Diss ETH No 23719

Continuous Parabolic Molecules

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH Zürich
(Dr. sc. ETH Zürich)

presented by

Željko Kereta

MSc in Mathematical Modelling and Scientific Computing
University of Oxford

born on 22 May 1987
citizen of Croatia

accepted on the recommendation of
Prof Dr Philipp Grohs, ETH Zürich, examiner
Prof Dr Helmut Bölcskei, ETH Zürich, co-examiner
Prof Dr Gitta Kutyniok, TU Berlin, co-examiner

2016

Abstract

Last decade saw the addition of a number of new representation dictionaries to the arsenal of computational harmonic analysis. A major source of inspiration that underlies their creation is the desire to address the weaknesses of the classical wavelet transform that arise due to its limited capacity for the analysis of edge like features of two-dimensional signals. The first dictionary fully adapted for addressing these tasks were the curvelets [1], followed closely by shearlets [2] and contourlets [3], to name but a few. Salient feature of these dictionaries is their adherence to a paradigm that aims to incorporate the directional selectivity and the anisotropic treatment of the axes, achieved through the parabolic scaling law, into the architecture of the dictionary.

Directional anisotropic dictionaries have subsequently garnered attention for their ability to answer questions regarding curvilinear singularities of functions of both theoretical and practical importance such as Fourier integral operators [4, 5], denoising [6] and inpainting [7]. Inferring about their properties each such dictionary required proofs tailored to the specifics of its construction, yet all are somehow repetitive. Nevertheless, there has been a sizeable body of evidence to suggest that these dictionaries answer the same fundamental questions and it should be possible to see them as parts of a much broader framework. Such a framework could help reduce the redundancy of proof techniques by establishing the universal features, and allow the usage of a dictionary that best fits any given situation, since some dictionaries are better for theoretical considerations whereas other dictionaries might be better suited for digital implementations.

In this thesis we will construct a comprehensive framework for directional parabolic dictionaries in the continuous setting, which we coin the continuous parabolic molecules. Our work builds on earlier studies such as the curvelet molecules [5] and the shearlet molecules [8], but the predominant influence is that of the discrete parabolic molecules [9]. In contrast with earlier studies, we will look at the questions and challenges that appear in the continuous setting. We will first show that our framework is general enough to include both the curvelet- and shearlet-type constructions, and that the pairs of molecules are almost-orthogonal with respect to a pseudo-distance function which is adapted to the ambient geometry. Using almost-orthogonality we will show that some fundamental notions of microlocal analysis such as the resolution of the wavefront set and microlocal Sobolev regularity can be detected by all suitable continuous parabolic molecules. Lastly, we will show that parabolic molecules can be used to identify edge and corner points of piecewise smooth closed curves in $\mathbb{R}^2$.

Prefazione

Il decennio scorso ha visto l'aggiunta all'arsenale dell'analisi armonica computazionale di nuovi dizionari di rappresentazione. La loro creazione trae ispirazione dalla volontà di colmare le carenze della classica trasformata *wavelet* dovute alla limitata capacità della suddetta nell'analisi di caratteristiche, quali spigoli, in segnali in due dimensioni. Il primo dizionario pienamente pensato a questo scopo furono le *curvelets*, seguite poco dopo dalle *shearlets* e dalle *contourlets*, per citarne alcune. Caratteristica saliente dei predetti dizionari è l'adozione di un paradigma rivolto a incorporare nell'architettura del dizionario la selettività direzionale e il trattamento anisotropo degli assi, quest'ultimo tramite la legge di *scaling* parabolico.

I dizionari parabolici anisotropi hanno successivamente attratto interesse per la loro abilità nel risolvere questioni, sia teoriche che pratiche, riguardanti singolarità curvilinee che si presentano, per esempio, nello studio degli operatori integrali di Fourier, nel *denoising* e nell' *inpainting*. Le argomentazioni matematiche necessarie allo studio delle proprietà di tali dizionari richiesero un adattamento alle specificità di ognuno di essi, risultando al contempo ripetitive. Ciò nondimeno, una considerevole evidenza matematica suggerisce che questi dizionari forniscono risposte alle medesime questioni fondamentali, per cui sembrerebbe possible inquadrarli quali parti di una struttura matematica più ampia e generalizzata. Tale struttura faciliterebbe l'identificazione di tratti universali, eliminando così tecniche di dimostrazione ridondanti; permetterebbe inoltre l'utilizzo di un dizionario singolo idoneo ad ogni situazione, tenuto conto che alcuni dizionari si prestano meglio a considerazioni di tipo teorico mentre altri risultano più adatti all'implementazione digitale.

In questa tesi costruiremo un quadro completo per dizionari parabolici direzionali nel continuo, che denominiamo molecole paraboliche continue. Il nostro lavoro di basa su studi precedenti quali le molecole curvelet [5] e le molecole shearlet, sebbene l'influsso maggiore venga dalle molecole paraboliche discrete [9]. Contrariamente a studi precedenti, rivolgeremo la nostra attenzione alle domande ed ai problemi che appaiono al continuo. Mostreremo anzitutto che la struttura matematica da noi definita è sufficientemente generale da includere sia le costruzioni di tipo curvelet, sia quelle di tipo shearlet, e che le coppie di molecole sono quasi-ortogonali rispetto ad una funzione di pseudo-distanza adattata alla geometria in questione. Utilizzando la quasi-ortogonalità, mostreremo che nozioni fondamentali dell'analisi microlocale, quali la risoluzione del wave front set e regolarità microlocale negli spazi di Sobolev, possono essere rintracciate da ogni tipo di

molecola parabolica continua propriamente scelta. In ultimo, mostreremo che le molecole paraboliche possono essere impiegate per identificare spigoli e angoli di curve chiuse regolari a tratti in $\mathbb{R}^2$.

Acknowledgments

I would like to start by thanking my doctoral supervisor, Professor Philipp Grohs, without whom this thesis would not have been possible. I am very grateful for being given the opportunity to join this group at ETH even though I came from a different background. Your foresight and deep intuition helped foster a successful collaboration, and your humanity made it all worthwhile. Thank you for your generosity and wholehearted support, and for introducing me to this rich and interesting branch of mathematics. I hope your time in Vienna will mark your best days yet.

Furthermore, I want express my deepest gratitude to Professor Bölcksei and to Professor Kutyniok for finding the time to assess my thesis, in the height of what is sure to be an exceptional summer. I am looking forward to our future discussions.

It is hard to find the words that would express the importance of the role my family played throughout my educational path, whose support and encouragement gave the much needed sense of perspective.

My stay here was made enjoyable by many great friends I've made here at SAM and in the broader ETH community. Our lunches, playful and creative banter, excursions and a great many other activities made it a pleasurable experience. I would like to issue my sincerest thanks to the patient souls of Cecilia, Laura and Markus, for having the time and the tenacity to read, translate and impart their wisdom regarding some parts of this thesis.

Lastly, thank you to my friends here and abroad, wherever you may be. I hope our paths will cross again, and often.

Contents

Introduction

The main topic of study in this thesis are families of analysing functions that obey the parabolic scaling law $width = length^2$. Our goal is to develop an abstract framework that unifies pre-existing constructs and puts them into a broader context, and then use the developed machinery to try and answer problems in microlocal analysis. We will call this framework the continuous parabolic molecules.

Let us begin by briefly describing the motivations of our work from a historical perspective. A more detailed account of the historical and mathematical context will be discussed in Chapter 1. In recent decades it became increasingly evident that the classical methods of harmonic analysis have a very limited capacity for addressing a variety of far-reaching features of higher-dimensional objects. Furthermore, the upsurge of computing power inadvertently automatised the collection of data which percolated through every aspect of modern society and science. This created a need for efficient methods of storing and analysing such a deluge of data due to limitations in memory, processing power, and other physical constraints. Therefore, what we desire are methods that provide efficient and robust means of representing everyday data.

A fundamental feature of real world signals or images are their discontinuities. The inadequacy of Fourier transform for signals with discontinuities is apparent already in dimension one; the Gibbs' phenomenon suggests that any partial sum of Fourier series will necessarily exhibit large oscillations near a point singularity. These issues are commonly attributed to the fact that the Fourier transform lacks the sensitivity to changes in the time domain. Time-frequency analysis attempts to address those issues by *windowing* the data: the domain is split into pieces and the Fourier transform is then computed on each piece. Short-time Fourier transform and Gabor frames are just some of the methods that adhere to this ideology of windowed time-frequency atoms.

Whereas the methods of time-frequency analysis use a combination of translations and modulations of a mother functions, wavelets use translations and *dilations*. Even though wavelets were conceptually already present in the work of Calderón [10], it was not until the 1980s that their efficacy for applied tasks began to be exploited. Due to their specific architecture, wavelets are well suited for a number of tasks in signal processing and are by now famed for their ability to detect pointwise singularities [11]. Like other families of transform, wavelets can be used to efficiently encode and compress data. This element of their functionality has found its way toward everyday life as can be seen through their role in the JPEG 2000 standard. Unfortunately, the applicability of the standard wavelet

transform in higher dimensions leaves a lot to be desired. This is because standard two-dimensional wavelets are still essentially products of a one-dimensional theory. Therefore, wavelets fail to grab hold of the subtleties that exist in the analysis of higher-dimensional discontinuities. Whereas in one dimension all discontinuities are just points, the picture is significantly more involved already in two dimensions; discontinuities can be either points or curves. These failures are often attributed to the fact that wavelets lack directional sensitivity, hence the inability to resolve directional features, and to the fact that they are inherently isotropic, hence the inadequacy for efficient representation of data.

In the 1980s various groups of researchers in computer vision decided to go in a certain direction driven by developments in the study of human biology [12, 13]. Human visual system exhibits behaviour which is of both multi-scale and multi-orientation fashion, and the human eye is a directionally biased tool: horizontal field of view of an average human is about 180 degrees, whereas the vertical field of view is about 135 degrees [14]. These observation fostered the development of new classes of transforms with a more nuanced directional sensitivity.

In the early 2000s Candés and Donoho created curvelets which were a first fully formal and accessible mathematical framework intended to resolve directional features. Continuous curvelets are a dictionary of analysing elements which unlike the wavelets are guided by not just the scale and location, but also by the orientation [15]. What really sets curvelets apart is their inherent anisotropy; the dilations obey the parabolic law $width{=}length^2$. Therefore, the dilations are twice as strong in one direction than in the other. Consequently, the result is a family of needle-like functions whose support gets increasingly thin at higher scales. A couple of years later Kutyniok et al [2] created shearlets to address some of the issues that are fundamentally unavoidable with curvelets, which we will discuss in Chapters 1 and 2. Since then a plethora of other families, such as Bandlets [16], Contourlets [3], and others, each obeying some version of the parabolic scaling law, have been constructed by a great number of researchers. Each new construction required a new set of proofs tailored to the occasion to reach all of the desired properties. This creates a possibly unnecessary backlog of essentially indistinguishable proofs needed to reach essentially equivalent results.

Discrete parabolic molecules were introduced in [9] to create a unified treatment of discrete frames that adhere to the parabolic scaling law. We will discuss discrete parabolic molecules in more detail in Chapter 2. The continuous parabolic molecules are in many ways an extension of discrete parabolic molecules to the continuous setting, which tries to answer the question more relevant for microlocal analysis. What we are trying to achieve is to create a framework that allows for a unified treatment of a broad class of existing parabolic dictionaries, has the functionality of converting the results pertinent to one dictionary to all other dictionaries, gives satisfactory answers to mathematically interesting questions and yet is simple enough to work with.

Let us briefly outline the structure of this thesis. We shall begin with Chapter 1 where we will attempt to elucidate the broader mathematical perspective that brought these issues about. Furthermore, we will define standard notions of harmonic analysis and distribution theory that will permeate throughout the thesis such as the Fourier

transform, frame theory, wavelet transform, second generation curvelets, and other notions. In Chapter 2 we will introduce continuous parabolic molecules, define all the requisite notions, show that it is a framework broad enough to include parabolic dictionaries of our interest, and at the end prove its essential feature in Theorem 2.4.1; almost orthogonality of continuous parabolic molecules. Almost orthogonality serves not just as a tool to be used in proofs, but it is also a justification for the intuitive point of view that parabolic analysing functions whose scale, orientation and location parameters are close to each other, with respect to a specific pseudo-distance function, are essentially the same. Chapter 3 will serve to put the tools developed in the later parts of Chapter 2 into action. We will first define and discuss some select notions of microlocal analysis, such as the wavefront set and microlocal Sobolev regularity, and then analyse those notions with the machinery of our framework. In Chapter 4 we will focus on a specific instance of wavefront sets, namely when the underlying distribution is an indicator function of a bounded domain in $\mathbb{R}^2$. The wavefront set of such functions is exactly the boundary of the domain and we will argue that it can be recognised by detecting the point-direction pairs that exhibit fast decay. We will look at edge and corner points, focusing our attention on the latter.

The thesis will be brought to an end with a short summary, conclusion and a discussion regarding possible future directions. Technical aspects of some proofs will be deferred to the Appendix A.

To conclude the Introduction let us spend a moment to discuss the vernacular that will be present in this work. Terms such as normal and orthogonal, family of functions and dictionary, direction and orientation, scaling and dilation, frequency domain and Fourier domain, and other analogous examples, will be used interchangeably. To standardise the notation we have included a list of glossary and symbols we shall be using, which can be found in Appendix B.

Chapter 1

Preliminaries

1.1 Fourier Transform

The theory of Fourier analysis began with the research of its namesake, Joseph Fourier, which he presented in two groundbreaking papers; the 1807 paper *Mémoire sur la propagation de la chaleur dans les corps solides* and the 1822 paper *Théorie analytique de la chaleur*. In those papers Fourier studied trigonometric series for the purposes of solving the heat diffusion in a metal plate. This is a physical process that models the transfer of thermal energy, and is in its simplest terms governed by an elementary linear differential equation, the heat equation. Before Fourier's pioneering work, the problem of finding the general solution of the heat equation was still largely unsolved. It was known at the time that the heat equation admits certain particular solutions, provided the heat source behaves like a sine or a cosine wave. Such solutions are in contemporary mathematical vernacular often called the eigensolutions. Fourier's approach was to model an arbitrary heat source as a superposition of sine and cosine waves, and then write the corresponding solution as a superposition of eigensolutions. It is this linear superposition that we have to come to call the *Fourier series*. For functions $f \in L^2([0,1]^d)$ the Fourier series is formally defined as

$$f(\mathbf{x}) = \sum_{\mathbf{n}\in\mathbb{Z}^d} \hat{f}(\mathbf{n}) e^{2\pi\imath\mathbf{x}\cdot\mathbf{n}}, \tag{1.1}$$

where $\{e^{2\pi\imath\mathbf{x}\cdot\mathbf{n}}\}_{\mathbf{n}\in\mathbb{Z}^d}$ is the corresponding family of eigensolutions and the coefficients are computed by $\hat{f}(\mathbf{n}) = \langle f, e^{2\pi\imath\mathbf{x}\cdot\mathbf{n}}\rangle_{L^2([0,1]^d)}$. As often is the case with great ideas, Fourier's approach seems painfully obvious from the modern perspective, which only serves to cement the ingenuity and the validity of his ideas.

Although Fourier was not the first to study trigonometric representations of functions, he was the first to suggest the audacious claim that an arbitrary function can be represented as a sum of sines and cosines. This fundamental insight (although not completely true in the broadest sense) has had a deep and a profound impact and led to the development of many fields of mathematics, physics and engineering. We have to add that Fourier's exposition lacked the rigour and formality of present day mathematics since the theory of integration would only be formalised later by Riemann, and then

refined and extended by Lebesgue at the turn of the 19^{th} century.

Let us now examine the Fourier transform in $\mathbb{R}^d$ and discuss some of its elementary properties. The *Fourier transform* $\hat{f}$ of a given function f is defined by

$$\hat{f}(\boldsymbol{\xi}) = \int_{\mathbb{R}^d} f(\mathbf{x}) e^{-2\pi\imath \mathbf{x}\cdot\boldsymbol{\xi}} d\mathbf{x}. \tag{1.2}$$

Intuitively, the value $\hat{f}(\boldsymbol{\xi})$ can be thought of as a measure of the amount of the component of f that oscillates at frequency $\boldsymbol{\xi}$. The *inverse Fourier transform* is for a function g defined by

$$\check{g}(\mathbf{x}) = \int_{\mathbb{R}^d} g(\boldsymbol{\xi}) e^{2\pi\imath \mathbf{x}\cdot\boldsymbol{\xi}} d\boldsymbol{\xi}. \tag{1.3}$$

There are several common conventions regarding the choice of constants that can be used to define the Fourier transform which may affect some of its properties. The choice we adhere to in (1.2) makes the Fourier transform an isometry.

What we still need to discuss are the functions spaces on which the Fourier transform is defined. The natural domain of definition of the Fourier transform for functions defined on the Euclidean space $\mathbb{R}^d$ is $L^1(\mathbb{R}^d)$, since for functions in $L^1(\mathbb{R}^d)$ there are no convergence issues with regards to the integral in (1.2), and the Fourier transform is well defined. With the help of Plancherel's theorem the Fourier transform can then be defined on $L^2(\mathbb{R}^d)$, which can be extended to $L^p(\mathbb{R}^d)$ for $1 \leqslant p \leqslant 2$ using the Riesz-Thorin interpolation theorem, and further to tempered distributions through continuity arguments [17,18]. The Fourier transform satisfies many useful properties, some of which we shall now list. The proofs can be found in many standard texts on Fourier transform, such as [18,19].

Proposition 1.1.1. *Let the* translation operator $\mathsf{T}_{\mathbf{x}_0}$ *and the* modulation operator $\mathsf{M}_{\boldsymbol{\xi}_0}$, *for* $\mathbf{x}_0, \boldsymbol{\xi}_0 \in \mathbb{R}^d$, *be defined for functions* $f : \mathbb{R}^d \to \mathbb{R}$ *by*

$$\mathsf{T}_{\mathbf{x}_0} f(\mathbf{x}) = f(\mathbf{x} - \mathbf{x}_0), \qquad \text{and } \mathsf{M}_{\boldsymbol{\xi}_0} f(\mathbf{x}) = e^{2\pi \mathbf{x}\cdot\boldsymbol{\xi}_0} f(\mathbf{x}).$$

The following properties hold for functions $f, g \in L^2(\mathbb{R}^d)$.

(i) $\widehat{[\mathsf{T}_{\mathbf{x}_0} f]} = \mathsf{M}_{-\mathbf{x}_0} \hat{f}$ *and* $\widehat{[\mathsf{M}_{\mathbf{x}_0} f]} = \mathsf{T}_{\mathbf{x}_0} \hat{f}$.

(ii) *Plancherel Formula:* $\langle f, g\rangle = \langle \hat{f}, \hat{g}\rangle$, *and specifically* $\|f\|_2 = \|\hat{f}\|_2$.

(iii) *Convolution Theorem:* *Let* $f, g \in L^1(\mathbb{R}^d)$. *Then* $\widehat{f \star g}(\boldsymbol{\xi}) = \hat{f}(\boldsymbol{\xi})\hat{g}(\boldsymbol{\xi})$.

(iv) *Let* $\boldsymbol{\alpha} \in \mathbb{N}_0^d$ *be a multi-index. Then* $\widehat{\partial^{\boldsymbol{\alpha}} f}(\boldsymbol{\xi}) = (-2\pi\imath)^{|\boldsymbol{\alpha}|} \boldsymbol{\xi}^{\boldsymbol{\alpha}} \hat{f}(\boldsymbol{\xi})$.

(v) *We have* $(\widehat{f(\mathsf{x}_1) g(\mathsf{x}_2)})(\boldsymbol{\xi}) = \hat{f}(\xi_1)\hat{g}(\xi_2)$.

(vi) *If* $f(\mathsf{x}_1, \mathsf{x}_2) = g(\mathsf{x}_2, \mathsf{x}_1)$ *then* $\hat{f}(\xi_1, \xi_2) = \hat{g}(\xi_2, \xi_1)$, *and also vice versa, if* $\hat{f}(\xi_1, \xi_2) = \hat{g}(\xi_2, \xi_1)$ *then* $f(\mathsf{x}_1, \mathsf{x}_2) = g(\mathsf{x}_2, \mathsf{x}_1)$.

(vii) *Let* $\mathrm{A} \in \mathbb{R}^{d\times d}$ *be a regular matrix. For* $f(\mathbf{x}) = g(\mathrm{A}\mathbf{x})$ *we have*

$$\hat{f}(\boldsymbol{\xi}) = |\det \mathrm{A}|^{-1} \hat{g}\left(\mathrm{A}^{-\top}\boldsymbol{\xi}\right).$$

(viii) *Complex conjugation:* *If* $f(\mathbf{x}) = g^-(\mathbf{x}) := \overline{g(-\mathbf{x})}$ *then* $\hat{f}(\boldsymbol{\xi}) = \overline{\hat{g}(\boldsymbol{\xi})}$.

1.2 Frames

One of the fundamental questions in mathematics is how to (efficiently) represent vectors, functions and other objects of interest. We can distinguish between two standard approaches. The first is a more recent development where we want to construct data-adapted dictionaries. In other words, we want to construct an algorithm, which has access to samples from a class of functions $\mathcal{F}$, that constructs a dictionary of elements, mostly by minimising some functional, which allows for sparse and rapid computations of coefficients that give a representation of the elements in the given class $\mathcal{F}$. This is called dictionary learning [20, 21]

We will focus on the more standard approach were the dictionary is precomputed. The classical notion that tries to answer this question in the study of vectors and vector spaces is that of the basis. Given a vector space $\mathcal{V}$, a set of vectors $\mathcal{B} = \{e_j\}_{j\in\mathcal{J}} \subset \mathcal{V}$ is said to be a *basis* of $\mathcal{V}$ if an arbitrary element $f \in \mathcal{V}$ can be expressed as a linear superposition of elements of $\mathcal{B}$ as

$$f = \sum_{j\in\mathcal{J}} c_j e_j, \tag{1.4}$$

where the coefficients $\{c_j\}_{j\in\mathcal{J}}$ are uniquely determined.

Having a basis allows us to restrict our attention to only the elements of $\mathcal{B}$ in a variety of contexts. In order to express this point of view more vividly, let $\mathsf{T}\colon \mathcal{V} \to \mathcal{W}$ be a linear operator between two vector spaces, $\mathcal{V}$ and $\mathcal{W}$. The value of $\mathsf{T}f$, for $f \in \mathcal{V}$, can be expressed as

$$\mathsf{T}f = \sum_{j\in\mathcal{J}} c_j \mathsf{T}e_j.$$

Therefore, in order to make inferences about the properties of T and its effects on arbitrary elements of $\mathcal{V}$, it is sufficient to study only the effects and properties of T on the elements of the basis. We have already seen the standard example of a basis in (1.1), the exponentials $\{e^{2\pi\imath\mathbf{n}\cdot\mathbf{x}}\}_{\mathbf{n}\in\mathbb{Z}^d}$, which form an orthonormal basis for $L^2([0,1]^d)$. Therefore, exponential functions are the fundamental building blocks of Fourier series and of the study of the whole of Fourier analysis.

Unfortunately, the conditions that qualify a set to be a basis can be very restrictive, and it is often exceedingly difficult to find sets of vectors that can successfully couple the constraints that come with being a basis together with other properties that we might desire in a certain scenario. The definition of the basis dictates two things; the closed linear span of $\mathcal{B}$ must be the entire vector space $\mathcal{V}$, and the coefficients $\{c_j\}_{j\in\mathcal{J}}$ in (1.4) must be unique, which means that $\mathcal{B}$ cannot have any superfluous elements.

Therefore, we often find ourselves in one of the following positions: the set we have is either a basis but does not posses properties we desire, or it does posses properties we desire but is not a basis. For example, there are several essential properties of a signal that Fourier transform fails to, and indeed cannot, capture that can traced back to the properties of functions in $\{e^{2\pi\imath\mathbf{n}\cdot\mathbf{x}}\}_{\mathbf{n}\in\mathbb{Z}^d}$. This arises from the fact that the Fourier transform only shows the frequency properties of the signal, while it loses the

information in the time domain. Consequently, it does not provide the information regarding, for example, the duration of the signal, or the time of its emission. There are theoretical results that make these claims rigorous in more general contexts. A standard example is the *Balian-Low* theorem [22], which shows that certain Riesz bases[1] cannot be simultaneously well localised in both time and frequency. These issue express a need for the development of a mathematical framework that would provide more flexibility, while retaining the crucial properties.

Time-frequency analysis is the study of the techniques and representations that allow for a simultaneous control of both the time and the frequency components of a signal. An elementary method of time-frequency analysis, and the first step in this direction, is the *short-time Fourier transform* (herein STFT). The line of thought underlying STFT is fairly simple; we want to improve the Fourier transform by restricting the frequencies that contribute to Fourier coefficients of a given function f to a finite interval. Naturally, this process should not burden the analysis of f with unwanted effects, such as, creating singularities that did not exist prior to the restriction. Therefore, the cut-off should be done smoothly, that is, the restriction to an interval ought to be done using a smooth window function. On the other hand, we want to get a hold of information regarding any significant local phenomena occurring in both f and $\hat{f}$. Therefore, we should be allowed to shift the focus between different regions in both time and frequency domains. In order to make this happen we should bring Proposition 1.1.1 back to mind, since it tells us that modulations and translations are each other's Fourier multipliers. Therefore, a composition of modulation and translation operators would achieve the flexibility we desire. The STFT of a function f with respect to a window function g is typically defined by

$$\mathsf{V}_g f(\mathbf{x},\boldsymbol{\xi}) := \langle f, \mathsf{M}_{\boldsymbol{\xi}}\mathsf{T}_{\mathbf{x}} g\rangle = \int_{\mathbb{R}^d} f(\mathbf{t})\overline{g(\mathbf{t}-\mathbf{x})}e^{-2\pi\imath\mathbf{t}\cdot\boldsymbol{\xi}}d\mathbf{t}, \qquad \text{for } \mathbf{x},\boldsymbol{\xi}\in\mathbb{R}^d. \tag{1.5}$$

The transform (1.5) gives a representation of an arbitrary function $f \in L^2(\mathbb{R}^d)$ with respect to an uncountable family of functions $\{\mathsf{M}_{\boldsymbol{\xi}}\mathsf{T}_{\mathbf{x}} g : (\mathbf{x},\boldsymbol{\xi})\in\mathbb{R}^d\times\mathbb{R}^d\}$. The domain of definition of STFT can be extended: $V_g f$ is a well-defined tempered distribution whenever $f, g \in \mathcal{S}'(\mathbb{R}^d)$. Special attention needs to be devoted to the choice of the window function g. All things considered, g should be smooth to ensure a nice, effect-free cut-off, and it should also have compact support so that the STFT yields localised information about the signal f.

Dennis Gabor went a step further using a specific form of the short-time Fourier transform [23]. In doing so he unwittingly introduced a new paradigm of signal analysis. Given the fact that $L^2(\mathbb{R}^d)$ is a separable Hilbert space, Gabor hypothesised that a series expansion with respect to only a countable subset of time-frequency shifts should be sufficient to represent any function $f \in L^2(\mathbb{R}^d)$. That is, that we have representations of the form

$$f = \sum_{\mathbf{j},\mathbf{k}\in\mathbb{Z}^d}\sum c_{\mathbf{j},\mathbf{k}}\mathsf{T}_{\alpha\mathbf{j}}\mathsf{M}_{\beta\mathbf{k}} g,$$

[1] A *Riesz basis* for a Hilbert space $\mathcal{H}$ is a family of the form $\{\mathsf{U}e_j\}_{j\in\mathbb{N}}$, where $\{e_j\}_{j\in\mathbb{N}}$ is an orthonormal basis for $\mathcal{H}$ and $\mathsf{U}\colon \mathcal{H}\to\mathcal{H}$ is a bounded bijective operator

where $\alpha, \beta > 0$ are the sampling parameters of the integer lattice $\mathbb{Z}^d$, and the coefficients $\{c_{\mathbf{j},\mathbf{k}}\}_{\mathbf{j},\mathbf{k}\in\mathbb{Z}^d}$ are to be determined. For the window function g Gabor proposed the Gaussian window, partly due to the fact that the exponential functions (that is, Gaussians) are eigenvectors of the Fourier transform.

Families of functions, such as those Gabor introduced, are over-complete, that is, they do not form a basis even though they span the ambient space. Such families of functions are a prototypical example of *frames*. Frames generalise the notion of the orthonormal basis for a Hilbert space as they allow non-unique decompositions of the elements of the ambient space. At a first glance it might come across as troublesome that the reconstruction formula frames provide is not unique, but it is exactly this redundancy that plays a key role in many applications. Gabor used frames only implicitly, and it took a further 6 years until the now classical definition of a discrete frame was formalised by Duffin and Schaeffer.

Definition 1.2.1. *A sequence $\mathcal{E} = \{e_j\}_{j\in\mathcal{J}}$, where $\mathcal{J}$ is countable, in a (separable) Hilbert space $\mathcal{H}$ is called a* discrete frame *if there exist constants $0 < A \leqslant B < \infty$ such that*

$$A\|f\|_{\mathcal{H}}^2 \leqslant \sum_{j\in\mathcal{J}} |\langle f, e_j\rangle|^2 \leqslant B\|f\|_{\mathcal{H}}^2, \qquad \text{for all } f \in \mathcal{H}$$

The constants A and B are called upper *and* lower frame bounds*, respectively. If $A = B$, then $\{e_j\}_{j\in\mathcal{J}}$ is said to be a* tight frame*, and if $A = B = 1$ we call $\{e_j\}_{j\in\mathcal{J}}$ a* Parseval frame*.*

It is easy to see that every basis of a Hilbert space is also a Parseval frame for that space. But the true power and beauty of frames for applications in signal processing was only later uncovered by Daubechies and Grossman [24]. Frames are nowadays a ubiquitous tool in signal processing and their applications can be found in many fields of both pure and applied mathematics.

There are two fundamental operators associated with a given frame. The *analysis operator* is an operator that maps each element of $\mathcal{H}$ to a sequence of *frame coefficients* by

$$\mathsf{C} : \mathcal{H} \to l^2(\mathcal{J}), \quad f \mapsto \{\langle f, e_j\rangle\}_{j\in\mathcal{J}}.$$

The adjoint of the analysis operator is called the *synthesis operator*, and to each sequence in $l^2(\mathcal{J})$ it assigns a linear superposition with respect to the frame $\{e_j\}_{j\in\mathcal{J}}$ through

$$\mathsf{C}^* : l^2(\mathcal{J}) \to \mathcal{H}, \quad \{c_j\}_{j\in\mathcal{J}} \mapsto \sum_{j\in\mathcal{J}} c_j e_j.$$

The most important operator associated with a frame is the *frame operator*, defined by $\mathsf{S} = \mathsf{C}^*\mathsf{C}$. The frame operator allows for a stable reconstruction of a signal from its frame coefficients. The operator S (as well as its inverse) is self-adjoint, positive and invertible, and for every $f \in \mathcal{H}$ we have

$$f = \sum_{j\in\mathcal{J}} \langle f, e_j\rangle \mathsf{S}^{-1} e_j = \sum_{j\in\mathcal{J}} \langle f, \mathsf{S}^{-1} e_j\rangle e_j. \tag{1.6}$$

It can be shown that the sequence $\{\mathsf{S}^{-1}e_j\}_{j\in\mathcal{J}}$ is also a frame, and is hence usually called the *canonical dual frame*. There could be an abundance of other frames for which an analogue of equation (1.6) holds. One active area of research in frame theory is the study of the properties of an arbitrary dual frame (not necessarily the canonical dual), if we have the knowledge regarding only the original frame.

The notion of the discrete frame can be generalised to families of functions indexed by locally compact Hausdorff spaces[2] endowed with a a positive Radon measure[3] [25]. In this thesis we will mostly restrict our attention to the corresponding *continuous frames*.

Definition 1.2.2. *Let $\mathcal{H}$ be a separable Hilbert space and $\mathcal{X}$ a locally compact Hausdorff space endowed with a positive Radon measure μ with* $\operatorname{supp}\mu = \mathcal{X}$. *A family of vectors $\mathcal{E} = \{e_x\}_{x\in\mathcal{X}}$ is called a* continuous frame *if there exist constants $0 < A \leqslant B < \infty$ such that*

$$A\|f\|_{\mathcal{H}}^2 \leqslant \int_{\mathcal{X}} |\langle f, e_x\rangle|^2 d\mu(x) \leqslant B\|f\|_{\mathcal{H}}^2, \qquad \text{for all } f \in \mathcal{H}.$$

If $A = B$, then $\{e_x\}_{x\in\mathcal{X}}$ is called a tight frame, and if $A = B = 1$ we say $\{e_x\}_{x\in\mathcal{X}}$ is a Parseval frame.

