.

The use of compact support eliminates ringing artifacts commonly appearing with the use of truncated infinite-length filters and at the same time promotes sparsity and localization of convolutional response, which is important for many applications. In that regard, our gold standard is the sparsity asymptotics of continuous curvelets and shearlets [13, 46]. Both families achieve their optimal sparsity by continuously increasing the orientation resolution with scale, something our constructs are not meant to do because we use a fixed number of filters common to every scale. However, the small compact support of our framelets in space gives them an advantage that curvelets and shearlets lack, because those are compactly supported in frequency,

with the notable exception of the compactly supported shearlets developed in [44].

An entirely different approach was proposed in [1, 56] where a filter-bank precursor of directional atoms was proposed, the steerable pyramids, aiming to define rotationally covariant multi-scale transforms. In theory, rotational covariance can be realized by continuous directional transforms such as the Curvelet and Shearlet transforms. For discrete transforms this is not always obviously true or even realizable. Nonetheless, some rotational covariance can be achieved also by directional atoms as in [11, 10, 12, 28]. In this context, the rotational covariance of the representation is important because it makes feature extraction resistant to misclassification of structures due to rotations [53]. With shearlets, rotational covariance is different because different orientations are implemented by powers of the shearing matrices and not by rotations. Results in [27] may help elucidate this fact. At any rate, if frame atoms are directional and orientable ([11, 10, 12, 22, 9, 13, 1, 47]), then rotational covariance is well-approximated because the induced transforms can be thought of as good approximations of their continuous counterparts.

More recently, a very interesting *projection method* has been proposed by Han *et al.* to define framelets with small supports in various orientations [35]. We reproduce the main results of [35] in Corollary 2.2.7.

The difficulty to construct orientable frame atoms with small spatial support motivated us to seek an alternative way to construct multi-scale framelets or, more generally, atoms with this kind of support in space, oriented to have targeted filtering selectivity along a single direction selected by us from a set of several, predetermined orientations. We can increase the number of those orientations by enlarging the spatial support of the generating refinable function.

The framelet construction method with respect to isotropic dyadic dilations we introduce here is based on Theorem 2.2.6, which bears no similarity to classical wavelet constructions. The refinable functions we use are tensor products of one-dimensional spline functions, which endows  $\Psi$  with axial symmetries, sufficient smoothness and compact support. We are bound to use refinable functions whose low-pass filter coefficients are positive. The essence of our approach is that framelets  $\Psi$  can be derived from any high pass filter  $H$  as long as  $H(0) = 0$  and the support of  $H$  is contained in the support of the low-pass filter (Section 2.5). Of course, there is an associated cost for this procedure because it is rather unlikely that we can construct sets of Parseval Framelets exclusively containing the high pass filters  $H$  of our choice. The multi-wavelet  $\Psi$  will likely contain other framelets introduced by the process Theorem 2.2.6 prescribes, but as we show in Theorem 2.3.2 these auxiliary elements of  $\Psi$  may end up having negligible contributions.

The framelets we construct have similar properties with parabolic molecules [27], but unlike the latter, the number of their orientations is fixed for all scales. The orientation of parabolic molecules is defined in the frequency domain. This is not suitable for us, since our framelets have compact support in space and are not  $C^\infty$ . Directional filter banks as well as atoms with higher order directional vanishing moments were studied in [1, 56, 49, 50, 18, 17]. All of them are constructed in the frequency domain. One of our novelties is the adaptation of these concepts in the spatial domain. We also provide a characterization of the Directional Vanishing Moment (DVM) orders of wavelets and an algorithmic construction to generate wavelets with maximum DVMs. Moreover, we can customize our DVMs to be directed toward a certain orientation which does not have to coincide with the orientation of its wavelet. This helps to increase local sensitivity to wavefronts with the same orientation.

Although directionality is a frequently used term in this work, we do not attempt to define it rigorously. In fact, a careful examination of the literature reveals that other authors who use the term also avoid to do so. We invoke directionality in a descriptive manner in the sense that such directional filters or framelets have pronounced anisotropies in certain orientations, but may also have directional vanishing moments not necessarily aligned with their pronounced orientation or its normal.

This chapter is divided into four main sections. In Section 2.2 we begin our discussion with the equations of the Unitary Extension Principle (UEP), where in lieu of Harmonic Analysis, we use basic Linear Algebra to derive a method which transforms the design problem of framelets to a problem of designing Parseval frames in finite-dimensional spaces. In Section 2.3 we develop an algorithm which allows us to custom-select the orientation and other properties of the filters defining these Parseval framelets in order to achieve high spatial orientation of the resulting high pass filters. In Section 2.4 we extend our results to cover the case of arbitrary dilation matrices in the setting of adaptive multiresolution analyses [34]. Finally, in Section 2.5 we show how to include high-pass filters of our choice in the high-pass filter set defining  $\Psi$  and present several typical examples of the filter design strategies we propose based on the methods we develop in the preceding two Sections.

## 2.2 The Geometry of the Proposed Construction

This part of our work explores a sufficient condition for solving the equations of the Unitary Extension Principle, which in essence is a system of polynomial equations with a large number of degrees of freedom and therefore quite hard to solve in closed form and in a way that yields compactly supported wavelets. In what follows  $H_0$  is

assumed to be a trigonometric polynomial of the form

$$H_0(\gamma) = \sum_{k=1}^N a_{n_k} e^{2\pi i n_k \cdot \gamma}$$

for  $a_{n_k} \in \mathbb{R} \setminus \{0\}$ ,  $N > 1$ , and  $n_k \in J \subset \mathbb{Z}^s$ , i.e., the exponents of the complex exponentials in the representation of such a low-pass filter are characterized by  $s$ -dimensional vectors with integer components. We also have

$$H_0(0) = 1,$$

or equivalently  $\sum_{k=1}^N a_{n_k} = 1$ . We rewrite  $H_0$  using the factorization

$$H_0 = aw$$

where  $a$  is the  $1 \times N$  vector of coefficients

$$a = (a_{n_k})_{k=1}^N$$

and  $w \in \mathbb{C}^{N \times 1}$  is the vector-valued function of complex exponentials given by

$$w(\gamma) = \left( e^{2\pi i n_k \cdot \gamma} \right)_{k=1}^N.$$

From now on we express the high-pass filter  $H_1 \in L_2^{v \times 1}(\mathbb{T}^s)$  as

$$H_1 = Bw$$

for some  $B \in \mathbb{R}^{v \times N}$ . Using these expressions for  $H_0$  and  $H_1$  we state the main problem this Section addresses.

**Problem [A]:** *Let  $H_0 = aw$  be a low-pass filter as above. Given a natural number  $v \geq 2^s - 1$ , we want to determine (if it exists) a real matrix  $B \in \mathbb{R}^{v \times N}$  such that the  $v \times 1$  vector-valued function  $H_1 = Bw$  satisfies Equation (2.3) and so its corresponding family  $X_\Psi$  forms a Parseval framelet for  $L_2(\mathbb{R}^s)$ .*

Focusing on Problem [A], we consider  $\{m_{kt}\}_{k,t=1}^N$  to be the elements of the  $N \times N$  matrix

$$M := a^T a + B^T B \quad (2.4)$$

and we notice that Equation (2.3) can now be written as

$$\begin{aligned} \delta_{0,q} &= w^*(\gamma + q)(a^T a + B^T B)w(\gamma) \\ &= \sum_{k=1}^N m_{kk} e^{-2\pi i n_k \cdot q} + \sum_{k,t=1, k \neq t}^N m_{kt} e^{-2\pi i n_k \cdot q} e^{2\pi i (n_t - n_k) \cdot \gamma}, \end{aligned} \quad (2.5)$$

for all  $q \in \{0, 1/2\}^s$  and for almost every  $\gamma \in \mathbb{T}^s$ . The second summand on the right hand side of Equation (2.5) is a linear combination of not necessarily distinct exponentials. Specifically, the second term may consist of several monomials associated with the same exponential, which means that uniqueness of coefficients cannot be directly assumed unless all terms associated with the same exponential are grouped. This gives rise to a rather complex system of non-linear equations, even in the case where the number of unknown parameters is not large. Equation (2.5) implies that Problem [A] has a solution if we can find appropriate entries for the matrix  $B$  (hence for  $M$ ) such that for all  $\gamma \in \mathbb{T}^s$  and for all  $q \in \{0, 1/2\}^s$  the following equations are satisfied:

$$\sum_{k=1}^N m_{kk} e^{-2\pi i n_k \cdot q} = \delta_{0,q}, \quad (2.6)$$

$$\sum_{k,t=1, k \neq t}^N m_{kt} e^{-2\pi i n_k \cdot q} e^{2\pi i (n_t - n_k) \cdot \gamma} = 0. \quad (2.7)$$

We provide insight into the analysis concerning the system of (2.6) and (2.7) in Example 2.6.1, but for the purpose of this work we study the case where  $M$  is a diagonal matrix or, in other words, the case where  $m_{kt} = 0$  for  $k \neq t$ . The second summand in Equation (2.5) vanishes for all  $\gamma$  and so Equation (2.7) is always satisfied. However, the hypothesis that  $M$  is diagonal imposes the constraint  $v \geq N - 1$  as the

next Lemma indicates. In other words, the number of non-zero Fourier coefficients of the low-pass filter  $H_0$  affects the dimensionality of the high-pass filter  $H_1$ .

**Lemma 2.2.1.** *Let  $H_0 = aw$  be a low-pass filter supported on a bounded set  $J$  as above and let  $v \geq 2^s - 1$ . If  $M = (m_{kt})_{k,t=1}^N$  is a diagonal matrix as in Equation (2.4), then*

(a)  $m_{kk} > 0$  for all  $k = 1, \dots, N$ .

(b)  $v + 1 \geq N$ .

*Proof.* (a) Since all the components of the vector  $a$  in the expression of  $H_0$  are non-zero, and since the  $k$ -th element in the diagonal of  $M$ ,  $m_{kk}$ , corresponds to the square of the norm of the  $k$ -th column vector of  $\begin{pmatrix} a \\ B \end{pmatrix} \in \mathbb{R}^{(v+1) \times N}$ , we have  $m_{kk} > 0$ .

(b) If  $v + 1 < N$ , then we would have at least one element of the diagonal of  $M$  being equal to zero, which by (a) leads to a contradiction.  $\square$

In view of Lemma 2.2.1 the pursuit of solutions for Problem [A] leads to the following modified formulation:

**Problem [A']:** *Let  $H_0 = aw$  be a low-pass filter with bounded support  $J$  such that  $H_0(0) = 1$ . Given a natural number*

$$v \geq \max \{N - 1, 2^s - 1\},$$

*we want to determine the real matrices  $B \in \mathbb{R}^{v \times N}$  for which the matrix  $M$  is diagonal and Equation (2.6) is satisfied.*

We now notice that if Problem [A'] admits a solution  $B$ , then  $B$  is a solution to Problem [A] as well. However, the solutions of Problem [A] are not exhausted by

the solutions of Problem [A'] since solutions of the former arise even when  $M$  is not diagonal. With this in mind, from now on we focus on Problem [A'] and we show that all its solutions define Parseval frames in finite dimensional spaces, which in turn define high-pass filters  $H_1$  for homogeneous Parseval wavelet frames  $X_\Psi$ . Lemma 2.2.2 helps us get a good picture of the underlying geometry.

**Lemma 2.2.2.** *Let  $\alpha, c \in \mathbb{R}^{1 \times N}$ ,  $c \neq 0$  and suppose  $D \in \mathbb{R}^{v \times N}$  is such that*

(a) *the rows of  $\begin{pmatrix} \alpha \\ D \end{pmatrix}$  form a Parseval frame for  $\mathbb{R}^N$ .*

(b)  $Dc^T = 0$ .

*Then  $\alpha$  and  $c$  are collinear vectors.*

*Proof.* Let  $d_i$  denote the  $i$ -th row vector of  $D$ . Then for  $c \in \mathbb{R}^N$  our assumptions imply that

$$c = \langle \alpha, c \rangle \alpha + \sum_{i=1}^v \langle c, d_i \rangle d_i = \langle \alpha, c \rangle \alpha.$$

Hence,  $\alpha$  and  $c$  are collinear. □

**Lemma 2.2.3.** *Let  $\alpha \in \mathbb{R}^{1 \times N}$  be such that  $\|\alpha\|_2 = 1$ . Then for any  $v \geq N - 1$ , there always exists a matrix  $D \in \mathbb{R}^{v \times N}$  such that the rows of  $\begin{pmatrix} \alpha \\ D \end{pmatrix}$  form a Parseval frame for  $\mathbb{R}^N$ .*

*Proof.* We prove the statement by presenting an explicit construction of such a matrix  $D$ . Suppose  $V \in \mathbb{R}^{N \times N}$  is such that its first row vector is equal to  $\alpha$  and its columns form an orthonormal set for  $\mathbb{R}^N$ . Therefore, we can write

$$\alpha = e_1^T V$$

where  $e_1 \in \mathbb{R}^{N \times 1}$  is the first vector of the standard basis for  $\mathbb{R}^N$ . We set

$$D = \left( 0_{v \times 1} \mid U_{v \times (N-1)} \right) V$$

and assume that the columns of  $U$  form an orthonormal set. Such a matrix  $U$  exists because  $v \geq N - 1$ . Then

$$\begin{aligned} \begin{pmatrix} \alpha \\ D \end{pmatrix}^T \begin{pmatrix} \alpha \\ D \end{pmatrix} &= \alpha^T \alpha + D^T D \\ &= V^T \left( e_1 e_1^T + \begin{pmatrix} 0_{v \times 1} & | & U_{v \times (N-1)} \end{pmatrix}^T \begin{pmatrix} 0_{v \times 1} & | & U_{v \times (N-1)} \end{pmatrix} \right) V \\ &= V^T I_N V = I_N \end{aligned}$$

Hence, the columns  $\begin{pmatrix} \alpha \\ D \end{pmatrix}$  are an orthonormal set of  $\mathbb{R}^N$  and so the rows of  $\begin{pmatrix} \alpha \\ D \end{pmatrix}$  form a Parseval frame for  $\mathbb{R}^N$ .  $\square$

**Remark 2.2.4.** The conclusion of Lemma 2.2.3 stems from the fact that if  $k \geq N$  and  $A$  is a  $k \times N$  matrix whose columns form an orthonormal set of vectors in  $\mathbb{R}^N$ , then the rows of  $A$  are a Parseval frame for  $\mathbb{R}^N$ . Indeed, let  $R = \{r_1, \dots, r_k\}$  be the rows of  $A = [a_{ij}]$ . Then for every  $x \in \mathbb{R}^N$ , we have

$$\begin{aligned} \sum_{i=1}^k |\langle x, r_i \rangle|^2 &= \sum_{i=1}^k \left( \sum_{j=1}^N x_j a_{ij} \right)^2 \\ &= \sum_{i=1}^k \sum_{j=1}^N \sum_{l=1}^N x_j a_{ij} x_l a_{il} \\ &= \sum_{j=1}^N \sum_{l=1}^N x_j x_l \sum_{i=1}^k a_{ij} a_{il} \\ &= \sum_{j=1}^N x_j^2 \\ &= \|x\|^2. \end{aligned}$$

We are now ready to present the complete solution of Problem  $[A']$ .

**Proposition 2.2.5.** *Problem  $[A']$  admits a solution if and only if*

- (a)  $a_{n_k} > 0$  for all  $k = 1, \dots, N$ .

(b)  $H_0(q) = \delta_{0,q}$  for  $q \in \{0, 1/2\}^s$ .

*Proof.* Based on the statement of Problem [A'], let  $M = a^T a + B^T B$  be a diagonal matrix and let  $B$  be such that Equation (2.6) is satisfied. We define the  $1 \times N$  vector  $c = (c_{n_k})_{k=1}^N$  by

$$c_{n_k} = \frac{a_{n_k}}{\sqrt{m_{kk}}}, \quad k = 1, \dots, N$$

where  $a_{n_k}$  are the low-pass filter coefficients and we notice that  $c$  is well defined since Lemma 2.2.1 implies  $m_{kk} > 0$ . Moreover, the low-pass filter condition  $H_0(0) = 1$  gives

$$\sum_{k=1}^N a_{n_k} = \sum_{k=1}^N c_{n_k} \sqrt{m_{kk}} = 1, \quad (2.8)$$

while by Equation (2.6) for  $q = 0$  we obtain  $\sum_{k=1}^N m_{kk} = 1$ . This means that the vector  $m = (\sqrt{m_{kk}})_{k=1}^N \in \mathbb{R}^{1 \times N}$  has unit norm. Next, we note that  $M$  is diagonal if and only if there exists a  $v \times N$  matrix  $D$  such that

$$B = D \operatorname{diag}(\sqrt{m_{11}}, \dots, \sqrt{m_{NN}})$$

and the rows of  $(\begin{smallmatrix} c \\ D \end{smallmatrix}) \in \mathbb{R}^{(v+1) \times N}$  form a Parseval frame for  $\mathbb{R}^N$ . This implies that for any  $\alpha \in \mathbb{R}^N$  we have

$$\|\alpha\|_2^2 = |\langle \alpha, c \rangle|^2 + \sum_{i=1}^v |\langle \alpha, d_i \rangle|^2. \quad (2.9)$$

Applying Equation (2.9) for  $\alpha = m$  and utilizing Equation (2.8) gives  $Dm^T = 0$ . Hence, Lemma 2.2.2 implies that  $c$  and  $m$  are collinear and so  $c_{n_k} = \lambda \sqrt{m_{kk}}$ , or equivalently,  $a_{n_k} = \lambda m_{kk}$  for some  $\lambda \in \mathbb{R}$ . By Equation (2.8) we deduce

$$1 = \sum_{k=1}^N a_{n_k} = \lambda \sum_{k=1}^N m_{kk} = \lambda,$$

so  $a_{n_k} = m_{kk} > 0$  for all  $k = 1, \dots, N$  by Lemma 2.2.1. Finally, this and Equation (2.6) also imply  $H_0(q) = \delta_{0,q}$ .

Conversely, if  $(a_{n_k})_{k=1}^N$  is a sequence of positive coefficients, then  $c = (\sqrt{a_{n_k}})_{k=1}^N$  is a well-defined unit vector of  $\mathbb{R}^N$ . For  $v \geq N - 1$ , Lemma 2.2.3 implies we can always find a real matrix  $D \in \mathbb{R}^{v \times N}$  so that the rows of  $\begin{pmatrix} c \\ D \end{pmatrix}$  form a Parseval frame for  $\mathbb{R}^N$ . Then for  $B = D \operatorname{diag}(\sqrt{a_{n_1}}, \dots, \sqrt{a_{n_N}})$  we have that  $c^T c + D^T D = I_N$  is equivalent to  $a^T a + B^T B = \operatorname{diag}(a_{n_1}, \dots, a_{n_N})$ . Hence  $M$  is diagonal and  $m_{kk} = a_{n_k}$ . Then

$$\delta_{0,q} = H_0(q) = \sum_{k=1}^N a_{n_k} e^{-2\pi i n_k \cdot q} = \sum_{k=1}^N m_{kk} e^{-2\pi i n_k \cdot q}$$

and the proof is complete.  $\square$

A surprising consequence of Proposition 2.2.5 is that in order to have a solution to Problem [A'], all the Fourier coefficients of the low-pass filter must be positive. Tensor products of spline refinable functions yield low-pass filters satisfying both conditions of Proposition 2.2.5. As a result, the first of the main results of this work summarizes the preceding discussion.

**Theorem 2.2.6.** *Let  $H_0 = aw \in L_2(\mathbb{T}^s)$  be a low-pass filter with positive coefficients supported on a finite set of indices  $J$  and suppose  $H_0(q) = \delta_{0,q}$  for all  $q \in \{0, 1/2\}^s$ . Then for  $v \geq \max\{N - 1, 2^s - 1\}$  and  $c = (\sqrt{a_{n_k}})_{k=1}^N$  we have the following:*

(a) *All solutions of Problem [A'] are of the form*

$$B = D \operatorname{diag}(\sqrt{a_{n_1}}, \dots, \sqrt{a_{n_N}})$$

*where the rows of  $\begin{pmatrix} c \\ D \end{pmatrix}$  form a Parseval frame for  $\mathbb{R}^N$ .*

(b) *Such matrices  $D$  always exist.*

(c) *Any solution  $B$  of Problem [A'] defines a high-pass filter  $H_1 = Bw$  whose associated family  $X_\Psi$  forms a homogeneous compactly supported framelet for  $L_2(\mathbb{R}^s)$  and therefore is a solution of Problem [A].*

*Proof.* As we see in the proof of the converse of Proposition 2.2.5 the assumptions imposed on  $H_0$  guarantee the existence of a diagonal matrix

$$M = \begin{pmatrix} a \\ B \end{pmatrix}^T \begin{pmatrix} a \\ B \end{pmatrix}$$

whose entries satisfy

$$\sum_{k=1}^N m_{kk} e^{-2\pi i n_k \cdot q} = \delta_{0,q}.$$

Now (a) follows from the equivalence between  $M$  being a diagonal matrix and the rows of  $\begin{pmatrix} \xi \\ D \end{pmatrix}$  forming a Parseval frame for  $\mathbb{R}^N$ . (b) follows directly from Lemma 2.2.3. Lastly, for (c) we have

$$\begin{aligned} \overline{H_0(\gamma + q)} H_0(\gamma) + H_1^*(\gamma + q) H_1(\gamma) &= W^*(\gamma + q) \begin{pmatrix} a \\ B \end{pmatrix}^T \begin{pmatrix} a \\ B \end{pmatrix} W(\gamma) \\ &= \sum_{k=1}^N m_{kk} e^{-2\pi i n_k \cdot q} \\ &= \delta_{0,q}. \end{aligned}$$

Thus  $X_\Psi$  is a Parseval frame for  $L_2(\mathbb{R}^s)$ . □

The next step we take is to generalize the construction of directional frame atoms with small spatial support presented in [35, Theorem 2], where the authors use a “projection method” to create orientations in the space domain essentially projected from higher dimensional Euclidean spaces to spaces with lower dimensionality. The resulting filters act like first order finite difference operators along the orientation of the atom. Here we recreate their main result in a somewhat more general framework, specifically for low-pass filters with positive coefficients satisfying  $H_0(q) = \delta_{0,q}$  for all  $q \in \{0, 1/2\}^s$ . This result was also generalized independently in [23], where the very interesting constructs of quasi-tight framelets were also first introduced.

