Università degli Studi di Genova

Dipartimento di Matematica

Corso di studi in

Matematica

Anno Accademico 2022/2023

Tesi di Laurea Magistrale

Wavelets, Shearlets, and Non-Linear Approximation

Candidato
Simone Sanna

Relatore
Prof. Giovanni Alberti

Correlatore
Prof. Filippo De Mari

Contents

Introduction

Due to advancements in technology that have made data acquisition easier and more cost-effective, we are currently dealing with a massive influx of data. This data necessitates efficient analysis and processing. The magnitude of this challenge is apparent not just from the big volume of data but also from the diversity of data types, and the range of processing tasks required. In order to address tasks that span from analyzing features to classifying and compressing data, advanced mathematical and computational methods are required. A fundamental characteristic of nearly all data encountered in real-world applications is that the essential information that needs to be extracted is sparse. In other words, the crucial information of a datum is often situated on low-dimensional structures. In principle, this allows us to efficiently represent the important information using only a small number of terms from a suitable dictionary. Moreover, discovering a dictionary that can efficiently represent a specific class of data in a sparse manner requires a deep understanding of the main characteristics of that data class, which are typically linked to their geometric features. For example, in natural two-dimensional images the essential of their information is contained along its edges, which, in general, are described by one dimensional curves. Hence, to effectively analyze and represent these data, we need to use a dictionary which is able to truly understand and characterize their geometric structures.

Applied harmonic analysis has become the central field within applied mathematics dedicated to the analysis and the representation of data. The primary task of this discipline is the practice of analyzing an object, to acquire a deeper understanding of it. For instance, if we aim to study signals from a class of data $\mathcal{D}$ in a separable Hilbert space $\mathcal{H}$, we need to choose a countable collection of analyzing functions $\{\psi_i\}_{i\in I} \subseteq \mathcal{H}$ such that, for every element $f \in \mathcal{D}$, we have

$$f = \sum_{i\in I} a_i(f)\psi_i.$$

This equation not only offers a way to break down the signal f into a set of measurements $\{a_i(f)\}_{i\in I}$, but also illustrates the procedure of reconstruction of f from its coefficients. Here, there is an important occurrence to note. If the dictionary $\{\psi_i\}_{i\in I}$ is an orthonormal basis, then the coefficients sequence is uniquely determined, and the reconstruction is stable. If we allow more flexibility, for example,

by opting for a redundant dictionary, we no longer have the uniqueness of the sequence $\{a_i(f)\}_{i\in I}$. On one hand, this allows us to choose a sparser coefficients sequence, but, on the other hand, it could generate instability while reconstructing. This bottleneck has been overcome thanks to *frame theory*, indeed frames systems are redundant dictionaries which guarantee stability in signal reconstruction, and for this reason are widely used in applied harmonic analysis and signal processing. Once the appropriate system to analyze a certain class of data is chosen, we aim to study its sparsifying properties. The degree of sparsity is measured as the decay rate of the error of *best N-term approximation*. Roughly speaking, this approximation consists in approximating f by selecting the indices associated with its N largest amplitude coefficients $a_i(f)$.

The first analyzing dictionaries we introduce in the thesis are one-dimensional wavelet bases. In this case, the system is obtained by dilating and translating a generating function ψ, called a *mother wavelet*,

$$\{\psi_{j,n} = 2^{\frac{j}{2}}\psi(2^j \cdot -n) : j, n \in \mathbb{Z}\}.$$

Chapters 1 and 2 are dedicated to formally introduce *continuous wavelet systems*, *discrete wavelet systems*, and the corresponding *wavelet transforms*. A wavelet system allows us to decompose a signal at different scales, and locations. This behaviour is mathematically explained by the link between wavelet bases and *multiresolution analysis*; an increasing sequence of closed subspaces of $L^2(\mathbb{R})$, invariant for translations and generated by dilations, which enable the construction of *orthonormal wavelets*. In addition, these chapters aim to underlie the main features we require a wavelet to possess in order to have better sparsifying properties: *vanishing moments*, and *compact support*.

In Chapter 3, we describe the *non-linear approximation* on a general Hilbert space $\mathcal{H}$, and then we present the main results for non-linear approximation of one-dimensional piecewise regular functions through wavelet bases, Theorem 3.3. Vanishing moments and compact supports enable to strongly compress these types of signals. Indeed, the first property guarantees that, if the wavelet is supported within an interval where the signal is regular, then the corresponding coefficient will be negligible. The second property allows us to control the number of wavelets whose supports intersect a singularity of the signal, i.e it guarantees that the number of relevant coefficient cannot be too large. These results emphasize that wavelet systems efficiently deal with one-dimensional signal possessing a finite number of pointwise singularities. In a natural way, one can extend the construction of wavelets to $\mathbb{R}^n$, $n > 1$, and study the corresponding approximation problem on higher dimensions. It comes out that the isotropic features of wavelet systems do not allow to obtain the same decay rate of the error as the one-dimensional case (see Theorem 3.4). Indeed, multidimensional wavelets are supported on cubes that, through the dilation parameter, can only be enlarged or reduced without changing their shape. This makes multidimensional wavelets

unable to efficiently deal with singularities distributed along curves, or in general, manifolds.

A research branch, focused on building new function systems capable of solving this problem, has emerged. In particular, in [Donoho, 2001], Donoho demonstrated that for every cartoon-like image f, i.e. a function that is C^2 away of a C^2 curve (Definition 4.1), and for any N, there exists a triangulation of $[0,1]^2$ with N triangles so that the piecewise linear interpolation f_N of these triangles obeys

$$\|f - f_N\|_2^2 \lesssim N^{-2}, \qquad N \to +\infty.$$

Moreover, it can be proved that this result is optimal. Therefore, on one hand, we have multidimensional wavelet bases, which are easy to handle numerically, but provide a poor error estimate. On the other hand, Donoho's theorem offers an optimal error estimate, but it is not very useful in applications. Nevertheless, the argument presented by Donoho shows that the capability of elongating and orienting the supports of the functions in the dictionary along the singularity set of the function we are analyzing is fundamental in order to achieve the optimal error decay estimate. This means that *anisotropic* dilations are crucial in order to obtain a good error estimate. Among the various attempts made in the early 2000s, *curvelets* [Candès and Donoho, 2004], and *shearlets* [Labate et al., 2005] are those which allow us to nearly obtain the best error estimate. Indeed, it has ben proved that for a cartoon-like image f, the non-linear approximation error, carried out by selecting the N largest amplitude coefficients, satisfies

$$\|f - f_N\|_2^2 \lesssim N^{-2} \log^3(N), \qquad N \to +\infty.$$

Compared to Donoho's method, shearlets and curvelets have the advantage of being *nonadaptive*, which means that the system we choose to analyze does not depend on the particular signal we are analyzing. Therefore, they are more useful in applications. In particular, in the thesis we present the construction of shearlet systems, and their non-linear approximation properties.

In Chapter 4, following the line of the first two chapters, we define *shearlet continuous systems*, *shearlet discrete systems*, and the *shearlet transform*. Then, we focus on the construction of a slightly different system which is the *cone-adapted shearlet system*. This latter was introduced in order to overcome a directional bias possessed by ordinary shearlet systems, which is discussed in Section 4.2. In Section 4.3, we present the main results about non-linear approximation of cartoon-like images.

As explained above, shearlet systems provide optimal error rate when approximating a function which is C^2 away of a C^2 curve, and so they are really efficient in sparsely representing these types of signals. For many purposes, it is not only interesting to know that images can be faithfully represented by sparse vectors, but it is of great interest also to study the structure of the set of the index corresponding to the largest coefficients. A priori, when carrying out a non-linear

approximation, we only know that if we choose the largest coefficients, then we get a good approximation of the signal, but we do not know anything about their location. The multiresolution properties of the one-dimensional wavelets allow us to think that it is possible to characterize more precisely the set of the relevant coefficients. Indeed, as previously mentioned, if we analyze a piecewise regular function, the largest coefficients are individuated by the location of the singularities. Therefore, when analyzing the signal through different scales, it is possible to define hierarchical relations between coefficients at scale j and those at scale $j+1$ with the property that: if a coefficient at scale $j+1$ is relevant, then so is its parent. These structures are referred to as *trees*, and in this case we speak of *tree approximation*, we refer the interested reader to [Cohen et al., 2001] and [Kekkonen et al., 2023]. Analogously, in the case of shearlet frames, their multiresolution properties, and their geometric structure invite us to a deeper study the locations of the relevant shearlet coefficients. One of the works that delves deeper into the tree structure in the case of shearlets is [Grohs, 2012]. In this paper, the author introduces a parent-child relation, and then he analyzes the corresponding tree approximation. Although it is shown that the error is optimal, the proofs presented do not involve the structure defined. In Chapter 5, we propose to explicitly exploit this structure in the case of a particular function, where the singularity curve is a straight line, and it is constant away of this line. We execute a quantitative analysis of its shearlet coefficients, and we prove that the relevant ones satisfy the hierarchical relations introduced in Definition 5.1, and so that the set of their location is a tree.

Chapter 1

Continuous Wavelet Transform

In this chapter we briefly introduce mother wavelets, the associated wavelet transforms, and we discuss the main results related to these tools. In particular, we focus on those results that state under which conditions on the mother wavelets we are able to reconstruct functions in $L^2(\mathbb{R})$.

We refer to [Mallat, 1999] for the detailed proofs of the section and for further details about the topic.

Before starting the central part of this chapter, keep in mind that throughout the thesis we use the following normalization for the Fourier transform on $\mathbb{R}^n$

$$\mathcal{F}f(\xi) := \int_{\mathbb{R}^n} f(x)\, \mathrm{e}^{-i\xi x}\, dx, \qquad \xi \in \mathbb{R}^n,$$

where the product ξx is to be understood, for $n > 1$, as a scalar product. Let us also recall that this normalization leads to the following Plancharel identity

$$\langle f, g\rangle = \frac{1}{2\pi}\langle \mathcal{F}f, \mathcal{F}g\rangle, \qquad f, g \in L^2(\mathbb{R}).$$

Definition 1.1. A function $\psi \in L^2(\mathbb{R})$ such that $\|\psi\|_2 = 1$ is said to be a *mother wavelet* if

$$C_\psi := \int_0^\infty \frac{|\mathcal{F}\psi(\xi)|^2}{\xi} d\xi < +\infty. \tag{1.1}$$

From a mother wavelet we can define the family of functions

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right), \qquad (u, s) \in \mathbb{R} \times \mathbb{R}_+.$$

The condition $\|\psi\|_2 = 1$ ensures, by a simple change of variable, that $\|\psi_{u,s}\|_2 = 1$. One can think of ψ as a function supported in a neighborhood of the origin, so the family $\psi_{u,s}$ can be used to analyze signal structures at different scales and time intervals through the *continuous wavelet transform*.

Definition 1.2. Let $\psi \in L^2(\mathbb{R})$ be a mother wavelet. For all $f \in L^2(\mathbb{R})$ the *continuous wavelet transform* is defined as

$$\mathcal{W}_\psi f(u,s) := \langle f, \psi_{u,s} \rangle = \int_R f(t) \frac{1}{\sqrt{s}} \overline{\psi}\left(\frac{t-u}{s}\right) dx, \qquad (u,s) \in \mathbb{R} \times \mathbb{R}_+. \quad (1.2)$$

The continuous wavelet transform can be reformulated as a convolution product

$$\mathcal{W}_\psi f(u,s) = f * \tilde{\psi}_s(u),$$

where $\tilde{\psi}_s(t) := \frac{1}{\sqrt{s}} \overline{\psi}\left(\frac{-t}{s}\right)$.

The technical condition (1.1), also known as *admissibility condition*, or *Calderón condition* is useful to reconstruct a signal $f \in L^2(\mathbb{R})$ via its wavelet coefficients.

Theorem 1.3. *[Mallat, 1999] Let $\psi \in L^2(\mathbb{R})$ be a real valued mother wavelet. Every function $f \in L^2(\mathbb{R})$ satisfies the following reconstruction formula*

$$f(t) = \frac{1}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} \mathcal{W}_\psi f(u,s) \psi_{u,s}(t) \, du \frac{ds}{s^2}, \ \text{a.e. } t \in \mathbb{R}. \quad (1.3)$$

Moreover

$$\int_{-\infty}^{+\infty} |f(t)|^2 dt = \frac{1}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} |\mathcal{W}_\psi f(u,s)|^2 \, du \frac{ds}{s^2}. \quad (1.4)$$

While the continuous wavelet transform can be defined for every $\psi \in L^2(\mathbb{R})$, the previous theorem shows that is important to choose wavelets that satisfy the admissibility condition (1.1). Hence, it is useful to find sufficient conditions that ensure (1.1). It is obvious that a necessary condition for (1.1) is that $\mathcal{F}\psi(0) = 0$ (i.e. ψ has zero average), but unfortunately it is not sufficient. In order to obtain a sufficient condition, we require also that $\mathcal{F}\psi$ is C^1 in a neighborhood of the origin, i.e. there exists $\epsilon > 0$ such that $\mathcal{F}\psi \in C^1(-\epsilon, \epsilon)$. Indeed, we have that there exists $M > 0$ such that

$$\left| \frac{d\mathcal{F}\psi}{d\xi}(\xi) \right| \leq M \text{ for every } \xi \in \left[0, \frac{\epsilon}{2}\right].$$

Hence, for each $\xi \in \left[0, \frac{\epsilon}{2}\right]$,

$$|\mathcal{F}\psi(\xi)| \leq \int_0^\xi \left| \frac{d\mathcal{F}\psi}{d\omega}(\omega) \right| d\omega \leq M\xi.$$

Now, by splitting the integral in (1.1), we obtain

$$\begin{aligned} C_\psi &= \int_0^{\frac{\epsilon}{2}} \frac{|\mathcal{F}\psi(\xi)|^2}{\xi} d\xi + \int_{\frac{\epsilon}{2}}^{+\infty} \frac{|\mathcal{F}\psi(\xi)|^2}{\xi} d\xi \\ &\leq M^2 \frac{\epsilon^2}{8} + \int_{\frac{\epsilon}{2}}^{+\infty} \frac{|\mathcal{F}\psi(\xi)|^2}{\xi} d\xi < +\infty, \end{aligned}$$

due to the Hölder inequality applied to the functions $\xi \mapsto \frac{1}{\xi}$, which is bounded over $\left[\frac{\epsilon}{2}, +\infty\right)$, and $|\mathcal{F}\psi|^2$, which belongs to $L^1\left(\left[\frac{\epsilon}{2}, +\infty\right)\right)$. Moreover, we can use the time decay of ψ as a sufficient condition for the regularity of its Fourier transform. In particular, one has that $\mathcal{F}\psi$ is C^1 over $\mathbb{R}$ if

$$\int_{\mathbb{R}} (1 + |t|) \, |\psi(t)| dt < +\infty.$$

So, if $\int_{\mathbb{R}} \psi(t) \, dt = 0$, and $\int_{\mathbb{R}} (1 + |t|) \, |\psi(t)| dt < +\infty$, then (1.1) holds.

An important function associated to a mother wavelet is the *scaling function*. When the wavelet transform is known only for $s < s_0$ for a certain $s_0 > 0$, to recover completely f we need the complement information contained in $\mathcal{W}_\psi f(u, s)$ for $s \geq s_0$. The scaling function helps us to obtain this information.

Definition 1.4. Let ψ be a mother wavelet. A *scaling function* $\phi \in L^2(\mathbb{R})$ is a function that satisfies

$$|\mathcal{F}\phi(\xi)|^2 = \int_{\xi}^{+\infty} \frac{|\mathcal{F}\psi(\omega)|^2}{\omega} d\omega, \tag{1.5}$$

and whose complex phase can be arbitrarily chosen.

Obviously (1.1) implies

$$\lim_{\xi \to 0} |\mathcal{F}\phi(\xi)|^2 = C_\psi, \tag{1.6}$$

and it is easy to verify that $\|\phi\|_2 = 1$.

Let us define

$$Lf(u, s) := \langle f, \phi_{u,s} \rangle = f * \tilde{\phi}_s(u). \tag{1.7}$$

Following the scheme of the proof of Theorem 1.3, one can prove that, for every $f \in L^2(\mathbb{R})$, the following formula holds:

$$f(t) = \frac{1}{C_\psi} \int_0^{s_0} \mathcal{W}_\psi f(\cdot, s) * \psi_s(t) \frac{ds}{s^2} + \frac{1}{C_\psi s_0} Lf(\cdot, s_0) * \phi_{s_0}(t) \ \ a.e.\ t \in \mathbb{R}, \tag{1.8}$$

where $\psi_s(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t}{s}\right)$.

Chapter 2

Orthonormal Wavelets

Having discussed the continuous transform and its reconstruction properties, we now want to introduce a discretised version of it. Indeed, since wavelets are widely used in applications to efficiently represent signals, we need to discuss wavelets properties in a discrete setting and to understand when these discrete collections of wavelets are useful to represent signals.

In this chapter, we define an *orthonormal wavelet* as a function such that the family

$$\{\psi_{j,n} := 2^{\frac{j}{2}}\psi\left(2^j \cdot -n\right) : j, n \in \mathbb{Z}\}$$

forms an orthonormal basis of $L^2(\mathbb{R})$. Then, we note that it can be related to a *multiresolution analysis* as pointed out in Section 2.1, and we discuss a general way for constructing orthonormal wavelets. From the wavelet construction in Section 2.1.1, we can understand that exist wavelets of such a different nature. Hence, we want to better understand which types of wavelets we are interested in, in particular which properties they have to satisfy in order to be efficient in signal representation. In Section 2.2, we address this question, and we discuss the main properties, such as *vanishing moments*, and *compact support*, and the main related results. We conclude the chapter by briefly presenting construction schemes for wavelet bases on intervals, and on multidimensional spaces. This chapter follows the topics covered in [Mallat, 1999, Hernández and Weiss, 1996, Adcock and Hansen, 2021]. Whenever the detailed proofs are not presented, we will provide the exact reference.

Definition 2.1. An *orthonormal wavelet* is a function $\psi \in L^2(\mathbb{R})$, $\|\psi\|_2 = 1$ such that the family of the translated and dilated functions

$$\{\psi_{j,n} := 2^{\frac{j}{2}}\psi\left(2^j \cdot -n\right) : j, n \in \mathbb{Z}\}$$

is an orthonormal basis of $L^2(\mathbb{R})$.

If ψ is supported in a neighborhood of the origin, then $\psi_{j,n}$ is supported in a neighborhood of $\frac{n}{2^j}$ with the support size of the order of $O(2^{-j})$. Hence, at

fixed $j \in \mathbb{Z}$, the wavelet coefficients $\langle f, \psi_{j,n} \rangle$ of a signal $f \in L^2(\mathbb{R})$ carry local information of f near $\frac{n}{2^j}$ at scale 2^{-j} through its average with $\psi_{j,n}$. We can say that the sequence $\{\langle f, \psi_{j,n} \rangle\}_{n \in \mathbb{Z}}$ contains the details of f at scale 2^{-j}, and a signal f is reconstructed by adding its details at scale 2^{-j} for each $j \in \mathbb{Z}$

$$f = \sum_{j,n \in \mathbb{Z}} \langle f, \psi_{j,n} \rangle \psi_{j,n}. \tag{2.1}$$

Similarly to the continuous wavelet transform, if we only have access to the details of f for $j \geq j_0$, then it is possible to retrieve the information for $j \leq j_0$ by considering a *scaling function* $\varphi \in L^2(\mathbb{R})$ such that the projection of f onto the space generated by

$$\{\varphi_{j_0,n} = 2^{\frac{j_0}{2}} \varphi\left(2^{j_0} \cdot -n\right) : n \in \mathbb{Z}\}$$

provides an approximation of f at scale 2^{-j_0}

$$f = f_{j_0} + \sum_{j=j_0}^{+\infty} \sum_{n \in \mathbb{Z}} \langle f, \psi_{j,n} \rangle \psi_{j,n}, \tag{2.2}$$

where

$$f_{j_0} = \sum_{n \in \mathbb{Z}} \langle f, \varphi_{j_0,n} \rangle \varphi_{j_0,n}.$$

Example 2.2 (The Haar wavelet)**.** A classical example of orthonormal wavelet is the *Haar wavelet*, introduced by Alfréd Haar in 1909. We will use this wavelet as an example to illustrate the results in the next sections. The Haar wavelet is defined by

$$\psi(x) := \begin{cases} 1 & \text{if } x \in [0, \frac{1}{2}), \\ -1 & \text{if } x \in (\frac{1}{2}, 1], \\ 0 & \text{otherwhise.} \end{cases} \tag{2.3}$$

It is associated to a scaling function

$$\varphi(x) := \begin{cases} 1 & \text{if } x \in [0, 1), \\ 0 & \text{otherwhise.} \end{cases} \tag{2.4}$$

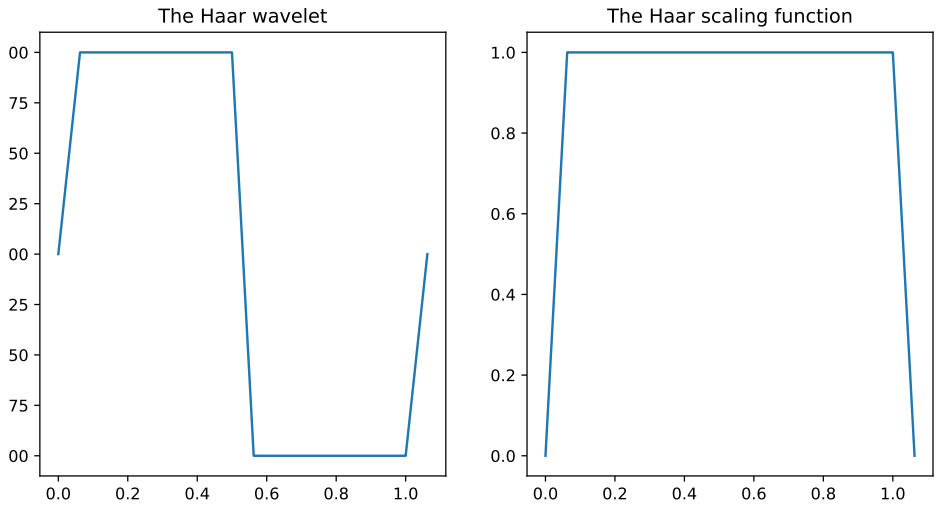


Figure 2.1: On the left the Haar wavelet ψ, on the left its scaling function φ.

The elements of the Haar basis are

$$\psi_{j,n}(x) = \begin{cases} 2^{\frac{j}{2}} & \text{if } x \in [\frac{n}{2^j}, \frac{n}{2^j} + \frac{1}{2^{j+1}}), \\ -2^{\frac{j}{2}} & \text{if } x \in (\frac{n}{2^j} + \frac{1}{2^{j+1}}, \frac{n+1}{2^j}], \\ 0 & \text{otherwhise.} \end{cases} \tag{2.5}$$

It is easy to prove that they form an orthonormal system. The fact that they are also a basis is a consequence of Theorem 2.9 below, see Example 2.10.

We can think of orthonormal wavelets as a way of analysing a signal at different resolutions; indeed the concept of orthonormal wavelets can be naturally related to a *multiresolution analysis* (MRA) that formalises the ideas expressed above. Moreover, an MRA provides a general method for constructing an orthonormal wavelet.

2.1 Multiresolution Analysis

Definition 2.3. A *multiresolution analysis* (MRA) is a sequence $\{V_j\}_{j\in\mathbb{Z}}$ of closed subspaces of $L^2(\mathbb{R})$ such that

(a) $V_j \subseteq V_{j+1}$ for all $j \in \mathbb{Z}$,

(b) $f \in V_j$ if and only if $f(2\cdot) \in V_{j+1}$ for all $j \in \mathbb{Z}$,

(c) $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$,

(d) $\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R})$,

(e) There exists a *scaling function* $\varphi \in V_0$, such that $\{\varphi(\cdot - n) : n \in \mathbb{Z}\}$ is an orthonormal basis of V_0.

Example 2.4 (The MRA of the Haar wavelet)**.** Let us define the spaces V_j as

$$V_j := \{f \in L^2(\mathbb{R}) : f|_{[\frac{n}{2^j}, \frac{n+1}{2^j}]} \text{ is constant}, n \in \mathbb{Z}\} \tag{2.6}$$

It is easily verified that the family $\{V_j\}_{j\in\mathbb{Z}}$ satisfies property $(a) - (d)$ of the definition above. Moreover, it is clear that the family of the integer translations of φ as in (2.4) forms an orthonormal basis of V_0.

An MRA is an increasing sequence of closed subspaces that are invariant by translations (property (e)), and generated by dilations (property (b)). Property (d) is useful because it allows us to approximate with arbitrary accuracy a function $f \in L^2(\mathbb{R})$ by its orthogonal projection on V_j. As an immediate consequence of (b) and (e), we also observe that

$$\{\varphi_{j,n} := 2^{j/2}\varphi(2^j \cdot -n) : n \in \mathbb{Z}\}$$

is an orthonormal basis for V_j.

The next Lemma, shows that the five properties of an MRA are not completely independent. Nevertheless, it is useful to enumerate them all to make the role of the V_j spaces clearer.

Lemma 2.5. *Properties* $(a), (b)$ *and* (e) *imply property* (c).

Proof. Suppose that there exists a non-zero function $f \in \bigcap_{j\in\mathbb{Z}} V_j$. Without loss of generality we can assume that $\|f\|_2 = 1$. In particular, $f \in V_{-j}$ for every $j \in \mathbb{Z}$. By using property (b) of the previous definition, we have that $f_j = 2^{\frac{j}{2}} f(2^j \cdot) \in V_0$ and, by a change of variable, we have that $\|f\|_2 = \|f_j\|_2 = 1$.