The frame operator is again denoted by S and continuous analogues of (1.6) can be shown to hold. For example, if the frame $\{e_x\}_{x\in\mathcal{X}}$ is tight, the following reconstruction formula holds in the weak sense

$$f = \frac{1}{A}\int_{\mathcal{X}} \langle f, \mathsf{S}^{-1}e_y\rangle e_y d\mu(y).$$

There again might be many frames $\{\tilde{e}_x\}_{x\in\mathcal{X}}$ such that

$$f = \int_{\mathcal{X}} \langle f, \tilde{e}_x\rangle e_x d\mu(x), \tag{1.7}$$

holds for all $f \in \mathcal{H}$. Such frames are called *dual frames*, while $\{\mathsf{S}^{-1}e_x\}_{x\in\mathcal{X}}$ is still referred to as the canonical dual frame. We will come back to reconstruction formulas later in the thesis.

1.3 Wavelets (and their Limitations)

In the previous chapter we discussed how small changes in the frequency domain can cause a tremendous change everywhere in the time domain which we cannot observe with standard Fourier machinery due to the fact that Fourier basis functions are not well localised in time, Time frequency analysis tries to address these and other flaws of the classical Fourier analysis by cutting the signal into slices and performing Fourier

[2]A topological space $\mathcal{X}$ is said to be Hausdorff if distinct points have disjoint neighbourhoods, and it is locally compact if every point of $\mathcal{X}$ has a compact neighbourhood

[3]A measure μ defined on the σ-algebra $\mathcal{B}(\mathcal{X})$ of Borel sets of a locally compact Hausdorff space $\mathcal{X}$ is called the *Radon measure* if the measure $\mu(B)$, for any Borel set B, is the supremum of $\mu(K)$, over all compact subsets K of B, and if it is finite on all compact sets.

analysis on each slice. Another important question we have to think about is that of the efficient representation of data. Functions with discontinuities and functions with sharp spikes cannot be compactly represented using Fourier basis functions, as can be traced back to the Gibbs phenomenon. Wavelet analysis can be seen as a surrogate candidate for addressing the said issues.

As the saying goes, the clue is in the name. *Wavelets* are functions that exhibit wave-like oscillations, have zero average, and are brief (hence the diminutive), that is, they are well localised. Let us be more specific and begin in the continuous setting, where we will work throughout this thesis.

Whereas the architecture of time-frequency analysis relies on the location $\mathbf{b}$ and the frequency $\boldsymbol{\xi}$ as its parameters, wavelets depend on the location $\mathbf{b}$ and the *scale* a. This means that instead of using the modulation operator $\mathsf{M}_{\boldsymbol{\xi}}$, wavelet theory uses the isotropic dilation operator, which is defined by

$$\mathsf{D}_a f(\mathbf{x}) = a^{-d/2} f\left(a^{-1}\mathbf{x}\right), \quad a \in \mathbb{R}^+,$$

for functions f on $\mathbb{R}^d$. The normalisation $a^{-d/2}$ ensures that the operator D_a is unitary.

The full wavelet dictionary can now be constructed by considering a family of dilated and translated versions of a single generating function $\psi \in L^2(\mathbb{R}^d)$, traditionally termed the *mother wavelet*, through

$$\psi_{a\,\mathbf{b}}(\mathbf{x}) = \mathsf{D}_a \mathsf{T}_{\mathbf{b}} \psi = a^{-d/2} \psi\left(\frac{\mathbf{x}-\mathbf{b}}{a}\right) : a \in \mathbb{R}^+,\, \mathbf{b} \in \mathbb{R}^d. \tag{1.8}$$

It is important to respect this order of operations; we first apply the translation operator, and then the dilation operator [26].

The wavelet transform of a function $f \in L^2(\mathbb{R}^d)$ can now be defined by

$$\mathcal{W}_f(a, \mathbf{b}) = \langle f, \psi_{a\,\mathbf{b}} \rangle = \langle f, \mathsf{D}_a \mathsf{T}_{\mathbf{b}} \psi \rangle.$$

One of the areas of research in wavelet theory is to find conditions that ensure that f can be reconstructed from its wavelet coefficients by the reproduction formula

$$f = \int_{\mathbb{R}^+ \times \mathbb{R}^d} \langle f, \psi_{a\,\mathbf{b}} \rangle \psi_{a\,\mathbf{b}}\, d\lambda(a, \mathbf{b}),$$

where λ is the (left) Haar measure[4] with $d\lambda(a, \mathbf{b}) = \frac{da\, d\mathbf{b}}{a^d}$. The first such condition was established by Calderón in 1964 [10], though from an entirely different point of view. We say that $\psi \in L^2(\mathbb{R}^d)$ is an *admissible wavelet* if

$$\int_{a>0} |\hat{\psi}(a\boldsymbol{\xi})|^2 \frac{da}{a} = 1, \quad \text{for a.e. } \boldsymbol{\xi} \in \mathbb{R}^d. \tag{1.9}$$

[4]Measure μ is a left Haar measure if it is a regular Borel measure on a locally compact group G which is finite on compact subsets and is left-invariant, that is, if for every set S in the Borel σ-algebra of G, $\mu(gS) = \mu(S)$, for all $g \in G$

This condition was rediscovered by Morlet and Grossman in the context of signal processing [26, 27]. Equation (1.9) is not particularly restrictive and there are a plethora of wavelet constructions that have been found to satisfy it, such as the classical Haar and Shannon wavelets, Daubechies wavelets, and the Lemarié-Meyer wavelets [11], to name but a few.

The wavelet transform has a particularly neat group theoretical interpretation, which is a feature it shares with short-time Fourier transform. An elementary computation gives that the composition of operators $\mathsf{T}_{\mathbf{b}}$ and D_a satisfies the following group law

$$(\mathsf{D}_{a_2}\mathsf{T}_{\mathbf{b}_2})(\mathsf{D}_{a_1}\mathsf{T}_{\mathbf{b}_1}) = \mathsf{D}_{a_1a_2}\mathsf{T}_{a_1\mathbf{b}_2+\mathbf{b}_1}, \quad \text{for } \mathbf{b}_1, \mathbf{b}_2 \in \mathbb{R}^d, \quad \text{and } a_1, a_2 \in \mathbb{R}^+.$$

This property defines an operation on the group $\mathbb{R}^+ \times \mathbb{R}^d$ through

$$(a_2, \mathbf{b}_2) \cdot (a_1, \mathbf{b}_1) := (a_1a_2, a_1\mathbf{b}_2 + \mathbf{b}_1).$$

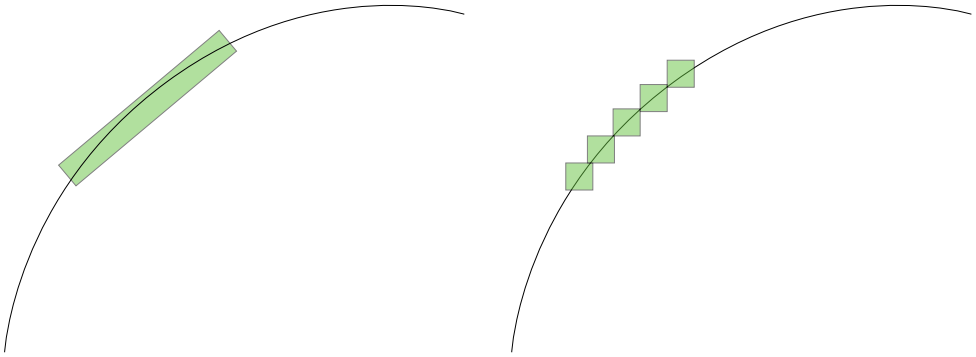
The $a\mathbf{b}_2 + \mathbf{b}_1$ group on $\mathbb{R}^d$ is the so called *group of affine transformations* on $\mathbb{R}^d$. Further analysis unveils that the wavelet transform is the representation coefficient of this affine group with respect to the unitary representation $\pi(a, \mathbf{b}) = \mathsf{D}_a\mathsf{T}_{\mathbf{b}}$, acting on $L^2(\mathbb{R}^d)$.

For the sake of completeness, let us briefly discuss the standard discretisation of wavelets, restricting our attention to $d = 1$. The question we have to answer is how to sample the dilations and the locations. Typical wavelet bases or frames use dyadic scaling and translations along the integer lattice

$$\{\psi_{j,k} = 2^{-j/2}\psi\left(2^{-j} \cdot -k\right) = \mathsf{D}_2^j\mathsf{T}_k\psi : j, k \in \mathbb{Z}\}.$$

There are two issues we have to address here. Let us say that we want to use the wavelet architecture to represent a function, that is, use an analogue of (1.6). The canonical dual frame associated with the wavelet frame $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ is given through the frame operator as $\{\mathsf{S}^{-1}\psi_{j,k}\}_{j,k\in\mathbb{Z}}$. Computing the canonical dual can be a challenge because inverting the frame operator S can be highly non-trivial. Assuming the canonical dual is also a wavelet frame, it could be obtained by first computing its mother wavelet, and then applying the dilation and translation operators. If that is not the case, we would have to apply S^{-1} to a double-infinite set of functions. The Achilles' heel of the problem is that the frame operator S and operators $\mathsf{D}_2^j\mathsf{T}_k$ do not necessarily commute. Computational difficulties aside, there are explicit examples that attest to the fact we cannot *a priori* say that the canonical dual of a wavelet frame will also have wavelet structure. To find a work-around to these problems we can look into the very heart of what brought frames about. If the family $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ is over-complete, we can exploit its redundancy and search for a different dual frame, one that has wavelet structure, but this then usually amounts to a case-by-case analysis.

Unfortunately though, for all of its upsides, the standard wavelet transform we just described comes with serious limitations when it comes to the analysis of higher dimensional functions. To show how these limitations come about we will focus on the two-dimensional case, which is the standard model used for the analysis of images. The problem is that the wavelet transform fails to reveal geometric properties of an image



(a) Anisotropic elements (b) Isotropic elements

Figure 1.1: Comparison of paving a curved edge using anisotropic and isotropic elements. We can see that the anisotropic structure of the underlying dictionaries results in a more efficient representation

since it lacks directional selectivity, as by Definition (1.8) wavelets take into account only the location and the scale. In other words, standard multi-variate wavelets are products of an essentially one-dimensional theory [11, 28]. Therefore, it should come as no surprise that there are unavoidable obstacles when dealing with features of multi-variate data.

One such feature, which is of directional nature, is the edge of an image. Therefore, in order to resolve directional features in a desired class of objects we need transforms that take into account different orientations. A straightforward attempt to rectify this dire predicament, could probably be reduced to simply incorporating orientation parameters in (1.8) by say, including the rotation operator R_θ into the definition of the wavelet family. Furthermore, wavelets are well known to yield high-amplitude coefficients in the vicinity of edges, but they cannot take advantage of any further regularity of an object to improve the sparsity of the representation. Thus, these new and improved wavelets ought to have elongated supports, which could be achieved by pre-stretching the mother wavelet in one direction, according to, for example, $\tilde{\psi}(\mathbf{x}) = \psi(x_1, x_2/100)$. Such an approach would certainly give oriented, directional wavelets that admit a reconstruction formula and satisfy a number of other desirable properties. The issue though is that this would conceptually be nothing new, and as we will see later, still fails to address many interesting questions, even some of which it was designed to answer. Hence, a different approach is needed.

1.4 Beyond the Wavelet Transform: Ridgelets, Curvelets, Shearlets, etc.

One of the remarkable features of wavelet bases and frames lies in their ability to analyse functions that are smooth apart from pointwise singularities [29]. Wavelet coefficients of such objects are sparse, whereas the coefficients of other classical transforms are not. This aspect of the functionality of the wavelet transform has had a wide and significant impact to various corners of both pure and applied mathematics. The significance it bears for compression of data that exhibits such singularities is due to the fact that the wavelet energy of the singularity is concentrated in only a few big coefficients. Consequently, a signal can be (partially) recovered, with good approximation quality, using only a few wavelet coefficients. However, these properties are often used for situations well outside of their original context, and their efficacy can often be overstated. The crux of this misuse can be attributed to the fact that wavelets are fully efficient only for pointwise singularities, whereas the singularities of objects in higher dimensions lie along predominantly higher dimensional structures. For example, the edge of a typical image (we can imagine an equilateral triangle) is a one dimensional singularity, and such singularities are dominant factors in the description of the object.

1.4.1 Best N-term Approximation of Cartoon-Like Images

A typical image is a collection of competing influences, from pointwise and curvilinear discontinuities, to various types of noise, blur and so on. In practice, it would be unfeasible to expect for any single transform to be fully efficient for all possible types of signals. We should instead use a collage of various transforms. When it comes to the mathematical analysis, we often restrict our attention to classes of particular interest. In imaging sciences we would want a class that models natural images, yet is simple enough to allow for a tractable mathematical analysis. A model of images that emphasises the edges is the class of so-called of *cartoon-like images*, and it was introduced by Donoho in [30].

Definition 1.4.1. *The class* $\mathcal{E}^2(\mathbb{R}^2)$ *of* cartoon-like images *is the class of functions* $f\colon \mathbb{R}^2 \to \mathbb{C}$ *of the form*

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

where $B \subset [0,1]^2$ *is a set whose boundary* ∂B *is a closed* C^2 *curve of bounded curvature, and* $f_i \in C^2(\mathbb{R}^2)$ *are functions such that* $\operatorname{supp} f_i \subset [0,1]^2$ *and* $\|f_i\|_{C^2} \leqslant 1$ *for* $i = 0, 1$.

The class of cartoon-like images is a simple prototype of everyday images; there are two regions that are separated by a smooth edge, thus emphasising edge-like features of images, which are by and large anisotropic.

Let us consider now a mathematical context which puts the limitations of the wavelet transform to the fore. Before we begin we ought to make a couple of observations.

The objects we want to analyse will often lie in a strange functional space, are non-intuitive, and can be too complex to assess directly. Drawing on our earlier discussion we

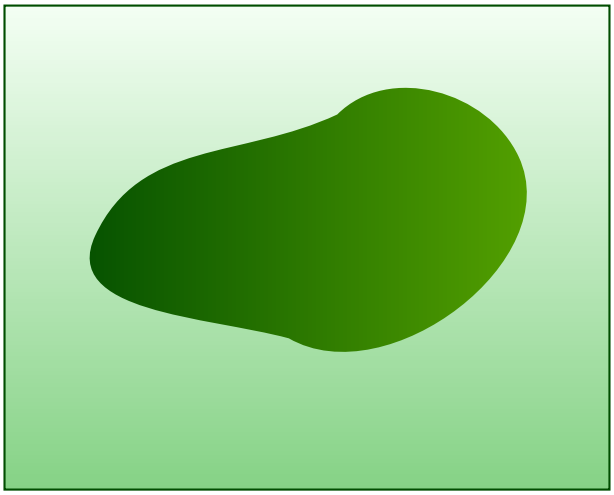


Figure 1.2: A Cartoon-Like Image

instead prefer to examine the coefficients of those objects with respect to some dictionary. Working with the coefficients associated to an object would then yield conclusions regarding the properties of the original object.

If we were to continue with this line of thought, we should add that it would also be desirable to deal with as few of the coefficients as possible. Coming to conclusions from infinitely many coefficients poses not only an unplausible, but also a technically impossible challenge. Therefore, it is in our best interest to have a dictionary that *sparsifies* the objects in a given class as efficiently as possible. This means that there would be only a few significant coefficients, and we could discard the rest. We can formalise these considerations as in [30].

Definition 1.4.2. *Consider a family of functions* $\{\psi_j\}_{j\in\mathcal{J}} \subseteq L^2(\mathbb{R}^2)$. *We say that* $f_N = \sum_{j\in\mathcal{J}_N} c_j\psi_j$, *where* $\mathcal{J}_N \subseteq \mathcal{J}, |\mathcal{J}_N| = N$, *is the* best N-term approximation *of* f *with respect to the family* $\{\psi_j\}_{j\in\mathcal{J}}$ *if*

$$\|f - f_N\|_{L^2} \leqslant \left\| f - \sum_{j\in\tilde{\mathcal{J}}} c_j\psi_j \right\|_{L^2},$$

holds for all $\tilde{\mathcal{J}} \subset \mathcal{J}$, $|\tilde{\mathcal{J}}| = N$, *and all sequences* $\{c_j\}_{j\in\tilde{\mathcal{J}}}$.

In order for Definition 1.4.2 to be of any practical significance, this pursuit of an optimal N-term representation within a given dictionary ought to have constraints. The rationale for having any type of constraints is simple. When there are no restrictions regarding the selection process, some truly pathological examples might emerge. For example, dictionaries that consist of countable and dense subsets of $L^2(\mathbb{R}^d)$ have universally sparse approximations [30]. This means that for an arbitrary element of $L^2(\mathbb{R}^d)$ there would always be a member of such a dictionary that achieves an arbitrarily

small approximation error. The downside is that an approximation result of this type would be of no practical consequence since finding the element that achieves this error is computationally unfeasible.

A standard constraint is to allow only the *polynomial depth search algorithms*. The polynomial depth search constraint essentially imposes a condition that the first N terms of any approximation of a given function f must come from one of the first $\pi(N)$ terms in the dictionary, where π is a polynomial independent of f. Without getting into technical details, which can be found in [30], this restriction eliminates some pathological cases and does not hinder the performance of standard methods in any tangible way.

The optimal asymptotic decay rate of the non-linear approximation error for objects in $\mathcal{E}^2(\mathbb{R}^2)$ was established in [30]. It is optimal in the sense that no other polynomial depth search algorithm can provide a better asymptotic rate.

Theorem 1.4.3. *Let $f \in \mathcal{E}^2(\mathbb{R}^2)$. There exists a constant $C > 0$ such that for any $N \in \mathbb{N}$ there is a triangulation of the box $[0,1]^2$ into N triangles so that the resulting piecewise linear interpolant f_N satisfies the error estimate*

$$\|f - f_N\|_{L^2}^2 \leqslant CN^{-2}, \quad N \to \infty. \tag{1.10}$$

Theorem 1.4.3 can serve as a benchmark for sparse approximations of two-dimensional data using a reconstruction dictionary. The construction used in the proof of Theorem 1.4.3 is in itself important: salient features of analysing elements is that they have elongated supports and exhibit directional selectivity.

A natural question now is to examine how do the classical dictionaries perform in this context. It can be shown that the partial reconstruction obtained by taking the N largest Fourier coefficients satisfies

$$\|f - f_N^{\mathcal{F}}\|_{L^2}^2 \leqslant C_{\mathcal{F}} N^{-1/2}.$$

On the other hand, the best N term approximation with respect to a wavelet basis is obtained by simply taking the N largest wavelet coefficients, that is, the coefficients are of the form $c_j = \langle f, \psi_j \rangle$. It can be shown [1] that the error obtained by using a wavelet family satisfies

$$\|f - f_N^{\mathcal{W}}\|_{L^2}^2 \leqslant C_{\mathcal{W}} N^{-1}.$$

Furthermore, the rate N^{-1} is tight, in the sense that one can construct cartoon-like images for which the approximation rate is bounded by N^{-1} not only from above, but also from below, up to a multiplicative constant. Even though this is significantly better than the asymptotics with Fourier series, wavelets nevertheless fail to approximate edges of cartoon-like images efficiently since the approximation rate N^{-1} is far from optimal.

Taking into account the stark dichotomy between the optimal approximation rate and the rate obtained with wavelets, it should be clear that we ought to pursue dictionaries that yield approximation rates that are closer to the optimal rate. To that end, the ideas from the proof of Theorem 1.4.3 serve as a natural starting point. The ridgelet transform is the first transform that we shall consider that goes in this direction.

1.4.2 Ridgelets

Ridgelets were first introduced by Emmanuel Candés's in his PhD thesis [31], and are typically constructed as follows [32]. Take a smooth function $\psi\colon \mathbb{R} \to \mathbb{R}$, with sufficient decay, a zero average $\int_{\mathbb{R}} \psi(t)dt = 0$ and a normalisation $\int_0^\infty |\hat{\psi}(\xi)|^2 |\xi|^{-2} d\xi = 1$. For each scale $a > 0$, orientation $\theta \in [0, 2\pi)$, and position $b \in \mathbb{R}$, we can now define the *ridgelet* $\psi_{a\,\theta\,b}\colon \mathbb{R}^2 \to \mathbb{R}$ by

$$\psi_{a\,\theta\,b}(\mathbf{x}) = a^{-1/2}\psi\left(\frac{x_1\cos(\theta) + x_2\sin(\theta) - b}{a}\right).$$

The function ψ is typically called the *mother ridgelet*. The ridgelet $\psi_{a\,\theta\,b}$ is constant along the ridges, that is lines, $x_1\cos(\theta) + x_2\sin(\theta) = \mathsf{const}$, and transverse to these ridges it is a wavelet. Given an integrable function $f \in L^1(\mathbb{R}^2)$ we then define ridgelet coefficients by

$$\mathcal{R}_f(a\,b\,\theta) := \langle f, \psi_{a\,\theta\,b}\rangle.$$

The continuous ridgelet transform then admits an exact reconstruction formula

$$f(\mathbf{x}) = \frac{1}{4\pi}\int_0^{2\pi}\int_{-\infty}^{\infty}\int_0^{\infty} \mathcal{R}_f(a,b,\theta)\psi_{a\,\theta\,b}(\mathbf{x})\frac{da\,db\,d\theta}{a^3}, \quad \text{for a.e. } \mathbf{x} \in \mathbb{R}^2,$$

where $f \in L^2(\mathbb{R}^2) \cap L^1(\mathbb{R}^2)$, and also satisfies a Parseval-like formula. The definition of the ridgelet transform can be extended to any dimension $d \geqslant 2$.

A useful way to think about the ridgelet transform is to reimagine it as the wavelet analysis in the Radon transform[5]. The rationale that underlies the construction of the ridgelet transform then becomes clear. Radon transform converts singularities that lie along the lines into point singularities. Using the wavelet transform in the next step yields optimally sparse representations, since wavelets provide optimal representations of point singularities. This reinterpretation of the ridgelet transform serves two further purposes. First, it gives a basic prescription about how to compute the ridgelet transform. Secondly, such an interpretation lends itself to generalisations in other contexts. For example, to define the ridgelet transform on a sphere, we can use spherical Radon and wavelet transforms [33].

Let us add, for the sake of completeness, that there are a number of discretisations of the ridgelet transform that correspond to either frames or orthobases [34]. Furthermore, it has been established that the non-linear approximation using the best N-term approximation with these discretisations achieves a near optimal decay rate for smooth images with discontinuities along straight lines.

1.4.3 Curvelets

The questions ridgelets are trying to address are considerably more difficult than the questions wavelets are trying to. On one hand, zero-dimensional singularities are all

[5]Radon transform of a function f is defined as $R_f(\rho,\tau) = \int_{\mathbb{R}} f(\mathsf{x}, \tau + \rho\mathsf{x})d\mathsf{x}$

exactly alike; they are just points. One-dimensional singularities can on the other hand be either straight lines or (proper) curves. The fact of the matter is that edges of real world images are not perfectly straight but are curved, that is, they have non-zero curvature. Therefore, since ridgelets hinge on the ability of the Radon transform to convert lines into points, they can efficiently represent only objects with singularities that lie along straight lines. A temporary solution is to use ridgelets in a localised way, that is, to make use of the fact that at fine scales curved singularities can be well approximated by straight lines. The method of localisation has a rich history in time-frequency analysis. The idea is to divide a given domain, that is an image, into a set of wavelet bands and localise the function into smooth pieces that are supported at or near those sub-bands by using either partition of unity or smooth orthonormal windowing. The next step consists of analysing each band with a local ridgelet transform. The size of the blocks can also be changed at each scale. It can be shown that using the sub-band decomposition of the domain forces a relationships between the width and length of frame elements. Specifically, the frame elements have to be anisotropic and must (roughly) obey the parabolic scaling law $\textit{width} \approx \textit{length}^2$. These ideas are the building blocks behind the *first generation curvelets*. Interested readers can refer to [35] for further details.

The first generation curvelet elements can form either a frame or a tight frame, are anisotropic by construction and get progressively more anisotropic at higher scales. Furthermore, they have directional selectivity and exhibit oscillatory behaviour across the ridges.

A discrete frame of first-generation curvelets can be constructed by sampling the continuous curvelets, but we will defer further details regarding the sampling until the discussion on second-generation curvelets. It can be shown that the best N-term approximation of a given function f using first generation curvelets obeys

$$\|f - f_N^{\mathcal{C}_1}\|_2^2 \leqslant C_\kappa N^{-2+\kappa}, \quad \text{as } N \to \infty,$$

for every $\kappa > 0$. This rate is near-optimal and is clearly an improvement compared to the best N-term approximation rates obtained with Fourier or wavelet elements.

Even though first generation curvelets have useful and interesting properties, they still leave a lot to be desired. The construction of the corresponding curvelet frame involves a complicated seven-index structure, has a redundancy factor $16M + 1$ whenever M scales are used, and the basic computational complexity of an $n \times n$ image is $\mathcal{O}((n \log n)^2)$, which is already very good but it could be improved [36]. Furthermore, the transform admits a wide range of aspect ratios, meaning that the parabolic scaling law $\textit{width} = \textit{length}^2$ does not always hold. Consequently, rigorous mathematical analysis is very delicate.

The *second generation curvelets* [15] attempt to address these issues. At scale a the second generation curvelet family of analysing element is defined by translating and rotating the generating element $\gamma_{a\,0\,\mathbf{0}}$ according to the rule

$$\gamma_{a\,\theta\,\mathbf{b}}(\mathbf{x}) = \gamma_{a\,0\,\mathbf{0}}(\mathsf{R}_\theta(\mathbf{x} - \mathbf{b})).$$

The generator $\gamma_{a\,0\,\mathbf{0}}$ is defined in polar Fourier coordinates by

$$\hat{\gamma}_{a\,0\,\mathbf{0}}(r, \alpha) = a^{3/4} W(a \cdot r) V\left(\frac{\alpha}{\sqrt{a}}\right).$$

The windows $W(r)$ and $V(\alpha)$ should both be smooth, nonnegative, real valued, with supports obeying

$$\operatorname{supp} W \subset \left(\frac{1}{2}, 2\right), \qquad \operatorname{supp} V \subset [-1, 1],$$

and satisfy some further admissibility conditions we will discuss later in Chapter 2. It follows from the definitions that the support of $\hat{\gamma}_{a\,0\,\mathbf{0}}$ is an oriented polar wedge with a scale-dependent width and length that becomes increasingly needle-like at fine scales. Furthermore, the elements $\hat{\gamma}_{a\,0\,\mathbf{0}}$ obey the parabolic scaling law in polar coordinates r and α, where α is the *thin* coordinate.

The *continuous curvelet transform* of f, denoted Γ_f is defined by

$$\Gamma_f(a, \theta, \mathbf{b}) = \langle f, \gamma_{a\,\theta\,\mathbf{b}} \rangle, \quad a < a_0,\, \theta \in [0, 2\pi),\, \mathbf{b} \in \mathbb{R}^2. \tag{1.11}$$

The value a_0 represents the coarsest scale, and needs to obey $a_0 < \pi^2$ in order for the construction to work. Curvelet coefficients (1.11) are by definition complex-valued, but real-valued curvelets can be obtained by simply combining the symmetric coefficients, that is, by using $\hat{\gamma}_{a\,\theta\,\mathbf{b}}(r, \omega) + \hat{\gamma}_{a\,\theta\,\mathbf{b}}(r, \omega + \pi)$. In order to extend the functionality to functions containing low frequencies we should add coarse scale isotropic father wavelets, but we will not discuss this presently as it is only the behaviour at fine scales that we are interested in.

The second generation curvelet transform defined in (1.11) has an abundance of desirable properties. The elements of the curvelet family have a simple and natural indexation that uses only three parameters, thus simplifying the mathematical analysis Moreover, second generation curvelets admit a reconstruction formula

$$f = \int_{[0,a_0]\times[0,2\pi)\times\mathbb{R}^2} \langle f, \gamma_{a\,\theta\,\mathbf{b}} \rangle \gamma_{a\,\theta\,\mathbf{b}}\; d\mu(a, \theta, \mathbf{b}), \tag{1.12}$$

where $d\mu(a, \theta, \mathbf{b}) = \frac{da\, d\theta\, d\mathbf{b}}{a^3}$ is the a left invariant Haar measure, and a Parseval-like formula that holds for all L^2 functions. The curvelet transform also follows a fully parabolic scaling law and defines a tight frame for $L^2(\mathbb{R}^2)$ with a much lower redundancy than the first-generation curvelets. Furthermore, unlike the first generation curvelets, the second generation curvelets do not use the ridgelet transform in neither continuous nor in discrete settings. Hence, the implementation is of the computational complexity $\mathcal{O}(n^2 \log n)$, which is the same as FFT and is faster than the first generation curvelets by a factor of $\log n$.

The last remaining question is regarding the approximation properties. The standard approach [37] of constructing a discrete tight frame involves the dyadic sampling of the scales $a_j = 2^{-j}$, equidistant sampling of directions with respect to the scale $\theta_{j,l} = 2\pi 2^{-j/2} l$, and the equispaced sampling of the locations that depends on both j and l. The construction consists of two steps. We first sample in scale and angle, leaving the location parameters continuous. This creates a semi-discrete transform from which we then discretise in location. It can be shown that the best N-term approximation of cartoon-like images using the second generation curvelet family satisfies

$$\|f - f_N^{\mathcal{C}_2}\|_2^2 \leqslant C_{\mathcal{C}} N^{-2} (\log N)^3, \quad N \to \infty.$$

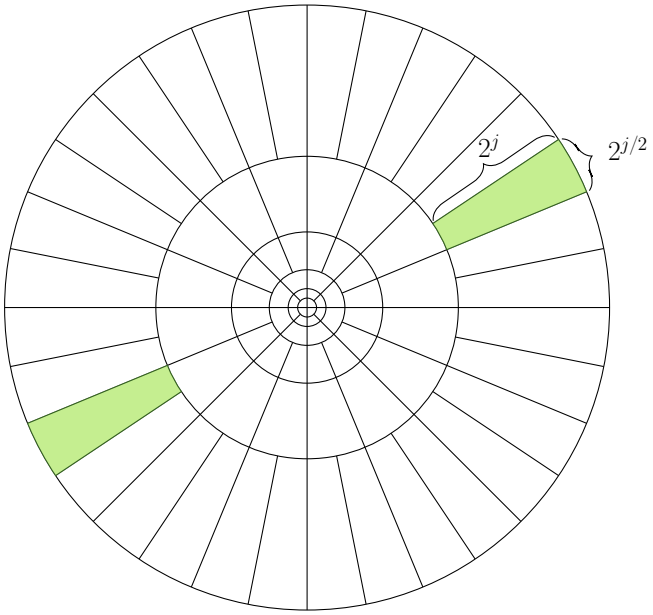


Figure 1.3: Sketch of a typical paving of the domain using a discretisation of the second generation curvelets

In terms of Theorem 1.4.3, this is an asymptotically optimal rate up to the $(\log N)^3$ factor. This result is remarkable having in mind the fact that curvelet analysis does not use adaptive elements. The drawback of this simplicity is that the approximation rate is unfortunately not near-optimal if the edges are not exactly C^2 curves. Moreover, if the edges are $C^{2+\kappa}$ curves, with $\kappa > 0$, the approximation rate does not improve, it remains N^{-2}. It might be relevant to mention that there are geometric representation families (though adaptive), such as contourlets [16], that manage to achieve the optimal decay rate $\mathcal{O}(N^{-2-\kappa})$ for images with $C^{2+\kappa}$ edges.

1.4.4 Shearlets

Curvelets are constructed through a meticulous analysis that combines the translation and the reorientation of a given generating element, while simultaneously dilating the elements of the curvelet dictionary so that they obey the law of parabolic scaling. The resultant transform has a simple structure that lends itself to elegant mathematical analysis, and admits an efficient digital representation. A careful examination of this process shows that somewhere along the way we lost some of the particularly useful properties we had with wavelets. First of all, whereas each wavelet is just an affine transformation of the mother wavelet, a curvelet element $\gamma_{a\,0\,\mathbf{0}}$ cannot be obtained through such a simple transformation of $\gamma_{a'\,0\,\mathbf{0}}$ for $a' \neq a$. Furthermore, directional selectivity of curvelet elements is achieved by using the most obvious choice, the rotation operator. Although rotations may be elegant and straight-forward, they are burdened with an

inherent drawback; rotation operators destroy the lattice $\mathbb{Z}^2$ whenever the rotation angle is not $k\frac{\pi}{2}$ for $k \in \mathbb{Z}$. This will be a serious impediment whenever we would want to transition between continuous and discrete settings. Therefore, as an alternative to the rotation operator we would desire an operator that preserves the integer lattice for a much broader range of inputs. To this end Kutyniok et al constructed *shearlets*, which first appeared in [2].

The *shearing operator* S_s is defined through the shearing matrix

$$\mathrm{S}_s := \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}, \quad s \in \mathbb{R}.$$

The matrix S_s changes the slope of a vector $\mathbf{x} \in \mathbb{R}^2$ by displacing its first coordinate with respect to the shearing variable s, which measures the amount of the displacement. Consequently, provided s is an integer, the lattice $\mathbb{Z}^2$ remains intact and we have $\mathsf{S}_s\mathbb{Z}^2 \subset \mathbb{Z}^2$.