**Corollary 2.2.7.** Let  $H_0 = aw \in L_2(\mathbb{T}^s)$  be a low-pass filter with positive coefficients supported on a finite set  $J$  and suppose  $H_0(q) = \delta_{0,q}$  for all  $q \in \{0, 1/2\}^s$ . Then the  $N(N-1)/2 \times 1$  high-pass filter vector  $H_1$  with components

$$\sqrt{a_{n_k} a_{n_t}} \left( -e^{2\pi i n_k \cdot} + e^{2\pi i n_t \cdot} \right)$$

for all  $k \neq t$  with  $k < t$  defines an affine Parseval framelet for  $L_2(\mathbb{R}^s)$ .

*Proof.* From the definition of  $H_1$  we have  $\begin{pmatrix} H_0 \\ H_1 \end{pmatrix}(\gamma) = \begin{pmatrix} a \\ B \end{pmatrix} w(\gamma)$  where

$$\begin{pmatrix} a \\ B \end{pmatrix} = \begin{pmatrix} a_{n_1} & a_{n_2} & a_{n_3} & \cdots & a_{n_{N-1}} & a_{n_N} \\ -\sqrt{a_{n_1} a_{n_2}} & \sqrt{a_{n_1} a_{n_2}} & 0 & \cdots & 0 & 0 \\ -\sqrt{a_{n_1} a_{n_3}} & 0 & \sqrt{a_{n_1} a_{n_3}} & \cdots & 0 & 0 \\ & & & \ddots & & \\ -\sqrt{a_{n_1} a_{n_N}} & 0 & 0 & \cdots & 0 & \sqrt{a_{n_1} a_{n_N}} \\ 0 & -\sqrt{a_{n_2} a_{n_3}} & \sqrt{a_{n_2} a_{n_3}} & \cdots & 0 & 0 \\ & & & \ddots & & \\ 0 & -\sqrt{a_{n_2} a_{n_N}} & 0 & \cdots & 0 & \sqrt{a_{n_2} a_{n_N}} \\ & & & \vdots & & \\ 0 & 0 & 0 & \cdots & -\sqrt{a_{n_{N-1}} a_{n_N}} & \sqrt{a_{n_{N-1}} a_{n_N}} \end{pmatrix}$$

for  $\gamma \in \mathbb{T}^s$ . Essentially, the rows of  $B$  are generated from all possible permutations of non-zero column pairs. This implies that  $M = a^T a + B^T B$  is a diagonal matrix since the columns of  $\begin{pmatrix} a \\ B \end{pmatrix}$  form an orthogonal set of  $N$  vectors in  $\mathbb{R}^N$ . Moreover, computing the norm of the  $k$ -th column of  $\begin{pmatrix} a \\ B \end{pmatrix}$  gives

$$a_{n_k} a_{n_1} + a_{n_k} a_{n_2} + \dots + a_{n_k}^2 + \dots + a_{n_k} a_{n_N} = a_{n_k} \sum_{i=1}^N a_{n_i} = a_{n_k},$$

for all  $k = 1, \dots, N$ . Therefore,  $M = \text{diag}(a)$  and  $B$  is a solution of Problem [A'].

The result follows by Theorem 2.2.6.  $\square$

## 2.3 Wavelets with Directional Vanishing Moments and Customizable Filters.

The core message of Section 2.2 is that under the assumptions of Theorem 2.2.6 one can construct affine Parseval framelets for  $L_2(\mathbb{R}^s)$  arising from a refinable function by constructing Parseval frames for  $\mathbb{R}^N$ . This theorem not only allows us to translate the difficult problem of solving the system of equations of the UEP into the much more algorithmically tractable problem of designing Parseval frames in finite dimensions, but furthermore enables us to custom-shape the filters defining the sought framelets. For example, sparse filters, edge detection filters, filters inducing wavelets with a high order of vanishing moments, etc. are some of the high-pass filter families we know of that produce informative results in a variety of applications.

Our goal here is to propose a theoretical framework that enables us to hand-pick the high-pass filters that define a Parseval framelet. We can also impose certain directional vanishing moments to increase their sensitivity to singularities in application-specific targeted orientations. These design choices, although not the only realizable ones, are the ones that drive the filter constructs in Section 2.5. The key tool is Theorem 2.2.6, which dictates that the matrix entries of the filters  $h_{1,i}$  are determined by the rows of the sub-matrix  $D$  of  $\begin{pmatrix} c \\ D \end{pmatrix} \in \mathbb{R}^{(v+1) \times N}$ ,  $v \geq N - 1$ , whose rows form a Parseval frame for  $\mathbb{R}^N$ , and  $c$  is a given unit norm  $1 \times N$  vector with positive components defined by the Fourier coefficients of  $H_0$ .

Customizing filters that define affine multi-dimensional Parseval frames and/or selecting the number and direction of their vanishing moments is not a straightforward task. It requires the development of a number of tools which will guarantee that in every Parseval frame filter ensemble we create we maximize the number of filters with

those desirable properties. Each such filter set may have to contain a number of filters acting as a complement to the set of filters with predesigned properties in order to derive a Parseval frame. A significant amount of this Section is devoted to making their contributions and their number as small as possible (Theorem 2.3.2). In order to achieve these goals, we first need to develop certain filter design tools utilizing Theorem 2.2.6.

- (i) We begin by presenting a sufficient condition for pre-determining  $L$  rows of  $D$ , or a sub-matrix  $D_1 \in \mathbb{R}^{L \times N}$  whose rows are orthogonal to  $c$  so that there exist appropriate matrices  $D_2$  for which the rows of

$$\begin{pmatrix} c \\ D_1 \\ D_2 \end{pmatrix}$$

form a Parseval frame for  $\mathbb{R}^N$  [Lemma 2.3.1]. The sub-matrix  $D_2$  determines the filters acting as a complement to the set of customized filters defined by  $D_1$ .

- (ii) Next, we seek a technique to optimize the rows of  $D_1$  to control redundancy and simultaneously minimize the reconstruction error when we choose to omit the framelets  $\psi_i$  resulting from  $D_2$  [Theorem 2.3.2]. The algorithm implementing (i) and (ii) can be found at the beginning of Section 2.5.
- (iii) Finally, we give a characterization of the directional vanishing moment orders (DVM) of framelets but also show how one can explicitly construct wavelets with up to  $N - 1$  DVM.

The next lemma addresses (i). In this setting the affine framelets induced by the rows of  $D_1$  are predesigned but it is not necessary that they form an affine frame for  $L_2(\mathbb{R}^s)$ . From now on we use the notation  $Q := \begin{pmatrix} c \\ D_1 \end{pmatrix}$ .

**Lemma 2.3.1.** *Let  $D_1$  be a fixed  $L \times N$  matrix with rows orthogonal to  $c$ . If the singular values of  $Q$  satisfy  $\sigma_i \leq 1$  for all  $i = 1, \dots, L+1$ , then there exists an  $N \times N$  matrix  $D_2$  such that the rows of*

$$\begin{pmatrix} Q \\ D_2 \end{pmatrix} = \begin{pmatrix} c \\ D_1 \\ D_2 \end{pmatrix}$$

*form a Parseval frame for  $\mathbb{R}^N$ . In this case the Parseval frame consists of  $v = L+N+1$  vectors in  $\mathbb{R}^N$ .*

*Proof.* We prove the case where  $L+1 \leq N$ . Using Singular Value Decomposition (SVD), we have  $Q = U\Sigma_1V^T$  for  $U \in \mathbb{R}^{(L+1) \times (L+1)}$  and  $V \in \mathbb{R}^{N \times N}$  unitary matrices and

$$\Sigma_1 = \left( \text{diag}(\sigma_1, \dots, \sigma_{L+1}) \mid \mathbf{0}_{(L+1) \times (N-L-1)} \right) \in \mathbb{R}^{(L+1) \times N}.$$

Now let  $D_2 = \Sigma_2V^T \in \mathbb{R}^{N \times N}$  with

$$\Sigma_2 = \text{diag} \left( \sqrt{1 - \sigma_1^2}, \dots, \sqrt{1 - \sigma_{L+1}^2}, 1, \dots, 1 \right) \in \mathbb{R}^{N \times N}.$$

This gives

$$Q^TQ + D_2^TD_2 = V(\Sigma_1^T\Sigma_1 + \Sigma_2^T\Sigma_2)V^T = VI_NV^T = I_N.$$

The case  $L+1 > N$  is similar and the proof is omitted. □

We remark that the number of non-zero singular values of  $Q$  is directly linked to the total number  $v$  of high-pass filters. The larger the number of singular values equal to 1, the smaller the number of rows of  $\Sigma_2$  is going to be, thus providing us with a tool to control the overall redundancy of the affine family  $X_\Psi$ .

However, this is not the only notable aspect of this construction. All singular values  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{L+1}$  come from the predesigned filters induced by  $D_1$ . If

$\sigma_i = 1$  for  $i = 1, \dots, L + 1$ , then whatever complementary filters we add using  $D_2$  can be considered as the only part of the framelet construction over which we have no control, for it is determined by  $V^T$ . This observation leads us to consider (ii), the second point mentioned at the beginning of this Section.

One way to control the  $D_2$ -contributions is to eliminate the chance of introducing zeros as singular values or, in other words, by ensuring that  $\text{rank}(Q) = N$ . As we will see in Theorem 2.3.2, this can be done in a way that keeps the resulting singular values  $\sigma_i$  as close to 1 as possible. Nevertheless, this is one aspect of the  $D_2$ -construction we do not control.

The next theorem shows there exist matrices  $D_1$  for which we can jointly maximize all singular values of  $Q$  under the constraint  $\sigma_{\max}(Q) \leq 1$ . Moreover, provided that  $\text{rank}(Q) = N$ , we want to see how accurate an approximation of an  $L_2$  function  $f$  one can obtain when disregarding the completion matrix  $D_2$ . For this, recall that if  $\Psi = (\psi_1, \dots, \psi_v)$  is a multi-wavelet whose corresponding affine family  $X_{\phi, \Psi}^0$  forms a Parseval frame for  $L_2(\mathbb{R}^s)$ , then the Calderon Condition states

$$|\widehat{\phi}(\gamma)|^2 + \sum_{j=0}^{\infty} \sum_{i=1}^v \left| \widehat{\psi}_i \left( \frac{\gamma}{2^j} \right) \right|^2 = 1.$$

We define

$$E := 1 - \sum_{j=0}^{\infty} \sum_{i=1}^L \left| \widehat{\psi}_i \left( \frac{\gamma}{2^j} \right) \right|^2 - |\widehat{\phi}(\gamma)|^2 = \sum_{j=0}^{\infty} \sum_{i=L+1}^v \left| \widehat{\psi}_i \left( \frac{\gamma}{2^j} \right) \right|^2,$$

as well as the reconstruction error of  $f$

$$\begin{aligned} E(f) &:= \|f\|_{L_2}^2 - \sum_{j=0}^{\infty} \sum_{k \in \mathbb{Z}^s} \sum_{i=1}^L |\langle f, \psi_{i,j,k} \rangle|^2 - \sum_{k \in \mathbb{Z}^s} |\langle f, T_k \phi \rangle|^2 \\ &= \sum_{j=0}^{\infty} \sum_{k \in \mathbb{Z}^s} \sum_{i=L+1}^v |\langle f, \psi_{i,j,k} \rangle|^2. \end{aligned}$$

We seek to establish a connection between the reconstruction error  $E(f)$  and the simultaneously maximized singular values of  $Q$ .

**Theorem 2.3.2.** (a) Let  $c$  be a  $1 \times N$  vector such that  $\|c\|_2 = 1$  and suppose the rows of  $D_1$ ,  $\{d_i\}_{i=1}^L$ , satisfy

$$d_i c^T = 0$$

for all  $i$ . For  $\lambda \in \mathbb{R}^L$ , we define  $Q(\lambda) := (\text{diag}(\lambda) D_1)$  and

$$f_c(\lambda) := \text{trace} \left( Q^T(\lambda) Q(\lambda) \right).$$

Then the problem

$$\mathcal{P} : \begin{cases} \max f_c(\lambda) \\ \text{subject to } \|Q^T(\lambda) Q(\lambda)\| \leq 1 \end{cases}$$

admits a solution.

(b) Let  $\tilde{\lambda} \in \mathbb{R}^L$  be a solution of problem  $\mathcal{P}$  and let  $\tilde{D}_1 = \text{diag}(\tilde{\lambda}) D_1 \in \mathbb{R}^{L \times N}$  be such that  $\text{rank}(Q) = N$ . Then

$$E(f) \leq \sigma \|f\|_{L_2}^2$$

where  $\sigma := 1 - \sigma_N^2$  and the truncated non-homogeneous affine wavelet family

$$\{D_2^j T_k \psi_i : j \in \mathbb{Z}, k \in \mathbb{Z}^s, i = 1, \dots, L\} \cup \{T_k \phi : k \in \mathbb{Z}^s\}$$

is a frame with lower frame bound  $\sigma_N^2$  and upper frame bound 1.

*Proof.* (a) We define  $\Gamma = \{\lambda \in \mathbb{R}^L : \|Q^T(\lambda) Q(\lambda)\| \leq 1\}$  and notice that if  $\lambda \in \Gamma$ , then for any  $D_1 \in \mathbb{R}^{L \times N}$  with rows in the orthogonal complement of  $c$  we have

$$\sigma_{\max}(Q(\lambda)) \leq 1.$$

Moreover,  $\Gamma$  is non-empty since  $0_L \in \Gamma$ , but also bounded. Now for a sequence  $(\lambda_n)_{n \in \mathbb{N}} \subset \Gamma$  such that  $\lambda_n \rightarrow \lambda_0$  as  $n \rightarrow \infty$ , we have

$$\begin{aligned} \|Q^T(\lambda_n) Q(\lambda_n) - Q^T(\lambda_0) Q(\lambda_0)\| &= \|D_1^T (\text{diag}(\lambda_n)^2 - \text{diag}(\lambda_0)^2) D_1\| \\ &\leq \|D_1\|^2 \left\| \text{diag}(\lambda_n^2 - \lambda_0^2) \right\| \rightarrow 0 \end{aligned}$$

as  $n \rightarrow \infty$  and so  $\Gamma$  is also closed. The result follows by the continuity of the trace function  $f_c$ .

(b) Since the rows of  $\widetilde{D}_1$  are orthogonal to  $c$  and since  $\|c\|_2 = 1$ , we have  $\sigma_1 = 1$ . Then by applying Lemma 2.3.1 to  $\widetilde{D}_1$ , we have

$$\Sigma_1^T \Sigma_1 = \text{diag}(1, \sigma_2^2, \dots, \sigma_N^2)$$

and

$$\Sigma_2^T \Sigma_2 = (0, 1 - \sigma_2^2, \dots, 1 - \sigma_N^2),$$

where  $\Sigma_1$  and  $\Sigma_2$  are defined as in Lemma 2.3.1. First, we claim

$$\sum_{i=L+1}^v |H_{i,1}(\gamma)|^2 \leq (1 - \sigma_N^2) \sum_{i=1}^v |H_{i,1}(\gamma)|^2. \quad (2.10)$$

Indeed, since  $\sigma_N^2 - \sigma_i^2 \leq 0$  for all  $i = 1, \dots, N$ , we notice that the matrix

$$\begin{aligned} \mathcal{S} &:= \Sigma_2^T \Sigma_2 - (1 - \sigma_N^2) \left( \Sigma_1^T \Sigma_1 + \Sigma_2^T \Sigma_2 - \text{diag}(1, 0, \dots, 0) \right) \\ &= \text{diag}(0, -\sigma_2^2 + \sigma_N^2, \dots, -\sigma_{N-1}^2 + \sigma_N^2, 0) \end{aligned}$$

is negative semi-definite. Hence

$$w^*(\gamma) \left( M^{1/2} \right)^T V S V^T M^{1/2} w(\gamma) = \sum_{i=L+1}^v |H_{i,1}(\gamma)|^2 - (1 - \sigma_N^2) \sum_{i=1}^v |H_{i,1}(\gamma)|^2 \leq 0$$

for  $M^{1/2} = \text{diag}(\sqrt{a_{n_1}}, \dots, \sqrt{a_{n_N}})$  and for almost every  $\gamma \in \mathbb{T}^s$ . Next, let  $\theta_j : \mathbb{T}^s \rightarrow \mathbb{C}$  be given by

$$\theta_j(\cdot) = \prod_{k=0}^{j-1} H_0(2^{j-1-k} \cdot) H_1(2^j \cdot)$$

for any  $j \geq 0$ . Recall that the Fundamental Function  $\Theta : \mathbb{T}^s \rightarrow \mathbb{R}^+$  associated with the family  $X_\Psi$  is given by

$$\Theta(\cdot) = \sum_{j=0}^{\infty} |\theta_j(\cdot)|^2$$

and recall that, [21, 16], for almost every  $\gamma \in \mathbb{T}^s$  we have

$$\lim_{j \rightarrow \infty} \Theta \left( \frac{\gamma}{2^j} \right) = 1. \quad (2.11)$$

We begin by considering the error of approximation for two scales of resolution. Specifically, using (2.10) and the definition of the Fundamental function above we have

$$\begin{aligned} \sum_{i=L+1}^v \sum_{j=0}^1 \left| \widehat{\psi}_i \left( \frac{\gamma}{2^j} \right) \right|^2 &= \sum_{i=L+1}^v \left( \left| \widehat{\psi}_i(\gamma) \right|^2 + \left| \widehat{\psi}_i \left( \frac{\gamma}{2} \right) \right|^2 \right) \\ &\leq \sigma \sum_{i=1}^v \left( \left| H_{1,i} \left( \frac{\gamma}{2} \right) \right|^2 \left| \widehat{\phi} \left( \frac{\gamma}{2} \right) \right|^2 + \left| H_{1,i} \left( \frac{\gamma}{4} \right) \right|^2 \left| \widehat{\phi} \left( \frac{\gamma}{4} \right) \right|^2 \right) \\ &= \sigma \left( \left| H_1 \left( \frac{\gamma}{2} \right) \right|^2 \left| H_0 \left( \frac{\gamma}{4} \right) \right|^2 + \left| H_1 \left( \frac{\gamma}{4} \right) \right|^2 \right) \left| \widehat{\phi} \left( \frac{\gamma}{4} \right) \right|^2 \\ &= \sigma \sum_{j=0}^1 \left| \theta_j \left( \frac{\gamma}{4} \right) \right|^2 \left| \widehat{\phi} \left( \frac{\gamma}{4} \right) \right|^2 \\ &\leq \sigma \Theta \left( \frac{\gamma}{4} \right) \left| \widehat{\phi} \left( \frac{\gamma}{4} \right) \right|^2 \end{aligned}$$

for almost every  $\gamma \in \mathbb{T}^s$ . Hence if  $j_0 \in \mathbb{N}$ , proceeding inductively using the same technique yields

$$\begin{aligned} \sum_{i=L+1}^v \sum_{j=0}^{j_0} \left| \widehat{\psi}_i \left( \frac{\gamma}{2^j} \right) \right|^2 &\leq \sigma \sum_{j=0}^{j_0} \left| \theta_j \left( \frac{\gamma}{2^{j_0+1}} \right) \right|^2 \left| \widehat{\phi} \left( \frac{\gamma}{2^{j_0+1}} \right) \right|^2 \\ &\leq \sigma \Theta \left( \frac{\gamma}{2^{j_0+1}} \right) \left| \widehat{\phi} \left( \frac{\gamma}{2^{j_0+1}} \right) \right|^2 \end{aligned}$$

Finally, using (2.11) and  $\widehat{\phi}(0) = 1$  and by letting  $j_0$  tend to infinity we obtain  $E \leq \sigma$ .

The result follows from Theorem 3.2 of [40] for Parseval frames.  $\square$

### 2.3.0.1 A Characterization of Directional Vanishing Moments (DVM)

Recall that for a given unit vector  $\beta \in \mathbb{R}^s$ , we say a compactly supported wavelet  $\psi$  has  $n$  vanishing moments in the direction of  $\beta$  if

$$D_{\beta}^r \widehat{\psi}(0) = 0$$

for all  $r = 0, 1, \dots, n - 1$ , where  $D_{\beta}^r$  represents the  $r$ -th order directional derivative in the direction of  $\beta$ . A routine calculation shows

$$D_{\beta}^r \widehat{f}(0) = \mathcal{F} \left( (-2\pi i(x \cdot \beta))^r f(x) \right) (0)$$

for every compactly supported  $f \in L_1$ , where  $\mathcal{F}$  denotes the Fourier transform. The previous equation shows that DVM act just like regular moments, primarily in the direction of  $\beta$ . As in the one-dimensional case, the number of directional vanishing moments of a wavelet  $\psi$  is expected to affect the rate of decay of the frame coefficients with respect to the scale  $j$  at various directions at any point, especially at points of singularity. We illustrate this effect with Figure 2.1 below. Specifically, we consider a cubic polynomial image and the high-pass filter

$$h = \begin{pmatrix} 0.1655 & -0.2372 & 0.0718 \\ -0.0073 & 0.0146 & -0.0073 \\ -0.0207 & 0.0414 & -0.0207 \end{pmatrix}$$

corresponding to a wavelet with four DVM in the direction of  $(0, 1)$  and we notice that  $2D$  convolution with  $h$  produces an output with no edges.

