Applied and Numerical Harmonic Analysis

$$\widehat{f}(\gamma) = \int f(x) e^{-2\pi i x \gamma} dx$$

Gitta Kutyniok
Demetrio Labate
Editors

Shearlets

Multiscale Analysis for Multivariate Data

Birkhäuser

Birkhäuser

Applied and Numerical Harmonic Analysis

Series Editor
John J. Benedetto
University of Maryland
College Park, MD, USA

Editorial Advisory Board

Akram Aldroubi
Vanderbilt University
Nashville, TN, USA

Andrea Bertozzi
University of California
Los Angeles, CA, USA

Douglas Cochran
Arizona State University
Phoenix, AZ, USA

Hans G. Feichtinger
University of Vienna
Vienna, Austria

Christopher Heil
Georgia Institute of Technology
Atlanta, GA, USA

Stéphane Jaffard
University of Paris XII
Paris, France

Jelena Kovačević
Carnegie Mellon University
Pittsburgh, PA, USA

Gitta Kutyniok
Technische Universität Berlin
Berlin, Germany

Mauro Maggioni
Duke University
Durham, NC, USA

Zuowei Shen
National University of Singapore
Singapore, Singapore

Thomas Strohmer
University of California
Davis, CA, USA

Yang Wang
Michigan State University
East Lansing, MI, USA

For further volumes:
http://www.springer.com/series/4968

Gitta Kutyniok • Demetrio Labate
Editors

Shearlets

Multiscale Analysis for Multivariate Data

Birkhäuser

Editors
Gitta Kutyniok
Institut für Mathematik
Technische Universität Berlin
Berlin, Germany

Demetrio Labate
Department of Mathematics
University of Houston
Houston, TX, USA

ISBN 978-0-8176-8315-3 e-ISBN 978-0-8176-8316-0
DOI 10.1007/978-0-8176-8316-0
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2012932597

Mathematical Subject Classification (2010): 42C15, 42C40, 65T60, 68U10, 94A08

© Springer Science+Business Media, LLC 2012
All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.
The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed on acid-free paper

Springer is part of Springer Science+Business Media
(www.birkhauser-science.com)

ANHA Series Preface

The *Applied and Numerical Harmonic Analysis (ANHA)* book series aims to provide the engineering, mathematical, and scientific communities with significant developments in harmonic analysis, ranging from abstract harmonic analysis to basic applications. The title of the series reflects the importance of applications and numerical implementation, but richness and relevance of applications and implementation depend fundamentally on the structure and depth of theoretical underpinnings. Thus, from our point of view, the interleaving of theory and applications and their creative symbiotic evolution is axiomatic.

Harmonic analysis is a wellspring of ideas and applicability that has flourished, developed, and deepened over time within many disciplines and by means of creative cross-fertilization with diverse areas. The intricate and fundamental relationship between harmonic analysis and fields such as signal processing, partial differential equations (PDEs), and image processing is reflected in our state-of-the-art *ANHA* series.

Our vision of modern harmonic analysis includes mathematical areas such as wavelet theory, Banach algebras, classical Fourier analysis, time–frequency analysis, and fractal geometry, as well as the diverse topics that impinge on them.

For example, wavelet theory can be considered an appropriate tool to deal with some basic problems in digital signal processing, speech and image processing, geophysics, pattern recognition, biomedical engineering, and turbulence. These areas implement the latest technology from sampling methods on surfaces to fast algorithms and computer vision methods. The underlying mathematics of wavelet theory depends not only on classical Fourier analysis, but also on ideas from abstract harmonic analysis, including von Neumann algebras and the affine group. This leads to a study of the Heisenberg group and its relationship to Gabor systems, and of the metaplectic group for a meaningful interaction of signal decomposition methods. The unifying influence of wavelet theory in the aforementioned topics illustrates the justification for providing a means for centralizing and disseminating information from the broader, but still focused, area of harmonic analysis. This will be a key role of *ANHA*. We intend to publish the scope and interaction that such a host of issues demands.