The standard continuous shearlet transform can now be defined by combining the translation, dilation and shear operators. Consider a function $\psi \in L^2(\mathbb{R}^2)$, dubbed the *mother shearlet*, and define the continuous shearlet family by

$$\psi_{a\,s\,\mathbf{b}}(\mathbf{x}) = a^{-3/4}\psi\left(\mathrm{D}_a\mathrm{S}_s^{-1}(\mathbf{x}-\mathbf{b})\right) : a > 0,\ s \in \mathbb{R},\ \mathbf{b} \in \mathbb{R}^2.$$

The corresponding *continuous shearlet transform* of a function $f \in L^2(\mathbb{R}^2)$ is now readily defined as the mapping

$$\mathcal{SH}_f(a, s, \mathbf{b}) = \langle f, \psi_{a\,s\,\mathbf{b}} \rangle.$$

A fundamental question is under what conditions does the continuous shearlet transform define an isometry, since this bears immediate implications on the availability of a reconstruction formula. The standard answer to this question is to say that $\psi \in L^2(\mathbb{R}^2)$ is an *admissible shearlet* if it satisfies

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

The *classical shearlet* is constructed by choosing functions ψ_1, ψ_2, such that

$$\operatorname{supp}\hat{\psi}_1 \subset \left[-\frac{1}{2}, -\frac{1}{16}\right] \cup \left[\frac{1}{16}, \frac{1}{2}\right], \quad \text{and} \quad \operatorname{supp}\hat{\psi}_2 \subset [-1, 1]. \tag{1.13}$$

The mother shearlet is then defined in the Fourier domain by

$$\hat{\psi}(\boldsymbol{\xi}) := \hat{\psi}_1(\xi_1)\hat{\psi}_2\left(\frac{\xi_2}{\xi_1}\right). \tag{1.14}$$

We should add that there is nothing fundamentally important regarding any particular choice of the frequency-domain supports of ψ_1, ψ_2. All that matters is that $\hat{\psi}_1$ is supported away from 0 and that $\hat{\psi}_2$ is supported around 0. This is usually interpreted by saying that ψ is wavelet-like along one coordinate axis and bump-like along the other.

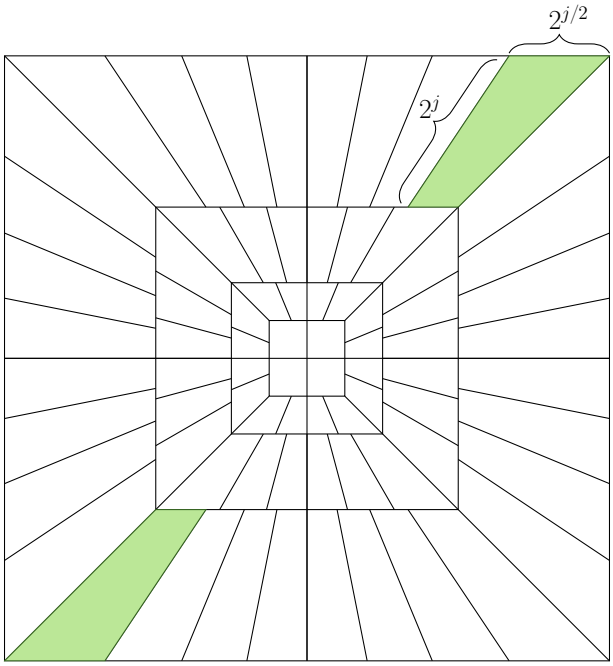


Figure 1.4: Sketch of a typical paving of the domain using a discretisation of classical shearlets

The continuous shearlet system we just described has a number of desirable properties: it allows the reconstruction of a signal and the corresponding Parseval-like formula, has elegant group structure [38], and is generated as an affine transformation of a single mother shearlet. There is however a problem that must be addressed. The continuous shearlets have a certain directional bias. In order to see this we can observe that a singularity that lies along the ξ_2 axis can only be uncovered by considering the limit $s \to \infty$, which poses questions about the suitability of the shearlet transform for practical applications. This issue can be rectified by partitioning the Fourier domain into four high-frequency cones, and a further low-frequency box centred around the origin (see Figure 1.5). In doing so we would, within each of the cones, effectively restrict the shearing parameter to a finite range.

The reimagination of the shearlet transform that adheres to such a construction yields an alternative shearlet system, called *cone-adapted continuous shearlet system*, and it is defined by

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

where $\varphi, \psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$ and

$$\Phi(\varphi) = \{\varphi_{\mathbf{b}} = \varphi(\cdot - \mathbf{b}) : \mathbf{b} \in \mathbb{R}^2\},$$
$$\Psi(\psi) = \{\psi_{a\,s\,\mathbf{b}} = a^{-3/4}\psi\left(\mathrm{D}_a\mathrm{S}_s^{-1}(\cdot - \mathbf{b})\right) : a \in (0,1],\ |s| \leqslant 1 + a^{1/2},\ \mathbf{b} \in \mathbb{R}^2\},$$

$$\tilde{\Psi}(\tilde{\psi}) = \{\tilde{\psi}_{a\,s\,\mathbf{b}} = a^{-3/4}\tilde{\psi}\left(\tilde{\mathrm{D}}_a \mathrm{S}_s^{-\top}(\cdot - \mathbf{b})\right) : a \in (0,1],\, |s| \leqslant 1 + a^{1/2},\, \mathbf{b} \in \mathbb{R}^2\},$$

where $\tilde{\mathrm{D}}_a = \mathrm{diag}\left(a^{1/2}, a\right)$ and $\hat{\tilde{\psi}}(\boldsymbol{\xi}) = \hat{\psi}(\xi_2, \xi_1)$.

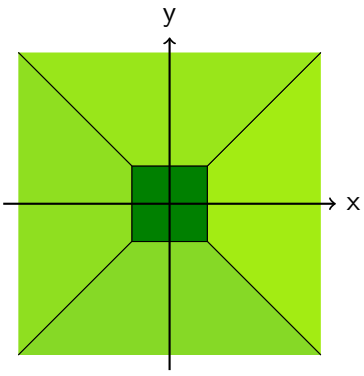


Figure 1.5: Partitioning of the frequency plane into a low-frequency box and four higher frequency cones

We can again define the shearlet transform associated with cone-adapted shearlets, find conditions on $\varphi, \psi, \tilde{\psi}$ which ensure that the corresponding transform is an isometry, and obtain reconstruction and a Parseval-like formulae. It is possible to construct a family of compactly supported shearlets and efficiently discretise (cone-adapted) shearlet systems [39--41]. The book [40] provides an extensive and a comprehensive account of the theory and applicability of the shearlet transform.

The unfortunate consequence of having to resort to cone-adapted shearlets is that we lose some of the initial simplicity. Furthermore, conducting a rigorous mathematical analysis becomes substantially less elegant, especially when compared to second generation curvelets. A further cause of problems lies in the fact that the dual frame of classical shearlets is not necessarily a family of shearlets, whereas the dual frame of second generation curvelets (consider the equation (1.12)) are again second generation curvelets. What we gain though is an affine system that admits an analogous treatment of continuous and discrete realms.

One final test that shearlets ought to be subjected to is the examination of their performance in terms of efficient representations of cartoon-like functions. The cone-adapted discrete shearlet systems can be constructed by sampling the continuous cone-adapted shearlets [40,42]. Due to their construction, shearlets allow for an easier treatment of the discrete setting. We again use dyadic sampling of the dilations $a_j = 2^{-j}$, and the shears satisfy $|s_j| \leqslant \lceil 2^{j/2} \rceil$, $s_j \in \mathbb{Z}$, whereas the locations can be simply taken from $\mathrm{M}_{\mathsf{c}}\mathbb{Z}^2$ where $\mathrm{M}_{\mathsf{c}} = \mathrm{diag}(\mathsf{c})$. It can be shown [43] that there exists a constant $C > 0$ such that the best N-term approximation of a cartoon-like function $f \in \mathcal{E}^2(\mathbb{R}^2)$ with respect to a shearlet system, obtained by choosing N largest coefficients, satisfies

$$\|f - f_N^{\mathcal{SH}}\|_2^2 \leqslant CN^{-2}(\log N)^3, \quad \text{as } N \to \infty.$$

This near-optimal rate is the same as what is achieved with curvelets, and it holds for cone-adapted shearlets, as well as for the compactly supported shearlets.

1.5 Discrete Parabolic Molecules

The notions, definitions, variety of proof techniques, and the ideology that will be discussed and elaborated throughout this thesis (predominantly in Chapter 2) are built on the work on *discrete parabolic molecules*, but adapted to the continuous setting. Discrete parabolic molecules were introduced in [9] and they define a comprehensive framework for discrete parabolic dictionaries and focus on questions and topics that can be answered with discrete frames. Grohs and Kutyniok showed in [9] that one can use discrete parabolic molecules to identify classes of discrete parabolic dictionaries which exhibit the same sparse approximation properties, in terms of the best N-term approximation rate (see Definition 1.4.2). Moreover, one can also show that there are classes of approximation spaces that can be characterised by all discrete parabolic molecules which are *sufficiently good* in some sense. In order to establish such results, Grohs and Kutyniok showed that pairs of frames of discrete parabolic molecules exhibit strong off-diagonal decay with respect to a certain notion of distance which is guided by the inherent geometry of the setting. The Theorem 2.4.1 is analogous to this particular form of almost-orthogonality. The idea of abstractifying parabolic dictionaries, and the result (and the proof) on almost-orthogonality, finds its most immediate precursors and influences in the earlier work of Candés and Demanet on wave propagators and representation of Fourier integral operators [5, 44], and similar studies on the topic conducted by Smith [4]. This functionality was later extended to shearlets in [8].

Most of the work in Chapter 2 builds up on ideas developed in [9] by adapting them to the continuous setting. The elementary difference between the present work and the studies conducted in [9] is that our focus is on continuous dictionaries and continuous frames and the corresponding set of challenges and questions, which we will address and develop in Chapters 3 and 4. We will also adopt elements of the ethos from [9] by looking at the universalities that can be found in the treatment of microlocal analysis with parabolic dictionaries.

Discrete parabolic molecules were later generalised in [45] by replacing the dilation matrix D_a with $\mathrm{D}_{a,\alpha} = \mathrm{diag}(a, a^\alpha)$ for $\alpha \in [0, 1]$. Such constructions are called the *α-molecules*, and we will briefly look at some select consequences of these considerations in Chapter 4. Parabolic molecules are a special case of α-molecules which corresponds to the case $\alpha = \frac{1}{2}$. Ridgelets and wavelets can also be classified as α-molecules for $\alpha = 0$ and $\alpha = 1$, respectively.

1.6 Distribution Theory

We shall devote this last section to a brief discussion of some elements of distribution theory. Distribution theory is one of the most important contributions of the 20^{th} century mathematics. The early history of the distribution theory dates back to the first half

of the 19^{th} century, when it was (somewhat informally) used to construct solutions of ordinary differential equations. Its modern history began with the work of Sergei Sobolev and others in 1930s and early 1940s, though its status as one of the most important contributions of 20^{th} century mathematics was not established until 1951, when it was formalised in Laurent Schwartz's definitive book; *The theory of distributions*. In its essence, distribution theory proposes means of generalising and reinterpreting the classical notions of a function. As a consequence, this allows us to define fundamental notions of mathematical analysis, such as differentiability of a function or solvability of partial differential equations, in situations where the classical notions do not apply.

Many parallels can be drawn between the basic ideas of distribution theory and those of modern measure theory. The starting point is the idea that a function is not required to have a defined value at every point. Instead, all we require is the ability to extract meaningful information about the behaviour of a function in a neighbourhood of any given point. In order to ensure that the output is a real number, functions can then be reinterpreted as linear functionals that act on a select set of functions, which are called *test functions*.

Definition 1.6.1. *A* distribution T *is a continuous linear functional* $\mathsf{T}\colon \mathcal{K} \to \mathbb{R}$*, where* $\mathcal{K}$ *is a space of test functions. The action of* T *on a test function* $\varphi \in \mathcal{K}$ *will be typically denoted as* $\langle \mathsf{T}, \varphi \rangle$*, instead of the more usual* $\mathsf{T}(\varphi)$*.*

The last sentence in Definition 1.6.1 is in part motivated by the fact that, under some assumptions on $\mathcal{K}$, a function $f \in L^1(\mathbb{R}^d)$ defines a distribution through

$$\mathsf{T}_f(\varphi) = \int_{\mathbb{R}^d} f(\mathbf{x})\varphi(\mathbf{x})d\mathbf{x} = \langle f, \varphi \rangle. \tag{1.15}$$

Equation (1.15) is important for one further reason. It reflects the reality of physical measurements; one can rarely (if ever) measure the state of a physical quantity at exactly the desired point. What we instead hope to capture is the local behaviour of that physical quantity, that is the in vicinity of a desired point, where the readings might be affected by surrounding agents. What this suggests is that we should pay particular attention to the types test functions that should be admissible.

For reasons of both mathematical and physical nature we typically choose test functions that are smooth and localised (or fast decaying). Defining the set of test functions, on which these generalised functions should act, is the crucial first step. The properties of the resulting class of distributions depend delicately on how is this step conducted.

The standard choice for the set of test functions which ensures that notions such as the derivative, support and (crucially) the Fourier transform of a distribution are defined is $\mathcal{S}(\mathbb{R}^d)$, the set of Schwartz functions. There are of course many other function spaces for which these notions make sense, and in many practical situations they can be readily adopted as long as the statements remain mathematically sound. The class of *tempered distributions* is then defined as the continuous dual of the Schwartz space $\mathcal{S}(\mathbb{R}^d)$, and is denoted by $\mathcal{S}'(\mathbb{R}^d)$. Choosing a larger set of test functions results in a relatively smaller

class of admissible distributions, but the consequence is that they are in many ways *nicer* and easier to work with. We shall now list some standard properties and notions of distributions.

(i) Derivative: $\langle \partial^{\boldsymbol{\alpha}} \mathsf{T}, \varphi \rangle = (-1)^{|\boldsymbol{\alpha}|} \langle \mathsf{T}, \partial^{\boldsymbol{\alpha}} \varphi \rangle$, for $\boldsymbol{\alpha} \in \mathbb{N}_0^d$.

(ii) Support: Let $\mathcal{K}$ be the set of test functions. The support of is defined by

$$\operatorname{supp} \mathsf{T} = \{ \mathbf{x} \in \mathbb{R}^d : \langle \mathsf{T}, \varphi \rangle = 0 \text{ for all } \varphi \in \mathcal{K} \\ \text{such that } \operatorname{supp} \varphi \subset \mathcal{U}, \text{ where } \mathcal{U} \ni \mathbf{x} \text{ is open } \}^c.$$

(iii) Linear Transform: Let $\mathrm{A} : \mathbb{R}^{d \times d}$ be a regular matrix. Then

$$\langle \mathsf{T}(\mathrm{A}\cdot), \varphi \rangle = |\det \mathrm{A}|^{-1} \langle \mathsf{T}, \varphi(\mathrm{A}^\top) \rangle.$$

(iv) Fourier Transform: $\langle \hat{\mathsf{T}}, \varphi \rangle = \langle \mathsf{T}, \hat{\varphi} \rangle$

(v) Direct Product: Let $\mathcal{X} \subseteq \mathbb{R}^{d_1}$ and $\mathcal{Y} \subseteq \mathbb{R}^{d_2}$ be open. The direct product of distributions $\mathsf{T}_{\mathbf{x}}$ and $\mathsf{T}_{\mathbf{y}}$ that act on the set of test functions on $\mathcal{X}$ and $\mathcal{Y}$, respectively, is defined on the set of test functions on $\mathcal{X} \times \mathcal{Y}$ through

$$\langle \mathsf{T}_{\mathbf{x}} \times \mathsf{T}_{\mathbf{y}}, \varphi \rangle = \langle \mathsf{T}_{\mathbf{x}}, \langle \mathsf{T}_{\mathbf{y}}, \varphi \rangle \rangle.$$

The direct product is both commutative and associative .

We will now define the distributions that will be used in Chapter 4. By φ we denote the arbitrary context-dependent test function, that is, in the following we will not be particularly specific regarding the underlying set of test functions. A more detailed account of distribution theory can be found in standard texts on the topic, for example [17--19].

(i) The Dirac delta distribution, denoted by δ, is defined for test functions on $\mathbb{R}^d$ as

$$\langle \delta, \varphi \rangle = \varphi(\mathbf{0}). \tag{1.16}$$

When $d > 1$ we formally have the direct product of one dimensional distributions

$$\delta(\mathbf{x}) = \delta(\mathsf{x}_1) \cdots \delta(\mathsf{x}_d).$$

The Dirac delta is the prototypical distribution. It is the standard abstractifier of physical notions such as point mass or point charge.

(ii) Heaviside function, or the step function, $H : \mathbb{R} \to \mathbb{R}$, defined by $H(\mathsf{x}) = \chi_{\mathsf{x} \geqslant 0}$, induces a distribution through

$$\langle \mathsf{H}, \varphi \rangle = \int_{\mathbb{R}} H(\mathsf{x}) \varphi(\mathsf{x}) d\mathsf{x} = \int_0^\infty \varphi(\mathsf{x}) d\mathsf{x}. \tag{1.17}$$

The Heaviside function can also be defined as the *integral* of the Dirac delta distribution by $H' = \delta$

(iii) Cauchy Principal Value is the standard method of defining values of improper integrals. As a distribution it is defined by

$$\left\langle \mathrm{PV}\frac{1}{\mathsf{x}}, \varphi \right\rangle = \lim_{\varepsilon \to 0^+} \int_{|\mathsf{x}|>\varepsilon} \frac{\varphi(\mathsf{x})}{\mathsf{x}} d\mathsf{x}. \tag{1.18}$$

It can be shown that the Cauchy principal value is well defined for a very broad class of functions [28].

Chapter 2

Continuous Parabolic Molecules

2.1 Motivation

Recent years have seen the focus of several areas of applied harmonic analysis, image analysis, and other related fields move towards the development of representation systems that are required to address a myriad of tasks and applications. These tasks range from classical image processing topics such as denoising [6], inpainting [7], and component separation [46] to topics in an increasingly broader family of problems such as in astronomical [47] and seismical [48] imaging and in the analysis of hyperbolic differential equations [5]. In these problems sparsity is often seen as the crucial feature that characterises efficient representations, since it allows for sharp estimates and fast numerical algorithms. As we discussed in the previous chapter, when it comes to the analysis of edges of natural images or to the propagation of singularities along the characteristics of hyperbolic PDEs, this led to the development of dictionaries that are aimed at addressing the weaknesses of wavelets and other classical transforms. The two main components that aim to achieve these goals are the directional sensitivity and the anisotropic structure of the frame (or basis) functions. The necessary changes are often executed by adhering to ideas from multiscale analysis and geometry. These new paradigms can be justified through a common-sense observation: the edges of natural images are an oriented and a prevalently anisotropic phenomena. Consequently, dictionaries that respect anisotropic structures should be better adapted to address questions that pertain to natural images. As we mentioned in Section 1.4.1, this discussion can be formalised within the framework of cartoon-like images.

Hence, two choices ought to be made. The first choice is how to address the directional sensitivity, while the second choice is how to endow our dictionaries with an anisotropic architecture. Whereas there is a wide selection of answers with regards to the first of these two questions, the second question seems to have reached a certain consensus more or less from the outset; curvelets, shearlets, and other similar dictionaries, adopt the law of parabolic scaling, $width = length^2$. Coupled with directional selectivity, parabolic scaling allows for dictionaries that are sensitive to curvilinear features of images.

Parabolic scaling and its uses in modern harmonic analysis can be traced back to the

work of Fefferman on spherical summation multipliers [49]. Fefferman used parabolic scaling to decompose the underlying convolution kernel into progressively finer pieces whose supports are obtained by dividing the plane $\mathbb{R}^2$ into a family of dyadic-parabolic rectangles in polar coordinates. These methods were first used for Bochner-Riesz spherical summation operators and then extended to the study of Fourier integral operators (herein FIOs) [50].

A more recent study of FIOs and parabolic scaling, and one which is significantly more pertinent to our intentions, was conducted by Smith in [4] and [51]. Smith used parabolic scaling to decompose functions into parabolic components, thus defining function spaces that are preserved by FIOs, and analysed the behaviour of certain wave equations. Furthermore, Smith introduced a curvelet-like parabolic system of functions and defined a pseudo-distance function that corresponds to the ambient geometry, which greatly influenced analogous notions not only in our work, but also in the study of curvelet [5, 44] and shearlet molecules [8].

A natural question to ask ourselves is whether there is anything truly exceptional about the parabolic scaling law, or would other anisotropic scaling laws be equally viable. A diligent examination of the literature would suggest that for a number of interesting questions this is not the case and that other options are admissible, though the evidence nevertheless seems to indicate that the parabolic scaling provides the optimal compromise, see for example the discussion in [52] or in [44]. These questions motivated the study of α-molecules, which we mentioned in Section 1.5. We will revisit some of these ideas in Chapter 4.

The goal of this chapter is to describe and discuss the basic properties of a general framework for representation systems that obey the law of parabolic scaling. As we mentioned in Section 1.5, our work builds on the machinery of discrete parabolic molecules [9], and adapts it to the continuous setting. To that end, we will introduce the notion of *continuous parabolic molecules* (herein CPMs), which will encompass a wide array of continuous parabolic dictionaries that share the same fundamental properties. We will principally focus on two families of transforms, curvelets and shearlets, since they are two of the most widely used existing parabolic families and there is a significant body of evidence in support of their properties and applicability. Our construction is directly related to, and as we will show, is a generalisation of, earlier constructions such as curvelet molecules [5] and shearlet molecules [8], which were used to describe optimally sparse representations of wave propagators. It is important to add that what inspired both the curvelet and the shearlet molecules is the notion of *vaguelettes* which is used in wavelet analysis in the study of Calderón-Zygmund operators [53].

There are two main motives that initiated the study of abstract parabolic dictionaries. Firstly, having a generalised framework aims to give a glimpse into the fundamental features that tie the parabolic representation systems together. Therefore, we hope to not only eliminate the redundancy of the proof techniques, but also use the abstractified framework to determine the universal properties and develop methods that allow for an easy transfer of conclusions about those properties from one family of CPMs to another. The second motivation is historically more significant. In 1957 Peter Lax [54] investigated

solution operators of linear and symmetric hyperbolic systems, namely, he developed *parametrices* that approximate the solution operators. Parametrices take the form of FIOs, which are operators of the form

$$\int e^{\imath\Psi(\mathsf{x},\xi)}\sigma(\mathsf{x},\xi)\hat{f}(\xi)d\xi, \tag{2.1}$$

under some conditions on the phase $\Psi(\mathsf{x},\xi)$ and the amplitude $\sigma(\mathsf{x},\xi)$. Lax conjectured that a superposition of integrals of the form (2.1) with certain amplitudes gives a good approximation of the solution of an initial value variable coefficient hyperbolic system. This insight stimulated the study of FIOs. Now, the fact of the matter is that certain classes of operators do not map the elements of a given dictionary back into that same dictionary. For example, FIOs transform curvelets not into curvelets themselves but rather into curvelet-like constructs. Even though they are not curvelets, these functions have very similar time-frequency localisation properties. Therefore, it is useful to detect and quantify exactly what are the essential properties that underlie the curvelet construction and formulate these through an abstract framework. Working with an abstract framework would ensure that the FIOs map curvelet molecules into curvelet molecules, which allows to further the study of their behaviour and applicability for the sparse representation of FIOs, and consequently of wave propagators as well. We will already see glimpses of this type of functionality in the proof of Theorem 2.4.1 where we will work with a differential operator that maps products of continuous parabolic molecules into a linear superposition of parabolic molecules of a decreased smoothness.

We will begin with Section 2.2 where we will define continuous parabolic molecules and other notions relevant for their study. In Section 2.3 we will introduce curvelet and shearlet molecules and show that both satisfy the conditions of continuous parabolic molecules (Theorem 2.3.2 and Theorem 2.3.7). Furthermore, we will show that the continuous families of both the compactly supported and band-limited shearlets, latter which is a somewhat novel construction that also includes a dual family which is also a family of CPMs, are elements of our framework (Proposition 2.3.10 and Theorem 2.3.13). We will conclude in Section 2.4, where we will establish the fact that the Gramians of continuous parabolic molecules, $\langle m_\lambda, n_\nu\rangle$ have a strong off-diagonal decay. This property is typically referred to as the *almost-orthogonality*, and it will be the essential tool for the analysis in Chapter 3.

2.2 Definitions and Elementary Notions

Continuous parabolic molecules (herein CPMs) are, roughly speaking, families of oriented $L^2(\mathbb{R}^2)$ functions that obey the law of parabolic scaling and are well localised in both time and frequency, which is given through the conditions on their smoothness and decay. The precise notion of time-frequency localisation that we will adhere to will be formalised shortly in Definition 2.2.2, and it ensures fast decay of each molecule once we get away from its effective support which is determined by the scale, orientation and location parameters. As we discussed in Section 1.5, out construction builds on the notion of

discrete parabolic molecules [9], and we will briefly examine the connection between discrete parabolic molecules and CPMs in Section 2.2.1.

Let us start by setting up the notation and introducing a couple of definitions. We define the *parameter space* $\mathbb{P}$ as

$$\mathbb{P} = \mathbb{R}^+ \times [0, 2\pi) \times \mathbb{R}^2,$$

where a point $(a, \theta, \mathbf{b}) \in \mathbb{P}$ describes a scale a, an orientation θ, and a location $\mathbf{b}$.

Let $\mathrm{R}_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$ be a rotation matrix associated with an angle θ, and let $\mathrm{D}_a = \mathrm{diag}(a, a^{1/2})$ be the parabolic scaling matrix, with a scaling parameter $a \in \mathbb{R}^+$. Since our goal is to create a framework that is broad enough to encompass a number of different dictionaries that adhere to the law of parabolic scaling, we need to allow for different geometries. This observation is a consequence of the fact that each of the existing parabolic dictionaries (such as curvelets and shearlets) hinges on a unique implementation of the directional selectivity and the parabolic scaling in its construction. Therefore, the first order of business will be to create a tool that allows us to tell when are elements belonging to different dictionaries qualitatively similar in a certain sense. As we will see in Theorem 2.4.1, this correspondence relies on the relationship between their respective scale, orientation and location parameters. In order to be able to quantify the dependence of this notion of *similarity* on the inner workings of dictionaries, we will relate the native geometry of each CPM family to the parameter space $\mathbb{P}$. This is achieved through *parameterisations*, which are, loosely speaking, subsets of the parameter space $\mathbb{P}$.

Definition 2.2.1. *A* parametrisation *is a pair* (Λ, Φ) *where* Λ *is an index set and* Φ *is a mapping* $\Phi\colon \Lambda \to \mathbb{P}$. *A parametrisation family is a family of parameterisations* $(\Lambda_i, \Phi_i)_{i \in \mathcal{I}}$, *where* $\mathcal{I}$ *is a finite index set.*

The role parameterisations play is now clear. The set Λ is defined by the native geometry of the CPM family and it serves to index its members, while Φ then associates that index with a specific scale, orientation, location triplet in $\mathbb{P}$. Definition 2.2.1 indeed goes a bit further by allowing a finite family of parameterisations. The idea is that a given CPM dictionary might have been constructed using several disjoint pieces, for example each piece might be constructed to deal with a different part of the frequency domain. One example of this type of a dictionary are the cone-adapted shearlets (Section 1.4.4), where the frequency domain is split into high-frequency cones (and a low-frequency box) and the shearlets are then defined separately on the vertical and on the horizontal cones. Furthermore, we ought to have a special parametrisation associated with low frequency regimes, since they are always defined according to a different set of rules. In doing so, we allow each parametrisation function Φ_i to have desirable properties, instead of dealing with a single and complicated parametrisation. Having said that, these considerations only serve to simplify some of the proofs and in the rest of this thesis we will only be concerned with the case $|\mathcal{I}| = 1$. This is due to the fact that all the questions we hope to address are to do with estimates on norms and Gramians. Hence, provided the index set

$\mathcal{I}$ is finite, the general case follows easily from the case $|\mathcal{I}| = 1$ by using for example, the triangle inequality and other similar estimates. A larger $\mathcal{I}$ would only serve to further complicate the notation, make the proofs lengthier, and would not require any conceptual changes to the exposition or the proofs.

We are now ready to define continuous parabolic molecules.

Definition 2.2.2. *Let (Λ, Φ) be a parametrisation. We say that family of functions $\Gamma = \{m_\lambda : \lambda \in \Lambda\}$ is a family of* continuous parabolic molecules *of order (R, M, N_1, N_2) if it can be written as*

$$m_\lambda(\mathbf{x}) = a_\lambda^{-3/4} \varphi^{(\lambda)} \left(\mathrm{D}_{1/a_\lambda} \mathrm{R}_{\theta_\lambda} (\mathbf{x} - \mathbf{b}_\lambda) \right), \tag{2.2}$$

where $(a_\lambda, \theta_\lambda, \mathbf{b}_\lambda) = \Phi(\lambda) \in \mathbb{P}$, and $\varphi^{(\lambda)}$ satisfies

$$|\partial^{\boldsymbol{\beta}} \hat{\varphi}^{(\lambda)}(\boldsymbol{\xi})| \lesssim \min\left(1, a_\lambda + |\xi_1| + a_\lambda^{1/2} |\xi_2|\right)^M \langle \|\boldsymbol{\xi}\| \rangle^{-N_1} \langle \xi_2 \rangle^{-N_2} \tag{2.3}$$

for all multi-indices $\boldsymbol{\beta} \in \mathbb{N}_0^2$, such that $|\boldsymbol{\beta}| \leqslant R$. The implicit constants can be chosen independently of λ so that the inequalities hold uniformly over all λ.

Definition 2.2.2 implies a number of useful estimates. First of all, a direct computation yields

$$|\hat{m}_\lambda(\boldsymbol{\xi})| \lesssim a_\lambda^{1/2} \min(1, a_\lambda(1 + \|\boldsymbol{\xi}\|))^M \langle a_\lambda \|\boldsymbol{\xi}\| \rangle^{-N_1} \langle a^{1/2} (\mathrm{R}_{\theta_\lambda} \boldsymbol{\xi})_2 \rangle^{-N_2}.$$

Similar estimates hold for the derivatives of $\hat{m}_\lambda$. Therefore, Definition 2.2.2 implies a (somewhat biased) directional decay as the coordinates tend to infinity, M almost vanishing moments , and frequency localisation of $\hat{m}_\lambda$ with respect to the dilation parameter a_λ. Furthermore, $R \in \mathbb{N}$ describes the spatial localisation of the molecule m_λ while N_1 and N_2 correspond to its smoothness.

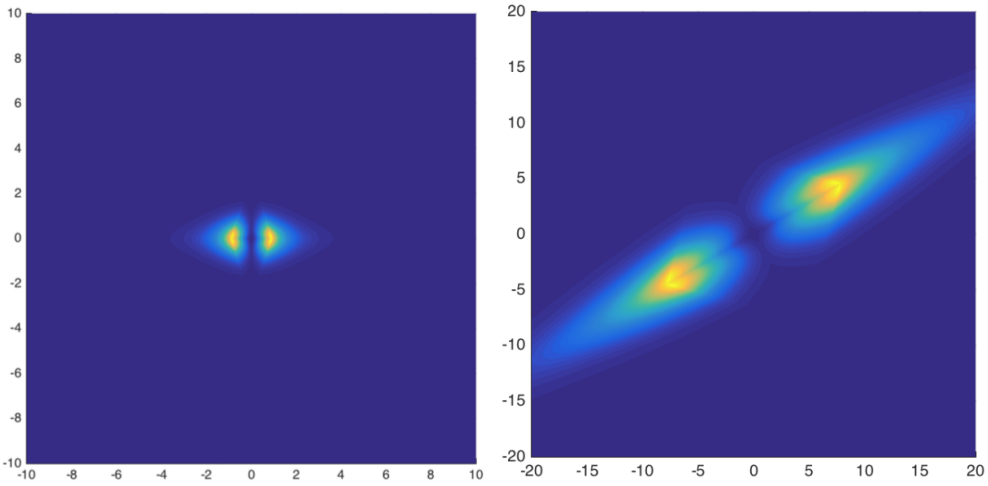
In order to measure the similarity of two CPM families we need a notion of distance that takes into account the specifics of the ambient geometry. This is determined by the relation between their dilation, orientation and location parameters. To that end we will now introduce a pseudo-distance function that decays as the indices λ and ν get farther apart.