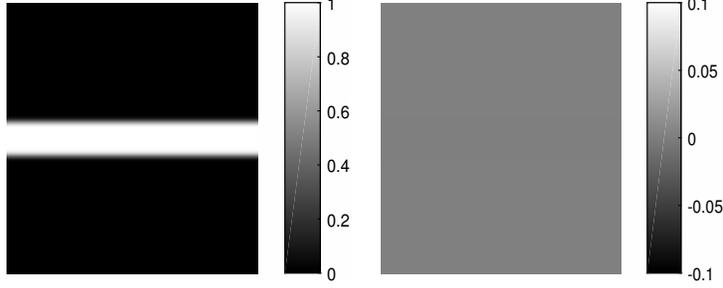


Figure 2.1: Left: Cubic spline interpolation of binary image containing a single horizontal non-zero band. Values in this image are constant in the direction of  $(0, 1)$ . Right:  $2D$  convolution with  $h$  defining a wavelet with 4 directional vanishing moments in the direction of  $(0, 1)$ . As expected, the lower polynomial degree of the intensity profile of the left panel relative to the number of DVM of  $h$  parallel to  $(0, 1)$  practically flattens the cubic spline bump in the middle of the left panel.

Next, assuming  $B \in \mathbb{R}^{v \times N}$  is a solution to Problem  $[A']$ , we translate the DVM orders of a wavelet  $\psi_i$  into certain geometric conditions in  $\mathbb{R}^N$  via the following characterization:

**Proposition 2.3.3.** *Let  $\beta \in \mathbb{R}^s$  and  $\Psi = (\psi_i)_{i=1}^v$  be a multi-wavelet arising from a matrix  $D$  as described in Theorem 2.2.6. Then if  $d_i$  denotes the  $i$ -th row vector of  $D$ , a given wavelet  $\psi_i$  has  $n$  vanishing moments in the direction of  $\beta$  if and only if*

$$cZ^r d_i^T = 0$$

for all  $r = 0, 1, \dots, n - 1$  and for  $Z := \text{diag}(\beta \cdot n_1, \dots, \beta \cdot n_N)$ .

*Proof.* Since the multi-wavelet  $\Psi$  satisfies the two-scale equation  $\widehat{\Psi}(2\cdot) = H_1(\cdot)\widehat{\phi}(\cdot)$ , we infer that  $\psi_i$  has  $n$  vanishing moments in the direction of  $\beta$  if and only if  $H_{1,i}$  has  $n$  vanishing moments in the direction of  $\beta$ , where  $H_{1,i}$  denotes the  $i$ -th component of  $H_1$ . Next, using  $D_\beta^2$  to denote the second order directional derivative in the direction

of  $\beta$ , we have

$$\begin{aligned}
D_\beta^2 (H_{1,i}(\gamma)) &= D_\beta \left( D_\beta (H_{1,i}(\gamma)) \right) \\
&= D_\beta \left( \left( \nabla \sum_{n_k \in J} \sqrt{a_{n_k}} d_{i,k} e^{2\pi i n_k \cdot \gamma} \right) \cdot \beta \right) \\
&= 2\pi i D_\beta \left( \sum_{n_k \in J} \sqrt{a_{n_k}} d_{i,k} e^{2\pi i n_k \cdot \gamma} (\beta \cdot n_k) \right) \\
&= (2\pi i)^2 \sum_{n_k \in J} \sqrt{a_{n_k}} d_{i,k} e^{2\pi i n_k \cdot \gamma} (\beta \cdot n_k)^2.
\end{aligned}$$

Proceeding inductively we find that the  $r$ -th order directional derivative in the direction of  $\beta$  is given by

$$D_\beta^r (H_{1,i}(\gamma)) = (2\pi i)^r \sum_{n_k \in J} \sqrt{a_{n_k}} d_{i,k} e^{2\pi i n_k \cdot \gamma} (\beta \cdot n_k)^r.$$

Therefore, a given wavelet  $\psi_i$  has  $n$  vanishing moments in the direction of  $\beta$  if and only if

$$\sum_{n_k \in J} \sqrt{a_{n_k}} d_{i,k} (\beta \cdot n_k)^r = 0$$

for all  $r = 0, 1, \dots, n - 1$ , or equivalently if and only if

$$cZ^r d_i^T = 0$$

for all  $r = 0, 1, \dots, n - 1$ . □

Next, based on Proposition 2.3.3, we claim that for a given set of low-pass filter polynomial exponents  $\{n_k\}_{k=1}^N$ , there exist uncountably many direction vectors for which one can construct wavelets with  $N - 1$  DVM inducing solutions to Problem [A']. The following proposition supports this claim.

**Proposition 2.3.4.** *There exists a unit vector  $\beta \in \mathbb{R}^s$  and a vector  $d \in \mathbb{R}^N$  such that the high-pass filter with coefficients  $(\sqrt{a_{n_k}} d_k)_{k=1}^N$  induces a wavelet with  $N - 1$  vanishing moments in the direction of  $\beta$ .*

*Proof.* First, we claim that there always exists a vector  $\beta \in \mathbb{R}^s$  such that all dot products

$$\beta \cdot n_k, \quad k = 1, \dots, N$$

are distinct. Equivalently, one can always find a  $\beta$  such that  $(n_k - n_t) \cdot \beta \neq 0$  for all  $k \neq t$ . Indeed, to not have  $(n_k - n_t) \cdot \beta = 0$  for some  $\beta$  and for all  $k \neq t$ , we have to exclude  $\binom{N}{2}$  hyperplanes from  $\mathbb{R}^s$ . However, by Baire's Category Theorem,  $\mathbb{R}^s$  is not the union of a finite number of hyperplanes and hence uncountably many such  $\beta$  vectors exist. Next, for such a  $\beta \in \mathbb{R}^s$  we consider the  $N \times N$  Vandermonde matrix

$$\mathcal{V} = \begin{pmatrix} 1 & \cdots & 1 \\ n_1 \cdot \beta & \cdots & n_N \cdot \beta \\ (n_1 \cdot \beta)^2 & \cdots & (n_N \cdot \beta)^2 \\ \vdots & \ddots & \vdots \\ (n_1 \cdot \beta)^{N-1} & \cdots & (n_N \cdot \beta)^{N-1} \end{pmatrix}$$

for which  $\det(\mathcal{V}) \neq 0$ , since all  $\beta \cdot n_k$ ,  $k = 1, \dots, N$  are distinct. Moreover, the matrix

$$R := \mathcal{V} \operatorname{diag}(\sqrt{a_{n_1}}, \dots, \sqrt{a_{n_N}}) = \begin{pmatrix} \sqrt{a_{n_1}} & \cdots & \sqrt{a_{n_N}} \\ (n_1 \cdot \beta) \sqrt{a_{n_1}} & \cdots & (n_N \cdot \beta) \sqrt{a_{n_N}} \\ (n_1 \cdot \beta)^2 \sqrt{a_{n_1}} & \cdots & (n_N \cdot \beta)^2 \sqrt{a_{n_N}} \\ \vdots & \ddots & \vdots \\ (n_1 \cdot \beta)^{N-1} \sqrt{a_{n_1}} & \cdots & (n_N \cdot \beta)^{N-1} \sqrt{a_{n_N}} \end{pmatrix}$$

is invertible, since  $a_{n_k} \neq 0$  and so the last column vector of  $R^{-1}$  is orthogonal to all first  $N - 1$  rows of  $R$ . Therefore, by Proposition 2.3.3, choosing  $d$  to be the last column vector of  $R^{-1}$  and applying Theorem 2.3.2(a) implies that the corresponding wavelet  $\psi$  has  $N - 1$  vanishing moments in the direction of  $\beta$ .  $\square$

**Remark 2.3.5.** Although we cannot expect the order of directional vanishing moments to exceed  $N - 1$ , the previous proposition shows that there are uncountably many direction vectors  $\beta$  for which this order of moments is realized.

## 2.4 Adaptive MRA and Compactly Supported Parseval Frames.

The dyadic multi-wavelet construction presented in the previous Sections is based on a single refinable generator  $\phi \in L_2(\mathbb{R}^s)$  corresponding to a low-pass filter  $H_0 \in L_2(\mathbb{T}^s)$ . In this Section we adopt a slightly different approach focusing on using the methodology of Section 2.2 solely on the digital setting. Our starting point is the work of Han, Kutyniok, and Shen [34], where the idea of an adaptive multiresolution analysis (AMRA) was introduced. The main objective of an AMRA is to provide a structure for a data-adapted fast decomposition strategy at each level of scaling. General affine-like systems are studied and a Unitary Extension Principle is derived for obtaining low and high-pass filters. However, this theoretical base lacks an algorithmic method for filter bank constructions. This is precisely the reason why our methods can be used to associate the filter design for such general affine-like systems with the construction of Parseval frames in finite-dimensional spaces.

The word "adaptive" refers to the fact in UEP low and high-pass filters are treated in the same way. It is therefore not necessary to use one low-pass filter only. Using a multitude of low-pass filters enables the construction of systems that achieve directionality in the discrete domain. Moreover, UEP generates filters implementable by a fast wavelet transform. In what follows we present a brief introduction of the AMRA structure and the treatment of each decomposition level. For a more detailed analysis the reader may refer to [34].

Let  $S_l, l = 1, \dots, r$  be  $s \times s$  integer matrices and let  $a_l, l = 1, \dots, r$  be  $\mathbb{Z}^s$ -periodic low-pass filter trigonometric polynomials supported over finite sets  $J_l \subset \mathbb{Z}^s$ . We consider  $J = \bigcup_{l=1}^r J_l := \{n_1, \dots, n_N\} \subset \mathbb{Z}^s$  and write the low-pass filters  $a_l$  with

respect to  $J$  as

$$a_l(\gamma) = \sum_{n_k \in J} a_{n_k}^l e^{2\pi i n_k \cdot \gamma}, \quad \gamma \in \mathbb{T}^s, \quad l = 1, \dots, r$$

where if  $n_k \notin J_l$ , we set  $a_{n_k}^l = 0$ ,  $l = 1, \dots, r$ . We also assume

$$a_l(0) \neq 0, \quad l = 1, \dots, r. \quad (2.12)$$

Equation (2.12) is the only distinction between low and high-pass filters. Once again we put in action the low-pass filter factorization

$$H_0(\gamma) = Aw(\gamma) \quad (2.13)$$

for

$$A = \begin{pmatrix} a_{n_1}^1 & \cdots & a_{n_N}^1 \\ \vdots & & \vdots \\ a_{n_1}^r & \cdots & a_{n_N}^r \end{pmatrix} \in \mathbb{R}^{r \times N}$$

and  $w(\cdot) = (e^{2\pi i n_1 \cdot}, \dots, e^{2\pi i n_N \cdot})^T$ . On the other hand, let  $b_l$ ,  $l = 1, \dots, v$  be  $\mathbb{Z}^s$ -periodic high-pass filter polynomials ( $b_l(0) = 0$ ) supported over  $J$ . We write each  $b_l$  as

$$b_l(\gamma) = \sum_{n_k \in J} b_{n_k}^l e^{2\pi i n_k \cdot \gamma}, \quad \gamma \in \mathbb{T}^s, \quad l = 1, \dots, v$$

and consider

$$H_1(\gamma) = Bw(\gamma) \quad (2.14)$$

with

$$B = \begin{pmatrix} b_{n_1}^1 & \cdots & b_{n_N}^1 \\ \vdots & & \vdots \\ b_{n_1}^v & \cdots & b_{n_N}^v \end{pmatrix} \in \mathbb{R}^{v \times N}.$$

For a multi-scale analysis in the discrete domain, let us consider  $f$  to be a signal in  $\ell_2(\mathbb{Z}^s)$ . We wish to discuss one level of decomposition of such a signal. So let  $A$  and  $B$  be the low and high frequency coefficients that we have acquired from the previous decomposition step. We then compute the next level low and high frequency

coefficients by  $f_l = T_{a_l, S_l} f$ ,  $l = 1, \dots, r$  and  $f_l = T_{b_l, S_l} f$ ,  $l = 1, \dots, v$ , respectively, where for any invertible integer matrix  $S \in \mathbb{R}^{s \times s}$  and any finitely supported sequence  $a : \mathbb{Z}^s \rightarrow \mathbb{C}$ , the transition operator  $T_{a, S} : \ell_2(\mathbb{Z}^s) \rightarrow \ell_2(\mathbb{Z}^s)$  is defined by

$$T_{a, S} f(n) := \sum_{k \in \mathbb{Z}^s} f(k) \overline{a(k - Sn)}.$$

The transition operator splits the given signal into different non-overlapping channels. We then enter the next level of analysis with the decomposition of  $f_l$ ,  $l = 1, \dots, v + r$  and we note that in the discrete domain, this type of analysis is used for a finite number of decomposition steps and is associated with a perfect reconstruction algorithm based on the reverse application of appropriate subdivision operators  $T'$  of the form

$$T'_{a, S} f(n) = |\det(S)| \sum_{k \in \mathbb{Z}^s} f(k) a(n - Sk)$$

for a suitable invertible integer matrix  $S \in \mathbb{R}^{s \times s}$  and a finitely supported sequence  $a : \mathbb{Z}^s \rightarrow \mathbb{C}$ .

**Lemma 2.4.1.** *Let  $S$  be an  $s \times s$  invertible integer matrix and  $a : \mathbb{Z}^s \rightarrow \mathbb{C}$  be a finitely supported sequence. Then*

$$\widehat{T'_{a, S} f}(\gamma) = |\det(S)| \hat{f}(S^T \gamma) \hat{a}(\gamma)$$

and

$$\widehat{T_{a, S} f}(S^T \gamma) = |\det(S)|^{-1} \sum_{q \in \Omega_S} \hat{f}(\gamma + q) \overline{\hat{a}(\gamma + q)},$$

where

$$\hat{a}(\gamma) := \sum_{k \in \mathbb{Z}^s} a(k) e^{-2\pi i k \cdot \gamma}$$

and

$$\Omega_S = \left[ (S^T)^{-1} \mathbb{Z}^s \right] \cap [0, 1)^s.$$

Han, Kutyniok, and Shen prove in [34] that a perfect reconstruction can be achieved if and only if each decomposition level is associated with a generalized version of the UEP system of equations.

**Theorem 2.4.2.** ([34]) *Let  $S_l$ ,  $l = 1, \dots, v + r$  be  $s \times s$  invertible integer matrices. Let  $a_l$ ,  $l = 1, \dots, r$  and  $b_l$ ,  $l = 1, \dots, v$  be finitely supported low and high-pass filters, respectively, and let  $H_0$  and  $H_1$  be defined as in Equations (2.13) and (2.14). Then for any  $f \in \ell_2(\mathbb{Z}^s)$  we have the perfect reconstruction*

$$\sum_{l=1}^r T'_{a_l, S_l} T_{a_l, S_l} f + \sum_{l=1}^v T'_{b_l, S_l} T_{b_l, S_l} f = f$$

*if and only if, for any  $q \in \Omega = \bigcup_{l=1}^{v+r} \Omega_{S_l}$ , where  $\Omega_{S_l} := [(S_l^T)^{-1} \mathbb{Z}^s] \cap [0, 1)^s$ ,*

$$\overline{H_0(\gamma + q)} H_0(\gamma) + \overline{H_1(\gamma + q)} H_1(\gamma) = \delta_{0,q}. \quad (2.15)$$

The proof of the above Theorem is achieved by utilizing the expressions of the Fourier transforms of the transition and subdivision operators given in Lemma 2.4.1. For more details one may refer to [34].

We notice that the generalized Unitary Extension Principle equations of Theorem 2.4.2 differ from then ones stated in Equation (2.3) only in their respective cosets defined by the sequence of dilation matrices  $S_l$ ,  $l = 1, \dots, v + r$ .

Our goal is to analyze the AMRA case of [34] using the tools presented in Proposition 2.2.5. Specifically, we note that the factorization of  $H_0$  and  $H_1$  implies that Equation (2.15) is equivalent to

$$\delta_{0,q} = \sum_{k,t=1}^N m_{kt} e^{-2\pi i n_k \cdot q} e^{2\pi i (n_t - n_k) \cdot \gamma}, \quad q \in \bigcup_{l=1}^{v+r} \Omega_{S_l}$$

where  $M = (m_{kt})_{k,t=1}^N = A^T A + B^T B$  for some appropriate  $B \in \mathbb{R}^{v \times N}$  (see Equation (2.5)). Just as in Section 2.2, we assume  $M$  is a diagonal matrix and so the above

system of equations takes the form

$$\delta_{0,q} = \sum_{k=1}^N m_{kk} e^{-2\pi i n_k \cdot q}, \quad q \in \bigcup_{l=1}^{v+r} \Omega_{S_l}. \quad (2.16)$$

As in Lemma 2.2.1,  $B \in \mathbb{R}^{v \times N}$  must satisfy  $v \geq N - r$  and  $m_{kk} > 0$  for all  $k$ . We then claim that  $M = A^T A + B^T B = \text{diag}(m_{11}, \dots, m_{NN})$ ,  $m_{kk} > 0$ ,  $k = 1, \dots, N$  if and only if there exists  $D \in \mathbb{R}^{v \times N}$ ,  $v \geq N - r$  such that

$$B = D \text{diag}(\sqrt{m_{11}}, \dots, \sqrt{m_{NN}})$$

and the rows of  $\begin{pmatrix} C \\ D \end{pmatrix} \in \mathbb{R}^{(v+r) \times N}$  form a Parseval frame for  $\mathbb{R}^N$  where

$$C = A \text{diag}(1/\sqrt{m_{11}}, \dots, 1/\sqrt{m_{NN}}).$$

Indeed, let  $Q := \begin{pmatrix} C \\ D \end{pmatrix}$ . Then  $Q^T Q = I_N$  and

$$\begin{aligned} A^T A + B^T B &= (Q \text{diag}(\sqrt{m_{11}}, \dots, \sqrt{m_{NN}}))^T (Q \text{diag}(\sqrt{m_{11}}, \dots, \sqrt{m_{NN}})) \\ &= \text{diag}(\sqrt{m_{11}}, \dots, \sqrt{m_{NN}}) Q^T Q \text{diag}(\sqrt{m_{11}}, \dots, \sqrt{m_{NN}}) \\ &= \text{diag}(m_{11}, \dots, m_{NN}). \end{aligned}$$

Additionally, Equation (2.16) gives  $\sum_{k=1}^N m_{kk} = 1$  for  $q = 0$ . This means that the vector  $m = (\sqrt{m_{11}}, \dots, \sqrt{m_{NN}})^T$  has unit norm. Then if the rows of  $Q$  form a Parseval frame for  $\mathbb{R}^N$ , then for any  $\alpha \in \mathbb{R}^N$  we have

$$\|\alpha\|^2 = \sum_{i=1}^r |\langle \alpha, c_i \rangle|^2 + \sum_{i=1}^v |\langle \alpha, d_i \rangle|^2.$$

We notice that the previous equation for  $\alpha = m$  gives

$$1 = \sum_{i=1}^r |a_i(0)|^2 + \sum_{i=1}^v |\langle m, d_i \rangle|^2.$$

A closer look at Equation (2.12) reveals that if we assume that  $a_i(0) = 1/\sqrt{r}$  for all  $i = 1, \dots, r$ , then the previous equation implies  $Dm = 0$ . We are now ready to state a generalized version of Proposition 2.2.5.

**Proposition 2.4.3.** *The system of Equations (2.16) admits a solution if*

(a)  $a_l(0) = 1/\sqrt{r}$  for all  $l = 1, \dots, r$ .

(b)  $\delta_{0,q} = \frac{1}{\sqrt{r}} \sum_{l=1}^r \overline{a_l(q)}$  for  $q \in \bigcup_{l=1}^{v+r} (S_l^T)^{-1} \mathbb{Z}^s \cap [0, 1)^s$ .

*Proof.* We look at the following configuration for the matrix  $M \in \mathbb{R}^{N \times N}$ : Let  $M = \text{diag}(m_{11}, \dots, m_{kk})$  with

$$m_{kk} = \frac{1}{\sqrt{r}} \sum_{l=1}^r a_{n_k}^l, \quad k = 1, \dots, N.$$

We then notice that the above choice implies that Equation (2.16) becomes

$$\delta_{0,q} = \sum_{k=1}^N \frac{1}{\sqrt{r}} \sum_{l=1}^r a_{n_k}^l e^{-2\pi i n_k \cdot q} = \frac{1}{\sqrt{r}} \sum_{l=1}^r \overline{a_l(q)}$$

for  $q \in \bigcup_{l=1}^{v+r} (S_l^T)^{-1} \mathbb{Z}^s \cap [0, 1)^s$ . □

## 2.5 Algorithmic Construction of Parseval Framelets

As indicated in Sections 2.2 and 2.3, the purpose of this work is to develop techniques to handcraft affine Parseval framelet sets, or at least handcraft the part of them which most significantly contributes to multidimensional image reconstructions. In this section we propose a four-step algorithmic process via which for any high-pass filter

$$H(\cdot) = (H_1(\cdot), \dots, H_L(\cdot))^T \in L_2^{L \times 1}(\mathbb{T}^s)$$

with components  $H_i(\cdot) = \sum_{k=1}^N b_{n_k}^i e^{2\pi i n_k \cdot \cdot}$ ,  $i = 1, \dots, L$ , one can force a Parseval framelet for  $L_2(\mathbb{R}^s)$  to comprise wavelets  $\psi_i$  with corresponding high-pass filters  $H_i$  (up to scalar multiplications). Using this algorithm, we construct classes of representative examples of explicit affine framelet sets containing atoms implementable by sparse

filters with directional characteristics. The algorithm below can easily be applied to every finite set of high-pass filters of our choice multiplied by an appropriate set of scalars.

Specifically, for  $n_k \in J \subset \mathbb{Z}^s$ , let  $H_0$  be a low-pass filter with positive coefficients  $a = (a_{n_k})_{k=1}^N$  and let  $H$  be any high-pass filter of the form

$$H(\cdot) = \begin{pmatrix} b_{n_1}^1 & \cdots & b_{n_N}^1 \\ \vdots & & \vdots \\ b_{n_1}^L & \cdots & b_{n_N}^L \end{pmatrix} \begin{pmatrix} e^{2\pi i n_1 \cdot} \\ \vdots \\ e^{2\pi i n_N \cdot} \end{pmatrix}$$

with  $H(0) = 0$ .