Given that $\{\varphi(\cdot - k) : k \in \mathbb{Z}\}$ forms a basis of V_0, we have that there exists a unique sequence $\{\alpha_k^j\}_{k\in\mathbb{Z}} \subset \ell^2(\mathbb{Z})$ such that

$$f_j = \sum_{k\in\mathbb{Z}} \alpha_k^j \varphi(\cdot - k),$$

with convergence in $L^2(\mathbb{R})$, and such that

$$\sum_{k\in\mathbb{Z}} |\alpha_k^j|^2 = \|f_j\|_2^2 = 1.$$

By taking the Fourier transform, we have

$$\mathcal{F}f_j(\xi) = 2^{-\frac{j}{2}} \mathcal{F}f\left(\frac{\xi}{2^j}\right),$$

$$\mathcal{F}f_j(\xi) = \mathcal{F}\left(\sum_{k\in\mathbb{Z}} \alpha_k^j \varphi(\cdot - k)\right)(\xi) = \sum_{k\in\mathbb{Z}} \alpha_k^j \mathcal{F}(\varphi(\cdot - k))(\xi) = \sum_{k\in\mathbb{Z}} \alpha_k^j e^{-ik\xi} \mathcal{F}\varphi(\xi).$$

By setting

$$m_j(\xi) := \sum_{k\in\mathbb{Z}} \alpha_k^j e^{-ik\xi},$$

we obtain

$$\mathcal{F}f(\xi) = 2^{\frac{j}{2}} m_j(2^j\xi)\mathcal{F}\varphi(2^j\xi). \tag{2.7}$$

Observe that m_j is 2π-periodic, and, since $\{\alpha_k^j\}_{k\in\mathbb{Z}} \subset \ell^2(\mathbb{Z})$, it is also square-integrable on the one-dimensional torus $\mathbb{T} := \mathbb{R}/\mathbb{Z}$. In particular, it obeys

$$\|m_j\|^2_{L^2(\mathbb{T})} \leq \sum_{k\in\mathbb{Z}} |\alpha_k^j|^2 = 1,$$

where for each $m \in L^2(\mathbb{T})$

$$\|m\|^2_{L^2(\mathbb{T})} := \frac{1}{2\pi}\int_{-\pi}^{\pi} |m(x)|^2 dx.$$

Therefore, from (2.7), we obtain

$$\int_{2\pi}^{4\pi} |\mathcal{F}f(\xi)|d\xi \leq 2^{\frac{j}{2}} \left(\int_{2\pi}^{4\pi} |\mathcal{F}\varphi(2^j\xi)|^2 d\xi\right)^{1/2} \left(\int_{2\pi}^{4\pi} |m_j(2^j\xi)|^2 d\xi\right)^{1/2}.$$

By the change of variable $\xi \mapsto 2^{-j}\xi$, we have

$$\begin{aligned}\int_{2\pi}^{4\pi} |\mathcal{F}f(\xi)|d\xi &\leq 2^{-\frac{j}{2}} \left(\int_{2^{j+1}\pi}^{2^{j+2}\pi} |\mathcal{F}\varphi(\xi)|^2 d\xi\right)^{1/2} \left(\int_{2^{j+1}\pi}^{2^{j+2}\pi} |m_j(\xi)|^2 d\xi\right)^{1/2} \\ &\leq \left(\int_{2^{j+1}\pi}^{\infty} |\mathcal{F}\varphi(\xi)|^2 d\xi\right)^{1/2} \left(\frac{1}{2^j}\int_{2^{j+1}\pi}^{2^{j+2}\pi} |m_j(\xi)|^2 d\xi\right)^{1/2}.\end{aligned}$$

Observe that

$$\int_{2^{j+1}\pi}^{2^{j+2}\pi} |m_j(\xi)|^2 d\xi = \sum_{l=0}^{2^j-1} \int_{2^{j+1}\pi+2l\pi}^{2^{j+1}\pi+2(l+1)\pi} |m_j(\xi)|^2 d\xi \leq 2^{j+1}\pi.$$

Hence,

$$\int_{2\pi}^{4\pi} |\mathcal{F}f(\xi)|d\xi \leq \left(2\pi\int_{2^{j+1}\pi}^{\infty} |\mathcal{F}\varphi(\xi)|^2 d\xi\right)^{1/2}.$$

Taking the limit for $j \to \infty$, we can conclude that $\mathcal{F}f = 0$ on $[2\pi, 4\pi]$ due to the fact that $\mathcal{F}\varphi \in L^2(\mathbb{R})$. We can apply the same argument to $2^{\frac{l}{2}}\mathcal{F}f(2^l\cdot)$ to obtain that $\mathcal{F}f = 0$ on $2^l[2\pi, 4\pi]$ for each $l \in \mathbb{Z}$. So $\mathcal{F}f = 0$ on $(0, \infty)$. Applying the same argument to $[-4\pi, -2\pi]$, we obtain $\mathcal{F}f = 0$ on $(-\infty, 0)$. □

Moreover, there exists a characterisation of property (d) in terms of the Fourier transform of φ.

Lemma 2.6. *[Hernández and Weiss, 1996] Let $\{V_j\}_{j\in\mathbb{Z}}$ be a sequence of closed subspaces of $L^2(\mathbb{R})$ satisfying properties $(a), (b)$ and (e). If φ is such that $|\mathcal{F}\varphi|$ is continuous at 0, then the following are equivalent:*

(a) $\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R})$,

(b) $\mathcal{F}\varphi(0) \neq 0$.

Moreover, when either is the case, $|\mathcal{F}\varphi(0)| = 1$.

2.1.1 Wavelet Construction from a Multiresolution Analysis

We now show a general method for constructing orthonormal wavelets from MRAs. Let us describe the construction.

First of all, we define the space W_0 as follows:

$$W_0 := \{g \in V_1 : g \perp f \text{ for each } f \in V_0\}.$$

So W_0 is the orthogonal complement of V_0 in V_1, namely

$$V_1 = V_0 \oplus W_0.$$

Hence, due to the property (b) of an MRA, we have that

$$W_j := \{g(2^j\cdot) : g \in W_0\} = \{g \in V_{j+1} : g \perp f \text{ for each } f \in V_j\}.$$

In other words, for each $j \in \mathbb{Z}$, we have

$$V_{j+1} = V_j \oplus W_j.$$

This means that the spaces W_j contain the details that are lost when a function in V_{j+1} is approximated by its projection on V_j. Hence, for any $j_0 \leq j$,

$$V_{j+1} = V_{j_0} \oplus W_{j_0} \oplus W_{j_0+1} \oplus \cdots \oplus W_j. \tag{2.8}$$

Therefore properties (c) and (d) of an MRA imply the following decomposition formulae

$$L^2(\mathbb{R}) = V_{j_0} \oplus W_{j_0} \oplus W_{j_0+1} \oplus \ldots \tag{2.9}$$

$$L^2(\mathbb{R}) = \bigoplus_{j\in\mathbb{Z}} W_j. \tag{2.10}$$

Observe that (2.9) is equivalent to the decomposition of f in (2.2), and (2.10) is equivalent to the decomposition in (2.1). In particular, due to (2.10), we have

that finding an orthonormal wavelet is equivalent to finding a function $\psi \in W_0$ such that the family of its translates

$$\{\psi_{0,n} = \psi(\cdot - n) : n \in \mathbb{Z}\} \tag{2.11}$$

forms an orthonormal basis of W_0. Indeed, similarly to the scaling function, if (2.11) is an orthonormal basis of W_0, then

$$\{\psi_{j,n} := 2^{\frac{j}{2}}\psi\left(2^j \cdot -n\right) : n \in \mathbb{Z}\}$$

is an orthonormal basis of W_j. Therefore, due to (2.10),

$$\{\psi_{j,n} : j, n \in \mathbb{Z}\}$$

is an orthonormal basis of $L^2(\mathbb{R})$.

Let us now consider the function

$$\frac{1}{\sqrt{2}}\varphi\left(\frac{\cdot}{2}\right) \in V_{-1} \subset V_0.$$

Due to (e), we have

$$\frac{1}{\sqrt{2}}\varphi\left(\frac{\cdot}{2}\right) = \sum_{n\in\mathbb{Z}} h_n \varphi(\cdot - n),$$

where

$$h_n = \frac{1}{\sqrt{2}}\int_{\mathbb{R}} \varphi\left(\frac{x}{2}\right)\overline{\varphi(x-n)}dx. \tag{2.12}$$

The sequence $h = (h_n)_{n\in\mathbb{Z}}$ is referred to as the *filter* associated to φ, and h_n are the *filter coefficients*. Applying the Fourier transform, and using

$$\mathcal{F}(\varphi(\cdot - n))(\xi) = \mathrm{e}^{-in\xi}\,\mathcal{F}\varphi(\xi),$$

$$\mathcal{F}\left(\varphi\left(\frac{\cdot}{2}\right)\right)(\xi) = 2\mathcal{F}\varphi(2\xi),$$

we obtain

$$\mathcal{F}\varphi(2\xi) = \frac{1}{\sqrt{2}}\sum_{n\in\mathbb{Z}} h_n\,\mathrm{e}^{-in\xi}\,\mathcal{F}(\varphi)(\xi) = m_0(\xi)\mathcal{F}\varphi(\xi), \tag{2.13}$$

where

$$m_0(\xi) = \frac{1}{\sqrt{2}}\sum_{n\in\mathbb{Z}} h_n\,\mathrm{e}^{-in\xi} \tag{2.14}$$

is the so-called *transfer function* associated to the filter h. The transfer function m_0 satisfies the following property:

Lemma 2.7. *The function m_0 is 2π-periodic and satisfies the following partition of unity formula:*

$$|m_0(\xi)|^2 + |m_0(\xi+\pi)|^2 = 1, \ a.e.\ \xi \in \mathbb{R}. \tag{2.15}$$

The proof of this lemma relies on the following lemma that we do not prove.

Lemma 2.8. *[Adcock and Hansen, 2021] Let $g \in L^2(\mathbb{R})$. Then the following are equivalent*

(a) $\{g(\cdot - n) : n \in \mathbb{Z}\}$ *is an orthonormal system ;*

(b) $\sum_{n\in\mathbb{Z}} |\mathcal{F}g\,(\xi + 2n\pi)|^2 = 1, \ a.e.\ \xi \in \mathbb{R}.$

Proof Lemma 2.7. The previous Lemma, and property (e) of an MRA imply

$$\sum_{n\in\mathbb{Z}} |\mathcal{F}\varphi\,(2\xi + 2n\pi)|^2 = 1, \ a.e.\ \xi \in \mathbb{R}.$$

Using (2.13) we have

$$1 = \sum_{n\in\mathbb{Z}} |\mathcal{F}\varphi\,(\xi + n\pi)|^2 |m_0\,(\xi + n\pi)|^2 \ a.e.\ \xi \in \mathbb{R}.$$

By splitting the sum over the even and the odd integers, we get

$$1 = |m_0\,(\xi)|^2 \sum_{n\in\mathbb{Z}} |\mathcal{F}\varphi\,(\xi + 2n\pi)|^2 + |m_0\,(\xi+\pi)|^2 \sum_{n\in\mathbb{Z}} |\mathcal{F}\varphi\,(\xi + (2n+1)\,\pi)|^2.$$

We complete the proof by applying Lemma 2.8 once more. □

We are now ready to present the following theorem that shows how to construct an orthonormal wavelet from an MRA.

Theorem 2.9. *Let $\{V_j\}_{j\in\mathbb{Z}}$ be an MRA with scaling function φ and transfer function m_0 defined as in* (2.12) *and* (2.14)*. Let $\psi \in L^2(\mathbb{R})$ be the function*

$$\psi = \sqrt{2} \sum_{n\in\mathbb{Z}} (-1)^{n-1} \overline{h_n} \varphi\,(2\cdot + n - 1)\,, \tag{2.16}$$

and define

$$W_0 := \overline{\operatorname{span}\{\psi_{0,n} = \psi\,(\cdot - n) : n \in \mathbb{Z}\}}.$$

Then

(a) $\psi \in V_1$*;*

(b) $\{\psi_{0,n} : n \in \mathbb{Z}\}$ *is an orthonormal basis of W_0;*

(c) $V_0 \oplus W_0 = V_1$*;*

(d) ψ *is an orthonormal wavelet.*

Proof. By definition of an MRA, it is clear that (2.16) defines a function in V_1.

In order to prove (b), we observe that, due to Lemma 2.8, $\{\psi_{0,n} : n \in \mathbb{Z}\}$ is an orthonormal basis of W_0 if and only if

$$\sum_{n\in\mathbb{Z}} |\mathcal{F}\psi(2\xi + 2n\pi)|^2 = 1, \qquad a.e.\ \xi \in \mathbb{R}.$$

By applying the Fourier transform to (2.16), we obtain

$$\mathcal{F}\psi(\xi) = -\,\mathrm{e}^{-i\frac{\xi}{2}}\,\overline{m_0\left(\frac{\xi}{2}+\pi\right)}\,\mathcal{F}\varphi\left(\frac{\xi}{2}\right). \tag{2.17}$$

Now, using this expression, and arguing as in the proof of Lemma 2.7, we obtain

$$\sum_{n\in\mathbb{Z}} |\mathcal{F}\psi(2\xi + 2n\pi)|^2 = |m_0(\xi)|^2 + |m_0(\xi+\pi)|^2 = 1.$$

We now prove (c). In order to do that, we first show that $V_0 \perp W_0$. Since $\{\varphi_{0,n} : n \in \mathbb{Z}\}$ and $\{\psi_{0,n} : n \in \mathbb{Z}\}$ are respectively bases of V_0 and W_0, this is equivalent to prove

$$\langle \psi, \varphi_{0,k}\rangle = 0, \qquad k \in \mathbb{Z}.$$

$$\begin{aligned}\langle \psi, \varphi_{0,k}\rangle &= \int_{\mathbb{R}} \psi(x)\overline{\varphi(x-k)}dx = \frac{1}{2\pi}\int_{\mathbb{R}} \mathcal{F}\psi(\xi)\overline{\mathcal{F}\varphi(\xi)}\,\mathrm{e}^{ik\xi}\,d\xi \\ &= \int_0^{2\pi}\left(\sum_{n\in\mathbb{Z}} \mathcal{F}\psi(\xi+2n\pi)\overline{\mathcal{F}\varphi(\xi+2n\pi)}\right)\mathrm{e}^{ik\xi}\,d\xi.\end{aligned}$$

By proving

$$\sum_{n\in\mathbb{Z}} \mathcal{F}\psi(\xi+2n\pi)\overline{\mathcal{F}\varphi(\xi+2n\pi)} = 0, \qquad a.e.\ \xi \in \mathbb{R},$$

we can conclude the proof of the orthogonality between V_0 and W_0. Let consider the 2π-periodic function defined by $g(\xi) = \mathrm{e}^{-i\xi}\,\overline{m_0(\xi+\pi)}$. By using (2.13), and (2.17), we have

$$\sum_{n\in\mathbb{Z}} \mathcal{F}\psi(\xi+2n\pi)\overline{\mathcal{F}\varphi(\xi+2n\pi)} = \sum_{n\in\mathbb{Z}} g(\xi+n\pi)\overline{m_0(\xi+n\pi)}|\mathcal{F}\varphi(\xi+n\pi)|^2.$$

Splitting the sum into even and odd integers, and proceeding as in Lemma 2.7, we get

$$\sum_{n\in\mathbb{Z}} \mathcal{F}\psi(\xi+2n\pi)\overline{\mathcal{F}\varphi(\xi+2n\pi)} = g(\xi)\overline{m_0(\xi)} + g(\xi+\pi)\overline{m_0(\xi+\pi)},$$

where the right-hand-side is 0 for definition of g. To conclude the proof of (c), it remains to prove that $V_0 + W_0 = V_1$. This is equivalent to showing that for every sequence $a \in \ell^2(\mathbb{Z})$ there exist sequences $b, c \in \ell^2(\mathbb{Z})$ such that

$$2\sum_{n\in\mathbb{Z}} a_n \varphi(2x-n) = \sum_{n\in\mathbb{Z}} b_n \varphi(x-n) + \sum_{n\in\mathbb{Z}} c_n \psi(x-n).$$

Applying the Fourier transform, we obtain

$$\sum_{n\in\mathbb{Z}} a_n \, \mathrm{e}^{-in\frac{\xi}{2}} \, \mathcal{F}\varphi\left(\frac{\xi}{2}\right) = \sum_{n\in\mathbb{Z}} b_n \, \mathrm{e}^{-in\xi} \, \mathcal{F}\varphi(\xi) + \sum_{n\in\mathbb{Z}} c_n \, \mathrm{e}^{-in\xi} \, \mathcal{F}\psi(\xi).$$

Let introduce the functions $A(\xi) = \sum_{n\in\mathbb{Z}} a_n \, \mathrm{e}^{-in\xi}$, $B(\xi) = \sum_{n\in\mathbb{Z}} b_n \, \mathrm{e}^{-in\xi}$, and $C(\xi) = \sum_{n\in\mathbb{Z}} c_n \, \mathrm{e}^{-in\xi}$. Due to (2.13), and (2.17), we can conclude the proof if

$$A\left(\frac{\xi}{2}\right) = B(\xi) m_0\left(\frac{\xi}{2}\right) + C(\xi) g\left(\frac{\xi}{2}\right).$$

Hence, given $A \in L^2(\mathbb{T})$, we have to find $B, C \in L^2(\mathbb{T})$ such that the previous equality holds. By using the definitions of g, and Lemma 2.7, we can verify that the choices

$$\begin{aligned} B(\xi) &= A\left(\frac{\xi}{2}\right)\overline{m_0\left(\frac{\xi}{2}\right)} + A\left(\frac{\xi}{2}+\pi\right)\overline{m_0\left(\frac{\xi}{2}+\pi\right)}, \\ C(\xi) &= A\left(\frac{\xi}{2}\right)\overline{g\left(\frac{\xi}{2}\right)} + A\left(\frac{\xi}{2}+\pi\right)\overline{g\left(\frac{\xi}{2}+\pi\right)}, \end{aligned}$$

allow us to conclude this part of the proof.

Finally, (d) follows from (b), and (2.10). □

Example 2.10. Let us consider the MRA defined in (2.6), as explained in Example 2.4, it is associated to the scaling function (2.4). Therefore, one can compute the filter coefficients h_n using (2.12). An easy calculation shows that $h_0 = h_1 = \frac{1}{\sqrt{2}}$, and $h_n = 0$ otherwise. In particular, the corresponding transfer function is

$$m_0(\xi) = \frac{1}{\sqrt{2}}\left(1 + \mathrm{e}^{-i\xi}\right).$$

Now, by applying (2.16) with the filter coefficients computed above, we obtain the Haar wavelet as in (2.3)

$$\psi(x) = -\varphi(2x-1) + \varphi(2x) = \begin{cases} 1 & \text{if } x \in [0, \frac{1}{2}), \\ -1 & \text{if } x \in (\frac{1}{2}, 1], \\ 0 & \text{otherwhise.} \end{cases}$$

2.2 Wavelet Design

From the previous sections of this chapter, one can infer that it is possible to construct a large range of wavelet bases generated by orthonormal wavelets with different features. In this section, we want to understand what type of wavelets we want to construct. In particular, we ask ourselves which properties we would like an orthonormal wavelet to satisfy. Our aim is to construct wavelets that allow signals to be represented by compressible sequences, i.e. such that most wavelet coefficients are close to zero, so that they can be ignored during reconstruction. In this way, we can think that the most information of the signal is contained in few of its wavelet coefficients. Hence, we can efficiently represent signals in $L^2(\mathbb{R})$ through sparse vectors.

As will be clear from the next sections, the idea is that wavelet coefficients are small when the support of $\psi_{j,n}$ does not intersect any singularity of the signal, while the relevant coefficients are those relative to the elements of the wavelet basis whose support intersects a singularity of the signal. The key features we require an orthonormal wavelet to have are *vanishing moments* and *compact support*, with as small a support as possible. The first property is useful to obtain small coefficients in the smooth regions of the signal, while the second to minimize, at fine scales, the number of elements of the basis with support containing a singularity of the signal.

In the next sections we discuss these properties and show some results to obtain wavelets with these features.

2.2.1 Vanishing Moments

Definition 2.11. A function $\psi \in L^2(\mathbb{R})$ has p *vanishing moments* if

$$\int_{\mathbb{R}} x^k \psi(x) dx = 0 \text{ for each } k = 0, \dots, p-1.$$

Having p vanishing moments means, for an orthonormal wavelet ψ, being orthogonal to polynomials of degree lower than p, i.e.

$$\langle P, \psi \rangle = 0 \text{ for each } P \in \mathbb{C}[x] \text{ such that } \deg(P) \leq p-1. \tag{2.18}$$

By a change of variable, we obtain that (2.18) holds for $\psi_{j,n}$ for each $j, n \in \mathbb{Z}$. Considering a piecewise regular function $f \in L^2(\mathbb{R})$, we have that in its smooth regions, it is well approximated by a Taylor polynomial of degree m, P_f^m. Hence, if we choose a wavelet with $p > m$ vanishing moments, we have that

$$\langle f, \psi_{j,n} \rangle \sim \langle P_f^m, \psi_{j,n} \rangle = 0.$$

This observation makes clear the fact that the number of vanishing moments is useful to get small coefficients in the smooth regions of a signal.

Example 2.12. Let us consider the Haar basis (2.5). The support of $\psi_{j,n}$ is increasingly localized around $x = \frac{n}{2^j}$. Let $f \in L^2(\mathbb{R})$ be sufficiently smooth on a neighborhood of the point $x = \frac{n}{2^j}$. Observe that $\int_{\mathbb{R}} \psi(z)dz = 0$, and $\int_{\mathbb{R}} z\psi(z)dz = -\frac{1}{4}$. Then one has

$$\begin{aligned}\langle f, \psi_{j,n}\rangle &= \int_{\mathbb{R}} f(x)\psi_{j,n}(x)dx = 2^{-\frac{j}{2}} \int_{\mathbb{R}} f\left(\frac{z}{2^j} + \frac{n}{2^j}\right)\psi(z)dz \\ &\sim 2^{-\frac{j}{2}} \int_{\mathbb{R}} \left(f\left(\frac{n}{2^j}\right) + \frac{z}{2^j} f'\left(\frac{n}{2^j}\right)\right)\psi(z)dz = -\frac{1}{4} 2^{-\frac{3}{2}j} f'\left(\frac{n}{2^j}\right).\end{aligned} \tag{2.19}$$

Therefore, the wavelet coefficients of the smooth regions of f decay as $O\left(2^{-\frac{3}{2}j}\right)$. On the other hand, if f is bounded, one always has

$$|\langle f, \psi_{j,n}\rangle| \leq \|f\|_\infty \|\psi_{j,n}\|_1 \leq 2^{-\frac{j}{2}} \|f\|_\infty. \tag{2.20}$$

This shows that the wavelet coefficients of the nonsmooth regions of f may have magnitude of the order $O\left(2^{-\frac{j}{2}}\right)$.

The next result gives equivalent conditions for a wavelet ψ to have p vanishing moments.

Proposition 2.13. *Let $\varphi \in L^2(\mathbb{R})$ be a scaling function of an MRA, and $\psi \in L^2(\mathbb{R})$ be the corresponding wavelet. Suppose that*

$$\begin{aligned}|\varphi(x)| &\lesssim (1 + |x|)^{-p-2}, \\ |\psi(x)| &\lesssim (1 + |x|)^{-p-2}.\end{aligned}$$

Then the following statements are equivalent

(a) ψ has p vanishing moments,

(b) $\frac{d^k \mathcal{F}\psi}{d\xi^k}(0) = 0$ for every $k = 0, \ldots, p-1$,

(c) $\frac{d^k m_0}{d\xi^k}(\pi) = 0$ for every $k = 0, \ldots, p-1$.

Proof. The assumptions of the theorem imply that $\mathcal{F}\psi$ and $\mathcal{F}\varphi$ are p times continuously differentiable. Hence, we have

$$\frac{d^k \mathcal{F}\psi}{d\xi^k}(\xi) = \int_{\mathbb{R}} (-ix)^k \psi(x) \, \mathrm{e}^{-ix\xi} \, dx.$$

Therefore, by evaluating in zero, we obtain the equivalence between (a) and (b).

In order to prove that (b) implies (c), consider (2.16), and apply the Fourier transform

$$\mathcal{F}\psi(\xi) = -\mathrm{e}^{-i\frac{\xi}{2}} \overline{m_0\left(\frac{\xi}{2} + \pi\right)} \mathcal{F}\varphi\left(\frac{\xi}{2}\right).$$

Since $\mathcal{F}\varphi(0) \neq 0$ due to Lemma 2.6, by setting $\xi = 0$, we have $m_0(\pi) = 0$. Now, by computing the derivative of the expression above,

$$\frac{d\mathcal{F}\psi}{d\xi}(\xi) = \frac{1}{2}\,\mathrm{e}^{-i\frac{\xi}{2}}\left[\overline{m_0\left(\frac{\xi}{2}+\pi\right)}\left(\mathcal{F}\varphi\left(\frac{\xi}{2}\right) - \frac{d\mathcal{F}\varphi}{d\xi}\left(\frac{\xi}{2}\right)\right) - \frac{d\overline{m_0}}{d\xi}\left(\frac{\xi}{2}+\pi\right)\mathcal{F}\varphi\left(\frac{\xi}{2}\right)\right],$$

and by setting $\xi = 0$, we obtain that $\frac{dm_0}{d\xi}(\pi) = 0$. Observe that when computing the k-th derivative of this expression, the only factor that multiplies $\frac{d^k\overline{m_0}}{d\xi^k}\left(\frac{\xi}{2}+\pi\right)$ is $\mathcal{F}\varphi\left(\frac{\xi}{2}\right)$, hence we can recursively apply Lemma 2.6 to conclude the proof. By using the same process it is easy to prove that (c) implies (b). □

An immediate consequence of this theorem is that an orthonormal wavelet that arises from an MRA satisfying the assumptions of the theorem has at least one vanishing moment. Indeed, by setting $\xi = 0$ in (2.13), and by Lemma 2.6, we have that $m_0(0) = 1$. Hence, applying Lemma 2.7, we obtain $m_0(\pi) = 0$.