Definition 2.2.3. *The* pseudo-distance function $\omega \colon \mathbb{P} \times \mathbb{P} \to \mathbb{R}$ *is for a pair of triplets $\lambda = (a_\lambda, \theta_\lambda, \mathbf{b}_\lambda),\ \nu = (a_\nu, \theta_\nu, \mathbf{b}_\nu) \in \mathbb{P}$ defined by*

$$\omega(\lambda, \nu) = \frac{a_M}{a_m} \left(1 + a_M^{-1} d(\lambda, \nu)\right), \tag{2.4}$$

where

$$\begin{aligned} a_m &= \min(a_\lambda, a_\nu), \\ a_M &= \max(a_\lambda, a_\nu), \\ d(\lambda, \nu) &= |\theta_\lambda - \theta_\nu|^2 + \|\mathbf{b}_\lambda - \mathbf{b}_\nu\|^2 + |\langle \mathbf{e}_\lambda, \mathbf{b}_\lambda - \mathbf{b}_\nu \rangle|, \\ \mathbf{e}_\lambda &= (\cos(\theta_\lambda), \sin(\theta_\lambda))^\top. \end{aligned}$$



(a) Support of $\hat{\varphi}^{(\lambda)}$ (b) Support of $\hat{m}_\lambda$

Figure 2.1: Examples of frequency supports given by conditions (2.1)

The quotient a_M/a_m suggests the multi-scale nature of CPM dictionaries, whereas $d(\lambda, \nu)$ accounts for the difference in orientations and locations. Therefore, Definition 2.2.3 implies that $\omega(\lambda, \nu)$ will be small only if the parameters λ and ν are associated with roughly the same scale and orientation, and if they are at nearby locations. We should note that (real-valued) curvelets and shearlets are associated with a ray. Therefore, angle differences $\theta_\lambda - \theta_\nu$ are understood modulo π. It should be clear from the definition that the function ω is not a proper distance function, though it still satisfies a number of useful properties that we will add for the sake of completeness. A more exhaustive list of properties of ω, and their proofs, can be found in [5] and [4].

Proposition 2.2.4. *The following properties hold for the pseudo-distance function ω.*

(i) *Almost Symmetry:* $\omega(\lambda, \nu) \sim \omega(\nu, \lambda)$.

(ii) *Almost Triangle Inequality:* $d(\lambda, \nu) \leqslant C(d(\lambda, \kappa) + d(\kappa, \nu))$, *for a uniform constant* $C > 0$.

(iii) *Composition: For every positive integer N and some $C_N > 0$ we have*

$$\sum_{\kappa} \omega(\lambda, \kappa)^{-N} \omega(\kappa, \nu)^{-N} \leqslant C_N \omega(\lambda, \nu)^{-N-1}.$$

When applying the pseudo-distance ω to parameterisations of some two families of continuous parabolic molecules, we will write $\omega(\lambda, \nu)$, for $\lambda \in \Lambda_1, \nu \in \Lambda_2$ when, if written properly, it should read $\omega(\Phi_1(\lambda), \Phi_2(\nu))$. This implicit notation is intended to make the notation a little less cumbersome, hopefully without jeopardising the clarity of exposition.

The fundamental property of continuous parabolic molecules, which we will establish in Section 2.4, is the notion that two families of continuous parabolic molecules are

almost orthogonal, in the sense that their inner product exhibits a strong off-diagonal decay. This decay will be given in terms of controlling their inter-Gramian with the decay of the pseudo-distance function ω between their respective indices. An elementary interpretation of this result is simple: two molecules m_λ and n_ν are essentially the same if their indices λ and ν are not too far apart, that is, if they are associated to a similar scale, orientation and location. We should note that it is still an open problem whether or not there exists an orthobasis of curvelets. Hence, we cannot say that curvelets (or shearlets) are orthogonal to each other, but what one can show is that they are almost orthogonal.

We will now introduce the notion of admissibility of a parametrisation. First we need to define the canonical parametrisation.

Definition 2.2.5. *The parametrisation pair* (Λ_0, Φ)*, where* Φ *is the identity map and the index set is* $\Lambda_0 = [0,1] \times [0, 2\pi) \times \mathbb{R}^2$*, is called the* canonical parametrisation.

The following notion of admissibility will be essential in the upcoming proofs.

Definition 2.2.6. *Index set* Λ *is said to be* k-admissible *if*

$$\sup_{\lambda \in \Lambda} \int_{\Lambda_0} \omega(\lambda, \nu)^{-k} d\mu(\nu) < \infty \quad \textit{and} \quad \sup_{\lambda \in \Lambda_0} \int_{\Lambda} \omega(\lambda, \nu)^{-k} d\mu(\lambda) < \infty \tag{2.5}$$

where $d\mu(\lambda) = \frac{d\lambda}{a^3}$.

The reference measure $d\mu(\lambda) = \frac{d\lambda}{a^3}$ is characteristic for parabolic dictionaries and will reappear in the next section when we will run through standard examples of CPMs. This is a left-invariant Haar measure that can be justified within the theory of unitary representations of affine groups [55]. It is perhaps better to think about $d\mu$ in more practical terms, as suggested in [1], by writing

$$d\mu(\lambda) = \frac{da}{a} \frac{d\theta}{a^{1/2}} \frac{d\mathbf{b}}{a^{3/2}}.$$

This goes hand in hand with the intuition about the underlying decomposition of the frequency domain: the range of $\mathbf{b}$, is divided into unit blocks of size a by $\sqrt{a}$, whereas the range of directions θ is subdivided into blocks of length a, and the space of dilations is subdivided into blocks of size a.

2.2.1 Revisiting Discrete Parabolic Molecules

As we already mentioned, the bulk of this section has been motivated and then adjusted to our needs from previous work, most importantly from the study of discrete parabolic molecules in [9]. For example, a pseudo-distance function of this type first appeared in [4], and was recently brought back to the arsenal of applied harmonic analysis in [5]. The primary adjustment that is required for Definition 2.2.3 in the context of our framework

is to take into account the continuous setting. Let us take a closer look at these changes by considering the pseudo-distance function, which was defined in [5] and [9] by

$$\omega(\lambda,\nu) = 2^{|j_\lambda - j_\nu|}\left(1 + 2^{j_0} d(\lambda,\nu)\right), \tag{2.6}$$

where $d(\lambda,\nu)$ is the same as in Definition 2.4 and $j_0 = \min(j_\lambda, j_\nu)$. Instead of working with dilations from a continuous spectrum $a \in [0,1]$, equation (2.6) adopts the dyadic sampling $a_j = 2^j$, which reflects the construction of discrete curvelet and shearlet frames.

There are many interesting questions, which are motivated by the studies conducted in [25], that correlate the work on discrete parabolic molecules to our work on CPMs. These are the questions such as; what can we infer about the properties of a discrete parabolic frame which was sampled from a continuous one? One question worth investigating is for example the study of the conditions on the sampling density from which it would be possible to assert that the dual frame of sampled family of discrete parabolic molecules is also a family of discrete parabolic molecules. If that were the case, then the dual frame would also satisfy the almost-orthogonality conditions which by the preceding discussion bears implications on its approximation properties, and has important implications regarding compression of various classes of operators [56--58]. This is sometimes called the *intrinsic localisation* of frames, or the *intrinsic localisation of an operator* with respect to a frame.

2.3 Examples of CPMs

Now that we have laid down the foundations of our framework the next goal is to study the directional dictionaries it contains. We will show that CPMs encompass both the curvelet and shearlet-type constructions and that it is a generalisation of both the curvelet and shearlet molecules. In the remainder of this section we will briefly reintroduce curvelets, define curvelet molecules, and show that curvelet molecules form a family of continuous parabolic molecules. Then we will repeat the same for shearlets. We will also give a description of specific families of band-limited and the compactly supported shearlets that admit some further properties and will be valuable in Chapter 3.

2.3.1 Curvelets and Curvelet Molecules

Let us begin with a short description of second-generation curvelets that were introduced in Section 1.4.3. Denote by r and α the polar coordinates in the frequency domain. Take a pair of smooth, non-negative and real-valued windows $W(r)$ and $V(\omega)$, which are called the *radial window* and the *angular window*, respectively. Furthermore, assume that W takes positive real arguments, and is supported on $(1/2, 2)$, while V takes real arguments and is supported on $[-1, 1]$. These windows must also satisfy the following admissibility conditions

$$\int_0^\infty W(ar)^2 \frac{da}{a} = 1, \ \forall r > 0,$$

$$\int_{-1}^{1} V(u)^2 du = 1.$$

At scale a the generating element $\gamma_{a\,0\,\mathbf{0}}$ is defined via polar coordinates in the frequency domain as

$$\hat{\gamma}_{a\,0\,\mathbf{0}}(r,\alpha) = a^{3/4} W(ar) V(\alpha/\sqrt{a}).$$

The dilation parameter a has to satisfy $a \leqslant a_0$, where a_0 represents the coarsest scale and must obey $a_0 \leqslant \pi^2$ for the construction to work, but we will take $a_0 = 1$. The curvelets at scale a are defined via rotations and translations of the generating element $\gamma_{a\,0\,\mathbf{0}}$ by

$$\gamma_\lambda(\mathbf{x}) = \gamma_{a_\lambda\,0\,\mathbf{0}}\left(\mathrm{R}_{\theta_\lambda}(\mathbf{x} - \mathbf{b}_\lambda)\right), \text{ where } \quad \lambda = (a_\lambda, \theta_\lambda, \mathbf{b}_\lambda) \in \Lambda_0 := [0, a_0] \times [0, 2\pi) \times \mathbb{R}^2.$$

The family $\Gamma = \{\gamma_\lambda : \lambda \in \Lambda_0\}$ is called the family of *second generation curvelets*.

Let us now define continuous curvelet molecules.

Definition 2.3.1. *Consider the index set $\Lambda_0 = [0, a_0] \times [0, 2\pi) \times \mathbb{R}^2$. A family of functions $\{m_\lambda : \lambda \in \Lambda_0\}$ is called a* family of curvelet molecules *of regularity R if it can be expressed as*

$$m_\lambda(\mathbf{x}) = a_\lambda^{-3/4} \varphi^{(\lambda)}\left(\mathrm{D}_{1/a_\lambda} \mathrm{R}_{\theta_\lambda}(\mathbf{x} - \mathbf{b}_\lambda)\right)$$

such that

$$|\partial^{\boldsymbol{\beta}} \varphi^{(\lambda)}(\mathbf{x})| \lesssim \langle \|\mathbf{x}\| \rangle^{-N}, \tag{2.7}$$

and

$$|\hat{\varphi}^{(\lambda)}(\boldsymbol{\xi})| \lesssim \min\left(1, a_\lambda + |\xi_1| + a_\lambda^{1/2} |\xi_2|\right)^M \tag{2.8}$$

hold for $|\boldsymbol{\beta}| \leqslant R$, $N = 0, 1, 2, \ldots$, and all $M = 0, 1, \ldots, R$. The implicit constants are uniform over $\lambda \in \Lambda_0$.

It should come as no surprise that second generation curvelets constitute a family of curvelet molecules for an arbitrary degree of regularity R, which has been established in [15]. In order to show that the same holds for CPMs, we shall begin by showing that every family of curvelet molecules is also a family of CPMs.

Proposition 2.3.2. *A system of curvelet molecules of regularity $3R/2$ constitutes a system of continuous parabolic molecules of order $(R, R, R/2, R/2)$.*

Proof. Curvelet molecules are by definition of the form (2.2). Hence, we only need to is establish that the decay estimates (2.3) hold. Regarding the vanishing moments we refer to the discussion on page 14 in [5] (similar claims can also be found in [8]), where it is stated that the definition of curvelet molecules implies that (2.8) also holds for the derivatives of $\varphi^{(\lambda)}$. On the other hand, since

$$|\partial^{\boldsymbol{\beta}} \varphi^{(\lambda)}(\mathbf{x})| \leqslant C_N \langle \|\mathbf{x}\| \rangle^{-N}$$

holds for all $N \in \mathbb{N}$ and for all multi-indices $\boldsymbol{\beta} \in \mathbb{N}_0^2$, such that $|\boldsymbol{\beta}| \leqslant R$, it follows

$$\mathbf{x}^{\boldsymbol{\alpha}} \partial^{\boldsymbol{\beta}} \varphi^{(\lambda)}(\mathbf{x}) \in L^1(\mathbb{R}^2),$$

where $\boldsymbol{\alpha} \in \mathbb{N}_0^2$ is an arbitrary multi-index. Thus, for $|\boldsymbol{\beta}| \leqslant R$ we have

$$\xi_1^{R/2} \xi_2^R \partial^{\boldsymbol{\beta}} \hat{\varphi}^{(\lambda)}(\boldsymbol{\xi}) = \Big(\partial^{(R/2,R)} (\mathbf{x}^{\boldsymbol{\beta}} \varphi^{(\lambda)}(\mathbf{x})) \Big)\widehat{\ }(\boldsymbol{\xi}) \in L^\infty(\mathbb{R}^2),$$

which gives

$$|\xi_1^{R/2} \xi_2^R \partial^{\boldsymbol{\beta}} \hat{\varphi}^{(\lambda)}(\boldsymbol{\xi})| \lesssim 1.$$

It follows

$$\langle \|\boldsymbol{\xi}\| \rangle^{R/2} \langle \xi_2 \rangle^{R/2} |\partial^{\boldsymbol{\beta}} \hat{\varphi}^{(\lambda)}(\boldsymbol{\xi})| \lesssim 1,$$

giving the desired high-frequency localisation. □

In order to complete the description of how do the curvelet molecules fit into the CPM framework we need do define the corresponding parametrisation. The canonical parametrisation, which was defined in Definition 2.2.5, has been constructed with precisely curvelets in mind. In other words, the curvelet parametrisation is the canonical parametrisation (Λ_0, Φ), where Φ is the identity map. As we mentioned earlier, admissibility of parameterisations will play a crucial role later on. We will now show that the curvelet parametrisation is admissible for all $k > 2$.

Lemma 2.3.3. *Canonical parametrisation (Λ_0, Φ) is k-admissible for all $k > 2$.*

Proof. We want to show that

$$\sup_{\nu \in \Lambda_0} \int_{\Lambda_0} \omega(\lambda, \nu)^{-k} d\mu(\lambda) < \infty$$

holds for all $k > 2$. Considering an arbitrary $\nu \in \Lambda_0$ we have

$$\int_{\Lambda_0} \omega(\lambda, \nu)^{-k} d\mu(\lambda) = \int_{[0,a_0]} \frac{a_m^k}{a_M^k} \int_{[0,2\pi)\times\mathbb{R}^2} \big(1 + a_M^{-1} d(\lambda, \nu)\big)^{-k} d\mu(\lambda).$$

It can be shown (Lemma A.1.1 in the Appendix) that

$$\int_{[0,2\pi)\times\mathbb{R}^2} \big(1 + q^{-1} d(\lambda, \nu)\big)^{-k} d\theta d\mathbf{b} \lesssim q^2 \tag{2.9}$$

holds for all $q \in \mathbb{R}^+$. Hence, it follows

$$\begin{aligned} \int_{\Lambda_0} \omega(\lambda, \nu)^{-k} d\mu(\lambda) &\lesssim \int_0^{a_0} \frac{a_m^k}{a_M^k} a_M^2 \frac{da_\lambda}{a_\lambda^3} = a_\nu^{-k+2} \int_0^{a_\nu} a_\lambda^{k-3} da + {a_\nu}^k \int_{a_\nu}^{a_0} a_\lambda^{-k-1} da_\lambda \\ &\lesssim \frac{1}{k-2} + \frac{1}{k} - \frac{1}{k} {a_\nu}^k a_0^{-k} \leqslant \frac{1}{k-2} + \frac{1}{k} < \infty \end{aligned}$$

which is true as long as $k > 2$. In other words, Λ_0 is k-admissible for all $k > 2$. □

We are finally prepared to show that second generation curvelets constitute a family of continuous parabolic molecules.

Proposition 2.3.4. *Second generation curvelets are a family of continuous parabolic molecules of order $(R, R, R/2, R/2)$, for an arbitrary $R \in \mathbb{N}$, and their parameterisation is admissible for all $k > 2$.*

Proof. By Section 2.3 in [5], second generation curvelets constitute a family of curvelet molecules of an arbitrary degree of regularity R. Hence, the statement follows by Proposition 2.3.2 and by Lemma 2.3.3. □

2.3.2 Shearlets and Shearlet Molecules

Let us first briefly reintroduce the basics of the shearlet transform. As we have seen, second generation curvelets cover the spectrum of all orientations by rotating the supports of the generating elements. Shearlets, on the other hand, handle orientations by using the shearing operator, which is given through the shearing matrix $\mathrm{S}_s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$, for $s \in \mathbb{R}$. The fundamental problem with using rotations is that they destroy the integer lattice $\mathbb{Z}^2$, unless the rotation angle is $k\pi/2$ for $k \in \mathbb{Z}$. Therefore, we must pay particular attention to how will the discrete curvelet frames be constructed. Shears on the other hand leave the lattice $\mathbb{Z}^2$ invariant for a much wider range of inputs which allows for a unified treatment of continuous and discrete settings. Classical shearlets are then defined as an affine transformation of a single mother shearlet.

Definition 2.3.5. *Consider a function $\psi \in L^2(\mathbb{R}^2)$ that satisfies the admissibility condition*

$$\int_{\mathbb{R}^2} \frac{|\hat{\psi}(\boldsymbol{\xi})|^2}{\xi_1^2} d\boldsymbol{\xi} < \infty.$$

The continuous shearlet family *is the family of functions*

$$\psi_{a\,s\,\mathbf{b}}(\mathbf{x}) = a^{-3/4}\psi(\mathrm{D}_{1/a}\mathrm{S}_s^{-1}(\mathbf{x} - \mathbf{b})),$$

where $a \in \mathbb{R}^+$, $s \in \mathfrak{S} \subseteq \mathbb{R}$ and $\mathbf{b} \in \mathbb{R}^2$.

For the construction to work the set $\mathfrak{S}$ should be a (symmetric) subset of $\mathbb{R}$ that contains $[-1, 1]$. As we discussed in Chapter 1, if we were to take $\mathfrak{S}$ to be the whole $\mathbb{R}$ this would result in serious limitations for practical applications since a singularity that lies on the y-axis can only be resolved by considering the shearing parameter in the limit $s \to \infty$.

To overcome the shortcomings of the construction of classical shearlets we can use the cone-adapted shearlets, as we discussed in the Section 1.4.4. Cone-adapted shearlets work by splitting the frequency domain into four high frequency cones and a low frequency box, as in Figure 1.5. The horizontal shearlets are then be defined by restricting $\mathfrak{S}$ to a finite set, say $\mathfrak{S} = [-\Xi, \Xi]$ with $1 \leqslant \Xi < \infty$. The shearlets that live on the vertical cones are then obtained by simply swapping the roles of x_1 and x_2. Further details regarding the construction of shearlets and related topics can be found in [40].

We are now ready to define continuous shearlet molecules.

Definition 2.3.6. *Define the dilation matrices* $\mathrm{D}_a^0 = \mathrm{diag}(a, a^{1/2})$, $\mathrm{D}_a^1 = \mathrm{diag}(a^{1/2}, a)$, *and the shearing matrices* $\mathrm{S}_s^0 = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$ *and* $\mathrm{S}_s^1 = \left(\mathrm{S}_s^0\right)^\top$. *Furthermore, consider the index set*

$$\Lambda = \left\{(\epsilon, a, s, \mathbf{b}) : \epsilon \in \{0,1\},\ a \in [0, a_0],\ s \in \mathfrak{S},\ \mathbf{b} \in \mathbb{R}^2\right\},$$

where $\mathfrak{S} = [-\Xi, \Xi]$, *for* $1 \leqslant \Xi < \infty$. *A family of functions* $\{\sigma_\lambda : \lambda \in \Lambda\}$ *is called a family of* continuous shearlet molecules *if*

$$\sigma_\lambda(\mathbf{x}) = a_\lambda^{-3/4} \psi^{(\lambda)}\left(\mathrm{D}_{1/a_\lambda}^\epsilon \mathrm{S}_{s_\lambda}^\epsilon (\mathbf{x} - \mathbf{b}_\lambda)\right),\ \textit{for } \lambda = (\epsilon, a_\lambda, s_\lambda, \mathbf{b}_\lambda).$$

Furthermore, we say that $\{\sigma_\lambda : \lambda \in \Lambda\}$ *is of order* (R, M, N_1, N_2) *provided the functions* $\psi^{(\lambda)}$ *satisfy*

$$|\partial^{\boldsymbol{\beta}} \hat{\psi}^{(\epsilon, a_\lambda, \theta_\lambda, \mathbf{b}_\lambda)}(\boldsymbol{\xi})| \lesssim \min(1, a_\lambda + |\xi_{1+\epsilon}| + a_\lambda^{1/2}|\xi_{2-\epsilon}|)^M \langle \|\boldsymbol{\xi}\| \rangle^{-N_1} \langle \xi_{2-\epsilon} \rangle^{-N_2} \tag{2.10}$$

for all multi-indices $\boldsymbol{\beta} \in \mathbb{N}_0^2$ *such that* $|\boldsymbol{\beta}| \leqslant R$.

A related notion of discrete shearlet molecules was introduced in [8] as a shearlet-based analogue of curvelet molecules. Discrete shearlet molecules were intended to address the same issues of efficient representations of FIOs. We will now show that continuous shearlet molecules fit into the CPM framework.

Proposition 2.3.7. *Assume that* $\Sigma = \{\sigma_\lambda : \lambda \in \Lambda\}$ *is a family of continuous shearlet molecules of order* (R, M, N_1, N_2). *The family* Σ *is then also a family of continuous parabolic molecules of the same order, with the parametrisation* $(\Lambda_{\epsilon,\Sigma}, \Psi_{\epsilon,\Sigma})_{\epsilon \in \{0,1\}}$ *defined by*

$$\begin{aligned} \Lambda_{0,\Sigma} = \Lambda_{1,\Sigma} &= [0, a_0] \times \mathfrak{S} \times \mathbb{R}^2, \\ \Psi_{\epsilon,\Sigma}(a, s, \mathbf{b}) &= \left(a, \epsilon\frac{\pi}{2} + \arctan(-s), \mathbf{b}\right), \end{aligned}$$

for $\epsilon = 0, 1$.

Proof. We will restrict our discussion to the case $\epsilon = 0$. The other case is entirely analogous. Let us recall that CPMs have to be written in the form

$$m_\lambda(\mathbf{x}) = a_\lambda^{-3/4} \varphi^{(\lambda)}(\mathrm{D}_{1/a_\lambda} \mathrm{R}_{\theta_\lambda}(\mathbf{x} - \mathbf{b}_\lambda)).$$

Without loss of generality let us take $\mathbf{b}_\lambda = \mathbf{0}$. Therefore, omitting the index λ, we write

$$\varphi(\mathbf{x}) = \psi(\mathrm{D}_{1/a} \mathrm{S}_s \mathrm{R}_\theta^{-1} \mathrm{D}_a \mathbf{x}).$$

The Fourier transform of φ is given by

$$\hat{\varphi}(\boldsymbol{\xi}) = \hat{\psi}(\mathrm{D}_a \mathrm{S}_s^{-\top} \mathrm{R}_{-\theta} \mathrm{D}_{1/a} \boldsymbol{\xi}).$$

Denote $\mathrm{A} = \mathrm{D}_a \mathrm{S}_s^{-\top} \mathrm{R}_{-\theta} \mathrm{D}_{1/a}$. Since $\theta = \arctan(-s)$, we have

$$\mathrm{A} = \begin{pmatrix} \tau_1(-s) & a^{1/2}\sin(\arctan(-s)) \\ 0 & \tau_2(-s) \end{pmatrix},$$

where $\tau_1(t) = \cos(\arctan(t))$ and $\tau_2(t) = t\sin(\arctan(t)) + \cos(\arctan(t))$. Using the trigonometric identities we can rewrite τ_1 and τ_2 as $\tau_1(t) = (t^2+1)^{-1/2}$ and $\tau_2(t) = (t^2+1)^{1/2}$. Therefore, since $\mathfrak{S}$ is a bounded subset of $\mathbb{R}$ it follows

$$c_1 \leqslant \tau_1(t) \leqslant C_1, \quad \text{and} \quad c_2 \leqslant \tau_2(t) \leqslant C_2, \tag{2.11}$$

where the constants c_1, c_2, C_1 and C_2 depend only on $\mathfrak{S}$.

In order to bound the derivatives of $\hat{\varphi}$, we will now use the assumptions regarding the decay of shearlet molecules (2.10). For a multi-index $\boldsymbol{\beta} \in \mathbb{N}_0^2$ the chain rule gives

$$\begin{aligned} |\partial^{\boldsymbol{\beta}} \hat{\varphi}(\boldsymbol{\xi})| &\lesssim \sup_{|\boldsymbol{\gamma}| \leqslant |\boldsymbol{\beta}|} |\partial^{\boldsymbol{\gamma}} \hat{\psi}\,(\mathrm{A}\boldsymbol{\xi})\,| \\ &\lesssim \min(1, a_\lambda + |(\mathrm{A}\boldsymbol{\xi})_1| + a_\lambda^{1/2}|(\mathrm{A}\boldsymbol{\xi})_2|)^M \langle \|\mathrm{A}\boldsymbol{\xi}\| \rangle^{-N_1} \langle (\mathrm{A}\boldsymbol{\xi})_2 \rangle^{-N_2}. \end{aligned}$$

Therefore, we have to estimate the terms in the previous equation to ensure that $\hat{\varphi}$ obeys the decay conditions (2.3). Since $\|\boldsymbol{\xi}\| \leqslant \|\mathrm{A}^{-1}\| \|\mathrm{A}\boldsymbol{\xi}\|$ it follows

$$\langle \|\mathrm{A}\boldsymbol{\xi}\| \rangle^{-N_1} \leqslant (\min(1, \|\mathrm{A}^{-1}\|^{-1})^{-N_1} \langle \|\boldsymbol{\xi}\| \rangle^{-N_1} \lesssim \langle \|\boldsymbol{\xi}\| \rangle^{-N_1},$$

where the matrix norm of A is bounded due to (2.11). A similar argument yields the bounds for the remaining two terms. In conclusion, putting all the estimates together, we have

$$|\partial^{\boldsymbol{\beta}} \hat{\varphi}(\boldsymbol{\xi})| \lesssim \min(1, a_\lambda + |\xi_1| + a_\lambda^{1/2}|\xi_2|)^M \langle \|\boldsymbol{\xi}\| \rangle^{-N_1} \langle \xi_2 \rangle^{-N_2},$$

as desired. □

The last step is to establish the k-admissibility of the shearlet parametrisation.

Proposition 2.3.8. *The set $\Lambda_\Sigma = [0, a_0] \times \mathfrak{S} \times \mathbb{R}^2$, where $\mathfrak{S} = [-\Xi, \Xi]$ and $1 \leqslant \Xi < \infty$, is k-admissible for all $k > 2$.*

Proof. We only need to show

$$\sup_{\nu \in \Lambda_0} \int_{\Lambda_\Sigma} \omega(\lambda, \nu)^{-k} d\mu(\lambda) < \infty,$$

since the other statement required by Definition 2.2.6 follows along exactly the same lines as the proof of Lemma 2.3.3. We have

$$\begin{aligned} \int_{\Lambda_\Sigma} \omega(\lambda, \nu)^{-k} d\mu(\lambda) &= \int_{\Lambda_\Sigma} \left(\frac{a_M}{a_m} (1 + a_M^{-1} d(\lambda, \nu)) \right)^{-k} d\mu(\lambda) \\ &= \int_{[0, a_0]} \frac{a_m^k}{a_M^k} \left(\int_{\mathfrak{S}} \int_{\mathbb{R}^2} (1 + a_M^{-1} d(\lambda, \nu))^{-k} ds_\lambda d\mathbf{b}_\lambda \right) \frac{da_\lambda}{a_\lambda^3}. \end{aligned}$$

Therefore, if we could prove an analogue of inequality (2.9), the rest of the proof would follow by the proof of Lemma 2.3.3. We have

$$\begin{aligned}
&\int_{\mathfrak{S}}\int_{\mathbb{R}^2}(1+qd(\lambda,\nu))^{-k}dsd\mathbf{b}_\lambda \\
&\qquad=\int_{\mathfrak{S}}\int_{\mathbb{R}^2}\left(1+q(|\mathbf{b}_\lambda-\mathbf{b}_\nu|^2+|\arctan(-s)-\theta_\nu|^2+|\langle\mathbf{e}_\lambda,\mathbf{b}_\lambda-\mathbf{b}_\nu\rangle|)\right)^{-k}dsd\mathbf{b}_\lambda \\
&\qquad\leqslant\int_{\arctan(-\mathfrak{S})}\int_{\mathbb{R}^2}(1+qd(\lambda,\nu))^{-k}(\theta_\lambda^2+1)d\theta_\lambda d\mathbf{b}_\lambda \\
&\qquad\leqslant C_{\mathfrak{S}}\int_{\mathbb{R}}\int_{\mathbb{R}^2}(1+qd(\lambda,\nu))^{-k}d\theta_\lambda d\mathbf{b}_\lambda \\
&\qquad\lesssim q^{-2}
\end{aligned}$$

where we used the change of variables $\theta=\arctan(-s)$, the boundedness of $\mathfrak{S}$, and Lemma 2.3.3. Hence, the claim follows. □

We will now take a look at two important subtypes of shearlets; the band-limited shearlets and the compactly supported shearlets. The discussion will include the construction of a band-limited, cone-adapted shearlet family which is a family of CPMs, These band-limited shearlets allow a reconstruction formula, and their dual is also a family of CPMs. The importance of having such a family will become clear in Chapter 3.1. We will begin with the compactly supported shearlets.

2.3.3 Compactly Supported Shearlets

Compactly supported shearlets are shearlets with a compact support in the space domain. The existing families of compactly supported shearlets are constructed with separable generators, that is, mother shearlets for horizontal and vertical cones are defined by

$$\psi^0(\mathbf{x})=\psi_1(\mathsf{x}_1)\psi_2(\mathsf{x}_2),\ \text{and}\ \psi^1(\mathbf{x})=\psi^0(\mathsf{x}_2,\mathsf{x}_1). \tag{2.12}$$

In the following we will only consider shearlets of the form (2.12) and will pursue the conditions which ensure that the resulting shearlet family is a family of CPMs and is a frame for functions supported on a cone in $\mathbb{R}^2$. For this to work we will have to endow the functions ψ_1 and ψ_2 with sufficient smoothness and vanishing moments.

Given a dilation parameter a, a shearing parameter s and a location $\mathbf{b}$, we can now define the cone-adapted shearlet family by

$$\psi^\epsilon_{a\,s\,\mathbf{b}}=a^{-3/4}\psi^\epsilon\left(\mathrm{D}_{1/a}\mathrm{S}_s(\mathbf{x}-\mathbf{b})\right),\quad \epsilon\in\{0,1\}. \tag{2.13}$$

Let us now show that (2.13) defines a family of CPMs. We will pay specific attention to the horizontal shearlets and show that a projection of compactly supported shearlets onto a frequency cone

$$\mathcal{C}_{u,v}=\left\{\boldsymbol{\xi}\in\mathbb{R}^2:|\xi_1|\geqslant u,\left|\frac{\xi_2}{\xi_1}\right|\leqslant v\right\},\quad u,v>0,$$

admits a representation formula for $f \in L^2(\mathcal{C}_{u,v})^{\check{}}$. We will come back to these ideas in Section 3.2.2. As a reminder, the space $L^2(\mathcal{D})^{\check{}}$ is defined as

$$L^2(\mathcal{D})^{\check{}} = \{g \in L^2(\mathbb{R}^2) : \operatorname{supp} \hat{g} \subseteq \mathcal{D}\}.$$

Let $\mathsf{P}_{\mathcal{D}}$ denote the frequency space projection of a function $f \in L^2(\mathbb{R}^2)$ onto $\mathcal{D}$, that is, a projection such that

$$\operatorname{supp} \widehat{(\mathsf{P}_{\mathcal{D}} f)} \subseteq \mathcal{D}.$$

We first need to define the notion of vanishing moments.

Definition 2.3.9. *We say that a function $f \in L^2(\mathbb{R})$ has $K \in \mathbb{N}$* vanishing moments *if*

$$\int_{\mathbb{R}} \frac{|\hat{f}(\xi)|^2}{|\xi|^{2K}} d\xi < \infty.$$

Notice that the Definition 2.3.9 is equivalent to saying that $f = \frac{\partial^K}{\partial x^K}\rho$, where $\hat{\rho} \in L^2(\mathbb{R})$.

Proposition 2.3.10. *Consider the shearlet family defined in (2.12) and (2.13), where $\psi_1 \in C^{N_1}(\mathbb{R})$ has compact support and $M + R$ vanishing moments, and $\psi_2 \in C^{N_1+N_2}(\mathbb{R})$ has compact support. Furthermore, we assume M, R, N_1 and N_2 are positive integers that satisfy*

$$2(M + R) - 1/2 > N_1 + N_2 > M + R > 1/2.$$

Under these conditions the shearlet family (2.12)-(2.13) constitutes a family of continuous parabolic molecules of order $(R, M + N_1, N_1, N_2)$. Furthermore, the family of functions

$$\{\mathsf{P}_{\mathcal{C}_{u,v}} \psi_{a\, s\, \mathbf{b}} : a \in [0, 1],\ s \in [-\Xi, \Xi],\ \mathbf{b} \in \mathbb{R}^2\} \tag{2.14}$$

is a tight frame for $L^2(\mathcal{C}_{u,v})^{\check{}}$, provided $u > 0$ and $\Xi > v$.