**Step 1:** We define the  $1 \times N$  vector  $c = \left(\sqrt{a_{n_k}}\right)_{k=1}^N$  and notice that for any  $\lambda \in \mathbb{R}^L$ , the matrix

$$D_1(\lambda) = \text{diag}(\lambda) \begin{pmatrix} b_{n_1}^1 & \cdots & b_{n_N}^1 \\ \vdots & & \vdots \\ b_{n_1}^L & \cdots & b_{n_N}^L \end{pmatrix} \begin{pmatrix} 1/c_{n_1} & & \\ & \ddots & \\ & & 1/c_{n_N} \end{pmatrix}$$

is well-defined and  $D_1(\lambda)c^T = 0$  since  $H$  is a high-pass filter and therefore satisfies  $\sum_{k=1}^N b_{n_k}^i = 0$  for all  $i = 1, \dots, L$ .

**Step 2:** We use Theorem 2.3.2(a) to obtain  $\lambda^*$  such that

$$\text{trace} \left( c^T c + D_1(\lambda^*)^T D_1(\lambda^*) \right) = \begin{cases} \max \text{trace} \left( c^T c + D_1(\lambda)^T D_1(\lambda) \right) \\ \text{subject to } \|c^T c + D_1(\lambda)^T D_1(\lambda)\| \leq 1. \end{cases}$$

**Step 3:** We use Lemma 2.3.1 to find a completion matrix  $D_2$  for which the rows of

$$\begin{pmatrix} c \\ D_1(\lambda^*) \\ D_2 \end{pmatrix} \in \mathbb{R}^{(v+1) \times N}, \quad v \geq N - 1$$

form a Parseval frame for  $\mathbb{R}^N$ .

**Step 4:** We use Theorem 2.2.6 to guarantee that the wavelets  $\psi_i$  with corresponding high-pass filters  $\lambda_i^* H_i$ ,  $i = 1, \dots, L$  are components of a multi-wavelet  $\Psi$  whose associated family  $X_\Psi$  is a Parseval framelet for  $L_2(\mathbb{R}^s)$ . Indeed, this follows from Theorem 2.2.6(a) since the high-pass filter matrix  $B$  is obtained by

$$B = \begin{pmatrix} D_1(\lambda^*) \\ D_2 \end{pmatrix} \text{diag}(c_{n_1}, \dots, c_{n_N})$$

**Remark 2.5.1.** (a) The cost of incorporating into  $\Psi$  the frame wavelets defined by  $\lambda_i^* H_i$  is paid in part by having to include in  $\Psi$  the filters that come from  $D_2$ . This cost can only be controlled if we select multiple high-pass filters of our choice for which we have  $\text{rank}(Q) = N$ . This particular process will become more clear in what follows.

(b) The previous algorithm demonstrates the potentially limited role of the refinable function in the construction of  $H_1$ . As we see, as long as  $H$  has enough hand-picked filters to exhaust the available dimensionality of the construction space  $\mathbb{R}^N$ , the  $D_2$ -contribution in the high-pass filter set  $H_1$  may be limited as measured by the reconstruction error  $E(\cdot)$ . Consequently, we are led to the conclusion that the significance of the refinable function is limited as the only role it seems to play is to set  $N$ .

In the spirit of the previous remark, we introduce the typical models of high-pass filter designs of our choice including high-pass filters acting as *first and second order directional finite-difference*, *Prewitt and Sobel operators*, known to produce desirable results in edge and singularity detection in two dimensional imaging applications.

## 2.6 Filter Bank Examples

We recall that first and second order directional finite-difference filters are associated with the operators  $\delta_{h,u}$  and  $\delta_{h,u}^2$ , respectively, where

$$\delta_{h,u}[f](\cdot) = f(\cdot + hu) - f(\cdot - hu),$$

and

$$\delta_{h,u}^2[f](\cdot) = f(\cdot + hu) - 2f(\cdot) + f(\cdot - hu).$$

In one dimension, the corresponding filter matrices are  $(1, 0, -1)$  and  $(1, -2, 1)$  (see [54]). Those are used to generate tensor product filters, such as the Prewitt and Sobel filters [38] given by

$$P_x = \begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix} \quad P_y = \begin{pmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

and

$$S_x = \begin{pmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{pmatrix} \quad S_y = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix},$$

respectively. Both the Prewitt and Sobel operators are used to approximate or detect horizontal and vertical intensity changes. They are obtained as tensor products of smoothing and finite-difference operators, hence they are separable. We are interested in directing the action of such operators to several orientations to promote sparse decompositions and use them in feature extraction applications. For example, we

notice that the matrices

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & -1 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & -2 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -2 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 \end{pmatrix}$$

are sparse and oriented at  $63.43^\circ$ ,  $135^\circ$ ,  $26.57^\circ$ , and  $116.57^\circ$ , respectively, but cannot be obtained as tensor products of one-dimensional kernels. This is where our algorithm comes in handy, since it permits filters like the above to be part of filter families inducing Parseval framelets. Next, we construct families of wavelet frames arising from Cardinal  $B$ -spline refinable functions, whose low-pass filters have positive coefficients.

For  $N_1 N_2 = N$ , let  $h$  be an  $N_1 \times N_2$  filter matrix. We define the map  $\Lambda : \mathbb{R}^{N_1 \times N_2} \rightarrow \mathbb{R}^N$  given by

$$\Lambda(h) = (h_{N_1,1}, \dots, h_{N_1,N_2}, h_{N_1-1,1}, \dots, h_{N_1-1,N_2}, \dots, h_{1,1}, \dots, h_{1,N_2}) \in \mathbb{R}^N$$

to turn  $h$  from a matrix to a vector, in accordance with Theorem 2.2.6. As will become clear in Examples 2.6.2, 2.6.3 and 2.6.4, we use  $\Lambda$  in the following way: first, we pre-specify the form of a desirable high-pass filter matrix, say  $h$ , and then we define

$$d(\lambda) := \lambda \left( \frac{\Lambda(h)_k}{c_{n_k}} \right)_{k=1}^N$$

for a given vector  $c = (c_{n_k})_{k=1}^N$ . We then apply Steps 2, 3 and 4 of our algorithm as stated above. When we do this for more than one filter  $h$ , then we must solve the optimization problem of Theorem 2.3.2(a). If the filters we intend to use give pairwise

orthogonal vectors through  $\Lambda$ , then the steps of the algorithm presented above can be applied to each filter individually.

The first case we examine is a high-pass filter family arising when we only apply Lemma 2.3.1 and Theorem 2.2.6. In other words, we do not pre-design any of the filters.

**Example 2.6.1.** Let  $\varphi$  be the one-dimensional second order cardinal  $B$ -spline refinable function with corresponding low-pass filter

$$\mu_0(\gamma) = \left( \frac{1 + e^{2\pi i \gamma}}{2} \right)^2 = \frac{1}{4} \left( 1 + 2e^{2\pi i \gamma} + e^{4\pi i \gamma} \right), \quad \gamma \in \mathbb{T}$$

and consider  $\phi$  to be the tensor product refinable function  $\varphi \otimes \varphi$ . Then  $H_0(\gamma) = \mu_0(\gamma_1)\mu_0(\gamma_2)$  for  $\gamma = (\gamma_1, \gamma_2) \in \mathbb{T}^2$  and the low-pass filter matrix is given by

$$h_0 = \frac{1}{16} \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}.$$

Using  $\Lambda$ , we define

$$c = \frac{1}{4} \left( 1, \sqrt{2}, 1, \sqrt{2}, 2, \sqrt{2}, 1, \sqrt{2}, 1 \right).$$

For symmetry purposes we translate  $\phi$  so as to obtain  $J = \{-1, 0, 1\} \times \{-1, 0, 1\}$ . If we merely apply the SVD method of Lemma 2.3.1 we obtain

$$B = 10^{-2} \begin{pmatrix} -8.84 & 31.8 & -1.77 & -3.54 & -7.07 & -3.54 & -1.77 & -3.54 & -1.77 \\ -6.25 & -2.5 & 23.8 & -2.5 & -5 & -2.5 & -1.25 & -2.5 & -1.25 \\ -8.84 & -3.54 & -1.77 & 31.8 & -7.07 & -3.54 & -1.77 & -3.54 & -1.77 \\ -12.5 & -5 & -2.5 & -5 & 40 & -5 & -2.5 & -5 & -2.5 \\ -8.84 & -3.54 & -1.77 & -3.54 & -7.07 & 31.8 & -1.77 & -3.54 & -1.77 \\ -6.25 & -2.5 & -1.25 & -2.5 & -5 & -2.5 & 23.8 & -2.5 & -1.25 \\ -8.84 & -3.54 & -1.77 & -3.54 & -7.07 & -3.54 & -1.77 & 31.8 & -1.77 \\ -6.25 & -2.5 & -1.25 & -2.5 & -5 & -2.5 & -1.25 & -2.5 & 23.8 \end{pmatrix}$$

We notice that the fifth column of  $B$  contains the constant terms in the generated high-pass filter polynomials. Based on this observation, we note that even though

Theorem 2.2.6 guarantees that  $B$  induces a Parseval frame for  $L_2(\mathbb{R}^2)$ , none of the high-pass filter matrices are sparse, symmetric, anti-symmetric, or directional.

SVD for the construction of the high-pass filter set was first used in [26] for proving the existence of periodic tight frame multiwavelets  $L_2([0, 2\pi]^s)$  arising from multi-refinable periodic functions. As we see, apart from generating compactly supported frame wavelets, there is essentially no luck in obtaining filters with some of the desirable properties by using SVD only.

**Example 2.6.2.** Let  $\varphi$  be an even-order cardinal  $B$ -spline refinable function and let  $\phi$  be the tensor product  $\varphi \otimes \varphi$  as before, centered at the origin. Using  $\Lambda$  and the fact that the symmetry of  $h_0$  implies  $a_{n_i} = a_{n_{N-i+1}}$  for  $i = 1, \dots, (N-1)/2$ , we define

$$Q = \begin{pmatrix} c \\ D_1 \end{pmatrix} = \begin{pmatrix} \sqrt{a_{n_1}} & \cdots & \sqrt{a_{n_{(N-1)/2}} & \sqrt{a_{n_{(N+1)/2}} & \sqrt{a_{n_{(N-1)/2}} & \cdots & \sqrt{a_{n_1}} \\ -\frac{\sqrt{2}}{2} & \cdots & 0 & 0 & 0 & \cdots & \frac{\sqrt{2}}{2} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & -\frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} & \cdots & 0 \end{pmatrix}.$$

We notice that  $D_1$  defines central-difference filters with orientations parallel to the vectors  $n_i$ ,  $i = 1, \dots, (N-1)/2$ . If  $\beta$  is an arbitrary unit vector in  $\mathbb{R}^2$ , then we write

$$cZ = \left( (n_1 \cdot \beta) \sqrt{a_{n_1}}, \dots, (n_{(N+1)/2} \cdot \beta) \sqrt{a_{n_{(N+1)/2}}, \dots, (n_N \cdot \beta) \sqrt{a_{n_1}} \right)$$

as in Proposition 2.3.3 and note that the symmetry of the vectors  $n_i$  and  $n_{N-i+1}$  about the origin implies

$$n_i \cdot \beta = -n_{N-i+1} \cdot \beta, \quad i = 1, \dots, \frac{N-1}{2}.$$

This means that if a vector belongs to the orthogonal complement of the linear span of the rows of  $Q$ , then it is automatically orthogonal to  $cZ$ . In this setting, the rows of  $Q$  are pairwise orthogonal unit vectors. Any choice of a  $D_2$  matrix for which the rows of  $\begin{pmatrix} Q \\ D_2 \end{pmatrix}$  form a Parseval frame for  $\mathbb{R}^N$  will define an affine Parseval framelet for

$L_2(\mathbb{R}^2)$ , where the  $\psi_i$  defined by the rows of  $D_2$  have exactly one directional vanishing moment for all  $\beta \in \mathbb{R}^2$ .

By Proposition 2.3.3, each of the high-pass filters generated by  $Q$  makes its corresponding wavelet insensitive to singularities parallel to  $\beta$  when  $\beta$  is perpendicular to  $n_k$ , since then the wavelet has infinite moments along these directions. In fact, by continuity of the inner product, each wavelet loses its sensitivity as  $\beta$  converges to the unit vector perpendicular to  $n_k$ .

**Example 2.6.3.** Starting with the same refinable function  $\phi$  as in Example 2.6.1, our next effort is to design  $B$  so that it is associated with four first-order and four second-order directional finite-difference high-pass filter matrices. Specifically, we consider the matrices

$$h_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, h_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix}, h_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}, h_4 = \begin{pmatrix} 0 & 0 & 0 \\ -1 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

$$h_5 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & -2 & 0 \\ 1 & 0 & 0 \end{pmatrix}, h_6 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & -2 & 0 \\ 0 & 1 & 0 \end{pmatrix}, h_7 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 1 \end{pmatrix}, h_8 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & -2 & 1 \\ 0 & 0 & 0 \end{pmatrix},$$

which we vectorize using the map  $\Lambda$  to obtain the rows of  $D_1(\lambda)$  given by  $d_k(\lambda)$ ,  $k = 1, \dots, 8$ . This gives the matrix

$$D_1(\lambda) := \text{diag}(\lambda) \begin{pmatrix} -4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 \\ 0 & -2\sqrt{2} & 0 & 0 & 0 & 0 & 0 & 2\sqrt{2} & 0 \\ 0 & 0 & -4 & 0 & 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & -2\sqrt{2} & 0 & 2\sqrt{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & -2\sqrt{2} & 4 & -2\sqrt{2} & 0 & 0 & 0 \\ 0 & 0 & -4 & 0 & 4 & 0 & -4 & 0 & 0 \\ 0 & -2\sqrt{2} & 0 & 0 & 4 & 0 & 0 & -2\sqrt{2} & 0 \\ -4 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & -4 \end{pmatrix}$$

whose rows are in the orthogonal complement of  $c$ . Here the rows of  $D_1(\cdot)$  are not pairwise orthogonal and so the largest singular value of

$$Q(\lambda) = \begin{pmatrix} c \\ D_1(\lambda) \end{pmatrix}$$

is expected to be strictly greater than 1, even in the case where the rows of  $Q$  are normalized. At this point, we invoke Theorem 2.3.2(a). Specifically, we can find an optimal  $\lambda^*$  so that  $D_1(\lambda^*)$  is a solution to

$$\begin{cases} \max \text{trace} (c^T c + D_1^T(\lambda) D_1(\lambda)) \\ \text{subject to } \|c^T c + D_1^T(\lambda) D_1(\lambda)\| \leq 1 \end{cases} .$$

We use Matlab's built-in function *fmincon* to solve this problem and obtain

$$\lambda^* = (0.0442, 0.0884, 0.0442, 0.0884, 0.0234, 0.0293, 0.0088, 0.0316),$$

but also the high-pass filter coefficients

$$B = 10^{-2} \begin{pmatrix} -17.7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 17.7 \\ 0 & -25 & 0 & 0 & 0 & 0 & 0 & 25 & 0 \\ 0 & 0 & -17.7 & 0 & 0 & 0 & 17.7 & 0 & 0 \\ 0 & 0 & 0 & -25 & 0 & 25 & 0 & 0 & 0 \\ 0 & 0 & 0 & -6.63 & 13.26 & -6.63 & 0 & 0 & 0 \\ 0 & 0 & -11.75 & 0 & 23.5 & 0 & -11.75 & 0 & 0 \\ 0 & -2.5 & 0 & 0 & 5 & 0 & 0 & -2.5 & 0 \\ -12.65 & 0 & 0 & 0 & 25.3 & 0 & 0 & 0 & -12.65 \\ 0.002 & 0 & 0.001 & 0.0003 & -0.008 & 0.0003 & 0.001 & 0 & 0.002 \\ -8.52 & 0.0288 & 9.59 & 0.233 & -2.66 & 0.233 & 9.59 & 0.0288 & -8.52 \\ 5.46 & -0.939 & 5.69 & -19 & 17.5 & -19 & 5.69 & -0.939 & 5.46 \\ 3.39 & -21.5 & 3.4 & 8.1 & 13.2 & 8.1 & 3.4 & -21.5 & 3.39 \end{pmatrix}$$

by Lemma 2.3.1 and Theorem 2.2.6. The SVD process of Lemma 2.3.1 introduces four new filters, from the lower four rows of  $B$ , in order to complete the Parseval frame for  $\mathbb{R}^9$ . Moreover, as shown in Example 2.6.2, the wavelets induced by the rows  $\{b_i\}_{i=5}^{13}$  have first-order directional vanishing moments in the direction of all  $\beta \in \mathbb{R}^2$ . If we decide to omit the four filters added by  $D_2$ , Theorem 2.3.2(b) implies that for an arbitrary function  $f \in L_2(\mathbb{R}^2)$ , we have

$$E(f) \leq (1 - \sigma_9^2) \|f\|_{L_2}^2 \approx 0.987 \|f\|_{L_2}^2.$$

Additionally, by Theorem 2.3.2(b), the family

$$\{D_2^j T_k \psi_i : j \in \mathbb{Z}, k \in \mathbb{Z}^s, i = 1, \dots, 8\}$$

is a frame, which guarantees the representation's injectivity. We also point out that, if all the row-vectors of  $D_1(\lambda)$  are pairwise orthogonal, then the optimal  $\lambda^*$  gives  $\sigma_i(D_1(\lambda^*)) = 1$  for all  $i$ . The reader may refer to [3] for a Parseval framelet induced by the first five rows of  $D_1(\lambda)$ . In that paper we also present an application of the high-pass filter matrices arising from rows 3, 4, 5 and 6 of  $B$  given by the first-order finite-difference filters

$$h_3 = 10^{-2} \begin{pmatrix} 17.7 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -17.7 \end{pmatrix} \quad h_4 = 10^{-2} \begin{pmatrix} 0 & 0 & 0 \\ -25 & 0 & 25 \\ 0 & 0 & 0 \end{pmatrix}$$

and the second-order finite-difference filters

$$h_4 = 10^{-2} \begin{pmatrix} 0 & 0 & 0 \\ -6.63 & 13.26 & -6.63 \\ 0 & 0 & 0 \end{pmatrix} \quad h_6 = 10^{-2} \begin{pmatrix} -11.75 & 0 & 0 \\ 0 & 23.5 & 0 \\ 0 & 0 & -11.75 \end{pmatrix}$$

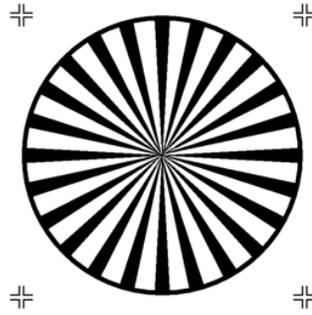


Figure 2.2: This is a 256x256 image freely available with Matlab 2017. We use it to demonstrate the interaction of the designed filters with singularities in various directions.

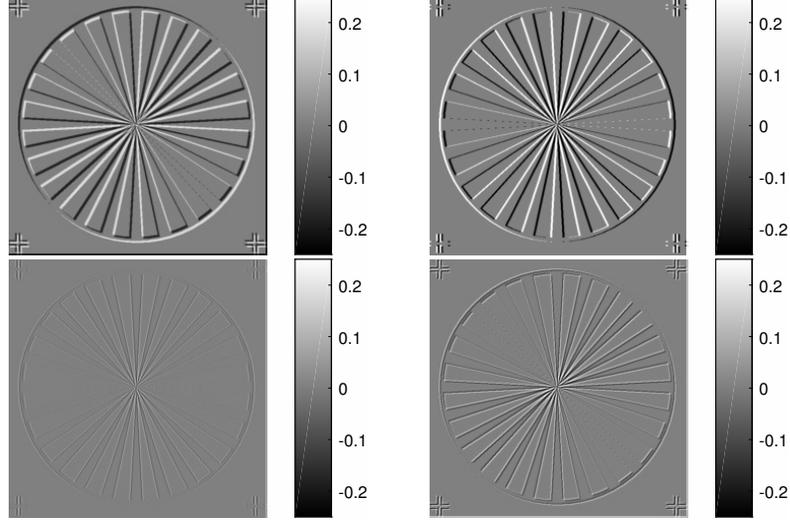


Figure 2.3: Application of  $h_i$ ,  $i = 3, 4, 5, 6$  constructed in Example 2.6.3 as discrete 2D-convolution kernels at native resolution. The first two filters act as first-order directional central-difference filters oriented at  $135^\circ$  and  $0^\circ$ , respectively. The last two act as second-order central-difference filters oriented at  $0^\circ$  and  $135^\circ$ , respectively. Note that singularity detection strength increases as edges are oriented closer to being perpendicular to the orientation of each filter. In Fig 2.5 we see that this effect may also be related to the anisotropy of the filter and its size.

