

# Introduction

Gitta Kutyniok and Demetrio Labate

Scientists frequently refer to the 21st century as the era of data. In fact, since technological advances make data acquisition easier and less expensive, we are facing today a deluge of data, including, for instance, astronomical, medical, seismic, meteorological and surveillance data, which require efficient analysis and processing. The enormity of the challenge this poses is evidenced not only by the sheer amount of data, but also by the diversity of data types and the variety of processing tasks which are required. To efficiently handle tasks ranging from feature analysis over classification to compression, highly sophisticated mathematical and computational methodologies are needed. From a mathematical standpoint, data can be modeled, for example, as functions, distributions, point clouds, or graphs. Moreover, data can be classified by membership in one of the two categories: Explicitly given data, such as imaging or measurement data, and implicitly given data, such as solutions of differential or integral equations.

A fundamental property of virtually all data found in practical applications is that the relevant information which needs to be extracted or identified is sparse, i.e., data are typically highly correlated and the essential information lies on low dimensional manifolds. This information can thus be captured, in principle, using just few terms in an appropriate dictionary. This observation is crucial not only for tasks such as data storage and transmission, but also for feature extraction, classification, and other high-level tasks. Indeed, finding a dictionary which sparsely represents a certain data class entails the intimate understanding of its dominant features, which are typically associated with singularities and other geometric properties. Related to this is the observation that virtually all multivariate data are typically dominated by anisotropic features such as singularities on lower dimensional embedded man-

---

Gitta Kutyniok

University of Osnabrück, Institute of Mathematics, 49069 Osnabrück, Germany, e-mail: kutyniok@math.uos.de

Demetrio Labate

University of Houston, Department of Mathematics, Houston, TX 77204, USA, e-mail: dlabate@math.uh.edu

ifolds. This is exemplified, for instance, by edges in natural images or shock fronts in the solutions of transport equations. Hence, to efficiently analyze and process these data, it is of fundamental importance to discover and truly understand their *geometric structures*.

The subject of this volume is a recently introduced multiscale framework, called *shearlets*, which allows optimal encoding of several classes of multivariate data through its ability to sparsely represent anisotropic features. As will be illustrated in the following, shearlets emerged as part of an extensive research activity developed during the last 10 years to create a new generation of analysis and processing tools for massive and higher dimensional data, which could go beyond the limitations of traditional Fourier and wavelet systems. One of the forerunners of this area of research is David L. Donoho, who observed that, in higher dimensions, traditional multiscale systems and wavelets ought to be replaced by a *Geometric Multiscale Analysis* where multiscale analysis is adapted to intermediate-dimensional singularities. It is important to remark that many of the ideas which are at the core of this approach can be traced back to key results in harmonic analysis from the 1990's, such as Hart Smith's Hardy space for Fourier Integral Operators and Peter Jones' Analyst's Traveling Salesman theorem. Both results concern the higher-dimensional setting, where geometric ideas are brought into play to discover 'new architectures for decomposition, rearrangement and reconstruction of operators and functions' [15].

This broader area of research is currently at the crossroad of applied mathematics, electrical engineering, and computer science, and has seen spectacular advances in recent years, resulting in highly sophisticated and efficient algorithms for image analysis and new paradigms for data compression and approximation. By presenting the theory and applications of shearlets obtained during last five years, this book is also a journey into one of the most active and exciting areas of research in applied mathematics.

## 1 The Rise of Shearlets

### 1.1 The Role of Applied Harmonic Analysis

Applied harmonic analysis has established itself as the main area in applied mathematics focused on the efficient representation, analysis, and encoding of data. The primary object of this discipline is the process of 'breaking into pieces' – from the Greek word *analysis* –, to gain insight into an object. Given a class of data  $\mathcal{C}$  in  $L^2(\mathbb{R}^d)$ , say, a collection of *analyzing* functions  $(\varphi_i)_{i \in I} \subseteq L^2(\mathbb{R}^d)$  with  $I$  being a countable indexing set is sought such that, for all  $f \in \mathcal{C}$ , we have the expansion

$$f = \sum_{i \in I} c_i(f) \varphi_i. \quad (1)$$

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

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 data class. Let us illustrate the two most successful types of analyzing systems in the one-dimensional setting. The *Gabor systems* are designed to best represent the joint time-frequency content of data. In this case, the analyzing elements  $(\varphi_i)_{i \in I}$  are obtained as translations and frequency shifts of a generating function  $\varphi \in L^2(\mathbb{R})$  as follows:

$$\{\varphi_{p,q} = \varphi(\cdot - p) e^{2\pi i q \cdot} : p, q \in \mathbb{Z}\}.$$

In contrast to this approach, *wavelet systems* represent the data as associated with different location and resolution levels. In this case, the analyzing elements  $(\varphi_i)_{i \in I}$  are obtained through the action of dilation and translation operators on a generating function  $\psi \in L^2(\mathbb{R})$ , called a *wavelet*, as:

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

Given a prescribed class of data  $\mathcal{C}$ , one major objective is to design an analyzing system  $(\varphi_i)_{i \in I}$  in such a way that, for each function  $f \in \mathcal{C}$ , the coefficient sequence  $(c_i(f))_{i \in I}$  in (1) can be chosen to be *sparse*. In the situation of an infinite-dimensional Hilbert space – which is our focus here – the degree of sparsity is customarily measured as the decay rate of the error of best  $n$ -term approximation. Loosely speaking, this means that we can approximate any  $f \in \mathcal{C}$  with high accuracy by using a coefficient sequence  $(\tilde{c}_i(f))_{i \in I}$  containing very few non-zero entries. In the finite-dimensional setting, such a sequence is called *sparse*, and this explains the use of the term *sparse approximations*. Intuitively, if a function can be sparsely approximated, it is conceivable that ‘important’ features can be detected by thresholding, i.e., by selecting the indices associated with the ‘largest’ coefficients in absolute values, or that high compression rates can be achieved by storing only few ‘large’ coefficients  $c_i(f)$ , see [18].

There is another fundamental phenomenon to observe here. If  $(\varphi_i)_{i \in I}$  is an orthonormal basis, the coefficient sequence  $(c_i(f))_{i \in I}$  in (1) is certainly uniquely determined. However, if we allow more freedom in the sense of choosing  $(\varphi_i)_{i \in I}$  to form a frame – a redundant, yet stable system (see Subsection 2.3) – the sequences  $(c_i(f))_{i \in I}$  might be chosen significantly sparser for each  $f \in \mathcal{C}$ . Thus, methodologies from *frame theory* will come into play, see Subsection 2.3 and [7, 5].

We can observe a close connection to yet another highly topical area. During the last four years, sparse recovery methodologies such as, in particular, Compressed Sensing have revolutionized the areas of applied mathematics, computer science, and electrical engineering by beating the traditional sampling theory limits, see [3, 22]. They exploit the fact that many types of signals can be represented using only a few non-vanishing coefficients when choosing a suitable basis or, more generally, a frame. Nonlinear optimization methods, such as  $\ell_1$  minimization, can then

be employed to recover such signals from ‘very few’ measurements under appropriate assumptions on the signal and on the basis or frame. These results can often be generalized to data which are merely sparsely approximated by a frame, thereby enabling Compressed Sensing methodologies for the situation we discussed above.

## 1.2 Wavelets and Beyond

Historically, the introduction of *wavelets* about 20 years ago represents a milestone in the development of efficient encoding of piecewise regular signals. The major 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 digital setting is a *Multiresolution Analysis*, which allows a direct transition between the realms of real variable functions and digital signals. This framework also combines very naturally with the theory of filter banks developed in the digital signal processing community. 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, including the algorithm of JPEG2000, the current standard for image compression. We refer the interested reader to [60] for more details about wavelets and their applications.

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 distributed singularities such as singularities along curves. The intuitive reason for this is that wavelets are *isotropic* objects, being generated by isotropically dilating a single or finite set of generators. However, in dimensions two and higher, 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.

The limitations of wavelets and traditional multiscale systems have stimulated a flurry of activity involving mathematicians, engineers, and applied scientists. Indeed, the need to increase the directional sensitivity of wavelets was already recognized in the early filter bank literature, and several versions of ‘directional’ wavelets were introduced, including the *steerable pyramid* by Simoncelli et al. [62], the *directional filter banks* by Bamberger and Smith [2], and the *2D directional wavelets* by Antoine et al. [1]. A more sophisticated approach was proposed more recently with the introduction of *complex wavelets* [40, 41]. However, even though they frequently outperform standard wavelets in applications, these methods do not provide

optimally sparse approximations of multivariate data governed by anisotropic features. The fundamental reason for this failure is that these approaches are not truly multidimensional extensions of the wavelet approach.

The real breakthrough occurred with the introduction of *curvelets* by Candès and Donoho [4] in 2004, which was the first system providing optimally sparse approximations for a class of bivariate functions exhibiting anisotropic features. Curvelets form a pyramid of analyzing functions defined not only at various scales and locations, as wavelets do, but also at various orientations, with the number of orientations increasing at finer scales. Another fundamental property is that their supports are highly anisotropic and become increasingly elongated at finer scales. Due to this anisotropy, curvelets are essentially as good as an adaptive representation system from the point of view of the ability to sparsely approximate images with edges. The two main drawbacks of the curvelet approach are that, firstly, this system is not singly generated, i.e., it is not derived from the action of countably many operators applied to a single (or finite set) of generating functions; secondly, since its construction involves rotations and this operator does not preserve the digital lattice, which prevents a direct transition from the continuum to the digital setting.

*Contourlets* were introduced in 2005 by Do and Vetterli [13] as a purely discrete filter-bank version of the curvelet framework. This approach offers the advantage of allowing a tree-structured filter bank implementation similar to the wavelet implementation, which was exploited to obtain remarkable numerical results. However, the directional selectivity is artificially imposed by the special sampling rule of a filter bank and a proper continuum theory is missing.

In the same year, *shearlets* were introduced by Guo, Kutyniok, Labate, Lim, and Weiss in [57, 28]. This approach was derived within a larger class of affine-like systems – the so-called *composite wavelets* [35, 37, 36] – as a truly multivariate extension of the wavelet framework. One of the distinctive features of shearlets is the use of shearing to control directional selectivity, in contrast to rotation used by curvelets. This is a fundamentally different concept, since it allows to derive shearlet systems as singly generated systems, and it also ensures a unified treatment of the continuum and digital world due to the fact that the shear matrix preserves the integer lattice. Indeed, as will be extensively discussed in the remainder of this volume, the shearlet representation offers a unique combination of the following list of desiderata:

- 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 will present a self-contained overview of the key results from the theory and applications of shearlets, focused primarily on the 2D setting. These results will be elaborated in much more detail in the various chapters of this volume, which will discuss both the continuum and digital aspects of shearlets. Before

starting our overview, it will be useful to establish the notation adopted throughout this volume and to present some background material from harmonic analysis and wavelet theory.

## 2 Notation and Background Material

### 2.1 Fourier Analysis

The Fourier transform is the most fundamental tool in harmonic analysis. Before stating the definition, we notice that, in the following, vectors in  $\mathbb{R}^d$  or  $\mathbb{C}^d$  will always be understood as column vectors, and their inner product – as also the inner product in  $L^2(\mathbb{R}^d)$  – shall be denoted by  $\langle \cdot, \cdot \rangle$ . For a function  $f \in L^1(\mathbb{R}^d)$ , the *Fourier transform* of  $f$  is defined by

$$\hat{f}(\xi) = \int f(x) e^{-2\pi i \langle x, \xi \rangle} dx,$$

and  $f$  is called a *band-limited* function if its Fourier transform is compactly supported. The *inverse Fourier transform* of a function  $g \in L^1(\mathbb{R}^d)$  is given as

$$\check{g}(x) = \int g(\xi) e^{2\pi i \langle x, \xi \rangle} d\xi.$$

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 ‘true’ inverse. 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  $f, g \in L^2(\mathbb{R}^n)$  reads

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

and, in particular,

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

We refer to [24] for additional background information on Fourier analysis.