Proof. Without loss of generality, let us take $\epsilon = 0$. The first step is to show that (2.12)-(2.13) meets the conditions from the definition of shearlet molecules. The statement then follows by Proposition 2.3.7. Therefore, we need to show that for $\boldsymbol{\beta} = (\beta_1, \beta_2) \in \mathbb{N}_0^2$ the generator ψ^0 satisfies the localisation conditions

$$|\partial^{\boldsymbol{\beta}} \hat{\psi}^0(\boldsymbol{\xi})| \lesssim \min\left(1, a + |\xi_1| + a^{1/2}|\xi_2|\right)^M \langle \|\boldsymbol{\xi}\| \rangle^{-N_1} \langle \xi_2 \rangle^{-N_2}.$$

We will split the proof in two parts. First, let $|\xi_1| \geqslant 1$ and write ψ instead of ψ^0 for the sake of brevity. It suffices to show

$$|\xi_1^{N_1} \xi_2^{N_1+N_2} \partial^{\boldsymbol{\beta}} \hat{\psi}(\boldsymbol{\xi})| \lesssim 1 \tag{2.15}$$

since this implies

$$\langle \xi_1 \rangle^{N_1} \langle \xi_2 \rangle^{N_1+N_2} |\partial^{\boldsymbol{\beta}} \hat{\psi}(\boldsymbol{\xi})| \lesssim 1.$$

The claim will then follow from the inequality

$$\langle \xi_1 \rangle^{N_1} \langle \xi_2 \rangle^{N_1+N_2} \geqslant \langle \|\boldsymbol{\xi}\| \rangle^{N_1} \langle \xi_2 \rangle^{N_2}.$$

To start, notice that we have

$$\xi_1^{N_1} \xi_2^{N_1+N_2} \partial^{\boldsymbol{\beta}} \widehat{\psi}(\boldsymbol{\xi}) = C \Big(\partial^{(N_1, N_1+N_2)} \mathsf{x}^{\boldsymbol{\beta}} \psi(\mathsf{x}) \Big)\widehat{\ }(\boldsymbol{\xi}),$$

where the constant C depends on N_1, N_2 and $\boldsymbol{\beta}$. Thus, due to the assumptions about the support and the smoothness of ψ, it follows that $\partial^{(N_1,N_1+N_2)} \mathsf{x}^{\boldsymbol{\beta}} \psi(\mathsf{x})$ is in $L^1(\mathbb{R}^2)$. Therefore, (2.15) holds and the desired conclusion follows.

Now, let $|\xi_1| \leqslant 1$. Using the separability of ψ we have

$$\mathsf{x}^{\boldsymbol{\beta}} \psi(\mathsf{x}) = \mathsf{x}_1^{\beta_1} \psi_1(\mathsf{x}_1) \mathsf{x}_2^{\beta_2} \psi_2(\mathsf{x}_2).$$

It now follows

$$\Big(\partial^{(N_1,N_1+N_2)} \mathsf{x}^{\boldsymbol{\beta}} \psi(\mathsf{x}) \Big)\widehat{\ }(\boldsymbol{\xi}) = \Big(\partial^{N_1} \mathsf{x}_1^{\beta_1} \psi_1(\mathsf{x}_1) \Big)\widehat{\ }(\xi_1) \Big(\partial^{N_1+N_2} \mathsf{x}_2^{\beta_2} \psi_2(\mathsf{x}_2) \Big)\widehat{\ }(\xi_2).$$

Following the same lines of argument as in the case $|\xi_1| \geqslant 1$ we can show that the second term satisfies

$$\left| \Big(\partial^{N_1+N_2} \mathsf{x}_2^{\beta_2} \psi_2(\mathsf{x}_2) \Big)\widehat{\ }(\xi_2) \right| \lesssim 1. \tag{2.16}$$

To deal with the first term we shall use the vanishing moments property. The assumption that ψ_1 has $M+R$ vanishing moments implies that $\psi = \frac{\partial^{M+R}}{\partial \mathsf{x}_1^{M+R}} \rho$, where $\hat{\rho} \in L^2(\mathbb{R})$. Taking the Fourier transform we then have $\hat{\psi}_1(\xi_1) = (2\pi \imath \xi_1)^{M+R} \hat{\rho}(\xi_1)$. Therefore, for $\beta_1 < R$, it follows

$$\frac{\partial^n}{\partial \xi_1^n} (\partial^{\beta_1} \hat{\psi}_1)(0) = 0, \text{ for all } n = 0, \ldots, M + R - \beta_1 - 1.$$

Furthermore, $\partial^{\beta_1} \hat{\psi}_1$ is an analytic function, since it is the Fourier transform of a compactly supported, continuous function whose derivatives (of order up to M) vanish at 0. It follows that the function $\partial^{\beta_1} \hat{\psi}_1$ is uniformly bounded for $|\xi_1| \geqslant 1$, while for small ξ_1 we have $|\partial^{\beta_1} \hat{\psi}_1(\xi_1)| \lesssim |\xi_1|^M$. In other words, we have

$$|\partial^{\beta_1} \hat{\psi}_1(\xi_1)| \lesssim \min\left(1, |\xi_1|\right)^M,$$

which gives

$$\left| \Big(\partial^{N_1} \mathsf{x}_1^{\beta_1} \psi_1(\mathsf{x}_1) \Big)\widehat{\ }(\xi_1) \right| \lesssim \min\left(1, |\xi_1|\right)^{M+N_1}. \tag{2.17}$$

Plugging (2.17) and (2.16) in we have

$$\Big(\partial^{(N_1,N_1+N_2)} \mathsf{x}^{\boldsymbol{\beta}} \psi(\mathsf{x}) \Big)\widehat{\ }(\boldsymbol{\xi}) \lesssim \min\left(1, |\xi_1|\right)^{M+N_1}.$$

which implies

$$|\xi_1^{N_1}\xi_2^{N_1+N_2}\partial^{\boldsymbol{\beta}}\hat{\psi}(\boldsymbol{\xi})| \lesssim \min\left(1, |\xi_1|\right)^{M+N_1},$$

Therefore, we have a system of shearlet molecules of order $(R, M+N_1, N_1, N_2)$, and by Proposition 2.3.7 it is a system of CPMs of the same order.

The second part of the claim follows from applying Theorem 4.9 of [41], which says that assuming ψ has M vanishing moments and Fourier decay of sufficient order in both of the variables, then it admits a representation formula for $f \in L^2(\mathcal{C}_{u,v}\check{)}$. The assumptions of this proposition are such that these conditions are immediately satisfied. Namely, ψ satisfies the Fourier decay conditions since ψ_1 is in $C^{N_1}(\mathbb{R})$ and ψ_2 is in $C^{N_1+N_2}(\mathbb{R})$, and it has $M+R$ vanishing moments by assumptions. Therefore, (2.14) is a tight frame for $L^2(\mathcal{C}_{u,v}\check{)}$. □

We should add that by the argumentation in [41], the cone-adapted shearlet system defined in (2.12)-(2.13) admits a reconstruction formula for $f \in L^2(\mathbb{R}^2)$. This essentially follows by showing that the horizontal shearlets deal with the horizontal cone $\mathcal{C}_{1,1}$, while the vertical shearlets deal with the corresponding vertical cone.

2.3.4 Band-Limited Shearlets

In this section we will develop a somewhat new dictionary of band-limited shearlets. The significance of this construction to the present discussion is twofold: first, its dual frame is also a family of parabolic molecules, and secondly it admits a useful reconstruction formula. Having such a family is essential for a full exploitation of the framework, as we will see in Chapter 3.

The construction presented here is motivated by [59, 60]. Let us begin with some further definitions.

Definition 2.3.11. *We say that a bivariate function $f \in L^2(\mathbb{R}^2)$ has* Fourier decay of order L_i in the i^{th} variable*, where $L_i \in \mathbb{N}$ and $i \in \{1,2\}$, if*

$$\hat{f}(\boldsymbol{\xi}) \lesssim |\xi_i|^{-L_i}. \tag{2.18}$$

Furthermore, we say that a multivariate f has K vanishing moments in the x_j direction*, where $K \in \mathbb{N}$ and $j \in \{1,2\}$, if*

$$\int_{\mathbb{R}^2} \frac{|\hat{f}(\boldsymbol{\xi})|^2}{|\xi_j|^{2K}} d\boldsymbol{\xi} < \infty. \tag{2.19}$$

Same as what we had with Definition 2.3.9, notice that by Proposition 1.1.1 it follows that (2.19) is equivalent to saying that there exists a function $\rho \in L^2(\mathbb{R}^2)$ such that $f = \left(\frac{\partial}{\partial x_1}\right)^K \rho$.

Let us begin our construction by taking a mother shearlet ψ^0 that has Fourier decay of order L_1 in the first variable, K vanishing moments in the x_1 direction, that is,

$\psi^0 = \left(\frac{\partial}{\partial x_1}\right)^M \vartheta$ with $\vartheta \in L^2(\mathbb{R}^2)$. Furthermore, we assume ϑ has Fourier decay of order L_2 in the second variable. The mother shearlet ψ^0 is the classical shearlet

$$\hat{\psi}^0(\boldsymbol{\xi}) = \hat{\psi}_1(\xi_1)\hat{\psi}_2\left(\frac{\xi_2}{\xi_1}\right),$$

where

$$\operatorname{supp}\hat{\psi}_1 \subset \left[-\frac{1}{4}, -\frac{1}{32}\right] \cup \left[\frac{1}{32}, \frac{1}{4}\right], \quad \text{and} \quad \operatorname{supp}\hat{\psi}_2 \subset \left[-\frac{4}{3}, \frac{4}{3}\right].$$

Adhering to the theory of cone-adapted shearlets, the generator for the shearlets that live in the vertical cone is defined via $\psi^1(\mathsf{x}_1, \mathsf{x}_2) = \psi^0(\mathsf{x}_2, \mathsf{x}_1)$.

We can now define a shearlet system through

$$\psi^{\epsilon}_{a\,s\,\mathbf{b}}(\mathbf{x}) = a^{-\frac{3}{4}}\psi^{\epsilon}(\mathrm{D}^{\epsilon}_{1/a}\mathrm{S}^{\epsilon}_{s}(\mathbf{x} - \mathbf{b})), \quad \text{for } \epsilon = 0, 1,$$

where we again the dilation matrices $\mathrm{D}^0_{1/a} = \operatorname{diag}(1/a, 1/\sqrt{a})$, $\mathrm{D}^1_{1/a} = \operatorname{diag}(1/\sqrt{a}, 1/a)$, and the shearing operators are given by $\mathrm{S}^0_s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$ and $\mathrm{S}^1_s = (\mathrm{S}^0_s)^\top$.

In order to create our dictionary we will use the construction from [60], which follows the ideas developed in [61] for the study of domain decompositions for wavelet frames. The idea is to split-up the frequency domain by using a series of partition functions, and then deconstruct each function and patch the corresponding partitions back together. Therefore, we need to define partition functions that can be used to patch horizontal and vertical shearlets together in such a way that they form a system of CPMs and more importantly, that they admit a useful reconstruction formula and that the dual frame is also a CPM family. As a word of warning, the construction we are about to describe is somewhat cumbersome in notation.

To begin, let us define the frequency cones the shearlets will be associated with. The horizontal cones are defined as follows

$$\mathcal{C}^0 := \left\{\boldsymbol{\xi} : |\xi_2| \geqslant \frac{1}{8}, \left|\frac{\xi_1}{\xi_2}\right| \leqslant \frac{4}{3}\right\}, \quad \text{and} \quad \bar{\mathcal{C}}^0 := \left\{\boldsymbol{\xi} : |\xi_2| \geqslant \frac{1}{4}, \left|\frac{\xi_1}{\xi_2}\right| \leqslant \frac{5}{4}\right\}.$$

The vertical cones $\mathcal{C}^1$ and $\bar{\mathcal{C}}^1$ are obtained by rotating the corresponding horizontal cones through the angle of $\pi/2$

$$\mathcal{C}^1 := \left\{\boldsymbol{\xi} : |\xi_1| \geqslant \frac{1}{8}, \left|\frac{\xi_2}{\xi_1}\right| \leqslant \frac{4}{3}\right\}, \quad \text{and} \quad \bar{\mathcal{C}}^1 := \left\{\boldsymbol{\xi} : |\xi_1| \geqslant \frac{1}{4}, \left|\frac{\xi_2}{\xi_1}\right| \leqslant \frac{5}{4}\right\}.$$

For the sake of completeness we will also include the two low-frequency boxes

$$\mathcal{C}^{\mathsf{box}} := \{\boldsymbol{\xi} : \|\boldsymbol{\xi}\|_\infty \leqslant 1\}, \quad \text{and} \quad \bar{\mathcal{C}}^{\mathsf{box}} := \left\{\boldsymbol{\xi} : \|\boldsymbol{\xi}\|_\infty \leqslant \frac{3}{4}\right\}.$$

Let us now define the partition functions that will effectuate this decomposition. Take

$$\gamma^0(\boldsymbol{\xi}) = g_1(\xi_1)g_2\left(\frac{\xi_2}{\xi_1}\right), \quad \text{and} \quad \gamma^1(\boldsymbol{\xi}) = \gamma^0(\xi_2, \xi_1),$$

with

$$\operatorname{supp} g_1 \subset \left[\frac{1}{8}, \infty\right), \quad \text{and} \quad \operatorname{supp} g_2 \subset \left[-\frac{4}{3}, \frac{4}{3}\right],$$

where g_1 and g_2 are smooth and real-valued functions. Notice that by construction we have $\operatorname{supp} \gamma^i \subseteq \mathcal{C}^i$. Similarly, we define

$$\tilde{\chi}^0(\boldsymbol{\xi}) = h_1(\xi_1)h_2\left(\frac{\xi_2}{\xi_1}\right), \quad \tilde{\chi}^1(\boldsymbol{\xi}) = h_1(\xi_2)\left(1 - h_2\left(\frac{\xi_2}{\xi_1}\right)\right), \quad \tilde{\chi}^{\mathsf{box}} = 1 - \tilde{\chi}^0 - \tilde{\chi}^1, \tag{2.20}$$

with

$$\operatorname{supp} h_1 \subset \left[\frac{1}{4}, \infty\right), \quad h_1|_{[\frac{1}{2},\infty)} \equiv 1, \text{ and } \operatorname{supp} h_2 \subset \left[-\frac{5}{4}, \frac{5}{4}\right], \quad h_2|_{[-\frac{4}{5},\frac{4}{5}]} \equiv 1, \tag{2.21}$$

where h_1 and h_2 are non-negative smooth functions and furthermore, h_1 is non-decreasing. It follows straight from the definitions that $\operatorname{supp} \tilde{\chi}^i \subset \bar{\mathcal{C}}^i$ for $i = 0, 1$.

We will now show that these functions are bounded on the required cones. This is essentially the same as the analogous proof in [60]. The difference lies in the fact that instead of dealing with $\tilde{\chi}^0, \tilde{\chi}^1$ and χ^{box}, we want to have χ^0 and χ^1 such that $\chi^0(\boldsymbol{\xi}) + \chi^1(\boldsymbol{\xi}) = 1$, for reasons which will be discussed in the proof of Theorem 2.3.15.

Lemma 2.3.12. *For $\epsilon = 0, 1$ we have*

$$\left\| \tilde{\chi}^\epsilon \left((\mathrm{D}^i_{1/a} \mathrm{S}^i_s)^\top \boldsymbol{\xi} \right) \right\|_{C^N(\operatorname{supp} \hat{\psi}^i)} \leqslant C_N, \ N \in \mathbb{N},$$

and

$$\left\| \chi^{\mathsf{box}} \left((\mathrm{D}^i_{1/a} \mathrm{S}^i_s)^\top \boldsymbol{\xi} \right) \right\|_{C^N(\operatorname{supp} \hat{\psi}^i)} \leqslant \tilde{C}_N, \ N \in \mathbb{N}.$$

Analogous estimates hold for γ^ϵ and $\frac{\chi^\epsilon}{\gamma^\epsilon}$.

Proof. Consider $\epsilon = 0$. We have

$$\tilde{\chi}^0 \left((\mathrm{D}^0_{1/a} \mathrm{S}^0_s)^\top \boldsymbol{\xi} \right) = h_1(a^{-1}\xi_1) h_2\left(s + a^{1/2} \frac{\xi_2}{\xi_1} \right).$$

Since $\boldsymbol{\xi} \in \operatorname{supp} \hat{\psi}^0$ we have

$$\frac{a^{-1}}{32} \leqslant |a^{-1}\xi_1| \leqslant \frac{a^{-1}}{4}.$$

Therefore, provided $a \leqslant 2^{-4}$ we have $h_1(a^{-1}\xi_1) = 1$, which gives

$$\tilde{\chi}^0(a^{-1}\xi_1, a^{-1}s\xi_1 + a^{-1/2}\xi_2) = h_2\left(s + a^{1/2} \frac{\xi_2}{\xi_1} \right).$$

Using (2.21) we have

$$\begin{aligned}\left|\partial^{\boldsymbol{\alpha}} h_2\left(s + a^{1/2}\frac{\xi_2}{\xi_1}\right)\right| &\lesssim a^{\frac{|\boldsymbol{\alpha}|}{2}}|\xi_1|^{-2|\boldsymbol{\alpha}|}|\xi_2|^{|\boldsymbol{\alpha}|}\sup_{\boldsymbol{\beta}\leqslant\boldsymbol{\alpha}}\left|\partial^{\boldsymbol{\beta}} h_2\left(s + a^{1/2}\frac{\xi_2}{\xi_1}\right)\right| \\ &\lesssim \left(\frac{|\xi_2|}{|\xi_1|}\right)^{|\boldsymbol{\alpha}|}|\xi_1|^{-|\boldsymbol{\alpha}|} \lesssim \left(\frac{4}{3}\right)^{|\boldsymbol{\alpha}|}\left(\frac{1}{4}\right)^{-|\boldsymbol{\alpha}|} \\ &\lesssim 1,\end{aligned}$$

for a multi-index $\boldsymbol{\alpha} \in \mathbb{N}_0^2$. What is left is to address the case $a > 2^{-4}$. We have

$$\begin{aligned}|\partial^{\boldsymbol{\alpha}}\tilde{\chi}^0\left((\mathrm{D}^0_{1/a}\mathrm{S}^0_s)^\top\boldsymbol{\xi}\right)| &\lesssim a^{-|\boldsymbol{\alpha}|/2}|\xi_1|^{-2|\boldsymbol{\alpha}|}|\xi_2|^{|\boldsymbol{\alpha}} \\ &\lesssim 3^{-|\boldsymbol{\alpha}|}a^{-|\boldsymbol{\alpha}|/2} \\ &\lesssim 1\end{aligned}$$

where the first inequality follows from the properties of $\operatorname{supp} h_1$ and $\operatorname{supp} h_2$.

Now we turn our attention to the case $\epsilon = 1$. We have

$$\tilde{\chi}^1\left((\mathrm{D}^1_{1/a}\mathrm{S}^1_s)^\top\boldsymbol{\xi}\right) = h_1(a^{-1}\xi_2)\left(1 - h_2\left(\frac{a^{-1}\xi_2}{a^{-1}s\xi_2 + a^{-1/2}\xi_1}\right)\right).$$

The question regarding $h_1(a^{-1}\xi_2)$ and its derivatives is readily addressed using the same arguments as we did in the case $i = 0$. Thus, due to the properties of h_2 and its support, we only need to consider the case

$$\frac{5}{4} \geqslant \left|a^{1/2}\frac{\xi_1}{\xi_2} + s\right| \geqslant \frac{4}{5}. \tag{2.22}$$

Define the functions $g(\boldsymbol{\xi}) = \xi_2$ and $h(\boldsymbol{\xi}) = (a^{1/2}\xi_1 + s\xi_2)^{-1}$ so that

$$(gh)(\boldsymbol{\xi}) = \frac{a^{-1}\xi_2}{a^{-1}s\xi_2 + a^{-1/2}\xi_1} = \frac{1}{a^{1/2}\frac{\xi_1}{\xi_2} + s}.$$

Consider a multi-index $\boldsymbol{\alpha} = (\alpha_1, \alpha_2) \in \mathbb{N}_0^2$. We have

$$\partial^{\boldsymbol{\alpha}}(gh)(\boldsymbol{\xi}) = \sum_{\boldsymbol{\beta}\leqslant\boldsymbol{\alpha}} C_{\boldsymbol{\beta}}\partial^{\boldsymbol{\alpha}-\boldsymbol{\beta}} g(\boldsymbol{\xi})\partial^{\boldsymbol{\beta}} h(\boldsymbol{\xi}).$$

Bounding the derivatives of g is trivial. Regarding the derivatives of h, we have

$$\partial^{\boldsymbol{\alpha}} h(\boldsymbol{\xi}) = (-1)^{|\boldsymbol{\alpha}|}a^{\frac{\alpha_1}{2}}s^{\alpha_2}(h(\boldsymbol{\xi}))^{2|\boldsymbol{\alpha}|}.$$

Therefore, the function $\partial^{\boldsymbol{\alpha}} h$ is bounded from above as long as h is, that is, as long as $1/h(\boldsymbol{\xi}) = a^{1/2}\xi_1 + s\xi_2$ is bounded from below which is ensured by (2.22). In conclusion, $\tilde{\chi}^1$ and its derivatives are bounded on $\operatorname{supp}\hat{\psi}^1$.

Let us now find the bounds for χ^{box}. Without loss of generality we will consider only the case $\epsilon = 0$. In order to show

$$\left\| \chi^{\mathsf{box}} \left((\mathrm{D}^0_{1/a} \mathrm{S}^0_s)^\top \boldsymbol{\xi} \right) \right\|_{C^N(\operatorname{supp} \hat{\psi}^0)} \leqslant C_N,$$

it is sufficient to show that derivatives of $\tilde{\chi}^1 \left((\mathrm{D}^0_{1/a} \mathrm{S}^0_s)^\top \boldsymbol{\xi} \right)$ are uniformly bounded on $\operatorname{supp} \hat{\psi}^0$. We have

$$\tilde{\chi}^1 \left((\mathrm{D}^0_{1/a} \mathrm{S}^0_s)^\top \boldsymbol{\xi} \right) = h_1(a^{-1} s \xi_1 + a^{-1/2} \xi_2) \left(1 - h_2 \left(\frac{a^{-1} s \xi_1 + a^{-1/2} \xi_2}{a^{-1} \xi_1} \right) \right).$$

Restrictions imposed by (2.20) suggest we only ought to consider

$$\frac{4}{5} \leqslant \left| \frac{a^{-1} s \xi_1 + a^{-1/2} \xi_2}{a^{-1} \xi_1} \right| \leqslant \frac{5}{4},$$

which gives

$$\frac{1}{40} \leqslant \left| s \xi_1 + a^{1/2} \xi_2 \right| \leqslant \frac{5}{16}. \tag{2.23}$$

On the other hand, the support of h_1 suggests that we only need to consider $\boldsymbol{\xi}$ that satisfy

$$\frac{1}{4} \leqslant a^{-1} |s \xi_1 + a^{1/2} \xi_2|.$$

Combined with (2.23), it follows that there is an $\tilde{a}$ such that if $a \leqslant \tilde{a}$ then

$$a^{-1} |s \xi_1 + a^{1/2} \xi_2| < \frac{1}{4}.$$

which means that such $\boldsymbol{\xi}$ are not in $\operatorname{supp} h_1$. Using now (2.22) for $a > \tilde{a}$ we have

$$\left| \partial^{\boldsymbol{\alpha}} h_1(a^{-1} s \xi_1 + a^{-1/2} \xi_2) \right| \lesssim a^{-\alpha_1 - \frac{\alpha_2}{2}} s^{\alpha_2} \left| \partial^{\boldsymbol{\alpha}} h_1(a^{-1} s \xi_1 + a^{-1/2} \xi_2) \right| \lesssim 1.$$

Bounding $h_2 \left(s + a^{1/2} \frac{\xi_2}{\xi_1} \right)$ is analogous to the case $\epsilon = 1$.

□

Let us now write $\chi^{\mathsf{box}} = \chi^{\mathsf{box},0} + \chi^{\mathsf{box},1}$, where $\chi^{\mathsf{box},0}$ and $\chi^{\mathsf{box},1}$ are smooth, and define $\chi^0 = \tilde{\chi}^0 + \chi^{\mathsf{box},0}$ and $\chi^1 = \tilde{\chi}^1 + \chi^{\mathsf{box},1}$. Using Lemma 2.3.12 it follows that the statements analogous to those in Lemma 2.3.12 hold for χ^0 and χ^1 as well. We can now define our shearlet system in the Fourier domain as

$$\hat{\sigma}^\epsilon_{a\, s\, \mathbf{b}}(\boldsymbol{\xi}) = \gamma^\epsilon(\boldsymbol{\xi}) \hat{\psi}^\epsilon_{a\, s\, \mathbf{b}}(\boldsymbol{\xi})$$

and its corresponding dual system

$$\hat{\tilde{\sigma}}^\epsilon_{a\, s\, \mathbf{b}}(\boldsymbol{\xi}) = \frac{\chi^\epsilon(\boldsymbol{\xi})}{\gamma^\epsilon(\boldsymbol{\xi})} \hat{\psi}^\epsilon_{a\, s\, \mathbf{b}}(\boldsymbol{\xi}),$$

for $\epsilon = 0, 1$.

Theorem 2.3.13. *The families*

$$\Psi = \left\{ \sigma^{\epsilon}_{a\, s\, \mathbf{b}}, \epsilon \in \{0,1\}, a \in [0,1], s \in \left(-\frac{3}{2}, \frac{3}{2}\right), \mathbf{b} \in \mathbb{R}^2 \right\}$$

and

$$\tilde{\Psi} = \left\{ \tilde{\sigma}^{\epsilon}_{a\, s\, \mathbf{b}}, \epsilon \in \{0,1\}, a \in [0,1], s \in \left(-\frac{3}{2}, \frac{3}{2}\right), \mathbf{b} \in \mathbb{R}^2 \right\}$$

are families of parabolic molecules.

Proof. This follows directly from Lemma 2.3.12 and the support properties of ψ^{ϵ}. □

Let us now establish a reconstruction formula by adopting the methodology of [59]. Define

$$C_{\psi_\epsilon} = \int_{\mathbb{R}} \int_{\mathbb{R}^+} \frac{|\hat{\psi}^{\epsilon}(\boldsymbol{\xi})|^2}{|\xi_{1+\epsilon}|^2} d\xi_{1+\epsilon} d\xi_{2-\epsilon}, \tag{2.24}$$

and

$$\Delta_{\psi_\epsilon}(\boldsymbol{\xi}) = \int_{-3/2}^{3/2} \int_0^1 \left| \hat{\psi}(a\xi_{1+\epsilon}, a^{\frac{1}{2}}(\xi_{2-\epsilon} - s\xi_{1+\epsilon})) \right|^2 a^{-\frac{3}{2}} da ds. \tag{2.25}$$

Lastly, we define functions φ_1 and φ_2 through

$$|\hat{\varphi}_0(\boldsymbol{\xi})|^2 = C_{\psi_0} - \Delta_{\psi_0} \text{ and } |\hat{\varphi}_1(\boldsymbol{\xi})|^2 = C_{\psi_1} - \Delta_{\psi_1}.$$

In order to establish the validity and existence of a reconstruction formula, we first have to show the smoothness of functions φ_0 and φ_1. This will be done using the standard arguments from Fourier analysis, since it is sufficient to show that the Fourier transform of φ_ϵ decays rapidly.

Lemma 2.3.14. *Assume ψ^0 has the Fourier decay of order L_1 in the 1^{st} variable, N in the 2^{nd} variable and has K vanishing moments in the x_1 direction so that $\psi^0 = \left(\frac{\partial}{\partial x_1}\right)^K \vartheta$ where ϑ has Fourier decay of order L_2 in the 2^{nd} variable. Then assuming $K \geqslant L_2 + 4$, $L_2 \geqslant N + 2$ and $L_1 \geqslant N$ we have*

$$\hat{\varphi}_0(\boldsymbol{\xi}) = \mathcal{O}(\|\boldsymbol{\xi}\|^{-N})$$

for $\left|\frac{\xi_2}{\xi_1}\right| \leqslant \frac{4}{3}$. An analogous statement holds for φ_1 in the cone $\left|\frac{\xi_1}{\xi_2}\right| \leqslant \frac{4}{3}$.

Proof. Let us drop first drop the indices and use the fact that C_ψ can be written as

$$C_\psi = \int_{\mathbb{R}} \int_0^\infty \left| \hat{\psi}\left(a\xi_1, a^{1/2}(\xi_2 - s\xi_1)\right) \right|^2 a^{-3/2} da ds,$$

by the change of variables $\eta_1 = -a\xi_1$ and $\eta_2 = a^{1/2}(\xi_2 - s\xi_1)$. Therefore, we have

$$|\hat{\varphi}(\boldsymbol{\xi})| = \underbrace{\int_{|s|>3/2}\int_{a>0}\left|\hat{\psi}(a\xi_1, a^{1/2}(\xi_2 - s\xi_1))\right|^2 a^{-3/2}dads}_{I_1} + \underbrace{\int_{|s|<3/2}\int_{a>1}\left|\hat{\psi}\left(a\xi_1, a^{1/2}(\xi_2 - s\xi_1)\right)\right|^2 a^{-3/2}dads}_{I_2}.$$

We will estimate I_1 by splitting the integration in two parts. Firstly, we have

$$\begin{aligned}\int_{|s|>3/2}\int_{a>1}&\left|\hat{\psi}\big(a\xi_1, a^{1/2}(\xi_2 - s\xi_1)\big)\right|^2 a^{-3/2}dads \\ &\lesssim \int_{|s|>3/2}\int_{a>1} a^{-N-3/2}\big(\xi_2 - s\xi_1\big)^{-2N}dads \\ &\lesssim \int_{|s|>3/2}\int_{a>1}|\xi_1|^{-2N}a^{-N-3/2}\Big|\frac{4}{3} - |s|\Big|^{-2N}dads \lesssim \|\boldsymbol{\xi}\|^{-N}.\end{aligned}$$

where the last inequality holds since by assumption $\left|\frac{\xi_2}{\xi_1}\right| \leqslant \frac{4}{3}$. On the other hand, we have

$$\begin{aligned}\int_{|s|>3/2}\int_{a>1}&\left|\hat{\psi}\big(a\xi_1, a^{1/2}(\xi_2 - s\xi_1)\big)\right|^2 a^{-3/2}dads \\ &\lesssim \int_{|s|>3/2}\int_{a>1} a^{K-3/2}|\xi_1|^K|\vartheta\big(a\xi_1, a^{1/2}(\xi_2 - s\xi_1)\big)|dads \\ &\lesssim \int_{|s|>3/2}\int_{a>1} a^{K-L_2-3/2}|\xi_1|^{K-2L_2}\Big|\frac{4}{3} - |s|\Big|dads \lesssim \|\boldsymbol{\xi}\|^{-N}.\end{aligned}$$

Now, to estimate I_2 we have through a simple argument

$$\int_{|s|<3/2}\int_{a>1}\left|\hat{\psi}\left(a\xi_1, a^{1/2}(\xi_2 - s\xi_1)\right)\right|^2 a^{-3/2}dads \lesssim \int_1^\infty a^{-L_1-3/2}|\xi_1|^{-L_1}da \lesssim \|\boldsymbol{\xi}\|^{-N}.$$

Collecting the terms yields the desired conclusion. □

To derive a reconstruction formula consider now the functions

$$f_{\mathsf{high}} = \sum_{\epsilon=0}^{1}\int_{[0,1]\times\left(-\frac{3}{2},\frac{3}{2}\right)\times\mathbb{R}^2}\langle f, \sigma^\epsilon_{a\,s\,\mathbf{b}}\rangle\tilde{\sigma}^\epsilon_{a\,s\,\mathbf{b}}\frac{da\,ds\,d\mathbf{b}}{a^3} \tag{2.26}$$

and

$$f_{\mathsf{low}} = \sum_{\epsilon=0}^{1}\int_{\mathbb{R}^2}\langle f, \check{\gamma}^0 \star \mathsf{T}_{\mathbf{b}}\varphi_\epsilon\rangle\widetilde{\left(\frac{\chi^\epsilon}{\gamma^0}\right)}d\mathbf{b}. \tag{2.27}$$

We now have all the ingredients required to show the reconstruction formula. The proof follows along the same lines as analogous proofs for wavelets, shearlets and curvelets, which we will include for the sake of completeness.