Along with our commitment to publish mathematically significant works at the frontiers of harmonic analysis, we have a comparably strong commitment to publish major advances in the following applicable topics in which harmonic analysis plays a substantial role:

Antenna theory	*Prediction theory*
Biomedical signal processing	*Radar applications*
Digital signal processing	*Sampling theory*
Fast algorithms	*Spectral estimation*
Gabor theory and applications	*Speech processing*
Image processing	*Time–frequency and*
Numerical partial differential equations	*time-scale analysis*
	Wavelet theory

The above point of view for the *ANHA* book series is inspired by the history of Fourier analysis itself, whose tentacles reach into so many fields.

In the last two centuries, Fourier analysis has had a major impact on the development of mathematics, on the understanding of many engineering and scientific phenomena, and on the solution of some of the most important problems in mathematics and the sciences. Historically, Fourier series were developed in the analysis of some of the classical PDEs of mathematical physics; these series were used to solve such equations. In order to understand Fourier series and the kinds of solutions they could represent, some of the most basic notions of analysis were defined, e.g., the concept of "function". Since the coefficients of Fourier series are integrals, it is no surprise that Riemann integrals were conceived to deal with uniqueness properties of trigonometric series. Cantors set theory was also developed because of such uniqueness questions.

A basic problem in Fourier analysis is to show how complicated phenomena, such as sound waves, can be described in terms of elementary harmonics. There are two aspects of this problem: first, to find, or even define properly, the harmonics or spectrum of a given phenomenon, e.g., the spectroscopy problem in optics; second, to determine which phenomena can be constructed from given classes of harmonics, as done, e.g., by the mechanical synthesizers in tidal analysis.

Fourier analysis is also the natural setting for many other problems in engineering, mathematics, and the sciences. For example, Wiener's Tauberian theorem in Fourier analysis not only characterizes the behavior of the prime numbers, but also provides the proper notion of spectrum for phenomena such as white light; this latter process leads to the Fourier analysis associated with correlation functions in filtering and prediction problems, and these problems, in turn, deal naturally with Hardy spaces in the theory of complex variables.

Nowadays, some of the theory of PDEs has given way to the study of Fourier integral operators. Problems in antenna theory are studied in terms of unimodular trigonometric polynomials. Applications of Fourier analysis abound in signal processing, whether with the fast Fourier transform (FFT), or filter design, or the adaptive modeling inherent in timefrequency-scale methods such as wavelet theory. The coherent states of mathematical physics are translated and modulated Fourier transforms, and these are used, in conjunctionwith the uncertainty principle, for dealing with signal reconstruction in communications theory.We are back to the raison d'être of the *ANHA* series!

University of Maryland
College Park

John J. Benedetto
Series Editor

Preface

The introduction of wavelets about 20 years ago has revolutionized applied mathematics, computer science, and engineering by providing a highly effective methodology for analyzing and processing univariate functions/signals containing singularities. However, wavelets do not perform equally well in the multivariate case due to the fact that they are capable of efficiently encoding only isotropic features. This limitation can be seen by observing that Besov spaces can be precisely characterized by decay properties of sequences of wavelet coefficients, but they are not capable of capturing those geometric features which could be associated with edges and other distributed singularities. Indeed, such geometric features are essential in the multivariate setting, since multivariate problems are typically governed by anisotropic phenomena such as singularities concentrated on lower dimensional embedded manifolds. To deal with this challenge, several approaches were proposed in the attempt to extend the benefits of the wavelet framework to higher dimensions, with the aim of introducing representation systems which could provide both optimally sparse approximations of anisotropic features and a unified treatment of the continuum and digital world. Among the various methodologies proposed, such as curvelets and contourlets, the shearlet system, which was introduced in 2005, stands out as the first and so far the only approach capable of satisfying this combination of requirements.

Today, various directions of research have been established in the theory of shearlets. These include, in particular, the theory of continuous shearlets—associated with a parameter set of continuous range—and its application to the analysis of distributions. Another direction is the theory of discrete shearlets—associated with a discrete parameter set—and their sparse approximation properties. Thanks to the fact that shearlets provide a unified treatment of the continuum and digital realm through the utilization of the shearing operator, digitalization and hence numerical realizations can be performed in a faithful manner, and this leads to very efficient algorithms. Building on these results, several shearlet-based algorithms were developed to address a range of problems in image and data processing.