### 2.2 Modeling of Signal Classes

In the continuum setting, the standard model of  $d$ -dimensional signals is the space of *square-integrable functions* on  $\mathbb{R}^d$ , denoted by  $L^2(\mathbb{R}^d)$ . However, this space also contains objects which are very far from ‘natural’ images and data. Hence, it is convenient to introduce subclasses and subspaces which can better model the types of data encountered in applications. One approach for doing this consists in im-

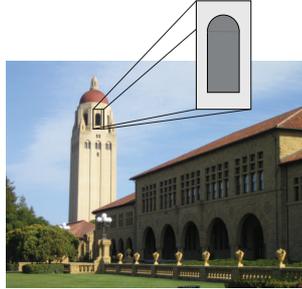
posing some degree of regularity. Therefore, we consider the *continuous functions*  $C(\mathbb{R}^d)$ , the  $k$ -times continuously differentiable functions  $C^k(\mathbb{R}^d)$ , and the infinitely many times continuously differentiable functions  $C^\infty(\mathbb{R}^d)$ , which are also referred to as *smooth functions*. Since images are compactly supported in nature, a notion for *compactly supported functions* is also required which will be indicated with the subscript 0, e.g.,  $C_0^\infty(\mathbb{R}^d)$ .

Sometimes it is useful to consider curvilinear singularities such as edges in images as singularities of distributions, which requires the *space of distributions*  $\mathcal{D}'(\mathbb{R}^d)$  as a model. For a distribution  $u$ , we say that  $x \in \mathbb{R}^d$  is a *regular point* of  $u$ , if there exists a function  $\phi \in C_0^\infty(U_x)$  with  $\phi(x) \neq 0$  and  $U_x$  being a neighborhood of  $x$ . This implies  $\phi u \in C_0^\infty(\mathbb{R}^d)$ , which is equivalent to  $(\phi u)^\wedge$  being rapidly decreasing. The complement of the set of regular points of  $u$  is called the *singular support* of  $u$  and is denoted by  $\text{sing supp}(u)$ . Notice that the singular support of  $u$  is a closed subset of  $\text{supp}(u)$ .

The anisotropic nature of singularities on one- or multi-dimensional embedded manifolds becomes apparent through the notion of a wavefront set. For simplicity, we illustrate the 2-dimensional case only. For a distribution  $u$ , a point  $(x, s) \in \mathbb{R}^2 \times \mathbb{R}$  is a *regular directed point*, if there exist neighborhoods  $U_x$  of  $x$  and  $V_s$  of  $s$  as well as a function  $\phi \in C_0^\infty(\mathbb{R}^2)$  satisfying  $\phi|_{U_x} \equiv 1$  such that, for each  $N > 0$ , there exists a constant  $C_N$  with

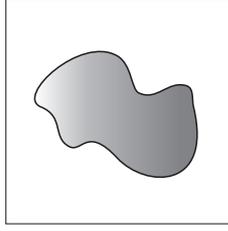
$$|(u\phi)^\wedge(\eta)| \leq C_N (1 + |\eta|)^{-N} \quad \text{for all } \eta \in \mathbb{R}^2 \text{ with } \frac{\eta_2}{\eta_1} \in V_s.$$

The complement in  $\mathbb{R}^2 \times \mathbb{R}$  of the regular directed points of  $u$  is called the *wavefront set* of  $u$  and is denoted by  $WF(u)$ . Thus, the singular support describes the location of the set of singularities of  $u$ , and the wavefront set describes both the location and local perpendicular orientation of the singularity set.



**Fig. 1** Natural images are governed by anisotropic structures.

A class of functions, which is of particular interest in imaging sciences, is the class of so-called *cartoon-like images*. This class was introduced in [14] to provide a simplified model of natural images, which emphasizes anisotropic features, most notably edges, and is consistent with many models of the human visual system. Consider, for example, the photo displayed in Figure 1. Since the image basically



**Fig. 2** Example of a cartoon-like image.

consists of smooth regions separated by edges, it is suggestive to use a model consisting of piecewise regular functions, such as the one illustrated in Figure 2. For simplicity, the domain is set to be  $[0, 1]^2$  and the regularity can be chosen to be  $C^2$ , leading to the following definition.

**Definition 1.** The class  $\mathcal{E}^2(\mathbb{R}^2)$  of cartoon-like image is the set of functions  $f: \mathbb{R}^2 \rightarrow \mathbb{C}$  of the form

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

where  $B \subset [0, 1]^2$  is a set with  $\partial B$  being a closed  $C^2$ -curve with bounded curvature and  $f_i \in C^2(\mathbb{R}^2)$  are functions with  $\text{supp } f_0 \subset [0, 1]^2$  and  $\|f_i\|_{C^2} \leq 1$  for each  $i = 0, 1$ .

Let us finally mention that, in the digital setting, the usual models for  $d$ -dimensional signals are either functions on  $\mathbb{Z}^d$  such as  $\ell^2(\mathbb{Z}^d)$  or functions on  $\{0, \dots, N-1\}^d$ , sometimes denoted by  $\mathbb{Z}_N^d$ .

### 2.3 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*, originally introduced by Duffin and Schaeffer in [19] and later revived by Daubechies in [12], guarantees stability while allowing non-unique decompositions. Let us recall the basic definitions from frame theory in the setting of a general (real or complex) Hilbert space  $\mathcal{H}$ .

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

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

The frame constants  $A$  and  $B$  are called *lower* and *upper frame bound*, respectively. The supremum over all  $A$  and the infimum over all  $B$  such that the frame inequalities hold are the *optimal frame bounds*. If  $A$  and  $B$  can be chosen as  $A = B$ , then the frame is called *A-tight*, and if  $A = B = 1$  is possible, then  $(\varphi_i)_{i \in I}$  is a *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  $c = 1$ .

Apart from providing redundant expansions, frames serve as an analysis tool. In fact, they allow the analysis of data by studying the associated *frame coefficients*  $(\langle x, \varphi_i \rangle)_{i \in I}$ , where the operator  $T$  defined by

$$T : \mathcal{H} \rightarrow \ell^2(I), \quad x \mapsto (\langle x, \varphi_i \rangle)_{i \in I}$$

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

$$T^* : \ell^2(I) \rightarrow \mathcal{H}, \quad ((c_i)_{i \in I}) \mapsto \sum_{i \in I} c_i \varphi_i.$$

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

$$S = T^*T : \mathcal{H} \rightarrow \mathcal{H}, \quad x \mapsto \sum_{i \in I} \langle x, \varphi_i \rangle \varphi_i.$$

The operator  $S$  is a positive, self-adjoint invertible operator on  $\mathcal{H}$  with  $A \cdot \text{Id}_{\mathcal{H}} \leq S \leq B \cdot \text{Id}_{\mathcal{H}}$ , where  $\text{Id}_{\mathcal{H}}$  denotes the identity operator on  $\mathcal{H}$ . In the case of a Parseval frame, this reduces to  $S = \text{Id}_{\mathcal{H}}$ .

In general, a signal  $x \in \mathcal{H}$  can be recovered from its frame coefficients through the reconstruction formula

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

The sequence  $(S^{-1} \varphi_i)_{i \in I}$ , 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  $(\varphi_i)_{i \in I}$ , we observe that, for each vector  $x \in \mathcal{H}$ ,

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

If the frame does not constitute a basis, i.e., it is redundant, the coefficient sequence  $(\langle x, S^{-1} \varphi_i \rangle)_{i \in I}$  of this expansion is certainly not unique. It is this property which then enables to derive much sparser expansions. It should also be noted that the sequence  $(\langle x, S^{-1} \varphi_i \rangle)_{i \in I}$  has the distinct property of being the smallest in  $\ell^2$  norm of all expansion coefficient sequences.

For more details on frame theory, we refer the interested reader to [5, 7].

## 2.4 Wavelets

Wavelet analysis plays a central role in this volume since, as will be made more precise in the following, shearlets arise naturally from this general framework. Hence

a full understanding of shearlets can only be derived through a sound understanding of wavelet theory.

We start by rewriting the definition of a *discrete wavelet system* in  $L^2(\mathbb{R})$ , stated at the beginning of the introduction in (2), as

$$\{\psi_{j,m} = D_2^j T_m \psi = 2^{j/2} \psi(2^j \cdot -m) : j, m \in \mathbb{Z}\}, \quad (3)$$

where  $\psi \in L^2(\mathbb{R})$ ,  $D_2$  is the *dyadic dilation operator* on  $L^2(\mathbb{R})$  defined by

$$D_2 \psi(x) = 2^{-1/2} \psi(2^{-1}x), \quad (4)$$

and  $T_t$  is the *translation operator* on  $L^2(\mathbb{R})$ , defined by

$$T_t \psi(x) = \psi(x-t), \quad \text{for } t \in \mathbb{R}. \quad (5)$$

The associated *Discrete Wavelet Transform* is then defined to be the mapping

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

If the system (3) is an orthonormal basis of  $L^2(\mathbb{R})$ , it is called an *orthonormal wavelet system*, and  $\psi$  is called a *wavelet*. Being a wavelet is by no means very restrictive and plenty of choices exist. In fact, it is possible to construct wavelets  $\psi$ , 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: the *Daubechies wavelets*, which have compact support and can be chosen to have high regularity, leading to good decay in the frequency domain; and the *Lemarié-Meyer wavelets*, which are band-limited and  $C^\infty$  in the frequency domain, forcing rapid decay in the spatial domain. 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 be shown below.

In fact, there is a general machinery to construct discrete wavelet systems, 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)  $\{0\} \subset \dots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \dots \subset L^2(\mathbb{R})$ .
- (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  $D_2 f \in V_{j+1}$ .
- (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<sup>1</sup> for  $V_0$ . is then

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

<sup>1</sup> This assumption can be replaced by the weaker assumption that  $\{T_m \phi : m \in \mathbb{Z}\}$  is Riesz basis for  $V_0$

by considering the orthogonal complements

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

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. Thus, 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 involving both the wavelet and the scaling function, of the form

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

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  $L^2(\mathbb{R}) \setminus V_0$ . We refer to [60] for additional information about the theory of MRA.