**Example 2.6.4.** We consider the fourth order Cardinal  $B$ -spline refinable function

$$\varphi(x) = \begin{cases} \frac{1}{6}x^3, & 0 < x \leq 1 \\ \frac{1}{6}(-x^3 + 12x^2 - 18x + 8), & 1 < x \leq 2 \\ \frac{1}{6}(-x^3 - 12x^2 + 78x - 88), & 2 < x \leq 3 \\ \frac{1}{6}(x^3 - 48x + 128), & 3 < x \leq 4 \end{cases}$$

with corresponding low-pass filter

$$\mu_0(\gamma) = \left( \frac{1 + e^{2\pi i \gamma}}{2} \right)^4 = \frac{1}{16} \left( 1 + 4e^{2\pi i \gamma} + 6e^{4\pi i \gamma} + 4e^{6\pi i \gamma} + e^{8\pi i \gamma} \right),$$

and we set  $\phi$  to be the tensor product  $\varphi \otimes \varphi$ . Then  $H_0(\gamma) = \mu_0(\gamma_1)\mu_0(\gamma_2)$ , the low-pass filter matrix is given by

$$h_0 = \frac{1}{64} \begin{pmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{pmatrix}$$

and  $c$  takes the form

$$c = \frac{1}{16} \left( 1, 2, \sqrt{6}, 2, 1, 2, 4, 2\sqrt{6}, 4, 2, \sqrt{6}, 2\sqrt{6}, 6, 2\sqrt{6}, \sqrt{6}, 2, 4, 2\sqrt{6}, 4, 2, 1, 2, \sqrt{6}, 2, 1 \right).$$

Centering  $\phi$  at the origin implies  $J = \{-2, \dots, 2\} \times \{-2, \dots, 2\}$ . We use our algorithm to create filters with different orientations from those along which their corresponding finite-difference kernels act. More specifically, we consider first and second-order filters of the form

$$\begin{pmatrix} 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & 1 \\ 0 & -1 & 0 & 1 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 & 0 \end{pmatrix} \quad \begin{pmatrix} -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & -2 & 1 \\ 0 & 1 & -2 & 1 & 0 \\ 1 & -2 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & -2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 \end{pmatrix}$$

First, with this new design approach we mimic one of the popular properties of curvelets and shearlets: We define filters that act as singularity detectors perpendicularly to the local orientation of a wave-front. Since our design is limited within  $J$ , the discreteness of this spatially limited integer sub-grid constrains our ability to direct the action of the associated differential operator perpendicularly to the filter's orientation. Moreover, the smaller number of bands of the filter matrix relative to the length along its orientation seems to better focus the direction of its action (see Fig. 2.5). This is something we also observe to a greater degree with shearlets and curvelets, because they are designed in the frequency domain where one can control their shape more easily.

The prototype of each of the two classes of the filters we design in this example is directed along the  $x$  or  $y$  axis. The third and seventh matrices above are the prototype filters for the first and second order directional central difference operators acting along the  $x$  direction. Both filters have vertical orientation. To switch these filters to another orientation, we re-position their central band by selecting one-by-one the lead point of the central band on the  $x$  and  $y$ -axis of the grid as shown in Figure 2.4.

This process gives a filter bank with 24 high-pass filters with hand-picked orientations. Next, SVD adds 24 more to complete a Parseval frame. The full list of all 48 filters of this example and of Example 2.6.3 can be found in the supplementary file which can be retrieved from [github.com/nkarantzas/multi-d-compactly-supported-PF](https://github.com/nkarantzas/multi-d-compactly-supported-PF) along with the codes used for the generation of the presented filter-banks.

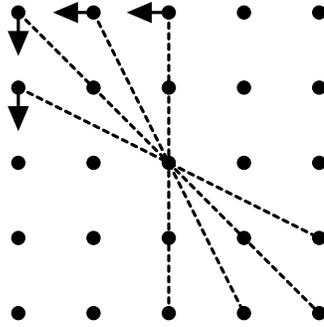


Figure 2.4: The dashed lines show four successive positions of central bands defining this predefined filter set. Once the central band has been set, we choose its nearest diametrically opposite bands to create all first and second-order finite difference filters allowed by this process.

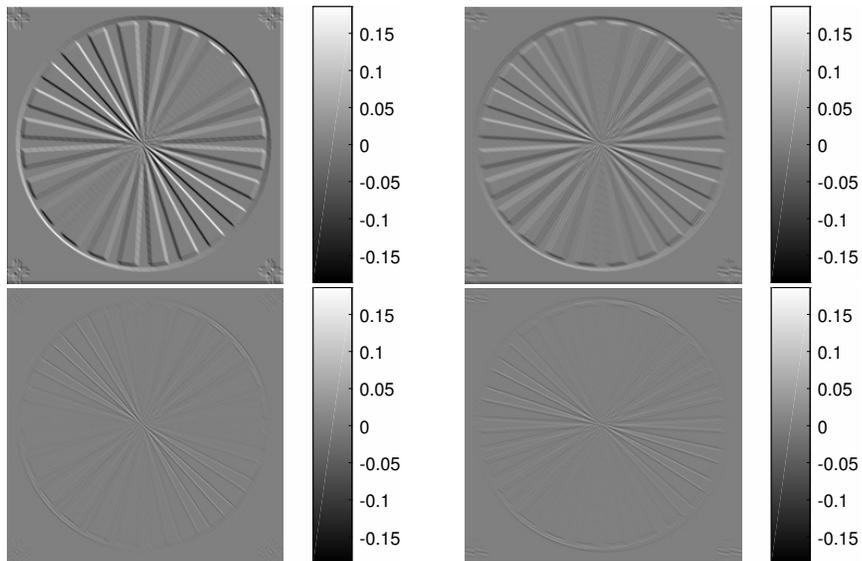


Figure 2.5: Application of  $h_i$ ,  $i = 5, 6, 17, 18$  at native resolution. The first two convolutions correspond to filters with orientations at  $135^\circ$  and  $153.43^\circ$ , respectively. The last two convolutions correspond to filters with orientations at  $135^\circ$  and  $153.43^\circ$ , respectively.

**Example 2.6.5.** As promised in Section 2.2, we illustrate the geometric implications

and complexities of solving the system of Equations (2.6) and (2.7). Equation (2.7) is relevant only when  $M$  is not a diagonal matrix. Recall that our analysis in Sections 2.2 and 2.3 is based on  $M$  being diagonal. To avoid computational complications, we consider the one-dimensional case, i.e.,  $s = 1$ . Without loss of generality, we assume  $\{n_k\}_{k=1}^N$  are consecutive integers. Then

$$\left\{ \begin{array}{l} e^{2\pi i(n_2-n_1)\gamma} = e^{2\pi i(n_3-n_2)\gamma} = \dots = e^{2\pi i(n_N-n_{N-1})\gamma} \\ e^{2\pi i(n_3-n_1)\gamma} = e^{2\pi i(n_4-n_2)\gamma} = \dots = e^{2\pi i(n_N-n_{N-2})\gamma} \\ \vdots \\ e^{2\pi i(n_{N-1}-n_1)\gamma} = e^{2\pi i(n_N-n_2)\gamma}. \end{array} \right.$$

The above equalities indicate that by rearranging and regrouping the monomials in (2.7) with respect to a fixed-valued  $n_t - n_k$ , we conclude that Equation (2.7) is satisfied if and only if

$$\sum_{k=1}^{N-t} m_{k,k+t} e^{-2\pi i n_{k+t} q} = 0,$$

for all  $t = 1, \dots, N - 1$ , which along with Equation (2.6) gives a full characterization of the problem.

However, even though the above equation indicates there is a relationship between the elements of the  $j$ -th off-diagonal of the matrix  $M$ , it does not provide us with any insight on the dimension of the desired high-pass vector, or a definite way of acquiring it.

For example, in the setting of the classical construction of orthonormal wavelets, let  $H_0$  be a low-pass filter with 4 coefficients given by  $a = [a_1, a_2, a_3, a_4]$  and  $H_1$  be a high-pass filter with coefficients  $B = [b_1, b_2, b_3, b_4]$ . Since  $M = a^T a + B^T B$  is

symmetric, the previous system of equations is equivalent to

$$\left\{ \begin{array}{l} m_{11} + m_{22} + m_{33} + m_{44} = 1, \\ m_{11} - m_{22} + m_{33} - m_{44} = 0, \\ m_{12} + m_{23} + m_{34} = 0, \\ m_{12} - m_{23} + m_{34} = 0, \\ m_{13} + m_{24} = 0, \\ m_{13} - m_{24} = 0, \\ m_{14} = 0, \end{array} \right.$$

from which we deduce  $m_{13} = m_{14} = m_{23} = m_{24} = 0$  and  $m_{12} = -m_{34}$ . Now let  $v_k \in \mathbb{R}^2$ ,  $k = 1, 2, 3, 4$  be the column vectors of

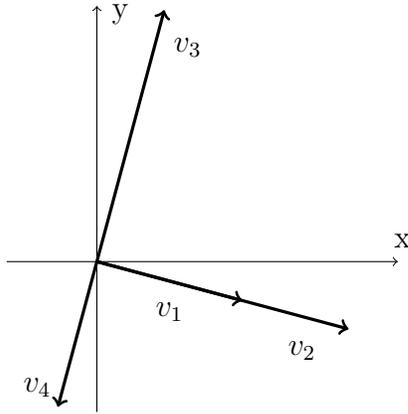
$$\begin{pmatrix} a \\ B \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 \\ b_1 & b_2 & b_3 & b_4 \end{pmatrix}.$$

Then the above linear system suggests

- $v_1$  is orthogonal to  $v_3$  and  $v_4$ , and  $v_2$  is orthogonal to  $v_3$  and  $v_4$ . Hence  $m_{12} \neq 0$ ,  $v_1 \parallel v_2$  and  $v_3 \parallel v_4$ .
- Finally, since  $m_{12} = -m_{34}$ , if  $v_1$  and  $v_2$  are parallel,  $v_3$  and  $v_4$  must be anti-parallel and vice versa.

This analysis indicates that the vectors  $v_k$  can only form a capital T-shaped configuration as indeed they do, for example in the Daubechies  $D4$  case [19] where the corresponding matrix  $\begin{pmatrix} a \\ B \end{pmatrix}$  is given by

$$\begin{pmatrix} a \\ B \end{pmatrix} = \frac{1}{8} \begin{pmatrix} 1 + \sqrt{3} & 3 + \sqrt{3} & 3 - \sqrt{3} & 1 - \sqrt{3} \\ 1 - \sqrt{3} & \sqrt{3} - 3 & 3 + \sqrt{3} & -1 - \sqrt{3} \end{pmatrix}.$$



Finally, we notice that if one wants to have additional high-pass filters or increase the length of the filters, the number of degrees of freedom increases significantly and the problem of maintaining a geometric intuition of the underlying properties becomes more complex. Moreover, we note that in the case of a four non-zero coefficient low-pass filter, we cannot have only non-negative coefficients.

## Chapter 3

# A Neural Network Application to the Kaggle Quick, Draw! Data Set

The final chapter of this dissertation is devoted to the application of the filter bank constructions of Chapter 2 in Deep Convolutional Neural Networks (CNN). This work is a collaborative effort with Mohamadkazem Safaripoorfatide, Mozahid Haque, and Saeed Sarmadi, all fellow doctoral students of the University of Houston. All of us toiled in the planning, preprocessing, and coding stages of this two month project, and special thanks are given to Kazem for developing the initial code implementation of the results presented in Chapter 2. The code for the implementation of what follows can be found at [github.com/nkarantzas/quick\\_draw](https://github.com/nkarantzas/quick_draw).

### 3.1 Project Description

Our project is based on an image classification competition posted on Kaggle. The specific challenge is the Quick, Draw! Doodle Recognition challenge ([www.kaggle.com](http://www.kaggle.com)).

`com/c/quickdraw-doodle-recognition`). The data set consists of 45,512,752 hand drawn images representing 340 classes. The task is to outperform the existing Quick, Draw! classifier. This would have potential impact on Handwriting Recognition along with Optical Character Recognition systems, Automatic Speech Recognition, and Natural Language Processing. Our strategy is twofold: We initially train architectures known to perform well on large data sets (ResNet18, ResNet34, ResNet50) to achieve a high enough Leaderboard score on the competition’s ranking page. We then used a combination of the ideas proposed in [42] (see Section 3.5.2) and the filter bank constructions of Chapter 2 to see if we could get comparable results to state-of-the-art architectures as well as to high-ranked algorithms developed by other contestants. The utilization of the directional filter banks of Chapter 2 is related to the nature of our data set, which consists solely of images characterized by geometric features (edges, ridges, etc.). This is also inextricably linked to the choice of residual networks (ResNet) because their special architecture allows for informative feature extraction at every level of decomposition (see Section 3.5.1).

## 3.2 The Data Set

The Quick, Draw! data set consists of 45,512,752 training images of certain objects (airplanes, hamburgers, hockey puck, etc.) belonging to 340 classes. They are given in a \*.csv format along with 112,199 test images. There are two versions of each of these sets: a simplified version and a raw version. The raw version contains additional information such as the timing of each stroke drawn by the person who sketched the image. Each stroke is given by sampled coordinates  $(x, y)$  that would need to be interpolated to give the original drawing. In the simplified version, unnecessary points

are removed from vector information (e.g., a straight line may have been originally recorded with 8 points, but since you only need 2 points to uniquely identify a line, 6 points can be dropped). We opted to use the simplified version due to limited computational resources and modest initial ambitions. The columns of the training data set are as follows:

- Country Code: two letter abbreviation of a drawer's country of origin;
- Drawing: a digitized object given in the form of an array with shape (stroke,  $x$ ,  $y$ ) where
  - (a) Stroke: a particular stroke drawn by the user without lifting their drawing stylus,
  - (b)  $x$ : the  $x$  coordinate of a subsample of points that can be interpolated to form a single stroke, and
  - (c)  $y$ : the  $y$  coordinate of a sub sample of points that can be interpolated to form a single stroke;
- Key ID: a unique identifier used by Quick, Draw! to catalogue the images;
- Recognized: a boolean consisting of whether the drawing matched the prompt given to the user to draw out;
- Time-stamp: date and time of the drawing; and
- Word: the word that describes the prompt that the user must draw (e.g., airplane, jail, hockey puck).

We ended up using just the drawing and word columns for training. One thing to note is that the  $(x, y)$  data given in the drawing column are a sample of the drawing

data which we are meant to interpolate to get the drawing data. Clearly, we could not work with this discrete set of data points as the convolutions in a deep CNN would zero out the data as it progresses through the network. Here are some examples of just drawing the discrete set of points.

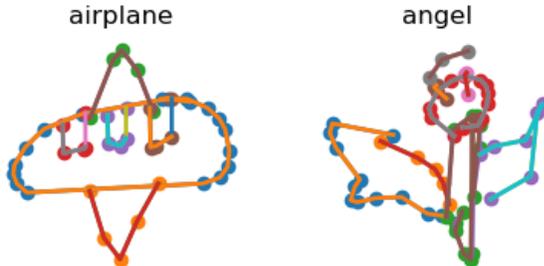


Figure 3.1: Strokes belonging to 2 classes drawn by linear interpolation.

We also note that the size of the data set is of major concern. The training data on its own was 7.6 gigabytes compressed and 27 gigabytes uncompressed. This is prior to drawing the images which would compound the size very quickly.

### 3.3 Scoring

The trained networks give the top three predictions for each image in the test set provided by the competition. After submitting our predictions, a Mean Average Precision @3 (MAP@3) is calculated by

$$MAP@3 = \frac{1}{N} \sum_{k=1}^N \sum_{s=1}^{\min(n,3)} P(s)$$

where  $N$  is the number of scored drawings in the test data,  $n = 340$  (number of predictions per drawing), and  $P(s)$  is the precision at cutoff  $k = 3$  assigned by the

classifier. In this case, our final layer is a softmax layer defined by

$$\text{softmax}(x)_i = \frac{e^{x_i}}{\sum_{j=1}^{340} e^{x_j}}, \quad i = 1, \dots, 340, \quad x = (x_1, \dots, x_{340}) \in \mathbb{R}^{340}.$$

For an instance  $x$  in the test set, let  $\text{Network}(x) \in \mathbb{R}^{340}$  denote the output of the softmax layer. We extract the 3 largest components of  $\text{Network}(x)$  and sort them from largest to smallest. Suppose these sorted components are associated with Classes  $s, t$  and  $u$  for  $s, t, u = 1, \dots, 340$ . Then

$$\text{score}(x) = \begin{cases} 1, & \text{if } x \in \text{Class}(s) \\ 1/2, & \text{if } x \in \text{Class}(t) \\ 1/3, & \text{if } x \in \text{Class}(u) \\ 0, & \text{otherwise.} \end{cases}$$

Then  $\text{MAP@3} = \frac{1}{N} \sum_{i=1}^N \text{score}(x_i)$ .

## 3.4 Preprocessing

Since our focus is on training a convolutional neural network, we have to come up with an efficient strategy for drawing the sketches from their under-sampled versions to generate the image arrays for our training and validation sets. However, the sheer size of the data set prevents us from reading from the interpolated versions of the images since that would require around 7 terabytes of space. So we opt to draw the images in batches during training. We use gray-scale images of size  $128 \times 128$  due to the depth of our chosen models. An important aspect of drawing the images using the raw strokes is that we are able to encode the timeline of strokes of each drawing, starting with 255, by assigning different intensity values to each stroke. For stroke  $t$ ,

its assigned intensity is given by  $255 - 2 \min(t, 10)$ . A normalization factor of  $1/255$  is applied to the images since the original data is given with intensities ranging from 0 to 255. Examples of these images are given in the figure below.

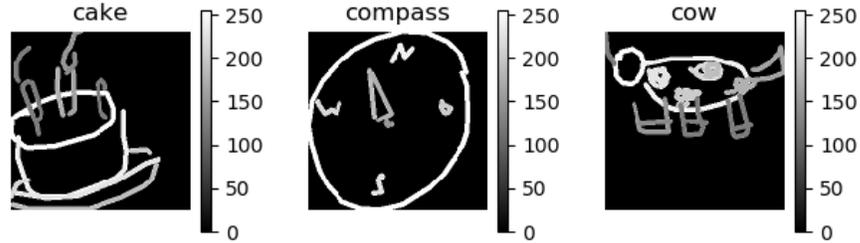


Figure 3.2: Drawing images from the sampled strokes for a variety of classes. Note that the images are not binary. There is a difference in intensity values from stroke to stroke.

## 3.5 Models

We train two different types of residual networks, namely standard residual networks [39] (ResNet18, 34, 50), but also a residual network based on convolutional kernels obtained as linear combinations of fixed Parseval frame filter banks [42].

### 3.5.1 ResNet

Deep residual neural nets [39] were introduced by He, Zhang, Ren, and Sun in 2015 to address accuracy saturation problems in Dense Convolutional Networks. It is known that the need to build deep networks for classification is unfortunately coupled with the notorious issue of vanishing gradients. As the gradient is back-propagated to earlier layers of the network, repeated multiplication operations might result in

very small gradients. This has evident adverse effects as performance gets saturated and often starts degrading rapidly. In ResNet architectures, residuals are learned as opposed to features. Residuals are subtractions of features learned from inputs of a particular layer by utilizing shortcut connections between the input of layers  $n$  and  $n + k$ . These shortcuts define residual blocks such as the one shown below.

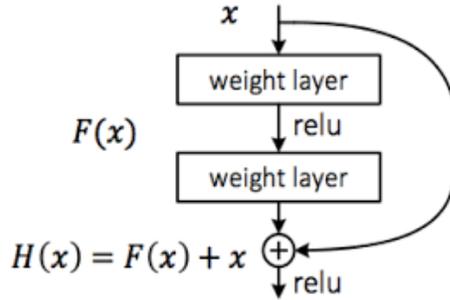


Figure 3.3: A residual block used for propagation of information over multiple layers. The image is taken from [39].

The rectified linear unit function is defined as

$$\text{relu}(x) = \begin{cases} x, & x > 0 \\ 0, & \text{otherwise.} \end{cases}$$

The motivation behind replacing the classical feature extraction block  $F$  with  $F + Id$  comes from the observation that a successive composition of feature extraction blocks followed by non-linearities often diminishes gradients during back propagation. One of the main arguments of [39] is that this type of architecture avoids these gradual degradations of gradients. For the purpose of this work we chose to work with the three smallest ResNet models, namely the ones with 18, 34, and 50 convolutional layers, respectively.

### 3.5.2 ResNetRF

We train our final model implementing the idea of Structured Receptive Fields (SRF) in CNNs [42]. This idea can be traced back to Mallat and Bruna [7] who introduced Invariant Scattering Convolution Networks to compute invariant image representations, stable to deformations and preserving high frequency information for classification. The basic ingredient of a SRF network is a fixed linear basis of 2D filters out of which the network learns the coefficients of their linear combinations. Specifically, for a fixed basis filter bank  $H = \{h_i\}_{i=1}^m \in \mathbb{R}$ , the convolutional node  $N$  is defined by

$$\left( \sum_{i=1}^m \lambda_i h_i \right) * I_N,$$

where  $I_N : \mathbb{Z}^2 \rightarrow \mathbb{R}$  denotes the input of the layer and  $\{\lambda_i\}_{i=1}^m$  are the scalar trainable parameters. The goal here is twofold: qualitative and quantitative. Qualitatively, when training data is limited, one can utilize predesigned priors into the model, such as tuned hyperparameters or even fully engineered representations like Scattering transforms. Quantitatively, it is possible to significantly reduce the number of trainable parameters which is a common practice to avoid over-fitting.

Withing the SRF framework, we test the effectiveness of the predesigned directional Parseval frame filter banks of Chapter 2 on the Quick Draw data set for all three ResNet models.

### 3.5.3 Model Architectures

We present the three ResNet architectures below. As mentioned above, they consist of bundled residual blocks and block-to-block shortcut identity mappings. Their residual blocks are defined as follows:

Table 3.1: (Left) Basic residual block. (Right) Bottleneck residual block.

<b>ResNet18 &amp; ResNet34</b>	<b>ResNet50</b>
3 × 3 Convolutional Layer	1 × 1 Convolutional Layer
Batch Normalization	Batch Normalization
relu	relu
3 × 3 Convolutional Layer	3 × 3 Convolutional Layer
Batch Normalization	Batch Normalization
relu	relu
	1 × 1 Convolutional Layer
	Batch Normalization
	relu

We note that ResNet50 differs from the other two in that it contains  $1 \times 1$  convolutions as well. This operation was first proposed in [48] and it was used by Google in the inception module as a means to reduce the layer input dimensionality across channels.

Batch normalization is a speed, performance, and stability improving technique used in artificial neural networks by normalizing the input layer and adjusting and scaling the activations appropriately [41]. We denote the mean and variance of a training set batch  $B$  consisting of  $n$  images  $x_1, \dots, x_n$  by

$$\mu_B = \frac{1}{n} \sum_{i=1}^n x_i$$

and

$$\sigma_B^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_B)^2.$$

Then each input  $x_i$  of a layer is normalized by

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}, \quad i = 1, \dots, n.$$

The  $\epsilon$  in the denominator is added for numerical stability and can be arbitrarily small.