This book is the first monograph devoted to shearlets. It is not only aimed at and accessible to a broad readership including graduate students and researchers in the areas of applied mathematics, computer science, and engineering, but it will also appeal to researchers working in any other field requiring highly efficient methodologies for the processing of multivariate data. Because of this fact, this volume can be used both as a state-of-the-art monograph on shearlets and advanced multiscale methods and as a textbook for graduate students.

This volume is organized into several tutorial-like chapters which cover the main aspects of theory and applications of shearlets and are written by the leading international experts in these areas. The first chapter provides a self-contained and comprehensive overview of the main results on shearlets and sets the basic notation and definitions which are used in the remainder of the book. The topics covered in the remaining chapters essentially follow the idea of going from the continuous setting, i.e., continuous shearlets and their microlocal properties, up to the discrete and digital setting, i.e., discrete shearlets, their digital realizations, and their applications. Each chapter is self-contained, which enables the reader to choose his/her own path through the book. Here is a brief outline of the content of each chapter.

The first chapter, written by the editors, provides an introduction and presents a self-contained overview of the main results on the theory and applications of shearlets. Starting with some background on frame theory and wavelets, it covers the definitions of continuous and discrete shearlets and the main results from the theory of shearlets, which are subsequently discussed in detail and expanded in the following chapters.

In the second chapter, Grohs focusses on the continuous shearlet transform. After making the reader familiar with concepts from microlocal analysis, he shows that the shearlet transform offers a simple and convenient way to characterize wavefront sets of distributions.

In the third chapter, Guo and Labate illustrate the ability of the continuous shearlet transform to characterize the set of singularities of multivariate functions and distributions. These properties set the groundwork for some of the imaging applications discussed in the eighth chapter.

In the fourth chapter, Dahlke et al. introduce the continuous shearlet transform for arbitrary space dimension. They further present the construction of smoothness spaces associated to shearlet representations and the analysis of their structural properties.

In the fifth chapter, Kutyniok et al. provide a comprehensive survey of the theory of sparse approximations of cartoon-like images using shearlets. Both the band-limited and the compactly supported shearlet frames are examined in this chapter.

In the sixth chapter, Sauer starts from the classical concepts of filterbanks and subband coding to present an entirely digital approach to shearlet multiresolution. This approach is not a discretization of the continuous transform, but is naturally connected to the filtering of digital data.

In the seventh chapter, Kutyniok et al. discuss the construction of digital realizations of the shearlet transform with a particular focus on a unified treatment of the continuum and digital realm. In particular, this chapter illustrates two distinct numerical implementations of the shearlet transform, one based on band-limited shearlets and the other based on compactly supported shearlets.

In the eighth chapter, Easley and Labate present the application of shearlets to several problems from imaging and data analysis to date. This includes the illustration of shearlet-based algorithms for image denoising, image enhancement, edge detection, image separation, deconvolution, and regularized reconstruction of Radon data. In all these applications, the ability of shearlet representations to handle anisotropic features efficiently is exploited in order to derive highly competitive numerical algorithms.

Finally, it is important to emphasize that the work presented in this volume would not have been possible without the interaction and discussions with many people during these years. We wish to thank the many students and researchers who over the years have given us insightful comments and suggestions, and helped this area of research to grow into its present form.

Berlin, Germany Gitta Kutyniok
Houston, USA Demetrio Labate

Contents

Introduction to Shearlets 1
Gitta Kutyniok and Demetrio Labate
1 Introduction 1
2 The Rise of Shearlets 3
2.1 The Role of Applied Harmonic Analysis 3
2.2 Wavelets and Beyond 4
3 Notation and Background Material 6
3.1 Fourier Analysis 6
3.2 Modeling of Signal Classes 7
3.3 Frame Theory 9
3.4 Wavelets 10
3.5 Wavelets for Multivariate Data and Their Limitations . . . 13
4 Continuous Shearlet Systems 15
4.1 Continuous Shearlet Systems and the Shearlet Group 17
4.2 The Continuous Shearlet Transform 17
4.3 Cone-Adapted Continuous Shearlet Systems 20
4.4 The Cone-Adapted Continuous Shearlet Transform 21
4.5 Microlocal Properties and Characterization of Singularities 22
5 Discrete Shearlet Systems 23
5.1 Discrete Shearlet Systems and Transforms 24
5.2 Cone-Adapted Discrete Shearlet Systems and Transforms 25
5.3 Compactly Supported Shearlets 28
5.4 Sparse Approximations by Shearlets 30
5.5 Shearlet Function Spaces 32
5.6 Extensions and Generalizations 32
6 Algorithmic Implementations of the Shearlet Transform 33
6.1 Fourier-Based Implementations 34
6.2 Spatial-Domain-Based Implementations 34
7 Shearlets in Applications 35
References 36