2.2.2 Compact Support

While vanishing moments give compressible coefficients in the smooth regions of a signal, they are not sufficient on their own. In order to get a sparse representation, we also want that the number of relevant coefficients, i.e. those intersecting singularities, cannot grow too far along the scales. Hence, by choosing an orthonormal wavelet with a small support size, we have that at fine scales, i.e. for j growing, the support size of $\psi_{j,n}$ is getting smaller and smaller. Therefore, the majority of the dilated wavelets do not intersect any singularity of the signal.

The following result gives a characterization of compact support for a wavelet and scaling function.

Theorem 2.14 ([Adcock and Hansen, 2021])**.** *Let N_1, N_2 be integers such that $N_1 \leq N_2$.*

If the scaling function φ of an MRA has compact support in $[N_1, N_2]$, then the transfer function m_0 defined in (2.14) *is a trigonometric polynomial of the form*

$$m_0(\xi) = \frac{1}{\sqrt{2}}\sum_{n=N_1}^{N_2} h_n\,\mathrm{e}^{-in\xi}. \tag{2.21}$$

Conversely, if (2.21) *holds, and $m_0(0) = 1$, then the scaling function φ has support contained in $[N_1, N_2]$.*

In either case, the corresponding wavelet ψ given by (2.16) *has support contained in $[\frac{N_1-N_2+1}{2}, \frac{N_2-N_1+1}{2}]$.*

Given the previous proposition, our aim is to construct orthonormal wavelets that have compact support. We do this by constructing a transfer function m_0 with suitable properties.

Theorem 2.15. *Suppose that* $m_0(\xi) = \frac{1}{\sqrt{2}} \sum_{n=N_1}^{N_2} h_n \, \mathrm{e}^{-in\xi}$ *satisfies the following properties*

(a) $|m_0(\xi)|^2 + |m_0(\xi+\pi)|^2 = 1$, *a.e.* $\xi \in \mathbb{R}$;

(b) $m_0(0) = 1$;

(c) $m_0(\xi) \neq 0$ *for all* $|\xi| \leq \frac{\pi}{2}$.

Then the function φ *with Fourier transform*

$$\mathcal{F}\varphi(\xi) = \prod_{j=1}^{+\infty} m_0\left(\frac{\xi}{2^j}\right), \qquad \xi \in \mathbb{R}, \tag{2.22}$$

is well defined and in $L^2(\mathbb{R})$. *Moreover,* φ *has support in* $[N_1, N_2]$ *and is the scaling function of an MRA with transfer function* m_0. *The corresponding wavelet*

$$\psi = \sqrt{2} \sum_{n=N_1}^{N_2} (-1)^{n-1} \overline{h_n} \varphi(2 \cdot + n - 1) \tag{2.23}$$

has support contained in $[\frac{N_1 - N_2 + 1}{2}, \frac{N_2 - N_1 + 1}{2}]$.

Let us observe that conditions $(a) - (c)$ are not so unreasonable. Indeed, condition (a) is essential for ensuring that m_0 can be a transfer function of an MRA, as mentioned in Lemma 2.7. Moreover, we should remember that (b) is satisfied when the scaling function φ is such that $|\mathcal{F}\varphi|$ is continuous at 0, making it essentially necessary. Regarding (c), one can prove that if it does not occur, then $\{\varphi(\cdot - n) : n \in \mathbb{Z}\}$ need not to be an orthogonal system. Furthermore, by applying recursively (2.13) N times, we have that

$$\mathcal{F}\varphi(\xi) = \mathcal{F}\varphi\left(\frac{\xi}{2^N}\right) \prod_{j=1}^{N} m_0\left(\frac{\xi}{2^j}\right), \qquad \xi \in \mathbb{R}.$$

Therefore, it is natural to expect a construction of $\mathcal{F}\varphi$ as in (2.22).

2.2.3 Vanishing Moments Versus Support Size

The previous two sections explain the importance of vanishing moments and compact support to obtain sparse representation of signals via its wavelet coefficients. A priori, these properties seem to be independent, and one may wish to construct orthonormal wavelets with an arbitrary number of vanishing moments and a very small support. Unfortunately this is not possible, and the next theorem shows that there is a dependence between these two features.

Theorem 2.16. *Let ψ be an orthonormal wavelet with p vanishing moments. Then ψ has support size at least $2p-1$.*

Therefore, when analysing a signal and choosing the wavelet, we always face a trade-off between the number of vanishing moments and the support size of ψ. In particular, if the signal to analyze has a few number of isolated singularity, and it is regular between them, then it is more important to have a lot of vanishing moments than a small support size. Indeed, in this case the signal is mostly regular, so it is very important to have many vanishing moments to cancel out all the coefficients related to the smooth regions. While, even if the wavelet does not have a very small support, the isolated singularities will still be well localised. Conversely, if the signal has a lot of close singularity, then it is more important to have as small the support as possible, than a large number of vanishing moments.

A natural question that arises from the statement in Theorem 2.16 is: are there orthonormal wavelets with p vanishing moments satisfying the optimal bound on the support size? The answer to this question is positive, a famous example of such wavelets are the *Daubechies wavelets*. These wavelets were constructed by Daubechies in [Daubechies, 1992], the orthonormal wavelet with p vanishing moments (DBp) has support equal to $[0, 2p-1]$, and the corresponding scaling function has support equal to $[-p+1, p]$. Except for the case $p = 1$, which provides the Haar wavelet, we do not have an explicit expressions for the Daubechies wavelet. However, it is possible to use them in applications because their filter coefficients allow us to perform numerical computations.

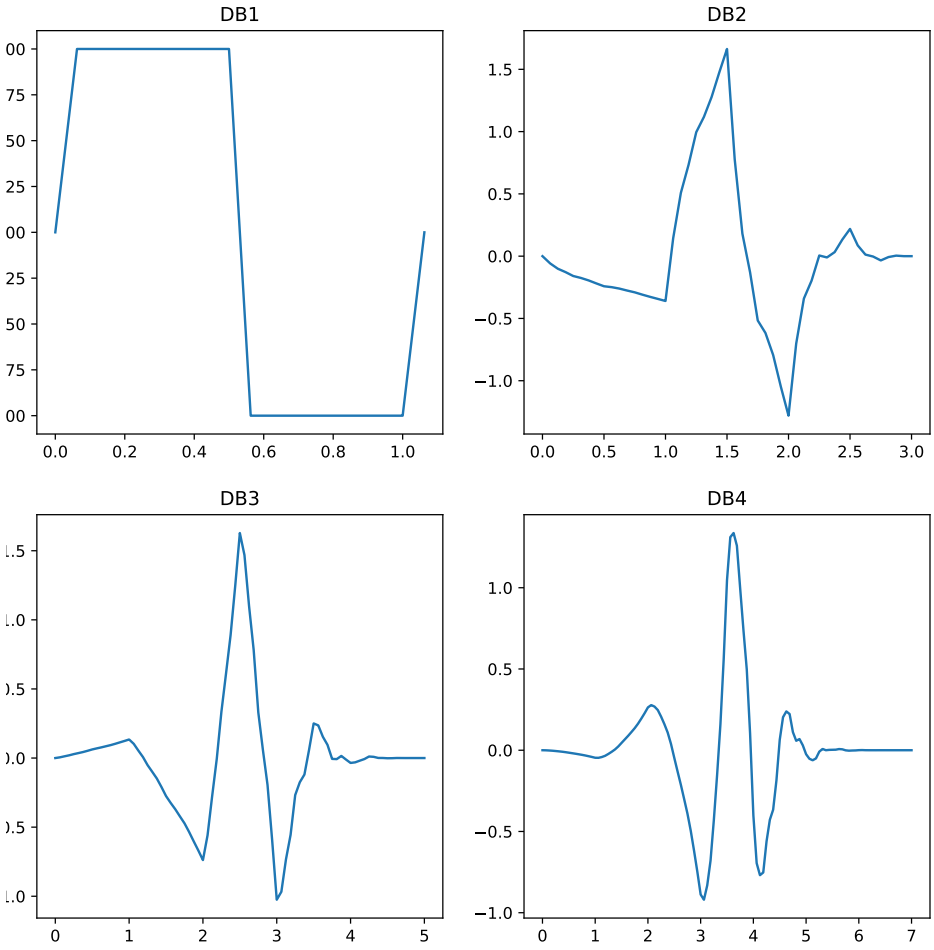


Figure 2.2: From top-left to bottom-right are plotted the Daubechies wavelets for $p = 1, 2, 3, 4$.

2.3 Wavelet basis of $L^2[0, 1]$

In many applications, the signals to study are compactly supported within an interval that, without loss of generality, could be supposed to be $[0, 1]$. Hence, the construction of wavelet bases of $L^2[0, 1]$ has a remarkable relevance. In this section, we briefly present the main ideas to obtain such bases from orthonormal wavelet bases of $L^2(\mathbb{R})$, and we refer to [Mallat, 1999, Adcock and Hansen, 2021] for further details. Let us start by analysing the Haar basis (2.5); in this case the process is straightforward. Indeed, the elements of the basis are either entirely supported within $[0, 1]$ or entirely supported outside of it. Consequently, the Haar

basis on the interval $[0,1]$ is simply obtained by keeping only those wavelets with support contained in $[0,1]$.

Unfortunately, this is a special case regarding only the Haar basis. In general, in a wavelet basis there are also elements whose supports intersect the boundary of $[0,1]$, but are not fully supported inside the interval. To construct an orthonormal basis of $L^2[0,1]$, we have to modify these boundary functions. We present two ways of doing this.

2.3.1 Periodic Wavelets

The most direct approach is periodizing an othonormal wavelet defined on the real line. Consider the periodizing operation

$$g \longmapsto g^{\mathrm{per}} := \sum_{n\in\mathbb{Z}} g(\cdot + n).$$

Observe that g^{per} is a 1-periodic function. By utilizing wavelets $\psi_{j,n}$, and scaling functions $\varphi_{j,n}$, we can construct $\varphi_{j,n}^{\mathrm{per}} = (\varphi^{\mathrm{per}})_{j,n}$, and $\psi_{j,n}^{\mathrm{per}} = (\psi^{\mathrm{per}})_{j,n}$, and define the spaces

$$V_j^{\mathrm{per}} := \mathrm{span}\{\varphi_{j,n}^{\mathrm{per}} : n = 0, \dots, 2^j - 1\},$$
$$W_j^{\mathrm{per}} := \mathrm{span}\{\psi_{j,n}^{\mathrm{per}} : n = 0, \dots, 2^j - 1\}.$$

Notice that in this case there is a finite number of elements since $\varphi_{j,n}^{\mathrm{per}} = \varphi_{j,n+2^j}^{\mathrm{per}}$, and likewise for $\psi_{j,n}^{\mathrm{per}}$. These spaces satisfy the same property of the original MRA, and so provide the analogous decomposition of $L^2[0,1]$ as in 2.9

$$L^2[0,1] = V_{j_0}^{\mathrm{per}} \oplus W_{j_0}^{\mathrm{per}} \oplus W_{j_0+1}^{\mathrm{per}} \oplus \dots, \tag{2.24}$$

for any j_0.

In particular, we have that

$$\{\varphi_{j_0,n}^{\mathrm{per}} : n = 0, \dots, 2^{j_0} - 1\} \cup \{\psi_{j,n}^{\mathrm{per}} : j \geq j_0,\ n = 0, \dots, 2^j - 1\} \tag{2.25}$$

is an orthonormal basis of $L^2[0,1]$ for each j_0.

When analyzing a signal $f\colon [0,1] \longrightarrow \mathbb{C}$ with a periodizated wavelet basis of $L^2[0,1]$, we are treating f as a 1-periodic signal on the real line, this procedure can generate discontinuities at $x = 1$, and so also in every $x \in \mathbb{N}$. Therefore, by periodizing, we add artificial discontinuities to the signal, which affects its compressibility. This behaviour is undesirable, so we now present another way of constructing wavelets on intervals, that avoid to create new discontinuities in the signal.

2.3.2 Boundary-Corrected Wavelets

This type of wavelets are introduced in order to overcome the problems arising from the periodization method.

Their construction is based on modifying those wavelets whose supports intersect the boundaries of the interval $[0,1]$. For example, beginning with the DBp wavelet basis, it is possible to define functions φ_n^{left}, φ_n^{right}, ψ_n^{left}, ψ_n^{right} that modify the wavelets and scaling functions at the left and right endpoints. [Mallat, 1999], and [Cohen et al., 1993] provide further details about the construction. Then, we consider new wavelets and scaling functions as

$$\psi_{j,n}^{\text{int}}(x) = \begin{cases} 2^{\frac{j}{2}}\psi_n^{\text{left}}(2^j x), & \text{if } 0 \le n < p \\ 2^{\frac{j}{2}}\psi(2^j x - n), & \text{if } p \le n < 2^j - p \\ 2^{\frac{j}{2}}\psi_n^{\text{right}}(2^j x), & \text{if } 2^j - p \le n < 2^j \end{cases}$$

and

$$\varphi_{j,n}^{\text{int}}(x) = \begin{cases} 2^{\frac{j}{2}}\varphi_n^{\text{left}}(2^j x), & \text{if } 0 \le n < p \\ 2^{\frac{j}{2}}\varphi(2^j x - n), & \text{if } p \le n < 2^j - p \\ 2^{\frac{j}{2}}\varphi_n^{\text{right}}(2^j x), & \text{if } 2^j - p \le n < 2^j \end{cases}$$

We omit the details of this construction. Similarly to the periodized case, the functions $\psi_{j,n}^{\text{int}}$, and $\varphi_{j,n}^{\text{int}}$ provide orthonormal bases for

$$V_j^{\text{int}} := \text{span}\{\varphi_{j,n}^{\text{int}} : n = 0, \dots, 2^j - 1\},$$
$$W_j^{\text{int}} := \text{span}\{\psi_{j,n}^{\text{int}} : n = 0, \dots, 2^j - 1\}.$$

Moreover, we always get the same properties, and the same space decomposition as in (2.9), applied to V_j^{int}, W_j^{int}, and $L^2[0,1]$

$$L^2[0,1] = V_{j_0}^{\text{int}} \oplus W_{j_0}^{\text{int}} \oplus W_{j_0+1}^{\text{int}} \oplus \dots. \tag{2.26}$$

for any j_0. Therefore, the corresponding wavelet basis is

$$\{\varphi_{j_0,n}^{\text{int}} : n = 0, \dots, 2^{j_0} - 1\} \cup \{\psi_{j,n}^{\text{int}} : j \ge j_0,\ n = 0, \dots, 2^j - 1\}. \tag{2.27}$$

Unlike periodizated wavelets, the boundary-corrected wavelets preserve vanishing moments. Indeed, the restriction of polynomials of degree less than p to the interval $[0,1]$ belongs to V_j^{int}, hence it is orthogonal to the detail space W_j^{int}. Therefore, this construction allows us to have fast decay of the coefficients relating to smooth regions, even if the support of the wavelet intersects $[0,1]$.

2.4 Multidimensional Wavelets

Having constructed wavelets bases for the analysis of one-dimensional signal, we now want to extend this construction to higher dimensions. We present the two-dimensional case, which will be the one of our interest in the next chapters. The constructions described below are simply generalisable to higher dimensions.

2.4.1 Wavelet Bases of $L^2\left(\mathbb{R}^2\right)$

Let us consider a wavelet basis $\{\psi_{j,n} : j,n \in \mathbb{Z}\}$ of $L^2(\mathbb{R})$ arising from an MRA $\{V_j : j \in \mathbb{Z}\}$ with scaling function φ. A possible approach to obtain an orthonormal basis of $L^2(\mathbb{R}^2)$ is by considering the tensor product of the spaces V_j

$$V_j^{(2)} := V_j \otimes V_j = \overline{\text{span}\{g \otimes h : g,h \in V_j\}} \subseteq L^2\left(\mathbb{R}^2\right), \tag{2.28}$$

where $(g \otimes h)(x_1,x_2) = g(x_1)h(x_2)$.

One can verify that the sequence $\{V_j^{(2)}\}_{j\in\mathbb{Z}}$ defines a two-dimensional MRA, i.e. a sequence of closed subspaces of $L^2(\mathbb{R}^2)$ verifying

(a) $V_j^{(2)} \subseteq V_{j+1}^{(2)}$ for all $j \in \mathbb{Z}$,

(b) $f \in V_j^{(2)}$ if and only if $f(2\cdot) \in V_{j+1}^{(2)}$,

(c) $\bigcap_{j\in\mathbb{Z}} V_j^{(2)} = \{0\}$,

(d) $\overline{\bigcup_{j\in\mathbb{Z}} V_j^{(2)}} = L^2(\mathbb{R}^2)$,

(e) the set $\{\varphi_{0,n} := \varphi_{0,n_1} \otimes \varphi_{0,n_2} : n = (n_1,n_2) \in \mathbb{Z}^2\}$ forms an orthonormal basis of $V_0^{(2)}$.

Similar to the one-dimensional case, we define the detail spaces $W_j^{(2)}$ as the orthogonal complement of $V_j^{(2)}$ within $V_{j+1}^{(2)}$. Hence, we get the following decompositions

$$L^2\left(\mathbb{R}^2\right) = V_{j_0}^{(2)} \oplus W_{j_0}^{(2)} \oplus W_{j_0+1}^{(2)} \oplus \dots \tag{2.29}$$

$$L^2\left(\mathbb{R}^2\right) = \bigoplus_{j\in\mathbb{Z}} W_j^{(2)}, \tag{2.30}$$

for any scale $j_0 \in \mathbb{Z}$.

Now we can construct wavelet bases for $L^2(\mathbb{R}^2)$. First observe that, due to properties (b) and (e),

$$\{\varphi_{j,n} = \varphi_{j,n_1} \otimes \varphi_{j,n_2} : n = (n_1,n_2) \in \mathbb{Z}^2\} \tag{2.31}$$

is an orthonormal basis of $V_j^{(2)}$. Now, one has

$$\begin{aligned} V_j^{(2)} \oplus W_j^{(2)} &= V_{j+1}^{(2)} = V_{j+1} \otimes V_{j+1} = (V_j \oplus W_j) \otimes (V_j \oplus W_j) \\ &= (V_j \otimes V_j) \oplus (V_j \otimes W_j) \oplus (W_j \otimes V_j) \oplus (W_j \otimes W_j) \\ &= V_j^{(2)} \oplus (V_j \otimes W_j) \oplus (W_j \otimes V_j) \oplus (W_j \otimes W_j). \end{aligned}$$

Therefore,

$$W_j^{(2)} = (V_j \otimes W_j) \oplus (W_j \otimes V_j) \oplus (W_j \otimes W_j),$$

so the functions $\varphi_{j,n_1} \otimes \psi_{j,n_2}$, $\psi_{j,n_1} \otimes \varphi_{j,n_2}$, and $\psi_{j,n_1} \otimes \psi_{j,n_2}$ provide a basis for these spaces. It is now useful to introduce some notation, let

$$\Psi_{j,n}^{(0)} = \varphi_{j,n} \ , \ \Psi_{j,n}^{(1)} = \psi_{j,n} \tag{2.32}$$

denote the one-dimensional wavelet and scaling functions. Then we set

$$\Psi_{j,n}^{(e)} = \Psi_{j,n_1}^{(e_1)} \otimes \Psi_{j,n_2}^{(e_2)}, \ e = (e_1, e_2) \in \{0,1\}^2, \ n = (n_1, n_2) \in \mathbb{Z}^2, \ j \in \mathbb{Z}.$$

Hence, orthonormal bases of $V_j^{(2)}$, and $W_j^{(2)}$ are respectively given by

$$\{\Psi_{j,n}^{(0,0)} : n \in \mathbb{Z}^2\},$$

and

$$\{\Psi_{j,n}^{(e)} : e \in \{0,1\}^2 \backslash \{(0,0)\}, \ n \in \mathbb{Z}^2\}.$$

Therefore, using (2.29), and (2.30), we obtain for any $j_0 \in \mathbb{Z}$

$$\{\Psi_{j_0,n}^{(0,0)} : n \in \mathbb{Z}^2\} \cup \{\Psi_{j,n}^{(e)} : j \geq j_0, \ e \in \{0,1\}^2 \backslash \{(0,0)\}, \ n \in \mathbb{Z}^2\}, \tag{2.33}$$

and

$$\{\Psi_{j,n}^{(e)} : e \in \{0,1\}^2 \backslash \{(0,0)\}, \ n \in \mathbb{Z}^2, \ j \in \mathbb{Z}\} \tag{2.34}$$

are orthonormal bases of $L^2(\mathbb{R}^2)$.

2.4.2 Wavelet Bases of $L^2([0,1]^2)$

In order to construct wavelet bases of $L^2([0,1]^2)$, we can define the spaces $V_j^{\text{type},(2)}$, $W_j^{\text{type},(2)}$ by following the construction in the previous section, where type denotes either the periodic (per), or the boundary-corrected (int) wavelets. Then, we get the decomposition of $L^2([0,1]^2)$

$$L^2([0,1]^2) = V_{j_0}^{\text{type},(2)} \oplus W_{j_0}^{\text{type},(2)} \oplus W_{j_0+1}^{\text{type},(2)} \oplus \dots \tag{2.35}$$

for any j_0. Now let

$$\Psi_{j,n}^{\text{type},(e)} = \Psi_{j,n_1}^{\text{type},(e_1)} \otimes \Psi_{j,n_2}^{\text{type},(e_2)}, \ e = (e_1, e_2) \in \{0,1\}^2, \ j \in \mathbb{Z}, \ n_1, n_2 = 0, \dots, 2^j - 1.$$

Similarly to the $\mathbb{R}^2$-case, we have that

$$\{\Psi_{j,n}^{\text{type},(0,0)} : n = 0, \dots, 2^j - 1\},$$

and

$$\{\Psi_{j,n}^{\text{type},(e)} : e \in \{0,1\}^2 \backslash \{(0,0)\}, \ n = 0, \dots, 2^j - 1\}$$

form bases of $V_j^{\text{type},(2)}$ and $W_j^{\text{type},(2)}$ respectively.

Then, due to (2.35), we have that

$$\begin{aligned} &\{\Psi_{j_0,n}^{\text{type},(0,0)} : n = 0, \dots, 2^j - 1\} \cup \\ &\{\Psi_{j,n}^{\text{type},(e)} : j \geq j_0, \ e \in \{0,1\}^2 \backslash \{(0,0)\}, \ n = 0, \dots, 2^j - 1\} \end{aligned} \tag{2.36}$$

is an orthonormal basis of $L^2([0,1]^2)$.

Chapter 3

Non-Linear Approximation

In signal processing, orthonormal bases are of interest because they can be used to efficiently approximate signals with just a few of their vectors. Approximation theory studies the error produced by different schemes of approximation.

A *linear approximation* projects the signal over M vectors of the basis chosen a priori. For instance, in the wavelets case, a linear approximation can be carried out by fixing a scale parameter j, and projecting the signal over the related space V_j. In general, given $\mathcal{H}$ an Hilbert space, $\mathcal{B} = \{g_m\}_{m\in\mathbb{N}}$ an orthonormal basis of $\mathcal{H}$, and a signal $f \in \mathcal{H}$, we can make a linear approximation by projecting f over the space generated by $\{g_m\}_{m=1}^{M}$, namely

$$f_M = \sum_{m=1}^{M} \langle f, g_m \rangle g_m.$$

In doing so, we bring in a linear approximation error

$$\epsilon_l[M] = \|f - f_M\|_{\mathcal{H}}^2 = \sum_{k=M+1}^{+\infty} |\langle f, g_m \rangle|^2. \tag{3.1}$$

Another approximation scheme is to choose the M vectors depending on the signal f to analyze. This means that we have to choose a set $I_M \subseteq \mathbb{N}$ of cardinality M, and projecting f over the space generated by $\{g_m\}_{m\in I_M}$

$$f_M = \sum_{m\in I_M} \langle f, g_m \rangle g_m.$$

With this procedure, we are projecting f over the space of the M-sparse vectors, i.e. those elements $f \in \mathcal{H}$ such that

$$f = \sum_{m\in I} \langle f, g_m \rangle g_m,$$

for some $I \subseteq \mathbb{N}$, $|I| \leq M$. Since this space is not a linear subspace of $\mathcal{H}$, we speak of *non-linear approximation*. In this case, the error is

$$\epsilon[M] = \sum_{m \notin I_M} |\langle f, g_m \rangle|^2.$$

To minimize the error, we have to choose I_M corresponding to the M vectors having the largest inner product amplitude $|\langle f, g_m \rangle|$. These are the vectors that contain the main features of f. The resulting error is necessarily smaller than the linear approximation error (3.1).

To simplify the notation, let us sort the sequence $\{|\langle f, g_m \rangle|\}_{m \in \mathbb{N}}$ in decreasing order. We denote with $f_{\mathcal{B}}[k] = \langle f, g_{m_k} \rangle$ the coefficient of rank k:

$$|f_{\mathcal{B}}[k]| \geq |f_{\mathcal{B}}[k+1]| \text{ for all } k \in \mathbb{N}.$$

Hence, the best non-linear approximation is

$$f_M = \sum_{k=1}^{M} f_{\mathcal{B}}[k] g_{m_k}, \tag{3.2}$$

and the corresponding non-linear approximation error is

$$\epsilon_n[M] = \|f - f_M\|_{\mathcal{H}}^2 = \sum_{k=M+1}^{+\infty} |f_{\mathcal{B}}[k]|^2. \tag{3.3}$$

The next theorem shows a characterization of the decay of the non-linear approximation error through the decay of the sorted coefficients.