An affine transformation step follows as

$$y_i = \gamma \hat{x}_i + \beta,$$

where the parameters  $\gamma$  and  $\beta$  are subsequently learned in the optimization process.

Max and Average pooling layers perform down-sampling by dividing the input into  $k \times l$  rectangular pooling regions and computing the max and average values of each region, respectively. Table 3.2 shows the full network architectures of all three networks.

Table 3.2: Network architectures. We note that there are shortcut identity connections from block to block in all three architectures.

<b>ResNet18</b>	<b>ResNet34</b>	<b>ResNet50</b>
7x7 conv 64 kernels	7x7 conv 64 kernels	7x7 conv 64 kernels
BatchNorm	BatchNorm	BatchNorm
relu	relu	relu
Maxpool	Maxpool	Maxpool
<b>2</b> x basic block 64 kernels	<b>3</b> x basic block 64 kernels	<b>3</b> x bottleneck 64 kernels
<b>2</b> x basic block 128 kernels	<b>4</b> x basic block 128 kernels	<b>4</b> x bottleneck 128 kernels
<b>2</b> x basic block 256 kernels	<b>6</b> x basic block 256 kernels	<b>6</b> x bottleneck 256 kernels
<b>2</b> x basic block 512 kernels	<b>3</b> x basic block 512 kernels	<b>3</b> x bottleneck 512 kernels
Average Pooling	Average Pooling	Average Pooling
Fully Connected	Fully Connected	Fully Connected
Softmax	Softmax	Softmax

To give a mathematical representation of a residual network’s architecture, we will define ResNet18. The other two networks can be defined similarly. Let  $\text{Bl} : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n}$  denote the ResNet18 basic block function and let  $\text{Bn}$  and  $T_1$  denote the batch

normalization operator and two dimensional  $3 \times 3$  discrete convolution, respectively. For  $x \in \mathbb{R}^{n \times n}$ , we can write Bl by

$$\text{Bl}(x) = \text{relu}(\text{Bn}(T_1(\text{relu}(\text{Bn}(T_1(x))))) + x$$

Based on Table 3.2, ResNet18:  $\mathbb{R}^{128 \times 128} \rightarrow \mathbb{R}^{340}$  is given by

$$\text{ResNet18}(x) = \text{softmax}(\text{fc}(\text{AvP}(\underset{8 \text{ times}}{\text{Bl} \cdots \text{Bl}}(\text{MaxPool}(\text{relu}(\text{Bn}(T_2(x)))) \cdots)))$$

where  $T_2$ , AvP, and MaxPool denote the two dimensional  $7 \times 7$  discrete convolution, the Average pooling, and the Max pooling operators, respectively.

## 3.6 Training Strategy

### 3.6.1 Feed-forward Scheme

The training strategy is customized based on the nature of the test set provided by Kaggle. The test set, which represents 10% of the entire test set, is uniformly distributed across all classes. Therefore, we decided to work with a uniformly sampled training set as well. We noticed that the smallest class in the data set contains 113,613 images and so we chose an equal number of samples of less than 113,613 images from each class. The results presented here are for a sample of 20,000 images per class. This makes sure that we avoid both over-represented and under-represented classes. We also generate a validation set consisting of 6,800 images (20 images per class). The under-sampled images are sketched and normalized in batches on the fly and then are propagated through the network.

### 3.6.2 Model Compilation

We used a Cross Entropy loss function and we compiled our models with an Adam optimizer with learning rate 0.001. Cross Entropy is given by

$$-\sum_{c=1}^{340} y_{x,c} \log(p_{x,c})$$

where  $y_{x,c}$  is a binary indicator (0 or 1) stating if the class  $c$  is the correct classification for observation  $x$ , and  $p_{x,c}$  is the predicted softmax probability that observation  $x$  is in class  $c$ . Obviously, one can try different configurations in an effort to optimize the hyperparameters of the networks, but for the purposes of this work we focused on a more theoretical approach for improving our results. We chose the same baseline for every network and explored the effects the switch from conventional ResNet to ResNetRF has. We tested the constructions of Chapter 2 by considering the following two opposite poles of the ResNetRF model:

- ResNetRF-One: A ResNet network where each 2D convolutional node is associated with a single predesigned fixed filter taken from a fixed Parseval frame filter bank. An arbitrary convolutional node would be defined as

$$\lambda h * I_N$$

where  $I_N : \mathbb{Z}^2 \rightarrow \mathbb{R}$  denotes the input of the Nth layer,  $\lambda$  is a scalar trainable parameter, and  $h$  a single fixed filter.

- ResNetRF-All: A ResNet network where if  $H = \{h_i\}_{i=1}^m$  is a Parseval frame filter bank, then the convolutional layer  $N$  is defined as

$$\left( \sum_{i=1}^m \lambda_i h_i \right) * I_N,$$

where  $I_N : \mathbb{Z}^2 \rightarrow \mathbb{R}$  denotes the input of the layer and  $\{\lambda_i\}_{i=1}^m$  are the scalar trainable parameters.

We saw that in all three ResNet models we trained, we came across two types of 2D filtering operations, namely  $7 \times 7$  and  $3 \times 3$  convolutions. We generated the corresponding filter banks as follows. For the conventional ResNet models all filters are randomly initialized. For ResNetRF we noticed that our data set comprises images dominated by edge and/or ridge characteristics. Hence, we wanted our frame filter banks to include directional finite-difference filters of various orders and orientations to steer our networks towards more meaningful features. Moreover, in the case of ResNetRF-One we attempted to significantly reduce the number of trainable parameters by considering sparse filter banks. The  $7 \times 7$  and  $3 \times 3$  filters used can be found in the Appendix.

### 3.7 Results

Finally, we present our results on all trained models. We first stress that all RF-One models reduce the number of trainable parameters by a significant amount. Table 3.3 below illustrates this fact.

Table 3.3: Number of trainable parameters.

	Conventional	RF-All	RF-One
ResNet18	11,344,660	11,344,660	1,576,724
ResNet34	21,452,820	21,452,820	2,706,452
ResNet50	24,198,420	24,198,420	14,135,572

We trained 9 classifiers on 20,000 images per class (6,800,000 images) for 2 epochs

with batch size 128. Each network took approximately 3 hours to train on 8 NVidia V100 GPUs provided by the Sabine cluster of the University of Houston ([uh.edu/cacds/resources/hpc/sabine/](http://uh.edu/cacds/resources/hpc/sabine/)). We obtained MAP@3 validation precisions and cross entropy losses and we present our findings in the Figures below. We noticed that during the training process of all three ResNet models, ResNetRf-All consistently outperformed both the conventional ResNet and ResNetRF-One. This trend persisted on the test set as well as will be shown shortly.

It is clear that RF-One does not perform as well as the other two, but when tested on ResNet50, the gap starts to close and its results improve. This could be due to the difference in the nature of the basic and bottleneck blocks. The bottleneck  $1 \times 1$  convolutions of a layer  $n$  can be seen as weighted averages of the activated filter responses of the previous RF-One convolutional layer, which creates a new post-activation linear combination of kernels.

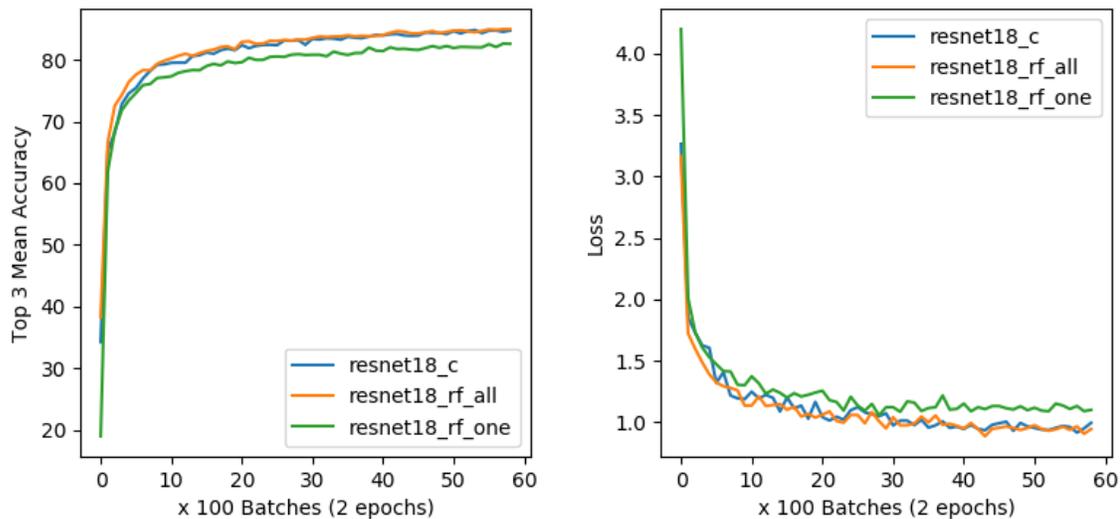


Figure 3.4: ResNet18 MAP@3 validation precision (left) and cross entropy loss (right).

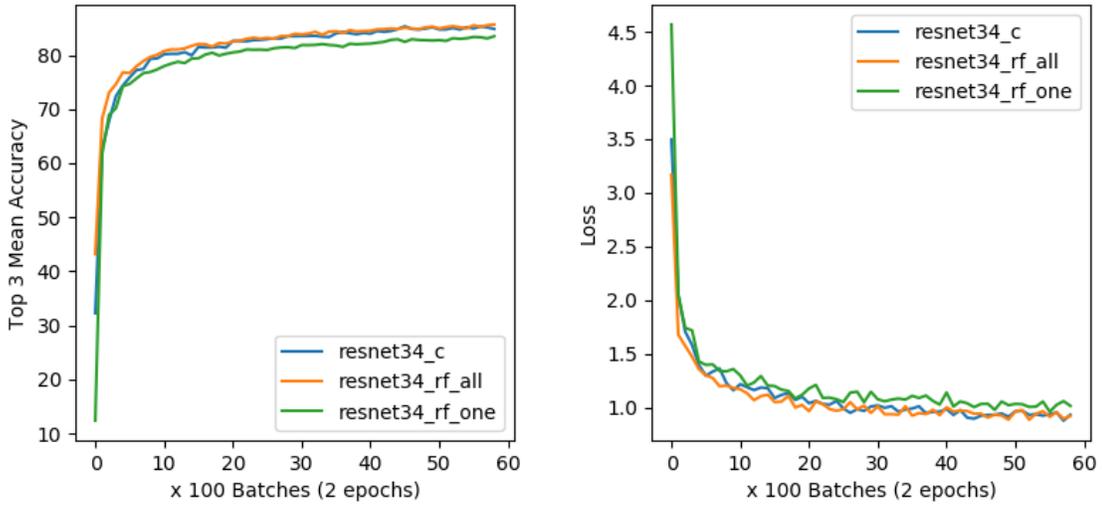


Figure 3.5: ResNet34 MAP@3 validation precision (left) and cross entropy loss (right).

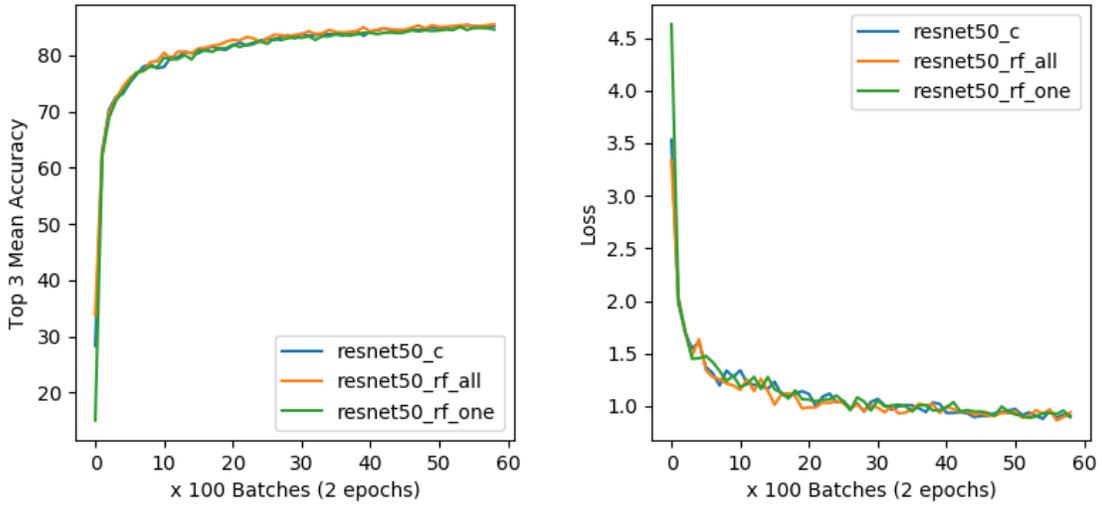


Figure 3.6: ResNet50 MAP@3 validation precision (left) and cross entropy loss (right).

### 3.7.1 A Qualitative Error Examination

We noticed interesting post-classification facts after a closer look at the quartiles of the validation confusion matrices of each network. To make our results more informative we tested our networks on a validation set consisting of 200 images per class. We give the two classes that a certain class is most confused by in a specific quartile for our best networks (RF-All). It is interesting to see how the networks commit errors between classes, the subtleties of which are sometimes difficult to incorporate in a sketch. We also present all confusion matrices in log scale for better visual representation.

- **ResNet18 RF-All:**

- Quartile(1): marker (77% miss-classified) is mostly confused with crayon (18% error) and pencil (13% error).
- Quartile(2): bear (59% miss-classified) is mostly confused with teddy-bear (10% error) and panda (8% error).
- Quartile(3): ambulance (27% miss-classified) is mostly confused with fire truck (8% error) and van (5% error).
- Quartile(4): airplane (12% miss-classified) is mostly confused with flying saucer (1% error) and crocodile (1% error).

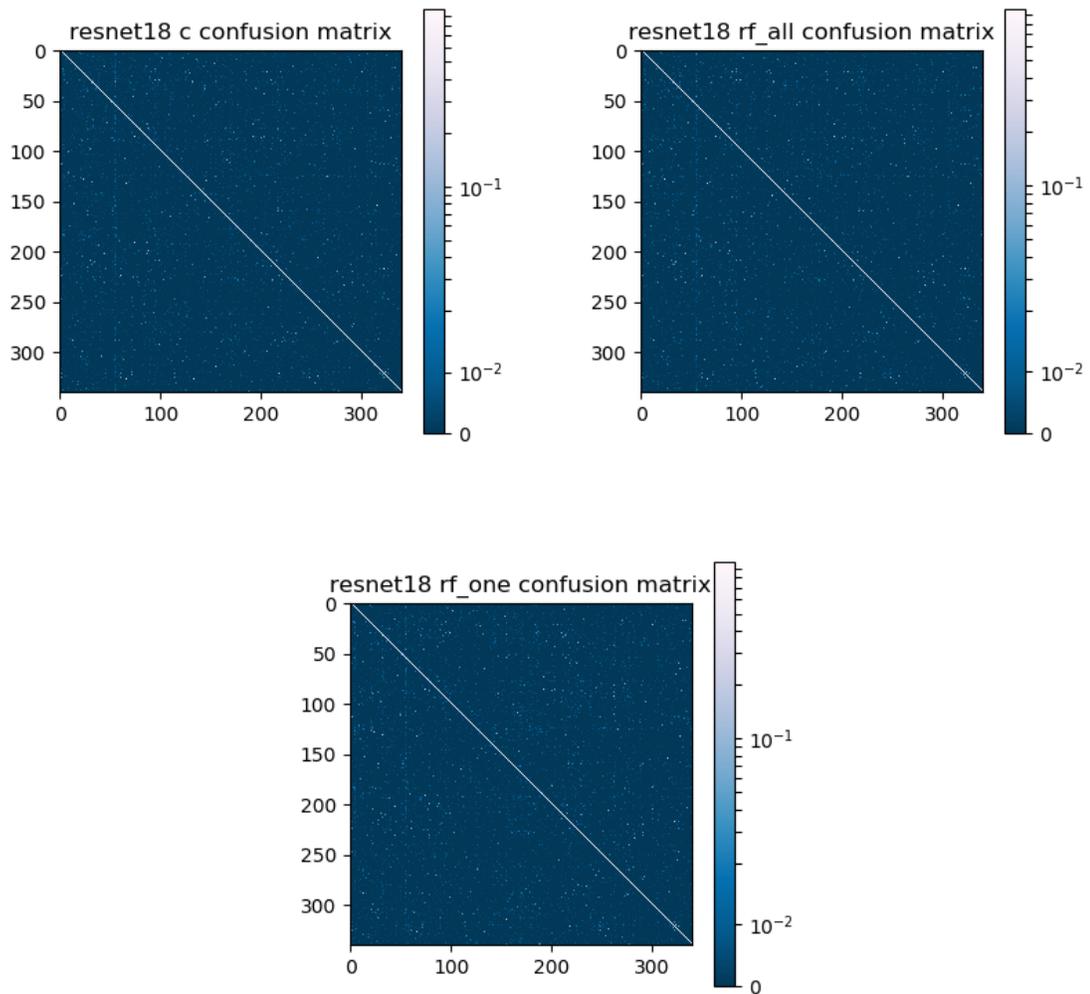


Figure 3.7: ResNet18 validation confusion matrices for 200 samples per class.

- **ResNet34 RF-All:**

- Quartile(2): bear (50% miss-classified) is mostly confused with teddy-bear (12% error) and dog (4% error).
- Quartile(3): ambulance (28% miss-classified) is mostly confused with van (6% error) and firetruck (5% error).

- Quartile(4): airplane (10% miss-classified) is mostly confused with ceiling fan (2% error) and shark (1% error).

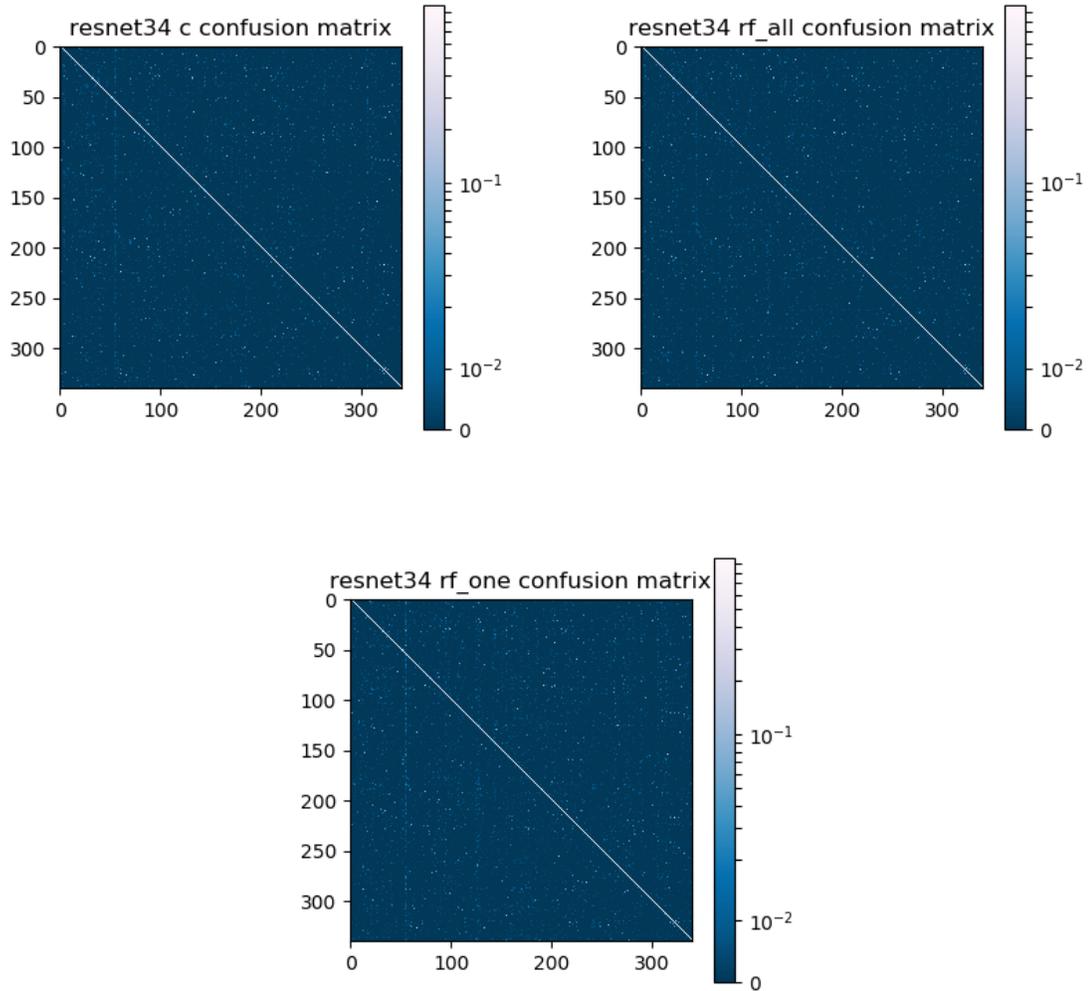


Figure 3.8: ResNet34 validation confusion matrices for 200 samples per class.

- **ResNet50 RF-All:**

- Quartile(2): bear (58% miss-classified) is mostly confused with teddy-bear (13% error) and monkey (4% error).

- Quartile(3): ambulance (42% miss-classified) is mostly confused with police car (16% error) and van (12% error).
- Quartile(4): airplane (8% miss-classified) is mostly confused with ceiling fan (3% error) and mosquito (1% error).

