Università degli Studi di Ferrara

Dipartimento di Matematica e Informatica

Shearlets: an overview

Tatiana A. Bubba

Preprint n. 381 – 2014

DIPARTIMENTO DI MATEMATICA E INFORMATICA

VIA MACHIAVELLI, 35 – tel. +39 0532 974002 – fax. +39 0532 974003

FERRARA, ITALY

Shearlets: an overview*

Tatiana Alessandra Bubba
Dept of Mathematics and Computer Science
second site, University of Ferrara
Sci.–Tec. Campus, via Saragat 1, 44122, Ferrara, Italy
bbbtnl@unife.it

February 28, 2014

Abstract

The aim of this report is a self-contained overview on shearlets, a new multiscale method emerged in the last decade to overcome some of the limitation of traditional multiscale methods, like wavelets.
Shearlets are obtained by translating, dilating and shearing a single mother function. Thus, the elements of a shearlet system are distributed not only at various scales and locations – as in classical wavelet theory – but also at various orientations. Thanks to this directional sensitivity property, shearlets are able to capture anisotropic features, like edges, that frequently dominate multidimensional phenomena, and to obtain optimally sparse approximations. Moreover, the simple mathematical structure of shearlets allows for the generalization to higher dimensions and to treat uniformly the continuum and the discrete realms, as well as fast algorithmic implementation.
For all these reasons, shearlets are one of the most successful tool for the efficient representation of multidimensional data and they are being employed in several numerical applications.

Keywords: Shearlets, Affine systems, Frames, Unitary representations of locally compact groups, Sparse approximation, Image processing, Wavelets

2010 MSC: 42C40, 22D10, 42C15, 94A08

*This work is supported by the Italian national research project "FIRB2012 – *Apprendere nel tempo: un nuovo approccio computazionale per l'apprendimento automatico di distemi dinamici*", grant n. RBFR12M3AC, and by the local research projects "FAR2012 – Optimization Methods for Inverse Problems" and "FAR2013 – Optimization methods for Imaging and Learning in dynamic systems" of the University of Ferrara.

Introduction

Nowadays, technology allows for easy acquisition, transmission and storage of huge amounts of data: medical imaging, astronomy, seismology, meteorology, air traffic control, internet traffic, audio and video applications and digital communications, just to mention a few. All this data require efficient analysis and processing in order to extract the relevant information from them. Moreover, it is important not only to provide the methodology to process various different types of data, but also to analyze the accuracy of such methods and to provide a deeper understanding of the underlying structures.

Usually, the first step to this end is to decompose the signal in suitable building blocks which should be well-suited for the specific application and should allow a fast and efficient extraction. This art of "breaking into pieces" to gain insight into an object is exactly the role of (applied) harmonic analysis.

Just to fix the idea, given a class of data $\mathscr{C} \in L^2(\mathbb{R}^d)$, $d \geq 1$, a collection of *analyzing functions* $(\phi_i)_{i\in\mathcal{I}} \subseteq L^2(\mathbb{R}^d)$, where $\mathcal{I}$ is a countable indexing set, is sought such that, for all $f \in \mathscr{C}$, we get the expansion

$$f = \sum_{i\in\mathcal{I}} c_i(f)\phi_i \,.$$

This formula not only provides a decomposition for any element $f \in \mathscr{C}$ into a countable collection of linear measurements $(c_i(f))_{i\in\mathcal{I}} \subseteq \ell^2(\mathcal{I})$, *i.e.*, its *analysis*; but also it illustrates the process of *synthesis*, where f is reconstructed from the expansion coefficients $(c_i(f))_{i\in\mathcal{I}}$.

Hence, one major goal of applied harmonic analysis is the construction of special classes of analyzing elements which can best capture the most relevant information in a certain class of data. Yet, this is the onset of one particular problem which is currently in great demand: directional information and sensitivity. Indeed, it is well known that, due to their limited directional sensitivity, traditional multiscale methods, like wavelets, are not very efficient in dealing with anisotropic features or distributed discontinuities such as the edges occurring in natural images or the boundaries of solid bodies, that however frequently dominate multidimensional phenomena. To address these issues, several variations of the wavelet scheme have been recently proposed, such as directional wavelets [21], complex wavelets [22], ridgelets [2], bandelets [30] and contourlets [12].

The real breakthrough occurred with the introduction of *curvelets* by Candès and Donoho [3]: it is the first system that provides a truly directional multiscale representation of multidimensional data. Roughly, curvelets form a pyramid of analyzing functions defined not only at various scales and locations, like wavelets, but also at various orientations, with the number of orientations increasing at finer scales. This property makes curvelets able to achieve an essentially optimal approximation rate for 2-D smooth functions with discontinuities along $\mathscr{C}^2$-curves. However, there are two main drawbacks: this system is not derived from the action of countably many operators applied to a single (or a finite set) of generating functions and its construction involves rotations that do not preserve the digital lattice, which prevents a direct transition from the continuum to the discrete setting.

At the same time, Guo, Kutyniok, Labate, Lim, and Weiss provided in [18, 29] an alternative approach to curvelets: the *shearlets*. Unlike curvelets, shearlets are derived within the class

of affine systems as a truly multivariate extension of the wavelet framework and they use a shearing parameter to control directional selectivity, in contrast to rotation used by curvelets. These are fundamental different concepts, since they allow shearlet systems to be derived from a single (or a finite set) of generators and they ensure a unified treatment of the continuum and discrete realms due to the fact that the shearing operation preserves integer lattices.
This combination of highly desirable properties make shearlets stand out:

- A single or a finite set of generating functions.
- Optimally sparse approximations of anisotropic features in multivariate data.
- Compactly supported analyzing elements.
- Fast algorithmic implementations.
- A unified treatment of the continuum and digital realms.
- Association with classical approximation spaces.

In the following, we present a brief overview of some key results from the theory and applications of shearlets, focused primarily on the 2D construction, discussing both the continuum and the discrete setting. We start, in Chapter 1, by establish the notation adopted throughout these notes and presenting some background material from harmonic analysis and wavelet theory. In Chapter 2, we give the definition of continuous shearlet transform; in Chapter 3 we discuss discrete shearlet systems and in Chapter 4 we focus on multidimensional extensions, primarily on 3-D continuous shearlet transform.

Chapter 1

Notation and Background

In the following, $\mathbb{R}^+ = \{x \in \mathbb{R} : x > 0\}$ and $\mathbb{R}_0^+ = \{x \in \mathbb{R} : x \geq 0\}$. Vectors in $\mathbb{R}^d$ or $\mathbb{C}^d$, $d \geq 1$, are always assumed to be column vectors, and their inner product shall be denoted by $\langle \cdot, \cdot \rangle$. $L^1(\mathbb{R}^d)$ is the space of Lebesgue integrable functions on $\mathbb{R}^d$ and $L^2(\mathbb{R}^d)$ is the Hilbert space of square Lebesgue integrable functions on $\mathbb{R}^d$ endowed with the inner product $\langle f, g \rangle = \int_{\mathbb{R}^d} \bar{f} g$.

1.1 Fourier Analysis

The Fourier transform is a fundamental tool in harmonic analysis. Since it is discussed in many textbooks, we need to spend only a few words just to establish the notations we use and recalling the properties we need.

Definition 1.1. The *Fourier transform*, denoted by $\mathcal{F}$, is the operator mapping a function $f \in L^1(\mathbb{R}^d)$ into $\mathcal{F}f = \hat{f}$ defined by

$$\hat{f}(\xi) = \int_{\mathbb{R}^d} f(x) \mathrm{e}^{-2\pi i \langle x, \xi \rangle} dx.$$

The *inverse Fourier transform*, denoted by $\mathcal{F}^{-1}$, is the operator mapping a function $g \in L^1(\mathbb{R}^d)$ into $\mathcal{F}^{-1}g = \check{g}$, where

$$\check{g}(x) = \hat{g}(-x) = \int_{\mathbb{R}^d} g(\xi) \mathrm{e}^{2\pi i \langle x, \xi \rangle} d\xi.$$

The function $\hat{f}(\xi)$ is continuous and bounded, since $|\hat{f}(\xi)| < \int |f(x)| dx$; moreover $|\hat{f}(\xi)|$ tends to zero when $|\xi| \to \infty$. We recall also that f is called a *band-limited* function if its Fourier transform is compactly supported.

If $f \in L^1(\mathbb{R}^d)$ with $\hat{f} \in L^1(\mathbb{R}^d)$, we have $f = (\hat{f})^\vee$, hence in this case - which is by far not the only possible case - the inverse Fourier transform is the inverse operator of $\mathcal{F}$ (that is why it is denoted by $\mathcal{F}^{-1}$).
It is well known that this definition can be extended to $L^2(\mathbb{R}^d)$ and, as usual, also these extensions will be denoted by $\hat{f}$ and $\check{g}$. By using this definition of the Fourier transform, the *Plancherel formula* for $f, g \in L^2(\mathbb{R}^d)$ reads

$$\langle f, g \rangle = \langle \hat{f}, \hat{g} \rangle$$

and, in particular,

$$||f||_2 = ||\hat{f}||_2.$$

Among the most important properties of Fourier analysis we recall the following list of results (cf. [15]).

Theorem 1.2. *Let $f, g \in L^1(\mathbb{R}^d)$ and*

$$T_t f(x) = f(x-t), \qquad t \in \mathbb{R}^d \tag{1.1}$$

$$D_M f(x) = |\det(M)|^{-\frac{1}{2}} f(M^{-1}x), \quad M \in GL_d(\mathbb{R}) \tag{1.2}$$

the translation and the dilation operator respectively. The following properties hold true:

(i) $(T_t f)^\wedge(\xi) = \mathrm{e}^{-2\pi i\langle \xi, t\rangle}\hat{f}(\xi)$ *and* $(D_M f)^\wedge(\xi) = D_N \hat{f}(\xi)$, *where* $N = (M^T)^{-1}$.

(ii) $(f * g)^\wedge = \hat{f}\hat{g}$.

(iii) *If* $x^\alpha f \in L^1(\mathbb{R}^d)$ *for* $|\alpha| \le k$, *then* $\partial^\alpha \hat{f} = ((-2\pi i x)^\alpha f)^\wedge$.

(iv) *If* $f \in C^k(\mathbb{R}^d)$, $\partial^\alpha \hat{f} \in L^1(\mathbb{R}^d)$, *for* $|\alpha| \le k$, *and* $\partial^\alpha \hat{f} \in C_0(\mathbb{R}^d)$, *for* $|\alpha| \le k-1$, *then* $(\partial^\alpha \hat{f})^\wedge(\xi) = (2\pi i \xi)^\alpha \hat{f}(\xi))$.

(v) $\mathcal{F}(L^1(\mathbb{R}^d)) \subset C_0(\mathbb{R}^d)$

The following proposition is a simple application of the Fourier transform showing that regularity on $\mathbb{R}^d$ implies decay in the Fourier domain (cf. [15]).

Proposition 1.3. *Suppose that $\psi \in L^2(\mathbb{R}^d)$ is such that $\hat{\psi} \in C_c^\infty(R)$, where $R = \operatorname{supp}(\hat{\psi}) \subset \mathbb{R}^d$. Then, for each $k \in \mathbb{N}$ there is a constant $C_k \in \mathbb{R}^+$ such that, for any $x \in \mathbb{R}^d$, we have*

$$|\psi(x)| \le C_k (1 + |x|^2)^{-k}.$$

In particular, $C_k = k\,\mu(R)(||\hat{\psi}||_\infty + ||\Delta^k \hat{\psi}||_\infty)$, where $\Delta = \sum_{i=1}^d \frac{\partial^2}{\partial \xi_i^2}$ is the frequency domain Laplacian operator and $\mu(R)$ is the Lebesgue measure of R.

We refer to [15] or [32] for more details.

1.2 Frame Theory

When designing representation systems of functions, it is sometimes advantageous or unavoidable to go beyond the setting of orthonormal bases and consider redundant systems. The notion of a *frame* guarantees stability while allowing non-unique decompositions: frames were invented in 1952 by Duffin and Schaeffer, but it took several years before the potential was realized by the scientific community; by now, frame theory is well established.
We will only give a glimpse of the general theory, recalling the basic definitions from frame theory in the setting of a general (real or complex) Hilbert space $\mathscr{H}$. For additional details we refer the reader to [5] and [4].

Definition 1.4. A sequence $(\varphi_i)_{i\in\mathcal{I}}$ in $\mathscr{H}$ is called a *frame* for $\mathscr{H}$, if there exist constants $0 < A \le B < \infty$ such that

$$A||x||^2 \le \sum_{i\in\mathcal{I}} |\langle x, \varphi_i\rangle|^2 \le B||x||^2 \qquad \text{for all } \; x \in \mathscr{H}.$$

The frame constants A and B are called *lower* and *upper frame bound*, respectively. They are not unique. The *optimal upper frame bound* is the infimum over all upper frame bounds, and the *optimal lower frame bound* is the supremum over all lower frame bounds. Note that the optimal bounds actually are frame bounds.

A special role is played by frames for which the optimal frame bounds coincide:

Definition 1.5. A sequence $(\varphi_i)_{i\in\mathcal{I}}$ in $\mathscr{H}$ is a *tight frame* if there exists a constant $A > 0$ such that

$$\sum_{i\in\mathcal{I}} |\langle x, \varphi_i\rangle|^2 = A||x||^2, \qquad \forall x \in \mathscr{H}.$$

If $A = 1$, then $(\varphi_i)_{i\in\mathcal{I}}$ is called *Parseval frame*. A frame is called *equal-norm* if there exists some $c > 0$ such that $||\varphi_i|| = c$ for all $i \in I$, and it is *unit-norm* if $c = 1$.

A simple example of a Parseval frame are three vectors of the same length in $\mathbb{R}^2$ forming a Mercedes-Benz star, which even led to its name 'Mercedes-Benz frame'.

Apart from providing redundant expansions, frames serve as an analysis tool. In fact, if $(\varphi_i)_{i\in\mathcal{I}}$ in $\mathscr{H}$ is a frame for $\mathscr{H}$ it allows the analysis of data through the study of the associated *frame coefficients* $(\langle x, \varphi_i\rangle)_{i\in\mathcal{I}}$, where the operator F defined by

$$F : \mathscr{H} \to \ell^2(\mathcal{I}), \qquad x \mapsto (\langle x, \varphi_i\rangle)_{i\in\mathcal{I}}$$

is called the *analysis operator*. The adjoint F^* of the analysis operator is referred to as the *synthesis operator* and satisfies

$$F^* : \ell^2(\mathcal{I}) \to \mathscr{H}, \qquad (c_i)_{i\in\mathcal{I}} \mapsto \sum_{i\in\mathcal{I}} c_i\varphi_i.$$

The main operator associated with a frame, which provides a stable reconstruction process, is the *frame operator*

$$S = F^*F : \mathscr{H} \to \mathscr{H}, \qquad x \mapsto \sum_{i\in\mathcal{I}} \langle x, \varphi_i\rangle\varphi_i.$$

Note that because $(\varphi_i)_{i\in\mathcal{I}}$ is a Bessel sequence[1], the series defining S converges unconditionally for all $x \in \mathscr{H}$ (cf. corollary 2.4 in [5]). Moreover, S is a positive, self-adjoint invertible operator on $\mathscr{H}$ with $A \cdot I_{\mathscr{H}} \leq S \leq B \cdot I_{\mathscr{H}}$, where $I_{\mathscr{H}}$ denotes the identity operator on $\mathscr{H}$ and A and B are the frame costants. In the case of a Parseval frame, this reduces to $S = I_{\mathscr{H}}$.
In general, $x \in \mathscr{H}$ can be recovered from its *frame coefficients* through the reconstruction formula

$$x = \sum_{i\in\mathcal{I}} \langle x, \varphi_i\rangle\, S^{-1}\varphi_i. \tag{1.3}$$

Notice that one of the cases where this decomposition is very useful is when $x \in \mathscr{H}$ represents a signal.
The sequence $(S^{-1}\varphi_i)_{i\in\mathcal{I}}$, which can be shown to form a frame itself, is referred to as the *canonical dual frame*. Taking a different viewpoint and regarding a frame as a means for expansion in the system $(\varphi_i)_{i\in\mathcal{I}}$, we observe that, for each vector $x \in \mathscr{H}$,

$$x = \sum_{i\in\mathcal{I}} \langle x, S^{-1}\varphi_i\rangle\, \varphi_i. \tag{1.4}$$

The *frame decomposition*, stated above, is one of the most important frame results. It shows that if $(\varphi_i)_{i\in\mathcal{I}}$ is a frame for $\mathscr{H}$, then every element in $\mathscr{H}$ has a representation as an infinite linear combination of the frame elements. Thus, it is natural to view a frame as some kind of "generalized basis".
In particular, when the frame $(\varphi_i)_{i\in\mathcal{I}}$ does not constitute a basis, *i.e.*, it is *redundant*, the coefficient sequence $(\langle x, S^{-1}\varphi_i\rangle)_{i\in\mathcal{I}}$ of this expansion is certainly not unique; but it is this

[1]Recall that a sequence $(\varphi_i)_{i\in\mathcal{I}}$ in $\mathscr{H}$ is called a *Bessel sequence* if there exists a constant $B > 0$ such that

$$\sum_{i\in\mathcal{I}} |\langle x, \varphi_i\rangle|^2 \leq B||x||^2 \qquad \text{for all } \ x \in \mathscr{H}.$$

property which enables to derive much sparser expansions. It should also be noticed that the sequence $(\langle x, S^{-1}\varphi_i\rangle)_{i\in\mathcal{I}}$ has the characterizing property of being the smallest of all expansion coefficient sequences with respect to the ℓ^2 norm.

Despite this, (1.3) and (1.4) reveals one of the main difficulties in frame theory: to make expansions (1.3) and (1.4) practically useful we need either to find the operator S^{-1}, or to compute its action on the whole $(\varphi_i)_{i\in\mathcal{I}}$. One way to bypass the problem is to consider only tight frames, for which the following result holds true:

Corollary 1.6. *If $(\varphi_i)_{i\in\mathcal{I}}$ is a tight frame with frame bound A, then the canonical dual frame is $(A^{-1}\varphi_i)_{i\in\mathcal{I}}$, and for all $x \in \mathscr{H}$*

$$x = \frac{1}{A}\sum_{i\in\mathcal{I}} \langle x, \varphi_i\rangle\, \varphi_i.$$

Tight frames have additional advantages. For the design of frames with prescribed properties, it is essential to control the behavior of the canonical dual frame, but the complicated structure of both the frame operator and its inverse makes it difficult. For example, if we consider a frame $(\varphi_i)_{i\in\mathcal{I}}$ for $L^2(\mathbb{R})$ consisting of functions with exponential decay, nothing guarantees that the functions in the canonical dual frame $(S^{-1}\varphi_i)_{i\in\mathcal{I}}$ have exponential decay. However, for tight frames, this type of matters have satisfying arguments. Also, for a tight frame, the canonical dual frame automatically has the same structure as the frame itself. For example, if the frame has a wavelet structure, also the canonical dual frame does. In contrast, the canonical dual frame of a nontight wavelet frame might not have the wavelet structure.