Shearlets and Microlocal Analysis 39
Philipp Grohs
1 Introduction 39
1.1 Notation 40
1.2 Getting to Know the Wavefront Set 41
1.3 Contributions 49
1.4 Other Ways to Characterize the Wavefront Set 49
2 Reproduction Formulas 50
3 Resolution of the Wavefront Set 55
3.1 A Direct Theorem 55
3.2 Properties of the Wavefront Set 59
3.3 Proof of the Main Result 61
References 67

Analysis and Identification of Multidimensional Singularities Using the Continuous Shearlet Transform 69
Kanghui Guo and Demetrio Labate
1 Introduction 69
1.1 Example: Line Singularity 70
1.2 General Singularities 75
2 Analysis of Step Singularities (2D) 76
2.1 Shearlet Analysis of Circular Edges 78
2.2 General 2D Boundaries 81
2.3 Proofs of Theorems 2 and 3 83
2.4 Extensions and Generalizations 97
3 Extension to Higher Dimensions 98
3.1 3D Continuous Shearlet Transform 99
3.2 Characterization of 3D Boundaries 100
References 103

Multivariate Shearlet Transform, Shearlet Coorbit Spaces and Their Structural Properties 105
Stephan Dahlke, Gabriele Steidl, and Gerd Teschke
1 Introduction 106
2 Multivariate Continuous Shearlet Transform 107
2.1 Unitary Representations of the Shearlet Group 107
2.2 Square Integrable Representations of the Shearlet Group . 110
2.3 Continuous Shearlet Transform 113
3 General Concept of Coorbit Space Theory 113
3.1 General Coorbit Spaces 115
3.2 Atomic Decompositions and Banach Frames 116
4 Multivariate Shearlet Coorbit Theory 117
4.1 Shearlet Coorbit Spaces 117
4.2 Shearlet Atomic Decompositions and Shearlet Banach Frames 118
4.3 Nonlinear Approximation 119

5 Structure of Shearlet Coorbit Spaces 121
5.1 Atomic Decomposition of Besov Spaces 125
5.2 A Density Result 126
5.3 Traces on the Real Axes 127
5.4 Embedding Results 131
6 Analysis of Singularities 134
6.1 Hyperplane Singularities 134
6.2 Tetrahedron Singularities 137
References 142

Shearlets and Optimally Sparse Approximations 145
Gitta Kutyniok, Jakob Lemvig, and Wang-Q Lim
1 Introduction 146
1.1 Choice of Model for Anisotropic Features 146
1.2 Measure for Sparse Approximation and Optimality 147
1.3 Why is 3D the Crucial Dimension? 147
1.4 Performance of Shearlets and Other Directional Systems . 148
1.5 Band-Limited Versus Compactly Supported Systems 148
1.6 Outline 149
2 Cartoon-Like Image Class 149
3 Sparse Approximations 151
3.1 (Nonlinear) N-term Approximations 151
3.2 A Notion of Optimality 155
3.3 Approximation by Fourier Series and Wavelets 158
4 Pyramid-Adapted Shearlet Systems 161
4.1 General Definition 162
4.2 Band-Limited 3D Shearlets 164
4.3 Compactly Supported 3D Shearlets 166
4.4 Some Remarks on Construction Issues 169
5 Optimal Sparse Approximations 170
5.1 Optimal Sparse Approximations in 2D 170
5.2 Optimal Sparse Approximations in 3D 191
References 196