The extension of wavelet theory to higher dimensions requires the introduction of some group theoretic tools. For this, it is useful to start by introducing the continuous *affine systems* of  $L^2(\mathbb{R}^d)$ , which are defined by

$$\{\psi_{M,t} = T_t D_M \psi = |\det M|^{-1/2} \psi(M^{-1}(\cdot - t)) : (M, t) \in G \times \mathbb{R}^d\}. \quad (6)$$

In this definition,  $\psi \in L^2(\mathbb{R}^d)$ ,  $G$  is a subset of  $GL_d(\mathbb{R})$ , the group of  $d$ -dimensional invertible matrices,  $D_M$  is the dilation operator on  $L^2(\mathbb{R}^d)$ , defined by

$$D_M \psi(x) = |\det M|^{-1/2} \psi(M^{-1}x), \quad \text{for } M \in GL_d(\mathbb{R}), \quad (7)$$

and  $T_t$  is the *translation operator* on  $L^2(\mathbb{R}^d)$ , defined by

$$T_t \psi(x) = \psi(x - t), \quad \text{for } t \in \mathbb{R}^d. \quad (8)$$

We now aim to derive conditions on  $\psi$  such that any  $f \in L^2(\mathbb{R}^d)$  can be recovered from its coefficients  $(\langle f, \psi_{M,t} \rangle)_{M,t}$ . For this, we first equip the parameter set of (6) with a group structure by setting

$$(M, t) \cdot (M', t') = (MM', t + Mt').$$

The resulting group, typically denoted by  $\mathcal{A}_d$ , is the so-called *affine group on  $\mathbb{R}^d$* . The mathematical structure of the affine systems becomes evident by observing that (6) can be generated by the action of the unitary representations  $\pi_{(M,t)} = D_M T_t$  of  $\mathcal{A}_d$  acting on  $L^2(\mathbb{R}^d)$  (cf. [38] for details on the theory of group representations). Then the following result on reproducibility of functions in  $L^2(\mathbb{R}^d)$  can be proven.

**Theorem 1 ([27, 58]).** *Retaining the notations introduced in this subsection, let  $\mu$  be a left-invariant Haar measure on  $G \subset GL_d(\mathbb{R})$ , and  $d\lambda$  be a left Haar measure*

of  $\mathcal{A}_d$ . 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

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

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

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

One interesting special case is obtained, when the dilation group  $G$  has the form  $G = \{aI_d : a > 0\}$ , which corresponds to the case of *isotropic dilations*. In this case, the admissibility condition for  $\psi$  becomes

$$\int_{a>0} |\hat{\psi}(a\xi)|^2 \frac{da}{a} = 1,$$

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

$$\mathcal{W}_\psi f(a,t) = a^{-d/2} \int_{\mathbb{R}^d} f(x) \overline{\psi(a^{-1}(x-t))} dx, \quad a > 0, t \in \mathbb{R}^d. \quad (9)$$

Notice that the discrete wavelet systems (3) are obtained by discretizing the continuous affine systems (6) for  $d = 1$ , when choosing isotropic dilations with  $G = \{2^j : j \in \mathbb{Z}\}$ .

## 2.5 Wavelets for Multivariate Data and their Limitations

The traditional theory of wavelets, which is based on the use of isotropic dilations, is essentially a one-dimensional theory. This can be illustrated by looking at the behavior of the isotropic Continuous Wavelet Transform of functions containing singularities. Indeed, consider a function or distribution  $f$ , which is regular everywhere except for a point singularity at  $x_0$ , and let us examine the behaviour of  $\mathcal{W}_\psi f(a,t)$ , given by (9). Provided  $\psi$  is smooth, a direct computations shows that  $\mathcal{W}_\psi f(a,t)$  has rapid asymptotic decay, as  $a \rightarrow 0$ , for all values of  $t$ , unless  $t = x_0$ . In this sense, the Continuous Wavelet Transform of  $f$  signals the location of the singularity through its asymptotic decay at fine scales. More generally, using this property, the Continuous Wavelet Transform can be used to characterize the singular support of a function or distribution [39].

However, due to its isotropic nature, the Continuous Wavelet Transform is unable to provide additional information about the geometry of the set of singularities of a function or distribution in terms of resolving the wavefront set. The key problem is that, although the isotropic wavelet transform has the advantage of simplicity, it lacks directional sensitivity and the ability to detect the geometry of  $f$ . The same phenomenon showing the limitation of the traditional wavelet framework can be illustrated using the Discrete Wavelet Transform.

Before doing this, let us recall the definition of *non-linear approximation* and, in particular, best  $N$ -term approximation, which is the proper notion of approximation in the context of wavelet bases. For a function  $f \in L^2(\mathbb{R}^2)$ , the *best  $N$ -term approximation*  $f_N$  of  $f$  with respect to a wavelet basis is obtained by reconstructing  $f$  from the  $N$  largest wavelet coefficients in magnitude – rather than from the ‘first  $N$ ’, which is the standard approach in linear Fourier approximations. Hence, denoting by  $\Lambda_N$  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  $f \in L^2(\mathbb{R}^2)$  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 instead of a basis, the best  $N$ -term approximation can usually not be explicitly determined. A more detailed discussion of non-linear approximation theory, encompassing the expansion in frames, is contained in Chapter [69].

We can now present a simple heuristic argument, which highlights the limitations of traditional wavelet approximations with respect to more sophisticated multiscale methods – such as the shearlet framework – when aiming at optimally sparse approximations of piecewise functions on  $\mathbb{R}^2$  and, in particular, cartoon-like images. Let  $f$  be a cartoon-like image (see Definition 1) containing a singularity along a smooth curve and  $\{\psi_{j,m}\}$  be a standard wavelet basis of  $L^2(\mathbb{R}^2)$ . For  $j$  sufficiently large, the only significant wavelet coefficients  $\langle f, \psi_{j,m} \rangle$  are those associated with the singularity. Since at scale  $2^{-j}$ , each wavelet  $\psi_{j,m}$  is supported or essentially supported inside a box of size  $2^{-j} \times 2^{-j}$ , there exist about  $2^j$  elements of the wavelet basis overlapping the singularity curve. The associated wavelet coefficients can be controlled by

$$|\langle f, \psi_{j,m} \rangle| \leq \|f\|_\infty \|\psi_{j,m}\|_{L^1} \leq C \cdot 2^{-j}.$$

It follows that the  $N$ -th largest wavelet coefficient in magnitude, which we denote by  $\langle f, \psi_{j,m} \rangle_{(N)}$ , is bounded by  $O(N^{-1})$ . Thus, if  $f$  is approximated by its best- $N$  term approximation  $f_N$ , the  $L^2$ -error obeys

$$\|f - f_N\|_{L^2}^2 \leq \sum_{\ell > N} |\langle f, \psi_{j,m} \rangle_{(\ell)}|^2 \leq CN^{-1}.$$

Indeed, this estimate can be proved rigorously and can be shown to be tight in the sense that there exist cartoon-like images for which the decay rate is of the order  $O(N^{-1})$ , see [60].

However, the approximation rate  $O(N^{-1})$  obtained using wavelet approximations is far from optimal for the class of cartoon-like images  $\mathcal{E}_v^2(\mathbb{R}^2)$ . Indeed, the following optimality result was proved in [14].

**Theorem 2 ([14]).** *Let  $f \in \mathcal{E}^2(\mathbb{R}^2)$ . There exists a constant  $C$  such that, for any  $N$ , a triangulation of  $[0, 1]^2$  with  $N$  triangles can be constructed so that the piecewise linear interpolation  $f_N$  of these triangles satisfies*

$$\|f - f_N\|_2^2 \leq C \cdot N^{-2}, \quad N \rightarrow \infty.$$

This result provides the optimal asymptotic decay rate of the nonlinear approximation error for objects in  $\mathcal{E}^2(\mathbb{R}^2)$ , in the sense that no other polynomial depth search algorithm can yield a better rate. In fact, it shows that the adaptive triangle-based approximation of the image is as good as as if the image had no singularities.

The approximation result from Theorem 2 provides a benchmark for optimally sparse approximation of 2-dimensional data. Furthermore, the argument in the proof of Theorem 2, which uses adapted triangulations, suggests that analyzing elements with elongated and orientable supports are required to achieve optimally sparse approximations of piecewise smooth bivariate functions. Indeed, this observation is at the core of the construction of curvelets and shearlets. Notice however that, unlike the triangulation approximations in Theorem 2, curvelet and shearlet systems are non-adaptive. It is a remarkable fact that, even though they are nonadaptive, curvelet and shearlet representations are able to achieve (essentially) the same optimal approximation rate of Theorem 2. This result will be discussed below, and, in more detail, in Chapter [69].

### 3 Continuous Shearlet Systems

After discussing the limitations of wavelet systems in higher dimensions, we will now introduce shearlet systems as a general framework to overcome these limitations. We will first focus on *continuous* shearlet systems; discrete shearlet systems will be discussed next. As mentioned above, we restrict ourselves to the 2D case.

Before defining the system of shearlets in a formal way, let us introduce intuitively the ideas which are at the core of its construction. Our observations from the previous section suggest that, in order to achieve optimally sparse approximations of signals exhibiting anisotropic singularities such as cartoon-like images, the analyzing elements must consist of waveforms ranging over several scales, orientations, and locations with the ability to become very elongated. 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 2D plane.

Since the scaling operator is required to generate waveforms with anisotropic support, we utilize 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 given by (7). This type of dilations correspond to the so-called *parabolic scaling*, which has a long history in the harmonic analysis literature and can be traced back to the ‘second dyadic decomposition’ from the theory of oscillatory integrals [23, 64] (see also the more recent work by Smith [63] on the decomposition of Fourier integral operators). It should be mentioned that, rather than  $A_a$ , the more general matrices  $\text{diag}(a, a^\alpha)$  with the parameter  $\alpha \in (0, 1)$  controlling the ‘degree of anisotropy’ could be used. 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. In fact, parabolic scaling is required in order to obtain optimally sparse approximations of cartoon-like images, since it is best adapted to the  $C^2$ -regularity of the discontinuity curves in this model class. For simplicity, in the remainder of this chapter, we will only consider the case  $\alpha = 1/2$ , which is required for the sparsity results discussed below. For generalizations and extensions, we refer to Chapters [68] and [69].

Next, we require an orthogonal transformation to change the orientations of the waveforms. The most obvious choice seems to be the rotation operator. However, rotations destroy the structure of the integer lattice  $\mathbb{Z}^2$ , whenever the rotation angle is different from  $0, \pm\frac{\pi}{2}, \pm\pi$ . This issue becomes a serious problem for the transition from the continuum to the digital setting. As an alternative orthogonal transformation, we choose the shearing operator  $D_{S_s}$ ,  $s \in \mathbb{R}$ , where the *shearing matrix*  $S_s$  is 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  $s$  is an integer.

Finally, for the translation operator we use the standard operator  $T_t$  given by (8).

Combining these three operators, we define continuous shearlet systems as follows.

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

$$SH(\psi) = \{\psi_{a,s,t} = T_t D_{A_a} D_{S_s} \psi : a > 0, s \in \mathbb{R}, t \in \mathbb{R}^2\}.$$

The next section will answer the question of how to choose a suitable generating function  $\psi$  so that the system  $SH(\psi)$  satisfies a reproducing formula for  $L^2(\mathbb{R}^2)$ .

### 3.1 Continuous Shearlet Systems and the Shearlet Group

One important structural property of the systems introduced in Definition 2 is their membership in the class of affine systems. Similar to the relation of wavelet systems

to group representation theory discussed in Subsection 2.4, the theory of continuous shearlet systems can also be developed within the theory of unitary representations of the affine group and its generalizations [9].

To state this relation precisely, we define the so-called *shearlet group*, denoted by  $\mathbb{S}$ , as the semi-direct product

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

equipped with group multiplication given by

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

A left-invariant Haar measure of this group is  $\frac{da}{a^3} ds dt$ . Letting the unitary representation  $\sigma : \mathbb{S} \rightarrow \mathcal{U}(L^2(\mathbb{R}^2))$  be defined by

$$\sigma(a, s, t)\psi = T_t D_{A_a} D_{S_s} \psi,$$

where  $\mathcal{U}(L^2(\mathbb{R}^2))$  denotes the group of unitary operators on  $L^2(\mathbb{R}^2)$ , a continuous shearlet system  $SH(\psi)$  can be written as

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

The representation  $\sigma$  is unitary, but not irreducible. If this additional property is desired, the shearlet group needs to be extended to  $(\mathbb{R}^* \times \mathbb{R}) \ltimes \mathbb{R}^2$ , where  $\mathbb{R}^* = \mathbb{R} \setminus \{0\}$ , yielding the continuous shearlet system

$$SH(\psi) = \{\sigma(a, s, t)\psi : a \in \mathbb{R}^*, s \in \mathbb{R}, t \in \mathbb{R}^2\}.$$

This point of view and its generalizations to higher dimensions will be examined in detail in Chapter [67].