1.3 Representation of locally compact groups

Representation theory of groups is a vast subject. Many of the aspects of this theory that are of interest in Applied Harmonic Analysis can be studied within the class of locally compact and second countable topological groups, even if the most interesting examples belongs to the smaller and nicer class of Lie groups. Most of the material presented here comes from the lectures attended in Genova at the Workshop *Three minicourses on Applied Harmonic Analysis.*
We start refreshing the basics of topological groups as well as giving the definition and the basic results regarding the Haar measure. For a deeper discussion on these topics the reader is referred to [14].

1.3.1 Topological Groups and Haar measure

Definition 1.7. A *topological group* is a group G endowed with a topology relative to which the group operations

$$(g, h) \mapsto gh, \qquad\qquad g \mapsto g^{-1}$$

are continuous as maps $G \times G \to G$ and $G \to G$, respectively. G is *locally compact* if every point has a compact neighborhood. We shall also assume our groups to be Hausdorff[2].

Definition 1.8. A Borel measure μ on the topological space X is called a *Radon measure* if:

(i) it is finite on compact sets;

(ii) it is outer regular on the Borel sets: for every Borel set E

$$\mu(E) = \inf\{\mu(U) : U \supset E, U \text{ open set}\}$$

[2]Recall that a *Hausdorff space* is a topological space in which distinct points have disjoint neighbourhoods.

(iii) it is inner regular on the open sets: for every open set U

$$\mu(U) = \sup\{\mu(K) : K \supset E, K \text{ compact set}\}$$

If G is a topological group, for $E \subset G$ and $x \in G$ we define

$$xE = \{xe : e \in E\}, \qquad Ex = \{ex : e \in E\}.$$

Definition 1.9. A *left Haar measure* on the topological group G is a non zero Radon measure μ such that $\mu(xE) = \mu(E)$ for every Borel set $E \subset G$ and every $x \in G$. A similar definition is given for right Haar measures.

Of course, the prototype of Haar measure is the Lebesgue measure on the additive group $\mathbb{R}^d$, which is invariant under left (and right) translations.

Theorem 1.10. *Every locally compact group G has a left Haar measure λ, which is essentially unique in the sense that if μ is any other left Haar measure, then there exists a positive constant c such that $\mu = c\lambda$.*

If we fix a left Haar measure λ on G, then for any $x \in G$ the measure λ_x defined by

$$\lambda_x(E) = \lambda(Ex)$$

is again a left Haar measure, because of the commutativity of left and right translation. Therefore there must exist a positive real number, denoted $\Delta(x)$ such that

$$\lambda_x = \Delta(x)\lambda.$$

The function $\Delta : G \to \mathbb{R}^+$ thus defined is independent of the choice of λ and it is called the *modular function* of G.

Proposition 1.11. *Let G be a locally compact group. The modular function $\Delta : G \to \mathbb{R}^+$ is a continuous homomorphism into the multiplicative group $\mathbb{R}^+$. Furthermore, for every $f \in L^1(G, \lambda)$ we have*

$$\int_G f(xy)dx = (\Delta(y))^{-1} \int_G f(x)dx.$$

A group for which every left Haar measure is also a right Haar measure, hence for which the modular function is identically equal to one, is called *unimodular*. Large classes of groups are unimodular, such as the abelian, compact, nilpotent, semisimple and reductive groups. Nevetherless, in Applied Harmonic Analysis non-unimodular groups play a prominent role, such as the affine group "$ax + b$" that we shall define below.

First, we need to recall the definition of Lie group. The class of Lie groups is smaller but nicer: indeed, since Lie groups are smooth manifolds, they can be studied using differential calculus and their geometric nature allows, for istance, to speak about dimension. For a detailed account on these matters the reader may consult [13].

Definition 1.12. A *Lie group* G is a C^∞ (smooth or differentiable) manifold endowed with a group structure such that the group operations $(g, h) \mapsto gh$ and $g \mapsto g^{-1}$ are smooth, that is, C^∞.

EXAMPLE 1.13 (The affine group "$ax + b$"). There are several possible versions of this group. Let $G = \mathbb{R}^+ \times \mathbb{R}$ as manifold. One can visualize it as the right half plane. The multiplication is obtained by thinking of the pair (a, b), with $a > 0$ and $b \in \mathbb{R}$, as identifying the affine

transformation of the real line given by $x \mapsto ax + b$, whence the name. The composition of maps

$$x \quad \mapsto \quad ax + b \quad \mapsto \quad a'(ax + b) + b' = [a'a]x + [a'b + b']$$

yields the product rule

$$(a', b')(a, b) = (a'a, a'b + b').$$

Evidently, both functions $a'a$ and $a'b + b'$ are smooth in the global coordinates on G, which is then a Lie group. Clearly, G is connected. When speaking of the "$ax + b$" group we refer to this group.
A non-connected version arises by taking $a \in \mathbb{R}$ instead of $a > 0$. Yet another slightly different construction comes from thinking of the pair (a, b) as identifying the affine transformation $x \mapsto a(x + b)$. This point of view yields both a connected and a non-connected Lie group.

Left Haar measures are very easy to construct on Lie groups.

Proposition 1.14. *If G is a Lie group whose underlying manifold is an open set in $\mathbb{R}^d$ and if the left translations are given by affine maps, that is*

$$xy = A(x)y + b(x),$$

where $A(x)$ is a linear transformation and $b(x) \in \mathbb{R}^d$, then $|\det A(x)|^{-1}dx$ is a Haar measure on G.

For example, in the group "$ax + b$" the left translations are

$$l_{(a,b)}(\alpha, \beta) = \begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} 0 \\ b \end{bmatrix},$$

so that by Proposition 1.14 we have

$$|\det A(a, b)|^{-1} da db = \frac{da}{a^2} db.$$

1.3.2 Representation Theory

Let $\mathscr{H}_1$ and $\mathscr{H}_2$ be two Hilbert spaces and suppose that $T : \mathscr{H}_1 \to \mathscr{H}_2$ is linear and bounded, that is $T \in \mathcal{B}(\mathscr{H}_1, \mathscr{H}_2)$. Recall that T is an isometry if $||Tu|| = ||u||$ for every $u \in \mathscr{H}_1$. Since $||Tu||^2 = \langle Tu, Tu\rangle = \langle T^*Tu, u\rangle$ e $||u||^2 = \langle u, u\rangle$, the polarization identity implies that T is an isometry if and only if $T^*T = id_{\mathscr{H}_1}$. Hence, isometries are injective, but they are not necessarily surjective. A bijective isometry is called a *unitary map*. If T is unitary, such is also T^{-1} and in this case $TT^* = id_{\mathscr{H}_2}$. In particular if $\mathscr{H}_1 = \mathscr{H}_2 = \mathscr{H}$, the set

$$\mathcal{U}(\mathscr{H}) = \{T \in \mathcal{B}(\mathscr{H}) : T \text{ is unitary}\}$$

forms a group.

Let now G be a locally compact Hausdorff topological (or a Lie) group.

Definition 1.15. A *unitary representation* of G on the Hilbert space $\mathscr{H}$ is a group homomorphism $\pi : G \to \mathcal{U}(\mathscr{H})$ continuous in the strong operator topology. This means:

(i) $\pi(gh) = \pi(g)\pi(h)$ for every $g, h \in G$;

(ii) $\pi(g^{-1}) = \pi(g)^{-1} = \pi(g)^*$ for every $g \in G$;

(iii) $g \mapsto \pi(g)u$ is continuous from G to $\mathscr{H}$, for every $u \in \mathscr{H}$.

EXAMPLE 1.16 (**Wavelet representation**). Let G be the "$ax+b$" group and $\mathscr{H} = L^2(\mathbb{R})$. Define

$$\pi(a,b)f(x) = \frac{1}{\sqrt{a}} f\left(\frac{x-b}{a}\right),$$

the so-called *wavelet representation*. Notice that it is just the composition of the two important and basic unitary maps

$$T_b f(x) = f(x-b) \qquad \text{(translation operator)}$$

$$D_a f(x) = \frac{1}{\sqrt{a}} f\left(\frac{x}{a}\right) \qquad \text{(dilation operator)}$$

for indeed

$$T_b D_a f(x) = T_b(D_a f)(x) = D_a f(x-b) = \frac{1}{\sqrt{a}} f\left(\frac{x-b}{a}\right).$$

Observe that

$$T_b T_{b'} = T_{b+b'}, \qquad D_a D_{a'} = D_{aa'}.$$

It is important to observe that $T_b D_a \neq D_a T_b$. More precisely,

$$D_a T_b f(x) = \frac{1}{\sqrt{a}} (T_b f)\left(\frac{x}{a}\right) = \frac{1}{\sqrt{a}} f\left(\frac{x}{a} - b\right) = \frac{1}{\sqrt{a}} f\left(\frac{x-ab}{a}\right) = T_{ab} D_a f(x).$$

In other words

$$D_a T_b = T_{ab} D_a.$$

It follows that

$$(T_\beta D_\alpha)(T_b D_a) = T_\beta (D_\alpha T_b) D_a = T_\beta (T_{\alpha b} D_\alpha) D_a = (T_\beta T_{\alpha b})(D_\alpha D_a) = T_{\beta+\alpha b} D_{\alpha a},$$

so that π is a homomorphism:

$$\pi(\alpha,\beta)\pi(a,b) = \pi(\alpha a, \beta + \alpha b) = \pi((\alpha,\beta)(a,b)).$$

Finally, it is easy to check the strong continuity.

Definition 1.17. Let $\mathcal{M}$ be a closed subspace of $\mathscr{H}$. Then $\mathcal{M}$ is called an *invariant subspace* for the unitary representation π, if $\pi(g)\mathcal{M} \subseteq \mathcal{M}$ for all $g \in G$. If there exists a nontrivial invariant subspace for π, then π is called reducible, otherwise π is *irreducible*.

REMARK 1.18. From the point of view of applications, a unitary representation π of a locally compact group G (with Haar measure dg) is particularly useful if it yields a **reproducing formula**, that is, a (weak) reconstruction of a function f in the representation space $\mathscr{H}$ from its *voice transform* $V_\eta : \mathscr{H} \to L^2(G)$ given by $V_\eta(f)(g) := \langle f, \pi(g)\eta \rangle$. In other words, a reproducing formula is a weak reconstruction of the form

$$f = \int_G \langle f, \pi(g)\eta \rangle \; \pi(g)\eta \; dg, \tag{1.5}$$

valid for every $f \in \mathscr{H}$, for some admissible $\eta \in \mathscr{H}$. We recall that, given a unitary representation π of G on $\mathscr{H}$, a function $\eta \in \mathscr{H}$ is called *admissible*, if

$$\int_G |\langle \eta, \pi(g)\eta \rangle|^2 \; dg < \infty.$$

In this case, (G, π, η) is called a *reproducing system*; otherwise, we simply say that G is a reproducing group. If π is irreducible, this is nothing else but the classical concept of square integrable representation. We stress that formula (1.5) is important also because it is the starting point for its discrete counterparts, within coorbit theory.
A classical example is when $\mathscr{H} = L^2(\mathbb{R}^d)$: in this case an admissible η is sometimes called a generating function or *wavelet*, as we shall see in the next section.

REMARK 1.19. It is clear from the above discussion that if a countable collection $(\varphi_i)_{i\in\mathcal{I}}$ in a Hilbert space $\mathscr{H}$ is a Parseval frame, then this is equivalent to the reproducing formula $f = \sum_{i\in\mathcal{I}} \langle f, \varphi_i\rangle \varphi_i$ for all $f \in \mathscr{H}$, where the series converges in the norm of $\mathscr{H}$.

1.4 Wavelets

Since *shearlets* arise naturally from the general framework of wavelet analysis, a full understanding of shearlets can only be derived through a sound understanding of wavelets theory.
The emergence of *wavelets* about 20 years ago represents a milestone in the development of efficient encoding of piecewise regular signals. The main reason for the spectacular success of wavelets consists not only in their ability to provide optimally sparse approximations of a large class of frequently occurring signals and to represent singularities much more efficiently than traditional Fourier methods, but also in the existence of fast algorithmic implementations which precisely digitalize the continuum domain transforms. The key property enabling such a unified treatment of the continuum and discrete setting is a Multiresolution Analysis, which allows a direct transition between the realms of real variable functions and digital signals. An additional aspect of the theory of wavelets, which has contributed to its success is its rich mathematical structure, which allows one to design families of wavelets with various desirable properties expressed in terms of regularity, decay, or vanishing moments. As a consequence of all these properties, wavelets have literally revolutionized image and signal processing and produced a large number of very successful applications.
In the following sections, we will present a self-contained overview of the key results from the theory and applications of wavelets.

1.4.1 One-dimensional Continuous Wavelet Transform

Let $\mathbb{A}_1$ be the *affine group* associated with $\mathbb{R}$, consisting of all pairs (a,t), $a,t \in \mathbb{R}, a > 0$, with group operation $(a,t)(a',t') = (aa', t + at')$. The *(continuous) affine systems* generated by $\psi \in L^2(\mathbb{R})$ are obtained from the action of the quasi-regular representation $\pi(a,t)$ of $\mathbb{A}_1$ on $L^2(\mathbb{R})$, that is

$$\left\{\psi_{a,t}(x) = \pi(a,t)\,\psi(x) = T_t D_a \psi(x) = \frac{1}{\sqrt{a}}\psi\left(\frac{x-t}{a}\right) : (a,t) \in \mathbb{A}_1\right\},$$

where the translation operator T_t and the dilation operator D_a are defined as above.
It was observed by Calderòn that, if ψ satisfies the admissibility condition

$$\int_0^\infty |\hat{\psi}(a\xi)|^2 \frac{da}{a} = 1 \qquad \text{for a. e. } \xi \in \mathbb{R}, \tag{1.6}$$

then any $f \in L^2(\mathbb{R})$ can be recovered via the reproducing formula:

$$f = \int_{\mathbb{A}_1} \langle f, \psi_{a,t}\rangle\ \psi_{a,t}\ d\mu(a,t),$$

where $d\mu(a,t) = dt\frac{da}{a^2}$ is the left Haar measure of $\mathbb{A}_1$. If ψ satisfies (2.6) the function ψ is called a *continuous wavelet* and

$$L^2(\mathbb{R}) \ni f \ \mapsto\ \mathcal{W}_\psi f(a,t) = \langle f, \psi_{a,t}\rangle \qquad (a,t) \in \mathbb{A}_1$$

is the *Continuous Wavelet Transform* of f. We refer to [33] for more details about this.

Discrete affine systems and wavelets are obtained by 'discretizing' appropriately the corresponding continuous systems. In fact, by replacing $(a,t) \in \mathbb{A}_1$ with the discrete set $(2^j, 2^j m)$, $j,m \in \mathbb{Z}$, one obtains the discrete dyadic affine system

$$\{\psi_{j,m}(x) = T_{2^j m} D_2^j \psi(x) = D_2^j T_m \psi(x) = 2^{-j/2}\psi(2^{-j}x - m) : (j,m) \in \mathbb{Z}\}, \tag{1.7}$$

and ψ is called a *wavelet* if (1.7) is an orthonormal basis or, more generally, a Parseval frame for $L^2(\mathbb{R})$. The associated *Discrete Wavelet Transform* is then defined to be the map

$$L^2(\mathbb{R}) \ni f \;\mapsto\; \mathcal{W}_\psi f(j,m) = \langle f, \psi_{j,m} \rangle, \qquad j,m \in \mathbb{Z}.$$

Being a wavelet is by no means very restrictive and a lot of choices exist. Indeed, it is possible to construct wavelets ψ which are *well localized*, in the sense that they have rapid decay both in the spatial and frequency domain, or which satisfy other regularity or decay requirements. Among the classical constructions, let us highlight the two most well-known wavelets. The *Daubechies wavelets* have compact support and can be chosen to have high regularity, leading to good decay in the frequency domain. It is not possible to write them down in closed form, but the interested reader can refer to [1, 20] for the construction. The *Lemarièe-Meyer* wavelet ψ_{LM} is defined by $\hat{\psi}_{LM}(\omega) = e^{i\pi\omega}\, b(\omega)$, where

$$b(\omega) = \begin{cases} \sin(\frac{\pi}{2}(3|\omega|-1)) & \frac{1}{3} \le |\omega| \le \frac{2}{3} \\ \sin(\frac{3\pi}{4}(\frac{4}{3}-|\omega|)) & \frac{2}{3} \le |\omega| \le \frac{4}{3} \\ 0 & \text{otherwise} \end{cases}$$

The Lemariè-Meyer wavelets are band-limited and C^∞ in the frequency domain, forcing rapid decay in the spatial domain (cf. [20]).
It should be emphasized that the localization properties of wavelet bases are among the major differences with respect to Fourier bases and play a fundamental role in their approximation properties, as we will show below.
There is a general machinery to construct orthonormal wavelet bases called *Multiresolution Analysis* (MRA). In dimension $d = 1$, this is defined as a sequence of closed subspaces $(V_j)_{j\in\mathbb{Z}}$ in $L^2(\mathbb{R})$ which satisfies the following properties:

(i) $\{0\} \subset \ldots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \ldots \subset L^2(\mathbb{R})$.

(ii) $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$ and $\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R})$.

(iii) $f \in V_j$ if and only if $D_2^{-1} f \in V_{j-1}$.

(iv) There exists a $\phi \in L^2(\mathbb{R})$, called *scaling function*, such that $\{T_m\phi : m \in \mathbb{Z}\}$ is an orthonormal basis for V_0.