Theorem 2.3.15. *Consider functions f_{high} and f_{low} as defined in (2.26) and (2.27). We then have*

$$f = \frac{1}{C_\psi}(f_{high} + f_{low}). \tag{2.28}$$

Proof. Define

$$g^i_{a\,s}(\mathbf{x}) = \int_{\mathbb{R}^2} \langle f, \sigma^i_{a\,s\,\mathbf{b}} \rangle \tilde{\sigma}^i_{a\,s\,\mathbf{b}} d\mathbf{b}.$$

It follows

$$\begin{aligned} g^\epsilon_{a\,s}(\mathbf{x}) &= \int_{\mathbb{R}} \tilde{\sigma}^\epsilon_{a\,s\,\mathbf{0}}(\mathbf{x}-\mathbf{b}) \int_{\mathbb{R}} f(\mathbf{y}) \sigma^{\epsilon,-}_{a\,s\,\mathbf{0}}(\mathbf{b}-\mathbf{y}) d\mathbf{y} d\mathbf{x} \\ &= \left(f \star \sigma^{\epsilon,-}_{a\,s\,\mathbf{0}} \star \tilde{\sigma}^\epsilon_{a\,s\,\mathbf{0}} \right)(\mathbf{b}). \end{aligned}$$

where $\sigma^{\epsilon,-}_{a\,s\,\mathbf{0}}(\mathbf{x}) = \overline{\sigma}^\epsilon_{a\,s\,\mathbf{0}}(-\mathbf{x})$. Therefore, it follows

$$\hat{g}^\epsilon_{a\,s}(\boldsymbol{\xi}) = \left(\hat{f} \cdot \hat{\sigma}^{\epsilon,-}_{a\,s\,\mathbf{0}} \cdot \hat{\tilde{\sigma}}^\epsilon_{a\,s\,\mathbf{0}} \right)(\boldsymbol{\xi}).$$

Taking now the Fourier transform of f_{high} yields

$$\begin{aligned} \hat{f}_{\mathsf{high}}(\boldsymbol{\xi}) &= \sum_{\epsilon=0}^{1} \int_{[0,1]\times\left(-\frac{3}{2},\frac{3}{2}\right)} \frac{\hat{g}^\epsilon_{a\,s}(\boldsymbol{\xi}) da ds}{a^3} \\ &= \hat{f}(\boldsymbol{\xi}) \sum_{\epsilon=0}^{1} \int_{[0,1]\times\left(-\frac{3}{2},\frac{3}{2}\right)} \frac{\hat{\sigma}^{\epsilon,-}_{a\,s\,\mathbf{0}}(\boldsymbol{\xi}) \hat{\tilde{\sigma}}^\epsilon_{a\,s\,\mathbf{0}}(\boldsymbol{\xi}) da ds}{a^3}. \end{aligned}$$

A simple calculation yields

$$\hat{\sigma}^{\epsilon,-}_{a\,s\,\mathbf{0}}(\boldsymbol{\xi}) \hat{\tilde{\sigma}}^\epsilon_{a\,s\,\mathbf{0}}(\boldsymbol{\xi}) = |\hat{\psi}^\epsilon_{a\,s\,\mathbf{0}}(\boldsymbol{\xi})|^2 \chi^\epsilon(\boldsymbol{\xi}).$$

Therefore,

$$\hat{f}_{\mathsf{high}}(\boldsymbol{\xi}) = \hat{f}(\boldsymbol{\xi}) \sum_{\epsilon=0}^{1} \chi^\epsilon(\boldsymbol{\xi}) \Delta_{\psi^\epsilon}(\boldsymbol{\xi}). \tag{2.29}$$

On the other hand, computing the Fourier transform of f_{low} we have

$$\hat{f}_{\mathsf{low}}(\boldsymbol{\xi}) = \hat{f}(\boldsymbol{\xi}) \left(\chi^0(\boldsymbol{\xi}) |\hat{\varphi}_0(\boldsymbol{\xi})|^2 + \chi^1(\boldsymbol{\xi}) |\hat{\varphi}_1(\boldsymbol{\xi})|^2 \right). \tag{2.30}$$

Collecting (2.29) and (2.30) we get

$$\begin{aligned} (\hat{f}_{\mathsf{high}}(\boldsymbol{\xi}) + \hat{f}_{\mathsf{low}}(\boldsymbol{\xi})) &= \hat{f}(\boldsymbol{\xi}) \left(\chi^0(\boldsymbol{\xi})(\Delta_{\psi^0}(\boldsymbol{\xi}) + |\hat{\varphi}_0(\boldsymbol{\xi})|^2) + \chi^1(\boldsymbol{\xi})(\Delta_{\psi^1}(\boldsymbol{\xi}) + |\hat{\varphi}_1(\boldsymbol{\xi})|^2) \right) \\ &= C_\psi \hat{f}(\boldsymbol{\xi})(\chi^0(\boldsymbol{\xi}) + \chi^1(\boldsymbol{\xi})) = \hat{f}(\boldsymbol{\xi}) \end{aligned}$$

since $C_{\psi^0} = C_{\psi^1}$, which is precisely what we wanted to show.

□

2.4 Almost Orthogonality

We will now state and prove our result on almost orthogonality of continuous parabolic molecules, which will be the essential tool for the applications in microlocal analysis we will need for the next chapter. The following theorem says that given two families of CPMs, the non-diagonal entries of their inter-Gramian matrix $\langle m_\lambda, n_\nu \rangle$ decay nearly exponentially fast, as measured by the pseudo-distance function ω. As we discussed earlier, this means that the Gramian is in a certain sense sparse, and that the parabolic molecules whose indices are close by in the parameter space $\mathbb{P}$ are *similar*.

Theorem 2.4.1. *Let $\Gamma = \{m_\lambda : \lambda \in \Lambda_\Gamma\}$ and $\Sigma = \{n_\nu : \nu \in \Lambda_\Sigma\}$ be two families of continuous parabolic molecules, both of order (at least) (R, M, N_1, N_2). Then*

$$|\langle m_\lambda, n_\nu \rangle| \leqslant C_N \, \omega(\lambda, \nu)^{-N}$$

holds for every $N \in \mathbb{N}$ such that

$$R \geqslant 2N, \quad M > 3N - \frac{5}{4}, \quad N_1 \geqslant N + \frac{3}{4}, \quad N_2 \geqslant 2N.$$

This result certainly is not that much of a surprise. Its precedents (in the analogue setting) can be found for example in [5] and [9], though it has not been shown in the continuous setting until now. The proof of this theorem follows almost entirely along the lines of its analogues in the case of curvelet [5] and shearlet molecules [8], and most importantly, discrete parabolic molecules [9] There are some technical differences though that need to be taken care of. The majority of minor proofs will be postponed to the Appendix A.

Proof. Since Γ and Σ are CPMs, we can write

$$\begin{aligned} m_\lambda(\mathbf{x}) &= a_\lambda^{-3/4} \varphi^{(\lambda)} \left(\mathrm{D}_{1/a_\lambda} \mathrm{R}_{\theta_\lambda} (\mathbf{x} - \mathbf{b}_\lambda) \right), \\ n_\nu(\mathbf{x}) &= a_\nu^{-3/4} \psi^{(\nu)} \left(\mathrm{D}_{1/a_\nu} \mathrm{R}_{\theta_\nu} (\mathbf{x} - \mathbf{b}_\nu) \right), \end{aligned}$$

where for notational convenience we take $(a_\lambda, \theta_\lambda, \mathbf{b}_\lambda), (a_\nu, \theta_\nu, \mathbf{b}_\nu) \in \mathbb{P}$, instead of referring to their respective parameterisations.

Parseval's formula then gives

$$\begin{aligned} \langle m_\lambda, n_\nu \rangle = \langle \hat{m}_\lambda, \hat{n}_\nu \rangle &= \int_{\mathbb{R}^2} \hat{m}_\lambda(\boldsymbol{\xi}) \overline{\hat{n}_\nu(\boldsymbol{\xi})} d\boldsymbol{\xi} \\ &= (a_\lambda a_\nu)^{3/4} \int_{\mathbb{R}^2} \hat{\varphi}^{(\lambda)} \left(\mathrm{D}_{a_\lambda} \mathrm{R}_{\theta_\lambda} \boldsymbol{\xi} \right) \overline{\hat{\psi}^{(\nu)} \left(\mathrm{D}_{a_\nu} \mathrm{R}_{\theta_\nu} \boldsymbol{\xi} \right)} e^{-2\pi \imath (\mathbf{b}_\lambda - \mathbf{b}_\nu) \cdot \boldsymbol{\xi}} d\boldsymbol{\xi}. \end{aligned} \tag{2.31}$$

Having the definition of the pseudo-distance function ω and expression (2.31) on our mind, it is immediately apparent that we ought to extract from (2.31) the desired decay in terms of the spatial, angular, and dilated parameters. The spatial decay can be achieved by a repeated use of the integration by parts with respect to an appropriate differential

operator. Therefore, the question is what would be a natural choice for a differential operator that would achieve this goal. Naturally, the general idea is to make use of the decay estimates (2.3). Integration by parts gives

$$\int_{\mathbb{R}^2} \hat{\varphi}^{(\lambda)}\left(\mathrm{D}_{a_\lambda}\mathrm{R}_{\theta_\lambda}\boldsymbol{\xi}\right)\overline{\hat{\psi}^{(\nu)}\left(\mathrm{D}_{a_\nu}\mathrm{R}_{\theta_\nu}\boldsymbol{\xi}\right)}e^{-2\pi\imath(\mathbf{b}_\lambda-\mathbf{b}_\nu)\cdot\boldsymbol{\xi}}d\boldsymbol{\xi} = \tag{2.32}$$

$$= \int_{\mathbb{R}^2} \mathcal{L}_{\lambda,\nu}^{k}\left(\hat{\varphi}^{(\lambda)}\left(\mathrm{D}_{a_\lambda}\mathrm{R}_{\theta_\lambda}\boldsymbol{\xi}\right)\overline{\hat{\psi}^{(\nu)}\left(\mathrm{D}_{a_\nu}\mathrm{R}_{\theta_\nu}\boldsymbol{\xi}\right)}\right)\mathcal{L}_{\lambda,\nu}^{-k}\left(e^{-2\pi\imath(\mathbf{b}_\lambda-\mathbf{b}_\nu)\cdot\boldsymbol{\xi}}\right)d\boldsymbol{\xi},$$

where the differential operator $\mathcal{L}_{\lambda,\nu}$ is defined via

$$\mathcal{L}_{\lambda,\nu} = \mathcal{I}_d - a_M^{-1}\Delta - \frac{a_M^{-2}}{1+a_M^{-1}|\theta_\lambda-\theta_\nu|^2}\frac{\partial^2}{\partial\mathbf{e}_\lambda^2}, \tag{2.33}$$

and $\mathcal{I}_d$ is the identity operator.

Let us introduce some shorthand notation. Denote $\delta\mathbf{b} = \mathbf{b}_\lambda - \mathbf{b}_\nu$ and $\delta\theta = \theta_\lambda - \theta_\nu$. On one hand Lemma A.1.2 states that the exponentials are eigenfunctions of $\mathcal{L}_{\lambda,\nu}$. In other words, we have

$$\mathcal{L}_{\lambda,\nu}^{-N}\left(e^{-2\pi\imath\boldsymbol{\xi}\cdot\delta\mathbf{b}}\right) = \left[1+4\pi^2 a_M^{-1}|\delta\mathbf{b}|^2+4\pi^2\frac{a_M^{-2}}{1+a_M^{-1}|\delta\theta|^2}\langle\mathbf{e}_\lambda,\delta\mathbf{b}\rangle^2\right]^{-k} e^{-2\pi\imath\boldsymbol{\xi}\cdot\delta\mathbf{b}}. \tag{2.34}$$

On the other hand, a simple computation (proved in Lemma A.10) gives that the term $\mathcal{L}_{\lambda,\nu}\left(\hat{\varphi}^{(\lambda)}\left(\mathrm{D}_{a_\lambda}\mathrm{R}_{\theta_\lambda}\boldsymbol{\xi}\right)\overline{\hat{\psi}^{(\nu)}\left(\mathrm{D}_{a_\nu}\mathrm{R}_{\theta_\nu}\boldsymbol{\xi}\right)}\right)$ can be written as a finite linear combination of the terms of the same form, that is, of terms $\hat{c}^{(\lambda)}\left(\mathrm{D}_{a_\lambda}\mathrm{R}_{\theta_\lambda}\boldsymbol{\xi}\right)\overline{\hat{d}^{(\nu)}\left(\mathrm{D}_{a_\nu}\mathrm{R}_{\theta_\nu}\boldsymbol{\xi}\right)}$, where c_λ and d_λ satisfy conditions analogous to (2.3). Applying $\mathcal{L}_{\lambda,\nu}$ results in a loss of two degrees of smoothness. Therefore, this process can be repeated at most N times, where $N \leqslant R/2$.

Lemma A.1.3 from the Appendix now provides a bound for $\mathcal{L}_{\lambda,\nu}^{N}$ for all N such that $N \leqslant \frac{R}{2}$

$$\left|\mathcal{L}_{\lambda,\nu}^{N}\left(\hat{\varphi}^{(\lambda)}\left(\mathrm{D}_{a_\lambda}\mathrm{R}_{\theta_\lambda}\boldsymbol{\xi}\right)\overline{\hat{\psi}^{(\nu)}\left(\mathrm{D}_{a_\nu}\mathrm{R}_{\theta_\nu}\boldsymbol{\xi}\right)}\right)\right| \lesssim S_{\lambda,M-N_2,N_1,N_2}(\boldsymbol{\xi})S_{\nu,M-N_2,N_1,N_2}(\boldsymbol{\xi}), \tag{2.35}$$

where in the polar coordinates we define

$$S_{\eta,M,N_1,N_2,N_2}(r,\alpha) = \min(1,a_\eta(1+r))^M\left(1+a_\eta^{-1/2}|\sin(\alpha+\theta_\eta)|\right)^{-N_2}(1+a_\eta r)^{-N_1}.$$

Plugging (2.34) and (2.35) into (2.32), and extracting the eigenvalues of $\mathcal{L}_{\lambda,\nu}$ yields

$$|\langle m_\lambda, n_\nu\rangle| \lesssim \mathcal{S}\cdot\left[1+a_M^{-1}\|\delta\mathbf{b}\|^2+\frac{a_M^{-2}}{1+a_M^{-1}|\delta\theta|^2}\langle\mathbf{e}_\lambda,\delta\mathbf{b}\rangle^2\right]^{-N}, \tag{2.36}$$

where

$$\mathcal{S} = (a_\lambda a_\nu)^{3/4} \int_{\mathbb{R}^2} S_{\lambda,M-N_2,N_1,N_2}(\boldsymbol{\xi}) S_{\nu,M-N_2,N_1,N_2}(\boldsymbol{\xi}) d\boldsymbol{\xi}. \tag{2.37}$$

The second term in equation (2.36) gives spatial localisation, whereas we can use Lemma A.1.5 from the Appendix to estimate $\mathcal{S}$ in terms of the dilation parameters

$$\mathcal{S} \lesssim \left(\frac{a_M}{a_m}\right)^{-N} \left(1 + a_M^{-1/2}|\delta\theta|\right)^{-N}.$$

Hence, it follows

$$\begin{aligned}
|\langle m_\lambda, n_\nu\rangle| &\lesssim \left(\frac{a_M}{a_m}\right)^{-N} \left(1 + a_M^{-1/2}|\delta\theta|\right)^{-N} \left[1 + a_M^{-1}\|\delta\mathbf{b}\|^2 + \frac{a_M^{-2}}{1 + a_M^{-1}|\delta\theta|^2}\langle \mathbf{e}_\lambda, \delta\mathbf{b}\rangle^2\right]^{-N} \\
&\lesssim \left(\frac{a_M}{a_m}\right)^{-N} \left(1 + a_M^{-1}|\delta\theta|^2 + a_M^{-1}\|\delta\mathbf{b}\|^2 + \frac{1}{a_M^2 + a_M|\delta\theta|^2}\langle \mathbf{e}_\lambda, \delta\mathbf{b}\rangle^2\right)^{-N}
\end{aligned}$$

Using now the quadratic mean-geometric mean inequality we have

$$\begin{aligned}
1 + a_M^{-1}|\delta\theta|^2 + \frac{1}{a_M^2 + a_M|\delta\theta|^2}\langle \mathbf{e}_\lambda, \delta\mathbf{b}\rangle^2 &= \left(\sqrt{1 + a_M^{-1}|\delta\theta|^2}\right)^2 + \left(\frac{|\langle \mathbf{e}_\lambda, \delta\mathbf{b}\rangle|}{\sqrt{a_M^2 + a_M|\delta\theta|^2}}\right)^2 \\
&\geqslant 2\sqrt{1 + a_M^{-1}|\delta\theta|^2}\frac{|\langle \mathbf{e}_\lambda, \delta\mathbf{b}\rangle|}{\sqrt{a_M^2 + a_M|\delta\theta|^2}} \\
&= 2a_M^{-1}|\langle \mathbf{e}_\lambda, \delta\mathbf{b}\rangle|
\end{aligned}$$

Therefore, plugging all the expressions back in we get

$$\begin{aligned}
|\langle m_\lambda, n_\nu\rangle| &\lesssim \left(\frac{a_M}{a_m}\right)^{-N} \left(1 + a_M^{-1}\|\delta\mathbf{b}\|^2 + a_M^{-1}|\langle \mathbf{e}_\lambda, \delta\mathbf{b}\rangle|\right)^{-N} \\
&\lesssim \left(\frac{a_M}{a_m}\right)^{-N} \left(1 + a_M^{-1}\left(\|\delta\mathbf{b}\|^2 + |\delta\theta|^2 + |\langle \mathbf{e}_\lambda, \delta\mathbf{b}\rangle|\right)\right)^{-N} \\
&\lesssim \omega(\lambda, \nu)^{-N},
\end{aligned}$$

which concludes the proof. □

Chapter 3

Microlocal Analysis

3.1 Microlocal Analysis

One of the fundamental interests of mathematical analysis is the study of the behaviour of a function near a given point, that is, we want to understand the local properties of a function. This proclivity towards local analysis is partly to do with the fact that classical notions such as continuity and differentiability of a function are essentially local properties. In other words, questions such as whether a function f is differentiable or not at a point x, and the value of the corresponding derivative, depend only on the values of f in an arbitrarily small neighbourhood of x. These and other classical notions have over the years been extended to a variety of contexts and reinterpretations that allow us to study increasingly complex mathematical structures.

Microlocal analysis is essentially the geometric theory of distributions which was first developed in the 1950s and 1960s in seminal works by Hörmander [19], Kohn, Maslov and Nirenberg [62] for the study of linear partial differential operators. Many of the ideas that underlie the study of microlocal analysis are motivated by the attempts to generalise pre-existing notions from classical analysis in order to investigate distributions and their singularities. Instead of studying general distributions (or general functions for that matter), we typically consider only the types of distributions that arise in the study of differential and integral equations. These problems include (but are not restricted to) pseudo-differential operators, Fourier integral operators, and oscillatory integral operators [19, 63, 64].

Distributions are usually defined by duality, and it is the nature of this duality that defines their behaviour; distributions that correspond to good functions have good properties, whereas distributions that correspond to bad functions typically have bad properties. Distributions we come across in reality are lie by and large somewhere in-between the two extremes. For example, *conormal distributions* have mostly good properties, excepts for flaws which are caused by their singularities. These singularities are the principal objects of study of microlocal analysis.

Let us take a step back and consider the ideas that underlie much of microlocal analysis. The basic ideas are based on the properties of the Fourier transform and its

relation with smoothness. If f is an integrable compactly supported function, we can tell whether it is in C_0^∞ by examining the behaviour of its Fourier transform $\hat{f}$ at infinity [19]. In other words, f is in C_0^∞ if and only if for all $N \in \mathbb{N}$ there exits a non-negative constant C_N such that

$$|\hat{f}(\boldsymbol{\xi})| \leqslant C_N \left(1 + \|\boldsymbol{\xi}\|\right)^{-N}, \quad \text{as } \|\boldsymbol{\xi}\| \to \infty. \tag{3.1}$$

Expression (3.1) essentially says that the information about the discontinuities of f is encoded in the high frequency components of $\hat{f}(\boldsymbol{\xi})$. We will revisit these ideas in Chapter 4. Similarly, let $u \in \mathcal{S}'(\Omega)$, where Ω is an open subset of $\mathbb{R}^d$. The classical argumentation (see Chapter VIII in [19]), which is related to the Paley-Wiener theorem [18], says that u is C^∞ in a neighbourhood of $\mathbf{x} \in \mathbb{R}^d$, if and only if there exists a test function $\phi \in C_c^\infty(\Omega)$ such that $\phi(\mathbf{x}) \neq 0$ and for all $N \in \mathbb{N}$ we have

$$|\widehat{u\phi}(\boldsymbol{\xi})| = |\langle u, e^{-2\pi\imath(\cdot)\boldsymbol{\xi}}\phi\rangle| \leqslant C_N \left(1 + \|\boldsymbol{\xi}\|\right)^{-N}, \quad \text{as } \|\boldsymbol{\xi}\| \to \infty. \tag{3.2}$$

Expressions such as (3.1) and (3.2) are called the *rapid decay* conditions. The set of all $\mathbf{x} \in \Omega$ for which (3.2) does not hold is known as the *singular support* of the distribution u (written as $\operatorname{sing\,supp} u$). Roughly speaking, singular support is the set of all points at which a distribution fails to be a smooth function. For example, the singular support of the Heaviside function, or of the Dirac delta function, is just the point 0, whereas the singular support of the characteristic function of region with smooth boundary is exactly the boundary of that region.

If a function or a distribution u fails to be smooth at a point, we can further analyse the singularity, that is, instead of locating the singularity only in space, we can also analyse its direction. We could then use the directions where $\hat{u}$ is not rapidly decaying to describe the high-frequency components of u that are causing the singularities. In other words, we end with some notion of directed smoothness of a distribution. The classical setup [19] for defining these notions goes as follows. We say that $\mathcal{V}$ is a *conic set* if for every $\boldsymbol{\xi} \in \mathcal{V}$ and every $t > 0$ we have $t\boldsymbol{\xi} \in \mathcal{V}$. Taking $\|\boldsymbol{\xi}\| = 1$ and $t > 0$, let us now rewrite the (3.2) as

$$|\widehat{u\phi}(t\boldsymbol{\xi})| = |\langle u, e^{-2\pi\imath t(\cdot)\boldsymbol{\xi}}\phi\rangle| \leqslant C_N(1+t)^{-N}. \quad t \to \infty, \tag{3.3}$$

In terms of (3.2), the expression (3.3) ought to hold or all $N \in \mathbb{N}$, and uniformly in $\|\boldsymbol{\xi}\| = 1$. Therefore, the regions of smoothness of a distribution u are uncovered by testing u against the oscillatory test function $e^{-2\pi\imath t\mathbf{x}\cdot\boldsymbol{\xi}}\phi(\mathbf{x})$, where ϕ is a smooth function that is localised in space, and examining the asymptotic behaviour in the Fourier domain. But we can also localise (3.3) in frequency, that is, localise in $\boldsymbol{\xi}$ where $\|\boldsymbol{\xi}\| = 1$. The idea is that a function can have rapid decay in some directions but not in other directions. Therefore, we want to find an open cone containing the desired set of directions in which the localised function has rapid decay, which indicates smoothness.

Adherence to conic sets allows us to restrict our attention to subsets of the co-sphere bundle $S^*(\mathbb{R}^d) = S^{d-1} \times \mathbb{R}^d$ of $\mathbb{R}^d$. Since we are concerned with the case $d = 2$ we can identify the elements of S^1 through just one parameter, the angle. This discussion leads to the classical definition of the wavefront set of a distribution.

Definition 3.1.1. *The* wavefront set *of a distribution u, denoted* $\mathrm{WF}(u)$*, is the complement of the set of all points $(\theta_0, \mathbf{x}_0)$ for which there exists a smooth window function $\phi \in C_0^\infty$, $\phi(\mathbf{x}_0) \neq 0$ and an open cone $\mathcal{C}$ such that $\theta_0 \in \mathcal{C}$, with the property that for all $N \in \mathbb{N}$*

$$|\widehat{u\phi}(\boldsymbol{\xi})| \leqslant C_N(1 + \|\boldsymbol{\xi}\|)^{-N}, \text{ for all } \boldsymbol{\xi} \in \mathcal{C}. \tag{3.4}$$

It is important to mention that Definition 3.1.1 does not depend on the choice of ϕ. Furthermore, note that $\mathbf{x} \cdot \boldsymbol{\xi} = const$ are the fronts of constant phase of the oscillating test function (that is, a wave) $e^{-2\pi\imath\mathbf{x}\cdot\boldsymbol{\xi}}\phi(\mathbf{x})$, which is part of the reason why the set we end up with is called the wavefront set.

There are many advantages in having the notion of the wavefront set. First of all (and perhaps least interestingly), it generalises the notion of singular support, that is, the projection of $\mathrm{WF}(u)$ onto its first coordinate is exactly the singular support

$$\operatorname{sing\,supp}(u) = \{\mathbf{x} : \exists\theta \text{ such that } (\theta, \mathbf{x}) \in \mathrm{WF}(u)\}.$$

Secondly, the notion of the wavefront set allows us to extend a number of operations on distributions. Notions such as the restriction to a manifold and the product of two distributions can be defined under certain conditions on the wavefront set of the given distributions [17, 19]. Thirdly, wavefront sets are particularly useful for the study of (pseudo)differential operators, their solutions and propagation of singularities [62].

Let us list the wavefronts sets of some very elementary distributions. Proofs of these properties can be found in [63, 65] and other texts.

- Dirac delta: Let δ be the Dirac delta distribution. Then

$$\mathrm{WF}(\delta) = \{(\theta, 0), \theta \in [0, 2\pi)\}.$$

- Heaviside function: Let H be the Heaviside distribution $H(\mathsf{x}) = \chi_{\mathsf{x}\geqslant 0}$. Then

$$\mathrm{WF}(H) = \{(\theta, 0), \theta \in [0, 2\pi)\}.$$

- Upper half-plane: Let u be the distribution associated with the indicator function of the upper half plane $\{\mathbf{x} \in \mathbb{R}^2 : \mathsf{x}_2 \geqslant 0\}$. Then

$$\mathrm{WF}(u) = \{(\pi/2, (\mathsf{x}, 0)) : \mathsf{x} \in \mathbb{R}\}.$$

- Region with a smooth surface: Let $\Omega \subseteq \mathbb{R}^d$ be a region with smooth boundary $\partial\Omega$, and let χ_Ω be the distribution associated with the indicator function of Ω. Then

$$\mathrm{WF}(\chi_\Omega) = \{(\theta, \mathbf{x}) : \mathbf{x} \in \partial\Omega, \theta \text{ normal to } \partial\Omega \text{ at } \mathbf{x}\}.$$

The last statement can be explained through a simple hand-waving argument. Using a test function with a very small support around $\mathbf{x} \in \partial\Omega$, the boundary around $\mathbf{x}$ looks like a line. We can then use the a modified version of the argument for the upper half plane to yield the result.

A concept related to the wavefront set is the microlocal Sobolev regularity, which we will now introduce.

Definition 3.1.2. *We say that a distribution f is microlocally in the L^2 Sobolev space H^k at $(\theta_0, \mathbf{x}_0)$, written $f \in H^k(\theta_0, \mathbf{x}_0)$, if for some smooth bump function $\phi \in C^\infty(\mathbb{R}^2)$, with $\phi(\mathbf{x}_0) \neq 0$, localised to a ball near $\mathbf{x}_0$, and for some smooth bump function $\beta \in C^\infty_{per}[0, 2\pi)$, obeying $\beta(\theta_0) = 1$ and localised to a ball near θ_0, the space/direction localised function $f_{\phi,\beta}$, defined in polar Fourier coordinates by $\beta(w)\widehat{\phi f}(r\cos(w), r\sin(w))$ belongs to the weighted L^2 space $L^2((1 + \|\xi\|^2)^{k/2} d\xi)$.*

3.2 Parabolic Molecules and Microlocal Analysis

As we mentioned in previous chapters, what we ultimately want to do is to use the framework that we have developed in Chapter 2 to say something about general properties of continuous parabolic molecules. More specifically, we want to show that there are universality-type results in microlocal analysis that hold for all admissible families of parabolic molecules.

The motivation behind this is two-fold. Firstly, some families within the vast array of directional representation systems have an elegant mathematical structure that allows for a painless theoretical analysis and inference of properties, whereas other systems are better suited for digital realisations or other tasks in applications. Therefore, in a properly constructed general framework that encompasses both types of families, we could establish various properties using the representation dictionary that provides the simplest, or most elegant, proofs and then extend these results to all other systems using the tools developed in our framework. Secondly, the currently existing proofs of various features of directional representation systems are often very similar, but have to yield ground to specifics of each construction. By generalising the framework we can distil the important qualities of directional systems, instead of focusing on their individual characteristics, thus avoiding the repetitiveness of proof techniques. It is needless to say that the idea abstractifying pre-existing models has a rich history in mathematics. Discussions of this type are echoes of the ideology argued in [9].

As it often is the case when dealing with frames, in order to analyse the properties of functions we will look at their corresponding frame coefficients. Therefore, in view of what we are trying to achieve, we need tools that enable bridging the gap between statements regarding frame coefficients of one family and statements regarding frame coefficients of some other family of CPMs. To that end, we will use reconstruction formulas of the type (1.7), where we will distinguish between two cases. The first case is that of dictionaries which admit a reconstruction formula which is valid for functions without any restrictions on the support of their Fourier transform. Reconstruction formulas of this type include for example the curvelets reconstruction formula (1.12) or the shearlets reconstruction formula (2.28). The other case concerns dictionaries that admit a reconstruction formula which is valid only for functions whose Fourier transform is supported inside some cone in the frequency plane.

Let us summarise the main results that will be argued in this chapter. Assume $\Gamma = \{m_\lambda : \lambda \in \Lambda_\Gamma\}$ and $\Sigma = \{n_\nu : \nu \in \Lambda_\Sigma\}$ are two families of CPMs of sufficiently high order and consider an angle-location pair $(\theta_0, \mathbf{x}_0)$. We will show that if there is an open

neighbourhood $\mathcal{M}$ of $(\theta_0, \mathbf{x}_0)$ such that $a_\nu^{-k}\langle f, n_\nu\rangle$ is in $L^p([0, a_0] \times \mathcal{M}, d\mu)$, then there is another open neighbourhood $\mathcal{N}$ of $(\theta_0, \mathbf{x}_0)$ such that the same holds for Γ, that is, such that $a_\lambda^{-k}\langle f, m_\lambda\rangle$ is in $L^p([0, a_0] \times \mathcal{N}, d\mu)$. This will hold for both the CPMs that admit a general reconstruction formula (Theorem 3.2.3), and for CPMs that admit a reconstruction formula which is valid for cone-supported functions (Theorem 3.2.8). The latter case requires additional assumptions to ensure that the contributions outside of the cone of interest can be readily neglected.

What is important is that these results help to uncover important notions of microlocal analysis. For example, it was established in [15] and [66] that the decay of curvelet, that is shearlet, coefficients allows the detection of the wavefront set of a function, which in our framework corresponds to the case $p = \infty$ (Theorem 3.2.7). On the other hand, microlocal Sobolev regularity can be inferred from a condition on curvelet frame coefficients [15], which here corresponds to the case $p = 2$ (Theorem 3.2.5). Therefore, our results generalise these notions and create a broader perspective from which we can get novel results regarding the ability of various parabolic dictionaries to uncover microlocal features of functions, since we consider all $1 \leqslant p \leqslant \infty$.

3.2.1 General Representation Formulas

Let us get back to the task at hand. It is well known that using the rapid decay of the standard two-dimensional continuous wavelet transform (for example, as described in Section 1.7) one can determine the singular support of a distribution [67]. Unfortunately, it is clear that due to the wavelets' lack of directional sensitivity we will not be able to detect the wavefront set, and thus examine the directions that are causing the singularities. It can even be shown that if we were to adjust the standard wavelet transform by adding the directionality through

$$\psi_{a\,\theta\,\mathbf{b}}(\mathbf{x}) = a^{-1}\psi\left(\mathrm{R}_\theta \frac{\mathbf{x} - \mathbf{b}}{a}\right),$$

the resulting transform will have slow decay at all directions θ for a given singularity $\mathbf{b}$ [1]. Thus, the addition of an orientation parameter has not proved to be particularly useful since it did not provide any further information about the directional structure of the singularity.

It turns out that coupling the directional selectivity with the anisotropic (in this case, parabolic) treatment of the axes is the right recipe to resolve these questions. This has been established in [15] and [66], in case of very specific curvelet and shearlet families. We shall for the moment focus on the microlocal Sobolev regularity, which we defined in Definition 3.1.2, and we will then turn our attention to the wavefront set. Candés and Donoho showed in [15] that the notion of microlocal Sobolev regularity can be determined by an L^2 condition, as stated in the following theorem.