Theorem 3.1. *Let $s > \frac{1}{2}$. If there exists $C > 0$ such that*

$$|f_{\mathcal{B}}[k]| \leq C k^{-s},$$

then

$$\epsilon_n[M] \leq \frac{C^2}{2s-1} M^{1-2s}. \tag{3.4}$$

Conversely, if $\epsilon_n[M]$ satisfies 3.4, then

$$|f_{\mathcal{B}}[k]| \leq C \left(1 - \frac{1}{2s}\right)^{-s} k^{-s}. \tag{3.5}$$

Proof. Let us suppose that $|f_{\mathcal{B}}[k]| \leq C k^{-s}$, then, from 3.3, we have

$$\epsilon_n[M] \leq C^2 \sum_{k>M} k^{-2s} \leq C^2 \int_M^{+\infty} t^{-2s} dt = \frac{C^2}{2s-1} M^{1-2s}.$$

Let us now observe that, for $\alpha < 1$,

$$\epsilon_n[\lceil \alpha M \rceil] = \sum_{k=\lceil \alpha M \rceil + 1}^{+\infty} |f_{\mathcal{B}}[k]|^2 \geq \sum_{k=\lfloor \alpha M \rfloor + 1}^{M+1} |f_{\mathcal{B}}[k]|^2 \geq |f_{\mathcal{B}}[M+1]|^2 M(1-\alpha).$$

Therefore,

$$|f_{\mathcal{B}}[M+1]|^2 \leq \frac{\epsilon_n[\lceil \alpha M \rceil]}{M(1-\alpha)}.$$

Now, by using 3.4, we have

$$|f_{\mathcal{B}}[M+1]|^2 \leq \frac{C^2}{2s-1} \frac{\alpha^{1-2s}}{1-\alpha} M^{-2s}.$$

Choosing $\alpha = 1 - \frac{1}{2s} < 1$, and $k = M+1$, we obtain

$$|f_{\mathcal{B}}[k]| \leq C \left(1 - \frac{1}{2s}\right)^{-s} (k-1)^{-s} = C \left(1 - \frac{1}{2s}\right)^{-s} \left(\frac{k}{k-1}\right)^{s} k^{-s}.$$

The fact that $\frac{k}{k-1} \leq 2$ for $k \geq 2$ allows us to obtain 3.5. □

The decay of the sorted inner products can be evaluated from the ℓ^p norm

$$\|f\|_{\mathcal{B},p}^p := \sum_{m \in \mathbb{N}} |\langle f, g_m \rangle|^p.$$

The following theorem relates the decay of the non-linear error approximation to the ℓ^p norm defined above.

Theorem 3.2. *Let $p < 2$. If $\|f\|_{\mathcal{B},p} < +\infty$, then*

$$|f_{\mathcal{B}}[k]| \leq \|f\|_{\mathcal{B},p} k^{-\frac{1}{p}}, \tag{3.6}$$

and $\epsilon_n[M] = o\left(M^{1-\frac{2}{p}}\right)$.

Proof. We prove 3.6 by observing that

$$\|f\|_{\mathcal{B},p}^p = \sum_{n \in \mathbb{N}} |f_{\mathcal{B}}[n]|^p \geq \sum_{n=1}^{k} |f_{\mathcal{B}}[n]|^p \geq k |f_{\mathcal{B}}[k]|^p.$$

Now, by setting

$$S[k] := \sum_{n=k}^{2k-1} |f_{\mathcal{B}}[n]|^p \geq k |f_{\mathcal{B}}[2k]|^p,$$

we obtain

$$|f_{\mathcal{B}}[k]| \leq S \left[\frac{k}{2}\right]^{\frac{1}{p}} \left(\frac{k}{2}\right)^{-\frac{1}{p}}.$$

Hence,

$$\begin{aligned}\epsilon_n[M] &= \sum_{k=M+1}^{+\infty} |f_{\mathcal{B}}[k]|^2 \leq \sum_{k=M+1}^{+\infty} S\left[\frac{k}{2}\right]^{\frac{2}{p}} \left(\frac{k}{2}\right)^{-\frac{2}{p}} \\ &\leq \left(\sup_{k>\frac{M}{2}} S[k]^{\frac{2}{p}}\right) \sum_{k=M+1}^{+\infty} \left(\frac{k}{2}\right)^{-\frac{2}{p}} \leq \left(\sup_{k>\frac{M}{2}} S[k]^{\frac{2}{p}}\right) \int_{M}^{+\infty} \left(\frac{t}{2}\right)^{-\frac{2}{p}} dt \\ &\leq C \left(\sup_{k>\frac{M}{2}} S[k]^{\frac{2}{p}}\right) M^{1-\frac{2}{p}}.\end{aligned}$$

Since $\|f\|^p_{\mathcal{B},p} < +\infty$, we have

$$\lim_{M\to+\infty} \sup_{k>\frac{M}{2}} S[k]^{\frac{2}{p}} = 0,$$

and so $\epsilon_n[M] = o\left(M^{1-\frac{2}{p}}\right)$. □

This theorem provides spaces of functions whose elements are well approximated by a few vectors of a basis $\mathcal{B}$

$$\mathrm{B}_{\mathcal{B},p} := \{f \in \mathcal{H} : \|f\|_{\mathcal{B},p} < +\infty\}.$$

If $f \in \mathrm{B}_{\mathcal{B},p}$, then the previous theorem shows that $\epsilon_n[M] = o\left(M^{1-\frac{2}{p}}\right)$. Conversely, if $\epsilon_n[M] = O\left(M^{1-\frac{2}{p}}\right)$, then the inequality (3.5) for $s = \frac{1}{p}$ shows that $f \in \mathrm{B}_{\mathcal{B},q}$ for any $q > p$.

3.1 Non-Linear Wavelets Approximation

We are now interested in applying these results to the case of wavelet bases. We consider the Hilbert space $L^2[0,1]$, and an orthonormal wavelet basis, such as (2.25), or (2.27), with compactly supported wavelets that are C^q, with q vanishing moments

$$\mathcal{B} = \{\varphi_{J,n} : n = 0, \ldots, 2^J - 1\} \cup \{\psi_{j,n} : j \geq J,\ n = 0, \ldots, 2^j - 1\}.$$

Therefore, for any $f \in L^2[0,1]$, we have

$$f = \sum_{n=0}^{2^J-1} \langle f, \varphi_{J,n}\rangle \varphi_{J,n} + \sum_{j=J}^{+\infty} \sum_{n=0}^{2^j-1} \langle f, \psi_{j,n}\rangle \psi_{j,n}.$$

To simplify the notation, let $\psi_{J-1,n} = \varphi_{J,n}$. Hence, the best non-linear approximation of f using M elements of $\mathcal{B}$ is

$$f_M = \sum_{(j,n)\in I_M} \langle f, \psi_{j,n}\rangle \psi_{j,n},$$

where I_M is a set of cardinality M that contains the indices corresponding to the M wavelet coefficients having the largest amplitude. The approximation error is

$$\epsilon_n[M] = \sum_{(j,n)\notin I_M} |\langle f, \psi_{j,n}\rangle|^2.$$

A class of functions that are well approximated by a non-linear wavelets approximation, i.e. such that their non-linear approximation error $\epsilon_n[M]$ has fast decay as M increases, is the class of *piecewise regular functions*. In fact, there are a few wavelet coefficients affected by isolated discontinuities, and the error decay rate depends on the *uniform Lipschitz regularity* (see Appendix A) of the function in its smooth areas. The next theorem formalizes such behaviour. Observe that, following the proofs presented in [Mallat, 1999], since we cannot control the scaling coefficients magnitudes, we assume to select them in the non-linear approximation.

Theorem 3.3. *If $f \in L^2[0,1]$ is a piecewise regular function, with a finite number of discontinuities, and is uniformly Lipschitz $\alpha < q$ between these discontinuities, then*

$$\epsilon_n[M] = O\left(M^{-2\alpha}\right).$$

Proof. In the following, we present the main ideas of the proof. We will prove that $f_{\mathcal{B}}[k] = O(k^{-\alpha-\frac{1}{2}})$, this allows for the application of the first part of Theorem 3.1, and this concludes the proof.

Let us split the wavelet coefficients into two families; we consider type one coefficients, related to those wavelets whose supports contain at least one singularity of the signal, and type two coefficients, related to those wavelets whose supports are contained in regions where the signal f is uniformly Lipschitz α. Consider now the associated decreasing sequences $f_{\mathcal{B},1}[k]$, and $f_{\mathcal{B},2}[k]$. We demonstrate $f_{\mathcal{B}}[k] = O(k^{-\alpha-\frac{1}{2}})$, by proving $f_{\mathcal{B},1}[k] = O(k^{-\alpha-\frac{1}{2}})$, and $f_{\mathcal{B},2}[k] = O(k^{-\alpha-\frac{1}{2}})$ separately. The principal result we need in order to conclude the proof is Theorem 6.3 in [Mallat, 1999] applied to the sample of the continuous wavelet transform $\mathcal{W}_\psi f(2^{-j}n, 2^{-j})$. This theorem states that, if f is uniformly Lipschitz α on the support of $\psi_{j,n}$, then there exists a constant $A > 0$ such that

$$|\langle f, \psi_{j,n}\rangle| \leq A 2^{-j(\alpha+\frac{1}{2})}. \tag{3.7}$$

Let us now consider type two coefficients. Fix an integer $l < 0$, and consider the scale parameter $-l > 0$. At coarser scales $2^{-j} > 2^l$ there are at most 2^{-l} type two coefficients. At finer scales $2^{-j} \leq 2^l$, due to (3.7), we have

$$|\langle f, \psi_{j,n}\rangle| \leq A 2^{l(\alpha+\frac{1}{2})}.$$

Hence, since the coefficients decrease for the scale parameter growing, we have that, for $k \sim 2^{-l}$,

$$f_{\mathcal{B},2}[k] \leq A2^{l(\alpha+\frac{1}{2})},$$

and this concludes the proof in this case. Let us now consider type one coefficients. Since f is uniformly Lipschitz α between the discontinuities, in particular f has to be uniformly bounded on $[0,1]$, and so it is uniformly Lipschitz 0 over $[0,1]$. Hence, (3.7) implies that

$$|\langle f, \psi_{j,n}\rangle| \leq A2^{-\frac{j}{2}}.$$

Let us suppose that the orthogonal wavelet ψ is supported within $[0, L]$ for some $L > 0$, and suppose that the function f has D discontinuities. At each scale 2^{-j}, the wavelets are supported within $[\frac{k}{2^j}, \frac{L+k}{2^j}]$, so every time k increases by 1, we shift the support to the right by $\frac{1}{2^j}$. Since the length of the support is constant $\frac{L}{2^j}$, we have that, for a fixed abscissa $v \in [0,1]$, there are at most L wavelets whose supports contain it. Since there are D discontinuities, then there exist at most LD wavelets whose supports contain at least one singularity. Fix an integer $l < 0$, and consider the scale parameter $-l > 0$. Since the coefficients decrease for the scale parameter growing, in the sequence $f_{\mathcal{B},1}[k]$ we find the coefficients at scale 2^l for $k \sim -lLD$. Moreover, at finer scales $2^{-j} \leq 2^l$ every type one coefficient is lower than $A2^{\frac{l}{2}}$. Therefore, we can conclude that

$$f_{\mathcal{B},1}[-lLD] \leq A2^{\frac{l}{2}}.$$

This implies that

$$f_{\mathcal{B},1}[k] = O(k^{-\beta-\frac{1}{2}}),$$

for every $\beta > 0$, and this concludes the proof. □

This theorem shows that a finite number of discontinuities does not influence the decay rate of the non-linear error, which only depends on the regularity between the singularities; the higher the regularity, the faster the error decays.

Let us now consider a piecewise regular image, i.e. a piecewise regular function that belongs to $L^2([0,1]^2)$, such as a *cartoon-like image*, which will be introducted formally in Chapter 4.

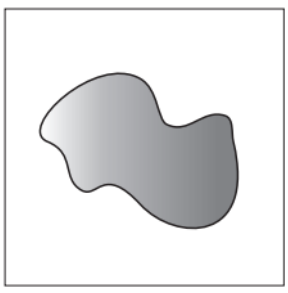


Figure 3.1: Cartoon-Like Image.

It has discontinuities along curves of dimension one, which create a non-negligible number of high magnitude wavelet coefficients. The next result shows that, if we consider the prototype of a piecewise regular image, than the non-linear error cannot decay faster than M^{-1}.

Proposition 3.4. *If $f = \mathbb{1}_\Omega$ is the indicator function of a set Ω whose boundary has a finite length, then*

$$|f_{\mathcal{B}}[k]| \sim \|f\|_V \, k^{-1}, \tag{3.8}$$

and hence

$$\epsilon_n[M] \sim \|f\|_V^2 \, M^{-1}, \tag{3.9}$$

where $\|f\|_V := \int_{[0,1]^2} |\nabla f(x)| dx$ is the total variation norm.

Proof. We give the main idea of the proof without all the details, and we refer to [Mallat, 1999] for more details. When analyzing the function f with a two-dimensional wavelet at a fixed scale parameter j, we are portioning the unitary square into dyadic squares of side length 2^{-j}. Since the boundary of Ω, $\partial\Omega$, has finite length L, one can prove (see Appendix C) that, for each scale parameter j, there are on the order of $L2^j$ wavelets whose supports intersect $\partial\Omega$. Now, if the wavelet does not intersect $\partial\Omega$, since f is constant there, we can conclude that the corresponding coefficient is 0. If the wavelet support intersects the boundary of Ω, we can prove that the corresponding wavelet coefficient has magnitude of the order of 2^{-j}. Since the amplitudes of these coefficients decrease as the scale 2^{-j} decreases, i.e. for $j \to +\infty$, and since there are on the order of $L2^j$ significant wavelet coefficients at coarser scales, then we meet the coefficients at scale 2^{-j} for $k \sim L2^j$. This means that $|f_{\mathcal{B}}[k]| \sim 2^{-j}$ for $k \sim L2^j$, which is

$$|f_{\mathcal{B}}[k]| \sim Lk^{-1}.$$

To conclude the proof of (3.8), one can prove $\|f\|_V = L$. Finally, we can conclude the proof of the theorem by substituting (3.8) into the definition of non-linear approximation error (3.3) to obtain (3.9). □

In particular, this theorem shows that despite the strong regularity of the function in its regular areas (in this case it is even C^∞), the non-linear error cannot decay arbitrarily fast as in the one-dimensional case. Therefore, while wavelets bases are optimal for non-linear approximation of one-dimensional signal, in higher dimensions they do not provide the same error decay results.

The reason why this happens lies in the geometry of the set of the signal singularities. Whereas in one dimension it is a finite collection of points, in dimension two it forms a curve, so the geometry of this curve also comes into play. The fact is that the support of a two-dimensional wavelet, (construction in Section 2.4.2) arising from a compactly supported one-dimensional wavelet, is a square that can only be translated or enlarged, but cannot change shape adaptively to the curvature of the curve.

Chapter 4

Shearlets

In Proposition 3.4, it is shown that wavelet bases are not optimal for non-linear approximation of multivariate functions. The underlying explanation for this is that wavelets possess *isotropic* features. In fact, wavelet bases are constructed by translating, and isotropically dilating a generating function. This makes wavelets unable to effectively dealing with multidimensional piecewise regular function such as the so-called *cartoon-like images* (see Figure 3.1).

Definition 4.1. The class $\mathcal{E}(\mathbb{R}^2)$ of cartoon-like images is the set of functions $f\colon \mathbb{R}^2 \longrightarrow \mathbb{C}$ of the form

$$f = f_0 + f_1 \mathbb{1}_B,$$

where $f_i \in C^2(\mathbb{R}^2)$ is supported in $[0,1]^2$, with $\|f_i\|_{C^2} := \sum_{|\alpha|\leq 2} \|\mathrm{D}^\alpha f_i\|_\infty \leq 1$ for $i = 0, 1$, and $B \subset [0,1]^2$ is a set with C^2 boundary ∂B.

In this case, the isotropic features of wavelet bases do not allow then to capture most of the signal information with a few coefficients. These limitations have motivated the study of new techniques for non-linear approximation of multivariate functions in order to outperform the error rate relative to multidimensional wavelets. The best result to date was proved in [Donoho, 2001]; it provides an error rate of the order of N^{-2}, but it has evident practical limitations since it requires to construct adapted triangulations of the unitary square. However, it provides a benchmark for optimally sparse approximation of cartoon-like images. Moreover, the argument in the proof indicates that to obtain the most efficient sparse representation, it is necessary to use analyzing elements with elongated and orientable supports. Throughout these years, starting from the concept of wavelets and the need to incorporate directional sensitivity, several attempts were carried out in order to achieve Donoho's error rate. In 2005, Kutyniok, Labate, Lim and Weiss introduced shearlets in [Labate et al., 2005]. Shearlet systems currently represent the finest generalization of one-dimensional wavelets since they offer a distinctive combination of the following desired attributes:

- A finite set of generating functions;

- Anisotropic dilations;
- Directional sensitivity which preserve the discrete lattice;
- Asymptotic optimal error rate.

In particular, the first property allows easy handling of shearlet systems in applications. The second and third properties enable to adaptively modify the supports of shearlets to the singularity curve. The fact that the discrete lattice is preserved is useful in applications when transitioning from the continuum to the discrete setting. Finally, as it will be shown in Theorem 4.15, shearlet systems reach, up to a logarithmic factor, the optimal asymptotic error decay of N^{-2}.

4.1 Continuous Shearlet Systems

Before providing the technical construction of shearlet systems, let us discuss the intuitive ideas that are at the core of the construction. We refer the interested reader to [Kutyniok and Labate, 2012] and [Labate et al., 2005]. As explained in the introduction of the chapter, we want to construct systems whose elements must be functions spanning over different locations, scales, and orientations. This necessitates the compositions of three different operators that are capable to translate, dilate, and orientate the supports of the generating functions. In order to translate, we use the standard translation operator T_t defined, for any $t \in \mathbb{R}^2$, as

$$T_t\psi := \psi(\cdot - t).$$

For the dilation, since we need to anisotropically dilate the supports, we use the family of *parabolic scaling matrices*

$$A_a := \begin{pmatrix} a & 0 \\ 0 & a^{\frac{1}{2}} \end{pmatrix}, \qquad a > 0,$$

where the corresponding operator is defined by

$$D_{A_a}\psi(x_1, x_2) := |\det A_a|^{-\frac{1}{2}}\psi\left(A_a^{-1}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) = a^{-\frac{3}{4}}\psi(a^{-1}x_1, a^{-\frac{1}{2}}x_2).$$

Now, we need an orthogonal operator to orientate the support of the generating functions. The most straightforward choice is to use rotations, such as in the curvelet construction (see [Candès and Donoho, 2004]), but unless we rotate by $\frac{\pi}{2}$, π, $\frac{3}{2}\pi$, or 2π, these operations destroy the integer lattice structure of $\mathbb{Z}^2$. This issue becomes a problem when transitioning to the discrete setting in applications. The alternative approach proposed for shearlet systems involves using *shearing matrices*,

$$S_s := \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}, \qquad s \in \mathbb{R}.$$

The shearing matrices have the advantage of preserving the integer lattice structure provided s is an integer. Likewise the parabolic scaling operators, the shearing operators are defined by

$$D_{S_s}\psi(x_1,x_2) := |\det S_s|^{-\frac{1}{2}}\psi\left(S_s^{-1}\begin{pmatrix}x_1\\x_2\end{pmatrix}\right) = \psi(x_1 - sx_2, x_2).$$

Combining these three operators, we can define a *continuous shearlet system* generated by a function $\psi \in L^2(\mathbb{R}^2)$. For any $a > 0$, $s \in \mathbb{R}$, $t = (t_1, t_2) \in \mathbb{R}^2$ we define

$$\psi_{a,s,t}(x_1,x_2) = T_t D_{A_a} D_{S_s}\psi(x_1,x_2) = a^{-\frac{3}{4}}\psi\left(A_a^{-1}S_s^{-1}\begin{pmatrix}x_1\\x_2\end{pmatrix} - \begin{pmatrix}t_1\\t_2\end{pmatrix}\right).$$

Definition 4.2. For $\psi \in L^2(\mathbb{R}^2)$, the *continuous shearlet system* is defined by

$$\mathbf{SH}(\psi) := \{\psi_{a,s,t} : a > 0,\ s \in \mathbb{R},\ t \in \mathbb{R}^2\}. \tag{4.1}$$

Similarly to the wavelet case, the *continuous shearlet transform* is the map that associates to a function $f \in L^2(\mathbb{R}^2)$ its components along $\mathbf{SH}(\psi)$.

Definition 4.3. For $\psi \in L^2(\mathbb{R}^2)$, the *continuous shearlet transform* of $f \in L^2(\mathbb{R}^2)$ is

$$\mathcal{SH}_\psi f(a,s,t) := \langle f, \psi_{a,s,t}\rangle,\ \text{for } a > 0,\ s \in \mathbb{R},\ t \in \mathbb{R}^2. \tag{4.2}$$

We now seek to find suitable generating functions ψ such that the system $\mathbf{SH}(\psi)$ satisfies a reproducing formula for $L^2(\mathbb{R}^2)$, i.e. the mapping $\mathcal{SH}_\psi$ is a multiple of an isometry.

To state this result precisely, we need to point out some group structure of $\mathbf{SH}(\psi)$. Let us introduce the *Shearlet group* $\mathbb{S}$ defined as the semi-direct product

$$(\mathbb{R}_+ \times \mathbb{R}) \ltimes \mathbb{R}^2,$$

equipped with the multiplication

$$(a,s,t)\cdot(a',s',t') := \left(aa', s + \sqrt{a}s', t + S_s A_a t'\right).$$

It is easy to see that $d\mu := \frac{da}{a^3}dsdt$ is a left Haar measure on $\mathbb{S}$. Let us define the representation $\sigma\colon \mathbb{S} \longmapsto \mathcal{U}(L^2(\mathbb{R}^2))$ as

$$\sigma(a,s,t)\psi := T_t D_{A_a} D_{S_s}\psi, \qquad \psi \in L^2\left(\mathbb{R}^2\right),$$

where $\mathcal{U}(L^2(\mathbb{R}^2))$ is the group of the unitary operators on $L^2(\mathbb{R}^2)$. Using this notation we can represent the shearlet system as

$$\mathbf{SH}(\psi) = \{\sigma(a,s,t)\psi : (a,s,t) \in \mathbb{S}\}.$$

Hence, we search for sufficient conditions on ψ ensuring that the continuous shearlet transform $\mathcal{SH}_\psi$ is a multiple of an isometry.

Definition 4.4. A function $\psi \in L^2(\mathbb{R}^2)$, $\psi \neq 0$, such that

$$\int_{\mathbb{R}^2} \frac{|\mathcal{F}\psi(\xi_1, \xi_2)|^2}{\xi_1^2} d\xi_2 d\xi_1 < +\infty, \tag{4.3}$$

is referred to as *admissible shearlet*.

The next result shows that an admissible shearlet gives that $\mathcal{SH}_\psi$ is a multiple of an isometry.

Theorem 4.5. *[Kutyniok and Labate, 2012] Let $\psi \in L^2(\mathbb{R}^2)$ be an admissible shearlet, and define*

$$C_\psi^+ := \int_0^{+\infty} \int_{\mathbb{R}} \frac{|\mathcal{F}\psi(\xi_1, \xi_2)|^2}{\xi_1^2} d\xi_2 d\xi_1,$$
$$C_\psi^- := \int_{-\infty}^0 \int_{\mathbb{R}} \frac{|\mathcal{F}\psi(\xi_1, \xi_2)|^2}{\xi_1^2} d\xi_2 d\xi_1.$$

If $C_\psi := C_\psi^+ = C_\psi^- < +\infty$, then, for every $f \in L^2(\mathbb{R}^2)$, $\mathcal{SH}_\psi$ satisfies

$$\int_{\mathbb{S}} |\mathcal{SH}_\psi f(a, s, t)|^2 d\mu = 2\pi C_\psi \int_{\mathbb{R}^2} |f(x_1, x_2)|^2 dx_1 dx_2.$$

In particular, if $C_\psi = \frac{1}{2\pi}$, then $\mathcal{SH}_\psi$ is an isometry.

An important class of admissible shearlets satisfying the assumption of the theorem is the class of the so-called *classical shearlets*.

Definition 4.6. A function $\psi \in L^2(\mathbb{R}^2)$ is said to be a classical shearlet if it is defined by

$$\mathcal{F}\psi(\xi_1, \xi_2) = \mathcal{F}\psi_1(\xi_1) \mathcal{F}\psi_2\left(\frac{\xi_2}{\xi_1}\right),$$

where $\psi_1 \in L^2(\mathbb{R})$ satisfies the discrete Calderón condition

$$\sum_{j \in \mathbb{Z}} |\mathcal{F}\psi_1\left(2^{-j}\xi\right)|^2 = 1 \qquad a.e.\ \xi \in \mathbb{R}, \tag{4.4}$$

with $\mathcal{F}\psi_1 \in C^\infty(\mathbb{R})$, and $\operatorname{supp} \mathcal{F}\psi_1 \subseteq \left[-\frac{1}{2}, -\frac{1}{16}\right] \cup \left[\frac{1}{16}, \frac{1}{2}\right]$, and $\psi_2 \in L^2(\mathbb{R})$ satisfies

$$\sum_{n=-1}^{1} |\mathcal{F}\psi_2(\xi + n)|^2 = 1 \qquad a.e.\ \xi \in [-1, 1], \tag{4.5}$$

with $\mathcal{F}\psi_2 \in C^\infty(\mathbb{R})$, and $\operatorname{supp} \mathcal{F}\psi_2 \subseteq [-1, 1]$.

Hence, a classical shearlet can be described as a function possessing wavelet properties along one axis, and bump-like properties along the other. There exist several constructions of classical shearlets. One possible choice is to consider ψ_1 to be a Lemariè-Meyer wavelet (see [Hernández and Weiss, 1996], Section 1.4), and $\mathcal{F}\psi_2$ to be a spline.