This approach enables the decomposition of functions into different resolution levels associated with the so-called wavelet spaces W_j, $j \in \mathbb{Z}$. These spaces are defined by considering the orthogonal complements

$$W_j = V_{j+1} \ominus V_j, \qquad j \in \mathbb{Z}.$$

That is, a function $f_{j+1} \in V_{j+1}$ is decomposed as $f_{j+1} = f_j + g_j \in V_j \oplus W_j$, where f_j contains, roughly, the lower frequency component of f_{j+1} and g_j its higher frequency component. It follows that $L^2(\mathbb{R})$ can be broken up as a direct sum of wavelet spaces. Also, given an MRA, there always exists a function $\psi \in L^2(\mathbb{R})$ such that $\{\psi_{j,m} : j,m \in \mathbb{Z}\}$ is an orthonormal basis for $L^2(\mathbb{R})$. In fact, the MRA approach allows to introduce an alternative orthonormal basis of the form

$$\{\phi_m = T_m\phi = \phi(\cdot - m) : m \in \mathbb{Z}\} \cup \{\psi_{j,m} : j \ge 0, m \in \mathbb{Z}\}$$

that involves both the wavelet and the scaling function.
In this case, the translates of the scaling function take care of the low frequency region - the subspace $V_0 \subset L^2(\mathbb{R})$ - and the wavelet terms of the high frequency region - the complementary space $L^2(\mathbb{R}) \ominus V_0$. For additional information about the theory of MRA see, *e.g.*, [31] and [6].

1.4.2 Higher-dimensional Continuous Wavelet Transform

The extension of wavelet theory to higher dimensions requires to extend the theory of affine systems to higher dimensions. The natural way to do it is by replacing $\mathbb{A}_1$ with the *full affine group of motions* on $\mathbb{R}^d$, $\mathbb{A}_d$, consisting of the pairs $(M,t) \in GL_d(\mathbb{R}) \ltimes \mathbb{R}^d$ with group operation $(M,t)\cdot(M',t') = (MM', t+Mt')$. Similarly to the one-dimensional case, the affine systems generated by $\psi \in L^2(\mathbb{R}^d)$ are given by

$$\{\psi_{(M,t)}(x) = T_t D_M \psi(x) = |\det M|^{-\frac{1}{2}} \psi(M^{-1}(x-t)) : (M,t) \in \mathbb{A}_d\}, \tag{1.8}$$

where the *dilation operator* D_M is defined by $D_M\psi(x) = |\det M|^{-\frac{1}{2}}\psi(M^{-1}x)$.
The mathematical structure of the affine systems becomes evident by observing that (1.8) can be generated by the action of the unitary representation $\pi : \mathbb{A}_d \to \mathcal{U}(L^2(\mathbb{R}^d))$ defined by $\pi(M,t) = T_t D_M$. This allows us to write the elements of a wavelet system as

$$\psi_{(M,t)} = \pi(M,t)\,\psi.$$

Then the following result on reproducibility of functions in $L^2(\mathbb{R}^d)$ holds true.

Theorem 1.20. *Retaining the notations introduced in this subsection, let $d\mu$ be a left-invariant Haar measure on $G \subset GL_d(\mathbb{R})$, and $d\lambda$ be a left Haar measure of $\mathbb{A}_d$. Further, suppose that $\psi \in L^2(\mathbb{R}^d)$ satisfies the admissibility condition*

$$\int_G |\hat{\psi}(M^T\xi)|^2 |\det M| d\mu(M) = 1.$$

Then any function $f \in L^2(\mathbb{R}^d)$ can be recovered via the reproducing formula

$$f = \int_{\mathbb{A}_d} \langle f, \psi_{(M,t)}\rangle \psi_{(M,t)} d\lambda(M,t),$$

interpreted weakly.

When the conditions of the above theorem are satisfied, $\psi \in L^2(\mathbb{R}^d)$ is called a *continuous wavelet*. The associated *Continuous Wavelet Transform* is defined to be the mapping

$$L^2(\mathbb{R}^d) \ni f \;\mapsto\; \mathcal{W}_\psi f(M,t) = \langle f, \psi_{M,t}\rangle, \qquad M,t \in \mathbb{A}_d.$$

One interesting special case is obtained, when the dilation group G has the form $G = \{a\mathbb{I}_d : a > 0\}$, which corresponds to the case of *isotropic dilations*, *i.e.* the dilation factor a acts in the same way for each coordinate direction. In this case, the admissibility condition for ψ becomes

$$\int_0^\infty |\hat{\psi}(a\xi)|^2 \frac{da}{a} = 1,$$

and the *(isotropic) Continuous Wavelet Transform* is the mapping of $f \in L^2(\mathbb{R}^d)$ into

$$\mathcal{W}_\psi f(a,t) = a^{-d/2} \int_{\mathbb{R}^d} f(x)\overline{\psi(a^{-1}(x-t))} dx, \qquad a > 0, t \in \mathbb{R}^d. \tag{1.9}$$

It is reasonable to expect that, by choosing more general dilation groups G, one obtains wavelet with more interesting geometric properties.

1.4.3 Drawbacks

Despite their success, wavelets are not very effective when dealing with multivariate data. In fact, wavelet representations are optimal for approximating data with pointwise singularities only and cannot handle equally well the case of distributed singularities, such as singularities along curves. The intuitive reason is that wavelets are *isotropic* objects, being generated by isotropically dilating a single generator or finite set of generators. However, in dimensions greater than one, distributed discontinuities, such as edges of surface boundaries, are usually present or even dominant and – as a result – wavelets are far from optimal in dealing with multivariate data.
More generally, it can be shown that the Continuous Wavelet Transform can be used to characterize the singular support of a function f, *i.e.*, the Continuous Wavelet Transform of a function f identify the location of the singularity through its asymptotic decay at fine scales. Indeed, if f is a function regular everywhere except for a singularity at x_0 and ψ is smooth, a direct computation shows that $\mathcal{W}_\psi f(a,t)$, given by (1.9), has rapid asymptotic decay, as $a \to 0$, for all values of t, but $t = x_0$. Despite this, the Continuous Wavelet Transform is unable to provide additional information about the geometry of the set of singularities, essentially because it lacks of directional sensitivity. A detailed proof is far from the aim of this introduction; the interested reader is referred to [25] for a simple heuristic argument and to [24] for a deeper study.
As far as Discrete Wavelet Systems concern, we recall here below the definition of *non-linear approximation*, just to outline the approximation error that can be achieved: this allows to show that, for a smooth function $f \in L^2(\mathbb{R})$ with pointwise discontinuities, Discrete Wavelet Systems – as stated above – have much better nonlinear approximation rates than the Fourier basis. Nevertheless, Discrete Wavelet Systems performances degrade with respect to both the one-dimensional case and the optimal error rate in the general d-dimensional case.
In the context of wavelet bases the *best N-term approximation* is the proper notion of approximation. Roughly, for a function $f \in L^2(\mathbb{R}^d)$, the best N-term approximation f_N of f with respect to a wavelet basis is obtained by approximating f from its N largest wavelet coefficients in magnitude, rather than from its "first" N coefficients, as it is the standard approach in linear Fourier approximations. Hence, denoting by Λ_N the index set corresponding to the N largest wavelet coefficients $|\langle f, \psi_\lambda \rangle|$ associated with some wavelet basis $(\psi_\lambda)_{\lambda \in \Lambda}$, the best N-term approximation of some $f \in L^2(\mathbb{R}^d)$ in $(\psi_\lambda)_{\lambda \in \Lambda}$ is defined as

$$f_N = \sum_{\lambda \in \Lambda_N} \langle f, \psi_\lambda \rangle \psi_\lambda.$$

If a function is expanded in a frame rather than in a basis, the best N-term approximation usually can not be explicitly determined.
The approximation error is measured by

$$E_N(f) = ||f - f_N||^2.$$

If $E_N(f)$ decays rapidly as N increases, then the representation system $\{\psi_\lambda : \lambda \in \Lambda\}$ is *sparse* and most of the information or essential features of f can be recovered by using a few representation terms only.
Now, if $f_N^F \in L^2(\mathbb{R})$ is the best N-term approximation with respect to the Fourier series, and likewise, considering an orthonormal basis $\{\psi_{j,m}\}$, $f_N^W \in L^2(\mathbb{R})$ is the best N-terms approximation from the wavelet coefficients $\{\langle f, \psi_{j,m} \rangle\}$ in absolute value, then it can be shown that wavelets provide the optimal approximation error rate [25, 28]:

$$\text{Fourier approximation error: } ||f - f_N^F||^2 \leq c \cdot N^{-1}, \quad N \to \infty,$$

$$\text{Wavelet approximation error: } ||f - f_N^W||^2 \leq c \cdot N^{-2}, \quad N \to \infty.$$

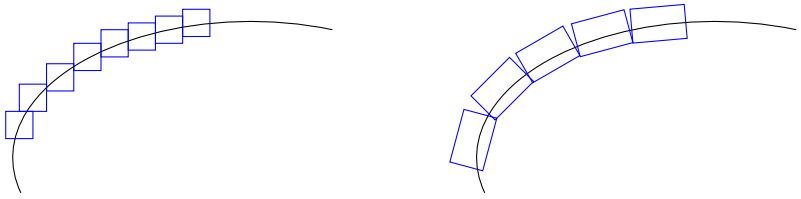
As already stressed, although the Wavelet Transform outperforms the Fourier Transform for one-dimensional signals, it does not perform equally well in higher dimensions, where anisotropic features such as curves begin to play a role. In fact, the Continuous Wavelet Transform is not able to precisely identify the wavefront set of a distribution, neither are the Discrete Wavelet Systems capable to reach the optimal approximation error. Given a function f which is in $\mathcal{C}^2(\mathbb{R}^2 \setminus \Gamma)$ where Γ is a $\mathcal{C}^2$ curve, the nonlinear approximation errors satisfy the following estimates:

$$\text{Fourier approximation error: } ||f - f_N^F||^2 \leq c \cdot N^{-\frac{1}{2}}, \quad N \to \infty,$$

$$\text{Wavelet approximation error: } ||f - f_N^W||^2 \leq c \cdot N^{-1}, \quad N \to \infty,$$

$$\text{Optimal approximation error: } ||f - f_N^O||^2 \leq c \cdot N^{-2}, \quad N \to \infty.$$

As already stressed, the key problem of the suboptimal behavior of Fourier series and wavelet bases is the fact that these systems are generated by isotropic elements. Intuitively, since the discontinuity is spatially distributed, it interacts extensively with the elements of the wavelet basis, and thus "many" wavelet coefficients are needed to represent the function accurately.



(a) Wavelet covering: isotropic elements capturing a discontinuity curve.

(b) Shearlet covering: rotated, anisotropic elements capturing a discontinuity curve.

Figure 1.1: Shearlets are sparser than wavelets in approximating discontinuity curves.

Nevertheless, considering wavelets with anisotropic scaling will not fix the situation because of the lack of control on the direction of the elements. Thus, to capture a discontinuity curve, one needs not only anisotropic elements but also a location parameter to locate the elements on the curve and a rotation parameter to align the elongated elements in the direction of the curve. As we shall see in the next chapter, *shearlets* were introduced by Guo, Kutyniok, Labate, Lim and Weiss in [18, 29] to address this problem.

Chapter 2

Continuous Shearlet Systems

Before formally defining the (Continuous) Shearlet Systems, let us introduce intuitively the main ideas of its construction. In this regard, we first restrict ourselves to the two-dimensional case (the multivariate one will be discussed later).
As pointed out in the previous section, to achieve optimally sparse approximations of functions exhibiting anisotropic singularities, the analyzing elements must consist of waveform elements ranging over several scales, orientations, and locations with the ability to become increasingly elongated at finer scales. This requires a combination of an appropriate scaling operator to generate elements at different scales, an orthogonal operator to change their orientations, and a translation operator to displace these elements over the space.
As can be easily imagined, for the translation operator one can use the standard operator T_t as defined in (1.1).
Next, we require a scaling operator to generate waveforms with anisotropic support. A natural choice is to use the family of dilation operators D_{A_a} , $a > 0$, based on *parabolic scaling matrices* A_a of the form

$$A_a = \begin{pmatrix} a & 0 \\ 0 & a^{1/2} \end{pmatrix},$$

where the dilation operator is defined as in (1.2). In particular, A_a produces *parabolic scaling*, that is $f(A_a x) = f\left(A_a \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right)$ leaves invariant the parabola $x_1 = x_2^2$. It should be mentioned that, rather than A_a, it could be used the more general matrices

$$\begin{pmatrix} a & 0 \\ 0 & a^{\alpha} \end{pmatrix},$$

where $\alpha \in (0, 1)$ controls the "degree of anisotropy". However, the value $\alpha = 1/2$ plays a special role in the discrete setting, *i.e.*, when the parameters of the shearlet system are discretized: indeed, parabolic scaling is required to obtain optimally sparse approximations (cf. [27]). For this reason, in the remainder of this chapter, we will only consider the case $\alpha = 1/2$.
Finally, as far as the orthogonal transformation concerns, the most obvious choice seems to be the rotation operator. However, this is not "practicable" since rotations destroy the structure of the integer lattice $\mathbb{Z}^2$ (whenever the rotation angle is different from $0, \pm\frac{\pi}{2}, \pm\pi, \pm\frac{3\pi}{2}$). This becomes a serious issue for the transition from the continuum to the discrete setting. An alternative choice for the orthogonal transformation is the shearing operator D_{S_s}, $s \in \mathbb{R}$, associated with the *shearing matrix* S_s given by

$$S_s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}.$$

The shearing matrix parameterizes the orientations using the variable s associated with the slopes rather than the angles, and has the advantage of leaving the integer lattice invariant,

provided that s is an integer.
Roughly speaking, the operation of shearing is a translation along an axis (*e.g.*, the abscissa axis x) by an amount that increases linearly with another axis (the ordinate axis y). Thus, the shear transformation leave the y coordinate of any point (x, y) unchanged while the x coordinate is stretched in a linear way, based on the height of the point above the x axis, *i.e.*, on y. The result is a shape distortions as if objects were composed of layers that slide one over another. The change of coordinates has the form:

$$\begin{cases} x' = x + sy \\ y' = y \end{cases}$$

where s is the constant that measures the degree of shearing. Clearly, if s is negative the shearing is in the opposite direction.
For example, in 2D a shear along the x direction changes a rectangle (with lower right corner at the origin) into a parallelogram, as it is shown in Figure 2.1:

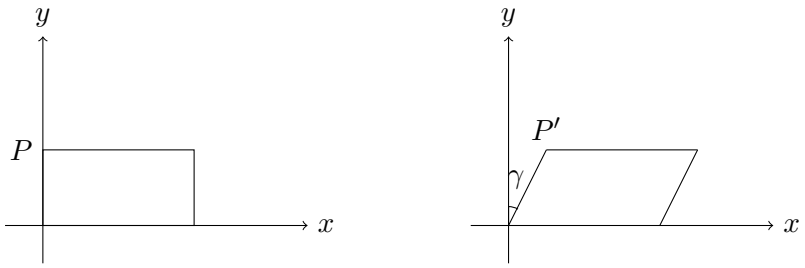


Figure 2.1: The shearing translation.

Note that the point $P(0, H)$ is taken into the point $P'(sH, H)$. It follows that the shearing angle γ (the angle through which the vertical edge is sheared) is given by:

$$\tan(\gamma) = \frac{sH}{H} = s.$$

So the parameter s is just the trigonometric tangent of the shearing angle.

Combining these three operators, we are ready to define the *Continuous Shearlet Systems*. These definitions are taken from [25].

Definition 2.1. For $\psi \in L^2(\mathbb{R}^2)$, the *Continuous Shearlet System* $\mathcal{SH}(\psi)$ is defined by

$$\mathcal{SH}(\psi) = \{\psi_{a,s,t}(x) = T_t D_{S_s} D_{A_a} \psi(x) = a^{-\frac{3}{4}} \psi(A_a^{-1} S_s^{-1}(x - t)) : a > 0, s \in \mathbb{R}, t \in \mathbb{R}^2\}. \quad (2.1)$$

The associated *Continuous Shearlet Transform* of some $f \in L^2(\mathbb{R}^2)$ is given by

$$L^2(\mathbb{R}^2) \ni f \to \mathscr{SH}_\psi f(a, s, t) = \langle f, \psi_{a,s,t} \rangle, \qquad (a, s, t) \in \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^2.$$

In other word, the Continuous Shearlet Transform projects the function f onto the functions $\psi_{a,s,t}$ at scale a, orientation s and location t. The anisotropic dilation A_a controls the 'scale' of the shearlets, by applying a different dilation factor along the two axes. This ensures that the frequency support of the shearlets becomes increasingly elongated at finer scales: indeed, as $a \to 0$ we obtain needlelike functions. The shear matrix S_s is not expansive and determines the orientation of the shearlets using the shear parameter s to detect different directions by slope. Finally, the location parameter t ensures position sensitivity.

2.1 The Shearlet Group

In this section we show that the elements of a Shearlet System can be generated by using a representation of a special group that we will refer to as the *Shearlet Group*.

Lemma 2.2. *The set* $\mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^2$ *equipped with multiplication given by*

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

forms a group.

Proof. It is easy to check that $(1, 0, 0)$ is the neutral element. The inverse of some $(a, s, t) \in \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^2$ is given by

$$(a,\, s,\, t)^{-1} = \left(\frac{1}{a},\, -\frac{s}{\sqrt{a}},\, -A_a^{-1} S_s^{-1} t\right).$$

Indeed,

$$(a,\, s,\, t) \cdot \left(\frac{1}{a}, -\frac{s}{\sqrt{a}}, -A_a^{-1} S_s^{-1} t\right) = \left(a\frac{1}{a}, s - \frac{s}{\sqrt{a}}\sqrt{a}, t - S_s A_a A_a^{-1} S_s^{-1} t\right) = (1, 0, 0)$$

and

$$\left(\frac{1}{a},\, -\frac{s}{\sqrt{a}},\, -A_a^{-1} S_s^{-1} t\right) \cdot (a,\, s,\, t) = \left(\frac{1}{a}a,\, -\frac{s}{\sqrt{a}} + s\frac{1}{\sqrt{a}},\, -A_a^{-1} S_s^{-1} t + S_{-\frac{s}{\sqrt{a}}} A_{\frac{1}{a}} t\right) = (1,\, 0,\, 0)$$

since

$$S_{-\frac{s}{\sqrt{a}}} A_{\frac{1}{a}} = \begin{pmatrix} 1 & -\frac{s}{\sqrt{a}} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{a} & 0 \\ 0 & \frac{1}{\sqrt{a}} \end{pmatrix} = \begin{pmatrix} \frac{1}{a} & -\frac{s}{a} \\ 0 & \frac{1}{\sqrt{a}} \end{pmatrix} = \begin{pmatrix} \frac{1}{a} & 0 \\ 0 & \frac{1}{\sqrt{a}} \end{pmatrix} \begin{pmatrix} 1 & -s \\ 0 & 1 \end{pmatrix} = A_a^{-1} S_s^{-1}.$$

As far as the associative property concerns, observe that

$$\begin{aligned} S_s\, A_a\, S_{s'}\, A_{a'} &= \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & \sqrt{a} \end{pmatrix} \begin{pmatrix} 1 & s' \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a' & 0 \\ 0 & \sqrt{a'} \end{pmatrix} \\ &= \begin{pmatrix} a & s\sqrt{a} \\ 0 & \sqrt{a} \end{pmatrix} \begin{pmatrix} a' & s'\sqrt{a'} \\ 0 & \sqrt{a'} \end{pmatrix} \\ &= \begin{pmatrix} aa' & s'a\sqrt{a'} + s\sqrt{aa'} \\ 0 & \sqrt{aa'} \end{pmatrix} \\ &= \begin{pmatrix} 1 & s + s'\sqrt{a} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} aa' & 0 \\ 0 & \sqrt{aa'} \end{pmatrix} \\ &= S_{s+s'\sqrt{a}}\, A_{aa'} \end{aligned}$$

So, the following computation holds true:

$$\begin{aligned} ((a,\, s,\, t) \cdot (a',\, s',\, t')) \cdot (a'',\, s'',\, t'') &= (aa',\, s + s'\sqrt{a},\, t + S_s A_a t') \cdot (a'',\, s'',\, t'') \\ &= (aa'a'',\, s + s'\sqrt{a} + s''\sqrt{aa'},\, t + S_s A_a t' + S_{s+s'\sqrt{a}} A_{aa'} t'') \\ &= (a(a'a''),\, s + (s' + s''\sqrt{a})\sqrt{a'},\, t + S_s A_a t' + S_s A_a S_{s'} A_{a'} t'') \\ &= (a(a'a''),\, s + (s' + s''\sqrt{a})\sqrt{a'},\, t + S_s A_a (t' + S_{s'} A_{a'} t'')) \\ &= (a,\, s,\, t) \cdot (a'a'',\, s' + s''\sqrt{a},\, t' + S_{s'} A_{a'} t'') \\ &= (a,\, s,\, t) \cdot ((a',\, s',\, t') \cdot (a'',\, s'',\, t'')). \end{aligned}$$

□

The previous lemma allows us to give the definition stated below.