In the following, we provide an overview of the main results and definitions related to continuous shearlet systems for  $L^2(\mathbb{R}^2)$ .

### 3.2 The Continuous Shearlet Transform

Similar to the Continuous Wavelet Transform, the Continuous Shearlet Transform defines a mapping of  $f \in L^2(\mathbb{R}^2)$  to the components of  $f$  associated with the elements of  $\mathbb{S}$ .

**Definition 3.** For  $\psi \in L^2(\mathbb{R}^2)$ , the *Continuous Shearlet Transform* of  $f \in L^2(\mathbb{R}^2)$  is the mapping

$$L^2(\mathbb{R}^2) \ni f \rightarrow \mathcal{SH}_\psi f(a, s, t) = \langle f, \sigma(a, s, t)\psi \rangle, \quad (a, s, t) \in \mathbb{S}.$$

Thus,  $\mathcal{SH}_\psi$  maps the function  $f$  to the coefficients  $\mathcal{SH}_\psi f(a, s, t)$  associated with the scale variable  $a > 0$ , the orientation variable  $s \in \mathbb{R}$ , and the location variable  $t \in \mathbb{R}^2$ .

Of particular importance are the conditions on  $\psi$  under which the Continuous Shearlet Transform is an isometry, since this is automatically associated with a reconstruction formula. For this, we define the notion of an *admissible* shearlet, also called *continuous shearlet*.

**Definition 4.** If  $\psi \in L^2(\mathbb{R}^2)$  satisfies

$$\int_{\mathbb{R}^2} \frac{|\hat{\psi}(\xi_1, \xi_2)|^2}{\xi_1^2} d\xi_2 d\xi_1 < \infty,$$

it is called an *admissible shearlet*.

Notice that it is very easy to construct examples of admissible shearlets, including examples of admissible shearlets which are well localized. Essentially any function  $\psi$  such that  $\hat{\psi}$  is compactly supported away from the origin is an admissible shearlet. Of particular importance is the following example, which is called *classical shearlet*. This was originally introduced in [35], and later slightly modified in [28, 57].

**Definition 5.** 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  $\psi_1 \in L^2(\mathbb{R})$  is a discrete wavelet in the sense that it satisfies the discrete Calderón condition, given by

$$\sum_{j \in \mathbb{Z}} |\hat{\psi}_1(2^{-j}\xi)|^2 = 1 \quad \text{for a.e. } \xi \in \mathbb{R}, \quad (10)$$

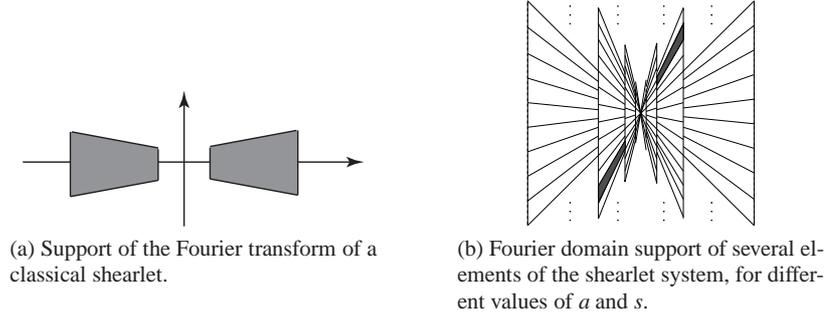
with  $\hat{\psi}_1 \in C^\infty(\mathbb{R})$  and  $\text{supp } \hat{\psi}_1 \subseteq [-\frac{1}{2}, -\frac{1}{16}] \cup [\frac{1}{16}, \frac{1}{2}]$ , and  $\psi_2 \in L^2(\mathbb{R})$  is a ‘bump function’ in the sense that

$$\sum_{k=-1}^1 |\hat{\psi}_2(\xi + k)|^2 = 1 \quad \text{for a.e. } \xi \in [-1, 1], \quad (11)$$

satisfying  $\hat{\psi}_2 \in C^\infty(\mathbb{R})$  and  $\text{supp } \hat{\psi}_2 \subseteq [-1, 1]$ . Then  $\psi$  is called a *classical shearlet*.

Thus, a classical shearlet  $\psi$  is a function which is wavelet-like in one direction and bump-like in the orthogonal one. The frequency support of a classical shearlet is illustrated in Figure 3a. Notice that there exist several choices of  $\psi_1$  and  $\psi_2$  satisfying conditions (10) and (11). One possible choice is to set  $\psi_1$  to be a Lemarié–Meyer wavelet and  $\psi_2$  to be a spline, see [29].

The notion of admissible shearlets allows us to state sufficient conditions for a reconstruction formula in  $L^2(\mathbb{R}^2)$ .



**Fig. 3** Classical shearlets.

**Theorem 3 ([9]).** Let  $\psi \in L^2(\mathbb{R}^2)$  be an admissible shearlet, and define

$$C_{\psi}^{+} = \int_0^{\infty} \int_{\mathbb{R}} \frac{|\hat{\psi}(\xi_1, \xi_2)|^2}{\xi_1^2} d\xi_2 d\xi_1 \quad \text{and} \quad C_{\psi}^{-} = \int_{-\infty}^0 \int_{\mathbb{R}} \frac{|\hat{\psi}(\xi_1, \xi_2)|^2}{\xi_1^2} d\xi_2 d\xi_1.$$

If  $C_{\psi}^{-} = C_{\psi}^{+} = 1$ , then  $\mathcal{SH}_{\psi}$  is an isometry.

*Proof.* By the Plancherel theorem, we obtain

$$\begin{aligned} & \int_{\mathbb{S}} |\mathcal{SH}_{\psi} f(a, s, t)|^2 \frac{da}{a^3} ds dt \\ &= \int_{\mathbb{S}} |f * \psi_{a,s,0}^*|^2 dt ds \frac{da}{a^3} \\ &= \int_0^{\infty} \int_{\mathbb{R}} \int_{\mathbb{R}^2} |\hat{f}(\xi)|^2 |\widehat{\psi_{a,s,0}^*}(\xi)|^2 d\xi ds \frac{da}{a^3} \\ &= \int_0^{\infty} \int_{\mathbb{R}^2} \int_{\mathbb{R}} |\hat{f}(\xi)|^2 a^{-\frac{3}{2}} |\hat{\psi}(a\xi_1, \sqrt{a}(\xi_2 + s\xi_1))|^2 ds d\xi da, \end{aligned}$$

where we used the notation  $\psi^*(x) = \overline{\psi(-x)}$ . By appropriate changes of variables,

$$\begin{aligned} & \int_{\mathbb{S}} |\mathcal{SH}_{\psi} f(a, s, t)|^2 \frac{da}{a^3} ds dt \\ &= \int_{\mathbb{R}} \int_0^{\infty} \int_0^{\infty} \int_{\mathbb{R}} |\hat{f}(\xi)|^2 a^{-2} \xi_1^{-1} |\hat{\psi}(a\xi_1, \omega_2)|^2 d\omega_2 da d\xi_1 d\xi_2 \\ &\quad - \int_{\mathbb{R}} \int_{-\infty}^0 \int_0^{\infty} \int_{\mathbb{R}} |\hat{f}(\xi)|^2 a^{-2} \xi_1^{-1} |\hat{\psi}(a\xi_1, \omega_2)|^2 d\omega_2 da d\xi_1 d\xi_2 \\ &= \int_{\mathbb{R}} \int_0^{\infty} |\hat{f}(\xi)|^2 d\xi_1 d\xi_2 \int_0^{\infty} \int_{\mathbb{R}} \frac{|\hat{\psi}(\omega_1, \omega_2)|^2}{\omega_1^2} d\omega_2 d\omega_1 \\ &\quad + \int_{\mathbb{R}} \int_{-\infty}^0 |\hat{f}(\xi)|^2 d\xi_1 d\xi_2 \int_{-\infty}^0 \int_{\mathbb{R}} \frac{|\hat{\psi}(\omega_1, \omega_2)|^2}{\omega_1^2} d\omega_2 d\omega_1. \end{aligned}$$

The claim follows from here.  $\square$

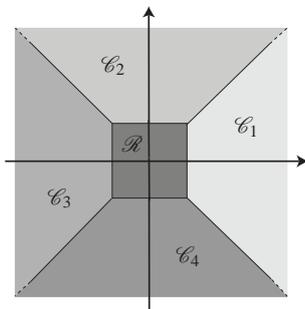
The classical shearlets, given in Definition 5, satisfy the hypothesis of admissibility, as the following result shows. The proof is straightforward, wherefore we omit it.

**Lemma 1 ([9]).** *Let  $\psi \in L^2(\mathbb{R}^2)$  be a classical shearlet. Retaining the notations from Theorem 3, we have  $C_{\psi}^- = C_{\psi}^+ = 1$ .*

### 3.3 Cone-Adapted Continuous Shearlet Systems

Although the continuous shearlet systems defined above exhibit an elegant group structure, they do have a directional bias, which is already recognizable in Figure 3b. To illustrate the impact of this directional bias, consider a function or distribution which is mostly concentrated along the  $\xi_2$  axis in the frequency domain. Then the energy of  $f$  is more and more concentrated in the shearlet components  $\mathcal{SH}_{\psi} f(a, s, t)$  as  $s \rightarrow \infty$ . Hence, in the limiting case in which  $f$  is a delta distribution supported along the  $\xi_2$  axis, – this typically serves as a model for an edge along the  $x_1$  axis in spatial domain –  $f$  can only be ‘detected’ in the shearlet domain as  $s \rightarrow \infty$ . It is clear that this behavior can be a serious limitation for some applications.

One way to address this problem is to partition the Fourier domain into four cones, while separating the low-frequency region by cutting out a square centered around the origin. This yields a partition of the frequency plane as illustrated in Figure 4. Notice that, within each cone, the shearing variable  $s$  is only allowed to vary over a finite range, hence producing elements whose orientations are distributed more uniformly.



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

Thus, we define the following variant of continuous shearlet systems.

**Definition 6.** For  $\phi, \psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$ , the *cone-adapted continuous shearlet system*  $SH(\phi, \psi, \tilde{\psi})$  is defined by

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

where

$$\Phi(\phi) = \{\phi_t = \phi(\cdot - t) : t \in \mathbb{R}^2\},$$

$$\Psi(\psi) = \{\psi_{a,s,t} = a^{-\frac{3}{4}} \psi(A_a^{-1} S_s^{-1}(\cdot - t)) : a \in (0, 1], |s| \leq 1 + a^{1/2}, t \in \mathbb{R}^2\},$$

$$\tilde{\Psi}(\tilde{\psi}) = \{\tilde{\psi}_{a,s,t} = a^{-\frac{3}{4}} \tilde{\psi}(\tilde{A}_a^{-1} S_s^{-T}(\cdot - t)) : a \in (0, 1], |s| \leq 1 + a^{1/2}, t \in \mathbb{R}^2\},$$

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

In the following, the function  $\phi$  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  $\mathcal{R} = \{(\xi_1, \xi_2) : |\xi_1|, |\xi_2| \leq 1\}$ . By choosing  $\psi$  to satisfy the conditions of Definition 5, the system  $\Psi(\psi)$  is associated with the horizontal cones  $\mathcal{C}_1 \cup \mathcal{C}_3 = \{(\xi_1, \xi_2) : |\xi_2/\xi_1| \leq 1, |\xi_1| > 1\}$ . The shearlet  $\tilde{\psi}$  can be chosen likewise with the roles of  $\xi_1$  and  $\xi_2$  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  $\mathcal{C}_2 \cup \mathcal{C}_4 = \{(\xi_1, \xi_2) : |\xi_2/\xi_1| > 1, |\xi_2| > 1\}$ .