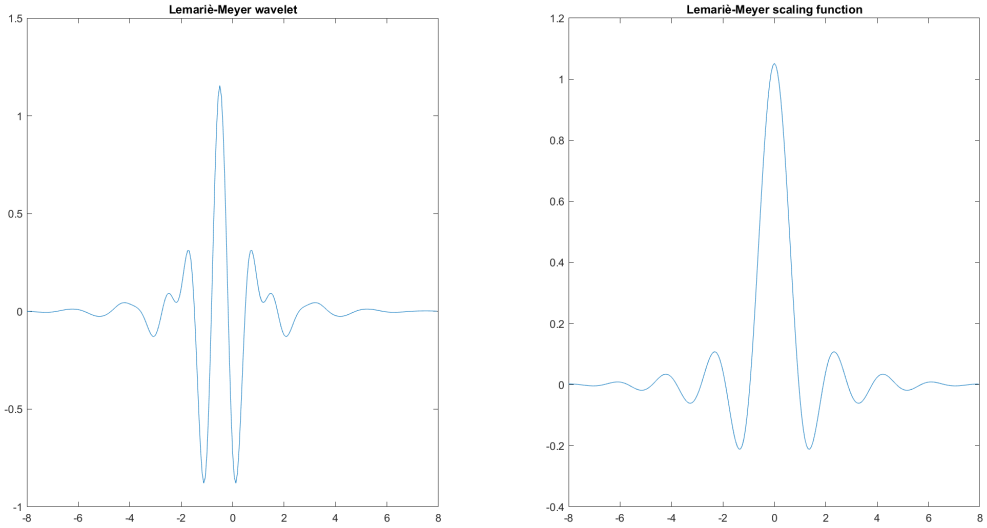


Figure 4.1: Lemariè-Meyer wavelet, and Lemariè-Meyer scaling function

As said above, the classical shearlets verify the assumption of the theorem.

Lemma 4.7. *Let $\psi \in L^2(\mathbb{R}^2)$ be a classical shearlet, then $C_\psi^+ = C_\psi^- = \frac{1}{2\pi}$.*

4.2 Discrete Shearlet Systems

By sampling the continuous shearlet systems, various discrete shearlet systems can be constructed. Throughout the following sections, we will use the sampling scheme proposed in [Guo and Labate, 2007].

Let us introduce the matrices

$$A := \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}, \quad S := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \tag{4.6}$$

We consider the shearlet system

$$\mathbf{SH}(\psi) = \{\psi_{j,l,k} = 2^{\frac{3}{2}j}\psi\left(S^l A^j \cdot -k\right) : j, l \in \mathbb{Z},\ k \in \mathbb{Z}^2\}, \tag{4.7}$$

where

$$A^j = \begin{pmatrix} 4^j & 0 \\ 0 & 2^j \end{pmatrix}, \quad S^l = \begin{pmatrix} 1 & l \\ 0 & 1 \end{pmatrix}, \qquad j, l \in \mathbb{Z}.$$

It is useful to observe that applying the Fourier transform, we obtain

$$\mathcal{F}\psi_{j,l,k}(\xi) = 2^{-\frac{3}{2}j}\mathcal{F}\psi\left(\xi A^{-j}S^{-l}\right)\mathrm{e}^{-i\xi A^{-j}S^{-l}k}. \tag{4.8}$$

As in the wavelet case, we are interested in shearlet systems forming an orthonormal basis or, more in general, a Parseval frame (see Appendix B) for $L^2(\mathbb{R}^2)$. In particular, in applications, classical shearlet systems are not frequently used. Indeed, although they have an elegant group structure, they also posses a directional bias.

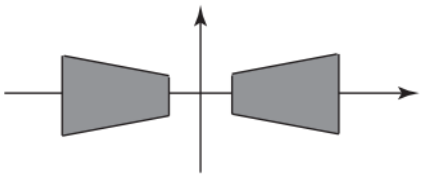


Figure 4.2: Classical shearlet support in the frequency domain.

In order to show the impact of this bias, consider a function f that is concentrated along the ξ_2 axis in the frequency domain. As shown in Figure 4.2, the shearlets are supported in pairs of trapezoids, and the slope of their edges is $\frac{l}{2^j}$. Therefore, we can observe that the energy of f is mostly concentrated in the shearlet components for the shearing parameter $l \to \infty$. Hence, it is clear that this can be a significant constraint for some applications. Here, the main problem is the fact that the shearing parameter l takes on values on an unbounded set. Therefore, to address this problem, we have to restrict the set over which the shearing parameter can take on values. The most common approach is to partition the frequency domain into four cones, and a square centered around the origin (see Figure 4.3).

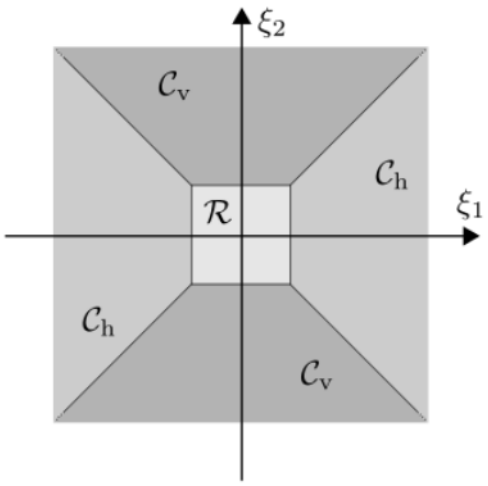


Figure 4.3: Partition of the frequency domain.

This strategy gives rise to the so-called *cone-adapted shearlet system*, where a frame of $L^2(\mathbb{R}^2)$ is constructed by defining three different frames corresponding to the three regions $\mathcal{C}_{\mathrm{h}}$, $\mathcal{C}_{\mathrm{v}}$, and $\mathcal{R}$. Specifically, in the following, we consider $\mathcal{R}$ to be the square centered at the origin of side length $\frac{1}{8}$, $\mathcal{C}_{\mathrm{h}} = \{(\xi_1, \xi_2) : |\xi_1| \geq \frac{1}{8},\ |\frac{\xi_2}{\xi_1}| \leq 1\}$ the horizontal cones, and $\mathcal{C}_{\mathrm{v}} = \{(\xi_1, \xi_2) : |\xi_2| \geq \frac{1}{8},\ |\frac{\xi_1}{\xi_2}| \leq 1\}$ the vertical ones. Then, we construct a Parseval frame of $L^2\left(\mathcal{C}_{\mathrm{h}}\right)^\vee$, a Parseval frame of $L^2\left(\mathcal{C}_{\mathrm{v}}\right)^\vee$, and one of $L^2\left(\mathcal{R}\right)^\vee$, where, for $\mathcal{A} \subset \mathbb{R}^2$,

$$L^2(\mathcal{A})^\vee = \{f \in L^2\left(\mathbb{R}^2\right) : \operatorname{supp}\left(\mathcal{F}f\right) \subseteq \mathcal{A}\}.$$

Finally, the union of this three Parseval frames will provide a Parseval frame of $L^2(\mathbb{R}^2)$.

We start by constructing a Parseval frame of $L^2\left(\mathcal{C}_{\mathrm{h}}\right)^\vee$. On this matter, we consider $\psi \in L^2(\mathbb{R}^2)$ given by

$$\mathcal{F}\psi(\xi_1, \xi_2) = \mathcal{F}\psi_1(\xi_1)\mathcal{F}\psi_2\left(\frac{\xi_2}{\xi_1}\right), \qquad (\xi_1, \xi_2) \in \mathbb{R}^2, \xi_1 \neq 0, \tag{4.9}$$

where $\mathcal{F}\psi_1, \mathcal{F}\psi_2 \in C^\infty(\mathbb{R})$, $\operatorname{supp}\mathcal{F}\psi_1 \subseteq \left[-\frac{1}{2}, -\frac{1}{16}\right] \cup \left[\frac{1}{16}, \frac{1}{2}\right]$, and $\operatorname{supp}\mathcal{F}\psi_2 \subseteq [-1, 1]$. In particular, it implies that $\mathcal{F}\psi \in C^\infty\left(\mathbb{R}^2\right)$, with $\operatorname{supp}\left(\mathcal{F}\psi\right) \subseteq \left[-\frac{1}{2}, \frac{1}{2}\right]^2$. Moreover, we assume that

$$\sum_{j\geq 0} |\mathcal{F}\psi_1(2^{-2j}\xi)|^2 = 1, \qquad \text{a.e. } |\xi| \geq \frac{1}{8}, \tag{4.10}$$

and

$$\sum_{l=-1}^{1} |\mathcal{F}\psi_2(\xi + l)|^2 = 1, \qquad \text{a.e. } |\xi| \leq 1. \tag{4.11}$$

We refer to Appendix D for an explicit construction of a function satisfying these properties.

Lemma 4.8. *Let $f\colon \mathbb{R} \longrightarrow \mathbb{C}$ be a function such that* $\operatorname{supp} f \subseteq [-1, 1]$. *Suppose that*

$$\sum_{l=-1}^{1} |f(x + l)|^2 = 1, \qquad \textit{a.e. } |x| \leq 1,$$

then for any integer $j \geq 0$

$$\sum_{l=-2^j}^{2^j} |f\left(2^j x + l\right)|^2 = 1, \qquad \textit{a.e. } |x| \leq 1.$$

Proof. The proof proceeds by induction on j. For $j = 0$ it is trivial. Now, for any $j \geq 1$, we suppose that the thesis holds for $j - 1$, and we demonstrate it for

j. Let us first show the proof for a.e. $x \in \left(-\frac{1}{2}, \frac{1}{2}\right)$. In this case, we have that $2^j x + l \in (l - 2^{j-1}, l + 2^{j-1})$. Hence, since f is supported within $[-1, 1]$, we have

$$\sum_{l=-2^j}^{2^j} |f\left(2^j x + l\right)|^2 = \sum_{l=-2^{j-1}}^{2^{j-1}} |f\left(2^j x + l\right)|^2.$$

By the change of variable $y = 2x \in (-1, 1)$, we can conclude due to the inductive hypothesis

$$\sum_{l=-2^{j-1}}^{2^{j-1}} |f\left(2^j x + l\right)|^2 = \sum_{l=-2^{j-1}}^{2^{j-1}} |f\left(2^{j-1} y + l\right)|^2 = 1.$$

Now, we show the result for $x \in \left[\frac{1}{2}, 1\right)$. The proof for the symmetric case $x \in \left(-1, -\frac{1}{2}\right]$ is analogous. Let us consider the following partition

$$\left[\frac{1}{2}, 1\right) = \bigcup_{k=0}^{2^{j-1}-1} \left[\frac{1}{2} + \frac{k}{2^j}, \frac{1}{2} + \frac{k+1}{2^j}\right)$$

Let us prove the thesis for $x \in \left[\frac{1}{2} + \frac{k}{2^j}, \frac{1}{2} + \frac{k+1}{2^j}\right)$. This implies that $2^j x + l \in [l + 2^{j-1} + k, l + 2^{j-1} + k + 1)$. Therefore, since f is compactly supported within $[-1, 1]$, we have

$$\sum_{l=-2^j}^{2^j} |f\left(2^j x + l\right)|^2 = \sum_{l=-2^{j-1}-k-1}^{-2^{j-1}-k+1} |f\left(2^j x + l\right)|^2.$$

By the change of variable $y = 2^j x - 2^{j-1} - k \in [0, 1)$, we can conclude the proof

$$\sum_{l=-2^{j-1}-k-1}^{-2^{j-1}-k+1} |f\left(y + l + 2^{j-1} + k\right)|^2 = \sum_{l=-1}^{1} |f\left(y + l\right)|^2 = 1.$$

□

Proposition 4.9. *The system*

$$\boldsymbol{SH}(\psi) = \{\psi_{j,l,k} = 2^{\frac{3}{2}j} \psi\left(S^l A^j \cdot - k\right) : j \geq 0,\ -2^j \leq l \leq 2^j,\ k \in \mathbb{Z}^2\} \tag{4.12}$$

is a Parseval frame of $L^2(\mathcal{C}_{\mathrm{h}})^\vee$.

Proof. Lemma 4.8 applied to $f = \mathcal{F}\psi_2$ shows that, for any integer $j \geq 0$,

$$\sum_{l=-2^j}^{2^j} |\mathcal{F}\psi_2\left(2^j \xi + l\right)|^2 = 1, \qquad \text{a.e. } |\xi| \leq 1. \tag{4.13}$$

By using (4.10), (4.13), and observing that $\xi A^{-j}S^{-l} = (2^{-2j}\xi_1, 2^{-j}\xi_2 - 2^{-2j}l\xi_1)$, we obtain

$$\begin{aligned}\sum_{j=0}^{+\infty}\sum_{l=-2^j}^{2^j}|\mathcal{F}\psi\left(\xi A^{-j}S^{-l}\right)|^2 &= \sum_{j=0}^{+\infty}\sum_{l=-2^j}^{2^j}|\mathcal{F}\psi_1\left(2^{-2j}\xi_1\right)|^2|\mathcal{F}\psi_2\left(2^j\frac{\xi_2}{\xi_1}-l\right)|^2\\ &= \sum_{j=0}^{+\infty}|\mathcal{F}\psi_1\left(2^{-2j}\xi_1\right)|^2\sum_{l=-2^j}^{2^j}|\mathcal{F}\psi_2\left(2^j\frac{\xi_2}{\xi_1}-l\right)|^2 = 1,\end{aligned} \tag{4.14}$$

for $\xi = (\xi_1, \xi_2) \in \mathcal{C}_{\rm h}$. The fact that, for any $f \in L^2(\mathcal{C}_{\rm h})^\vee$,

$$\sum_{j=0}^{+\infty}\sum_{l=-2^j}^{2^j}\sum_{k\in\mathbb{Z}^2}|\langle f, \psi_{j,l,k}\rangle|^2 = \|f\|_2^2,$$

is a consequence of (4.14), and the fact that $\operatorname{supp}(\mathcal{F}\psi) \subseteq \left[-\frac{1}{2}, \frac{1}{2}\right]^2$. Indeed, due to Plancharel identity, using (4.8) we have

$$\begin{aligned}&\sum_{j=0}^{+\infty}\sum_{l=-2^j}^{2^j}\sum_{k\in\mathbb{Z}^2}|\langle f, \psi_{j,l,k}\rangle|^2 = \frac{1}{(2\pi)^2}\sum_{j=0}^{+\infty}\sum_{l=-2^j}^{2^j}\sum_{k\in\mathbb{Z}^2}|\langle \mathcal{F}f, \mathcal{F}\psi_{j,l,k}\rangle|^2\\ &= \frac{1}{(2\pi)^2}\sum_{j=0}^{+\infty}\sum_{l=-2^j}^{2^j}\sum_{k\in\mathbb{Z}^2}|\int_{\mathcal{D}_C}\mathcal{F}f\left(\xi\right)2^{-\frac{3}{2}j}\mathcal{F}\psi\left(\xi A^{-j}S^{-l}\right)\mathrm{e}^{-i\xi A^{-j}S^{-l}k}\,d\xi|^2.\end{aligned}$$

Now, by the change of variable $\omega = \xi A^{-j}S^{-l}$, we obtain

$$\begin{aligned}&\int_{\mathcal{D}_C}\mathcal{F}f\left(\xi\right)2^{-\frac{3}{2}j}\mathcal{F}\psi\left(\xi A^{-j}S^{-l}\right)\mathrm{e}^{-i\xi A^{-j}S^{-l}k}\,d\xi\\ &= \int_{\left[-\frac{1}{2},\frac{1}{2}\right]^2}\mathcal{F}f\left(\omega S^l A^j\right)\mathcal{F}\psi\left(\omega\right)2^{\frac{3}{2}j}\,\mathrm{e}^{-i\omega k}\,d\omega.\end{aligned}$$

Therefore,

$$\begin{aligned}&\sum_{j=0}^{+\infty}\sum_{l=-2^j}^{2^j}\sum_{k\in\mathbb{Z}^2}|\langle f, \psi_{j,l,k}\rangle|^2\\ &= \frac{1}{(2\pi)^2}\sum_{j=0}^{+\infty}\sum_{l=-2^j}^{2^j}\sum_{k\in\mathbb{Z}^2}|\int_{\left[-\frac{1}{2},\frac{1}{2}\right]^2}\mathcal{F}f\left(\omega S^l A^j\right)\mathcal{F}\psi\left(\omega\right)2^{\frac{3}{2}j}\,\mathrm{e}^{-i\omega k}\,d\omega|^2\\ &= \frac{1}{(2\pi)^2}\sum_{j=0}^{+\infty}\sum_{l=-2^j}^{2^j}2^{3j}\sum_{k\in\mathbb{Z}^2}|\int_{\left[-\frac{1}{2},\frac{1}{2}\right]^2}\mathcal{F}f\left(\omega S^l A^j\right)\mathcal{F}\psi\left(\omega\right)\mathrm{e}^{-i\omega k}\,d\omega|^2.\end{aligned}$$

By using the Parseval identity, we have

$$\sum_{k\in\mathbb{Z}^2}|\int_{\left[-\frac{1}{2},\frac{1}{2}\right]^2}\mathcal{F}f\left(\omega S^lA^j\right)\mathcal{F}\psi\left(\omega\right)\mathrm{e}^{-i\omega k}\,d\omega|^2=2\pi\int_{\left[-\frac{1}{2},\frac{1}{2}\right]^2}|\mathcal{F}f\left(\omega S^lA^j\right)|^2|\mathcal{F}\psi\left(\omega\right)|^2d\omega.$$

Hence,

$$\sum_{j=0}^{+\infty}\sum_{l=-2^j}^{2^j}\sum_{k\in\mathbb{Z}^2}|\langle f,\psi_{j,l,k}\rangle|^2=\frac{1}{2\pi}\sum_{j=0}^{+\infty}\sum_{l=-2^j}^{2^j}\int_{\left[-\frac{1}{2},\frac{1}{2}\right]^2}2^{3j}|\mathcal{F}f\left(\omega S^lA^j\right)|^2|\mathcal{F}\psi\left(\omega\right)|^2d\omega.$$

Applying the change of variable $\xi=\omega S^lA^j$, we have

$$\int_{\left[-\frac{1}{2},\frac{1}{2}\right]^2}2^{3j}|\mathcal{F}f\left(\omega S^lA^j\right)|^2|\mathcal{F}\psi\left(\omega\right)|^2d\omega=\int_{\mathcal{D}_C}|\mathcal{F}f\left(\xi\right)|^2|\mathcal{F}\psi\left(\xi A^{-j}S^{-l}\right)|^2d\xi.$$

Now, using (4.14), we can conclude the proof

$$\begin{aligned}\sum_{j=0}^{+\infty}\sum_{l=-2^j}^{2^j}\sum_{k\in\mathbb{Z}^2}|\langle f,\psi_{j,l,k}\rangle|^2&=\frac{1}{2\pi}\sum_{j=0}^{+\infty}\sum_{l=-2^j}^{2^j}\int_{\mathcal{D}_C}|\mathcal{F}f\left(\xi\right)|^2|\mathcal{F}\psi\left(\xi A^{-j}S^{-l}\right)|^2d\xi\\&=\frac{1}{2\pi}\int_{\mathcal{D}_C}|\mathcal{F}f(\xi)|^2d\xi=\frac{1}{2\pi}\|\mathcal{F}f\|_2^2=\|f\|_2^2.\end{aligned}$$

□

To obtain a Parseval frame of $L^2\left(\mathbb{R}^2\right)$, we need to construct a Parseval frame of $L^2\left(\mathcal{C}_{\mathrm{v}}\right)^\vee$, and a Parseval frame of $L^2\left(\mathcal{R}\right)^\vee$. The first one can be constructed similarly to the one constructed for $\mathcal{C}_{\mathrm{h}}$ reversing the roles of ψ_1, and ψ_2. Namely, we can define $\mathcal{F}\tilde{\psi}(\xi_1,\xi_2)=\mathcal{F}\psi_1(\xi_2)\mathcal{F}\psi_2\left(\frac{\xi_1}{\xi_2}\right)$, where ψ_1,ψ_2 are defined as above. Then, following the proof of the previous proposition, it can be proved that the system

$$\mathbf{SH}\left(\tilde{\psi}\right)=\{\tilde{\psi}_{j,l,k}=2^{\frac{3}{2}j}\tilde{\psi}\left(\tilde{S}^l\tilde{A}^j\cdot-k\right):j\geq0,\ -2^j\leq l\leq2^j,\ k\in\mathbb{Z}^2\},\quad(4.15)$$

where

$$\tilde{A}:=\begin{pmatrix}2&0\\0&4\end{pmatrix},\quad\tilde{S}:=\begin{pmatrix}1&0\\1&1\end{pmatrix},$$

is a Parseval frame of $L^2\left(\mathcal{C}_{\mathrm{v}}\right)^\vee$. The second one can be constructed through a *shearlet scaling function* φ such that its Fourier transform is C^∞ and compactly supported near the origin, so that the family of its translated $\{\varphi\left(\cdot-k\right):k\in\mathbb{Z}^2\}$ forms a Parseval frame of $L^2(\mathcal{R})^\vee$. More details about the construction can be find in [Easley et al., 2008, Kutyniok and Labate, 2012]. With this in hand, we can represent a function $f\in L^2(\mathbb{R}^2)$ as a sum of three components,

$$f=f_{\mathcal{C}_{\mathrm{h}}}+f_{\mathcal{C}_{\mathrm{v}}}+f_{\mathcal{R}},\quad(4.16)$$

where each component is the orthogonal projection of f onto one of the three subspaces, namely

$$f_{\mathcal{C}_\mathrm{h}} = \mathcal{F}^{-1}[\mathcal{F}(f)\mathbb{1}_{\mathcal{C}_\mathrm{h}}],$$
$$f_{\mathcal{C}_\mathrm{v}} = \mathcal{F}^{-1}[\mathcal{F}(f)\mathbb{1}_{\mathcal{C}_\mathrm{v}}],$$
$$f_{\mathcal{R}} = \mathcal{F}^{-1}[\mathcal{F}(f)\mathbb{1}_{\mathcal{R}}].$$

4.3 Non-Linear Shearlets Approximation

In this section, our focus is on investigating the non-linear representation of cartoon-like images using the shearlet decomposition. Given that the primary goal of this thesis is to examine the hierarchical connections between the indices associated with significant shearlet coefficients across various scales (as explained in Chapter 5), our attention is directed towards analyzing the non-linear approximation specific to the shearlet elements exclusively. This section is based on the arguments presented in [Guo and Labate, 2007].

Let $f \in \mathcal{E}(\mathbb{R}^2)$ be a cartoon-like imagage, introduced in Definition 4.1, and consider its decomposition as the sum of $f_{\mathcal{C}_\mathrm{h}}$, $f_{\mathcal{C}_\mathrm{v}}$, and $f_{\mathcal{R}}$. The next results analyze the best N term approximation of $f_{\mathcal{C}_\mathrm{h}} + f_{\mathcal{C}_\mathrm{v}}$. Because the construction of the shearlet on the vertical cones is symmetrical to that of the horizontal cones, it is sufficient to investigate the non-linear approximation for the horizontal cones.

Let $M = \{(j,l,k) : j \geq 0, -2^j \leq l \leq 2^j, k \in \mathbb{Z}^2\}$, and $\{\psi_\mu\}_{\mu\in M}$ be the Parseval frame defined in (4.12). The sequence of the shearlet coefficients of f is $s(f) = \{\langle f, \psi_\mu\rangle : \mu \in M\}$. We rearrange this sequence in a decreasing way, and we denote by $|s(f)|_{(N)}$ the N-th entry of the reordered sequence, i.e. the N-th largest shearlet coefficient. In order to analyze the sparsity of the shearlet coefficients, we will use the *weak-ℓ^p-quasi-norm*. Let us consider $s = (s_\mu)$ a sequence, and let $|s_\mu|_{(N)}$ be its N-th largest entry. The weak-ℓ^p-quasi-norm is defined as

$$\|s\|_{w\ell^p} := \left(\sup_{\epsilon>0} |\{\mu : |s_\mu| > \epsilon\}|\epsilon^p\right)^{\frac{1}{p}}.$$

Equivalently, it can be defined as

$$\|s\|_{w\ell^p} = \sup_{N>0} N^{\frac{1}{p}}|s_\mu|_{(N)}.$$

Further details about the weak-ℓ^p-norms can be found in [Grafakos et al., 2008].

We first analyze the decay of the coefficients at a given scale 2^{-j}. In order to do that, we need to localize the function on dyadic squares. Fix the scale parameter $j \geq 0$, and consider the sequence of shearlet coefficients at scale 2^{-j}, denoted by $s_j(f) = \{\langle f, \psi_\mu\rangle : \mu \in M_j\}$, where $M_j = \{(j,l,k) : -2^j \leq l \leq 2^j, k \in \mathbb{Z}^2\}$. Let

us partition $\mathbb{R}^2$ in dyadic squares with side length of 2^{-j}, i.e. squares of the form $Q = [\frac{k_1}{2^j}, \frac{k_1+1}{2^j}] \times [\frac{k_2}{2^j}, \frac{k_2+1}{2^j}]$, $k_1, k_2 \in \mathbb{Z}$. Let us consider a smooth partition of unity

$$\sum_Q w_Q(x) = 1 \quad \text{for every } x \in \mathbb{R}^2,$$

where $w_Q(x_1, x_2) = w(2^j x_1 - k_1, 2^j x_2 - k_2)$, and w is a C^∞ non-negative function supported within $[-1, 1]^2$. Let $\mathcal{Q}_j$ be the collection of the 2^{2j} dyadic squares contained in $[0, 1]^2$. Now we study the decay of the shearlet coefficients of localized function $f_Q = f w_Q$. Since f is compactly supported within $[0, 1]^2$, the only squares that matter are those in $\mathcal{Q}_j$. We will see that there is a different decay rate of the coefficients depending on whether Q intersects or not the boundary of B. Let us split $\mathcal{Q}_j$ into two disjoint families $\mathcal{Q}_j^0$ and $\mathcal{Q}_j^1$, where $\mathcal{Q}_j^0$ is the collection of those squares such that the support of w_Q intersects ∂B, and $\mathcal{Q}_j^1 = \mathcal{Q}_j \backslash \mathcal{Q}_j^0$. In particular, we have that $|\mathcal{Q}_j^0| \leq C_0 2^j$ (see Appendix C), while obviously $|\mathcal{Q}_j^1| \leq 2^{2j}$. Let us consider the sequence of the localized shearlet coefficients $s_j^Q(f) = \{\langle f_Q, \psi_\mu \rangle : \mu \in M_j\}$ for some $Q \in \mathcal{Q}_j$. Then the following lemmas hold. Their proofs are omitted.