Definition 2.3. The *Shearlet Group* $\mathbb{S}$ is defined to be the set $\mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^2$ along with the multiplication law given by

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

REMARK 2.4. It has been shown that the Shearlet Group is isomorphic to the locally compact group $G \ltimes \mathbb{R}^2$, where $G \subset GL_2(\mathbb{R})$ is the set of matrices

$$G = \left\{ M_{a,s} = \begin{pmatrix} a & s\sqrt{a} \\ 0 & \sqrt{a} \end{pmatrix} : a \in \mathbb{R}^+, s \in \mathbb{R} \right\}.$$

Thus, it is a subgroup of the full group of motions $\mathbb{A}_d = GL_d(\mathbb{R}) \ltimes \mathbb{R}^d$. It is evident that the matrix $M_{a,s}$ is the superposition of the parabolic scaling matrix A_a and the shear matrix S_s defined above.

Now, we are ready to prove that the map $\sigma : \mathbb{S} \to \mathcal{U}(L^2(\mathbb{R}^2))$ is a unitary representation of the Shearlet Group, where $\mathcal{U}(L^2(\mathbb{R}^2))$ denotes, as usual, the group of unitary operators on $L^2(\mathbb{R}^2)$. Clearly, likewise the wavelets case, this representation can be related to the Continuous Shearlet Transform in the following way:

$$\mathscr{SH}_\psi f(a,s,t) = \langle f, \psi_{a,s,t} \rangle = \langle f, \sigma(a,s,t)\psi \rangle \qquad \text{for all } f \in L^2(\mathbb{R}^2).$$

Lemma 2.5. *Let $\sigma : \mathbb{S} \to \mathcal{U}(L^2(\mathbb{R}^2))$ be defined by*

$$\sigma(a,s,t)\psi(x) = \psi_{a,s,t} = a^{-\frac{3}{4}} \psi(A_a^{-1} S_s^{-1}(x-t)).$$

Then σ is a unitary representation of $\mathbb{S}$ on $L^2(\mathbb{R}^2)$.

Proof. Observe that

$$A^{-1}_{aa'}\, S^{-1}_{s+s'\sqrt{a}} = \begin{pmatrix} \frac{1}{aa'} & 0 \\ 0 & \frac{1}{\sqrt{aa'}} \end{pmatrix} \begin{pmatrix} 1 & -s - s'\sqrt{a} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{aa'} & -\frac{s}{aa'} - \frac{s'}{a'\sqrt{a}} \\ 0 & \frac{1}{\sqrt{aa'}} \end{pmatrix}$$

Thus,

$$\begin{aligned} A^{-1}_{a'}\, S^{-1}_{s'}\, A^{-1}_{a}\, S^{-1}_{s} &= \begin{pmatrix} \frac{1}{a'} & 0 \\ 0 & \frac{1}{\sqrt{a'}} \end{pmatrix} \begin{pmatrix} 1 & -s' \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{a} & 0 \\ 0 & \frac{1}{\sqrt{a}} \end{pmatrix} \begin{pmatrix} 1 & -s \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{a'} & -\frac{s'}{a'} \\ 0 & \frac{1}{\sqrt{a'}} \end{pmatrix} \begin{pmatrix} \frac{1}{a} & -\frac{s}{a} \\ 0 & \frac{1}{\sqrt{a}} \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{aa'} & -\frac{s}{aa'} - \frac{s'}{a'\sqrt{a}} \\ 0 & \frac{1}{\sqrt{aa'}} \end{pmatrix} \\ &= A^{-1}_{aa'}\, S^{-1}_{s+s'\sqrt{a}} \end{aligned}$$

Now, let $\psi \in L^2(\mathbb{R}^2)$ and $x \in \mathbb{R}^2$. Then

$$\begin{aligned} \sigma(a,s,t)\,(\sigma(a',s',t')\psi)(x) &= a^{-\frac{3}{4}}\, \sigma(a',s',t')\, \psi(A_a^{-1} S_s^{-1}(x-t)) \\ &= (aa')^{-\frac{3}{4}}\, \psi(A^{-1}_{a'} S^{-1}_{s'} (A_a^{-1} S_s^{-1}(x-t) - t')) \\ &= (aa')^{-\frac{3}{4}}\, \psi(A^{-1}_{a'} S^{-1}_{s'} A_a^{-1} S_s^{-1}(x - (t + S_s A_a t'))) \\ &= (aa')^{-\frac{3}{4}}\, \psi(A^{-1}_{aa'} S^{-1}_{s+s'\sqrt{a}}(x - (t + S_s A_a t'))) \\ &= \sigma((aa',\, s + s'\sqrt{a},\, t + S_s A_a t'))\psi(x) \\ &= \sigma((a,s,t) \cdot (a',s',t'))\psi(x) \end{aligned}$$

Thus, σ is a representation. The second assertion follows immediately by that fact that $\sigma(a,s,t)^*\, \sigma(a,s,t) = \mathrm{Id}$ which in turn yields $\sigma(a,s,t)^* = \sigma((a,s,t)^{-1})$. □

In order to formulate the admissibility condition associated with the Shearlet Group $\mathbb{S}$, we need first the left Haar measure for this group. We already stressed that, for locally compact groups, there always exist a left invariant Haar integrals. For the Shearlet group, this invariance implies

$$\int_{\mathbb{S}} f(a,s,t)\,d\mu = \int_{\mathbb{S}} f(a'a,\, s'+s\sqrt{a'},\, t'+S_{s'}A_{a'}t)\,d\mu$$

where $d\mu = \nu(a,s,t)da\,ds\,dt$ and ν is a function of the parameters of the Shearlet transform. The calculation of ν can be done by calculating the Jacobian of the change of variables:

$$a'' = a'a, \quad s'' = s' + s\sqrt{a'}, \quad t'' = t' + S_{s'}A_{a'}t,$$

which turns out to be $\frac{1}{a^3}$. Thus, the left Haar measure is $\frac{da\,ds\,dt}{a^3}$.
A heuristic explanation for the power of -3 in the density is the fact that this measure divides the parameter space into unit cells of side a by $\sqrt{a}$ in space (hence a factor $a^{-3/2}$), unit intervals of length $\sqrt{a}$ on the space of directions (hence, a factor $a^{-1/2}$), and finally a factor of a^{-1} since a is a scale parameter:

$$d\mu = \frac{dt}{a^{3/2}}\,\frac{ds}{a^{1/2}}\,\frac{da}{a} = \frac{da\,ds\,dt}{a^3}.$$

This point of view will be important in understanding the discretization of the transform.

Now, we are ready to derive the admissibility condition associated with the representation σ of the Shearlet Group $\mathbb{S}$. We recall that the admissibility condition is important, since this is automatically associated with a reconstruction formula. This approach can be found in [8].

Theorem 2.6. *If $f, \psi \in L^2(\mathbb{R}^2)$, then*

$$\begin{aligned}\int_{\mathbb{S}} |\langle f, \psi_{a,s,t}\rangle|^2 \frac{da\,ds\,dt}{a^3} &= \int_{\mathbb{R}}\int_0^\infty |\hat{f}(\omega)|^2 d\omega_1\,d\omega_2 \int_0^\infty\int_{\mathbb{R}} \frac{|\hat{\psi}(\nu_1,\nu_2)|^2}{\nu_1^2} d\nu_2\,d\nu_1 \\ &\quad + \int_{\mathbb{R}}\int_{-\infty}^0 |\hat{f}(\omega)|^2 d\omega_1\,d\omega_2 \int_{-\infty}^0\int_{\mathbb{R}} \frac{|\hat{\psi}(\nu_1,\nu_2)|^2}{\nu_1^2} d\nu_2\,d\nu_1\end{aligned}$$

Proof. Notice that the Shearlet transform of some function $f \in L^2(\mathbb{R}^2)$ can be regarded as a convolution product. Indeed,

$$\mathscr{SH}_\psi f(a,s,t) = \langle f, a^{-\frac{3}{4}}\psi(A_a^{-1}S_s^{-1}(\cdot - t))\rangle = f * \psi^*_{a,s,0}(t), \tag{2.2}$$

where $\psi^*_{a,s,0}(x) = \overline{\psi_{a,s,0}(-x)}$ for all $x \in \mathbb{R}^2$. The Fourier transform of an element of the shearlets system can be computed easily using (i) of Theorem 1.2:

$$\widehat{\psi_{a,s,t}}(\omega) = a^{\frac{3}{4}}\mathrm{e}^{-2\pi i\langle t,\omega\rangle}\hat{\psi}(A_a^T S_s^T\omega) = a^{\frac{3}{4}}\mathrm{e}^{-2\pi i\langle t,\omega\rangle}\hat{\psi}(a\omega_1, \sqrt{a}(\omega_2 + s\omega_1)), \tag{2.3}$$

where B^T denotes the transpose of a matrix B. Using (2.2), the Plancherel theorem, and (2.3),

we obtain

$$\begin{aligned}
\int_{\mathbb{S}} |\langle f, \psi_{a,s,t}\rangle|^2 \frac{da\,ds\,dt}{a^3} &= \int_{\mathbb{S}} |f * \psi^*_{a,s,0}(t)|^2 \, dt\, ds\, \frac{da}{a^3} \\
&= \int_0^\infty \int_{\mathbb{R}} \int_{\mathbb{R}^2} |\hat{f}(\omega)|^2 \, |\widehat{\psi^*_{a,s,0}}(\omega)|^2 \, d\omega\, ds\, \frac{da}{a^3} \\
&= \int_0^\infty \int_{\mathbb{R}} \int_{\mathbb{R}^2} |\hat{f}(\omega)|^2 \, a^{\frac{3}{2}} \, |\hat{\psi}(A_a^T S_s^T \omega)|^2 d\omega\, ds\, \frac{da}{a^3} \\
&= \int_0^\infty \int_{\mathbb{R}^2} \int_{\mathbb{R}} |\hat{f}(\omega)|^2 \, a^{-\frac{3}{2}} \, |\hat{\psi}(a\omega_1, \sqrt{a}(\omega_2 + s\omega_1))|^2 ds\, d\omega\, da \\
&= \int_{\mathbb{R}} \int_0^\infty \int_0^\infty \int_{\mathbb{R}} |\hat{f}(\omega)|^2 \, a^{-2} \omega_1^{-1} |\hat{\psi}(a\omega_1, \nu_2)|^2 d\nu_2\, da\, d\omega_1\, d\omega_2 \\
&\quad - \int_{\mathbb{R}} \int_{-\infty}^0 \int_0^\infty \int_{\mathbb{R}} |\hat{f}(\omega)|^2 \, a^{-2} \omega_1^{-1} |\hat{\psi}(a\omega_1, \nu_2)|^2 d\nu_2\, da\, d\omega_1\, d\omega_2 \\
&= \int_{\mathbb{R}} \int_0^\infty |\hat{f}(\omega)|^2 d\omega_1\, d\omega_2 \int_0^\infty \int_{\mathbb{R}} \frac{|\hat{\psi}(\nu_1, \nu_2)|^2}{\nu_1^2} d\nu_2\, d\nu_1 \\
&\quad + \int_{\mathbb{R}} \int_{-\infty}^0 |\hat{f}(\omega)|^2 d\omega_1\, d\omega_2 \int_{-\infty}^0 \int_{\mathbb{R}} \frac{|\hat{\psi}(\nu_1, \nu_2)|^2}{\nu_1^2} d\nu_2\, d\nu_1
\end{aligned}$$

The last two equivalences stem from the change of variables

$$\nu_2 = \sqrt{a}(\omega_2 + s\omega_1), \qquad \nu_1 = a\omega_1.$$

□

The next two corollaries follow immediately from Theorem 2.6.

Corollary 2.7. *Let $\psi \in L^2(\mathbb{R}^2)$ be such that*

$$\int_{\mathbb{R}^2} \frac{|\hat{\psi}(\nu_1, \nu_2)|^2}{\nu_1^2} d\nu_2\, d\nu_1 < \infty \tag{2.4}$$

is satisfied. Then ψ is admissible.

Corollary 2.8. *Given an admissible $\psi \in L^2(\mathbb{R}^2)$, define*

$$c_\psi^+ = \int_0^\infty \int_{\mathbb{R}} \frac{|\hat{\psi}(\nu_1, \nu_2)|^2}{\nu_1^2} d\nu_2\, d\nu_1, \qquad c_\psi^- = \int_{-\infty}^0 \int_{\mathbb{R}} \frac{|\hat{\psi}(\nu_1, \nu_2)|^2}{\nu_1^2} d\nu_2\, d\nu_1.$$

If $c_\psi^- = c_\psi^+ = c_\psi$, then the shearlet transform is a c_ψ-multiple of an isometry. Clearly, if $c_\psi = 1$, then the shearlet transform is an isometry.

Definition 2.9. A function $\psi \in L^2(\mathbb{R}^2)$ is called a *Continuous Shearlet* if it satisfies the admissibility condition (2.4).

Notice that examples of admissible shearlets are very easy to construct, including examples of admissible shearlets which are well localized. Essentially, any function ψ bandlimited (*i.e.*, such that $\hat{\psi}$ is compactly supported) away from the origin is an admissible shearlet. The following example, which can be found in [18, 29], is very important.

Example 2.10 (Classical Shearlet). Let $\psi \in L^2(\mathbb{R}^2)$ be defined by

$$\hat{\psi}(\xi) = \hat{\psi}(\xi_1, \xi_2) = \hat{\psi}_1(\xi_1)\hat{\psi}_2\left(\frac{\xi_2}{\xi_1}\right),$$

where $\psi_1 \in L^2(\mathbb{R}^2)$ is a discrete wavelet in the sense that it satisfies the discrete Calderòn condition, given by

$$\sum_{j\in\mathbb{Z}} |\hat{\psi}_1(2^{-j}\xi)|^2 = 1 \qquad \text{for a.e. } \xi \in \mathbb{R}, \tag{2.5}$$

with $\psi_1 \in C^\infty(\mathbb{R})$ and $\text{supp}(\hat{\psi}_1) \subseteq [-\frac{1}{2}, -\frac{1}{16}] \cup [\frac{1}{16}, \frac{1}{2}]$, and $\psi_2 \in L^2(\mathbb{R})$ is a bump function in the sense that

$$\sum_{k=-1}^{1} |\hat{\psi}_2(\xi + k)|^2 = 1 \qquad \text{for a.e. } \xi \in [-1, 1], \tag{2.6}$$

satisfying $\psi_2 \in C^\infty(\mathbb{R})$ and $\text{supp}(\hat{\psi}_2) \subseteq [-1, 1]$. Then ψ is called a *Classical Shearlet*.
Thus, a classical shearlet ψ is a function which is wavelet-like along one axis and bump-like along another one.

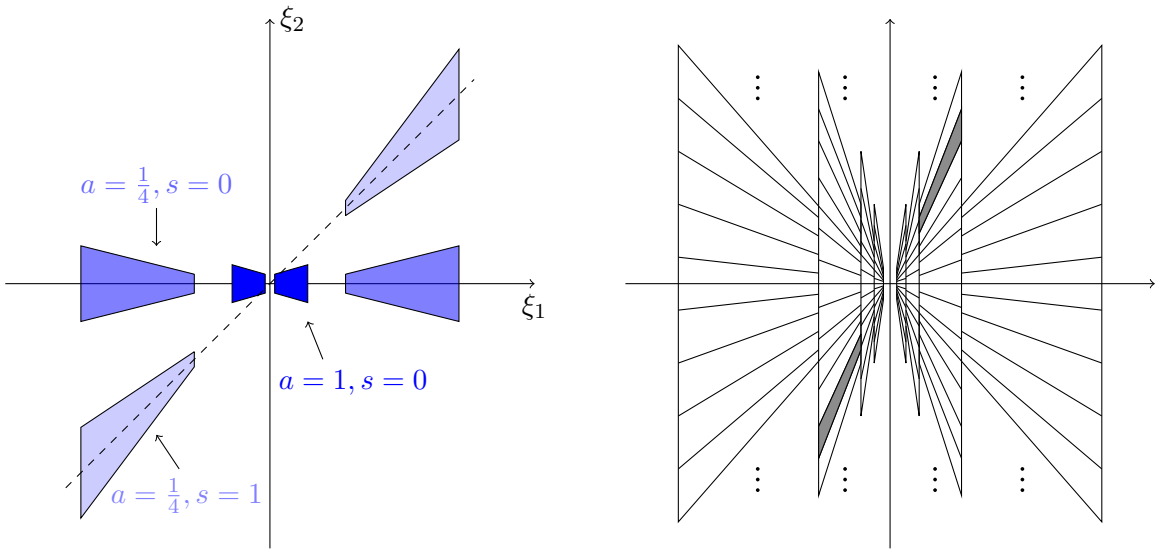


Figure 2.2: Classical shearlets. Fourier domain support of several elements of the shearlet system, for different values of a and s.