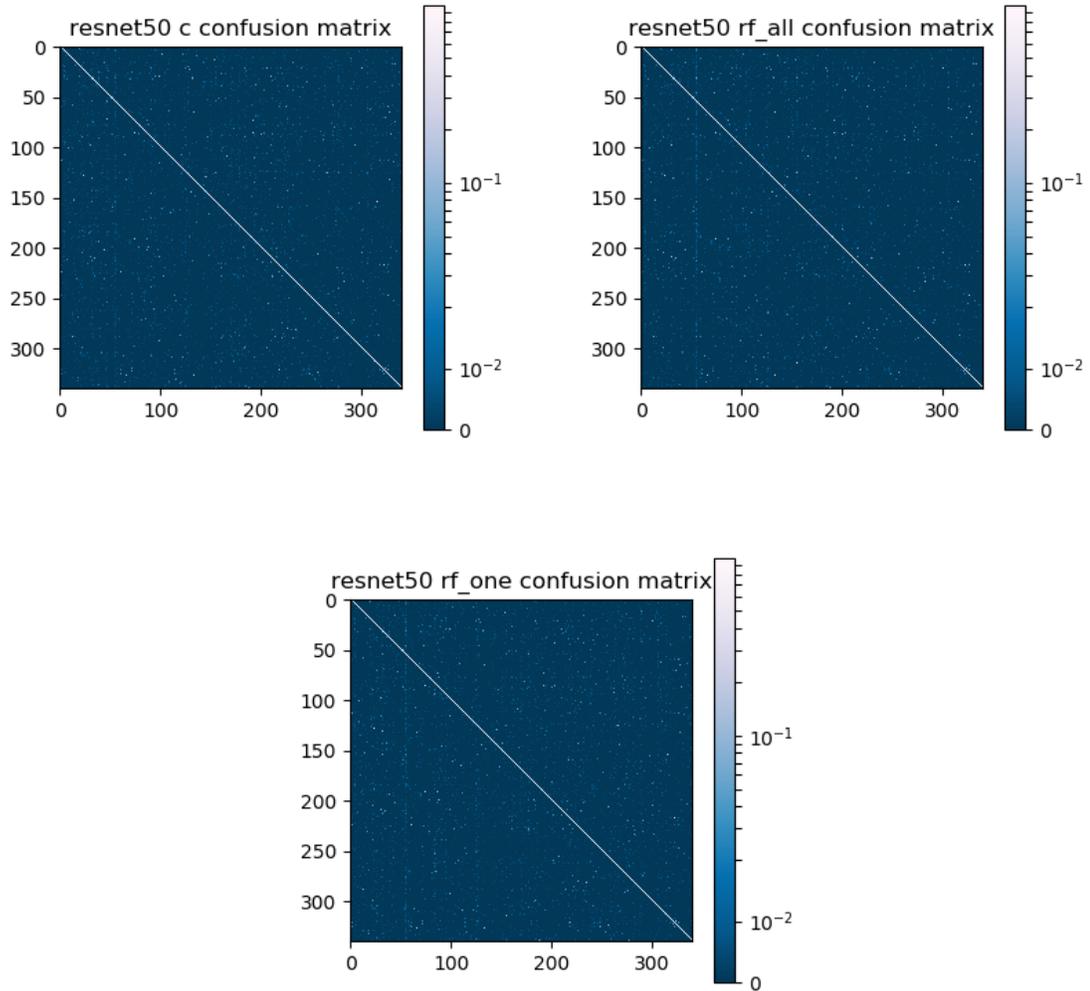


Figure 3.9: ResNet50 validation confusion matrices for 200 samples per class.

### 3.7.2 Test Results

We tested our results at the end of epochs 1 and 2 and present our MAP@3 precisions in Table 3.4. All accuracy results are provided by Kaggle on the competition’s webpage. It is once again evident that the RF-All variant is the go-to variant when comparing the three different ResNet models.

Table 3.4: Test accuracy provided by Kaggle

Test Accuracy		Epoch 1	Epoch2
ResNet18	Conventional	0.867	0.886
	RF-All	<b>0.872</b>	<b>0.891</b>
	RF-One	0.83	0.855
ResNet34	Conventional	0.875	0.894
	RF-All	<b>0.878</b>	<b>0.896</b>
	RF-One	0.845	0.867
ResNet50	Conventional	0.872	0.895
	RF-All	<b>0.875</b>	<b>0.898</b>
	RF-One	0.859	0.888

Lastly, based on this work, we trained one final ResNet50 RF-All network on 100,000 samples per class (34,000,000 images). We achieved a test MAP@3 precision of 93.68%, which would place our team in the top 8% of all participating teams.

# Chapter 4

## Conclusions

In Chapter 2, we developed theoretical tools and an algorithmic process to construct compactly supported multi-dimensional Parseval Frame and/or frame filter banks comprising filters designed to address the specific nature of image decomposition tasks. Our methods are easily implementable and provide a framework within which we are able to better understand the geometric nuances of the Unitary Extension Principle equation for a useful class of its solutions. We presented explicit examples of filter banks consisting of well-known first and second order central-difference filters, as well as filters with directional vanishing moments. We tested our construction on the Quick, Draw! image classification task and obtained comparable results to popular convolutional neural network architectures.

### Limitations

As shown in Chapter 2, designing application-specific Parseval Frame filter banks will often produce filters with unwanted characteristics. Even though Theorem 2.3.2

helps us minimize their contribution, fully designed filter banks still remain difficult to achieve. Another limitation of our construction is that it does not allow us to increase the size of the high-pass filters generated, which can only have at most as many non-zero coefficients as the low-pass filter has. Removing this limitation would be very useful in many applications, e.g., when one wishes to increase the number of orientations of the high-pass filters.

## **Future work**

We wish to further explore the applications of our methods in tasks that we believe are crucial. For example, we wish to extend our research by constructing 3D filter banks for the analysis and accurate segmentation of dendritic spines in neuronal structures. This would have potential impact in the monitoring and understanding of the evolution of brain diseases such as autism, etc. Moreover, our neural network application indicated that directional representations can boost the performance of well-known architectures. We aim to continue testing these ideas in more complex data sets to produce efficient and informative learning algorithms.

# Appendix A

## Filter Banks

### A.1 $3 \times 3$ Filters for ResNetRF-One

#### A.1.1 Low-pass Filter

$$\begin{pmatrix} 0.0625 & 0.1250 & 0.0625 \\ 0.1250 & 0.2500 & 0.1250 \\ 0.0625 & 0.1250 & 0.0625 \end{pmatrix}$$

#### A.1.2 High-pass Filters

##### A.1.2.1 First-order Finite-difference Filters

$$\begin{pmatrix} 0 & 0 & -0.1750 \\ 0 & 0 & 0 \\ 0.1750 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 & 0 \\ 0.2475 & 0 & -0.2475 \\ 0 & 0 & 0 \end{pmatrix}$$
$$\begin{pmatrix} 0.1750 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -0.1750 \end{pmatrix} \quad \begin{pmatrix} 0 & -0.2475 & 0 \\ 0 & 0 & 0 \\ 0 & 0.2475 & 0 \end{pmatrix}$$

### A.1.2.2 Second-order Finite-difference filters

$$\begin{pmatrix} 0 & -0.1170 & 0 \\ 0 & 0.2341 & 0 \\ 0 & -0.1170 & 0 \end{pmatrix} \quad \begin{pmatrix} -0.0988 & 0 & 0 \\ 0 & 0.1976 & 0 \\ 0 & 0 & -0.0988 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 0 & 0 \\ -0.1220 & 0.2440 & -0.1220 \\ 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 & -0.0908 \\ 0 & 0.1817 & 0 \\ -0.0908 & 0 & 0 \end{pmatrix}$$

## A.2 $7 \times 7$ Filters for ResNetRF-One

### A.2.1 Low-pass Filter

$$\begin{pmatrix} 0.0002 & 0.0015 & 0.0037 & 0.0049 & 0.0037 & 0.0015 & 0.0002 \\ 0.0015 & 0.0088 & 0.0220 & 0.0293 & 0.0220 & 0.0088 & 0.0015 \\ 0.0037 & 0.0220 & 0.0549 & 0.0732 & 0.0549 & 0.0220 & 0.0037 \\ 0.0049 & 0.0293 & 0.0732 & 0.0977 & 0.0732 & 0.0293 & 0.0049 \\ 0.0037 & 0.0220 & 0.0549 & 0.0732 & 0.0549 & 0.0220 & 0.0037 \\ 0.0015 & 0.0088 & 0.0220 & 0.0293 & 0.0220 & 0.0088 & 0.0015 \\ 0.0002 & 0.0015 & 0.0037 & 0.0049 & 0.0037 & 0.0015 & 0.0002 \end{pmatrix}$$

### A.2.2 High-pass Filters

#### A.2.2.1 First-order Finite-difference Filters

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & -0.0110 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0110 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -0.0270 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0270 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$









$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ -0.022 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.044 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -0.022 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -0.030 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.061 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -0.030 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \\
\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -0.033 & 0 & 0 & 0.066 & 0 & -0.033 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -0.030 \\ 0 & 0 & 0 & 0.061 & 0 & 0 \\ -0.030 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \\
\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -0.022 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.044 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -0.022 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & -0.010 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.020 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -0.010 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

## A.3 $3 \times 3$ Filters for ResNetRF-All

### A.3.1 Low-pass Filter

$$\begin{pmatrix} 0.0625 & 0.1250 & 0.0625 \\ 0.1250 & 0.2500 & 0.1250 \\ 0.0625 & 0.1250 & 0.0625 \end{pmatrix}$$

### A.3.2 High-pass Filters

#### A.3.2.1 First-order Finite-difference Filters

$$\begin{pmatrix} 0.1768 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -0.1768 \end{pmatrix} \quad \begin{pmatrix} 0 & 0.2500 & 0 \\ 0 & 0 & 0 \\ 0 & -0.2500 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 0 & 0.1768 \\ 0 & 0 & 0 \\ -0.1768 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0.2500 & 0 & -0.2500 \\ 0 & 0 & 0 \end{pmatrix}$$

### A.3.2.2 Second-order Finite-difference Filter with 2 Vanishing Moments

$$\begin{pmatrix} -0 & -0 & -0 \\ 0.1768 & -0.3536 & 0.1768 \\ 0 & -0 & -0 \end{pmatrix}$$

### A.3.2.3 Filters with 2 Vanishing Moments

$$\begin{pmatrix} 0.0250 & 0.0333 & 0.1417 \\ -0.1000 & -0.2000 & -0.1000 \\ 0.1417 & 0.0333 & 0.0250 \end{pmatrix} \begin{pmatrix} 0.0177 & 0.2003 & -0.0766 \\ -0.0707 & -0.1414 & -0.0707 \\ -0.0766 & 0.2003 & 0.0177 \end{pmatrix}$$

$$\begin{pmatrix} 0.1625 & -0.0750 & -0.0375 \\ -0.0250 & -0.0500 & -0.0250 \\ -0.0375 & -0.0750 & 0.1625 \end{pmatrix}$$

## A.4 $7 \times 7$ Filters for ResNetRF-All

### A.4.1 Low-pass Filter

$$\begin{pmatrix} 0.0002 & 0.0015 & 0.0037 & 0.0049 & 0.0037 & 0.0015 & 0.0002 \\ 0.0015 & 0.0088 & 0.0220 & 0.0293 & 0.0220 & 0.0088 & 0.0015 \\ 0.0037 & 0.0220 & 0.0549 & 0.0732 & 0.0549 & 0.0220 & 0.0037 \\ 0.0049 & 0.0293 & 0.0732 & 0.0977 & 0.0732 & 0.0293 & 0.0049 \\ 0.0037 & 0.0220 & 0.0549 & 0.0732 & 0.0549 & 0.0220 & 0.0037 \\ 0.0015 & 0.0088 & 0.0220 & 0.0293 & 0.0220 & 0.0088 & 0.0015 \\ 0.0002 & 0.0015 & 0.0037 & 0.0049 & 0.0037 & 0.0015 & 0.0002 \end{pmatrix}$$





$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.0428 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.0428 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0494 & 0 & 0 & 0 & 0 & 0 & -0.0494 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.1210 & 0 & 0 & 0 & -0.1210 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.1914 & 0 & -0.1914 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

#### A.4.2.2 Filters with 2 Vanishing Moments

$$\begin{pmatrix} -0.0003 & -0.0006 & -0.0008 & -0.0007 & -0.0005 & -0.0002 & -0.0001 \\ -0.0001 & -0.0005 & -0.0011 & -0.0014 & -0.0010 & -0.0004 & -0.0001 \\ -0.0002 & -0.0010 & -0.0024 & -0.0032 & -0.0024 & -0.0010 & -0.0002 \\ -0.0002 & -0.0013 & 0.1322 & -0.2252 & 0.1322 & -0.0013 & -0.0002 \\ -0.0002 & -0.0010 & -0.0024 & -0.0032 & -0.0024 & -0.0010 & -0.0002 \\ -0.0001 & -0.0004 & -0.0010 & -0.0014 & -0.0011 & -0.0005 & -0.0001 \\ -0.0001 & -0.0002 & -0.0005 & -0.0007 & -0.0008 & -0.0006 & -0.0003 \end{pmatrix}$$

$$\begin{pmatrix} 0.0012 & 0.0023 & 0.0029 & 0.0026 & 0.0017 & 0.0007 & 0.0002 \\ 0.0005 & 0.0018 & 0.0039 & 0.0049 & 0.0036 & 0.0015 & 0.0003 \\ 0.0006 & 0.0034 & 0.0085 & 0.0113 & 0.0085 & 0.0034 & 0.0006 \\ 0.0008 & 0.0901 & -0.0845 & -0.1413 & -0.0845 & 0.0901 & 0.0008 \\ 0.0006 & 0.0034 & 0.0085 & 0.0113 & 0.0085 & 0.0034 & 0.0006 \\ 0.0003 & 0.0015 & 0.0036 & 0.0049 & 0.0039 & 0.0018 & 0.0005 \\ 0.0002 & 0.0007 & 0.0017 & 0.0026 & 0.0029 & 0.0023 & 0.0012 \end{pmatrix}$$

$$\begin{pmatrix} 0.0021 & 0.0039 & 0.0048 & 0.0044 & 0.0029 & 0.0012 & 0.0003 \\ 0.0008 & 0.0030 & 0.0065 & 0.0082 & 0.0061 & 0.0025 & 0.0004 \\ 0.0010 & 0.0058 & 0.0143 & 0.0190 & 0.0142 & 0.0057 & 0.0010 \\ 0.0362 & -0.0530 & -0.0488 & -0.0853 & -0.0488 & -0.0530 & 0.0362 \\ 0.0010 & 0.0057 & 0.0142 & 0.0190 & 0.0143 & 0.0058 & 0.0010 \\ 0.0004 & 0.0025 & 0.0061 & 0.0082 & 0.0065 & 0.0030 & 0.0008 \\ 0.0003 & 0.0012 & 0.0029 & 0.0044 & 0.0048 & 0.0039 & 0.0021 \end{pmatrix}$$

$$\begin{pmatrix} 0.0017 & 0.0032 & 0.0041 & 0.0037 & 0.0024 & 0.0010 & 0.0003 \\ 0.0007 & 0.0025 & 0.0054 & 0.0069 & 0.0051 & 0.0021 & 0.0004 \\ 0.0008 & 0.0048 & 0.0119 & 0.0159 & 0.0119 & 0.0048 & 0.0311 \\ -0.0236 & -0.0365 & -0.0320 & -0.0571 & -0.0320 & -0.0365 & -0.0236 \\ 0.0311 & 0.0048 & 0.0119 & 0.0159 & 0.0119 & 0.0048 & 0.0008 \\ 0.0004 & 0.0021 & 0.0051 & 0.0069 & 0.0054 & 0.0025 & 0.0007 \\ 0.0003 & 0.0010 & 0.0024 & 0.0037 & 0.0041 & 0.0032 & 0.0017 \end{pmatrix}$$

$$\begin{pmatrix} 0.0008 & 0.0015 & 0.0018 & 0.0017 & 0.0011 & 0.0005 & 0.0001 \\ 0.0003 & 0.0011 & 0.0025 & 0.0031 & 0.0023 & 0.0009 & 0.0002 \\ 0.0004 & 0.0022 & 0.0054 & 0.0072 & 0.0054 & 0.0763 & -0.0210 \\ -0.0170 & -0.0274 & -0.0266 & -0.0457 & -0.0266 & -0.0274 & -0.0170 \\ -0.0210 & 0.0763 & 0.0054 & 0.0072 & 0.0054 & 0.0022 & 0.0004 \\ 0.0002 & 0.0009 & 0.0023 & 0.0031 & 0.0025 & 0.0011 & 0.0003 \\ 0.0001 & 0.0005 & 0.0011 & 0.0017 & 0.0018 & 0.0015 & 0.0008 \end{pmatrix}$$

$$\begin{pmatrix} 0.0004 & 0.0007 & 0.0009 & 0.0008 & 0.0005 & 0.0002 & 0.0001 \\ 0.0002 & 0.0006 & 0.0012 & 0.0015 & 0.0011 & 0.0005 & 0.0001 \\ 0.0002 & 0.0011 & 0.0026 & 0.0035 & 0.1198 & -0.0514 & -0.0150 \\ -0.0121 & -0.0200 & -0.0204 & -0.0344 & -0.0204 & -0.0200 & -0.0121 \\ -0.0150 & -0.0514 & 0.1198 & 0.0035 & 0.0026 & 0.0011 & 0.0002 \\ 0.0001 & 0.0005 & 0.0011 & 0.0015 & 0.0012 & 0.0006 & 0.0002 \\ 0.0001 & 0.0002 & 0.0005 & 0.0008 & 0.0009 & 0.0007 & 0.0004 \end{pmatrix}$$

$$\begin{pmatrix} 0.0007 & 0.0013 & 0.0016 & 0.0015 & 0.0010 & 0.0004 & 0.0001 \\ 0.0003 & 0.0010 & 0.0022 & 0.0028 & 0.0020 & 0.0008 & 0.0001 \\ 0.0003 & 0.0019 & 0.0048 & 0.1417 & -0.0781 & -0.0351 & -0.0104 \\ -0.0083 & -0.0126 & -0.0106 & -0.0192 & -0.0106 & -0.0126 & -0.0083 \\ -0.0104 & -0.0351 & -0.0781 & 0.1417 & 0.0048 & 0.0019 & 0.0003 \\ 0.0001 & 0.0008 & 0.0020 & 0.0028 & 0.0022 & 0.0010 & 0.0003 \\ 0.0001 & 0.0004 & 0.0010 & 0.0015 & 0.0016 & 0.0013 & 0.0007 \end{pmatrix}$$

$$\begin{pmatrix} 0.0015 & 0.0029 & 0.0036 & 0.0033 & 0.0021 & 0.0009 & 0.0002 \\ 0.0006 & 0.0022 & 0.0048 & 0.0061 & 0.0045 & 0.0018 & 0.0003 \\ 0.0007 & 0.0043 & 0.1277 & -0.0817 & -0.0481 & -0.0220 & -0.0069 \\ -0.0052 & -0.0051 & 0.0020 & -0.0010 & 0.0020 & -0.0051 & -0.0052 \\ -0.0069 & -0.0220 & -0.0481 & -0.0817 & 0.1277 & 0.0043 & 0.0007 \\ 0.0003 & 0.0018 & 0.0045 & 0.0061 & 0.0048 & 0.0022 & 0.0006 \\ 0.0002 & 0.0009 & 0.0021 & 0.0033 & 0.0036 & 0.0029 & 0.0015 \end{pmatrix}$$

$$\begin{pmatrix} 0.0024 & 0.0044 & 0.0055 & 0.0051 & 0.0033 & 0.0014 & 0.0004 \\ 0.0009 & 0.0034 & 0.0074 & 0.0094 & 0.0069 & 0.0028 & 0.0005 \\ 0.0011 & 0.0807 & -0.0666 & -0.0461 & -0.0252 & -0.0121 & -0.0043 \\ -0.0029 & 0.0010 & 0.0131 & 0.0149 & 0.0131 & 0.0010 & -0.0029 \\ -0.0043 & -0.0121 & -0.0252 & -0.0461 & -0.0666 & 0.0807 & 0.0011 \\ 0.0005 & 0.0028 & 0.0069 & 0.0094 & 0.0074 & 0.0034 & 0.0009 \\ 0.0004 & 0.0014 & 0.0033 & 0.0051 & 0.0055 & 0.0044 & 0.0024 \end{pmatrix}$$

$$\begin{pmatrix} 0.0027 & 0.0051 & 0.0063 & 0.0058 & 0.0038 & 0.0016 & 0.0004 \\ 0.0011 & 0.0039 & 0.0085 & 0.0108 & 0.0080 & 0.0032 & 0.0006 \\ 0.0316 & -0.0449 & -0.0399 & -0.0230 & -0.0107 & -0.0057 & -0.0025 \\ -0.0014 & 0.0045 & 0.0187 & 0.0232 & 0.0187 & 0.0045 & -0.0014 \\ -0.0025 & -0.0057 & -0.0107 & -0.0230 & -0.0399 & -0.0449 & 0.0316 \\ 0.0006 & 0.0032 & 0.0080 & 0.0108 & 0.0085 & 0.0039 & 0.0011 \\ 0.0004 & 0.0016 & 0.0038 & 0.0058 & 0.0063 & 0.0051 & 0.0027 \end{pmatrix}$$

$$\begin{pmatrix} 0.0023 & 0.0042 & 0.0052 & 0.0048 & 0.0031 & 0.0014 & 0.0003 \\ 0.0009 & 0.0033 & 0.0071 & 0.0089 & 0.0066 & 0.0027 & 0.0196 \\ -0.0203 & -0.0308 & -0.0260 & -0.0133 & -0.0053 & -0.0031 & -0.0016 \\ -0.0008 & 0.0044 & 0.0162 & 0.0204 & 0.0162 & 0.0044 & -0.0008 \\ -0.0016 & -0.0031 & -0.0053 & -0.0133 & -0.0260 & -0.0308 & -0.0203 \\ 0.0196 & 0.0027 & 0.0066 & 0.0089 & 0.0071 & 0.0033 & 0.0009 \\ 0.0003 & 0.0014 & 0.0031 & 0.0048 & 0.0052 & 0.0042 & 0.0023 \end{pmatrix}$$