Shearlet Multiresolution and Multiple Refinement 199
Tomas Sauer
1 Introduction 199
2 Filters and Filterbanks 201
2.1 Filterbanks 201
2.2 Symbols and Transforms 204
2.3 Filterbanks by Matrix Completion 208
2.4 Subbands and Multiresolution 210
3 Subdivision and Refinability 212
3.1 Convergence and Basic Properties 212
3.2 Interpolatory Subdivision and Filterbanks 214
3.3 Multiresolution 216

4 Multiple Subdivision and Multiple Refinability ... 218
4.1 Basic Properties ... 219
4.2 The Multiple MRA ... 221
4.3 Filterbanks, Cascades, Trees ... 223
4.4 Things Work Along Trees ... 226
4.5 A Canonical Interpolatory Construction ... 227
5 Shearlet Subdivision and Multiresolution ... 229
5.1 Shears and Scaling ... 230
5.2 Shears of Codimension 1: Hyperplane Shearlets ... 231
5.3 Orthogonal Shearlets by Tensor Product ... 233
5.4 Implementation ... 234
References ... 236

Digital Shearlet Transforms ... 239
Gitta Kutyniok, Wang-Q Lim, and Xiaosheng Zhuang
1 Introduction ... 240
1.1 A Unified Framework for the Continuum and Digital World ... 240
1.2 Band-Limited vs. Compactly Supported Shearlet Transforms ... 241
1.3 Related Work ... 242
1.4 Framework for Quantifying Performance ... 242
1.5 ShearLab ... 243
1.6 Outline ... 243
2 Digital Shearlet Transform Using Band-Limited Shearlets ... 244
2.1 Pseudo-Polar Fourier Transform ... 245
2.2 Density-Compensation Weights ... 248
2.3 Digital Shearlets on Pseudo-Polar Grid ... 251
2.4 Algorithmic Realization of the FDST ... 259
3 Digital Shearlet Transform Using Compactly Supported Shearlets ... 261
3.1 Digital Separable Shearlet Transform ... 262
3.2 Digital Non-separable Shearlet Transform ... 270
4 Framework for Quantifying Performance ... 272
4.1 Algebraic Exactness ... 273
4.2 Isometry of Pseudo-Polar Transform ... 273
4.3 Parseval Frame Property ... 274
4.4 Space-Frequency-Localization ... 275
4.5 Shear Invariance ... 277
4.6 Speed ... 278
4.7 Geometric Exactness ... 279
4.8 Stability ... 280
References ... 282

Image Processing Using Shearlets 283
Glenn R. Easley and Demetrio Labate
1 Introduction 283
2 Image Denoising 284
2.1 Discrete Shearlet Transform 286
2.2 Shearlet Thresholding 289
2.3 Denoising Using Shearlet-Based Total Variation Regularization 292
2.4 Complex-Valued Denoising 294
2.5 Other Shearlet-Based Denoising Techniques 295
3 Inverse Problems 296
3.1 Inverting the Radon Transform 296
3.2 Deconvolution 299
3.3 Inverse-Halftoning 303
4 Image Enhancement 305
5 Edge Analysis and Detection 309
5.1 Edge Analysis Using Shearlets 310
5.2 Edge Detection Using Shearlets 311
5.3 Edge Analysis Using Shearlets 313
6 Image Separation 315
6.1 Image Model 315
6.2 Geometric Separation Algorithm 316
7 Shearlets Analysis of 3D Data 318
8 Additional Applications 320
References 321

Index 327

Contributors

Stephan Dahlke Philipps-Universität Marburg, FB12 Mathematik und Informatik, Hans-Meerwein Straße, Marburg, Germany

Glenn R. Easley System Planning Corporation, Arlington, VA, USA

Philipp Grohs ETH Zürich, Zürich, Switzerland

Kanghui Guo Department of Mathematics, Missouri State University, Springfield, MO, USA

Gitta Kutyniok Institut für Mathematik, Technische Universität Berlin, Berlin, Germany

Demetrio Labate Department of Mathematics, University of Houston, Houston, TX, USA

Jakob Lemvig Department of Mathematics, Technical University of Denmark, Lyngby, Denmark

Wang-Q Lim Institut für Mathematik, Technische Universität Berlin, Berlin, Germany