As illustrated in Figure 2.2, each element $\psi_{a,s,t}$ of Classical Shearlet System has frequency support on a pair of trapezoids, symmetric with respect to the origin, oriented along a line of slope s. The support becomes increasingly thin as $a \to 0$. Indeed, observe that

$$\text{supp}(\hat{\psi}_{a,s,t}) \subset \left\{(\xi_1, \xi_2) : \xi_1 \in \left[-\frac{1}{2a}, -\frac{1}{16a}\right] \cup \left[\frac{1}{16a}, \frac{1}{2a}\right], \left|\frac{\xi_2}{\xi_1} + s\right| \leq \sqrt{a}\right\}$$

We stress that the specific choices for the supports of the functions ψ_1, ψ_2 have no deeper meaning. Indeed, all that is needed (for the detection of anisotropic structures) is that $\hat{\psi}_1$ is supported away from zero (*i.e.*, ψ_1 is a wavelet) and $\hat{\psi}_2$ is supported around zero (cf. [25]).
Notice that there exist several choices of ψ_1 and ψ_2 satisfying conditions (2.5) and (2.6). One possible choice is to set ψ_1 to be a Lemariè–Meyer wavelet and ψ_2 to be a spline.
Finally, it is not difficult to prove that the Shearlet Transform associated with a classical shearlet $\psi \in L^2(\mathbb{R}^2)$ is an isometry since $c_\psi^- = c_\psi^+ = 1$.

Due to the admissibility condition, it is possible to obtain an inversion formula for the Shearlet Transform.

Theorem 2.11. *Suppose $\psi \in L^2(\mathbb{R}^2)$ is admissible with $c_\psi^- = c_\psi^+ = 1$. Let $\{\rho_n\}_{n=1}^\infty$ be an approximate identity such that $\rho_n \in L^2(\mathbb{R}^2)$ and $\rho_n(x) = \rho_n(-x)$ for all x. Then, for all*

$f \in L^2(\mathbb{R}^2)$ *it is* $\lim_{n\to\infty} ||f - f_n||_2 = 0$, *where*

$$f_n(x) = \int_{\mathbb{S}} \mathscr{SH}_\psi f(a,s,t)\,(\rho_n * \psi_{a,s,t})(x)\; dt\, ds\, \frac{da}{a^3}.$$

Proof. Since ρ_n is even and and the Shearlet Transform is an isometry, we obtain

$$\begin{aligned}
(f * \rho_n)(x) &= \int_{\mathbb{R}^2} f(y)\rho_n(x-y)\, dy \\
&= \langle f, \overline{T_x \rho_n}\rangle \\
&= \langle \mathscr{SH}_\psi f, \mathscr{SH}_\psi(\overline{T_x \rho_n})\rangle \\
&= \int_{\mathbb{S}} \mathscr{SH}_\psi f(a,s,t)\overline{\langle \overline{\rho_n(\cdot - x)}, \psi_{a,s,t}(\cdot)\rangle}\, dt\, ds\, \frac{da}{a^3} \\
&= \int_{\mathbb{S}} \mathscr{SH}_\psi f(a,s,t)\,(\rho_n * \psi_{a,s,t})(x)\; dt\, ds\, \frac{da}{a^3}
\end{aligned}$$

Now, $\{\rho_n\}_{n=1}^{\infty}$ is an approximate identity, thus $\lim_{n\to\infty} ||f - f_n||_2 = 0$. □

2.2 Cone-Adapted Continuous Shearlet Systems

Although the Continuous Shearlet Systems defined above exhibit an elegant group structure, there exist a directional bias related to the shear parameter. It is easy to see that the distribution of directions becomes infinitely dense as s grows (cf. Figure 2.2). Indeed, consider a function f which is mostly concentrated along the ordinate axis in the frequency domain: it is clear that the energy of f is more and more concentrated in the shearlet components $\mathcal{SH}_\psi f(a,s,t)$ as $s \to \infty$. This behaviour can be a serious limitation for some applications.
One way to address this problem is to partition the frequency domain into four cones $\mathscr{C}_i$, $i = 1,\ldots,4$, while separating the low-frequency region by cutting out $\mathscr{R} = \{(\xi_1,\xi_2) : |\xi_1|, |\xi_2| \le 1\}$, *i.e.*, a square centered around the origin. This yields a partition of the frequency plane as illustrated in Figure 2.3, that leads to the definition of Continuous Shearlet on the cone, a variant of Continuous Shearlet System.

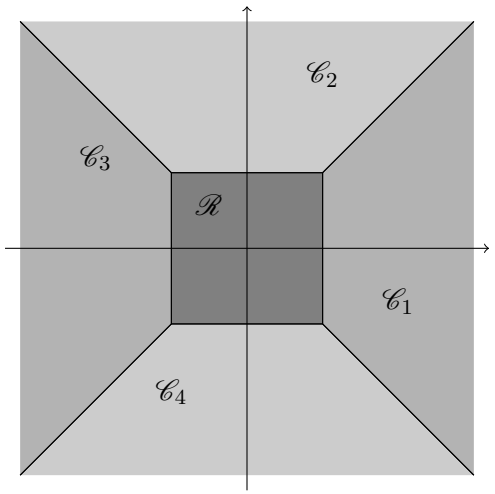


Figure 2.3: Resolving the problem of biased treatment of directions by continuous shearlet systems. The frequency plane is partitioned into four cones $\mathscr{C}_i$, $i = 1,\ldots,4$, and the low frequency box $\mathscr{R} = \{(\xi_1,\xi_2) : |\xi_1|, |\xi_2| \le 1\}$.

Definition 2.12. For $\phi, \psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$, the *Cone-Adapted Continuous Shearlet System* $\mathcal{SH}(\phi, \psi, \tilde{\psi})$ is defined by

$$\mathcal{SH}(\phi, \psi, \tilde{\psi}) = \Phi(\phi) \cup \Psi(\psi) \cup \tilde{\Psi}(\tilde{\psi}),$$

where

$$\begin{aligned}
\Phi(\phi) &= \{\phi_t = \phi(\cdot - t) : t \in \mathbb{R}^2\}, \\
\Psi(\psi) &= \{\psi_{a,s,t} = a^{-\frac{3}{4}} \psi(A_a^{-1} S_s^{-1}(\cdot - t)) : a \in (0,1], |s| \leq 1 + \sqrt{a}, t \in \mathbb{R}^2\}, \\
\tilde{\Psi}(\tilde{\psi}) &= \{\tilde{\psi}_{a,s,t} = a^{-\frac{3}{4}} \psi(\tilde{A}_a^{-1} S_s^{-T}(\cdot - t)) : a \in (0,1], |s| \leq 1 + \sqrt{a}, t \in \mathbb{R}^2\},
\end{aligned}$$

and $\tilde{A}_a = \text{diag}(a^{1/2}, a)$.

Notice that, within each cone, the shearing variable s is only allowed to vary over a finite range and thus we are able to detect only a certain subset of all possible directions.
Furthermore, the function ϕ will be chosen to have compact frequency support near the origin, which ensures that the system $\Phi(\phi)$ is associated with the low frequency region $\mathscr{R}$.

EXAMPLE 2.13. If ψ is a Classical Shearlet, the system $\Psi(\psi)$ is associated with the horizontal cones $\mathscr{C}_1 \cup \mathscr{C}_3 = \{(\xi_1, \xi_2) : |\frac{\xi_2}{\xi_1}| \leq 1, |\xi_1| > 1\}$. The shearlet $\tilde{\psi}$ can be chosen likewise with the roles of ξ_1 and ξ_2 reversed, *i.e.*, $\tilde{\psi}(\xi_1, \xi_2) = \psi(\xi_2, \xi_1)$. Then the system $\tilde{\Psi}(\tilde{\psi})$ is associated with the vertical cones $\mathscr{C}_2 \cup \mathscr{C}_4 = \{(\xi_1, \xi_2) : |\frac{\xi_2}{\xi_1}| > 1, |\xi_2| > 1\}$.

The previous example suggest that we can imagine the function f splitted into $f = P^{(h)}f + P^{(v)}f$, where $P^{(h)}$ is the frequency projection onto the cone with slope $s \leq 1$ and $P^{(v)}$ is the frequency projection onto the cone with slope $\frac{1}{s} \leq 1$. Thus, fixed a shearlet $\psi^{(h)}$, we can analyze only $P^{(h)}f$ while $P^{(v)}f$ is analyzed defining $\psi^{(v)}(\xi_1, \xi_2) = \psi^{(h)}(\xi_2, \xi_1)$.

Similar to the situation of Continuous Shearlet systems, an associated transform can be defined for Cone-adapted Continuous Shearlet Systems.

Definition 2.14. Set

$$\mathbb{S}_{\text{cone}} = \{(a, s, t) : a \in (0,1], |s| \leq 1 + \sqrt{a}, t \in \mathbb{R}^2\}.$$

Then, for $\phi, \psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$, the *Cone-Adapted Continuous Shearlet Transform* of $f \in L^2(\mathbb{R}^2)$ is the map

$$f \to \mathscr{SH}_{\phi,\psi,\tilde{\psi}} f(t', (a, s, t), (\tilde{a}, \tilde{s}, \tilde{t})) = (\langle f, \varphi_t \rangle, \langle f, \psi_{a,s,t} \rangle, \langle f, \tilde{\psi}_{\tilde{a},\tilde{s},\tilde{t}} \rangle)$$

where

$$(t', (a, s, t), (\tilde{a}, \tilde{s}, \tilde{t})) \in \mathbb{R}^2 \times \mathbb{S}_{\text{cone}}^2.$$

A similar argument to the one used in the proof of Theorem 2.6 can be used to show that the map $\mathscr{SH}_{\phi,\psi,\tilde{\psi}}$ is an isometry, under suitable conditions on ϕ, ψ and $\tilde{\psi}$.

Theorem 2.15. *Retaining the notation of Theorem 2.6, let $\psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$ be admissible shearlets satisfying $c_\psi^- = c_\psi^+ = 1$ and $c_{\tilde{\psi}}^- = c_{\tilde{\psi}}^+ = 1$, respectively, and let $\phi \in L^2(\mathbb{R}^2)$ be such that, for a.e. $\xi = (\xi_1, \xi_2) \in \mathbb{R}^2$,*

$$|\hat{\phi}(\xi)|^2 + \chi_{\mathscr{C}_1 \cup \mathscr{C}_3}(\xi) \int_0^1 |\hat{\psi}_1(a\xi_1)|^2 \frac{da}{a} + \chi_{\mathscr{C}_2 \cup \mathscr{C}_4}(\xi) \int_0^1 |\hat{\psi}_1(a\xi_2)|^2 \frac{da}{a} = 1.$$

Then, for each $f \in L^2(\mathbb{R}^2)$,

$$||f||^2 = \int_{\mathbb{R}} |\langle f, T_t \phi \rangle|^2 \, dt + \int_{\mathbb{S}_{\text{cone}}} |\langle (\hat{f} \chi_{\mathscr{C}_1 \cup \mathscr{C}_3})^\vee, \psi_{a,s,t} \rangle|^2 \frac{da}{a^3} \, ds \, dt + \int_{\mathbb{S}_{\text{cone}}} |\langle (\hat{f} \chi_{\mathscr{C}_2 \cup \mathscr{C}_4})^\vee, \tilde{\psi}_{\tilde{a},\tilde{s},\tilde{t}} \rangle|^2 \frac{d\tilde{a}}{\tilde{a}^3} \, d\tilde{s} \, d\tilde{t}$$

Notice that the functions ϕ, ψ and $\tilde{\psi}$ can be chosen to be in $C_c^\infty(\mathbb{R}^2)$. In addition, the cone-adapted shearlet system can be designed so that the low frequency and high frequency parts are smoothly combined.
A more detailed analysis of the Cone-Adapted Continuous Shearlet Transform and its generalizations can be found in [16, 17].

2.3 Resolution of edges using the Continuous Shearlet Systems

We already stressed that, if ψ is a 'nice' continuous wavelet, then the Continuous Wavelet Transform $\mathcal{W}_\psi f(a,t)$ is able to localize the singularities of f, *i.e.*, for $a \to 0$, the function $\mathcal{W}_\psi f(a,t)$ tends rapidly to zero when t is outside the singularity and $\mathcal{W}_\psi f(a,t)$ tends to zero slowly when t is at the singularity.
One major property of Continuous Shearlets is their ability to resolve the discontinuities of 2D functions by identifying not only the location, but also the orientation of the discontinuity. More precisely, if f is a 2D function that is smooth away from a discontinuity along a curve Γ, then for $a \to 0$ the Continuous Shearlet Transform satisfies

$$|\mathcal{SH}_\psi f(a,s,t)| \leq k\, a^n, \qquad \text{for each } n = 1, 2, \ldots,$$

unless t is at the singularity and s describes the direction that is perpendicular to the discontinuity curve.

Furthermore, the Cone-Adapted Continuous Shearlet Transform can be used to provide a precise characterization of edge-discontinuities of functions of two variables. In particular, consider a function $f = \chi_B \subset L^2(\mathbb{R}^2)$, where $B \subset \mathbb{R}^2$ is a planar region with piecewise smooth boundary. Then $\mathcal{SH}_{\phi,\psi,\tilde{\psi}} f$ characterizes both the location and orientation of the boundary edge ∂B by its decay at fine scales. This property is very useful in applications which require the analysis or detection of edge discontinuities. For example, using these observations, a shearlet-based algorithm for edge detection and analysis was developed and related ideas were exploited to develop algorithms for the regularized inversion of the Radon transform in [7].
A more detailed discussion of these issues, including the extensions to higher dimensions, can be found in [24].

Chapter 3

Discrete Shearlet Systems

Starting from continuous shearlet systems as defined in (2.1), several discrete versions of shearlet systems can be constructed by an appropriate sampling of the continuous parameter set $\mathbb{S}$ or $\mathbb{S}_{\text{cone}}$. Various approaches have been suggested, aiming for discrete shearlet systems which preferably form an orthonormal basis or a tight frame for $L^2(\mathbb{R}^2)$.
One approach applies a powerful methodology called *Coorbit Theory*, which is used to derive different discretizations while ensuring frame properties. In particular, the regular shearlet frame which will be introduced in the next subsection can be derived using this machinery. More details about this topic can be found in [9, 10]. A different path derives sufficient condition studying t_q*-equations* from the theory of wavelets. These equations are part of the sufficient conditions needed for an affine system to form a wavelet orthonormal basis or a tight frame ([20]). Thus, due to the close relationship between shearlet systems and affine systems, this approach can be transferred to the situation of cone-adapted continuous shearlet systems ([25]).

3.1 Discrete Shearlet Systems and Transforms

Discrete Shearlet Systems are formally defined by sampling Continuous Shearlet Systems on a discrete subset of the shearlet group $\mathbb{S}$. The following definition is taken from [25].

Definition 3.1. Let $\psi \in L^2(\mathbb{R}^2)$ and $\Lambda \subseteq \mathbb{S}$. An *irregular discrete shearlet system*, associated with ψ and Λ and denoted by $\mathcal{SH}(\psi, \Lambda)$, is defined by

$$\mathcal{SH}(\psi, \Lambda) = \{\psi_{a,s,t} = a^{-\frac{3}{4}}\psi(A_a^{-1}S_s^{-1}(\cdot - t)) : (a, s, t) \in \Lambda\}.$$

A *(regular) discrete shearlet system*, associated with ψ and denoted by $\mathcal{SH}(\psi)$, is defined by

$$\mathcal{SH}(\psi) = \{\psi_{j,k,m} = 2^{\frac{3}{4}j}\psi(S_k A_{2^j} \cdot -m) : j, k \in \mathbb{Z}, m \in \mathbb{Z}^2\}.$$

Observe that the regular versions of discrete shearlet systems are derived from the irregular systems by choosing

$$\Lambda = \{(2^{-j}, -k2^{-j/2}, S_{-k2^{-j/2}}A_{2^{-j}}m) : j, k \in \mathbb{Z}, m \in \mathbb{Z}^2\}.$$

Indeed, keeping the greatest generality, we can choose an arbitrary set of scales $\{a_j\}_{j\in\mathbb{Z}} \subset \mathbb{R}^+$; next, we pick the shear parameters $\{s_{j,k}\}_{k\in\mathbb{Z}} \subset \mathbb{R}$ dependent on j, so that the directionality of the representation is allowed to change with the scale. Then, in order to provide a "uniform covering", we allow the location parameter to describe a different grid depending on j, and hence, on k: we can select $t_{j,k,m} = S_{s_{j,k}}A_{a_j}m$, where $m \in \mathbb{Z}^2$. Finally, the translation parameter can be chosen to belong to $c_1\mathbb{Z} \times c_2\mathbb{Z}$ for some $(c_1, c_2) \in (\mathbb{R}^+)^2$. This provides some additional flexibility which is useful for some constructions.
As far as the previous definition concerns, we used the dyadic sampling for the scaling parameter,

i.e., $a_j = 2^{-j},\ j \in \mathbb{Z}$. Next, we set $s_{j,k} = -k2^{-j/2},\ k \in \mathbb{Z}$, in order to get a larger number of directions as j is getting smaller. Finally, the location parameter was determined by adjusting the canonical grid $\mathbb{Z}^2$ to the particular scaling and shear parameter, *i.e.*, we choose $t_{j,k,m} = S_{s_{j,k}} A_{a_j} m = S_{-k2^{-j/2}} A_{2^{-j}} m,\ m \in \mathbb{Z}^2$. Combining all this and observing that $A^{-1}_{2^{-j}} S^{-1}_{-k2^{-j/2}} = A_{2^j} S_{k2^{-j/2}} = S_k A_{2^j}$ we obtain the regular discrete shearlet system as defined above.

Similarly to the continuous case, we define the Discrete Shearlet Transform, for the regular case, as follows.

Definition 3.2. For $\psi \in L^2(\mathbb{R}^2)$, the *Discrete Shearlet Transform* of $f \in L^2(\mathbb{R}^2)$ is the map defined by

$$f \to \mathscr{SH}_\psi f(j,k,m) = \langle f, \psi_{j,k,m} \rangle, \qquad (j,k,m) \in \mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}^2.$$

Thus, $\mathscr{SH}_\psi$ maps the function f to the coefficients $\mathscr{SH}_\psi f(j,k,m)$ associated with the scale index j, the orientation index k, and the position index m. Clearly, the previous definition can be extended to the irregular shearlet systems in the most natural way.