Lemma 4.10. *[Guo and Labate, 2007] Let $f \in \mathcal{E}(\mathbb{R}^2)$. For $Q \in \mathcal{Q}_j^0$, the sequence $s_j^Q(f)$ obeys*

$$\|s_j^Q(f)\|_{w\ell^{\frac{2}{3}}} \leq C 2^{-\frac{3}{2}j},$$

for some positive constant C independent of Q and j.

Lemma 4.11. *[Guo and Labate, 2007] Let $f \in \mathcal{E}(\mathbb{R}^2)$. For $Q \in \mathcal{Q}_j^1$, the sequence $s_j^Q(f)$ obeys*

$$\|s_j^Q(f)\|_{w\ell^{\frac{2}{3}}} \leq C 2^{-3j},$$

for some positive constant C independent of Q and j.

As a consequence, we obtain the following result.

Corollary 4.12. *Let $f \in \mathcal{E}(\mathbb{R}^2)$. The sequence $s_j(f)$ obeys*

$$\|s_j(f)\|_{w\ell^{\frac{2}{3}}} \leq C,$$

for some positive constant C independent of j.

In order to prove this corollary, we need to recall a property of the weak-ℓ^p-quasi-norm.

Lemma 4.13. *Consider (X, μ) a measure space, $0 < p < 1$, and $f_1, \ldots, f_m$ measurable functions defined on X. Then*

$$\|\sum_{j=1}^m f_j\|_{w\ell^p}^p \leq \frac{2-p}{1-p} \sum_{j=1}^m \|f_j\|_{w\ell^p}^p,$$

where

$$\|f\|_{w\ell^p}^p = \sup_{\alpha>0} \alpha^p \mu\left(\{x \in X : |f(x)| > \alpha\}\right).$$

Proof. Observe that, if $\|f_j\|_{w\ell^p}^p = +\infty$ for some j, then the proof is trivial. Hence, we suppose that $\|f_j\|_{w\ell^p}^p < +\infty$ for every j. Firstly, we prove that, for any $s > 0$, and for any measurable function f on X,

$$\int_{E_s} |f(x)| d\mu(x) \leq \frac{s^{1-p}}{1-p} \|f\|_{w\ell^p}^p, \tag{4.17}$$

where $E_s = \{x \in X : |f(x)| \leq s\}$. By definition of Lebesgue integral and weak ℓ^p norm, we can observe that

$$\begin{aligned}
\int_{E_s} |f(x)| d\mu(x) &= \int_0^{+\infty} \mu(\{x \in E_s : |f(x)| > \alpha\}) d\alpha \\
&\leq \int_0^s \mu(\{x \in X : |f(x)| > \alpha\}) d\alpha \\
&\leq \|f\|_{w\ell^p}^p \int_0^s \alpha^{-p} d\alpha = \frac{s^{1-p}}{1-p} \|f\|_{w\ell^p}^p.
\end{aligned}$$

Now, we prove that, for any $\alpha > 0$,

$$\alpha^p \mu(E_\alpha) \leq \frac{1}{1-p} \sum_{j=1}^m \|f_j\|_{w\ell^p}^p, \tag{4.18}$$

where

$$E_\alpha = \{x \in X : |\sum_{j=1}^m f_j| > \alpha, \max_{j=1,\dots,m} |f_j(x)| \leq \alpha\}.$$

Due to (4.17), we note

$$\begin{aligned}
\alpha\mu(E_\alpha) = \alpha \int_{E_\alpha} d\mu(x) &\leq \sum_{j=1}^m \int_{\{x \in X : \max_{i=1,\dots,m} |f_i(x)| \leq \alpha\}} |f_j(x)| d\mu(x) \\
&\leq \sum_{j=1}^m \int_{\{x \in X : |f_j(x)| \leq \alpha\}} |f_j(x)| d\mu(x) \leq \frac{\alpha^{1-p}}{1-p} \sum_{j=1}^m \|f_j\|_{w\ell^p}^p.
\end{aligned}$$

By multiplying both sides for α^{p-1}, we conclude the proof of (4.18). Let us also note that

$$\mu(\{x \in X : \max_{j=1,\dots,m} |f_j(x)| > \alpha\}) \leq \sum_{j=1}^m \mu(\{x \in X : |f_j(x)| > \alpha\}). \tag{4.19}$$

In order to prove the statement, we observe that, due to (4.18), and (4.19),

$$\alpha^p\mu(\{x\in X:|\sum_{j=1}^m f_j(x)|>\alpha\})\leq \alpha^p\mu(E_\alpha)+\alpha^p\mu(\{x\in X:\max_{j=1,\dots,m}|f_j(x)|>\alpha\})$$
$$\leq \frac{1}{1-p}\sum_{j=1}^m\|f_j\|^p_{w\ell^p}+\sum_{j=1}^m\alpha^p\mu(\{x\in X:|f_j(x)|>\alpha\}).$$

By taking the sup over all the positive number α, we conclude the proof. □

Proof of Corollary 4.12. Using the previous lemmas, and writing the coefficients for $\mu\in M_j$ as

$$\begin{aligned}\langle f,\psi_\mu\rangle=\langle f\sum_Q w_Q,\psi_\mu\rangle&=\sum_{Q\in\mathcal{Q}_j}\langle f_Q,\psi_\mu\rangle\\&=\sum_{Q\in\mathcal{Q}_j^0}\langle f_Q,\psi_\mu\rangle+\sum_{Q\in\mathcal{Q}_j^1}\langle f_Q,\psi_\mu\rangle,\end{aligned}$$

we obtain

$$\begin{aligned}\|s_j(f)\|^{\frac{2}{3}}_{w\ell^{\frac{2}{3}}}&\lesssim\sum_{Q\in\mathcal{Q}_j^0}\|\langle f_Q,\psi_\mu\rangle\|^{\frac{2}{3}}_{w\ell^{\frac{2}{3}}}+\sum_{Q\in\mathcal{Q}_j^1}\|\langle f_Q,\psi_\mu\rangle\|^{\frac{2}{3}}_{w\ell^{\frac{2}{3}}}\\&\leq C_1|\mathcal{Q}_j^0|2^{-j}+C_2|\mathcal{Q}_j^1|2^{-2j}.\end{aligned}$$

We complete the proof by using the upper bounds on the cardinalities of $\mathcal{Q}_j^0$, and $\mathcal{Q}_j^1$. □

With these results in mind, we can now state and prove the main results of this section.

Theorem 4.14. *We have*

$$\sup_{f\in\mathcal{E}(\mathbb{R}^2)}|s(f)|_{(N)}\leq CN^{-\frac{3}{2}}\log^{\frac{3}{2}}(N).$$

Proof. For each $\epsilon>0$ and $j\in\mathbb{N}$, define

$$R(j,\epsilon):=|\{\mu\in M_j:|\langle f,\psi_\mu\rangle|>\epsilon\}|.$$

By Corollary 4.12, we have that

$$C_1\geq\|s_j(f)\|^{\frac{2}{3}}_{w\ell^{\frac{2}{3}}}\geq R(j,\epsilon)\epsilon^{\frac{2}{3}},$$

therefore

$$R(j,\epsilon)\leq C_1\epsilon^{-\frac{2}{3}}.$$

Moreover, for each $\mu \in M_j$

$$|\langle f, \psi_\mu \rangle| = |\int_{\mathbb{R}^2} f(x) 2^{\frac{3}{2}j} \psi\left(S^l A^j x - k\right) dx| \leq 2^{-\frac{3}{2}j} \|f\|_\infty \int_{\mathbb{R}^2} |\psi(x)| dx \leq C' 2^{-\frac{3}{2}j}.$$

In particular, we have that

$$C' 2^{-\frac{3}{2}j} \leq \epsilon \iff j \geq \frac{2}{3}\left(\log_2(C') + \log_2\left(\frac{1}{\epsilon}\right)\right) =: j_\epsilon.$$

Hence, we can conclude that $R(j, \epsilon) = 0$ for every $j > j_\epsilon$. Let us define

$$R(\epsilon) := |\{\mu \in M : |\langle f, \psi_\mu \rangle| > \epsilon\}|,$$

we can observe

$$R(\epsilon) \leq \sum_{j \geq 0} R(j, \epsilon) \leq C_2 \epsilon^{-\frac{2}{3}} \log_2\left(\frac{1}{\epsilon}\right),$$

which is equivalent to

$$R(\epsilon) \epsilon^{\frac{2}{3}} \log_2^{-1}\left(\frac{1}{\epsilon}\right) \leq C_2.$$

Let us consider the function $g(t) = t \log_2^{-\frac{3}{2}}\left(\frac{1}{t}\right)$ for $t > 0$. It is a positive, strictly increasing function such that $\lim_{t \to 0^+} g(t) = 0^+$. Fix $\eta = g(\epsilon)$, hence we have

$$R(g^{-1}(\eta)) \eta^{\frac{2}{3}} \leq C_2.$$

By observing that

$$R(g^{-1}(\eta)) = |\{\mu \in M : |\langle f, \psi_\mu \rangle| > g^{-1}(\eta)\}| = |\{\mu \in M : g(|\langle f, \psi_\mu \rangle|) > \eta\}|,$$

we can conclude $\|g(|\langle f, \psi_\mu \rangle|)\|_{w\ell^{\frac{2}{3}}} \leq C_2$. Therefore, by applying the equivalent definition of weak-ℓ^p-norm, we have

$$g(|s(f)|_{(N)}) \leq C_1 N^{-\frac{3}{2}}.$$

Up to rescaling, we can assume without loss of generality $\|s(f)\|_\infty \leq \frac{1}{2}$. Notice that, if $0 < t \leq \frac{1}{2}$, and $y = t \log_2^{-\frac{3}{2}}\left(\frac{1}{t}\right)$, we have $\log_2^{\frac{3}{2}}\left(\frac{1}{t}\right) \geq 1$, and $t = y \log_2^{\frac{3}{2}}\left(\frac{1}{t}\right) \geq y$. In particular, $\log_2^{\frac{3}{2}}\left(\frac{1}{t}\right) \leq \log_2^{\frac{3}{2}}\left(\frac{1}{y}\right)$. Thus,

$$g^{-1}(y) = t = y \log_2^{\frac{3}{2}}\left(\frac{1}{t}\right) \leq y \log_2^{\frac{3}{2}}\left(\frac{1}{y}\right).$$

Finally, using the monotonicity of g, we can conclude

$$|s(f)|_{(N)} \leq g^{-1}(C_1 N^{-\frac{3}{2}}) \leq C N^{-\frac{3}{2}} \log^{\frac{3}{2}}(N).$$

□

Theorem 4.15. *Let $f \in L^2(\mathcal{C}_{\mathrm{h}})^\vee$ be the horizontal cones component of a cartoon-like image, and f_N be its best N term approximation with respect to the Parseval frame* (4.12)*, namely*

$$f_N := \sum_{\mu \in I_N} \langle f, \psi_\mu \rangle \psi_\mu, \tag{4.20}$$

where $I_N \subset M$ is the set of indices corresponding to the N largest entries. Then

$$\|f - f_N\|_2^2 \lesssim \sum_{m>N} |s(f)|_{(m)}^2 \leq CN^{-2} \log^3(N) \qquad \text{as } N \to +\infty.$$

Proof. By using Theorem 4.14, and the fact that the function $t \mapsto \frac{\log^3(t)}{t}$ decreases for $t \geq \mathrm{e}^3$, we have

$$\begin{aligned} \|f - f_N\|_2^2 &= \sum_{m>N} |s(f)|_{(m)}^2 \leq C \sum_{m>N} m^{-3} \log^3(m) \\ &\leq C \frac{\log^3(N)}{N} \sum_{m>N} \frac{1}{m^2} \leq C \frac{\log^3(N)}{N} \int_N^{+\infty} t^{-2} dt \\ &= CN^{-2} \log^3(N). \end{aligned}$$

□

Now, let $f \in \mathcal{E}(\mathbb{R}^2)$, and consider its decomposition (4.16). Let us define the following N term approximation of f:

$$f_N = f_{\mathcal{R}} + f_{\mathcal{C}_{\mathrm{h}} \cup \mathcal{C}_{\mathrm{v}}, N},$$

where $f_{\mathcal{C}_{\mathrm{h}} \cup \mathcal{C}_{\mathrm{v}}, N}$ is the best N term approximation of $f_{\mathcal{C}_{\mathrm{h}}} + f_{\mathcal{C}_{\mathrm{v}}}$ with respect to the union of the Parseval frames (4.12), and (4.15). In other words, we define

$$f_{\mathcal{C}_{\mathrm{h}} \cup \mathcal{C}_{\mathrm{v}}, N} := \sum_{(j,l,k) \in I_\psi} \langle f_{\mathcal{C}_{\mathrm{h}}}, \psi_{j,l,k} \rangle + \sum_{(j,l,k) \in I_{\tilde{\psi}}} \langle f_{\mathcal{C}_{\mathrm{v}}}, \tilde{\psi}_{j,l,k} \rangle,$$

with I_ψ, $I_{\tilde{\psi}}$ corresponding to the N largest coefficients, and such that $|I_\psi| + |I_{\tilde{\psi}}| = N$. The previous results ensure that

$$\|f - f_N\|_2^2 \lesssim \sum_{(j,l,k) \notin I_\psi} |\langle f_{\mathcal{C}_{\mathrm{h}}}, \psi_{j,l,k} \rangle|^2 + \sum_{(j,l,k) \notin I_{\tilde{\psi}}} |\langle f_{\mathcal{C}_{\mathrm{v}}}, \tilde{\psi}_{j,l,k} \rangle|^2 \leq CN^{-2} \log^3(N).$$

Chapter 5

Shearlets Trees

5.1 Introduction

The non-linear approximation results presented in the previous chapters show that certain types of functions can be efficiently represented through sparse vectors, by decomposing them with respect to wavelet, or shearlet frames. This means that the sequences of the coefficients have just a few relevant components. In various mathematical fields, such as *Compressed Sensing* [Foucart et al., 2013], it is of great interest not only to know that the functions admit sparse representations, but also to have an indication about where the relevant components are located.

In this section, we aim to study this property with a deeper focus in the case of shearlet frames. Since two-dimensional shearlets are based on the idea of orthogonal wavelets, let us start by describing the problem in this latter case. As explained in Chapter 3, the number of vanishing moments of an orthogonal wavelet enables to compress the coefficients corresponding to a smooth region of the function, while its support size is helpful to localize the singularity, and to control the number of wavelets whose support intersects a singularity at each scale. Therefore, we have that the relevant coefficients correspond to those wavelets whose support intersects a singularity. As we can observe in Figure 5.1, this suggests that it is possible to construct structures, in the literature referred to as *trees*, which contain the indices related to the coefficients corresponding to a singularity. The trees are based on hierarchical relations *parent-child* between the indices of the coefficients at scale j, and those at scale $j+1$. Given such a relation, a tree is defined as a set T of indices satisfying the property:

if an index belongs to T, then its parent belongs to T.

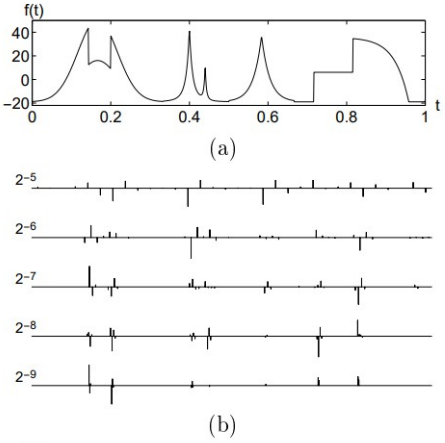


Figure 5.1: Multiscale analysis of a piecewise regular , [Mallat, 1999].

The non-linear approximation forcing the set of indices being a tree is called *tree approximation*. Intuitively, one can think that the set of the N largest coefficients (or, equivalently, the set of the coefficients larger than some threshold $\eta > 0$) is a tree itself, and so, that the best N-term approximation, and the tree approximation are equivalent. Unfortunately, this is not the case. Indeed, although the coefficients decrease to 0 for the scale parameter j growing, they do not decrease monotonically. In other words, it could happen that the amplitude of a child is larger than the parent's one. Therefore, by fixing a threshold $\eta > 0$ and selecting only the coefficients larger than η, there is no guarantee that the set of the indices forms a tree. An easy example of this phenomenon is given by the Haar wavelet (2.3), and the function $f = \mathbb{1}_{\{\frac{1}{5} \leq x \leq 1\}}$. In this case, since at each fixed scale j the Haar basis partition the interval $[0, 1]$ into dyadic intervals, the only reasonable parent-child relation is:

$$(j, k) \text{ is the parent of } (j + 1, 2k), \text{ and } (j + 1, 2k + 1).$$

An easy computation shows that

$$\frac{1}{5} = |\langle f, \psi_{0,0} \rangle| \leq |\langle f, \psi_{1,0} \rangle| = \frac{\sqrt{2}}{5}.$$

More generally, in the case of a piecewise constant function with discontinuity in $x = a \in (0, 1)$, and the Haar wavelet, it is possible to explicitly characterize the parent-child relations. It can be observed that, when the discontinuity point is sufficiently close to the edge of the support of the parent, then the magnitude of the child coefficient will be $\sqrt{2}$ times that of the parent. The situation could be even worse. Indeed, the Haar wavelet does not have any oscillation. Let us consider, for instance, the Daubechies wavelet with 2 vanishing moments ψ (see Figure 2.2), and a piecewise constant function $f_a = \mathbb{1}_{\{x \geq a\}}$, $a \in (0, 3)$. The oscillations

of ψ cause zeros in the function $a \mapsto |\langle f_a, \psi \rangle|$. This gives rise to undesirable phenomena. Indeed, as we can observe in Figure 5.1, there exist values of a such that the parent coefficient is zero (blue line), and the child is at its maximum (red line).

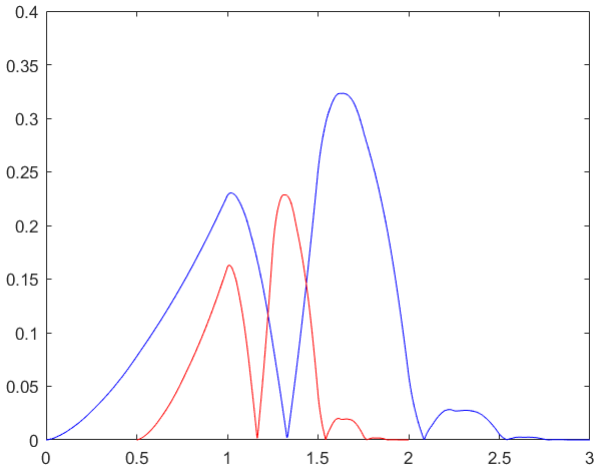


Figure 5.2: Plots of the functions $a \mapsto |\langle f_a, \psi_{0,0} \rangle|$ (blue line), and $a \mapsto |\langle f_a, \psi_{1,1} \rangle|$ (red line).

From this argument, we can conclude that it is reasonable to expect that the relevant coefficients are contained in a tree, but also that they do not form a tree themselves in general. This means that forcing the indices in the approximation to belong to a tree, leads to selecting also negligible coefficients. Following this philosophy, in [Cohen et al., 2001], it is proved that the tree approximation is as efficient as the classical non-linear approximation. There, the strategy is:

- Fix a threshold $\eta > 0$,
- Select the coefficients larger than η,
- Consider the smallest tree containing these coefficients,
- Estimate the error brought in by approximating only with the coefficients contained in the tree.

Another approach is to suppose that, if a wavelet does not intersect a singularity, then the associated coefficient has to be negligible, and so all its descendants. In this case, since it is studied the structure of the negligible coefficients, we speak of *zero-trees*. Based on this idea, in [Shapiro, 1993], an efficient coding algorithm for images is presented .

Let us now switch to the shearlets case. The multiresolution and geometric properties of the shearlets suggest that an analogous argument could be used in order to get a better understanding of the set of the relevant indices also in this case. One of the works that delves deeper into the tree structure in the case of shearlets is [Grohs, 2012]. The strategy used follows that used in [Cohen et al., 2001], where the tree is formed by adding all the missing parents to the set of the relevant coefficients. This strategy has the advantage of leading to an optimal error for any cartoon-like image, but it has some limitations that we intend to delve into further. Let us observe Figure 5.3, that shows the parent-child relation considered.

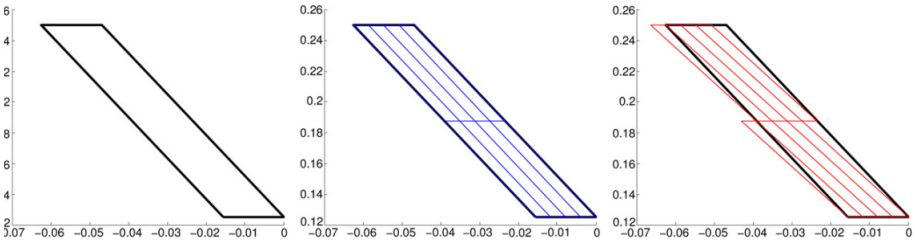


Figure 5.3: Tree structure on a shearlet frame: the shearlet essentially supported in the black parallelogram in the left is the parent of those essentially supported on the blue and red smaller parallelograms. [Grohs, 2012].

With this definition we have that if the edge curve overlaps some of the children, then it overlaps also the parent. This makes the structure defined extremely reasonable, but the argument, used by the author to obtain the approximation result, does not involve it in any way. The only property used is the uniqueness of the parent for every index. In this regard, we want to introduce a method that, through a more quantitative analysis of the coefficients, leads more naturally to the construction of a tree.

In the following, we work with the shearlet frame generated by a shearlet of the form (4.9). We recall that it is C^∞, and compactly supported in the frequency domain. Therefore, in the spatial domain, it cannot be compactly supported, but it has fast decay. This allows us to suppose that the shearlet ψ is *essentially supported* within a unit square. In other words, the decay of the shearlet outside of this square will depend on the particular choice of ψ, and, in general, it is not possible to claim that, outside of the square, the shearlet is very close to 0. For instance, if ψ is similar to a gaussian with a small variance, then the approximation made above is good; if the variance is larger, then this approximation is worse. Since we want to analyze the tree approximation for a general ψ, we have to consider the worst case scenario, i.e. the second one. In this case, when analyzing the magnitude of the coefficients, we have to keep in mind that there can be overlaps between shearlets essentially supported in adjacent regions of the space, see Figure 5.4. Hence, if the edge curve of a cartoon-like

image overlaps the essential support of a shearlet, then it is not obvious that the adjacent shearlets have to be insignificant.

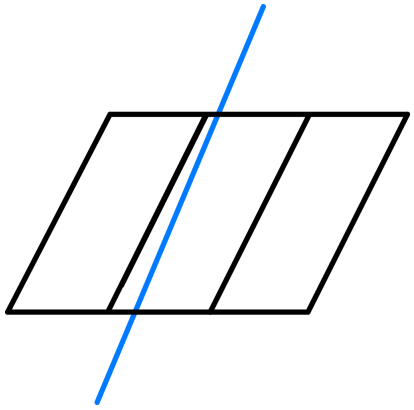


Figure 5.4: The black parallelograms represent the essential supports of adjacent shearlets. The blue line represents the edge curve of a cartoon-like image.

In the next sections, we consider the cone-adapted shearlet frame introduced in Chapter 4. In Section 5.2, we define the tree structure we will consider in the following. It is worth to observe that we will define parent-child relations only on the shearlet components, without considering the scaling components. This is motivated by the fact that there is no relation between the scaling coefficients and the shearlets coefficients at scale zero, hence it is not interesting to define a tree structure on this set of indices. Moreover, we will introduce the structure only on the horizontal cones. Because of the symmetry between horizontal and vertical cones, the relations on the vertical ones will be analogous. In Section 5.3, we present the main results of the thesis, concerning the construction of a tree which allows optimal approximation, motivated by a quantitative analysis of the coefficients. In the last section, we provide all the proofs needed.

5.2 Tree Structure

In this section, we introduce and motivate the tree structure we will use in this chapter. It is worth noting that it is the same as defined in [Grohs, 2012], the difference being that in these results, it is obtained after a quantitative analysis of the coefficients. Let us consider the shearlet Parseval frame introduced in Chapter 4, and consider the horizontal cones component

$$\{\psi_{j,l,k} = 2^{\frac{3}{2}j}\psi(S^l A^j \cdot -k) : j \geq 0, -2^j \leq l \leq 2^j, k \in \mathbb{Z}^2\}.$$

We observe that, by construction of ψ,

$$\operatorname{supp}(\mathcal{F}\psi_{j,l,k}) \subseteq \{(\xi_1, \xi_2) : \xi_1 \in [-2^{2j-1}, -2^{2j-4}] \cup [2^{2j-4}, 2^{2j-1}], |2^j \frac{\xi_2}{\xi_1} - l| \leq 1\}.$$

This implies that, in the frequency domain, the supports of the shearlets are contained within the portion of plane contained between two lines of slope $\frac{l-1}{2^j}$, and $\frac{l+1}{2^j}$. Since $l \in \{-2^j, \dots, 2^j\}$, at each scale j, the slopes of these lines constitute a partition of the interval $\left[-1-\frac{1}{2^j}, 1+\frac{1}{2^j}\right]$. Therefore, if at scale j, the edge curve is contained between the lines of slope $\frac{l-1}{2^j}$ and $\frac{l+1}{2^j}$, i.e the relevant indices are found at shearing parameter l, then, at scale $j+1$, the edge curve is contained either between the lines with slopes $\frac{2l-2}{2^{j+1}}$ and $\frac{2l}{2^{j+1}}$, between those of slopes $\frac{2l-1}{2^{j+1}}$ and $\frac{2l+1}{2^{j+1}}$, or between those with slopes $\frac{2l}{2^{j+1}}$ and $\frac{2l+2}{2^{j+1}}$. This means that, at scale $j+1$, the relevant indices are found at shearing parameter $l' \in \{2l-1, 2l, 2l+1\}$. In order to avoid that an index has more than one parent, we select the shearing parameters $2l$, $2l+1$ as children of l. Observe that it is true that if the edge curve overlaps the children supports, then it necessarily overlaps also the parent support.