### 3.4 The Cone-Adapted Continuous Shearlet Transform

Similar as in the situation of continuous shearlet systems, also for cone-adapted continuous shearlet systems an associated transform can be defined.

**Definition 7.** Set

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

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 mapping

$$f \rightarrow \mathcal{SH}_{\phi, \psi, \tilde{\psi}} f(t', (a, s, t), (\tilde{a}, \tilde{s}, \tilde{t})) = (\langle f, \phi_{t'} \rangle, \langle f, \psi_{a,s,t} \rangle, \langle f, \tilde{\psi}_{\tilde{a}, \tilde{s}, \tilde{t}} \rangle),$$

where

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

Similar to the situation above, conditions on  $\psi, \tilde{\psi}$  and  $\phi$  can be formulated for which the mapping  $\mathcal{SH}_{\phi, \psi, \tilde{\psi}}$  is an isometry. In fact, a similar argument to the one used in the proof of Theorem 3, yields the following result.

**Theorem 4 ([48]).** *Retaining the notation of Theorem 3, 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.  $\xi = (\xi_1, \xi_2) \in \mathbb{R}^2$ ,*

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

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

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

In this result, the functions  $\phi$ ,  $\psi$ , and  $\tilde{\psi}$  can in fact 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 [26] and in Chapter [66].

### 3.5 Microlocal Properties and Characterization of Singularities

As observed in Subsection 2.5, the Continuous Wavelet Transform is able to precisely characterize the singular support functions or distributions. However, due to its isotropic nature, this approach fails to provide additional information about the geometry of the set of singularities in the sense of resolving the wavefront set.

In contrast to this behavior, the anisotropic shape of elements of a cone-adapted continuous shearlet system enables the Continuous Shearlet Transform to very precisely characterize the geometric properties of the singularity set. For illustration purposes, let us examine the linear delta distribution  $\mu_p(x_1, x_2) = \delta(x_1 + px_2)$ ,  $p \in \mathbb{R}$ , defined by

$$\langle \mu_p, f \rangle = \int_{\mathbb{R}} f(-px_2, x_2) dx_2,$$

as a simple model for a distributed singularity. For simplicity, we assume that  $|p| \leq 1$ . Letting  $\phi$  be a scaling function and  $\psi, \tilde{\psi}$  be classical shearlets, the asymptotic analysis of its Cone-Adapted Continuous Shearlet Transform  $\mathcal{SH}_{\phi, \psi, \tilde{\psi}} \mu_p$  shows that this transform precisely determines both the position and the orientation of the linear singularity by its decay behavior at fine scales. Specifically, we have the following result.

**Proposition 1 ([48]).** *Let  $t' \in \mathbb{R}^2$  and  $(\tilde{a}, \tilde{s}, \tilde{t}) \in \mathbb{S}_{\text{cone}}$  be a fixed value. For  $t_1 = -pt_2$  and  $s = p$ , we have*

$$\mathcal{SH}_{\phi, \psi, \tilde{\psi}} \mu_p(t', (a, s, t), (\tilde{a}, \tilde{s}, \tilde{t})) \sim a^{-\frac{1}{4}} \quad \text{as } a \rightarrow 0.$$

*In all other cases,  $\mathcal{SH}_{\phi, \psi, \tilde{\psi}} \mu_p(t', (a, s, t), (\tilde{a}, \tilde{s}, \tilde{t}))$  decays rapidly as  $a \rightarrow 0$ .*

In fact, it can be proven that the Cone-Adapted Continuous Shearlet Transform precisely resolves the wavefront set for more general distributions [48, 25]. Furthermore, it 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  $\partial B$  by its decay at fine scales [30, 34]. This property is very useful in applications which require the analysis or detection of edge discontinuities. For example, using on these observations, a shearlet-based algorithm for edge detection and analysis was developed in [65], and related ideas were exploited to develop algorithms for the regularized inversion of the Radon transform in [6].

A more detailed discussion of these issues, including the extensions to higher dimensions, will be the content of Chapters [66] and [68].

## 4 Discrete Shearlet Systems

Starting from continuous shearlet systems defined in Definition 2, several discrete versions of discrete shearlet systems can be constructed by an appropriate sampling of the continuous parameter set  $\mathbb{S}$  or  $\mathbb{S}_{cone}$ . 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 proposed in [8] and continued in [10] and [11] applies a powerful methodology called *coorbit theory*, which is used to derive different discretizations while ensuring frame properties. In fact, the regular shearlet frame which will be introduced in the next subsection can be derived using this machinery, and this approach will be further discussed in Chapter [67]. A different path, which also relies on the group properties of continuous shearlet systems, is taken in [46]. In this paper, a quantitative density measure for discrete subsets of the shearlet group  $\mathbb{S}$  is introduced, adapted to the group multiplication, which is inspired by the well-known Beurling density for subsets of the Abelian group  $\mathbb{R}^2$ . These measures are shown to provide necessary conditions on the density of the sampling set for the existence of shearlet generators which yield a frame; thereby linking geometric properties of the sampling set to the frame properties of the resulting shearlet system. Notice, however, that the conditions derived using this approach are necessary but not sufficient. In a third approach [47], sufficient conditions are derived by studying the classical  $t_q$ -equations from the theory of wavelets. Recall that these equations are part of the sufficient conditions needed for an affine system to form a wavelet orthonormal basis or a tight frame (see [43] for a detailed discussion on this topic). Due to the close relationship between shearlet systems and affine systems discussed in Subsection 3.1, this ansatz can be transferred to the situation of cone-adapted continuous shearlet systems [45]. This will be further discussed in Chapter [69].

### 4.1 Discrete Shearlet Systems and Transforms

Discrete shearlet systems are formally defined by sampling continuous shearlet systems on a discrete subset of the shearlet group  $\mathbb{S}$ . This leads to the following definition.

**Definition 8.** Let  $\psi \in L^2(\mathbb{R}^2)$  and  $\Lambda \subseteq \mathbb{S}$ . An *irregular discrete shearlet system* associated with  $\psi$  and  $\Lambda$ , denoted by  $SH(\psi, \Lambda)$ , is defined by

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

A (*regular*) *discrete shearlet system* associate with  $\psi$ , denoted by  $SH(\psi)$ , is defined by

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

Notice that the regular versions of discrete shearlet systems are derived from the irregular systems by choosing  $\Lambda = \{(2^{-j}, -k, S_{-k} A_{2^j} m) : j, k \in \mathbb{Z}, m \in \mathbb{Z}^2\}$ . We also remark that, in the definition of a regular discrete shearlet system, the translation parameter is sometimes 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.

Our goal is to apply shearlet systems as analysis and synthesis tools. Hence, it is of particular interest to examine the situation in which a discrete shearlet system  $SH(\psi)$  forms a basis or, more generally, a frame. Similar to the wavelet case, we are particularly interested not only in finding generic generator functions  $\psi$ , but also in selecting a generator  $\psi$  with special properties, e.g., regularity, vanishing moments, compact support, so that the corresponding basis or frame of shearlets has satisfactory approximation properties. A particularly useful example are the classical shearlets from Definition 5. As the following result shows, these shearlets generate shearlet Parseval frames for  $L^2(\mathbb{R}^2)$ .

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

*Proof.* Using the properties of classical shearlets as stated in Definition 5, a direct computations gives that, for *a.e.*  $\xi \in \mathbb{R}^2$ ,

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

The claim follows immediately from this observation and the fact that  $\text{supp } \hat{\psi} \subset [-\frac{1}{2}, \frac{1}{2}]^2$ .  $\square$

Since a classical shearlet  $\psi$  is a well localized function, Proposition 2 implies that there exist Parseval frames  $SH(\psi)$  of well localized discrete shearlets. The well-localization property is critical for deriving superior approximation properties of

shearlet systems, and will be required for deriving optimally sparse approximations of cartoon-like images (cf. Subsection 4.4).

By removing the assumption that  $\psi$  is well localized in Definition 5, one can construct a discrete shearlet systems which form not only tight frames, but also orthonormal basis, as indicated in [35, 37]. This naturally raises the question about whether well localized shearlet orthonormal bases do exist. Unfortunately, the answer seems to be negative, according to the recent work in [44], which virtually excludes this possibility. Thus, loosely speaking, a well localized discrete shearlet system can form a frame or a tight frame, but not an orthonormal basis.

To achieve spatial domain localization, compactly supported discrete shearlet systems are required. It was recently shown that one can formulate sufficient conditions on  $\psi$  to generate a discrete shearlet frame of compactly supported functions with controllable frame bounds. This will be discussed in Subsection 4.3.

Finally, similar to the continuous case, we define a Discrete Shearlet Transform as follows. We state this definition only for the regular case, with obvious extension to the irregular shearlet systems.

**Definition 9.** For  $\psi \in L^2(\mathbb{R}^2)$ , the *Discrete Shearlet Transform* of  $f \in L^2(\mathbb{R}^2)$  is the mapping defined by

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

Thus,  $\mathcal{SH}_\psi$  maps the function  $f$  to the coefficients  $\mathcal{SH}_\psi f(j, k, m)$  associate with the scale index  $j$ , the orientation index  $k$ , and the position index  $m$ .

## 4.2 Cone-Adapted Discrete Shearlet Systems and Transforms

Similar to the situation of continuous shearlet systems, also discrete shearlet systems suffer from a biased treatment of the directions. As expected, this problem can be addressed by dividing the frequency plane into cones similar to Subsection 3.3. For the sake of generality, let us start by defining cone-adapted discrete shearlet systems with respect to an irregular parameter set.

**Definition 10.** Let  $\phi, \psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$ ,  $\Delta \subset \mathbb{R}^2$  and  $\Lambda, \tilde{\Lambda} \subset \mathbb{S}_{cone}$ . Then the *irregular cone-adapted discrete shearlet system*  $SH(\phi, \psi, \tilde{\psi}; \Delta, \Lambda, \tilde{\Lambda})$  is defined by

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

where

$$\Phi(\phi; \Delta) = \{\phi_t = \phi(\cdot - t) : t \in \Delta\},$$

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

$$\tilde{\Psi}(\tilde{\psi}; \tilde{\Lambda}) = \{\tilde{\psi}_{a, s, t} = a^{-\frac{3}{4}} \tilde{\psi}(\tilde{A}_a^{-1} S_s^{-T}(\cdot - t)) : (a, s, t) \in \tilde{\Lambda}\}.$$

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

**Definition 11.** 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  $SH(\phi, \psi, \tilde{\psi}; c)$  is defined by

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

where

$$\Phi(\phi; c_1) = \{\phi_m = \phi(\cdot - c_1 m) : m \in \mathbb{Z}^2\},$$

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

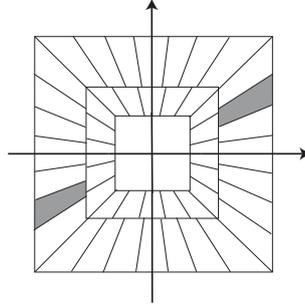
$$\tilde{\Psi}(\tilde{\psi}; c) = \{\tilde{\psi}_{j,k,m} = 2^{\frac{3}{4}j} \tilde{\psi}(S_k^T \tilde{A}_{2^j} \cdot - \tilde{M}_c m) : j \geq 0, |k| \leq \lceil 2^{j/2} \rceil, m \in \mathbb{Z}^2\},$$

with

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

If  $c = (1, 1)$ , the parameter  $c$  is omitted in the formulae above.