Now, since our aim is to apply shearlet systems as analysis and synthesis tools, we need to derive the condition under which a discrete shearlet system $\mathcal{SH}(\psi)$ forms a basis or, more generally, a frame. Similar to the wavelet case, we are particularly interested in selecting a generator ψ with special properties, *e.g.*, regularity, vanishing moments, and compact support, so that the corresponding basis or frame of shearlets has satisfactory approximation properties. Particularly useful examples are the classical shearlets since, as the following result shows, they generate a Parseval frames for $L^2(\mathbb{R}^2)$.

Proposition 3.3. *Let $\psi \in L^2(\mathbb{R}^2)$ be a classical shearlet. Then $\mathcal{SH}(\psi)$ is a Parseval frame for $L^2(\mathbb{R}^2)$.*

Proof. We recall that $\mathcal{SH}(\psi)$ is a Parseval frame for $L^2(\mathbb{R}^2)$ if

$$\sum_{j,k,m} |\langle f, \psi_{j,k,m} \rangle|^2 = ||f||^2 \qquad \text{for all } f \in L^2(\mathbb{R}^2).$$

Using (2.5) and (2.6) from the definition of classical shearlet, we obtain

$$\begin{aligned}
\sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} |\hat{\psi}(S^T_{-k} A_{2^{-j}} \xi)|^2 &= \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} |\hat{\psi}_1(2^{-j}\xi_1)|^2 \left|\hat{\psi}_2\left(2^{\frac{j}{2}} \frac{\xi_2}{\xi_1} - k\right)\right|^2 \\
&= \sum_{j\in\mathbb{Z}} |\hat{\psi}_1(2^{-j}\xi_1)|^2 \sum_{k\in\mathbb{Z}} \left|\hat{\psi}_2\left(2^{\frac{j}{2}} \frac{\xi_2}{\xi_1} + k\right)\right|^2 \\
&= 1
\end{aligned}$$

The claim follows immediately from the previous computation by using Theorem 3.1 in [23]. □

We already stressed that a classical shearlet ψ is a well localized function. Thus, the previous proposition implies that there exist Parseval frames $\mathcal{SH}(\psi)$ of well localized discrete shearlets. The well localization property is critical for deriving superior approximation properties for shearlet systems and will be required for deriving optimally sparse approximations.
Removing the assumption of well localization for ψ, it is possibile to construct discrete shearlet systems which form not only tight frames but also orthonormal bases. Unfotunately, it seems that well localized shearlet orthonormal bases does not exist. Indeed, in a recent work, it has been shown that a well localized discrete shearlet system can form a frame or a tight frame but not an orthonormal basis ([25]).
Finally, we observe that the shearlet systems generated by classical shearlets are band-limited,

i.e., they have compact support in the frequency domain and, hence, cannot be compactly supported in the spatial domain. However, to achieve spatial domain localization, compactly supported discrete shearlet systems are required. We will not focus on this topic in this presentation.

3.2 Cone-Adapted Discrete Shearlet Systems and Transforms

As one can imagine, also discrete shearlet systems suffer from a biased treatment of the directions. As in the situation of continuous shearlet systems, this problem can be addressed by dividing the frequency plane into cones.
Let us start by defining cone-adapted discrete shearlet systems with respect to an irregular parameter set: obviously, the group to sample will be $\mathbb{S}_{\text{cone}}$.

Definition 3.4. Let $\phi, \psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$, $\Delta \subset \mathbb{R}^2$, and $\Lambda, \tilde{\Lambda} \subset \mathbb{S}_{\text{cone}}$. Then the *irregular cone-adapted discrete shearlet system* $\mathcal{SH}(\varphi, \psi, \tilde{\psi}; \Delta, \Lambda, \tilde{\Lambda})$ is defined by

$$\mathcal{SH}(\phi, \psi, \tilde{\psi};\, \Delta, \Lambda, \tilde{\Lambda}) = \Phi(\phi; \Delta) \cup \Psi(\psi; \Lambda) \cup \tilde{\Psi}(\tilde{\psi}; \tilde{\Lambda})$$

where

$$\begin{aligned}
\Phi(\phi, \Delta) &= \{\phi_t = \phi(\cdot - t) : t \in \Delta\},\\
\Psi(\psi, \Lambda) &= \{\psi_{a,s,t} = a^{-\frac{3}{4}}\, \psi(A_a^{-1} S_s^{-1}(\cdot - t)) : (a, s, t) \in \Lambda\},\\
\tilde{\Psi}(\tilde{\psi}, \tilde{\Lambda}) &= \{\tilde{\psi}_{a,s,t} = a^{-\frac{3}{4}}\, \tilde{\psi}(\tilde{A}_a^{-1} S_s^{-T}(\cdot - t)) : (a, s, t) \in \tilde{\Lambda}\}.
\end{aligned}$$

The regular variant of the cone-adapted discrete shearlet systems is much more frequently used. To allow more flexibility and enable changes to the density of the translation grid, one can introduce a sampling factor $c = (c_1, c_2) \in (\mathbb{R}^+)^2$ in the translation index. Thus, we have the following definition.

Definition 3.5. For $\phi, \psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$ and $c = (c_1, c_2) \in (\mathbb{R}^+)^2$, the *(regular) cone-adapted discrete shearlet system* $\mathcal{SH}(\varphi, \psi, \tilde{\psi}; c)$ is defined by

$$\mathcal{SH}(\varphi, \psi, \tilde{\psi}; c) = \Phi(\phi; c_1) \cup \Psi(\psi; c) \cup \tilde{\Psi}(\tilde{\psi}; c),$$

where

$$\begin{aligned}
\Phi(\phi; c_1) &= \{\phi_m = \phi(\cdot - c_1 m) : m \in \mathbb{Z}^2\},\\
\Psi(\psi; c) &= \{\psi_{j,k,m} = 2^{\frac{3}{4}j}\, \psi(S_k A_{2^j} \cdot - M_c m) : j \geq 0, |k| \leq \lceil 2^{j/2} \rceil, m \in \mathbb{Z}^2\},\\
\tilde{\Psi}(\tilde{\psi}; c) &= \{\tilde{\psi}_{j,k,m} = 2^{\frac{3}{4}j}\, \tilde{\psi}(S_k^T \tilde{A}_{2^j} \cdot - \tilde{M}_c m) : j \geq 0, |k| \leq \lceil 2^{j/2} \rceil, m \in \mathbb{Z}^2\}.
\end{aligned}$$

with

$$M_c = \begin{pmatrix} c_1 & 0 \\ 0 & c_2 \end{pmatrix} \qquad \text{and} \qquad \tilde{M}_c = \begin{pmatrix} c_2 & 0 \\ 0 & c_1 \end{pmatrix}.$$

Clearly, if $c = (1, 1)$, the parameter c can be omitted in the formulae above.

The generating functions ϕ will be referred to as *shearlet scaling functions* and the generating functions $\psi, \tilde{\psi}$ as *shearlet generators*. As in the continuous case, the system $\Phi(\phi; c_1)$ is associated with the low frequency region, and the systems $\Psi(\psi; c)$ and $\tilde{\Psi}(\tilde{\psi}; c)$ are associated with the conic regions $\mathscr{C}_1 \cup \mathscr{C}_3$ and $\mathscr{C}_2 \cup \mathscr{C}_4$, respectively (cf. Figure 2.3).

Even for the cone-adapted case, we give the definition just of shearlet transform associated with the regular cone-adapted discrete shearlet systems. The extension to the irregular case is done, again, in the most natural way.

Definition 3.6. Set $\Lambda = \mathbb{N}_0 \times \{-\lceil 2^{j/2} \rceil, \dots, \lceil 2^{j/2} \rceil\} \times \mathbb{Z}^2$. For $\phi, \psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$, the *Cone-Adapted Discrete Shearlet Transform* of $f \in L^2(\mathbb{R}^2)$ is the map defined by

$$f \to \mathscr{SH}_{\phi,\psi,\tilde{\psi}} f(m', (j,k,m), (\tilde{j}, \tilde{k}, \tilde{m})) = (\langle f, \phi_{m'} \rangle, \langle f, \psi_{j,k,m} \rangle, \langle f, \tilde{\psi}_{\tilde{j},\tilde{k},\tilde{m}} \rangle),$$

where

$$(m', (j,k,m), (\tilde{j}, \tilde{k}, \tilde{m})) \in \mathbb{Z}^2 \times \Lambda \times \Lambda.$$

We already stressed that it seems impossible to construct a discrete shearlet orthonormal basis. Hence, our goal is to derive Parseval frames.
First of all, in the next proposition we prove that a classical shearlet is a shearlet generator of a Parseval frame for the subspace of $L^2(\mathbb{R}^2)$ of functions whose frequency support lies in the union of two cones $\mathscr{C}_1 \cup \mathscr{C}_3$.

Proposition 3.7. *Let $\psi \in L^2(\mathbb{R}^2)$ be a classical shearlet. Then the shearlet system*

$$\Psi(\psi) = \{\psi_{j,k,m} = 2^{\frac{3}{4}j}\, \psi(S_k A_{2^j} \cdot - m) : j \geq 0, |k| \leq \lceil 2^{j/2} \rceil, m \in \mathbb{Z}^2\}$$

is a Parseval frame for $L^2(\mathscr{C}_1 \cup \mathscr{C}_3)^\vee = \{f \in L^2(\mathbb{R}^2) : \mathrm{supp}(\hat{f}) \subset \mathscr{C}_1 \cup \mathscr{C}_3\}$.

Clearly, a result similar to Proposition 3.7 holds true for the subspace $L^2(\mathscr{C}_2 \cup \mathscr{C}_4)^\vee$ if ψ is replaced by $\tilde{\psi}$. Thus, one can build up a Parseval frame for the whole space $L^2(\mathbb{R}^2)$ by piecing together Parseval frames associated with different cones on the frequency domain together with a coarse scale system which takes care of the low frequency region. From this idea arise the following result.

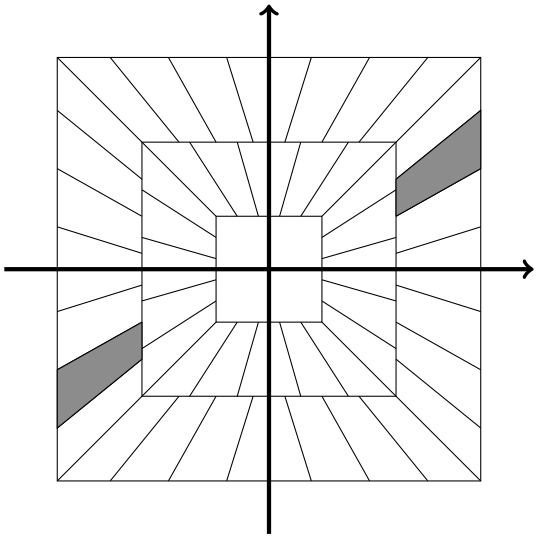


Figure 3.1: Tiling of the frequency plane induced by a cone-adapted Parseval frame of shearlets.

Theorem 3.8. *Let $\psi \in L^2(\mathbb{R}^2)$ be a classical shearlet, and let $\phi \in L^2(\mathbb{R}^2)$ be chosen so that, for a.e. $\xi \in \mathbb{R}^2$,*

$$|\hat{\phi}(\xi)|^2 + \sum_{j \geq 0} \sum_{|k| \leq \lceil 2^{j/2} \rceil} |\hat{\psi}(S_{-k}^T A_{s^{-j}} \xi)|^2 \chi_C + \sum_{j \geq 0} \sum_{|k| \leq \lceil 2^{j/2} \rceil} |\hat{\psi}(S_{-k}^T \tilde{A}_{s^{-j}} \xi)|^2 \chi_{\tilde{C}} = 1.$$

Let $P_C \Psi(\psi)$ denote the set of shearlet elements in $\Psi(\psi)$ after projecting their Fourier transforms onto $C = \left\{(\xi_1, \xi_2) \in \mathbb{R}^2 : \left|\frac{\xi_2}{\xi_1}\right| \leq 1\right\}$, and let $P_{\tilde{C}} \tilde{\Psi}(\tilde{\psi})$, where $\tilde{C} = \mathbb{R}^2 \setminus C$, be defined analogously. Then, the modified cone-adapted discrete shearlet system $\Phi(\phi) \cup P_C \Psi(\psi) \cup P_{\tilde{C}} \tilde{\Psi}(\tilde{\psi})$ is a Parseval frame for $L^2(\mathbb{R}^2)$.

Notice that, despite its simplicity, the Parseval frame construction above has one drawback. When the cone-based shearlet systems are projected onto C and $\tilde{C}$, the shearlet elements overlapping the boundary lines $\xi_1 = \pm\xi_2$ in the frequency domain are cut so that the "boundary" shearlets lose their regularity properties. To avoid this problem, it is possible to redefine the "boundary" shearlets in such a way that their regularity is preserved. This require to slightly modify the definition of the classical shearlet. However, this topic is far from the aim of this introduction.

In conclusion, in Figure 3.1 is illustrated the tiling of the frequency plane induced by a cone-adapted Parseval frame of shearlets.

3.3 Sparse Approximations by Shearlets

We already stressed that one of the main motivations for the introduction of the shearlet framework is the derivation of optimally sparse approximations of multivariate functions.
More precisely, let f be a 2-D function that is $\mathcal{C}_2$ apart from discontinuities along a $\mathcal{C}_2$ curve. Then, denoting by f_N the approximation obtained by taking the best N terms in the discrete shearlets expansion of f, the error $\|f - f_N\|^2$ of such approximation decays asymptotically as $\mathcal{O}(N^{-2}(\log N)^3)$, as $N \to \infty$. Since a log-like factor is negligible with respect to the other terms for large N, the optimal error decay rate is essentially achieved. Indeed, the following result holds true.

Theorem 3.9. *Let $\Phi(\phi) \cup P_C\Psi(\psi) \cup P_{\tilde{C}}\tilde{\Psi}(\tilde{\psi})$ be a Parseval frame for $L^2(\mathbb{R}^2)$ as defined in Theorem 3.8, where $\psi \in L^2(\mathbb{R}^2)$ is a classical shearlet and $\hat{\psi} \in C_0^\infty(\mathbb{R}^2)$. Let f be a cartoon-like image and f_N be its nonlinear N-term approximation obtained by selecting the N largest coefficients in the expansion of f with respect to this shearlet system. Then there exists a constant $C > 0$, independent of f and N, such that*

$$\|f - f_N\|^2 \leq C\, N^{-2}(\log N)^3 \qquad as \quad N \to \infty.$$

It has been shown that the same error decay rate is achieved also using approximations based on compactly supported shearlet frames.

Chapter 4

Multidimensional Extensions

In the last decades, scientists faced a rapidly growing deluge of data, which are becoming increasingly complex and higher dimensional and, thus, require highly sophisticated methodologies for analysis and compression.
In particular, many current high-impact applications require to deal with multidimensional data, especially 3-D data, such as seismic or biological ones. The computational challenges in this setting are much more demanding than in 2-D case, and sparse approximations are strongly needed. Due to the simplicity of the mathematical structure of shearlets, their extensions to higher dimensions is quite natural. Some basic ideas were already introduced in [19], where it was observed that there exist several ways to extend the shearing matrix to larger dimensions. Several other results, such as the extension of the optimally sparse approximation results and the analysis and detection of surface singularities, recently appeared (see [25] and the references therein).

We work out the fundamental results in the greatest generality and then we focus on the 3-D case. Indeed, the two crucial situations are the 2-D and the 3-D cases, because all the others derive from the analysis of these two. Once it is known how to handle anisotropic features of different dimensions, the step from 3-D to 4-D can be dealt with in a similar way. The same happens for the extension to even larger dimensions.

4.1 Multivariate Continuous Shearlet Transform

To analyze data in $\mathbb{R}^d$, $d \geq 3$, we have to generalize the Shearlet Transform to higher dimensions. The first step toward this end is the identification of a suitable shear matrix and the generalization of the parabolic dilation matrix.
In the following, let $\mathbb{1}_d$ denote the d-dimensional identity matrix and $\mathbf{0}_d$, respectively $\mathbf{1}_d$, the vectors with d entries equal to 0, respectively 1.
Given a d-dimensional vector space V and a k-dimensional subspace W of V, a reasonable model for the shear matrix is the following: the shear should fix the space W and translate all vectors parallel to W. Hence, for $V = W \bigoplus W'$ and $v = w + w'$, the shear operation S can be described as

$$S(v) = w + (w' + s(w')) = (w + s(w')) + w',$$

where s is a linear mapping from W' to W. Then, with respect to an appropriate basis of V, the shear operation S corresponds to a block matrix of the form

$$S = \begin{pmatrix} \mathbb{1}_k & s^T \\ \mathbf{0}_{(d-k)\times k} & \mathbb{1}_{d-k} \end{pmatrix}, \qquad s \in \mathbb{R}^{(d-k)\times k}.$$

Now, the crucial point is how to choose the block s, since we want to end up with a square integrable group representation. Usually, the number of parameters has to fit together with the

space dimension, otherwise the resulting group would be either too large or too small. Since we have d degrees of freedom related with the translates and 1 degree of freedom related with the dilation, $d-1$ degrees of freedom for the shear component would be optimal. Therefore, one natural choice would be $s \in \mathbb{R}^{(d-1)\times 1}$, *i.e.*, $k = 1$:

$$S = \begin{pmatrix} 1 & s^T \\ \mathbf{0}_{d-1} & \mathbb{1}_{d-1} \end{pmatrix}, \qquad s \in \mathbb{R}^{d-1}.$$

We will prove that, with this choice, the associated multivariate shearlet transform can be interpreted as a square integrable group representation of a $(2d)$-parameter group, that we will refer to as the *full shearlet group*.
As far as the dilation matrix concerns, it should depend on a parameter $a \in \mathbb{R}^* = \mathbb{R} \setminus \{0\}$ in such a way that

$$A_a = \operatorname{diag}(a_1(a), \dots, a_d(a)),$$

where $a_1(a) = a$ and $a_j(a) = a^{\alpha_j}$ with $\alpha_j \in (0,1)$, $j = 2, \dots, d$. In order to have directional selectivity, the dilation factors on the diagonal of A_a should be chosen in an anisotropic way, *i.e.*, $|a_j(a)|$, $j = 2, \dots, d$, should increase less than linearly in a as $a \to \infty$. Our choice is

$$A_a = \begin{pmatrix} a & \mathbf{0}_{d-1}^T \\ \mathbf{0}_{d-1} & \operatorname{sgn}(a)\,|a|^{\frac{1}{d}}\,\mathbb{1}_{d-1} \end{pmatrix}.$$

As in the 2-D case, we need to combine dilation matrices and shear matrices. A straightforward computation shows that:

$$S_s^{-1} = \begin{pmatrix} 1 & -s^T \\ \mathbf{0}_{d-1} & \mathbb{1}_{d-1} \end{pmatrix} \qquad \text{and} \qquad S_s A_a S_{s'} A_{a'} = S_{s+|a|^{1-\frac{1}{d}} s'} A_{aa'}.$$

Lemma 4.1. *The set $\mathbb{R}^* \times \mathbb{R}^{d-1} \times \mathbb{R}^d$ endowed with the operation*

$$(a,\, s,\, t) \circ (a',\, s',\, t') = (aa',\, s + |a|^{1-\frac{1}{d}} s',\, t + S_s A_a t')$$

is a locally compact group $\mathbb{S}$. The left Haar measures on $\mathbb{S}$ is given by

$$d\mu(a,\, s,\, t) = d\mu_l(a,\, s,\, t) = \frac{1}{|a|^{d+1}} da\, ds\, dt$$

Proof. See [10, 11]. □

For $f \in L^2(\mathbb{R}^d)$ we define

$$\pi(a,\, s,\, t) f(x) = f_{a,s,t}(x) = |\det A_a|^{-\frac{1}{2}}\, f(A_a^{-1} S_s^{-1}(x-t)) = |a|^{\frac{1}{2d}-1}\, f(A_a^{-1} S_s^{-1}(x-t))$$

It is easy to check that $\pi : \mathbb{S} \to \mathcal{U}(L^2(\mathbb{R}^d))$ is a map from $\mathbb{S}$ into the group $\mathcal{U}(L^2(\mathbb{R}^d))$ of unitary operators on $L^2(\mathbb{R}^d)$ and, what is more important, the following lemma holds true.