Let us now describe the relations between the position indices $k = (k_1, k_2)$. Following the idea explained in Section 5.1, let us reason in an ideal way for a moment, and suppose that the shearlet ψ is supported within the unit square $[0,1]^2$. If this is the case, then the shearlets $\psi_{j,l,k}$ are supported within the region

$$\{(x_1, x_2) : 0 \leq 4^j x_1 + 2^j l x_2 - k_1 \leq 1, 0 \leq 2^j x_2 - k_2 \leq 1\}.$$

Therefore, we can observe that, at scale j, along the x_2-axis, the unit interval $[0,1]$ is divided into dyadic intervals $\left[\frac{k_2}{2^j}, \frac{k_2+1}{2^j}\right]$ of amplitude 2^{-j}. Each of these intervals, at scale $j+1$, is divided into two intervals of length 2^{-j-1}. Hence, it is natural to consider as children of the sherlets supported within the horizontal stripes $\frac{k_2}{2^j} \leq x_2 \leq \frac{k_2+1}{2^j}$, those supported within the stripes $\frac{k_2}{2^j} \leq x_2 \leq \frac{2k_2+1}{2^{j+1}}$, and those in $\frac{2k_2+1}{2^{j+1}} \leq x_2 \leq \frac{k_2+1}{2^j}$, which correspond to the position indices $2k_2, 2k_2+1$. Let us now describe the relations between the parameters k_1. Along the x_1-axis, the supports are limited by the lines $\frac{k_1}{4^j} - \frac{l}{2^j}x_2 \leq x_1 \leq \frac{k_1+1}{4^j} - \frac{l}{2^j}x_2$. This implies that the relations between these parameters will depend also on the shearing parameter l. If the shearing parameter at scale $j+1$ is the double of that at scale j, then the lines which limit the supports are parallel, hence, it is sufficient just to scale k_1 with the parabolic scaling to obtain the positions $\{4k_1 + n : n = 0, 1, 2, 3\}$. Whenever the shearing parameter is $2l+1$, then the lines limiting the supports at scale $j+1$ have slopes $\frac{l}{2^j} + \frac{1}{2^{j+1}}$. We can observe that they are slightly rotated with respect to the lines limiting the supports at scale j. Therefore, here the idea is to apply a shearing operation also to the position parameters, namely

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} k_1 \\ k_2 \end{pmatrix} = \begin{pmatrix} 4k_1 + 2k_2 \\ * \end{pmatrix}.$$

This argument leads to expecting that, in the case of shearing parameter equal to $2l+1$, the children are located within $\{4k_1 + 2k_2 + n : n = 0, 1, 2, 3\}$.

These ideas give rise to the following parent-child definition.

Definition 5.1. An index $(j, l, (k_1, k_2))$ is said to be a child of $(j', l', (k_1', k_2'))$ if

- $j = j' + 1$,
- $l \in \{2l', 2l' + 1\}$,
- $k_2 \in \{2k_2', 2k_2' + 1\}$,
- $k_1 \in \begin{cases} \{4k_1' + n : n = 0, 1, 2, 3\} & l \text{ even} \\ \{4k_1' + 2k_2' + n : n = 0, 1, 2, 3\} & l \text{ odd} \end{cases}$

We will write $\lambda \prec \lambda'$, whenever the index λ is a descendant of λ'.

Definition 5.2. A set of indices T is said to be a tree if

$$\lambda \prec \lambda', \lambda \in T \implies \lambda' \in T.$$

By the previous definition, for each index we have up to two families of children, each one composed by 8 elements. Hence, each index has up to 16 children, and, for construction, every index at scale $j \geq 1$ has exactly one parent at scale $j - 1$.

Moreover, we can obtain the explicit expression of the unique parent of $(j, l, (k_1, k_2))$ depending on whether l is even or odd:

$$\begin{aligned} &(j-1, \frac{l}{2}, (\lfloor \frac{k_1}{4} \rfloor, \lfloor \frac{k_2}{2} \rfloor)) && \text{if } l \text{ is even,} \\ &(j-1, \frac{l-1}{2}, (\lfloor \frac{k_1 - 2\lfloor \frac{k_2}{2} \rfloor}{4} \rfloor, \lfloor \frac{k_2}{2} \rfloor)) && \text{if } l \text{ is odd.} \end{aligned}$$

5.3 Main Results

In this section, we present the main results of the thesis. In the following, we are going to analyze the structure of the relevant shearlet coefficients in the case of a piecewise constant function, which has an edge curve coinciding with a straight line. Unlike the approach in [Grohs, 2012], described in Section 5.1, our method involves searching for relevant coefficients by excluding those that are certainly insignificant.

Consider the functions $f = \frac{1}{2}\mathbb{1}_{\{x \geq 0\}} - \frac{1}{2}\mathbb{1}_{\{x < 0\}}$, $g = 1$, and define the function $G(x_1, x_2) = f(x_1)g(x_2)$. Fix an angle $\theta \in \left(\frac{\pi}{4}, \frac{3}{4}\pi\right)$, and consider the rotation matrix

$$P = \begin{pmatrix} \sin(\theta) & -\cos(\theta) \\ \cos(\theta) & \sin(\theta) \end{pmatrix}.$$

The function we will analyze throughout this section is $F(x_1, x_2) = G(P\binom{x_1}{x_2})$. Observe that its edge curve is distributed along the line $\sin(\theta)x_1 - \cos(\theta)x_2 = 0$, see Figure 5.5. The choice of this function was made to enable an explicit analysis of the coefficients and to facilitate the identification of relevant indices. Furthermore, it can be viewed as the fine-scale case of a generic cartoon-like image.

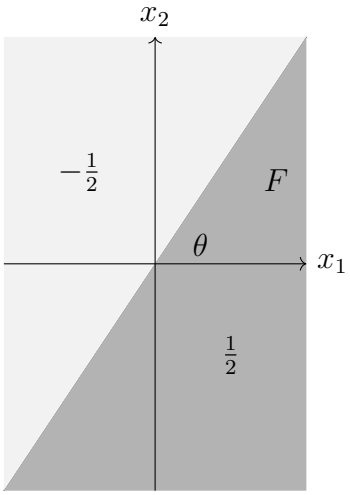


Figure 5.5: The function F.

First of all, we aim to discard all the shearing parameters associated with negligible coefficients. In this particular case, given the simple expression of the function F, we will conclude that there are at most two shearing parameters at each scale which can be associated to non-zero coefficients. Since we are working with band limited shearlets, and the shearing parameters play an important role in delimiting the shearlets supports, we analyze the scalar product at fixed scale j in the frequency domain. Notice that, since $F \notin L^2(\mathbb{R}^2)$, we interpret the Fourier transform in the distributional sense. Let us compute the Fourier transform of F. We start from the Fourier transform of the Heaviside function $h = \mathbb{1}_{\{x \geq 0\}}$, which is

$$\mathcal{F}h(\xi) = \pi\delta(\xi) - i\,\mathrm{P.\,V.}\left(\frac{1}{\xi}\right),$$

where $\mathrm{P.\,V.}\left(\frac{1}{\xi}\right)$ is the principal value of $\frac{1}{\xi}$, i.e. the distribution acting on a test function φ as

$$\langle \mathrm{P.\,V.}\left(\frac{1}{\xi}\right), \varphi\rangle = \lim_{\epsilon \to 0^+} \int_{|\xi| \geq \epsilon} \frac{\varphi(\xi)}{\xi} d\xi.$$

Since

$$f = \frac{1}{2}h - \frac{1}{2}h^{\vee},$$
$$\mathcal{F}(h^{\vee}) = (\mathcal{F}h)^{\vee},$$

where $h^{\vee}(x) = h(-x)$, we can compute the Fourier transform of the function f

$$\mathcal{F}f(\xi) = -i\,\mathrm{P.\,V.}\left(\frac{1}{\xi}\right).$$

Moreover, it is easy to see that $\mathcal{F}g(\xi) = 2\pi\delta(\xi)$. Therefore, since $G(x_1, x_2) = f(x_1)g(x_2)$, we have that

$$\mathcal{F}G(\xi_1, \xi_2) = \mathcal{F}f(\xi_1)\mathcal{F}g(\xi_2) = -2\pi i \,\mathrm{P.V.}\left(\frac{1}{\xi_1}\right)\delta(\xi_2). \tag{5.1}$$

Observe that G has discontinuity along the line $x_1 = 0$, and its Fourier transform is distributed along the orthogonal direction $\xi_2 = 0$. Analogously, given that $F = G(P\cdot)$, it holds that $\mathcal{F}F = \mathcal{F}G(P\cdot)$, hence the Fourier transform of the function F is distributed along the line $\xi_1\cos(\theta) + \xi_2\sin(\theta) = 0$, which is orthogonal to the discontinuity line of F, $x_1\sin(\theta) - x_2\cos(\theta) = 0$. Since $\theta \in \left(\frac{\pi}{4}, \frac{3}{4}\pi\right)$, we have that the line $\xi_1\cos(\theta) + \xi_2\sin(\theta) = 0$ is completely contained in the horizontal cones, hence we will analyze the decomposition of F with respect to the Parseval frame (4.12). In the next lemma, we analyze the coefficients with the scale parameter j fixed, and we show that most of the coefficients are definitely zero.

Lemma 5.3. *Let $j \geq 0$, $l_j^1 = \lfloor -2^j \operatorname{cotan}(\theta)\rfloor$, $l_j^2 = \lceil -2^j \operatorname{cotan}(\theta)\rceil$, $L_j = \{l_j^1, l_j^2\}$, and $k \in \mathbb{Z}^2$. Then, for every $l \notin L_j$,*

$$\langle F, \psi_{j,l,k}\rangle = 0.$$

Due to this lemma, we can discard from the set of the possible indices all the indices corresponding to the scale j, and to a shearing parameter $l \notin L_j$.

Now, we are interested in finding the right position parameters k_1, k_2. Given that the edge curve has infinite length, we have infinitely many shearlets that lie on the line. Therefore, to study the error, we need to restrict the set of the position indices. Following the idea explained in the previous section, by supposing that the shearlets are essentially supported within an unit square, we have that at scale j the x_2-axis is divided into horizontal stripes of amplitude 2^{-j}. Hence, we consider only the values $(k_1, k_2) \in \mathbb{Z} \times \{0, \dots, 2^j - 1\}$. In the following, we fix $j \geq 0$, $l \in L_j$, and $k_2 \in \{0, \dots, 2^j - 1\}$. The next lemma shows the decay for k_1 varying in $\mathbb{Z}$. Intuitively, we are studying the coefficients associated to the shearlets lying within the same horizontal stripe.

Lemma 5.4. *Let $j \geq 0$, $l \in L_j$, and $k_2 \in \{0, \dots, 2^j - 1\}$. Then,*

$$|\langle F, \psi_{j,l,k}\rangle| \leq \frac{C2^{-\frac{3}{2}j}}{|k_1 - k_2(2^j \operatorname{cotan}(\theta) + l)|^2}, \qquad k_1 \in \mathbb{Z}. \tag{5.2}$$

This lemma suggests that the largest amplitude values on each stripe are those for $k_1 \sim k_2(2^j \operatorname{cotan}(\theta) + l)$. Now, we need to decide how many indices k_1 to keep. In order to do this, we fix a threshold $\eta > 0$. The idea, here, is to discard the indices with amplitude that is certainly smaller than or equal to η. Let us study when the upper bound in (5.2) is larger than η.

$$\frac{C2^{-\frac{3}{2}j}}{|k_1 - k_2(2^j \operatorname{cotan}(\theta) + l)|^2} > \eta \iff |k_1 - k_2(2^j \operatorname{cotan}(\theta) + l)| < \frac{C2^{-\frac{3}{4}j}}{\sqrt{\eta}}.$$

Therefore, for every $j \geq 0$, $l \in L_j$, $k_2 \in \{0, \dots, 2^j - 1\}$, and for a fixed threshold $\eta > 0$, we choose the indices $k_1 \in \mathbb{Z}$ such that

$$k_1 \in A_{j,l,k_2}(\eta) = \left\{ k \in \mathbb{Z} : |k - k_2(2^j \operatorname{cotan}(\theta) + l)| < \frac{C 2^{-\frac{3}{4}j}}{\sqrt{\eta}} \right\}. \tag{5.3}$$

Observe that $|A_{j,l,k_2}(\eta)| \sim \frac{2^{-\frac{3}{4}j}}{\sqrt{\eta}}$. Notice that the number of k_1 we select at each scales depends on j, but it is uniformly bounded by $\frac{C}{\sqrt{\eta}}$. This ensures that we are not selecting too many indices. So far, we have constructed a set that, for every resolution scale parameter j, discards all those coefficients that are certainly smaller than η. Since the amplitude of the coefficients has to decay along the scales, we use the threshold η, in order to discard the scales that only contain small coefficients. In other words, due to Hölder inequality and a simple change of variable, we have

$$|\langle F, \psi_{j,l,k} \rangle| \leq \|F\|_\infty \|\psi_{j,l,k}\|_1 \leq C' 2^{-\frac{3}{2}j},$$

and

$$C' 2^{-\frac{3}{2}j} > \eta \iff j < \frac{2}{3} \log_2 \left(\frac{C'}{\eta} \right) = j_\eta.$$

This means that, for $j \geq j_\eta$, every coefficient has small amplitude with respect to η, hence we discard all the coefficients for $j \geq j_\eta$. To summarize, we have constructed the set of the most relevant indices $T(\eta)$ as

$$T(\eta) = \bigcup_{j=0}^{\lfloor j_\eta \rfloor} T_j(\eta),$$

where

$$\lambda = (j, l, (k_1, k_2)) \in T_j(\eta) \iff l \in L_j, k_2 \in \{0, \dots, 2^j - 1\}, k_1 \in A_{j,l,k_2}(\eta). \tag{5.4}$$

The next lemma shows that the set $T(\eta)$ is a tree, and provides an upper bound on its cardinality.

Theorem 5.5. *For every $\eta > 0$, $T(\eta)$ is a tree, and $|T(\eta)| \lesssim \eta^{-\frac{2}{3}}$.*

Our target, now, is to show that the error committed by selecting only the indices in $T(\eta)$ is optimal (see Theorem 4.15).

Since we are limiting the indices k_2, we are not approximating exactly F, but its projection over the space generated by

$$\{\psi_{j,l,k} : j \geq 0, -2^j \leq l \leq 2^j, 0 \leq k_2 < 2^j, k_1 \in \mathbb{Z}\}.$$

Let us denote by $\mathcal{P}F$ this approximation, and consider the approximating function

$$\mathcal{S}(F,\eta) = \sum_{\lambda \in T(\eta)} \langle F, \psi_\lambda \rangle \psi_\lambda.$$

The error committed is

$$\|\mathcal{P}F - \mathcal{S}(F,\eta)\|_2^2 \lesssim \sum_{\lambda \notin T(\eta)} |\langle F, \psi_\lambda \rangle|^2.$$

The next proposition shows that the error decaying is optimal.

Theorem 5.6. *If $|T(\eta)| \leq N$, then*

$$\|\mathcal{P}F - \mathcal{S}(F,\eta)\|_2^2 \lesssim N^{-2}.$$

5.4 Proofs

Proof of Lemma 5.3. First of all let us observe that, as a distribution, $\mathcal{F}F$ acts on a test function φ as

$$\langle \mathcal{F}F, \varphi \rangle = \langle \mathcal{F}G, \varphi(P^T \cdot) \rangle.$$

Therefore,

$$\langle F, \psi_{j,l,k} \rangle = \frac{1}{2\pi} \langle \mathcal{F}F, \mathcal{F}\psi_{j,l,k} \rangle = \frac{1}{2\pi} \langle \mathcal{F}G, \mathcal{F}\psi_{j,l,k}(P^T \cdot) \rangle.$$

From (4.8), and (4.9), we obtain that

$$\begin{aligned} &\mathcal{F}\psi_{j,l,k}\left(P^T \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}\right) \\ &= 2^{-\frac{3}{2}j} \mathcal{F}\psi_1(4^{-j}(\xi_1 \sin(\theta) + \xi_2 \cos(\theta)) \mathcal{F}\psi_2\left(2^j \frac{\xi_2 \sin(\theta) - \xi_1 \cos(\theta)}{\xi_1 \sin(\theta) + \xi_2 \sin(\theta)} - l\right) \mathrm{e}^{-i\alpha(\xi_1,\xi_2)}, \end{aligned}$$

where α will depend on all the parameters. Consequently, using the expression of $\mathcal{F}G$ given in (5.1), we obtain

$$\langle F, \psi_{j,l,k} \rangle = -i 2^{-\frac{3}{2}j} \mathcal{F}\psi_2(-2^j \operatorname{cotan}(\theta) - l) \lim_{\epsilon \to 0^+} \int_{|\xi_1| \geq \epsilon} \frac{\mathcal{F}\psi_1(4^{-j}\xi_1 \sin(\theta))}{\xi_1} \mathrm{e}^{-i\alpha(\xi_1,0)} \, d\xi_1.$$

Now, we recall that $\mathcal{F}\psi_2$ is C^∞, and compactly supported within $[-1,1]$, and we observe that

$$|-2^j \operatorname{cotan}(\theta) - l| < 1 \iff l \in L_j.$$

Therefore, we can conclude that if $l \notin L_j$, then $\langle F, \psi_{j,l,k} \rangle = 0$. □

Proof of Lemma 5.4. Let us notice that, by construction, $\mathcal{F}\psi(0,0) = 0$, so the shearlet ψ has zero average. Moreover, we recall that ψ has fast decay in the spatial domain. The idea, here, is that when k_1 is sufficiently small, then ψ is mostly concentrated on a region where $F = -\frac{1}{2}$. Therefore, we use the zero average property, and we add $\frac{1}{2}$ to F. In this way, we have that the shearlet coefficient coincides with the integral of ψ on the other side of the edge line, where the shearlet has to be close to 0. Hence, we expect that small values of k_1 generate small amplitude coefficients. We will argue in the same manner for k_1 sufficiently large.

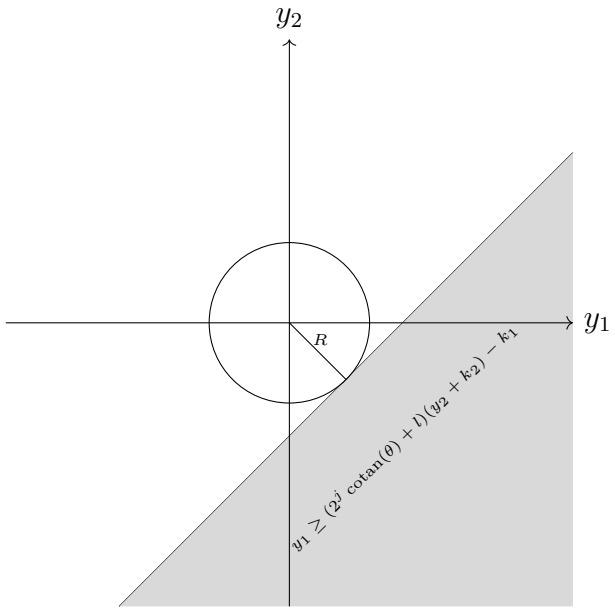
$$\begin{aligned}|\langle F, \psi_{j,l,k}\rangle| &= |\langle F + \frac{1}{2}, \psi_{j,l,k}\rangle| \\ &\leq 2^{\frac{3}{2}j} \int_{-\infty}^{+\infty} \int_{\operatorname{cotan}(\theta)x_2}^{+\infty} \left|\psi\left(S^l A^j \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} k_1 \\ k_2 \end{pmatrix}\right)\right| dx_1 dx_2.\end{aligned}$$

Due to the change of variable $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = S^l A^j \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} k_1 \\ k_2 \end{pmatrix}$, we obtain

$$|\langle F, \psi_{j,l,k}\rangle| \leq 2^{-\frac{3}{2}j} \int_{-\infty}^{+\infty} \int_{(2^j \operatorname{cotan}(\theta)+l)(y_2+k_2)-k_1}^{+\infty} |\psi(y_1, y_2)| dy_1 dy_2.$$

Let us suppose $k_1 < k_2(2^j \operatorname{cotan}(\theta) + l)$. Consider the ball centered at the origin of radius R, B_R, tangent to the line $y_1 = (2^j \operatorname{cotan}(\theta) + l)(y_2 + k_2) - k_1$. Since $k_1 < k_2(2^j \operatorname{cotan}(\theta) + l)$, we have that

$$\{(y_1, y_2) \in \mathbb{R}^2 : y_1 \geq (2^j \operatorname{cotan}(\theta) + l)(y_2 + k_2) - k_1\} \subset \mathbb{R}^2 \setminus B_R = B_R^C.$$



Moreover, since ψ is fast decaying, we have that, for every $N \in \mathbb{N}$, there exists a constant $C_N > 0$ such that

$$|\psi(y_1, y_2)| \leq \frac{C_N}{1 + \|y\|_2^N}, \qquad y = (y_1, y_2) \in \mathbb{R}^2.$$

Therefore, we have

$$|\langle F, \psi_{j,l,k} \rangle| \lesssim C_N 2^{-\frac{3}{2}j} \int_R^{+\infty} \frac{\rho}{1 + \rho^N} d\rho \leq \frac{C_N}{N-2} 2^{-\frac{3}{2}j} R^{2-N}, \qquad N > 2.$$

We can obtain the same estimate for $k_1 > k_2(2^j \operatorname{cotan}(\theta) + l)$, indeed, subtracting rather than adding $\frac{1}{2}$ to F leads to the symmetric case that can be analyzed in the same way.

$$\begin{aligned}
|\langle F, \psi_{j,l,k} \rangle| = |\langle F - \frac{1}{2}, \psi_{j,l,k} \rangle| &\leq 2^{\frac{3}{2}j} \int_{-\infty}^{+\infty} \int_{-\infty}^{\operatorname{cotan}(\theta) x_2} \left| \psi \left(S^l A^j \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} k_1 \\ k_2 \end{pmatrix} \right) \right| dx_1 dx_2 \\
&\leq 2^{-\frac{3}{2}j} \int_{-\infty}^{+\infty} \int_{-\infty}^{(2^j \operatorname{cotan}(\theta) + l)(y_2 + k_2) - k_1} |\psi(y_1, y_2)| dy_1 dy_2 \lesssim C_N 2^{-\frac{3}{2}j} \int_R^{+\infty} \frac{\rho}{1 + \rho^N} d\rho \\
&\leq \frac{C_N}{N-2} 2^{-\frac{3}{2}j} R^{2-N}, \qquad N > 2.
\end{aligned}$$

Observing that $l \in L_j$, we have

$$R = \frac{|(2^j \operatorname{cotan}(\theta) + l)k_2 - k_1|}{\sqrt{1 + (l + 2^j \operatorname{cotan}(\theta))^2}} \geq \frac{1}{\sqrt{2}} |(2^j \operatorname{cotan}(\theta) + l)k_2 - k_1|.$$

Therefore, we have

$$|\langle F, \psi_{j,l,k} \rangle| \leq \frac{C'_N 2^{-\frac{3}{2}j}}{|(2^j \operatorname{cotan}(\theta) + l)k_2 - k_1|^{N-2}}.$$

The choice $N = 4$ concludes the proof, but we note that to prove Theorem 5.5, one could select any $N > 3$. □

Proof of Theorem 5.5. We start by showing the estimate on the cardinality of $T(\eta)$. By construction of $T_j(\eta)$ given in (5.4), we can observe that

$$|T_j(\eta)| \lesssim \frac{2^{\frac{j}{4}}}{\sqrt{\eta}}.$$

Therefore, by definition of $T(\eta)$,

$$|T(\eta)| \lesssim \frac{1}{\sqrt{\eta}} \sum_{j=0}^{\lfloor j_\eta \rfloor} 2^{\frac{j}{4}} \lesssim \frac{1}{\sqrt{\eta}} 2^{\frac{j_\eta}{4}} \lesssim \eta^{-\frac{2}{3}}. \tag{5.5}$$

Let us now show that $T(\eta)$ is a tree. We recall that we have an explicit expression of the unique parent of the index $(j,l,(k_1,k_2))$ depending on whether l is even, or odd.

$$\begin{aligned}&\left(j-1,\frac{l}{2},\left(\left\lfloor\frac{k_1}{4}\right\rfloor,\left\lfloor\frac{k_2}{2}\right\rfloor\right)\right) && l \text{ even},\\ &\left(j-1,\frac{l-1}{2},\left(\left\lfloor\frac{k_1-2\left\lfloor\frac{k_2}{2}\right\rfloor}{4}\right\rfloor,\left\lfloor\frac{k_2}{2}\right\rfloor\right)\right) && l \text{ odd}.\end{aligned}$$

Observe that, in general, the parent has shearing parameter $\left\lfloor\frac{l}{2}\right\rfloor$. It is straightforward to see that

$$\begin{aligned}&l\in L_j \implies \left\lfloor\frac{l}{2}\right\rfloor\in L_{j-1},\\ &k_2\in\{0,\dots,2^j-1\} \implies \left\lfloor\frac{k_2}{2}\right\rfloor\in\{0,\dots,2^{j-1}-1\}.\end{aligned}$$

Let us suppose l even. We have to prove that

$$k_1\in A_{j,l,k_2}(\eta) \implies \left\lfloor\frac{k_1}{4}\right\rfloor\in A_{j-1,\frac{l}{2},\left\lfloor\frac{k_2}{2}\right\rfloor}(\eta).$$

Applying the definition of $A_{j,l,k_2}(\eta)$, it follows that

$$-\frac{C2^{-\frac{3}{4}j}}{4\sqrt{\eta}}+(2^{j-1}\operatorname{cotan}(\theta)+\frac{l}{2})\frac{k_2}{2}\le\frac{k_1}{4}\le(2^{j-1}\operatorname{cotan}(\theta)+\frac{l}{2})\frac{k_2}{2}+\frac{C2^{-\frac{3}{4}j}}{4\sqrt{\eta}}.$$

Using $\left\lfloor\frac{k_2}{2}\right\rfloor\le\frac{k_2}{2}\le\left\lfloor\frac{k_2}{2}\right\rfloor+1$, and $2^{j-1}\operatorname{cotan}(\theta)+\frac{l}{2}\le\frac{1}{2}$, we have

$$-\frac{C2^{-\frac{3}{4}j}}{4\sqrt{\eta}}+(2^{j-1}\operatorname{cotan}(\theta)+\frac{l}{2})\left\lfloor\frac{k_2}{2}\right\rfloor\le\frac{k_1}{4}\le(2^{j-1}\operatorname{cotan}(\theta)+\frac{l}{2})\left\lfloor\frac{k_2}{2}\right\rfloor+\frac{C2^{-\frac{3}{4}j}}{4\sqrt{\eta}}+\frac{1}{2}.$$

Now, we use $\left\lfloor\frac{k_1}{4}\right\rfloor\le\frac{k_1}{4}\le\left\lfloor\frac{k_1}{4}\right\rfloor+1$ to obtain

$$\left|\left\lfloor\frac{k_1}{4}\right\rfloor-(2^{j-1}\operatorname{cotan}(\theta)+\frac{l}{2})\left\lfloor\frac{k_2}{2}\right\rfloor\right|\le\frac{C2^{-\frac{3}{4}j}}{4\sqrt{\eta}}+1=2^{-\frac{11}{4}}\frac{C2^{-\frac{3}{4}(j-1)}}{\sqrt{\eta}}+1\le\frac{C2^{-\frac{3}{4}(j-1)}}{\sqrt{\eta}},$$

where the last estimate is valid for $\frac{C2^{-\frac{3}{4}(j-1)}}{\sqrt{\eta}}\ge\frac{2^{\frac{11}{4}}}{2^{\frac{11}{4}}-1}\sim 1.175$. We observe that, since $\frac{C2^{-\frac{3}{4}(j-1)}}{\sqrt{\eta}}\ge\frac{C2^{-\frac{3}{4}j\eta}}{\sqrt{\eta}}=\frac{C}{\sqrt{C'}}$, up to choosing C sufficiently large, the last estimate holds. This concludes the first part of the proof.