The generating functions  $\phi$  will be referred to as *shearlet scaling functions* and the generating functions  $\psi, \tilde{\psi}$  as *shearlet generators*. Notice that 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 region  $\mathcal{C}_1 \cup \mathcal{C}_3$  and  $\mathcal{C}_2 \cup \mathcal{C}_4$ , respectively (cf. Figure 4).



**Fig. 5** Tiling of the frequency plane induced by cone-adapted Parseval frame of shearlets.

We already discussed the difficulties – or even the impossibility – to construct a discrete shearlet orthonormal basis. Hence, one aims to derive Parseval frames. A first step towards this goal is the observation that a classical shearlet, according to

Definition 5, is a shearlet generator of a Parseval frame for the subspace of  $L^2(\mathbb{R}^2)$  of function whose frequency support lies in the union of two cones  $\mathcal{C}_1 \cup \mathcal{C}_3$ .

**Theorem 5 ([28]).** *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  $L^2(\mathcal{C}_1 \cup \mathcal{C}_3)^\vee = \{f \in L^2(\mathbb{R}^2) : \text{supp } \hat{f} \subset \mathcal{C}_1 \cup \mathcal{C}_3\}$ .

*Proof.* Let  $\psi$  be a classical shearlet. Then equation (11) implies that, for any  $j \geq 0$ ,

$$\sum_{|k| \leq \lceil 2^{j/2} \rceil} |\hat{\psi}_2(2^{j/2} \xi + k)|^2 = 1, \quad |\xi| \leq 1.$$

Thus, using this observation together with equation (10), a direct computations gives that, for a.e.  $\xi = (\xi_1, \xi_2) \in \mathcal{C}_1 \cup \mathcal{C}_3$ ,

$$\begin{aligned} \sum_{j \geq 0} \sum_{|k| \leq \lceil 2^{j/2} \rceil} |\hat{\psi}(S_{-k}^T A_{2^{-j}} \xi)|^2 &= \sum_{j \geq 0} \sum_{|k| \leq \lceil 2^{j/2} \rceil} |\hat{\psi}_1(2^{-j} \xi_1)|^2 |\hat{\psi}_2(2^{j/2} \frac{\xi_2}{\xi_1} - k)|^2 \\ &= \sum_{j \geq 0} |\hat{\psi}_1(2^{-j} \xi_1)|^2 \sum_{|k| \leq \lceil 2^{j/2} \rceil} |\hat{\psi}_2(2^{j/2} \frac{\xi_2}{\xi_1} + k)|^2 = 1. \end{aligned}$$

The claim follows immediately from this observation and the fact that  $\text{supp } \hat{\psi} \subset [-\frac{1}{2}, \frac{1}{2}]^2$ .  $\square$

It is clear that, if  $\psi$  is replaced by  $\tilde{\psi}$ , a result very similar to Theorem 5 holds for the subspace of  $L^2(\mathcal{C}_2 \cup \mathcal{C}_4)^\vee$ . This strongly suggests that 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. However, this ‘piecing together’ needs to be performed with care, in particular, by projecting the shearlet elements on the respective cones. Then we have the following result.

**Theorem 6 ([28]).** *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.  $\xi \in \mathbb{R}^2$ ,*

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

Let  $P_C \Psi(\psi)$  denote the set of elements in  $\Psi(\psi)$  after projecting their Fourier transform onto  $C = \{(\xi_1, \xi_2) \in \mathbb{R}^2 : |\xi_2/\xi_1| \leq 1\}$  with a similar definition for  $P_{\tilde{C}} \tilde{\Psi}(\tilde{\psi})$  with  $\tilde{C} = \mathbb{R}^2 \setminus C$ . 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  $L^2(\mathbb{R}^2)$ .

The tiling of the frequency plane induced by this Parseval frame of shearlets is illustrated in Figure 5. The shearlet transform associated to regular cone-adapted discrete shearlet systems is defined as follows.

**Definition 12.** 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 mapping defined by

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

where

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

### 4.3 Compactly Supported Shearlets

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. Thus, a different approach is needed for the construction of compactly supported shearlet systems.

We start our discussion by examining sufficient conditions for the existence of cone-adapted discrete shearlet systems which are compactly supported and form a frame for  $L^2(\mathbb{R}^2)$ . These conditions can be derived by extending the classical  $t_q$ -equations from the theory of wavelets to this situation. Before stating the main result, let us first introduce the following notation.

For functions  $\phi, \psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$ , we define  $\Theta : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$  by

$$\Theta(\xi, \omega) = |\hat{\phi}(\xi)| |\hat{\phi}(\xi + \omega)| + \Theta_1(\xi, \omega) + \Theta_2(\xi, \omega),$$

where

$$\Theta_1(\xi, \omega) = \sum_{j \geq 0} \sum_{|k| \leq \lceil 2^{j/2} \rceil} |\hat{\psi}(S_k^T A_{2^{-j}} \xi)| |\hat{\psi}(S_k^T A_{2^{-j}} \xi + \omega)|$$

and

$$\Theta_2(\xi, \omega) = \sum_{j \geq 0} \sum_{|k| \leq \lceil 2^{j/2} \rceil} |\hat{\tilde{\psi}}(S_k \tilde{A}_{2^{-j}} \xi)| |\hat{\tilde{\psi}}(S_k \tilde{A}_{2^{-j}} \xi + \omega)|.$$

Also, for  $c = (c_1, c_2) \in (\mathbb{R}_+)^2$ , let

$$R(c) = \sum_{m \in \mathbb{Z}^2 \setminus \{0\}} (\Gamma_0(c_1^{-1} m) \Gamma_0(-c_1^{-1} m))^{\frac{1}{2}} + (\Gamma_1(M_c^{-1} m) \Gamma_1(-M_c^{-1} m))^{\frac{1}{2}} \\ + (\Gamma_2(\tilde{M}_c^{-1} m) \Gamma_2(-\tilde{M}_c^{-1} m))^{\frac{1}{2}},$$

where

$$\Gamma_0(\omega) = \operatorname{ess\,sup}_{\xi \in \mathbb{R}^2} |\hat{\phi}(\xi)| |\hat{\phi}(\xi + \omega)| \quad \text{and} \quad \Gamma_i(\omega) = \operatorname{ess\,sup}_{\xi \in \mathbb{R}^2} \Theta_i(\xi, \omega) \quad \text{for } i = 1, 2.$$

Using this notation, we can now state the following theorem from [45].

**Theorem 7 ([45]).** *Let  $\phi, \psi \in L^2(\mathbb{R}^2)$  be such that*

$$\hat{\phi}(\xi_1, \xi_2) \leq C_1 \cdot \min\{1, |\xi_1|^{-\gamma}\} \cdot \min\{1, |\xi_2|^{-\gamma}\}$$

and

$$|\hat{\psi}(\xi_1, \xi_2)| \leq C_2 \cdot \min\{1, |\xi_1|^\alpha\} \cdot \min\{1, |\xi_1|^{-\gamma}\} \cdot \min\{1, |\xi_2|^{-\gamma}\},$$

for some positive constants  $C_1, C_2 < \infty$  and  $\alpha > \gamma > 3$ . Define  $\tilde{\psi}(x_1, x_2) = \psi(x_2, x_1)$ , and let  $L_{\inf}, L_{\sup}$  be defined by

$$L_{\inf} = \operatorname{ess\,inf}_{\xi \in \mathbb{R}^2} \Theta(\xi, 0) \quad \text{and} \quad L_{\sup} = \operatorname{ess\,sup}_{\xi \in \mathbb{R}^2} \Theta(\xi, 0).$$

Then there exists a sampling parameter  $c = (c_1, c_2) \in (\mathbb{R}^+)^2$  with  $c_1 = c_2$  such that  $SH(\phi, \psi, \tilde{\psi}; c)$  forms a frame for  $L^2(\mathbb{R}^2)$  with frame bounds  $A$  and  $B$  satisfying

$$0 < \frac{1}{|\det M_c|} [L_{\inf} - R(c)] \leq A \leq B \leq \frac{1}{|\det M_c|} [L_{\sup} + R(c)] < \infty.$$

It can be easily verified that the conditions imposed on  $\phi$  and  $\psi$  by Theorem 7 are satisfied by many suitably chosen scaling functions and classical shearlets. In addition, one can construct various compactly supported *separable* shearlets that satisfy these conditions.

The difficulty however arises when aiming for compactly supported *separable* functions  $\phi$  and  $\psi$  which ensure that the corresponding cone adapted discrete shearlet system is a tight or almost tight frame. Separability is useful to achieve fast algorithmic implementations. In fact, it was shown in [45] that there exists a class of functions generating almost tight frames, which have (essentially) the form

$$\hat{\psi}(\xi) = m_1(4\xi_1) \hat{\phi}(\xi_1) \hat{\phi}(2\xi_2), \quad \xi = (\xi_1, \xi_2) \in \mathbb{R}^2,$$

where  $m_1$  is a carefully chosen bandpass filter and  $\phi$  an adaptively chosen scaling function. The proof of this fact is of highly technical nature and will be omitted. We refer the reader to Chapter [69] and to the survey paper [49] for more details.

#### 4.4 Sparse Approximations by Shearlets

One of the main motivations for the introduction of the shearlet framework is the derivation of optimally sparse approximations for multivariate functions. In Subsection 2.5, we presented a heuristic argument to justify why traditional wavelets are unable to take advantage of the geometry of typical functions of two variables. In fact, since traditional wavelets are not very efficient at dealing with anisotropic features, they do not provide optimally sparse approximations of images containing edges. As discussed above, shearlet systems are able to overcome these limitations.

Before stating the main results, it is enlightening to present an heuristic argument similar to the one used in Subsection 2.5, in order to describe how shearlet expansions are able to achieve optimally sparse approximations of cartoon-like images.

For this, consider a cartoon-like function  $f$ , and let  $SH(\phi, \psi, \tilde{\psi}; c)$  be a shearlet system. Since the elements of  $SH(\phi, \psi, \tilde{\psi}; c)$  are essentially or – in case of compactly supported elements – exactly supported inside a box of size  $2^{-j/2} \times 2^{-j}$ , it follows that, at scale  $2^{-j}$ , there exist about  $O(2^{j/2})$  such waveforms whose support is tangent to the discontinuity curve. Similar to the wavelet case, for  $j$  sufficiently large, the shearlet elements associated with the smooth region of  $f$ , as well as the elements whose overlap with the discontinuity curve is non-tangential, yield negligible shearlet coefficients  $\langle f, \psi_{j,k,m} \rangle$  (or  $\langle f, \tilde{\psi}_{j,k,m} \rangle$ ). Each shearlet coefficient can be controlled by

$$|\langle f, \psi_{j,k,m} \rangle| \leq \|f\|_\infty \|\psi_{j,k,m}\|_{L^1} \leq C 2^{-3j/4},$$

similarly for  $\langle f, \tilde{\psi}_{j,k,m} \rangle$ . Using this estimate and the observation that there exist at most  $O(2^{j/2})$  significant coefficients, we can conclude that the  $N$ -th largest shearlet coefficient, which we denote by  $|s_N(f)|$ , is bounded by  $O(N^{-3/2})$ . This implies that

$$\|f - f_N\|_{L^2}^2 \leq \sum_{\ell > N} |s_\ell(f)|^2 \leq CN^{-2},$$

where  $f_N$  denotes the  $N$ -term shearlet approximation using the  $N$  largest coefficients in the shearlets expansion. This is exactly the optimal approximation error rate of Theorem 2. Even though this is simple heuristic argument, it provides an error rate which – up to a log-like factor – coincides exactly with what can be proved using a rigorous argument.