Lemma 4.2. *The map π defined above is a unitary representation of $\mathbb{S}$.*

Proof. See [10, 11]. □

We recall that a nontrivial function $\psi \in L^2(\mathbb{R}^d)$ is called *admissible* if

$$\int_{\mathbb{S}} |\langle \psi, \pi(a,\, s,\, t)\psi \rangle|^2 \, d\mu(a,\, s,\, t) < \infty.$$

If π is irreducible and there exits at least one admissible function $\psi \in L^2(\mathbb{R}^d)$, then π is called *square integrable*. The following result shows that the unitary representation π defined above is, actually, square integrable.

Theorem 4.3. *A function $\psi \in L^2(\mathbb{R}^d)$ is admissible if and only if it fulfills the admissibility condition*

$$C_\psi = \int_{\mathbb{R}^d} \frac{|\hat{\psi}(\omega)|^2}{|\omega_1|^d}\, d\omega < \infty.$$

If ψ is admissible, then, for any $f \in L^2(\mathbb{R}^d)$, the following equality holds true:

$$\int_{\mathbb{S}} |\langle f, \psi_{a,s,t}\rangle|^2 \, d\mu(a,\, s,\, t) = C_\psi \, \|f\|^2_{L^2(\mathbb{R}^d)}.$$

In particular, the unitary representation π is irreducible and hence square integrable.

Proof. See [10, 11]. □

A function $\psi \in L^2(\mathbb{R}^d)$ that fulfills the admissibility condition is called *continuous shearlet*, the transform $\mathscr{SH}_\psi : L^2(\mathbb{R}^d) \to L^2(\mathbb{S})$ such that

$$\mathscr{SH}_\psi f(a, s, t) = \langle f, \psi_{a,s,t}\rangle = (f * \psi^*_{a,s,0})(t),$$

is called *continuous shearlet transform* and $\mathbb{S}$ defined in Lemma (4.1) is called *full shearlet group.*

Example 4.4. An example of a continuous shearlet can be constructed using the same idea shown in the 2-D case. Let ψ_1 be an admissible wavelet with $\hat{\psi}_1 \in C^\infty(\mathbb{R})$ and $\mathrm{supp}(\hat{\psi}_1) \subseteq [-2,-1] \cup [1,2]$, and let ψ_2 be such that $\hat{\psi}_2 \in C^\infty(\mathbb{R}^{d-1})$ and $\mathrm{supp}(\hat{\psi}_2) \subseteq [-1,1]^{d-1}$. Then, the function $\psi \in L^2(\mathbb{R}^d)$ defined by

$$\hat{\psi}(\omega) = \hat{\psi}(\omega_1, \tilde{\omega}) = \hat{\psi}_1(\omega_1)\hat{\psi}_2\left(\frac{1}{\omega_1}\,\tilde{\omega}\right)$$

is a continuous shearlet.

Remark 4.5. It is clear from the above that, also in the multidimensional case, the continuous shearlet transform is nothing else but the continuous wavelet transform associated with a special subgroup $\mathbb{A}_d^S$ of the affine group $\mathbb{A}_d$. For a fixed $\alpha = (\alpha_1, \ldots, \alpha_{d-1})$ where $\alpha_j \in (0,1)$, $1 \le j \le d-1$, $\mathbb{A}_d^S$ consists of the elements $(M_{a,s}, t)$, where

$$M_{a,s} = \begin{pmatrix} a & a^{\alpha_1}s_1 & \ldots & a^{\alpha_{d-1}}s_{d-1} \\ 0 & a^{\alpha_1} & \ldots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \ldots & a^{\alpha_{d-1}} \end{pmatrix}$$

$a > 0$, $s = (s_1, \ldots, s_{d-1}) \in \mathbb{R}^{d-1}$, and $t \in \mathbb{R}^d$. Clearly, each matrix $M_{a,s}$ is the product of the shear matrix S_s and the dilation matrix A_a:

$$S_s = \begin{pmatrix} 1 & s_1 & \ldots & s_{d-1} \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \ldots & 1 \end{pmatrix} \qquad A_a = \begin{pmatrix} a & 0 & \ldots & 0 \\ 0 & a^{\alpha_1} & \ldots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \ldots & a^{\alpha_{d-1}} \end{pmatrix}.$$

Obviously, the continuous shearlet transform is given by

$$L^2(\mathbb{R}^d) \ni f \longrightarrow \langle f, \psi_{M_{a,s},t}\rangle, \qquad (M_{a,s}, t) \in \mathbb{A}_d^S,$$

and the analyzing elements $\psi_{M_{a,s},t}$ are the affine functions defined by

$$\psi_{M_{a,s},t}(x) = |\det M_{a,s}|^{-\frac{1}{2}}\, \psi(M_{a,s}^{-1}(x-t)).$$

Remark 4.5 is the starting point for the 3-D setting.

4.2 3-D Continuous Shearlet Transform

The construction of shearlet systems in 3-D follows essentially the same ideas as the 2-D construction. Indeed, also in this case, it is convenient to use separate shearlet systems defined in different subregions of the frequency space.

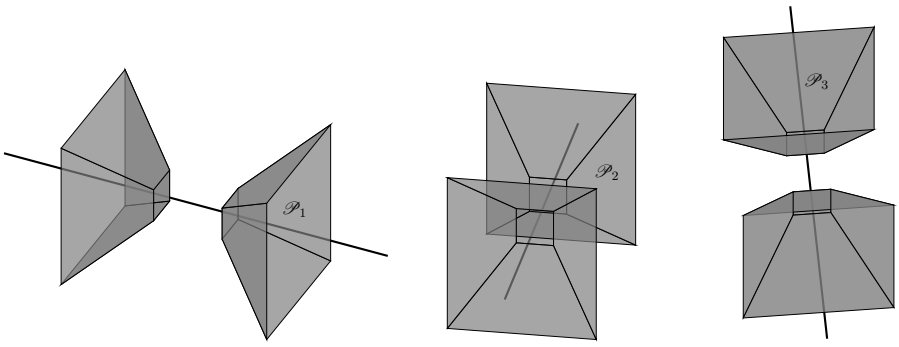


Figure 4.1: The partition of the frequency domain: sketch of the three couple of pyramids.

Recall that, in the definition of cone-adapted discrete 2-D shearlets, the 2-D frequency domain was partitioned into two pairs of high-frequency cones and one low-frequency rectangle. Analogously, we partition the 3-D frequency domain into the three pairs of pyramidal regions given by

$$\mathscr{P}_1 = \left\{(\xi_1,\xi_2,\xi_3) \in \mathbb{R}^3 : |\xi_1| \geq 1, \left|\frac{\xi_2}{\xi_1}\right| \leq 1 \text{ and } \left|\frac{\xi_3}{\xi_1}\right| \leq 1\right\},$$
$$\mathscr{P}_2 = \left\{(\xi_1,\xi_2,\xi_3) \in \mathbb{R}^3 : |\xi_2| \geq 1, \left|\frac{\xi_1}{\xi_2}\right| \leq 1 \text{ and } \left|\frac{\xi_3}{\xi_2}\right| \leq 1\right\},$$
$$\mathscr{P}_3 = \left\{(\xi_1,\xi_2,\xi_3) \in \mathbb{R}^3 : |\xi_3| \geq 1, \left|\frac{\xi_1}{\xi_3}\right| \leq 1 \text{ and } \left|\frac{\xi_2}{\xi_3}\right| > 1\right\},$$

and the centered cube

$$\mathscr{C} = \left\{(\xi_1,\xi_2,\xi_3) \in \mathbb{R}^3 : \|(\xi_1,\xi_2,\xi_3)\|_\infty < 1\right\}.$$

This partition is illustrated in Figures 4.1 and 4.2: in Figure 4.1 one can observe the three pairs of pyramids, while in Figure 4.2 the centered cube surrounded by the three pairs of pyramids $\mathscr{P}_1$, $\mathscr{P}_2$, and $\mathscr{P}_2$ is depicted.
Now, for $\xi = (\xi_1,\xi_2,\xi_3) \in \mathbb{R}^3$, $\xi_1 \neq 0$, let $\psi^{(d)}$, $d = 1,2,3$ be defined by

$$\hat{\psi}^{(1)}(\xi) = \hat{\psi}^{(1)}(\xi_1,\xi_2,\xi_3) = \hat{\psi}_1(\xi_1)\,\hat{\psi}_2\left(\frac{\xi_2}{\xi_1}\right)\,\hat{\psi}_2\left(\frac{\xi_3}{\xi_1}\right),$$
$$\hat{\psi}^{(2)}(\xi) = \hat{\psi}^{(2)}(\xi_1,\xi_2,\xi_3) = \hat{\psi}_1(\xi_2)\,\hat{\psi}_2\left(\frac{\xi_1}{\xi_2}\right)\,\hat{\psi}_2\left(\frac{\xi_3}{\xi_2}\right),$$
$$\hat{\psi}^{(3)}(\xi) = \hat{\psi}^{(3)}(\xi_1,\xi_2,\xi_3) = \hat{\psi}_1(\xi_3)\,\hat{\psi}_2\left(\frac{\xi_1}{\xi_3}\right)\,\hat{\psi}_2\left(\frac{\xi_2}{\xi_3}\right),$$

where ψ_1, ψ_2 satisfy the same assumptions as in the 2-D case. Note that $\psi^{(d)}$, $d = 1,2,3$, is nothing else but a special case of Example 4.4, *i.e.*, a 3-D generalization of the 2-D classical shearlet. We give the definition of *3-D pyramid-based continuous shearlet systems* in this special case.

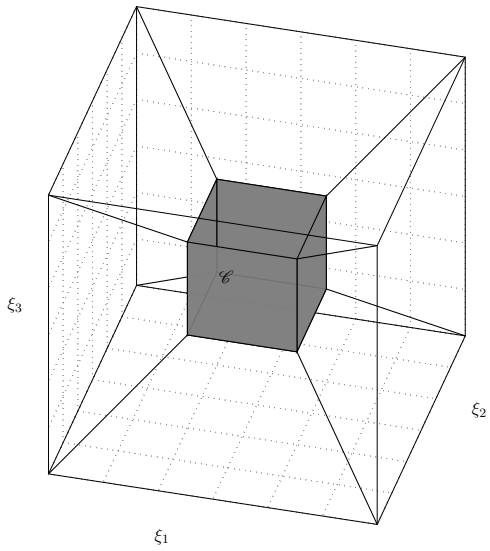


Figure 4.2: The partition of the frequency domain: the centered cube $\mathscr{C}$. See Figure 4.1 for a sketch of the pyramids.

Definition 4.6. For $d = 1, 2, 3$, let $\psi^{(d)} \in L^2(\mathbb{R}^3)$ and $\mathscr{P}_d$ be defined as above. The *3-D pyramid-based continuous shearlet systems* $\mathcal{SH}(\phi, \psi^{(1)}, \psi^{(2)}, \psi^{(3)})$ for $L^2(\mathscr{P}_d)^\vee$, generated by $\phi, \psi^{(1)}, \psi^{(2)}, \psi^{(3)} \in L^2(\mathbb{R}^3)$, is defined by

$$\mathcal{SH}(\phi, \psi^{(1)}, \psi^{(2)}, \psi^{(3)}) = \Phi(\phi) \cup \Psi^{(1)}(\psi^{(1)}) \cup \Psi^{(2)}(\psi^{(2)}) \cup \Psi^{(3)}(\psi^{(3)}),$$

where

$$\Phi(\phi) = \{\phi_t = \phi(\cdot - t) : t \in \mathbb{R}^3\},$$
$$\Psi^{(d)}(\psi^{(d)}) = \left\{\psi^{(d)}_{a,s_1,s_2,t} : 0 \le a \le \frac{1}{4}, -\frac{3}{2} \le s_1 \le \frac{3}{2}, -\frac{3}{2} \le s_2 \le \frac{3}{2}, t \in \mathbb{R}^3\right\},$$

with

$$\psi^{(d)}_{a,s_1,s_2,t}(x) = |\det M^{(d)}_{a,s_1,s_2}|^{-\frac{1}{2}}\, \psi^{(d)}((M^{(d)}_{a,s_1,s_2})^{-1}(x - t)),$$

and

$$M^{(1)}_{a,s_1,s_2} = \begin{pmatrix} a & a^{1/2}s_1 & a^{1/2}s_2 \\ 0 & a^{1/2} & 0 \\ 0 & 0 & a^{1/2} \end{pmatrix}, \qquad M^{(2)}_{a,s_1,s_2} = \begin{pmatrix} a^{1/2} & 0 & 0 \\ a^{1/2}s_1 & a & a^{1/2}s_2 \\ 0 & 0 & a^{1/2} \end{pmatrix},$$

$$M^{(3)}_{a,s_1,s_2} = \begin{pmatrix} a^{1/2} & 0 & 0 \\ 0 & a^{1/2} & 0 \\ a^{1/2}s_1 & a^{1/2}s_2 & a \end{pmatrix}.$$

Note that the elements of the shearlet systems $\psi^{(d)}_{a,s_1,s_2,t}$ are well-localized waveforms associated with various scales, controlled by a, various orientations, controlled by the two shear variables s_1, s_2 and various locations, controlled by t.
Similar to the 2-D case, in each pyramidal region the shearing variables are only allowed to vary over a compact set. This approach is essential to provide an almost uniform treatment of different directions in a sense of a good approximation to rotation.

Finally, for $f \in L^2(\mathbb{R}^3)$ we define the *3-D (fine-scale) pyramid-based continuous shearlet transform* $f \longrightarrow \mathscr{SH}_\psi f(a, s_1, s_2, t)$, for $a > 0$, s_1, $s_2 \in \mathbb{R}$, $t \in \mathbb{R}^3$ by

$$\mathscr{SH}_\psi f(a, s_1, s_2, t) = \begin{cases} \langle f, \psi^{(1)}_{a,s_1,s_2,t} \rangle & \text{if } |s_1|, |s_2| \leq 1 \\ \langle f, \psi^{(2)}_{a,\frac{1}{s_1},\frac{s_2}{s_1},t} \rangle & \text{if } |s_1| > 1, |s_2| \leq |s_1| \\ \langle f, \psi^{(3)}_{a,\frac{s_1}{s_2},\frac{1}{s_2},t} \rangle & \text{if } |s_2|, |s_2| > |s_1| \end{cases}$$

Depending on the values of the shearing variables, the 3-D continuous shearlet transform corresponds to one specific pyramid-based shearlet system.

4.3 3-D Discrete Shearlet Transform

The partition defined in the previous section enables restriction also in the discrete setting: the shear parameters is restricted to $[-\lceil 2^{j/2} \rceil, \lceil 2^{j/2} \rceil]$, as in the definition of cone-adapted discrete shearlet systems.
Pyramid-adapted discrete shearlets are scaled according to the paraboloidal scaling matrices $A^{(d)}_{2^j}$, $d = 1, 2, 3$ and $j \in \mathbb{Z}$ defined by

$$A^{(1)}_{2^j} = \begin{pmatrix} 2^j & 0 & 0 \\ 0 & 2^{j/2} & 0 \\ 0 & 0 & 2^{j/2} \end{pmatrix}, \quad A^{(2)}_{2^j} = \begin{pmatrix} 2^{j/2} & 0 & 0 \\ 0 & 2^j & 0 \\ 0 & 0 & 2^{j/2} \end{pmatrix}, \quad A^{(3)}_{2^j} = \begin{pmatrix} 2^{j/2} & 0 & 0 \\ 0 & 2^{j/2} & 0 \\ 0 & 0 & 2^j \end{pmatrix},$$

and directionality is encoded by the shear matrices $S^{(d)}_k$, $d = 1, 2, 3$ and $k = (k_1, k_2) \in \mathbb{Z}^2$, given by

$$S^{(1)}_k = \begin{pmatrix} 1 & k_1 & k_2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad S^{(2)}_k = \begin{pmatrix} 1 & 0 & 0 \\ k_1 & 1 & k_2 \\ 0 & 0 & 1 \end{pmatrix}, \quad S^{(3)}_k = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ k_1 & k_2 & 1 \end{pmatrix}$$

respectively. Clearly, these definitions are (discrete) special cases of the general setup discussed in the previous subsections. The translation lattices will be defined through the matrices $M^{(1)}_c = \mathrm{diag}(c_1, c_2, c_2)$, $M^{(2)}_c = \mathrm{diag}(c_2, c_1, c_2)$, and $M^{(3)}_c = \mathrm{diag}(c_2, c_2, c_1)$, where $c_1 > 0$ and $c_2 > 0$. In the following, we use the vector notation $|k| \leq K$ for $k = (k_1, k_2)$ and $K > 0$ to denote $|k_1| \leq K$ and $|k_2| \leq K$.
Now, we are ready to introduce 3-D shearlet systems for the discrete setting.