Suppose, now, that l is odd. In the same way, we obtain

$$-\frac{C2^{-\frac{3}{4}j}}{4\sqrt{\eta}}+(2^{j-1}\operatorname{cotan}(\theta)+\frac{l}{2})\left\lfloor\frac{k_2}{2}\right\rfloor\le\frac{k_1}{4}\le(2^{j-1}\operatorname{cotan}(\theta)+\frac{l}{2})\left\lfloor\frac{k_2}{2}\right\rfloor+\frac{C2^{-\frac{3}{4}j}}{4\sqrt{\eta}}+\frac{1}{2},$$

which implies

$$-\frac{C2^{-\frac{3}{4}j}}{4\sqrt{\eta}}+(2^{j-1}\operatorname{cotan}(\theta)+\frac{l-1}{2})\left\lfloor\frac{k_2}{2}\right\rfloor\leq\frac{k_1-2\left\lfloor\frac{k_2}{2}\right\rfloor}{4}$$
$$\leq(2^{j-1}\operatorname{cotan}(\theta)+\frac{l-1}{2})\left\lfloor\frac{k_2}{2}\right\rfloor+\frac{C2^{-\frac{3}{4}j}}{4\sqrt{\eta}}+\frac{1}{2}.$$

Now, we can argue as in the other case to conclude the proof. □

Proof of Theorem 5.6. In order to prove this result, we use the argument presented in the proof of Theorem 4.1 in [Cohen et al., 2001]. Observe that, for construction, if $\eta_1 \leq \eta_2$, then $T(\eta_1) \supseteq T(\eta_2)$. This implies

$$T(2^{-m-1}\eta) \supseteq T(2^{-m}\eta), \qquad m \geq 1.$$

This, with the fact that

$$\lambda \notin T(2^{-l}\eta) \implies |\langle F, \psi_\lambda\rangle| \leq 2^{-l}\eta,$$

and Proposition 5.5, implies that

$$\|\mathcal{P}F - \mathcal{S}(F,\eta)\|_2^2 \lesssim \sum_{\lambda\notin T(\eta)} |\langle F,\psi_\lambda\rangle|^2 \leq \sum_{m=0}^{+\infty} \sum_{\lambda\in T(2^{-m-1}\eta)\setminus T(2^{-m}\eta)} |\langle F,\psi_\lambda\rangle|^2$$
$$\leq \eta^2 \sum_{m=0}^{+\infty} 2^{-2m}|T(2^{-m-1}\eta)| \leq \eta^{\frac{4}{3}} \sum_{m=0}^{+\infty} 2^{\frac{-4m+2}{3}} \lesssim \eta^{\frac{4}{3}}.$$

By considering N a natural number such that $N \sim \eta^{-\frac{2}{3}}$, we have that $|T(\eta)| \lesssim N$. This allows to conclude the proof

$$\|\mathcal{P}F - \mathcal{S}(F,\eta)\|_2^2 \lesssim N^{-2}.$$

□

Appendix

A Lipschitz Regularity

In Section 3.1, we see that the regularity of a function is useful to characterize the non-linear error decay. In order to formalize this, we need to extend the concept of regularity to non-integers. The starting point to do this is the Taylor formula. Let $f\colon \mathbb{R} \longrightarrow \mathbb{C}$ be a function, and suppose that it is m times differentiable in a neighborhood I of a certain point $x \in \mathbb{R}$. Let us consider the Taylor polynomial of degree $m-1$ at x

$$P_x(t) = \sum_{k=0}^{m-1} \frac{1}{k!} \frac{d^k f}{dt^k}(x)(t-x)^k.$$

The Lagrange error bound of a Taylor polynomial gives that

$$|f(t) - P_x(t)| \leq \frac{|t-x|^m}{m!} \sup_{s \in I} \left| \frac{d^m f}{ds^m}(s) \right|, \qquad t \in I.$$

The Lipschitz regularity generalizes this inequality to non-integer exponents.

Definition A.1. Let $f\colon \mathbb{R} \longrightarrow \mathbb{C}$ be a function.

- f is said to be Lipschitz $\alpha \geq 0$ at $x \in \mathbb{R}$ if there exist a constant $K > 0$, and a polynomial P_x of degree $m = \lfloor \alpha \rfloor$ such that

$$|f(t) - P_x(t)| \leq K|t-x|^\alpha, \qquad t \in \mathbb{R};$$

- f is said to be uniformly Lipschitz $\alpha \geq 0$ over $[a,b] \subset \mathbb{R}$ if it is Lipschitz $\alpha \geq 0$ at every $x \in [a,b]$, with a constant K which does not depend on x;

- The Lipschitz regularity of f at x or over $[a,b]$ is the supremum of the α such that f is Lipschitz α.

The polynomial P_x is uniquely defined, indeed if f is $m = \lfloor \alpha \rfloor$ differentiable in a neighborhood of x, then P_x is the Taylor polynomial of f at x. In particular,

if $0 \leq \alpha < 1$, then $P_x(t) = f(x)$, and the Lipschitz condition becomes the Hölder condition with exponent α

$$|f(t) - f(x)| \leq K|t - x|^\alpha, \qquad t \in \mathbb{R}.$$

A function f that is bounded, and discontinuous at x is Lipschitz 0 at x. If $\alpha < 1$, then f is not differentiable at x, and the exponent α characterizes the singularity type. If f is uniformly Lipschitz $\alpha > m$ in a neighborhood of x, then one can prove that f is necessarily m times continuously differentiable in the neighborhood of x.

B Frame Theory

When studying the decomposition of functions with respect to certain systems of functions, it is sometimes necessary surpassing the concept of orthonormal basis, and considering redundant systems. The concept of *frame*, introduced for the first time in [Duffin and Schaeffer, 1952], often comes into play because it guarantees stability while allowing redundancy. In this section, we recall the main definitions and properties of frames, without providing all the proofs. We refer the interested reader to Chapter 5 of [Mallat, 1999], and to Chapter 8 of [Hernández and Weiss, 1996] for further details.

Definition B.1. Let $\mathcal{H}$ be an Hilbert space. A family of functions $\{\psi_i\}_{i\in I} \subset \mathcal{H}$ is said to be a frame if there exist constants $A, B > 0$ such that, for every $f \in \mathcal{H}$,

$$A\|f\|_{\mathcal{H}}^2 \leq \sum_{i\in I} |\langle f, \psi_i\rangle_{\mathcal{H}}|^2 \leq B\|f\|_{\mathcal{H}}^2. \tag{B.1}$$

If A and B can be chosen with $A = B$, then $\{\psi_i\}_{i\in I}$ is said to be a tight frame. If $A = B = 1$ is possible, then $\{\psi_i\}_{i\in I}$ is a Parseval frame.

We can analyze an element $f \in \mathcal{H}$ through its sequence of frame coefficients $(\langle f, \psi_i\rangle)_{i\in I}$. In this regard, we consider the *analysis operator*

$$T\colon \mathcal{H} \longrightarrow \ell^2(I)$$

defined as $Tf := (\langle f, \psi_i\rangle)_{i\in I}$. Its adjoint

$$T^*\colon \ell^2(I) \longrightarrow \mathcal{H}$$

is the so-called *synthesis operator*, and its analytic expression is $T^*c = \sum_{i\in I} c_i\psi_i$ for $c = (c_i)_{i\in I} \in \ell^2(I)$. The third operator that comes into play is the *frame operator*

$$T^*T\colon \mathcal{H} \longrightarrow \mathcal{H},$$

where, for each $f \in \mathcal{H}$, $T^*Tf = \sum_{i\in I}\langle f, \psi_i\rangle\psi_i$. The frame condition (B.1) can be written as

$$A\|f\|_{\mathcal{H}}^2 \leq \langle T^*Tf, f\rangle_{\mathcal{H}} \leq B\|f\|_{\mathcal{H}}^2,$$

this directly implies that the frame operator T^*T is self-adjoint, positive, and such that $A\,\mathrm{Id}_{\mathcal{H}} \leq T^*T \leq B\,\mathrm{Id}_{\mathcal{H}}$. In particular, if the frame is tight, i.e. $A = B$, we have

$$\langle T^*Tf, f\rangle_{\mathcal{H}} = A\|f\|_{\mathcal{H}}^2.$$

This implies that, in case of a tight frame, $T^*T = A\,\mathrm{Id}$.

The reconstruction of f is calculated through the *pseudo inverse*. Let us observe that the frame property (B.1) guarantees that T is a bounded, injective operator. On the other hand, there is no guarantee on the surjectivity of T. Hence, in general, $\mathrm{Im}\,T^{\perp} \neq \{0\}$. We define the pseudo inverse of T,

$$T^{\dagger}\colon \ell^2(I) \longrightarrow \mathcal{H},$$

as the left inverse of T assuming 0 on the orthogonal complement of $\mathrm{Im}\,T$, namely

$$T^{\dagger}c = 0, \qquad c \in \mathrm{Im}\,T^{\perp}.$$

Theorem B.2. *[Mallat, 1999] The pseudo inverse satisfies*

$$T^{\dagger} = (T^*T)^{-1}T^*.$$

Moreover, it is the left inverse with minimum norm, and

$$\|T^{\dagger}\| = \sup_{c\in\ell^2(I), c\neq 0} \frac{\|T^{\dagger}c\|_{\mathcal{H}}}{\|c\|_{\ell^2(I)}} \leq A^{-\frac{1}{2}}.$$

The pseudo inverse of a frame is related to a dual frame $\{\tilde{\psi}_i\}_{i\in I}$, where

$$\tilde{\psi}_i := (T^*T)^{-1}\psi_i.$$

The next theorems specifies some of its properties.

Theorem B.3. *[Mallat, 1999] The dual frame $\{\tilde{\psi}_i\}_{i\in I}$ satisfies for $f \in \mathcal{H}$*

$$\frac{1}{B}\|f\|_{\mathcal{H}}^2 \leq \sum_{i\in I}|\langle f, \tilde{\psi}_i\rangle_{\mathcal{H}}|^2 \leq \frac{1}{A}\|f\|_{\mathcal{H}}^2, \tag{B.2}$$

and

$$f = T^{\dagger}Tf = \sum_{i\in I}\langle f, \psi_i\rangle_{\mathcal{H}}\tilde{\psi}_i = \sum_{i\in I}\langle f, \tilde{\psi}_i\rangle_{\mathcal{H}}\psi_i. \tag{B.3}$$

If the frame is tight, then $\tilde{\psi}_i = A^{-1}\psi_i$.

In particular, in Sections 4.2, and 4.3 of Chapter 4, we deal with a Parseval frame ($A = B = 1$) of the Hilbet space $L^2(\mathbb{R}^2)$. In this case, due to the previous results, we have that

$$f = \sum_{i \in I} \langle f, \psi_i \rangle \psi_i, \qquad f \in \mathcal{H},$$

and

$$\|f\|_{\mathcal{H}}^2 = \sum_{i \in I} |\langle f, \psi_i \rangle|^2, \qquad f \in \mathcal{H}.$$

C Hausdorff Measure $\mathcal{H}^1$ and the Notion of Length

The Hausdorff measure $\mathcal{H}^1$ is an efficient tool which allows to measure the length of a curve without requiring a parametrization. In the following, we present the definition of $\mathcal{H}^1$ on $\mathbb{R}^n$, we discuss its relation with the notion of length, and finally we discuss how we apply these arguments in sections 3.1, and 4.3. We refer to [Maggi, 2012], and [Evans, 2018] for a more in-depth argumentation.

Given a set $E \subseteq \mathbb{R}^n$, we define the diameter of E as

$$\operatorname{diam}(E) = \sup\{\|x - y\|_2 : x, y \in E\}.$$

Definition C.1. Let $E \subseteq \mathbb{R}^n$, and $0 < \delta \leq +\infty$. Set

$$\mathcal{H}^1_\delta(E) = \inf \left\{ \sum_{i \in I} \operatorname{diam}(E_i) : E \subseteq \bigcup_{i \in I} E_i, \operatorname{diam}(E_i) \leq \delta, |I| \leq \aleph_0 \right\}.$$

The Hausdorff measure $\mathcal{H}^1$ of E is defined as

$$\mathcal{H}^1(E) = \lim_{\delta \to 0^+} \mathcal{H}^1_\delta(E).$$

Remark C.2. The Hausdorff measure $\mathcal{H}^1$ belongs to the family of Hausdorff measures on $\mathbb{R}^n$, $\mathcal{H}^k$, which can be defined for every $0 \leq k < +\infty$. Moreover, it can be proved that the Hausdorff measures are Borel regular measures on $\mathbb{R}^n$.

We are interested in the case $k = 1$, because it is strictly related to the length of a curve. Let us briefly recall the main definitions.

A set $\Gamma \subset \mathbb{R}^n$ is said to be a curve if there exist $a < b$, and a continuous injective function $\gamma \colon [a, b] \longrightarrow \mathbb{R}^n$ such that $\Gamma = \gamma([a, b])$. The function γ is said to be a parametrization of Γ. The length of Γ is defined as

$$\ell(\Gamma) = \sup \left\{ \sum_{i=0}^{N} |\gamma(t_i) - \gamma(t_i)| : a = t_0 < t_1 < \cdots < t_{N-1} < t_N = b, N \in \mathbb{N} \right\}.$$

The next result shows that the length of a curve in $\mathbb{R}^n$ is equal to its Hausdorff measure $\mathcal{H}^1(\Gamma)$.

Theorem C.3. *[Maggi, 2012] Let $\Gamma \subset \mathbb{R}^n$ be a curve. Then*

$$\mathcal{H}^1(\Gamma) = \ell(\Gamma).$$

Let us now consider a closed curve Γ of finite length $L > 0$. Without loss of generality, we can suppose it to be supported inside the unitary square $[0,1]^2$. Since Γ is compact, we can assume the covering $\{E_i\}_{i\in I}$ to be finite, and we set $N(\delta) = |I|$. Choosing $\delta_j \sim 2^{-j}$, due to theorem C.3, we have that

$$\lim_{j\to+\infty} \mathcal{H}^1_{\delta_j} = L.$$

Therefore, in this case, we necessarily have

$$N(\delta_j) \sim L2^j.$$

In particular, since a dyadic square Q of side length 2^{-j} satisfies

$$\operatorname{diam}(Q) \sim 2^{-j}, \qquad Q \in \mathcal{Q}_j,$$

we can apply the previous argumentation to a covering of dyadic squares. This provides us the order of the number of dyadic squares intersecting a given curve Γ.

D Classical Shearlet Construction

In this section, we show a particular construction of ψ_1, and ψ_2 satisfying the properties (4.10), and (4.11) in Section 4.2.

Let us start with the construction of ψ_1. Consider an even function $h \in C^\infty(\mathbb{R})$, supported within $\left(-\frac{1}{6}, \frac{1}{6}\right)$, and satisfying $\int_{\mathbb{R}} h(t)dt = \frac{\pi}{2}$. Now, define a function

$$\Theta(\xi) = \int_{-\infty}^{\xi} h(t)dt,$$

and a smooth bell function

$$b(\xi) := \begin{cases} \sin\left(\Theta\left(|\xi| - \frac{1}{2}\right)\right) & \text{if } \frac{1}{3} \le |\xi| \le \frac{2}{3}, \\ \cos\left(\Theta\left(\frac{|\xi|}{2} - \frac{1}{2}\right)\right) & \text{if } \frac{2}{3} \le |\xi| \le \frac{4}{3}, \\ 0 & \text{otherwise.} \end{cases}$$

From the previous definition, it is not difficult to see that

$$\sum_{j=-1}^{+\infty} b^2(2^{-j}\xi) = 1, \qquad |\xi| \ge \frac{1}{3}.$$

Now, defining $u^2(\xi) := b^2(2\xi) + b^2(\xi)$, it follows that

$$\sum_{j=0}^{+\infty} u^2(2^{-2j}\xi) = \sum_{j=-1}^{+\infty} b^2(2^{-j}\xi) = 1, \qquad |\xi| \geq \frac{1}{3}.$$

Finally, we can define $\mathcal{F}\psi_1(\xi) := u\left(\frac{8}{3}\xi\right)$. Therefore, we have that $\operatorname{supp}\mathcal{F}\psi_1 \subseteq \left[-\frac{1}{2}, -\frac{1}{16}\right] \cup \left[\frac{1}{16}, \frac{1}{2}\right]$, and that equation 4.10 is satisfied.

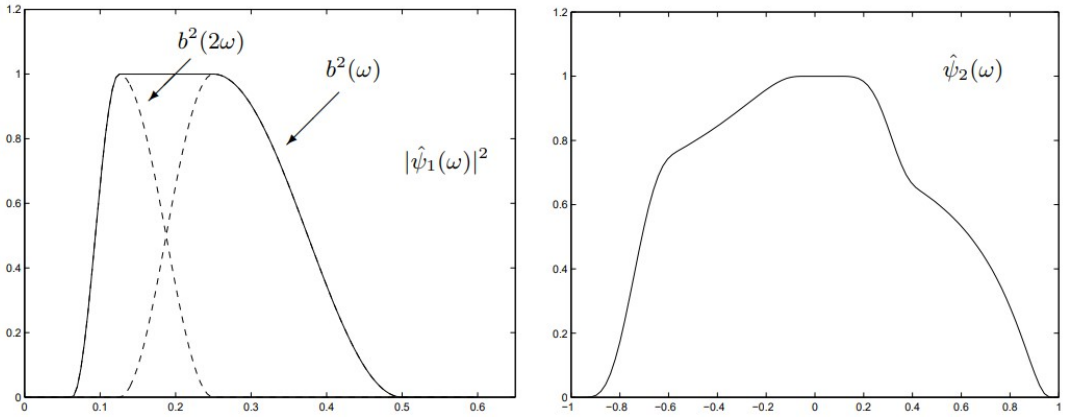


Figure 6: [Guo and Labate, 2007]. On the left, the solid line represent the positive side of the function $|\mathcal{F}\psi_1|^2$, the negative side is symmetrical. On the right, the function $\mathcal{F}\psi_2$.

Let us now discuss the construction of ψ_2. Consider a C^∞ function f_1 compactly supported inside $(-1, 1)$, such that $0 \leq f_1 \leq 1$, and $f_1 = 1$ on $\left[-\frac{1}{2}, \frac{1}{2}\right]$. Now, consider the function $f_2(t) = \sqrt{1 - \mathrm{e}^{\frac{1}{t}}}$, and define $f(t) = f_1(t)f_2(t)$, for $t \in [-1, 0)$. Due to the properties of f_1 and f_2, we can observe that, in the limit sense, $\frac{d^k f}{dt^k}(-1) = 0$ for $k \geq 0$, $f(0) = 1$, and $\frac{d^k f}{dt^k}(0) = 0$ for $k \geq 1$. Since $0 \leq f \leq 1$, we can define the function $g(t) = \sqrt{1 - f^2(t-1)}$ for $t \in (0, 1)$. It is easy to see that g has the same behaviour as f for $t = 0$. Moreover, since $f(t) = f_2(t)$ for $|t| \leq \frac{1}{2}$, then $g(t) = \mathrm{e}^{\frac{1}{2(t-1)}}$ for $t \in \left[\frac{1}{2}, 1\right)$, and $\frac{d^k g}{dt^k}(1) = 0$ for $k \geq 0$ in the left limit sense. Finally, we can define

$$\mathcal{F}\psi_2(\xi) := \begin{cases} f(\xi) & \text{if } \xi \in [-1, 0), \\ g(\xi) & \text{if } \xi \in [0, 1), \\ 0 & \text{otherwise.} \end{cases}$$

By construction, we have $\mathcal{F}\psi_2 \in C^\infty(\mathbb{R})$ with compact support in $[-1, 1]$. Moreover, an easy computation shows that (4.11) holds.

Bibliography

[Adcock and Hansen, 2021] Adcock, B. and Hansen, A. C. (2021). *Compressive imaging: structure, sampling, learning*. Cambridge University Press.

[Candès and Donoho, 2004] Candès, E. J. and Donoho, D. L. (2004). New tight frames of curvelets and optimal representations of objects with piecewise c2 singularities. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 57(2):219–266.

[Cohen et al., 2001] Cohen, A., Dahmen, W., Daubechies, I., and DeVore, R. (2001). Tree approximation and optimal encoding. *Applied and Computational Harmonic Analysis*, 11(2):192–226.

[Cohen et al., 1993] Cohen, A., Daubechies, I., and Vial, P. (1993). Wavelets on the interval and fast wavelet transforms. *Applied and computational harmonic analysis*.

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

[Donoho, 2001] Donoho, D. L. (2001). Sparse components of images and optimal atomic decompositions. *Constructive Approximation*, 17:353–382.

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

[Easley et al., 2008] Easley, G., Labate, D., and Lim, W.-Q. (2008). Sparse directional image representations using the discrete shearlet transform. *Applied and Computational Harmonic Analysis*, 25(1):25–46.

[Evans, 2018] Evans, L. (2018). *Measure theory and fine properties of functions*. Routledge.

[Foucart et al., 2013] Foucart, S., Rauhut, H., Foucart, S., and Rauhut, H. (2013). *An invitation to compressive sensing*. Springer.

[Grafakos et al., 2008] Grafakos, L. et al. (2008). *Classical fourier analysis*, volume 2. Springer.

[Grohs, 2012] Grohs, P. (2012). Tree approximation with anisotropic decompositions. *Applied and Computational Harmonic Analysis*, 33(1):44–57.

[Guo and Labate, 2007] Guo, K. and Labate, D. (2007). Optimally sparse multidimensional representation using shearlets. *SIAM journal on mathematical analysis*, 39(1):298–318.

[Hernández and Weiss, 1996] Hernández, E. and Weiss, G. (1996). *A first course on wavelets*. CRC press.

[Kekkonen et al., 2023] Kekkonen, H., Lassas, M., Saksman, E., and Siltanen, S. (2023). Random tree besov priors - towards fractal imaging. *Inverse Problems and Imaging*, 17(2):507–531.

[Kutyniok and Labate, 2012] Kutyniok, G. and Labate, D. (2012). Introduction to shearlets. *Shearlets: Multiscale analysis for multivariate data*, pages 1–38.

[Labate et al., 2005] Labate, D., Lim, W.-Q., Kutyniok, G., and Weiss, G. (2005). Sparse multidimensional representation using shearlets. In *Wavelets XI*, volume 5914, pages 254–262. SPIE.

[Maggi, 2012] Maggi, F. (2012). *Sets of finite perimeter and geometric variational problems: an introduction to Geometric Measure Theory*. Number 135. Cambridge University Press.

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

[Shapiro, 1993] Shapiro, J. M. (1993). Embedded image coding using zerotrees of wavelet coefficients. *IEEE Transactions on signal processing*, 41(12):3445–3462.