Indeed, the following result holds.

**Theorem 8 ([29]).** *Let  $\Phi(\phi) \cup P_C\Psi(\psi) \cup P_C\tilde{\Psi}(\tilde{\psi})$  be a Parseval frame for  $L^2(\mathbb{R}^2)$  as defined in Theorem 6, where  $\psi \in L^2(\mathbb{R}^2)$  is a classical shearlet and  $\hat{\phi} \in C_0^\infty(\mathbb{R}^2)$ .*

*Let  $f \in \mathcal{E}^2(\mathbb{R}^2)$  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*

$$\|f - f_N\|_2^2 \leq C \cdot N^{-2} \cdot (\log N)^3 \quad \text{as } N \rightarrow \infty.$$

Since a log-like factor is negligible with respect to the other terms for large  $N$ , the optimal error decay rate is essentially achieved. It is remarkable that, by using a non-adaptive system, an approximation rate which is essentially as good as the one obtained using an adaptive construction can be achieved. The same approximation rate – with the same additional log-like factor – is obtained using a Parseval frame of curvelets, see [4].

Interestingly, the same error decay rate is also achieved using approximations based on compactly supported shearlet frames, as stated below.

**Theorem 9 ([51]).** *Let  $SH(\phi, \psi, \tilde{\psi}; c)$  be a frame for  $L^2(\mathbb{R}^2)$ , where  $c > 0$ , and  $\phi, \psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$  are compactly supported functions such that, for all  $\xi = (\xi_1, \xi_2) \in \mathbb{R}^2$ , the shearlet  $\psi$  satisfies*

$$(i) \quad |\hat{\psi}(\xi)| \leq C_1 \cdot \min\{1, |\xi_1|^\alpha\} \cdot \min\{1, |\xi_1|^{-\gamma}\} \cdot \min\{1, |\xi_2|^{-\gamma}\} \quad \text{and}$$

$$(ii) \quad \left| \frac{\partial}{\partial \xi_2} \hat{\psi}(\xi) \right| \leq |h(\xi_1)| \cdot \left( 1 + \frac{|\xi_2|}{|\xi_1|} \right)^{-\gamma},$$

where  $\alpha > 5$ ,  $\gamma \geq 4$ ,  $h \in L^1(\mathbb{R})$ ,  $C_1$  is a constant, and the shearlet  $\tilde{\psi}$  satisfies (i) and (ii) with the roles of  $\xi_1$  and  $\xi_2$  reversed.

Let  $f \in \mathcal{E}^2(\mathbb{R}^2)$  and  $f_N$  be its nonlinear  $N$ -term approximation obtained by selecting the  $N$  largest coefficients in the expansion of  $f$  with respect to the shearlet frame  $SH(\phi, \psi, \tilde{\psi}; c)$ . Then there exists a constant  $C > 0$ , independent of  $f$  and  $N$ , such that

$$\|f - f_N\|_2^2 \leq C \cdot N^{-2} \cdot (\log N)^3 \quad \text{as } N \rightarrow \infty.$$

Conditions (i) and (ii) are rather mild conditions and might be regarded as a weak version of directional vanishing moment conditions.

The topic of sparse shearlet approximations, including extensions to higher dimensions, will be the main topic of Chapter [69].

#### 4.5 Shearlet Function Spaces

As already mentioned in Subsection 1.2, the study of the smoothness spaces associated with shearlet coefficients is particularly useful to thoroughly understand and take advantage of the approximation properties of shearlet representations. Intuitively, shearlet systems can be described as directional versions of wavelet systems. Hence, since wavelets are known to be naturally associated with Besov spaces (in the sense that Besov spaces are characterized by the decay of wavelet coefficients), it seems conceivable that shearlet systems could be effective at characterizing some anisotropic version of Besov spaces.

The theory of coorbit spaces was applied as a systematic approach towards the construction of ‘shearlet spaces’ in the series of papers [8, 10, 11]. This ansatz leads to the so-called *shearlet coorbit spaces*, which are associated to decay properties of shearlet coefficients of discrete shearlet frames. The main challenge then consists in relating these spaces to known function spaces such as anisotropic Besov spaces and deriving embedding results. Chapter [67] provides a thorough introduction to this topic and a survey of recent results.

#### 4.6 Extensions and Generalizations

A number of recent studies have focused on the construction of shearlet systems, which are tailored to specific tasks or applications.

- *Shearlet on bounded domains.* Some applications such as, for example, the construction of numerical solvers of certain partial differential equations require systems defined on bounded domains. This could be a rectangle or, more generally, a polygonal-shaped domain. When shearlets are used for the expansion of functions – explicitly or implicitly given – defined on a bounded domain, the treatment of the boundary is crucial. One typical challenge is to set zero boundary conditions without destroying necessary (directional) vanishing moment conditions. A first attempt in this direction was undertaken in [53], but many challenges are still open.
- *Multidimensional extensions.* Many current high-impact applications such as, for example, the analysis of seismic or biological data require to deal with 3-dimensional data. The computational challenges in this setting are much more demanding than in two dimensions, and sparse approximations are in great demand. Due to the simplicity of the mathematical structure of shearlets, their extensions to higher dimensions is very natural. Indeed, some basic ideas were already introduced in [37], where it was observed that there exist several ways to extend the shearing matrix to higher dimensions. Additional observations and results have more recently appeared, including results on sparse approximations and detection of singularities [10, 31, 32, 33, 50].

In 3-dimensional data, different types of anisotropic features occur, namely, singularities on 1-dimensional and 2-dimensional manifolds. This situation is therefore very different from the situation in 2 dimensions, since anisotropic features of two different dimensions are involved. This is reflected in the following two main approaches to extend the parabolic scaling matrix:

$$\begin{pmatrix} 2^j & 0 & 0 \\ 0 & 2^{j/2} & 0 \\ 0 & 0 & 2^j \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 2^j & 0 & 0 \\ 0 & 2^{j/2} & 0 \\ 0 & 0 & 2^{j/2} \end{pmatrix}.$$

The first choice leads to ‘needle-like’ shearlets, which are intuitively better suited to capture 1-dimensional singularities. The second choice leads to ‘plate-like’ shearlets, which are more suited to 2-dimensional singularities. Intriguingly, both systems are needed if the goal is to distinguish these two types of singularities. However, for the construction of (nearly) optimally sparse approximations which extend the results of Subsection 4.4, it turns out that the ‘plate-like’ shearlets are the right approach [32, 33, 50].

These topics will be further discussed in the Chapters [67], [68], and [69].

## 5 Algorithmic Implementations of the Shearlet Transform

One major feature of the shearlet approach is a unified treatment of the continuum and digital setting. The numerical implementations, which have been developed in the literature, aim – and succeed – to faithfully digitalize the continuum domain Dis-

crete Shearlet Transform. This ensures that microlocal and approximation properties of shearlet expansions, which are proven in the continuum realm, can be carried over to the digital setting.

To date, several distinct numerical implementations of the Discrete Shearlet Transform exist [21, 59, 54, 42, 56] and some additional implementations are designed to address specific applications such as, for example, edge detection [65]. Furthermore, several attempts were made to develop a multiresolution analysis similar to the one associated with wavelets, in an effort to develop MRA-based implementations [37, 42, 54]. It is useful to mention that the shearlet algorithms associated with [59] and [56] are available and downloadable from the webpage [www.ShearLab.org](http://www.ShearLab.org).

Let us next briefly describe the different approaches developed so far, by grouping these into two categories: The approaches which are Fourier-domain based and those which are spatial-domain based. All these topics will be discussed in much more detail in Chapters [72] and [70].

### 5.1 *Fourier Based Implementations*

The Cone-adapted Discrete Shearlet Transform provides a particular decomposition of the frequency plane into frequency regions associated with different scales and orientations, as illustrated in Figure 5. Hence, a very natural and direct approach to a digitalization of the Discrete Shearlet Transform is a *Fourier-based approach*, which aims to directly produce the same frequency tiling. This approach was adopted in the following two contributions.

- One such algorithmic realization was introduced in [21] as a cascade of a subband decomposition, based on the Laplacian Pyramid filter followed by a directional filtering stage which applied the Pseudo-Polar Discrete Fourier Transform.
- A different approach, which was introduced in [55, 56], consists of a carefully weighted Pseudo-Polar transform ensuring isometry followed by windowing and inverse FFT. This transform is associated with band-limited tight shearlet frames, thereby allowing the adjoint frame operator for reconstruction.

### 5.2 *Spatial Domain Based Implementations*

A *spatial domain approach* is a method where the filters associated with the transform are implemented by a convolution in the spatial domain. This approach is exploited from different viewpoints in the following four contributions.

- In [21], a numerical algorithm implementing the Discrete Shearlet Transform is introduced, in which the directional filters are approximations of the inverse

Fourier transforms of digitalized band-limited window functions in Fourier domain. With respect to the corresponding Fourier-based implementation, this approach ensures that the filters have good spatial localization.

- In contrast to the method in [21], separable window functions – thereby allowing compactly supported shearlets – are exploited in [59]. This algorithm enables the application of fast transforms ‘along both axes’, even if the corresponding transform is not associated with a tight frame.
- Yet another approach is adopted in [54], which explores the theory of subdivision schemes, leading to an associated multiresolution analysis. The main idea here is to adapt the construction of a Multiresolution analysis for wavelets, which can also be regarded as being generated by subdivision schemes. This approach comprises the possibility to obtain scaling functions ‘along the way’.
- Related to in [54], the approach developed in [42] introduces a general unitary extension principle, which – applied to the shearlet setting – determines the conditions on the filters needed for deriving a shearlet frame.

## 6 Shearlets in Applications

Shearlets were introduced to tackle a number of challenges in the representation and processing of multivariate data, and they have been successfully employed in several numerical applications. Let us briefly summarize the main areas of application below, and refer to Chapter [71] for a detailed overview.

- *Imaging Applications.* The sparsity of shearlet expansions can be exploited for various problems of data regularization and feature extraction. One class of imaging applications in which shearlets succeeded are denoising problems. For this, several shearlet-based image denoising algorithms were introduced, including those in [21, 59], which adapt wavelet thresholding to the shearlet setting, and the method in [20], which combines thresholding with minimization of bounded variation. Another class of imaging applications, for which the microlocal properties of shearlets have been found useful is the analysis and detection of edges [65].
- *Data Separation.* In several practical applications, it is important to separate data in their subcomponents. In astronomical imaging, it is very useful to separate stars from galaxies, or in neurobiological imaging, spines from dendrites. In both cases, the goal is the separation of point- and curve-like structures. Using methodologies from sparse approximation and combining wavelet and shearlet expansions, a very effective method for data separation was developed in [16, 17, 52].
- *Inverse Problems.* Shearlet-based methods have also been applied to construct a regularized inversion algorithm for the Radon transform. This transform is at the basis of computerized tomography [6]. Similar ideas were also shown to be useful when dealing with more general classes of inverse problems, such as deblurring and deconvolution [61].