Theorem 3.2.1. *Let $S_2^k(\theta, \mathbf{x})$ denote the (normal-approach, parabolic scaling) square function*

$$S_2^k(\theta, \mathbf{x}) = \left(\int_0^{a_0} |\langle f, \gamma_{a\,\theta\,\mathbf{x}}\rangle|^2 a^{-2k} \frac{da}{a^3}\right)^{1/2}, \tag{3.5}$$

where $\{\gamma_{a\,\theta\,\mathbf{x}}, a \in [0, a_0], \theta \in [0, 2\pi), \mathbf{x} \in \mathbb{R}^2\}$ *are second generation curvelets.*

The distribution f *is in* $H^k(\theta_0, \mathbf{x}_0)$ *if and only if for some neighbourhood* $\mathcal{N}$ *of* $(\theta_0, \mathbf{x}_0)$ *we have*

$$\int_{\mathcal{N}} (S_2^k(\theta, \mathbf{x}))^2 d\theta d\mathbf{x} < \infty.$$

The case we want to argue throughout this thesis is that the results such as Theorem 3.2.1 are not restricted to any particular representation family, but that they are rather a part of a much broader framework. To our knowledge this result has up to now been shown for only curvelets, though it stands to reason that this should not be the case.

A slightly different reading of Theorem 3.2.1 would be to interpret it as

$$f \text{ is in } H^k(\theta_0, \mathbf{x}_0) \quad \text{if and only if} \quad a^{-k}\langle f, \gamma_{a\,\theta\,\mathbf{x}}\rangle \in L^2([0, a_0] \times \mathcal{N}, \mu), \tag{3.6}$$

for some open neighbourhood $\mathcal{N}$ of $(\theta_0, \mathbf{x}_0)$. This is the interpretation we will use. Our goal now is to extend this result, and other similar results, to all admissible CPM dictionaries. To begin, take $\Gamma = \{m_\lambda : \lambda \in \Lambda_\Gamma\}$ and $\Sigma = \{n_\nu : \nu \in \Lambda_\Sigma\}$ to be two families of CPMs, with parameterisations $(\Lambda_\Gamma, \Phi_\Gamma)$ and $(\Lambda_\Sigma, \Phi_\Sigma)$, respectively, and denote their inter-Gramian by

$$G_{\Gamma,\Sigma}(\lambda, \nu) = \langle n_\nu, m_\lambda\rangle. \tag{3.7}$$

As we have mentioned earlier, reconstruction formulas of representation systems will play a crucial role in converting statements such as (3.6) to other CPMs. Reconstruction formulas of high-frequency functions are generally of the form

$$f = \int_{\Lambda_\Sigma} \langle f, n_\nu\rangle \tilde{n}_\nu d\mu(\nu), \tag{3.8}$$

which is valid in (at least) the weak sense. The functions $\{\tilde{n}_\nu : \nu \in \Lambda_\Sigma\}$ in (3.8) denote the elements of the dual frame, which we will also require to be a family of CPMs. We have already encountered such reconstruction formulas in Sections 2.3.1 and 2.3.4, though many other examples can be found, for example in [41]. The idea now is to use (3.8) to establish a relationship between frame coefficients associated to the CPMs Γ and Σ. Taking the inner product of (3.8) with m_λ we immediately have

$$a_\lambda^{-k}\langle f, m_\lambda\rangle = a_\lambda^{-k} \int_{\Lambda_\Sigma} \langle f, n_\nu\rangle\langle \tilde{n}_\nu, m_\lambda\rangle d\mu(\nu). \tag{3.9}$$

Therefore, to get a result analogous that would be analogous to Theorem 3.2.1, we would need to get bounds on the weighted L^2 norm of the coefficients in (3.9). We can actually show a bit more, namely, that we can consider not only the L^2 norm but rather any L^p norm, for $p \in [1, \infty]$.

Let us choose an arbitrary angle-location pair $(\theta_0, \mathbf{x}_0)$. We can notice that the integral in (3.9) can be split up in two parts; an integral over $[0, a_0] \times \mathcal{N}$, where $\mathcal{N}$ is an open neighbourhood around $(\theta_0, \mathbf{x}_0)$, and an integral over $[0, a_0] \times \mathcal{N}^c$. The first of these integrals is directly related to the square function S_2^k through a bounded integral operator.

Lemma 3.2.2. *Let $\Gamma = \{m_\lambda : \lambda \in \Lambda_\Gamma\}$ and $\Sigma = \{n_\nu : \nu \in \Lambda_\Sigma\}$ be two families of continuous parabolic molecules and take $N \in \mathbb{N}$ as given in Theorem 2.4.1, and let $\mathcal{N}$ and $\mathcal{M}$ be two open and bounded subsets of $[0, 2\pi) \times \mathbb{R}^2$. The operator $\mathsf{T} : L^p([0, a_0] \times \mathcal{N}, \mu) \to L^p([0, a_0] \times \mathcal{M}, \mu)$, where $p \in [1, \infty]$, defined via*

$$(\mathsf{T}u)(\nu) = \int_{[0,a_0]\times\mathcal{N}} \left(\frac{a_\lambda}{a_\nu}\right)^k G_{\Gamma,\Sigma}(\lambda, \nu) u(\lambda) d\mu(\lambda),$$

where $G_{\Gamma,\Sigma}$ is the inter-Gramian defined in (3.7), is bounded (in L^p) provided the parameterisations $(\Lambda_\Gamma, \Phi_\Gamma)$ and $(\Lambda_\Sigma, \Phi_\Sigma)$ of Γ and Σ are $N - k$ admissible.

Proof. In order to show the boundedness of T we will use Schur test (see for example Theorem 5.6 in [68]) which says that T is bounded in the induced operator norm, and the bound is given by

$$\|\mathsf{T}\| \leqslant \left[\sup_\nu \int \left(\frac{a_\lambda}{a_\nu}\right)^k |G_{\Gamma,\Sigma}(\lambda, \nu)| d\mu(\lambda)\right]^{1/p} \left[\sup_\lambda \int \left(\frac{a_\lambda}{a_\nu}\right)^k |G_{\Gamma,\Sigma}(\lambda, \nu)| d\mu(\nu)\right]^{(p-1)/p},$$

provided that the right hand side of the expression is finite. To show that these integrals are indeed bounded we will use Theorem 2.4.1, which gives bounds on the integral kernel $G_{\Gamma,\Sigma}(\lambda, \nu)$. Therefore, for every $\nu \in \Lambda_\Sigma$ we have

$$\begin{aligned}\int_{[0,a_0]\times\mathcal{N}} &\left(\frac{a_\lambda}{a_\nu}\right)^k |G_{\Gamma,\Sigma}(\lambda, \nu)| d\mu(\lambda) \\ &\lesssim \int_{[0,a_0]\times\mathcal{N}} \left(\frac{a_\lambda}{a_\nu}\right)^k \left(\frac{a_M}{a_m}\right)^{-N} \left(1 + a_M^{-1} d(\lambda, \nu)\right)^{-N} d\mu(\lambda).\end{aligned}$$

Since $a_m \leqslant a_\lambda, a_\nu \leqslant a_M$ hold trivially and since

$$\left(1 + a_M^{-1} d(\lambda, \nu)\right)^{-N} \leqslant \left(1 + a_M^{-1} d(\lambda, \nu)\right)^{-(N-k)},$$

holds for all $k \geqslant 0$ it follows

$$\begin{aligned}\int_{[0,a_0]\times\mathcal{N}} \left(\frac{a_\lambda}{a_\nu}\right)^k |G_{\Gamma,\Sigma}(\lambda, \nu)| d\mu(\lambda) &\lesssim \int_{[0,a_0]\times\mathcal{N}} \left[\frac{a_M}{a_m}\left(1 + a_M^{-1} d(\lambda, \nu)\right)\right]^{-(N-k)} d\mu(\lambda) \\ &\leqslant \int_{\Lambda_\Gamma} \omega(\lambda, \nu)^{-(N-k)} d\mu(\lambda) \qquad (3.10) \\ &< \infty.\end{aligned}$$

The boundedness of the last expression follows from the admissibility of the parametrisation. The other integral is treated analogously. Hence, T is a bounded operator. □

We are now ready to prove our first universality-type result. It will allow us to infer what we claimed: assuming the frame coefficients of one CPM family are in a certain weighted L^p space, then the frame coefficients of all other admissible CPM families lie in an analogous weighted L^p space. In order for the proof to work we need a further technical assumption on the parametrisation mapping Φ_Σ, namely, we require $(\Phi_\Sigma)^{-1}$ to have a uniformly bounded Jacobian.

Theorem 3.2.3. *Let $k \in \mathbb{N}$ and take $\Gamma = \{m_\lambda : \lambda \in \Lambda_\Gamma\}$ and $\Sigma = \{n_\nu : \nu \in \Lambda_\Sigma\}$ to be two families of continuous parabolic molecules, with parameterisations $(\Lambda_\Gamma, \Phi_\Gamma)$ and $(\Lambda_\Sigma, \Phi_\Sigma)$, such that Σ admits a reproduction formula of the form (3.8) and the corresponding Plancherel-like formula, that the Jacobian of Φ_Σ^{-1} is uniformly bounded and that the conditions of Lemma 3.2.2 are satisfied. Also, we assume that the dual frame $\tilde{\Sigma} = \{\tilde{n}_\nu : \nu \in \Lambda_\Sigma\}$ is also a CPM family. Take $p \in [1, \infty]$. Then if for some open and bounded neighbourhood $\mathcal{N}$ of $(\theta_0, \mathbf{x}_0)$*

$$a^{-k}_{\Phi_\Sigma^{-1}(\cdot)} \langle f, n_{\Phi_\Sigma^{-1}(\cdot)} \rangle \in L^p\left([0, a_0] \times \mathcal{N}, d\mu\right),$$

holds then

$$a^{-k}_{\Phi_\Gamma^{-1}(\cdot)} \langle f, m_{\Phi_\Gamma^{-1}(\cdot)} \rangle \in L^p\left([0, a_0] \times \mathcal{M}, d\mu\right),$$

holds for some open and bounded neighbourhood $\mathcal{M}$ of $(\theta_0, \mathbf{x}_0)$.

To reduce the notational load we will write a_ν instead of $a_{\Phi_\Sigma{}^{-1}(\nu)}$. Analogous abbreviations will be used in other cases.

Proof. Take $\mathcal{M}$ to be an open and bounded neighbourhood of $(\theta_0, \mathbf{x}_0)$ such that $\operatorname{dist}(\mathcal{M}; \mathcal{N}^c) > 0$. We have

$$a_\lambda^{-k} \langle f, m_\lambda \rangle = a_\lambda^{-k} \int_{\Lambda_\Sigma} \langle f, n_\nu \rangle \langle \tilde{n}_\nu, m_\lambda \rangle d\mu(\nu) = A_1 + A_2,$$

where

$$\begin{aligned} A_1 &= a_\lambda^{-k} \int_{\Phi_\Sigma^{-1}([0,a_0]\times\mathcal{N})} \langle f, n_\nu \rangle \langle \tilde{n}_\nu, m_\lambda \rangle d\mu(\nu), \\ A_2 &= a_\lambda^{-k} \int_{\Phi_\Sigma^{-1}([0,a_0]\times\mathcal{N}^c)} \langle f, n_\nu \rangle \langle \tilde{n}_\nu, m_\lambda \rangle d\mu(\nu). \end{aligned}$$

For A_2 we have

$$\begin{aligned} A_2 &= a_\lambda^{-k} \int_{\Phi_\Sigma^{-1}([0,a_0]\times\mathcal{N}^c)} \langle f, n_\nu \rangle \langle \tilde{n}_\nu, m_\lambda \rangle d\mu(\nu) \\ &\lesssim a_\lambda^{-k} \int_{[0,a_0]\times\mathcal{N}^c} \langle f, n_\nu \rangle \langle \tilde{n}_\nu, m_\lambda \rangle d\mu(\nu) \\ &\lesssim a_\lambda^{-k} \|f\|_{L^2(\mathbb{R}^2)} \left(\int_{[0,a_0]\times\mathcal{N}^c} \omega(\nu, \lambda)^{-2N} d\mu(\nu) \right)^{1/2}, \end{aligned}$$

where we used the boundedness of the Jacobian in the second line, and Theorem 2.4.1 in the third line. We can now write

$$\omega(\nu, \lambda) = \frac{a_M}{a_m} \left(1 + a_M^{-1} d(\nu, \lambda)\right) \geqslant a_m^{-1} \left(|\theta_\lambda - \theta_\nu|^2 + \|\mathbf{b}_\lambda - \mathbf{b}_\nu\|^2\right),$$

to get

$$A_2 \lesssim a_\nu^{-k} \left(\int_0^{a_0} a_m^{2N} \frac{da}{a^3} \right)^{1/2} \left(\int_{\mathcal{N}^c} \left(|\theta_\lambda - \theta_\nu|^2 + \|\mathbf{b}_\lambda - \mathbf{b}_\nu\|^2 \right)^{-2N} \right)^{1/2}.$$

The integral

$$\left(\int_{\mathcal{N}^c} \left(|\theta_\lambda - \theta_\nu|^2 + \|\mathbf{b}_\lambda - \mathbf{b}_\nu\|^2 \right)^{-2N} \right)^{1/2},$$

is clearly uniformly bounded as long as $\operatorname{dist}(\mathcal{M}; \mathcal{N}^c) \geqslant \epsilon > 0$. It follows

$$A_2 \lesssim a_\lambda^{-k} \left(\int_0^{a_0} a_m^{2N} \frac{da}{a^3} \right)^{1/2} \lesssim a_\lambda^{N-1-k}.$$

In other words, $A_2 = A_2(a_\lambda)$ will have a finite $L^p([0, a_0] \times \mathcal{M}, \mu)$ norm provided $N \geqslant k$ for $p = \infty$, and $N > \frac{2}{p} + k + 1$, for $p \in [1, \infty)$.

Turning our attention to A_1, notice first that due to the boundedness of the Jacobian we have

$$A_1 \lesssim (\mathsf{T}u)(\lambda),$$

with T as in Lemma 3.2.2 and $u(\lambda) = a_\lambda^{-k} \langle f, n_\nu \rangle$ and $\tilde{G}_{\Gamma,\Sigma} = \langle \tilde{n}_\nu, m_\lambda \rangle$. Hence, by Lemma 3.2.2, it follows that A_1 also has a finite $L^p([0, a_0] \times \mathcal{M}, \mu)$ norm for $p \in [1, \infty]$ and the statement follows.

□

Remark 3.2.4. *Notice that the argument for the boundedness of A_2 can be also seen as trying to find a bound for the operator $\tilde{\mathsf{T}}$, defined by*

$$\left(\tilde{\mathsf{T}} u \right)(\lambda) = \int_{[0,a_0] \times \mathcal{N}^c} \left(\frac{a_\nu}{a_\lambda} \right)^k G_{\Sigma,\Gamma}(\nu, \lambda) u(\nu) d\mu(\nu).$$

In that case, we would need a further assumption, namely, we need $a_\nu^k u(\nu)$ to be in L^2. Alternatively, we may assume $a_\nu^k u(\nu)$ is in L^q, for some q. This would give $N > \frac{2}{p} + \frac{2}{q'} + k$, where q' is such that $1/q + 1/q' = 1$.

From the proof of Theorem 3.2.3 it is clear why is it important to have a family of CPMs such that its dual frame is also a CPM family, since we can use almost-orthogonality of both frames. Another reason is that it allows us to go both ways, that is, assuming that the dual frames of both CPM families Γ and Σ, are also CPMs we could replace *if ... then* in the statement of Theorem 3.2.3 with *if and only if*.

Depending on the choice of p, we can now apply Theorem 3.2.3 to different situations. For the first application we will address the question of the universality of microlocal Sobolev regularity with respect to CPMs. In other words, we will consider $p = 2$. The following theorem says that we can infer whether f is in $H^k(\theta_0, \mathbf{x}_0)$ by looking at the L^2 condition on S_2^k, of the type (3.5), with respect to not just curvelets but also other families of continuous parabolic molecules.

Theorem 3.2.5. *Let $\Sigma = \{n_\nu : \nu \in \Lambda_\Sigma\}$ be a family of CPMs of order (R, M, N_1, N_2), that satisfies the assumptions of Theorem 3.2.3. Then f is in $H^k(\theta_0, \mathbf{x}_0)$ if and only if*

$$a^{-k}_{\Phi_\Sigma^{-1}(\cdot)} \langle f, n_{\Phi_\Sigma^{-1}(\cdot)} \rangle \in L^2\left([0, a_0] \times \mathcal{N}, \mu\right), \tag{3.11}$$

for some open neighbourhood $\mathcal{N}$ of $(\theta_0, \mathbf{x}_0)$.

Proof. Assume (3.11) holds. Take $\Gamma = \{m_\lambda : \lambda \in \Lambda_\Gamma\}$ to be a family of second generation curvelets. It follows by Proposition 2.3.4 that Γ is a family of CPMs of order $(R, R, R/2, R/2)$, where we can take any R. Applying now Theorem 3.2.3, it follows that

$$a^{-k}_{\Phi_\Gamma^{-1}(\cdot)} \langle f, m_{\Phi_\Gamma^{-1}(\cdot)} \rangle \in L^p\left([0, a_0] \times \mathcal{M}, \mu\right)$$

holds for some neighbourhood $\mathcal{M}$ of $(\theta_0, \mathbf{x}_0)$. Hence, the claim follows by Theorem 3.2.1. The converse follows trivially since if f is in $H^k(\theta_0, \mathbf{x}_0)$ then by Theorem 3.2.1 we have

$$a^{-k} \langle f, \gamma_{a\,\theta\,\mathbf{x}} \rangle \in L^2\left([0, a_0] \times \mathcal{N}, \mu\right).$$

The parametrisation mapping for curvelets is the identity. Hence, the requisite Jacobian is clearly bounded and we can apply Theorem 3.2.3, yielding the statement. □

We can be more specific and take Σ to be shearlets. Having Theorem 3.2.3 in mind, we need to ensure that the shearlet family we will use admits a representation formula.

Corollary 3.2.6. *Let Σ denote a family of cone-adapted, band-limited shearlets which are a family of CPMs (look at [60, 69] or at Section 2.3.4). Then $f \in H^k(\theta_0, \mathbf{x}_0)$ if and only if*

$$\int_{\mathcal{N}_s} \left(\int_0^{a_0} |\langle f, \sigma_{as\mathbf{x}} \rangle|^2 a^{-2k} \frac{da}{a^3} \right) ds d\mathbf{x} < \infty$$

where $\mathcal{N}_s$ is some neighbourhood of $(s_0, \mathbf{x}_0)$ and s_0 is the corresponding shearing parameter.

Proof. Without loss of generality we can assume $\theta \in \left(\frac{-\pi}{4}, \frac{\pi}{4}\right) \cup \left(\frac{3\pi}{4}, \frac{5\pi}{4}\right)$. The other case is analogous. This assumption equates to $s = -\tan(\theta_0) \in (-1, 1)$, which means that there exists $\varepsilon > 0$ such that

$$(s_0 - \varepsilon, s_0 + \varepsilon) \subseteq (-1, 1)$$

and we can use only the horizontal shearlets for the analysis (this helps to simplify the expression for the parametrisation). Therefore, by Theorem 2.3.13, the corresponding shearlet system is a family of CPMs of arbitrary order, and is admissible for all $k > 2$. Furthermore, the Jacobian of Φ_Σ^{-1} is equal to $\frac{-2}{\cos(2\theta)+1}$, which is uniformly bounded for θ in $\left(\frac{-\pi}{4}, \frac{\pi}{4}\right) \cup \left(\frac{3\pi}{4}, \frac{5\pi}{4}\right)$. Thus, by Theorem 3.2.5 it suffices to show that

$$a^{-k}_{\Phi_\Sigma^{-1}(\cdot)} \langle f, \sigma_{\Phi_\Sigma^{-1}(\cdot)} \rangle \in L^2\left([0, a_0] \times \mathcal{N}, \mu\right).$$

Without loss of generality we can assume that $\mathcal{N}_s = (s_0 - \varepsilon, s_0 + \varepsilon) \times \mathcal{B}_\varepsilon(\mathbf{x}_0)$. Let us define

$$\mathcal{N} = \arctan(s_0 - \varepsilon, s_0 + \varepsilon) \times \mathcal{B}_\varepsilon(\mathbf{x}_0)$$

and notice that $\mathcal{N}$ is an open neighbourhood of $(\theta_0, \mathbf{x}_0)$. It follows

$$\begin{aligned}
\left\| a^{-k}_{\Phi_\Sigma^{-1}(\cdot)} \langle f, \sigma_{\Phi_\Sigma^{-1}(\cdot)} \rangle \right\|^2_{L^2([0,a_0]\times\mathcal{N}, d\mu)} &= \int_{\mathcal{N}} \int_0^{a_0} \left| \langle f, \sigma_{\Phi_\Sigma^{-1}(a\theta\mathbf{x})} \rangle \right|^2 a^{-2k} \frac{da}{a^3} d\theta d\mathbf{x} \\
&= \int_{\mathcal{N}_s} \int_0^{a_0} |\langle f, \sigma_{as\mathbf{x}} \rangle|^2 \, a^{-2k} \, |\det J\Phi_\Sigma(as\mathbf{x})| \, \frac{da}{a^3} ds d\mathbf{x} \\
&\lesssim \int_{\mathcal{N}_s} \int_0^{a_0} |\langle f, \sigma_{as\mathbf{x}} \rangle|^2 \, a^{-2k} \frac{da}{a^3} ds d\mathbf{x} < \infty.
\end{aligned}$$

Here we used the fact that $|J\Phi_\Sigma(a, s, \mathbf{x})| = \left| \frac{1}{1+s^2} \right|$ is uniformly bounded. Therefore, by Theorem 3.2.5 the claim follows. □

It is important to note at this point that the argument used in the proofs of Corollary 3.2.6 and Theorem 3.2.5 would not go through for the non cone-adapted shearlets. The reason behind this lies in the fact the Jacobian of Φ_Σ^{-1} for non cone-adapted shearlets is not uniformly bounded on $[0, 2\pi)$. This agrees with the intuition, since regular shearlets exhibit a directional bias in the sense that the singularities on the y-axis can only be resolved as the shearing parameter tends to infinity. Cone-adapted shearlets on the other hand, always take the shearing parameters from a bounded set, which, in terms of Theorem 3.2.3, means that the corresponding Jacobian will necessarily be bounded.

For the second application of the Theorem 3.2.3 we will look at the resolution of the wavefront set. It is has been established that curvelets and shearlets can both resolve the wavefront set [15, 40] by detecting the locations were the frame coefficients are of fast decay with respect to the dilation parameter. In the case of shearlets this means that the wavefront set of f is given as the complement of pairs $(s_0, \mathbf{x}_0)$ for which $\langle f, \psi_{a\,s\,\mathbf{x}} \rangle \lesssim a^N$ for all $N \in \mathbb{N}$, as $a \to 0$, and pairs $(s, \mathbf{x})$ in some neighbourhood of $(s_0, \mathbf{x}_0)$. Let us now generalise this fact within our framework. In terms of Theorem 3.2.3 this corresponds to taking p to be ∞.

Theorem 3.2.7. *Let $\Sigma = \{n_\nu : \nu \in \Lambda_\Sigma\}$ be a family of continuous parabolic molecules of order (R, M, N_1, N_2), and $(\Phi_\Sigma, \Lambda_\Sigma)$ its parametrisation, satisfying the conditions of Theorem 3.2.3. The wavefront set of f is the complement of*

$$\mathrm{R}_\Sigma = \Bigg\{ (\theta_0, \mathbf{x}_0) : \text{ for all } k \in \mathbb{N} \text{ we have } |\langle f, n_{\Phi_\Sigma^{-1}(a,\theta,\mathbf{x})} \rangle| = \mathcal{O}(a^k) \text{ as } a \to 0, \\ \text{for some neighbourhood } \mathcal{N} \text{ of } (\theta_0, \mathbf{x}_0) \Bigg\}. \tag{3.12}$$

Proof. The condition in the definition of R_Σ says that $(\theta_0, \mathbf{x}_0)$ is in R_Σ provided

$$a^{-k} \langle f, n_{\Phi_\Sigma^{-1}(a,\theta,\mathbf{x})} \rangle \in L^\infty \left([0, a_\epsilon] \times \mathcal{N}, \mu \right), \text{ for all } k \in \mathbb{N},$$

for some open neighbourhood $\mathcal{N}$ of $(\theta_0, \mathbf{x}_0)$ and $0 < a_\epsilon < a_0$. On the other hand, by [15], second generation curvelets have the property that $\mathrm{WF}(f)$ is the complement of R_Γ, where by Γ we denote the family of second generation curvelets, and R_Γ is defined analogously to (3.12). We can rewrite this as

$$a^{-k}\langle f, \gamma_{a\,\theta\,\mathbf{x}}\rangle \in L^\infty\left([0, a_\epsilon] \times \mathcal{N}, \mu\right).$$

Hence, the claim follows by Theorem 3.2.3. □

3.2.2 Representation Systems for Cone-Supported Functions

Results of the previous section relied heavily upon the requirement that we have CPMs with good reconstruction formulas. To be more precise, we needed families of parabolic molecules that can reconstruct functions whose Fourier transform has a support which covers (possibly) the entire frequency plane. However, there are families of functions that admit a representation formula which is valid only for functions with frequency support inside a certain cone. Thus, the approach described in the previous subsection cannot be immediately applied because we have no means of controlling the decay of the frame coefficients outside of the cone in which the representation formula is valid.

Our goal is to show that we can work around this problem and that analysis of the same type is still applicable. In order to get a grip on those bounds we will need stronger assumptions, as the conditions (2.3) are not sufficient.

Let us begin by describing the situation at hand. As was the case in previous chapters, we will be concerned with only the high frequency case. Define the cones $\mathcal{C}_{u,v}$ and $\mathcal{C}^c_{u,v}$ in the frequency domain by

$$\mathcal{C}_{u,v} = \left\{\boldsymbol{\xi} \in \mathbb{R}^2 : |\xi_1| \geqslant u, \left|\frac{\xi_2}{\xi_1}\right| \leqslant v\right\},$$

and

$$\mathcal{C}^c_{u,v} = \left\{\boldsymbol{\xi} \in \mathbb{R}^2 : |\xi_2| \geqslant u, \left|\frac{\xi_1}{\xi_2}\right| \leqslant v\right\},$$

where $u, v > 0$. Same as in Chapter 2, for a set $\mathcal{D} \subseteq \mathbb{R}^2$ we define the corresponding L^2 space

$$L^2(\check{\mathcal{D}}) = \{f \in L^2(\mathbb{R}^2) : \operatorname{supp} \hat{f} \subseteq \mathcal{D}\}.$$

The standard situation is when $u = v = 1$. In that case let us denote $\mathcal{C} := \mathcal{C}_{u,v}$ and $\mathcal{C}^c := \mathcal{C}^c_{u,v}$. Furthermore, we will denote by $\mathsf{P}_{\mathcal{C}}$ and $\mathsf{P}_{\mathcal{C}^c}$ the respective frequency-domain projections onto the cones $\mathcal{C}$ and $\mathcal{C}^c$. In the rest this section we will consider a point $(\theta_0, \mathbf{x}_0)$ and its open and bounded neighbourhoods $\mathcal{N}$ and $\mathcal{M}$, whose closures are contained inside the cone $\mathcal{C}$. The analysis for general $\mathcal{C}_{u,v}$ is analogous.

Let us assume that we have two families of parabolic molecules. The first CPM family, $\Gamma = \{m_\lambda : \lambda \in \Lambda_\Gamma\}$, admits a representation formula

$$g = \int_{\Lambda_\Gamma} \langle g, m_\lambda\rangle \tilde{m}_\lambda d\mu(\lambda), \tag{3.13}$$

which holds for $g \in L^2(\check{\mathcal{C}})$, at least in the weak sense. In the following we will assume that we have a microlocal result which holds for Γ and what we want is to extend this result to some other CPM family $\Sigma = \{n_\nu : \nu \in \Lambda_\Sigma\}$.

An arbitrary function $f \in L^2(\mathbb{R}^2)$ can now be partitioned as

$$f = \mathsf{P}_{\mathcal{C}} f + \mathsf{P}_{\mathcal{C}^c} f.$$

The term $\mathsf{P}_{\mathcal{C}} f$ can be treated in a manner which is entirely analogous to the discussion in the previous section, where we had a representation formula which was valid on the entire plane $\mathbb{R}^2$. Hence, the problem lies in bounding the frame coefficients for the projection on $\mathcal{C}^c$, that is, outside of the cone in which we can represent f using frame coefficients given through the members of Γ. To do that we will require some further assumptions on the CPM family Σ. Let us recall the Definition 2.2.2, which gives

$$n_\nu(\boldsymbol{\xi}) = a_\nu^{-3/4} \varphi^{(\nu)} \left(\mathrm{D}_{1/a_\nu} \mathrm{R}_{\theta_\nu} (\mathbf{x} - \mathbf{b}_\nu) \right).$$

The first assumption we need is that the functions $\varphi^{(\nu)}$ have M vanishing moments in x_1 direction, which is equivalent to $\hat{\varphi}^{(\nu)}(\boldsymbol{\xi}) = \xi_1^M \hat{\rho}^{(\nu)}(\boldsymbol{\xi})$, with $\rho^{(\nu)} \in L^2(\mathbb{R}^2)$. We also require a Sobolev condition on $\varphi^{(\nu)}$, namely, $\frac{\partial^L}{\partial^L \mathsf{x}_2} \varphi^{(\nu)} \in L^2(\mathbb{R}^2)$. The last assumption we need is that all the L^2 norms are uniformly bounded as we go over all indices ν. Standard parabolic dictionaries such as curvelets, shearlets and a most other systems of note satisfy these boundedness assumption trivially, since they either have only finitely many generators (shearlets), or they have generators which are simple variations of one another (curvelets).

Let us briefly summarise the assumptions.

(i)
$$\hat{\varphi}^{(\nu)}(\boldsymbol{\xi}) = \xi_1^M \hat{\rho}^{(\nu)}(\boldsymbol{\xi}), \text{ with } \sup_\nu \|\hat{\rho}^{(\nu)}\|_{L^2(\mathbb{R}^2)} < \infty. \tag{3.14}$$

(ii)
$$\frac{\partial^L}{\partial^L \mathsf{x}_2} \varphi^{(\nu)} \in L^2, \text{ with } \sup_\nu \left\| \frac{\partial^L}{\partial^L \mathsf{x}_2} \varphi^{(\nu)} \right\|_{L^2(\mathbb{R}^2)} < \infty. \tag{3.15}$$

We are now ready to state and prove our result.

Theorem 3.2.8. *Let $\Gamma = \{m_\lambda : \lambda \in \Lambda_\Gamma\}$ and $\Sigma = \{n_\nu : \nu \in \Lambda_\Sigma\}$ be two families of continuous parabolic molecules with parameterisations $(\Phi_\Gamma, \Lambda_\Gamma)$ and $(\Phi_\Sigma, \Lambda_\Sigma)$. Furthermore, assume the family Σ satisfies the assumptions (3.14)-(3.15) for L and M for which there exists $1/2 < \alpha < 1$ such that*

$$L(\alpha - 1/2) \geqslant k, \text{ and } M(1 - \alpha) \geqslant k,$$

and that Γ admits a reproduction formula of the form (3.13) for functions $g \in L^2(\check{\mathcal{C}})$. Take a function $f \in L^2(\mathbb{R}^2)$. Then if for some open and bounded neighbourhood $\mathcal{N}$ of $(\theta_0, \mathbf{x}_0)$ we have

$$a_{\Phi_\Gamma^{-1}(\cdot)}^{-k} \langle f, m_{\Phi_\Gamma^{-1}(\cdot)} \rangle \in L^p\left([0, a_0] \times \mathcal{N}, d\mu\right)$$

then

$$a^{-k}_{\Phi^{-1}_{\Sigma}(\cdot)}\langle f, n_{\Phi^{-1}_{\Sigma}(\cdot)}\rangle \in L^p\left([0,a_0]\times\mathcal{M}, d\mu\right),$$

holds for some neighbourhood $\mathcal{M}$ *of* $(\theta_0, \mathbf{x}_0)$.

Proof. Using the triangle inequality on the space $L^p\left([0,a_0]\times\mathcal{M}, d\mu\right)$ we have

$$\left\|a_\nu^{-k}\langle f, n_\nu\rangle\right\| \leqslant \left\|a_\nu^{-k}\langle \mathsf{P}_{\mathcal{C}} f, n_\nu\rangle\right\| + \left\|a_\nu^{-k}\langle \mathsf{P}_{\mathcal{C}^c} f, n_\nu\rangle\right\|.$$

Using the assumptions of the theorem we can apply the reconstruction formula (3.13) to $a_\nu^{-k}\langle \mathsf{P}_{\mathcal{C}} f, n_\nu\rangle$, and then follow in the same lines as in the proof of Theorem 3.2.3 to obtain the required bound on $\left\|a_\nu^{-k}\langle \mathsf{P}_{\mathcal{C}} f, n_\nu\rangle\right\|_{L^p([0,a_0]\times\mathcal{M},d\mu)}$. Therefore, what is left is to find the bounds for the coefficients pertaining to $\mathcal{C}^c$.