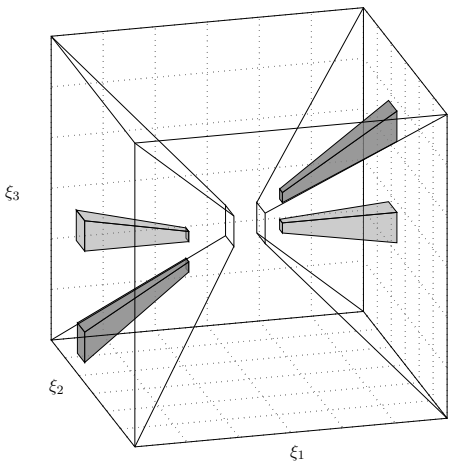


Figure 4.3: Support of two shearlet elements $\psi_{j,k,m}$ in the frequency domain. The two shearlet elements have the same scale parameter j, but different shearing parameters $k = (k_1, k_2)$.

Definition 4.7. For $c = (c_1, c_2) \in (\mathbb{R}^+)^2$, the *Pyramid-adapted Discrete Shearlet System* $\mathcal{SH}(\phi, \psi^{(1)}, \psi^{(2)}, \psi^{(3)}; c)$ generated by $\phi, \psi^{(1)}, \psi^{(2)}, \psi^{(3)} \in L^2(\mathbb{R}^3)$ is defined by

$$\mathcal{SH}(\phi, \psi^{(1)}, \psi^{(2)}, \psi^{(3)}; c) = \Phi(\phi; c_1) \cup \Psi^{(1)}(\psi^{(1)}; c) \cup \Psi^{(2)}(\psi^{(2)}; c) \cup \Psi^{(3)}(\psi^{(3)}; c),$$

where

$$\begin{aligned}\Phi(\phi; c_1) &= \{\phi_m = \phi(\cdot - m) : m \in c_1\mathbb{Z}^3\},\\ \Psi^{(d)}(\psi^{(d)}; c) &= \{\psi^{(d)}_{j,k,m} = 2^j\, \psi^{(d)}(S_k^{(d)} A_{2^j}^{(d)} \cdot - m) : j \geq 0, |k| \leq \lceil 2^{j/2} \rceil, m \in M_c^{(d)}\mathbb{Z}^3\},\end{aligned}$$

where $d = 1, 2, 3$, $j \in \mathbb{N}_0$ and $k \in \mathbb{Z}^2$.

Once again, our aim is to derive Parseval frames. For this purpose, let us consider the shearlet generator $\psi \in L^2(\mathbb{R}^3)$ defined by

$$\hat{\psi}(\xi) = \hat{\psi}_1(\xi_1)\, \hat{\psi}_2\left(\frac{\xi_2}{\xi_1}\right)\, \hat{\psi}_2\left(\frac{\xi_3}{\xi_1}\right),$$

We stress that, $\psi \in L^2(\mathbb{R}^3)$ as defined above is a canonical generalization of the classical band-limited 2-D shearlets. Recall that ψ_1 and ψ_2 satisfy the following assumptions:

a) $\hat{\psi}_1 \in C^\infty(\mathbb{R})$, $\mathrm{supp}(\hat{\psi}_1) \subseteq [4, -\frac{1}{2}] \cup [\frac{1}{2}, 4]$ and

$$\sum_{j \geq 0} |\hat{\psi}_1(2^{-j}\xi)|^2 = 1 \qquad \text{for } |\xi| \geq 1, \xi \in \mathbb{R}.$$

b) $\psi_2 \in C^\infty(\mathbb{R})$, $\mathrm{supp}(\hat{\psi}_1) \subset [-1, 1]$, and

$$\sum_{k=-1}^{1} |\hat{\psi}_2(\xi + k)|^2 = 1 \qquad \text{for } |\xi| \leq 1, \xi \in \mathbb{R}.$$

In frequency domain, the band-limited function $\psi \in L^2(\mathbb{R}^3)$ is almost a tensor product of one wavelet with two "bump" functions. This implies that the support in frequency domain has a needle-like shape with the wavelet acting in radial direction ensuring high directional selectivity, as can be observed in Figure 4.3. Indeed, the derivation from being a tensor product, *i.e.*, the substitution of ξ_2 and ξ_3 by the quotients ξ_2/ξ_1 and ξ_2/ξ_1, respectively, ensures a favorable behavior with respect to the shearing operator, and thus a tiling of frequency domain which leads to a tight frame for $L^2(\mathbb{R}^3)$.
A first step toward this result is the following observation.

Theorem 4.8. *Let ψ be a band-limited shearlet as defined in this subsection. Then the family of functions $P_{\mathscr{P}_1}\Psi(\psi)$ forms a tight frame for $L^2(\mathscr{P}_1)^\vee = \{f \in L^2(\mathbb{R}^3) : \mathrm{supp}\hat{f} \subset \mathscr{P}_1\}$, where $P_{\mathscr{P}}$ denotes the orthogonal projection onto $L^2(\mathscr{P}_1)^\vee$ and*

$$\Psi(\psi) = \{\psi_{j,k,m} : j \geq 0, |k| \leq \lceil 2^{j/2} \rceil, m \in \frac{1}{8}\mathbb{Z}^3\}.$$

By Theorem 4.8 and a change of variables, it is possible to construct shearlet tight frames for $L^2(\mathscr{P}_1)^\vee$, $L^2(\mathscr{P}_2)^\vee$ and $L^2(\mathscr{P}_3)^\vee$, respectively. Furthermore, wavelet theory provides us with many choices of $\phi \in L^2(\mathbb{R}^3)$ such that $\Phi(\phi; \frac{1}{8})$ forms a tight frame for $L^2(\mathbb{R})$. Since $\mathbb{R}^3 = \mathscr{C} \cup \mathscr{P}_1 \cup \mathscr{P}_2 \cup \mathscr{P}_3$ as a disjoint union, we can express any function $f \in L^2(\mathbb{R}^3)$ as $f = P_{\mathscr{C}} f + P_{\mathscr{P}_1} f + P_{\mathscr{P}_2} f + P_{\mathscr{P}_3} f$, where P_C denotes the orthogonal projection onto the closed subspace $L^2(C)^\vee$ for some measurable set $C \subset \mathbb{R}^3$.
Then, we can expand the projection $P_{\mathscr{P}_1} f$ in terms of the corresponding tight frame $P_{\mathscr{P}_1}\Psi(\psi)$

and similarly for the other three projections. Hence, the representation of f is the sum of these four expansions.
Finally, we remark that the projection of f and the shearlet frame elements onto the four subspaces can lead to artificially slow decaying shearlet coefficients; this is the case, *e.g.*, if f is in the Schwartz class. This problem does not occur in the construction of compactly supported shearlets.

REMARK 4.9. As already stressed, the situation for 3-D data is quite different from the situation for 2-D since anisotropic features of different dimensions are involved, namely, singularities on 1-D and 2-D manifolds. For this reason, beside the parabolic scaling matrix previously considered there is not another possibility:

$$\begin{pmatrix} 2^j & 0 & 0 \\ 0 & 2^{j/2} & 0 \\ 0 & 0 & 2^j \end{pmatrix} \qquad \text{instead of} \qquad \begin{pmatrix} 2^j & 0 & 0 \\ 0 & 2^{j/2} & 0 \\ 0 & 0 & 2^{j/2} \end{pmatrix}.$$

The first choice leads to needle-like shearlets, which are intuitively better suited to capture 1-D singularities. The second choice leads to plate-like shearlets, which are more suited to 2-D singularities. Both systems are needed if the goal is to distinguish these two types of singularities. However, for the construction of (nearly) optimally sparse approximations, it can be shown that the plate-like shearlets are the right approach ([25]).

4.4 Optimally Sparse Approximations

We already stressed that capturing anisotropic phenomenon in 3-D is quite different from capturing anisotropic features in 2-D. Indeed, in 2-D we have "only" curves to handle with, while in the 3-D setting can occur two different geometric structures: curves (1-D anisotropic features) and surfaces (2-D features).
In Remark 4.9 we observed that intuitively both needle-like and plate-like shearlets are needed to distinguish between these two types of singularities. However, in [27] it is proved that 3-D shearlet elements plate-like in spatial domain are able to perform optimally when representing and analyzing 3-D data containing both curve and surface singularities.
Before giving the general result, we need two more definition.

Definition 4.10. For a fixed $\mu > 0$, the class $\mathscr{E}^2(\mathbb{R}^d)$ of *cartoon-like images* is the set of functions $f : \mathbb{R}^d \to \mathbb{C}$ of the form

$$f = f_0 + f_1\chi_B,$$

where $B \subset [0,1]^d$ and $f_i \in \mathscr{C}^2(\mathbb{R}^d)$ are functions with $\operatorname{supp}(f_i) \subset [0,1]^d$ and $\|f_i\|_{\mathscr{C}^2} \leq \mu$ for each $i = 0, 1$.

This class was introduced to provide a simplified model of natural images, which emphasizes anisotropic features, in particular edges, whence the name. This is the standard model class used to derive results in approximation theory.

Definition 4.11. Let $\Phi = (\phi_i)_{i\in\mathcal{I}}$ be a frame for $L^2(\mathbb{R}^d)$ with $d = 2$ or $d = 3$. We say that Φ *provides optimally sparse approximations* of cartoon-like images if, for each $f \in \mathscr{E}^2(\mathbb{R}^d)$, the associated N-term approximation f_N obtained by keeping the N largest coefficients of $c = c(f) = (\langle f, \phi_i\rangle)_{i\in\mathcal{I}}$ satisfies

$$\|f - f_N\|_{L^2}^2 < N^{-\frac{2}{d-1}} \qquad \text{as } N \to \infty,$$

and

$$|c_n^*| < n^{-\frac{d+1}{2(d-1)}} \qquad \text{as } n \to \infty,$$

up to a log-factor.

Now, we are ready to state the general result in the case of band-limited generators. We observe that a similar result can be proved for compactly supported generators (see [27]).

Theorem 4.12. *Assume that $\phi, \psi^{(1)}, \psi^{(2)}, \psi^{(3)} \in L^2(\mathbb{R}^3)$ are band-limited and $\mathscr{C}^\infty$ in the frequency domain and that the shearlet system $\mathcal{SH}(\phi, \psi^{(1)}, \psi^{(2)}, \psi^{(3)}; c)$ forms a frame for $L^2(\mathbb{R}^3)$. For any $\mu > 0$, the shearlet frame $\mathcal{SH}(\phi, \psi^{(1)}, \psi^{(2)}, \psi^{(3)}; c)$ provides optimally sparse approximations of functions $f \in \mathscr{E}^2(\mathbb{R}^3)$ in the sense of previous definition, i.e.,*

$$\|f - f_N\|_{L^2}^2 < N^{-1}(\log N)^2 \qquad \text{as } N \to \infty,$$

and

$$|c_n^*| < n^{-1} \log n \qquad \text{as } n \to \infty.$$

Proof. See [27]. □

The previous result can extended to the so-called *extended class of cartoon-like image* $\mathscr{E}_L^2(\mathbb{R}^3)$, where $L \in \mathbb{N}$ denotes the number of $\mathscr{C}^2$-smooth pieces in which the discontinuity surface ∂B is divided.

In conclusion, we observed that, using cartoon-like images as model class, we can measure the approximation properties by considering the decay rate of the L^2 error of the best N-term approximation. Up to date, shearlet systems are the only representation system able to provide optimally sparse approximations of this model class in 2-D as well as 3-D.

Conclusions

This overview has the purpose to give a brief introduction on shearlets. Although there are no new results, the contribution lies in having collected, organized, explained and tied up a lot of material that is currently scattered in many different books and articles, even in different areas of mathematics, since this subject is pretty recent. The starting point for this work was the nice book on shearlets published in 2012, edited by Labate and Kutyniok [26], that had the merit of collecting a number of relevant results on shearlets. However, this book is not aimed at providing a whole reorganization of the material conceived until then, in particular for what concerns multivariate and 3-D shearlet systems and transform, and the related theory from continuum to discrete setting.
The overview presented here is intended as a theoretical guide to address numerical issues related to the practical application of sherlets in various contexts. The work in progress involve applications in the numerical solution of inverse problems, starting from some important results in image application and data separation recently published.

Bibliography

[1] Albert Boggess and Francis J. Narcowich. *A first course in wavelets with Fourier analysis.* John Wiley & Sons Inc., Hoboken, NJ, second edition, 2009.

[2] Emmanuel J. Candès and David L. Donoho. Ridgelets: a key to higher-dimensional intermittency? *R. Soc. Lond. Philos. Trans. Ser. A Math. Phys. Eng. Sci.*, 357(1760):2495–2509, 1999.

[3] Emmanuel J. Candès and David L. Donoho. New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities. *Comm. Pure Appl. Math.*, 57(2):219–266, 2004.

[4] Ole Christensen. *An introduction to frames and Riesz bases.* Boston, MA: Birkhäuser, 2003.

[5] Ole Christensen. B-spline generated frames. In *Four short courses on harmonic analysis. Wavelets, frames, time-frequency methods, and applications to signal and image analysis*, pages 51–86. Basel: Birkhäuser, 2010.

[6] Charles K. Chui. *An introduction to wavelets.* Boston, MA: Academic Press, 1992.

[7] Flavia Colonna, Glenn Easley, Kanghui Guo, and Demetrio Labate. Radon transform inversion using the shearlet representation. *Appl. Comput. Harmon. Anal.*, 29(2):232–250, 2010.

[8] Stephan Dahlke, Gitta Kutyniok, Peter Maass, Chen Sagiv, Hans-Georg Stark, and Gerd Teschke. The uncertainty principle associated with the continuous shearlet transform. *Int. J. Wavelets Multiresolut. Inf. Process.*, 6(2):157–181, 2008.

[9] Stephan Dahlke, Gitta Kutyniok, Gabriele Steidl, and Gerd Teschke. Shearlet coorbit spaces and associated banach frames. *Appl. Comput. Harmon. Anal.*, 27(2):195–214, 2009.

[10] Stephan Dahlke, Gabriele Steidl, and Gerd Teschke. The continuous shearlet transform in arbitrary space dimensions. *J. Fourier Anal. Appl.*, 16(3):340–364, 2010.

[11] Stephan Dahlke, Gabriele Steidl, and Gerd Teschke. Multivariate shearlet transform, shearlet coorbit spaces and their structural properties. In *Shearlets. Multiscale analysis for multivariate data.*, pages 105–144. Boston, MA: Birkhäuser, 2012.

[12] Minh N. Do and Martin Vetterli. The contourlet transform: an efficient directional multiresolution image representation. *IEEE Trans. Image Process.*, 14(4):2091–2106, 2005.

[13] Gerald B. Folland. *Harmonic analysis in phase space.* Princeton, NJ: Princeton University Press, 1989.

[14] Gerald B. Folland. *A course in abstract harmonic analysis.* Boca Raton, FL: CRC Press, 1995.

[15] Gerald B. Folland. *Real analysis. Modern techniques and their applications. 2nd ed.* New York, NY: Wiley, 2nd ed. edition, 1999.

[16] Philipp Grohs. Continuous shearlet tight frames. *J. Fourier Anal. Appl.*, 17(3):506–518, 2011.

[17] Philipp Grohs. Shearlets and microlocal analysis. In *Shearlets. Multiscale analysis for multivariate data.*, pages 39–67. Boston, MA: Birkhäuser, 2012.

[18] Kanghui Guo, Gitta Kutyniok, and Demetrio Labate. Sparse multidimensional representations using anisotropic dilation and shear operators. In *Wavelets and splines: Athens 2005. Papers based on lectures presented at the international conference on interactions between wavelets and splines, Athens, GA, USA, May 16–19, 2005. In honor of Professor Charles K. Chui on the occasion of his 65th birthday*, pages 189–201. Brentwood: Nashboro Press, 2006.

[19] Kanghui Guo, Demetrio Labate, Wang-Q. Lim, Guido Weiss, and Edward Wilson. Wavelets with composite dilations and their MRA properties. *Appl. Comput. Harmon. Anal.*, 20(2):202–236, 2006.

[20] Eugenio Hernández and Guido Weiss. *A first course on wavelets.* Boca Raton, FL: CRC Press, 1996.

[21] Antoine Jean-Pierre, Romain Murenzi, and Pierre Vandergheynst. Directional wavelets revisited: Cauchy wavelets and symmetry detection in patterns. *Appl. Comput. Harmon. Anal.*, 6(3):314–345, 1999.

[22] Nick Kingsbury. Complex wavelets for shift invariant analysis and filtering of signals. *Appl. Comput. Harmon. Anal.*, 10(3):234–253, 2001.

[23] Gitta Kutyniok and Demetrio Labate. Construction of regular and irregular shearlet frames. *J. Wavelet Theory and Appl.*, 1:1–10, 2007.

[24] Gitta Kutyniok and Demetrio Labate. Resolution of the wavefront set using continuous shearlets. *Trans. Am. Math. Soc.*, 361(5):2719–2754, 2009.

[25] Gitta Kutyniok and Demetrio Labate. Introduction to shearlets. In *Shearlets. Multiscale analysis for multivariate data.*, pages 1–38. Boston, MA: Birkhäuser, 2012.

[26] Gitta Kutyniok and Demetrio Labate, editors. *Shearlets.* Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, New York, 2012. Multiscale analysis for multivariate data.

[27] Gitta Kutyniok, Jakob Lemvig, and Wang-Q Lim. Shearlets and optimally sparse approximations. In *Shearlets. Multiscale analysis for multivariate data.*, pages 145–197. Boston, MA: Birkhäuser, 2012.

[28] Gitta Kutyniok and Tomas Sauer. From wavelets to shearlets and back again. In *Approximation theory XII. Proceedings of the 12th international conference, San Antonio, TX, USA, March 4–8, 2007*, pages 201–209. Brentwood, TN: Nashboro Press, 2008.

[29] Demetrio Labate, Wang-Q. Lim, Gitta Kutyniok, and Guido Weiss. Sparse multidimensional representation using shearlets. In Ca San Diego, editor, *SPIE Proc.*, volume 5914 of *SPIE*, pages 254–262, Bellingham, 2005.

[30] Erwan Le Pennec and Stéphane Mallat. Sparse geometric image representations with bandelets. *IEEE Trans. Image Process.*, 14(4):423–438, 2005.

[31] Stéphane Mallat. *A wavelet tour of signal processing. The sparse way. 3rd ed.* Amsterdam: Elsevier/Academic Press, 3rd ed. edition, 2009.

[32] Walter Rudin. *Real and complex analysis. 3rd ed.* New York, NY: McGraw-Hill, 3rd ed. edition, 1987.

[33] Guido Weiss and Edward N. Wilson. The mathematical theory of wavelets. In *Twentieth century harmonic analysis—a celebration. Proceedings of the NATO Advanced Study Institute, Il Ciocco, Italy, July 2–15, 2000*, pages 329–366. Dordrecht: Kluwer Academic Publishers, 2001.