## References

1. J. P. Antoine, P. Carrette, R. Murenzi, and B. Piette, *Image analysis with two-dimensional continuous wavelet transform*, Signal Process. **31** (1993), 241–272.
2. R. H. Bamberger and M. J. T. Smith, *A filter bank for the directional decomposition of images: theory and design*, IEEE Trans. Signal Process. **40** (1992), 882–893.
3. A. M. Bruckstein, D. L. Donoho, and M. Elad, *From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images*, SIAM Review **51** (2009), 34–81.
4. E. J. Candés and D. L. Donoho, *New tight frames of curvelets and optimal representations of objects with piecewise  $C^2$  singularities*, Comm. Pure and Appl. Math. **56** (2004), 216–266.
5. P. G. Casazza and G. Kutyniok, *Finite Frames: Theory and Applications*, Birkhäuser, Boston, to appear.
6. F. Colonna, G. R. Easley, K. Guo, and D. Labate, *Radon Transform Inversion using the Shearlet Representation*, Appl. Comput. Harmon. Anal., 29(2), p. 232-250 (2010).
7. O. Christensen, *An Introduction to Frames and Riesz Bases*, Birkhäuser, Boston, 2003.
8. S. Dahlke, G. Kutyniok, G. Steidl, and G. Teschke, *Shearlet coorbit spaces and associated Banach frames*, Appl. Comput. Harmon. Anal. **27** (2009), 195–214.
9. S. Dahlke, G. Kutyniok, P. Maass, C. Sagiv, H.-G. Stark, and G. Teschke, *The uncertainty principle associated with the continuous shearlet transform*, Int. J. Wavelets Multiresolut. Inf. Process. **6** (2008), 157–181.
10. S. Dahlke, G. Steidl, and G. Teschke, *The continuous shearlet transform in arbitrary space dimensions*, J. Fourier Anal. Appl. **16** (2010), 340–364.
11. S. Dahlke and G. Teschke, *The continuous shearlet transform in higher dimensions: variations of a theme*, in Group Theory: Classes, Representation and Connections, and Applications, edited by C. W. Danellis, Math. Res. Develop., Nova Publishers, 2010, 167–175.
12. I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1992.
13. M. N. Do and M. Vetterli, *The contourlet transform: an efficient directional multiresolution image representation*, IEEE Trans. Image Process. **14** (2005), 2091–2106.
14. D. L. Donoho, *Sparse components of images and optimal atomic decomposition*, Constr. Approx. **17** (2001), 353–382.
15. D. L. Donoho, *Emerging Applications of Geometric Multiscale Analysis*, Proceedings International Congress of Mathematicians **Vol. I** (2002), 209–233.
16. D. L. Donoho and G. Kutyniok, *Geometric separation using a wavelet-shearlet dictionary*, SampTA'09 (Marseille, France, 2009), Proc., 2009.
17. D. L. Donoho and G. Kutyniok, *Microlocal analysis of the geometric separation problem*, preprint.
18. D. L. Donoho, M. Vetterli, R. DeVore, and I. Daubechies, *Data Compression and Harmonic Analysis*, IEEE Trans. Info. Theory **44** (1998), 2435–2476.
19. R.J. Duffin and A.C. Schaeffer, *A class of nonharmonic Fourier series*, Trans. Amer. Math. Soc. **72** (1952), 341–366.
20. G. R. Easley, D. Labate, and F. Colonna, *Shearlet Based Total Variation for Denoising*, IEEE Trans. Image Process. **18** (2) pp. 260-268 (2009)
21. G. Easley, D. Labate, and W.-Q Lim, *Sparse directional image representations using the discrete shearlet transform*, Appl. Comput. Harmon. Anal. **25** (2008), 25–46.
22. M. Elad, *Sparse and redundant representations*, Springer, New York, 2010.
23. C. Fefferman, *A note on spherical summation multipliers*, Israel J. Math. **15** (1973), 44–52.
24. G. Folland, *Fourier Analysis and Its Applications*, American Mathematical Society, Rhode Island, 2009.
25. P. Grohs, *Continuous shearlet frames and resolution of the wavefront set*, Monatsh. Math., to appear.
26. P. Grohs, *Continuous shearlet tight frames*, J. Fourier Anal. Appl., to appear.
27. A. Grossmann, J. Morlet, and T. Paul, *Transforms associated to square integrable group representations I: General Results*, J. Math. Phys. **26** (1985), 2473–2479.

28. K. Guo, G. Kutyniok, and D. Labate, *Sparse multidimensional representations using anisotropic dilation and shear operators*, in Wavelets and Splines (Athens, GA, 2005), Nashboro Press, Nashville, TN, 2006, 189–201.
29. K. Guo and D. Labate, *Optimally sparse multidimensional representation using shearlets*, SIAM J. Math. Anal. **39** (2007), 298–318.
30. K. Guo, and D. Labate, *Characterization and analysis of edges using the Continuous Shearlet Transform*, SIAM on Imaging Sciences 2, p. 959-986 (2009).
31. K. Guo and D. Labate, *Analysis and detection of surface discontinuities using the 3D continuous shearlet transform*, Appl. Comput. Harmon. Anal. **30** (2011) 231-242.
32. K. Guo, and D. Labate, *Optimally Sparse 3D Approximations using Shearlet Representations*, Electronic Research Announcements in Mathematical Sciences, 17 p. 126-138 (2010).
33. K. Guo, and D. Labate, *Optimally Sparse Representations of 3D Data with  $C^2$  Surface Singularities using Parseval Frames of Shearlets* preprint (2011).
34. K. Guo, D. Labate and W. Lim, *Edge Analysis and identification using the Continuous Shearlet Transform*, Appl. Comput. Harmon. Anal. **27** (2009), 24–46.
35. K. Guo, D. Labate, W. Lim, G. Weiss, and E. Wilson, *Wavelets with composite dilations*, Electron. Res. Announc. Amer. Math. Soc. **10** (2004), 78–87.
36. K. Guo, D. Labate, W.-Q Lim, G. Weiss, and E. Wilson, *The theory of wavelets with composite dilations*, Harmonic analysis and applications, Appl. Numer. Harmon. Anal., Birkhäuser Boston, Boston, MA, 2006, 231–250.
37. K. Guo, W.-Q Lim, D. Labate, G. Weiss, and E. Wilson, *Wavelets with composite dilations and their MRA properties*, Appl. Comput. Harmon. Anal. **20** (2006), 220–236.
38. E. Hewitt and K.A. Ross, *Abstract Harmonic Analysis I, II*, Springer-Verlag, Berlin/ Heidelberg/New York, 1963.
39. M. Holschneider, *Wavelets. Analysis tool*, Oxford University Press, Oxford, 1995.
40. N. Kingsbury, *Image processing with complex wavelets*, Phil. Trans. Royal Society London A, **357** (1999), 2543–2560.
41. N. Kingsbury, *Complex wavelets for shift invariant analysis and filtering of signals*, Appl. Computat. Harmon. Anal. **10** (2001), 234–253.
42. B. Han, G. Kutyniok, and Z. Shen. *A unitary extension principle for shearlet systems*, preprint.
43. E. Hernandez and G. Weiss, *A First Course on Wavelets*, CRC, Boca Raton, FL, 1996
44. R. Houska, *The nonexistence of shearlet scaling functions*, to appear in Appl. Comput Harmon. Anal. (2011)
45. P. Kittipoom, G. Kutyniok, and W.-Q Lim, *Construction of compactly supported shearlet frames*, preprint.
46. P. Kittipoom, G. Kutyniok, and W.-Q Lim, *Irregular shearlet frames: Geometry and approximation properties*, J. Fourier Anal. Appl., to appear.
47. G. Kutyniok and D. Labate, *Construction of regular and irregular shearlets*, J. Wavelet Theory and Appl. **1** (2007), 1–10.
48. G. Kutyniok and D. Labate, *Resolution of the wavefront set using continuous shearlets*, Trans. Amer. Math. Soc. **361** (2009), 2719–2754.
49. G. Kutyniok, J. Lemvig, and W.-Q Lim, *Compactly supported shearlets*, preprint.
50. G. Kutyniok, J. Lemvig, and W.-Q Lim, *Compactly supported shearlet frames and optimally sparse approximations of functions in  $L^2(\mathbb{R}^3)$  with piecewise  $C^2$  singularities*, preprint.
51. G. Kutyniok and W.-Q Lim, *Compactly supported shearlets are optimally sparse*, J. Approx. Theory, to appear.
52. G. Kutyniok and W.-Q Lim, *Image separation using shearlets and wavelets*, preprint.
53. G. Kutyniok and W.-Q Lim, *Shearlets on bounded domains*, in Approximation Theory XIII (San Antonio, TX, 2010), Springer, to appear.
54. G. Kutyniok and T. Sauer, *Adaptive directional subdivision schemes and shearlet multiresolution analysis*, SIAM J. Math. Anal. **41** (2009), 1436–1471.
55. G. Kutyniok, M. Shahram, and D. L. Donoho, *Development of a digital shearlet transform based on pseudo-polar FFT*, in Wavelets XIII, edited by V. K. Goyal, M. Papadakis, D. Van De Ville, SPIE Proc. **7446** (2008), SPIE, Bellingham, WA, 2009, 7446-12.

56. G. Kutyniok, M. Shahram, and X. Zhuang, *ShearLab: A Rational Design of a Digital Parabolic Scaling Algorithm*, preprint.
57. D. Labate, W.-Q Lim, G. Kutyniok, and G. Weiss. *Sparse multidimensional representation using shearlets*, in Wavelets XI, edited by M. Papadakis, A. F. Laine, and M. A. Unser, SPIE Proc. **5914** (2005), SPIE, Bellingham, WA, 2005, 254–262.
58. Laugesen, R. S., N. Weaver, G. Weiss, and E. Wilson, *A characterization of the higher dimensional groups associated with continuous wavelets*, J. Geom. Anal. **12** (2001), 89–102.
59. W.-Q Lim, *The discrete shearlet transform: A new directional transform and compactly supported shearlet frames*, IEEE Trans. Image Process. **19** (2010), 1166–1180.
60. Mallat, S., *A Wavelet Tour of Signal Processing*, Academic Press, San Diego 1998.
61. V.M. Patel, G. Easley, D.M. Healy, *Shearlet-Based Deconvolution* IEEE Trans. Image Process. **18**(12) (2009), 2673-2685
62. E. P. Simoncelli, W. T. Freeman, E. H. Adelson, D. J. Heeger, *Shiftable multiscale transforms*, IEEE Trans. Inform. Theory **38** (1992), 587–607.
63. H. F. Smith, *A Hardy space for Fourier integral operators*, J. Geom. Anal. **8**, 629–653.
64. Stein, E., *Harmonic Analysis: Real-Variable Methods, Orthogonality and Oscillatory Integrals*, Princeton University Press, Princeton, 1993.
65. S. Yi, D. Labate, G. R. Easley, and H. Krim, *A Shearlet Approach to Edge Analysis and Detection*, IEEE Trans. Image Process. **18** (2009), 929–941.
66. Chapter Microlocal
67. Chapter Coorbit Spaces
68. Chapter Edge Detection
69. Chapter Optimal Sparsity
70. Chapter Shearlet MRA
71. Chapter Image Processing
72. Chapter Digital Shearlet Transform