$$\begin{pmatrix} 0.0013 & 0.0024 & 0.0031 & 0.0028 & 0.0018 & 0.0008 & 0.0002 \\ 0.0005 & 0.0019 & 0.0041 & 0.0052 & 0.0039 & 0.0484 & -0.0133 \\ -0.0145 & -0.0226 & -0.0203 & -0.0119 & -0.0056 & -0.0029 & -0.0013 \\ -0.0007 & 0.0021 & 0.0090 & 0.0111 & 0.0090 & 0.0021 & -0.0007 \\ -0.0013 & -0.0029 & -0.0056 & -0.0119 & -0.0203 & -0.0226 & -0.0145 \\ -0.0133 & 0.0484 & 0.0039 & 0.0052 & 0.0041 & 0.0019 & 0.0005 \\ 0.0002 & 0.0008 & 0.0018 & 0.0028 & 0.0031 & 0.0024 & 0.0013 \end{pmatrix}$$

$$\begin{pmatrix} 0.0008 & 0.0015 & 0.0019 & 0.0018 & 0.0011 & 0.0005 & 0.0001 \\ 0.0003 & 0.0012 & 0.0026 & 0.0033 & 0.0765 & -0.0322 & -0.0094 \\ -0.0103 & -0.0163 & -0.0151 & -0.0094 & -0.0047 & -0.0024 & -0.0010 \\ -0.0006 & 0.0011 & 0.0053 & 0.0065 & 0.0053 & 0.0011 & -0.0006 \\ -0.0010 & -0.0024 & -0.0047 & -0.0094 & -0.0151 & -0.0163 & -0.0103 \\ -0.0094 & -0.0322 & 0.0765 & 0.0033 & 0.0026 & 0.0012 & 0.0003 \\ 0.0001 & 0.0005 & 0.0011 & 0.0018 & 0.0019 & 0.0015 & 0.0008 \end{pmatrix}$$

$$\begin{pmatrix} 0.0008 & 0.0016 & 0.0020 & 0.0018 & 0.0012 & 0.0005 & 0.0001 \\ 0.0003 & 0.0012 & 0.0027 & 0.0890 & -0.0499 & -0.0224 & -0.0066 \\ -0.0072 & -0.0107 & -0.0088 & -0.0042 & -0.0015 & -0.0009 & -0.0006 \\ -0.0003 & 0.0018 & 0.0062 & 0.0079 & 0.0062 & 0.0018 & -0.0003 \\ -0.0006 & -0.0009 & -0.0015 & -0.0042 & -0.0088 & -0.0107 & -0.0072 \\ -0.0066 & -0.0224 & -0.0499 & 0.0890 & 0.0027 & 0.0012 & 0.0003 \\ 0.0001 & 0.0005 & 0.0012 & 0.0018 & 0.0020 & 0.0016 & 0.0008 \end{pmatrix}$$

$$\begin{pmatrix} 0.0013 & 0.0023 & 0.0029 & 0.0027 & 0.0018 & 0.0008 & 0.0002 \\ 0.0005 & 0.0018 & 0.0781 & -0.0555 & -0.0334 & -0.0151 & -0.0045 \\ -0.0047 & -0.0058 & -0.0017 & 0.0030 & 0.0034 & 0.0011 & -0.0001 \\ 0.0002 & 0.0036 & 0.0104 & 0.0135 & 0.0104 & 0.0036 & 0.0002 \\ -0.0001 & 0.0011 & 0.0034 & 0.0030 & -0.0017 & -0.0058 & -0.0047 \\ -0.0045 & -0.0151 & -0.0334 & -0.0555 & 0.0781 & 0.0018 & 0.0005 \\ 0.0002 & 0.0008 & 0.0018 & 0.0027 & 0.0029 & 0.0023 & 0.0013 \end{pmatrix}$$

$$\begin{pmatrix} 0.0017 & 0.0032 & 0.0040 & 0.0036 & 0.0024 & 0.0010 & 0.0003 \\ 0.0007 & 0.0493 & -0.0471 & -0.0360 & -0.0212 & -0.0097 & -0.0030 \\ -0.0030 & -0.0018 & 0.0044 & 0.0095 & 0.0080 & 0.0030 & 0.0003 \\ 0.0006 & 0.0055 & 0.0147 & 0.0194 & 0.0147 & 0.0055 & 0.0006 \\ 0.0003 & 0.0030 & 0.0080 & 0.0095 & 0.0044 & -0.0018 & -0.0030 \\ -0.0030 & -0.0097 & -0.0212 & -0.0360 & -0.0471 & 0.0493 & 0.0007 \\ 0.0003 & 0.0010 & 0.0024 & 0.0036 & 0.0040 & 0.0032 & 0.0017 \end{pmatrix}$$

$$\begin{pmatrix} 0.0019 & 0.0035 & 0.0043 & 0.0040 & 0.0026 & 0.0011 & 0.0003 \\ 0.0199 & -0.0304 & -0.0312 & -0.0229 & -0.0131 & -0.0061 & -0.0020 \\ -0.0018 & 0.0005 & 0.0076 & 0.0128 & 0.0102 & 0.0039 & 0.0005 \\ 0.0009 & 0.0063 & 0.0164 & 0.0217 & 0.0164 & 0.0063 & 0.0009 \\ 0.0005 & 0.0039 & 0.0102 & 0.0128 & 0.0076 & 0.0005 & -0.0018 \\ -0.0020 & -0.0061 & -0.0131 & -0.0229 & -0.0312 & -0.0304 & 0.0199 \\ 0.0003 & 0.0011 & 0.0026 & 0.0040 & 0.0043 & 0.0035 & 0.0019 \end{pmatrix}$$

$$\begin{pmatrix} 0.0016 & 0.0030 & 0.0037 & 0.0034 & 0.0022 & 0.0010 & 0.0081 \\ -0.0129 & -0.0211 & -0.0212 & -0.0151 & -0.0085 & -0.0040 & -0.0014 \\ -0.0011 & 0.0011 & 0.0072 & 0.0115 & 0.0090 & 0.0035 & 0.0005 \\ 0.0008 & 0.0054 & 0.0140 & 0.0186 & 0.0140 & 0.0054 & 0.0008 \\ 0.0005 & 0.0035 & 0.0090 & 0.0115 & 0.0072 & 0.0011 & -0.0011 \\ -0.0014 & -0.0040 & -0.0085 & -0.0151 & -0.0212 & -0.0211 & -0.0129 \\ 0.0081 & 0.0010 & 0.0022 & 0.0034 & 0.0037 & 0.0030 & 0.0016 \end{pmatrix}$$

$$\begin{pmatrix} 0.0010 & 0.0019 & 0.0024 & 0.0022 & 0.0014 & 0.0197 & -0.0054 \\ -0.0092 & -0.0151 & -0.0154 & -0.0111 & -0.0063 & -0.0030 & -0.0010 \\ -0.0008 & 0.0005 & 0.0043 & 0.0071 & 0.0056 & 0.0022 & 0.0003 \\ 0.0005 & 0.0034 & 0.0089 & 0.0118 & 0.0089 & 0.0034 & 0.0005 \\ 0.0003 & 0.0022 & 0.0056 & 0.0071 & 0.0043 & 0.0005 & -0.0008 \\ -0.0010 & -0.0030 & -0.0063 & -0.0111 & -0.0154 & -0.0151 & -0.0092 \\ -0.0054 & 0.0197 & 0.0014 & 0.0022 & 0.0024 & 0.0019 & 0.0010 \end{pmatrix}$$

$$\begin{pmatrix} 0.0007 & 0.0012 & 0.0016 & 0.0014 & 0.0312 & -0.0131 & -0.0038 \\ -0.0065 & -0.0107 & -0.0110 & -0.0080 & -0.0046 & -0.0021 & -0.0007 \\ -0.0006 & 0.0002 & 0.0028 & 0.0046 & 0.0037 & 0.0014 & 0.0002 \\ 0.0003 & 0.0023 & 0.0059 & 0.0078 & 0.0059 & 0.0023 & 0.0003 \\ 0.0002 & 0.0014 & 0.0037 & 0.0046 & 0.0028 & 0.0002 & -0.0006 \\ -0.0007 & -0.0021 & -0.0046 & -0.0080 & -0.0110 & -0.0107 & -0.0065 \\ -0.0038 & -0.0131 & 0.0312 & 0.0014 & 0.0016 & 0.0012 & 0.0007 \end{pmatrix}$$

$$\begin{pmatrix} 0.0006 & 0.0011 & 0.0014 & 0.0362 & -0.0206 & -0.0092 & -0.0027 \\ -0.0046 & -0.0074 & -0.0074 & -0.0053 & -0.0029 & -0.0014 & -0.0005 \\ -0.0004 & 0.0005 & 0.0027 & 0.0043 & 0.0033 & 0.0013 & 0.0002 \\ 0.0003 & 0.0020 & 0.0052 & 0.0068 & 0.0052 & 0.0020 & 0.0003 \\ 0.0002 & 0.0013 & 0.0033 & 0.0043 & 0.0027 & 0.0005 & -0.0004 \\ -0.0005 & -0.0014 & -0.0029 & -0.0053 & -0.0074 & -0.0074 & -0.0046 \\ -0.0027 & -0.0092 & -0.0206 & 0.0362 & 0.0014 & 0.0011 & 0.0006 \end{pmatrix}$$

$$\begin{pmatrix} 0.0007 & 0.0013 & 0.0318 & -0.0233 & -0.0142 & -0.0064 & -0.0019 \\ -0.0031 & -0.0049 & -0.0044 & -0.0027 & -0.0013 & -0.0007 & -0.0003 \\ -0.0001 & 0.0011 & 0.0038 & 0.0055 & 0.0042 & 0.0017 & 0.0003 \\ 0.0004 & 0.0024 & 0.0061 & 0.0081 & 0.0061 & 0.0024 & 0.0004 \\ 0.0003 & 0.0017 & 0.0042 & 0.0055 & 0.0038 & 0.0011 & -0.0001 \\ -0.0003 & -0.0007 & -0.0013 & -0.0027 & -0.0044 & -0.0049 & -0.0031 \\ -0.0019 & -0.0064 & -0.0142 & -0.0233 & 0.0318 & 0.0013 & 0.0007 \end{pmatrix}$$

$$\begin{pmatrix} 0.0008 & 0.0206 & -0.0195 & -0.0157 & -0.0096 & -0.0043 & -0.0013 \\ -0.0021 & -0.0030 & -0.0021 & -0.0006 & 0.0001 & -0.0001 & -0.0001 \\ 0.0001 & 0.0017 & 0.0049 & 0.0069 & 0.0052 & 0.0021 & 0.0003 \\ 0.0005 & 0.0029 & 0.0073 & 0.0097 & 0.0073 & 0.0029 & 0.0005 \\ 0.0003 & 0.0021 & 0.0052 & 0.0069 & 0.0049 & 0.0017 & 0.0001 \\ -0.0001 & -0.0001 & 0.0001 & -0.0006 & -0.0021 & -0.0030 & -0.0021 \\ -0.0013 & -0.0043 & -0.0096 & -0.0157 & -0.0195 & 0.0206 & 0.0008 \end{pmatrix}$$

$$\begin{pmatrix} 0.0087 & -0.0120 & -0.0132 & -0.0105 & -0.0064 & -0.0029 & -0.0008 \\ -0.0014 & -0.0017 & -0.0006 & 0.0007 & 0.0008 & 0.0003 & -0 \\ 0.0002 & 0.0019 & 0.0053 & 0.0073 & 0.0055 & 0.0022 & 0.0004 \\ 0.0005 & 0.0030 & 0.0076 & 0.0102 & 0.0076 & 0.0030 & 0.0005 \\ 0.0004 & 0.0022 & 0.0055 & 0.0073 & 0.0053 & 0.0019 & 0.0002 \\ -0 & 0.0003 & 0.0008 & 0.0007 & -0.0006 & -0.0017 & -0.0014 \\ -0.0008 & -0.0029 & -0.0064 & -0.0105 & -0.0132 & -0.0120 & 0.0087 \end{pmatrix}$$

# Bibliography

- [1] E. H. Adelson, E. Simoncelli, and R. Hingoranp. Orthogonal pyramid transforms for image coding. *Visual Communications and Image Processing II*, 845:50–58, 1987.
- [2] A. S. Antolín and R. A. Zalik. A family of nonseparable scaling functions and compactly supported tight framelets. *Journal of Mathematical Analysis and Applications*, 404(2):201–211, 2013.
- [3] N. Atreas, N. Karantzas, M. Papadakis, and T. Stavropoulos. Exploring neuronal synapses with directional and symmetric frame filters with small support. In *Proc.SPIE*, volume 10394, 2017.
- [4] A. Ayache. Construction de bases orthonormés d’ondelettes de  $\mathbb{L}^2(\mathbb{R}^2)$  non séparables, support compact et de régularité arbitrairement grande. *Comptes Rendus Académie des Sciences de Paris*, 325:17–20, 1997.
- [5] A. Ayache. Some methods for constructing nonseparable, orthonormal, compactly supported wavelet bases. *Applied and Computational Harmonic Analysis*, 10:99–111, 2001.
- [6] E. Belogay and Y. Wang. Arbitrarily smooth orthogonal nonseparable wavelets in  $\mathbb{R}^2$ . *SIAM J. Math. Anal.*, 30(3):678–697, 1999.
- [7] J. Bruna and S. Mallat. Invariant scattering convolution networks. *CoRR*, abs/1203.1513, 2012.
- [8] C. A. Cabrelli and M. L. Gordillo. Existence of multiwavelets in  $\mathbb{R}^n$ . *Proc. Amer. Math. Soc.*, 130(5):1413–1424, 2000.
- [9] E. J. Candès. Harmonic analysis of neural networks. *Appl. Comput. Harmon. Anal.*, 6:197–218, 1999.
- [10] E. J. Candès and L. Demanet. The curvelet representation of wave propagators is optimally sparse. *Comm. Pure Appl. Math.*, 58(11):1472–1528, 2005.
- [11] E. J. Candès, L. Demanet, D. Donoho, and L. Ying. Fast discrete curvelet transforms. *Multiscale Model. Simul.*, 5(3):861–899, 2006.

- [12] E. J. Candes and D. L. Donoho. Ridgelets: A key to higher dimensional intermittency? *Phil. Trans. R. Soc. London, A*:2495–2509, 1999.
- [13] E. J. Candès and D. L. Donoho. New tight frames of curvelets and optimal representations of objects with piecewise  $C^2$  singularities. *Comm. Pure Appl. Math.*, 57(2):219–266, 2004.
- [14] O. Christensen. *An introduction to frames and Riesz bases*. Applied and Numerical Harmonic Analysis. Birkhäuser Boston, 2002.
- [15] O. Christensen. *Functions, spaces, and expansions: Mathematical tools in Physics and Engineering*. Birkhäuser Verlag, 1st edition, 2010.
- [16] C. K. Chui, W. He, and J. Stöckler. Compactly supported tight and sibling frames with maximum vanishing moments. *Appl. Comput. Harmon. Anal.*, 13(3):224–262, 2002.
- [17] A. L. Da Cunha and M. N. Do. Bi-orthogonal filter banks with directional vanishing moments. In *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, volume 4, pages iv/553–iv/556, March 2005.
- [18] A. L. Da Cunha and M. N. Do. On two-channel filter banks with directional vanishing moments. *IEEE Trans. Image Process.*, 16(5):1207–1219, May 2007.
- [19] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, 41(7):909–996, 1988.
- [20] I. Daubechies, A. Grossmann, and Y. Meyer. Painless nonorthogonal expansions. *Journal of Mathematical Physics*, 27:1271–1283, May 1986.
- [21] I. Daubechies, B. Han, A. Ron, and Z. Shen. Framelets: Mra-based constructions of wavelet frames. *Appl. Comput. Harmon. Anal.*, 14(1):1 – 46, 2003.
- [22] L. Demanet and P. Vandergheynst. Gabor wavelets on the sphere. In *Proc. SPIE Int. Soc. Opt. Eng.*, volume 5207, 2003.
- [23] C. Diao and B. Han. Quasi-tight framelets with high vanishing moments derived from arbitrary refinable functions. *Applied and Computational Harmonic Analysis*, 2018.
- [24] R. J. Duffin and A. C. Schaeffer. A class of nonharmonic Fourier series. *Transactions of the American Mathematical Society*, 72(2):341–341, Feb. 1952.
- [25] M. Ehler. Compactly supported multivariate pairs of dual wavelet frames obtained by convolution. *International Journal of Wavelets, Multiresolution and Information Processing*, 06(02):183–208, 2008.

- [26] S. S. Goh and K. M. Teo. Extension principles for tight wavelet frames of periodic functions. *Appl. Comput. Harmon. Anal.*, 25(2):168–186, 2008.
- [27] P. Grohs and G. Kutyniok. Parabolic molecules. *Found. Comput. Math.*, 14(2):299–337, Apr. 2014.
- [28] K. Guo and D. Labate. Optimally sparse multidimensional representation using shearlets. *SIAM J. Math. Anal.*, 39:298–318, 2007.
- [29] A. Haar. Zur Theorie der orthogonalen Funktionensysteme. *Mathematische Annalen*, 69(3):331–371, Sep 1910.
- [30] B. Han. On dual wavelet tight frames. *Applied and Computational Harmonic Analysis*, 4(4):380–413, 1997.
- [31] B. Han. Compactly supported tight wavelet frames and orthonormal wavelets of exponential decay with a general dilation matrix. *Journal of Computational and Applied Mathematics*, 155(1):43–67, 2003. Approximation Theory, Wavelets, and Numerical Analysis.
- [32] B. Han. Nonhomogeneous wavelet systems in high dimensions. *Appl. Comput. Harmon. Anal.*, 32(2):169–196, 2012.
- [33] B. Han, Q. Jiang, Z. Shen, and X. Zhuang. Symmetric canonical quincunx tight framelets with high vanishing moments and smoothness. *Math. Comput.*, 87:347–379, 2018.
- [34] B. Han, G. Kutyniok, and Z. Shen. Adaptive multiresolution analysis structures and shearlet systems. *SIAM Journal on Numerical Analysis*, 49(5):1921–1946, 2011.
- [35] B. Han, T. Li, and X. Zhuang. Directional compactly supported box spline tight framelets with simple structure. *CoRR*, abs/1708.08421, 2017.
- [36] B. Han, Q. Mo, and Z. Zhao. Compactly supported tensor product complex tight framelets with directionality. *SIAM Journal on Mathematical Analysis*, 47(3):2464–2494, 2015.
- [37] B. Han and Z. Zhao. Tensor product complex tight framelets with increasing directionality. *SIAM Journal on Imaging Sciences*, 7(2):997–1034, 2014.
- [38] A. Hast. Simple filter design for first and second order derivatives by a double filtering approach. *Pattern Recognit. Lett.*, 42:65–71, 2014.
- [39] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

- [40] E. Hernandez and G. Weiss. *A first course on wavelets*. CRC Press, Boca Raton, FL, 1996.
- [41] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *CoRR*, abs/1502.03167, 2015.
- [42] J. H. Jacobsen, J. C. Van Gemert, Z. Lou, and A. W. M. Smeulders. Structured receptive fields in CNNs. *CoRR*, abs/1605.02971, 2016.
- [43] N. Kingsbury. Image processing with complex wavelets. *Phil. Trans. R. Soc. London A*, 357:2543–2560, 1999.
- [44] P. Kittipoom, G. Kutyniok, and W. Q. Lim. Construction of compactly supported shearlet frames. *Constr. Approx.*, 35(1):21–72, Feb 2012.
- [45] J. Kovacevic and M. Vetterli. Nonseparable multidimensional perfect reconstruction filter-banks. *IEEE Trans. Inf. Theory*, 38:533–555, 1992.
- [46] G. Kutyniok and D. Labate. Resolution of the wavefront set using continuous shearlets. *Trans. Amer. Math. Soc.*, 361(5):2719–2754, 2009.
- [47] D. Labate, W. Lim, G. Kutyniok, and G. Weiss. Sparse multidimensional representation using shearlets. *SPIE Proc. 5914, SPIE, Bellingham*, 2005.
- [48] M. Lin, Q. Chen, and S. Yan. Network In Network. *arXiv e-prints*, page arXiv:1312.4400, Dec 2013.
- [49] Y. Lu and M. N. Do. The finer directional wavelet transform. In *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, volume 4, pages iv/573–iv/576, March 2005.
- [50] Y. M. Lu and M. N. Do. Multidimensional directional filter banks and surfacelets. *IEEE Trans. Image Process.*, 16(4):918–931, April 2007.
- [51] S. G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11(7):674–693, Jul 1989.
- [52] Y. Meyer. *Wavelets and Operators*, volume 1 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1993.
- [53] M. Papadakis, B. G. Bodmann, S. K. Alexander, D. Vela, S. Baid, A. A. Gittens, D. J. Kouri, S. D. Gertz, S. Jain, J. R. Romero, X. Li, P. Cherukuri, D. D. Cody, G. W. Gladish, I. Aboshady, J. L. Conyers, and S. W. Casscells. Texture-based tissue characterization for high-resolution CT-scans of coronary arteries. *Commun. Numer. Methods Eng.*, 25(6):597–613, 2009.

- [54] A. Ron and Z. Shen. Affine system in  $\mathcal{L}^2(\mathbb{R}^d)$ : The analysis of the analysis operator. *J. Funct. Anal.*, (148):408–447, 1997.
- [55] I. W. Selesnick and L. Sendur. Iterated oversampled filter banks and wavelet frames. In M. Unser A. Aldroubi, A. Laine, editor, *Proc. Wavelet Applications in Signal and Image Processing VIII*, volume 4119 of *Proceedings of SPIE*, 2000.
- [56] E.P. Simoncelli and W.T. Freeman. The steerable pyramid: A flexible architecture for multi-scale derivative computation. *Proc. IEEE International Conference on Image Processing*, 1995.
- [57] M. Vetterli and J. Kovacevic. *Wavelets and subband coding*. Prentice Hall PTR, Englewood Cliffs, NJ, 1995.