Tomas Sauer Lehrstuhl für Numerische Mathematik, Justus–Liebig–Universität Gießen, Gießen, Germany

Gabriele Steidl Universität Kaiserslautern, Fachbereich Mathematik, Kaiserslautern, Germany

Gerd Teschke Hochschule Neubrandenburg - University of Applied Sciences, Institute for Computational Mathematics in Science and Technology, Neubrandenburg, Germany

Xiaosheng Zhuang Institut für Mathematik, Technische Universität Berlin, Berlin, Germany

Introduction to Shearlets

Gitta Kutyniok and Demetrio Labate

Abstract Shearlets emerged in recent years among the most successful frameworks for the efficient representation of multidimensional data. Indeed, after it was recognized that traditional multiscale methods are not very efficient at capturing edges and other anisotropic features which frequently dominate multidimensional phenomena, several methods were introduced to overcome their limitations. The shearlet representation stands out since it offers a unique combination of some highly desirable properties: it has a single or finite set of generating functions, it provides optimally sparse representations for a large class of multidimensional data, it is possible to use compactly supported analyzing functions, it has fast algorithmic implementations and it allows a unified treatment of the continuum and digital realms. In this chapter, we present a self-contained overview of the main results concerning the theory and applications of shearlets.

Key words: Affine systems, Continuous wavelet transform, Image processing, Shearlets, Sparsity, Wavelets

1 Introduction

Scientists sometimes refer to the twenty-first century as the Age of Data. As a matter of fact, since technological advances make the acquisition of data easier and less expensive, we are coping today with a deluge of data including astronomical, medical, seismic, meteorological, and surveillance data, which require efficient analysis

G. Kutyniok
Institut für Mathematik, Technische Universität Berlin, 10623 Berlin, Germany
e-mail: kutyniok@math.tu-berlin.de

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

G. Kutyniok and D. Labate (eds.), *Shearlets: Multiscale Analysis for Multivariate Data*,
Applied and Numerical Harmonic Analysis, DOI 10.1007/978-0-8176-8316-0_1,
© Springer Science+Business Media, LLC 2012

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 their geometric properties. This is closely related to the observation that virtually all multivariate data are typically dominated by anisotropic features such as singularities on lower dimensional embedded manifolds. 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, the theory of *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* in which 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 1990s, 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" [16].

This broader area of research is currently at the crossroads 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 the last 5 years, this book is also a journey into one of the most active and exciting areas of research in applied mathematics.

2 The Rise of Shearlets

2.1 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" (this is the literal meaning of the Greek word *analysis*) to gain insight into an object. For example, given a class of data $\mathscr{C}$ in $L^2(\mathbb{R}^d)$, 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 $f \in \mathscr{C}$, we have the expansion

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

This formula provides not only a decomposition for any element $f \in \mathscr{C}$ into a countable collection of linear measurements $(c_i(f))_{i\in I} \subseteq \ell^2(I)$, i.e., 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. *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:

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

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:

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

Given a prescribed class of data $\mathscr{C}$, one major objective is to design an analyzing system $(\varphi_i)_{i\in I}$ in such a way that, for each function $f \in \mathscr{C}$, 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 $f \in \mathscr{C}$ with high accuracy by using a coefficient sequence $(\tilde{c}_i(f))_{i\in I}$ containing very few nonzero entries. In the finite-dimensional setting, such a sequence is called *sparse*, and this explains the use of the term *sparse approximation*. 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 [19].

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 Sect. 3.3)—the sequences $(c_i(f))_{i\in I}$ might be chosen significantly sparser for each $f \in \mathscr{C}$. Thus, methodologies from *frame theory* will come into play, see Sect. 3.3 and [5, 7].

We can observe a close connection to yet another highly topical area. During the last 4 years, sparse recovery methodologies such as *Compressed Sensing* in particular have revolutionized the areas of applied mathematics, computer science, and electrical engineering by beating the traditional sampling theory limits, see [3, 23]. They exploit the fact that many types of signals can be represented using only a few nonvanishing coefficients when choosing a suitable basis or, more generally, a frame. Nonlinear optimization methods, such as ℓ_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.

2.2 Wavelets and Beyond

The emergence of *wavelets* about 25 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 [65] 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 introduce some form of directional sensitivity[1] in the wavelet framework 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. [71], 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* [44, 45]. 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, its construction involves rotations and these operators do 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 [14] as a purely discrete filter-bank version of the curvelet framework. This approach offers the advantage of

[1] It is important to recall that the importance of directional sensitivity in the efficient processing of natural images by the human brain has been a major finding in neuropsycological studies such as the work of Field and Olshausen [68], and a significant inspiration for some of the research developed in the harmonic analysis and image processing literature.

allowing a tree-structured filter bank implementation similar to the standard wavelet implementations which was exploited to obtain very efficient numerical algorithms. However, a proper continuum theory is missing in this approach.

In the same year, *shearlets* were introduced by Guo, Kutyniok, Labate, Lim, and Weiss in [30, 61]. This approach was derived within a larger class of affine-like systems—the so-called *composite wavelets* [39, 40, 41]—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 shearlet systems to be derived from a single or finite set of generators, 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 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.

For completeness, it is important to recall yet another class of representation systems which are able to overcome the limitations of traditional wavelets and produce optimally efficient representations for a large class of images, namely the *bandelets* [70] and the *grouplets* [66]. Also in these methods, the idea is to take advantage of the geometry of the data. However, in this case, this is done *adaptively*, that is, by constructing a special data decomposition which is especially designed for each data set, rather than by using a fixed representation system as it is done using wavelets or shearlets. While one can achieve very efficient data decompositions using such an adaptive approach, this is usually numerically more intensive than using nonadaptive methods.

In the following sections, 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.

3 Notation and Background Material

3.1 Fourier Analysis

The Fourier transform is the most fundamental tool in harmonic analysis. Before stating the definition, we remark that, in the following, vectors in $\mathbb{R}^d$ or $\mathscr{C}^d$ 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

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

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

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

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 [25] for additional background information on Fourier analysis.

3.2 *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 imposing 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* $\mathscr{D}'(\mathbb{R}^d)$ 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 sing supp(u). Notice that the singular support of u is a closed subset of supp(u).

The anisotropic nature of singularities on one- or multidimensional embedded manifolds becomes apparent through the notion of a wavefront set. For simplicity, we illustrate the two-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

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

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.

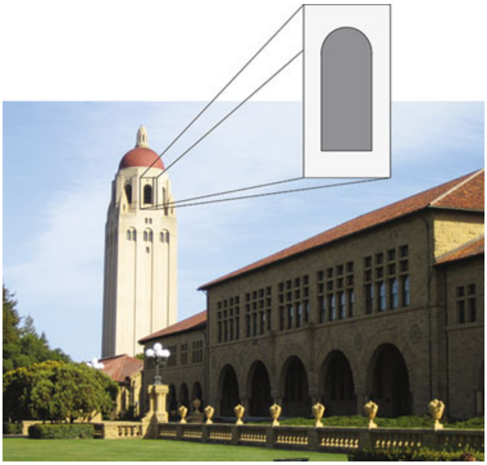

Fig. 1 Natural images are governed by anisotropic structures

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 [15] 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 Fig. 1. Since the image basically 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 Fig. 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.

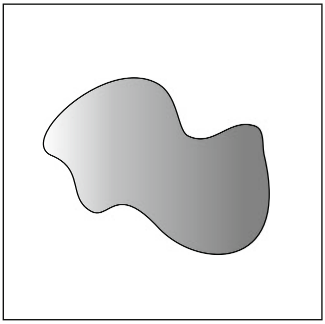

Fig. 2 Example of a cartoon-like image (function values represented using a gray scale map)

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

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

where $B \subset [0,1]^2$ is a set with ∂B being a closed C^2-curve with bounded curvature and $f_i \in C^2(\mathbb{R}^2)$ are functions with $\operatorname{supp} f_i \subset [0,1]^2$ 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 $\{0,\ldots,N-1\}^d$, sometimes denoted by $\mathbb{Z}_N^d$.

3.3 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 [20] and later revived by Daubechies in [13], guarantees stability while allowing nonunique 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 supremun 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 with $A = B$, 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 $c = 1$.