Let us write $g = \mathsf{P}_{\mathcal{C}^c} f$. By the Parseval equality we have

$$\begin{aligned}
\langle g, n_\nu\rangle &= a_\nu^{3/4}\int_{\mathbb{R}^2} \hat{g}(\boldsymbol{\xi})\hat{\varphi}^{(\nu)}\left(\mathrm{D}_{a_\nu}\mathrm{R}_{\theta_\nu}\boldsymbol{\xi}\right)\right) d\boldsymbol{\xi} \\
&= a_\nu^{3/4}\left(\int_{|\xi_2|<a_\nu^{-\alpha}} \hat{g}(\boldsymbol{\xi})\hat{\varphi}^{(\nu)}\left(\mathrm{D}_{a_\nu}\mathrm{R}_{\theta_\nu}\boldsymbol{\xi}\right)\right) d\boldsymbol{\xi} + \int_{|\xi_2|>a_\nu^{-\alpha}} \hat{g}(\boldsymbol{\xi})\hat{\varphi}^{(\nu)}\left(\mathrm{D}_{a_\nu}\mathrm{R}_{\theta_\nu}\boldsymbol{\xi}\right)\right) d\boldsymbol{\xi}\Bigg) \\
&= I_1 + I_2,
\end{aligned} \tag{3.16}$$

where $1/2 < \alpha < 1$.

Since $\varphi^{(\nu)}$ has M vanishing moments in the x_1 direction, it follows

$$|I_1| \leqslant a_\nu^{3/4}\int_{|\xi_2|<a_\nu^{-\alpha}} a_\nu^M |\hat{g}(\boldsymbol{\xi})||\cos(\theta_\nu)\xi_1 - \sin(\theta_\nu)\xi_2|^M |\hat{\rho}^{(\nu)}\left(\mathrm{D}_{a_\nu}\mathrm{R}_{\theta_\nu}\boldsymbol{\xi}\right)| d\boldsymbol{\xi}$$

For $\boldsymbol{\xi} \in \mathcal{C}^c$ with $|\xi_2| < a_\nu^{-\alpha}$ we have

$$|\cos(\theta)\xi_1 - \sin(\theta)\xi_2| \leqslant \|\boldsymbol{\xi}\| \lesssim a_\nu^{-\alpha}.$$

Hence, it follows

$$\begin{aligned}
|I_1| &\leqslant a_\nu^{M-M\alpha}\int_{|\xi_2|<a^{-\alpha}} a_\nu^{3/4}|\hat{g}(\boldsymbol{\xi})||\hat{\rho}^{(\nu)}\left(\mathrm{D}_{a_\nu}\mathrm{R}_{\theta_\nu}\boldsymbol{\xi}\right)| d\boldsymbol{\xi} = a_\nu^{M(1-\alpha)}\langle |\hat{g}|, |\hat{\rho}^{(\nu)}\left(\mathrm{D}_{a_\nu}\mathrm{R}_{\theta_\nu}\cdot\right)|\rangle \\
&\leqslant a_\nu^{M(1-\alpha)}\|f\|_2\|\hat{\rho}^{(\nu)}\|_2.
\end{aligned}$$

Regarding I_2 we have

$$\begin{aligned}
|I_2| &= a_\nu^{3/4}\int_{|\xi_2|<a_\nu^{-\alpha}} |\hat{g}(\boldsymbol{\xi})||\hat{\varphi}^{(\nu)}\left(\mathrm{D}_{a_\nu}\mathrm{R}_{\theta_\nu}\boldsymbol{\xi}\right)| d\boldsymbol{\xi} \\
&= a_\nu^{-3/4}\int_{|a_\nu^{-1/2}\cos(\theta_\nu)\tilde{\xi}_2 - a_\nu^{-1}\sin(\theta_\nu)\tilde{\xi}_1|<a_\nu^{-\alpha}} |\hat{g}(\mathrm{R}_{-\theta_\nu}\mathrm{D}_{1/a_\nu}\tilde{\boldsymbol{\xi}})||\hat{\varphi}^{(\nu)}(\boldsymbol{\xi})| d\boldsymbol{\xi}
\end{aligned}$$

$$= a_\nu^{-3/4} \int_{|a_\nu^{-1/2}\cos(\theta_\nu)\tilde{\xi}_2 - a_\nu^{-1}\sin(\theta_\nu)\tilde{\xi}_1|<a_\nu^{-\alpha}} |\hat{g}(\mathrm{R}_{-\theta_\nu}\mathrm{D}_{1/a_\nu}\tilde{\boldsymbol{\xi}})||\tilde{\xi}_2|^{-L} \left| \widehat{\left(\frac{\partial^L}{\partial \mathsf{x}_2^L}\varphi^{(\nu)}\right)} \right| (\tilde{\boldsymbol{\xi}}) d\tilde{\boldsymbol{\xi}}$$

Using the fact that $\theta_\nu \in [\theta_0 - \varepsilon, \theta_0 + \varepsilon] \subsetneq \mathcal{C}$ we have $|\mathrm{R}_{\theta_\nu}\boldsymbol{\xi}|_2 > C(\theta_0)|\xi_2|$. Thus, for $|\xi_2| > a_\nu^{-\alpha}$ it follows

$$|\tilde{\xi}_2| = a_\nu^{1/2}|\mathrm{R}_{\theta_\nu}\boldsymbol{\xi}|_2 > C(\theta_0)a_\nu^{1/2-\alpha}.$$

Therefore, we have

$$|I_2| \leqslant a_\nu^{L(\alpha-1/2)}\|f\|_2 \left\| \frac{\partial^L}{\partial \mathsf{x}_2^L}\varphi^{(\nu)} \right\|_2.$$

Plugging it all back into (3.16) it follows

$$|\langle g, n_\nu\rangle| \lesssim a_\nu^{L(\alpha-1/2)} + a_\nu^{M(1-\alpha)}.$$

which yields

$$\left\| a_\nu^{-k}\langle \mathsf{P}_{\mathcal{C}^c} f, n_\nu\rangle \right\|_{L^p([0,a_0]\times\mathcal{M},d\mu)} \lesssim 1.$$

Using the assumptions (3.14)-(3.15), the assumptions of the theorem, and the result on $\left\|a_\nu^{-k}\langle \mathsf{P}_{\mathcal{C}} f, n_\nu\rangle\right\|_{L^p([0,a_0]\times\mathcal{M},d\mu)}$ the claim of the theorem follows. □

We have everything in place and we are ready to state a result about the resolution of microlocal features for families that admit a restricted reconstruction formula. In Corollary 3.2.6 we had to ensure that the shearlet family we are working with admits a global reconstruction formula (and that the dual frame is also a CPM family), which we can now avoid. In order to see this new functionality in action we will consider a family of compactly-supported shearlets that were introduced in Section 2.3.3. This family satisfies the assumptions (3.14)-(3.15) and is a tight frame for cone-supported functions.

Proposition 3.2.9. *Let $\psi(\mathbf{x}) = \psi_1(\mathsf{x}_1)\psi_2(\mathsf{x}_2)$ be a generator of the shearlet system defined in (2.13), where the functions ψ_1, ψ_2 are both compactly supported, ψ_1 has $M + R$ vanishing moments and Fourier decay of order N_1, and furthermore ψ_2 has Fourier decay of order $N_1 + N_2$. Assume the angle θ_0 lies in the cone $\mathcal{C}_{u,v}$. Then f is in $H^k(\theta_0, \mathbf{x}_0)$ if and only if*

$$\int_{\mathcal{N}_s} \left(\int_0^{a_0} |\langle f, \psi_{a\,s\,\mathbf{x}}\rangle|^2 a^{-2k} \frac{da}{a^3} \right) ds d\mathbf{x} < \infty \tag{3.17}$$

where $\mathcal{N}_s$ is some open and bounded neighbourhood of $(s_0, \mathbf{x}_0)$ and s_0 is the corresponding shearing parameter.

Proof. Let us first assume that (3.17) holds. Proposition 2.3.10 tells us that under the assumptions of Proposition 3.2.9, the family of compact shearlets (2.13) constitute a family of CPMs of order $(R, M + N_1, N_1, N_2)$. Furthermore, by the discussion in Section 2.3.3 it also follows that the family (2.14), which is a projection onto the cone $\mathcal{C}_{u,v}$, constitutes a frame for $L^2(\mathcal{C}_{u,v})$ and admits the following representation formula

$$f = \frac{1}{C_\psi}\int_{\mathbb{R}^2} \langle f, \mathsf{T}_{\mathbf{b}} W\rangle \mathsf{P}_{\mathcal{C}_{u,v}} \mathsf{T}_{\mathbf{b}} W d\mathbf{b} + \frac{1}{C_\psi}\int_{\mathbb{R}^2}\int_{s\in[-\Xi,\Xi]}\int_{a\in[0,1]} \langle f, \psi_{as\mathbf{b}}\rangle \mathsf{P}_{\mathcal{C}_{u,v}}\psi_{as\mathbf{b}} da ds d\mathbf{b}.$$

The conditions (3.14)-(3.15) are satisfied by construction and since we have only one generator the supremum goes over a set which contains only one element. Therefore, it is bounded. Applying the same arguments as in the proof of Corollary 3.2.6, we can apply Theorem 3.2.8 and the claim follows by Theorem 3.2.1. As in the previous proofs, the converse follows trivially. □

These conditions can be considerably weakened, for example the condition on the separability of ψ simplifies the computations, but can it be avoided.

Chapter 4

Edge and Corner Point Detection

4.1 Introduction

The problem of detecting the edge, or other geometrical discontinuities, of a function is a topic of fundamental interest for many problems in image analysis, numerical solutions of partial differential equations and approximation theory. For example, many techniques of contemporary digital medical imaging, such as magnetic resonance imaging, encode a signal by digitising it with the Fourier transform. The image is then obtained by one of many methods of Fourier reconstruction [70, 71] which ought to take into account that due to various imperfections that exist in the signal acquisition process, the data might be noisy and incomplete. In many cases though, to reach a useful diagnosis clinical specialists are interested only in the shapes that outline and separate the areas of interest, rather than in the fine details of the image. Therefore, robust methods of edge detection are vital in conducting this stage of the process.

Due to its importance in a host of applications, this topic has naturally received a lot of attention over the years. The majority of contemporary high-dimensional edge detection methods are essentially one-dimensional, that is, they rely on the computation of the singular support of a given function through an application of one-dimensional detection schemes in each coordinate direction. As we have seen in Section 3.1 of Chapter 3, singular support consists of points $\boldsymbol{\xi}$ where $\hat{f}(\boldsymbol{\xi})$ does not decay rapidly as $\|\boldsymbol{\xi}\| \to \infty$. Therefore, we can say that the information regarding the discontinuities of an image is encoded in its high-frequency content, which can be extracted with high-pass filters [52, 70].

Let us look at a simple example. Assume that the input image can be modelled as the indicator function of the unit ball, $f(\mathbf{x}) = \chi_{\mathcal{B}_0(1)}(\mathbf{x})$. The singular support of f is then exactly the unit sphere $\{\mathbf{x} : \|\mathbf{x}\| = 1\}$, that is, the boundary of the image. In order to recover the boundary curve, the standard approach would then consist of two steps: the computation of the points that lie on the boundary and an application of the algorithm that connects the computed points into the boundary curve(s).

It is not hard to see that the reconstruction of the original curve from the computed points of discontinuity is subject to many practical issues. For example, the set of

computed points might contain spurious information if the initial image contained noise, resulting in a poor reconstruction. On the other hand, if the original image consists of two (or more) boundary curves, for example a unit ball and a unit rectangle whose closures are disjoint, it might be exceedingly hard to tell the two boundaries apart and associate the points to the correct boundary curve, if the boundary curves are close by.

We can try to rectify, or at the very least improve, this predicament by using additional information provided by the geometry of the underlying image. The type of information that would be apposite in the context of edges are the directions of the normals at points on the boundary. In terms of microlocal analysis, this would equate to computing the wavefront set of the image, instead of the singular support. As a reminder, the wavefront set of the unit ball is given as

$$\mathrm{WF}(\chi_{\mathcal{B}_0(\mathbf{1})}) = \{(\theta, (\cos(\theta), \sin(\theta))) : \theta \in [0, 2\pi)\}.$$

Computing both the locations and the directions of the points of discontinuity would serve a number of purposes. Firstly, it is highly unlikely that the direction of the normal of spurious, that is random, points would be consistent with the directions of the normals of the points on the actual boundary curve. Hence, this additional geometrical datum can be used to essentially denoise the signal and consequently dispose of any spurious data before we proceed with the reconstruction. Furthermore, the normals of the points that lie on the same boundary curve will locally not deviate too much, thus the separation of two disjoint boundary curves, and properly associating each point to its respective part of the boundary, will be easier to conduct.

We have argued in Chapter 3 that CPMs are well suited for tasks in microlocal analysis, such as indicating the wavefront set of a function. By the preceding argumentation, this feature of parabolic system is clearly a valuable asset for edge detection and there is a certain body of evidence to support the claim of the efficacy of parabolic systems in this context. These studies are based on detecting the boundary through the decay of the dilation parameter. For example, the seminal work on second generation curvelets [15] includes a brief investigation of the singularity features of sets in $\mathbb{R}^2$, though the relevant analysis is restricted to polygons and similar objects. The same type of analysis was later performed for band-limited shearlets in [66], and then extended to more general sets in [72]. A comparable analysis xwas done for specific constructions of compactly supported shearlets recently in [73]. On the other hand, in [52] the authors constructed a family of high-pass filters whose relationship between angle and radial filters satisfy the parabolic scaling law, though the focus of the work is somewhat different.

Our goal now is to extend and generalise these considerations. The methods we shall develop are somewhat of an amalgam of the ideas from [15], [66] and [72], while the discussion in Section 4.5 was mostly motivated by [74]. The primary contribution of our work lies in the level of generality. In other words, our results on corner and edge detection hold for families of parabolic molecules that satisfy some rather mild assumptions. Most importantly, we pose no assumptions on how are those families constructed (be it in the frequency or in the space domain) nor do we assume anything about their supports. Consequently, the results obtained herein can be applied to general

parabolic families, including those for which a detailed analysis of such features has not yet been conducted, such as curvelets.

Let us summarise our results. When $\mathbf{p} \in \partial\Omega$ is just an edge point we will show that if the molecule m_λ is not orthogonal to $\partial\Omega$ at $\mathbf{p}$ then we have arbitrarily fast decay $\langle \chi_\Omega, m_\lambda \rangle \lesssim a^N$, by appealing to the results of Chapter 3 (Theorem 4.2.1). On the other hand, when m_λ is orthogonal to $\partial\Omega$ at $\mathbf{p}$ the rate $\langle \chi_\Omega, m_\lambda \rangle \lesssim a^{3/4}$ will follow through a very simple argument (Equation (4.2)). The corner points on the other hand demand considerably more attention and we will need to use an approximation argument to address general domains Ω. In the end, we will show that for a corner point $\mathbf{p} \in \partial\Omega$ we again have $\langle \chi_\Omega, m_\lambda \rangle \lesssim a^{3/4}$, if m_λ is orthogonal to $\partial\Omega$ at $\mathbf{p}$, and otherwise $\langle \chi_\Omega, m_\lambda \rangle \lesssim a^{5/4}$ (Theorem 4.4.4). Therefore, whereas the frame coefficients for edge points have one direction of slow decay, and all other directions give fast decay, in case of corner points the decay will be slow for the two directions that correspond to the normals, and it will be marginally faster for all other directions.

The drawbacks that come when the framework is as general as it is here are rather obvious. The issue is that most of the results are of only qualitative nature (though a more detailed analysis could yield better quantitative estimates), and we can only produce upper bounds on frame coefficients. The present research, for example [73] and [8], seems to suggest that it would be reasonable to expect that the upper bounds on the frame coefficients are unattainable at this level of generality, since the existing proofs depend delicately on the specifics of each construction.

We will begin in Section 4.2 where we will briefly put the results from Chapter 3 in the context of analysis of edges. Section 4.3 will be devoted to the analysis of corner points of angular wedges and polygons. The analysis of angular wedges will form a basis for the analysis of general sets through an approximation procedure. In Section 4.4 we will describe and extended the results of Section 4.3 to general sets. Section 4.5 will be devoted to the study of the decay rates of the frame coefficients when the indicator function of the set in question is multiplied by a smooth function. The purpose of Section 4.6 is to ask (and answer) the question of what would change if we chose a different type of dilation. We will finish off with Sections 4.7 and 4.8 with a simple numerical study and a brief discussion of possible further directions that could be pursued.

Throughout this chapter we will use $\mathbf{p}$ to denote the arbitrary point of interest in $\mathbb{R}^2$. Furthermore, when we say that an angle θ_λ, is normal or orthogonal to a bounded open set $\Omega \subseteq \mathbb{R}^2$, or to its boundary $\partial\Omega$, at a point $\mathbf{p}$, or to some angle, what me mean is that $\cos(\eta + \theta_\lambda) = 0$, where η is the angle of the normal at $\mathbf{p} \in \partial\Omega$. Analogously, we will say that a molecule m_λ, with $\lambda = (a_\lambda, \theta_\lambda, \mathbf{p})$, is orthogonal to Ω, or to $\partial\Omega$, if $\mathbf{p} \in \partial\Omega$ and θ_λ is normal to $\partial\Omega$ at $\mathbf{p}$.

4.2 Detection of Edges

Consider a bounded, open set $\Omega \subseteq \mathbb{R}^2$ with a continuous and piecewise-smooth boundary $\partial\Omega$ that has non-vanishing and bounded curvature, and let χ_Ω denote its indicator

function. As we have mentioned in Chapter 3, the wavefront set of χ_Ω is given by

$$\mathrm{WF}(\chi_\Omega) = \{(\theta, \mathbf{x}) : \mathbf{x} \in \partial\Omega, \theta \text{ normal to } \partial\Omega \text{ at } \mathbf{x}\}. \tag{4.1}$$

The proof of this result can be found in [62]. On the other hand, in Theorem 3.2.7 we showed that the wavefront set of a function f is the complement of the set of angle-location pairs for which the frame coefficients with respect to a family of CPMs decay rapidly as the dilation parameter goes to zero. Let us summarise this in a theorem.

Theorem 4.2.1. *Let $\Omega \subseteq \mathbb{R}^2$ be a bounded and open set with continuous and piecewise-smooth boundary $\partial\Omega$ that has non-vanishing and bounded curvature, and let $\Gamma = \{m_\lambda : \lambda \in \Lambda_\Gamma\}$ be a family of continuous parabolic molecules that satisfy the assumptions of Theorem 3.2.7. Then if a point $\mathbf{p} \in \mathbb{R}^2$ does not lie on the boundary curve $\partial\Omega$ and the angle $\theta_\lambda \in [0, 2\pi)$ is arbitrary, we have*

$$\langle \chi_\Omega, m_\lambda \rangle \lesssim a_\lambda^N, \quad \text{for all } N \in \mathbb{N},$$

for $\lambda = (a_\lambda, \theta_\lambda, \mathbf{p})$. Otherwise, if $\mathbf{p}$ lies on the boundary curve $\partial\Omega$, and $\partial\Omega$ is C^∞ in the neighbourhood of $\mathbf{p}$, then taking θ_λ which is normal to $\partial\Omega$ at $\mathbf{p}$, we again have

$$\langle \chi_\Omega, m_\lambda \rangle \lesssim a_\lambda^N, \quad \text{for all } N \in \mathbb{N},$$

for $\lambda = (a_\lambda, \theta_\lambda, \mathbf{p})$.

Proof. Theorem 3.2.7 says that the wavefront set $\mathrm{WF}(\chi_\Omega)$ is the complement of the set of points where the frame coefficients are of rapid decay, that is, it is a complement of points $(\theta_\lambda, \mathbf{p})$ for which $\langle \chi_\Omega, m_\lambda \rangle \lesssim a_\lambda^N$ for all $N \in \mathbb{N}$ where $\lambda = (a_\lambda, \theta_\lambda, \mathbf{p})$ and a_λ is small enough. Therefore, the statement follows by a comparison with (4.1) □

Therefore, if a given point $\mathbf{p}$ is either not on the boundary of Ω or is on the smooth part of the boundary but the orientation of the parabolic molecule is not parallel with the normal at $\mathbf{p}$, then the frame coefficients $\langle \chi_\Omega, m_\lambda \rangle$ are of arbitrarily fast decay. Of course, this result is of only qualitative nature, that is to say, the dependence of the decay rates with respect to the order of the given family of CPMs and to the smoothness of the boundary in the vicinity of $\mathbf{p}$ could be made more explicit, but such considerations are not pertinent for the present discussion.

This still leaves a couple of questions unanswered. The first question concerns the decay rates of the frame coefficients $\langle \chi_\Omega, m_\lambda \rangle$ as the dilation parameter a tends to zero when $\mathbf{p}$ lies on the boundary of Ω and the angle θ_λ is normal to $\partial\Omega$ at $\mathbf{p}$. Studies conducted in [8, 15, 73] indicate that the answer is exactly $a_\lambda^{3/4}$. We will not focus on this question since we cannot obtain lower bounds through our framework, whereas the upper bound can be achieved through a very simple argument

$$|\langle \chi_\Omega, m_\lambda \rangle| \leqslant a_\lambda^{3/4} \int_{\mathbb{R}^2} |\varphi^{(\lambda)}(\mathbf{x})| d\mathbf{x} \lesssim a_\lambda^{3/4}, \tag{4.2}$$

which follows since the indicator function χ_Ω is in L^∞. The rate reached in (4.2) is up to the multiplicative constant the same as what is obtained through substantially more delicate means, such as in [8, 73]. Therefore, we will not pursue this question any further.

The second interesting question concerns the points were the boundary is not smooth. This is the question we will focus on in the rest of this chapter. These are the points where the boundary curve is continuous but the derivative is not uniquely defined, or in other words, the tangents from the left and from the right at those points on the boundary are not aligned.

4.3 Detection of Corner Points of Wedges and Polygons

The first order of business is to define what is a corner point of a set $\Omega \subseteq \mathbb{R}^2$. We will follow the definitions from [72].

Definition 4.3.1. *Let $\Omega \subseteq \mathbb{R}^2$ be a bounded and open set with continuous and piecewise-smooth boundary that has non-vanishing and bounded curvature. Denote by $\alpha_\Omega \colon [0,1] \to \mathbb{R}^2$ the parametrisation of the boundary $\partial\Omega$ (which we may assume to be an arc-length parametrisation). We say that a point $\mathbf{p} \in \partial\Omega$ is a* corner point *of Ω if $\alpha'_\Omega(t_0^+) \neq \pm\alpha'_\Omega(t_0^-)$, where $\mathbf{p} = \alpha_\Omega(t_0)$.*

The condition $\alpha'_\Omega(t_0^+) = \pm\alpha'_\Omega(t_0^-)$ is also meant to exclude the sets whose boundary is Möbius-like. That is, when following the path of the normal all the way back to the starting point we end up on the same line but facing the opposite direction. This can happen for example if the number of crossings in a curve is odd, or equivalently, if the winding number of a point in the interior of the curve is even, thus reversing the orientation of the normals.

The first step in computing the decay rates of the frame coefficients is showing that they are a local property, that is, that for a given function f it is only the behaviour in a small neighbourhood of a point $\mathbf{p}$ that is important for the rate of decay of the corresponding frame coefficients.

Lemma 4.3.2. *Consider two tempered distributions f_1 and f_2 such that $f_1 = f_2$ in an open neighbourhood of $\mathbf{p} \in \mathbb{R}^2$ and let $\Gamma = \{m_\lambda : \lambda \in \Lambda_\Gamma\}$ be a family of continuous parabolic molecules of order (R, M, N_1, N_2) such that $N_1 > 4$. Then for $\lambda = (a_\lambda, \theta_\lambda, \mathbf{p})$, the coefficients $\langle f_1, m_\lambda \rangle$ decay at rate $\rho \in \mathbb{N}$ if and only if $\langle f_2, m_\lambda \rangle$ decays at rate ρ, provided $R \geqslant 2\rho$, that is*

$$\langle f_1, m_\lambda \rangle \sim a^\rho \text{ if and only if } \langle f_2, m_\lambda \rangle \sim a^\rho, \text{ as } a \to 0.$$

Proof. Let us first assume that f_1 and f_2 are bounded compactly supported functions. The general result for tempered distributions then follows by standard methods. We can assume without loss of generality that f_1 and f_2 coincide on a ball $\mathcal{B}_\epsilon(\mathbf{p})$. We have

$$|\langle f_1, m_\lambda \rangle - \langle f_2, m_\lambda \rangle| = \left| \int_{\mathbb{R}^2} \overline{m_\lambda}(\mathbf{x})(f_1 - f_2)(\mathbf{x}) d\mathbf{x} \right|$$

$$\leqslant \|f_1 - f_2\|_{L^\infty(\mathcal{B}^c(\mathbf{p}))} \int_{\mathcal{B}^c_\epsilon(\mathbf{p})} |m_\lambda(\mathbf{x})| d\mathbf{x}.$$

We will omit the λ indices from now on. Writing $\mathrm{M} = \mathrm{D}_{1/a}\mathrm{R}_\theta$ we have

$$m(\mathbf{x}) = a^{-3/4}\varphi(\mathrm{M}(\mathbf{x} - \mathbf{p})).$$

Through standard methods we get

$$\begin{aligned}\varphi(\mathbf{x}) &= \int_{\mathbb{R}^2} \hat{\varphi}(\boldsymbol{\xi}) e^{2\pi\imath\mathbf{x}\cdot\boldsymbol{\xi}} d\boldsymbol{\xi} = -(2\pi\|\mathbf{x}\|)^{-2} \int_{\mathbb{R}^2} \Delta\hat{\varphi}(\boldsymbol{\xi}) e^{2\pi\imath\mathbf{x}\cdot\boldsymbol{\xi}} d\boldsymbol{\xi} \\ &= (-1)^k (2\pi\|\mathbf{x}\|)^{-2k} \int_{\mathbb{R}^2} \Delta^k\hat{\varphi}(\boldsymbol{\xi}) e^{2\pi\imath\mathbf{x}\cdot\boldsymbol{\xi}} d\boldsymbol{\xi}.\end{aligned}$$

Using Jensen's inequality we have $(1 + \mathsf{x}^2)^k \leqslant 2^k(1 + (c\mathsf{x})^{2k})$ for any $c > 1$ and for $\mathsf{x} \in \mathbb{R}$. Therefore,

$$|\varphi(\mathbf{x})| \leqslant C_k 2^{k-1} \left(1 + \|\mathbf{x}\|^2\right)^k = \tilde{C}_k \langle\|\mathbf{x}\|\rangle^{-2k},$$

holds for all $k \in \mathbb{N}$ for which $\Delta^k\hat{\varphi}$ exists and such that

$$\int_{\mathbb{R}^2} |\hat{\varphi}(\boldsymbol{\xi})| d\boldsymbol{\xi} < \infty, \quad \text{and} \quad \int_{\mathbb{R}^2} |\Delta^k\hat{\varphi}(\boldsymbol{\xi})| d\boldsymbol{\xi} < \infty. \tag{4.3}$$

Due to time frequency localisation conditions (2.3), the terms in (4.3) will hold as long as the CPM family Γ is of order (R, M, N_1, N_2), where $R \geqslant 2k$ and $N_1 \geqslant 4$. Therefore, we have

$$|\varphi(\mathrm{M}(\mathbf{x} - \mathbf{p}))| \leqslant C_k 2^{k-1} \langle\|\mathrm{M}(\mathbf{x} - \mathbf{p})\|\rangle^{-2k}. \tag{4.4}$$

Now, since

$$\|\mathrm{M}\mathbf{u}\| \geqslant \|\mathrm{M}^{-1}\|^{-1}\|\mathbf{u}\| = a^{-1/2}\|\mathbf{u}\|,$$

it follows

$$\int_{\mathcal{B}^c_\epsilon(\mathbf{p})} |m(\mathbf{x})| d\mathbf{x} \leqslant C_k a^k \int_{\mathcal{B}^c_\epsilon(\mathbf{p})} (a + \|\mathbf{u}\|^2)^{-k} d\mathbf{u} \leqslant C_{k,\epsilon} a^k,$$

as long as $k > 1$. Therefore, the conclusion follows. □

From this point forward we will assume that any family of CPMs satisfies the conditions of Lemma 4.3.2.

Let us now describe the strategy that we shall adhere to in the rest of the chapter for the purposes of corner detection. We will begin our analysis by considering the prototypical examples of sets in $\mathbb{R}^2$ with corner points; the angular wedges, such as in Figure 4.3. A simple argument using the localisation Lemma 4.3.2 then allows for the corner detection of polygons. In order to address more general sets we will use an approximation argument in which the boundary of a given set Ω will near a corner point be approximated with straight lines determined by the two tangents that define the corner point. We will also show that the information regarding the decay rates is preserved throughout this process.

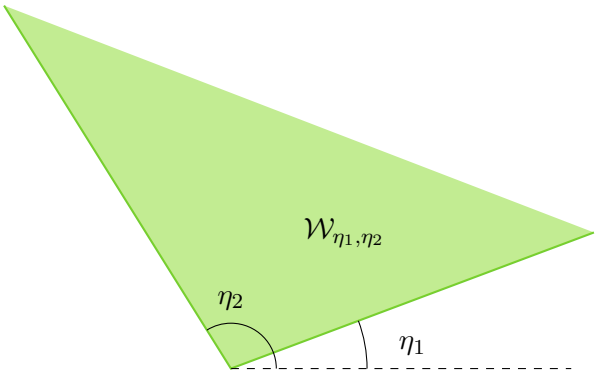


Figure 4.1: Angular wedge defined through its two angles, η_1 and η_2

To begin, let $\mathcal{W}_{\eta_1,\eta_2} \subseteq \mathbb{R}^2$ denote an angular wedge centred at the origin, where $0 < \eta_1 < \eta_2 < 2\pi$ and $\eta_2 - \eta_1 \neq \pi$, which is defined by

$$\mathcal{W}_{\eta_1,\eta_2} = \left\{ \mathbf{x} \in \mathbb{R}^2 : \eta_1 \leqslant \left| \operatorname{arctan2}\left(\frac{\mathsf{x}_2}{\mathsf{x}_1}\right) \right| \leqslant \eta_2 \right\}. \tag{4.5}$$

Here by arctan2 we denote the extension of the arc tangent so that its range is $[0, 2\pi)$. Let $H_{\eta_1,\eta_2}(\mathbf{x}) = \chi_{\mathcal{W}_{\eta_1,\eta_2}}(\mathbf{x})$ denote the indicator function of such a wedge and assume without loss of generality that $0 < \eta_2 - \eta_1 < \pi$. The function H_{η_2,η_2} induces a distribution whose Fourier transform can be obtained by first computing the Fourier transform of $H_{0,\pi/2}$ and then squeezing and rotating the angular wedge $\mathcal{W}_{0,\pi/2}$ as required. The function $H_{0,\pi/2}$ is the indicator function of the first quadrant, and we have

$$H_{0,\pi/2}(\mathbf{x}) = \chi_{\mathsf{x}_1>0}(\mathsf{x}_1)\chi_{\mathsf{x}_2>0}(\mathsf{x}_2).$$

Thus, $H_{0,\pi/2}$ is a direct product of two Heaviside distributions. Hence, by Proposition 1.1.1 its Fourier transform is also a direct product and is given as

$$\hat{H}_{0,\pi/2}(\boldsymbol{\xi}) = \frac{1}{4}\delta(\xi_1)\delta(\xi_2) - \frac{\imath}{4\pi}\left(\delta(\xi_2)\mathsf{PV}\frac{1}{\xi_1} + \delta(\xi_1)\mathsf{PV}\frac{1}{\xi_2}\right) - \frac{1}{4\pi^2}\mathsf{PV}\frac{1}{\xi_1}\mathsf{PV}\frac{1}{\xi_2}, \tag{4.6}$$

where PV marks the Cauchy's principal value. In the rest of the text we will (mostly) omit writing PV for the sake of reducing the notational load, though it will always be lurking somewhere in the background. Denoting now

$$\mathrm{A} = \mathrm{R}_{\frac{\eta_2+\eta_1}{2}}\,\mathrm{diag}\left(1, \tan\left(\frac{\eta_2-\eta_1}{2}\right)\right)\mathrm{R}_{\pi/4}, \tag{4.7}$$

we have $\mathcal{W}_{\eta_1,\eta_2} = \mathrm{A}\,\mathcal{W}_{0,\pi/2}$. It follows

$$\hat{H}_{\eta_1,\eta_2}(\boldsymbol{\xi}) = \tan\left(\frac{\eta_2-\eta_1}{2}\right)\hat{H}_{0,\pi/2}(\mathrm{A}^\top\boldsymbol{\xi}). \tag{4.8}$$

The case $\pi < \eta_1 - \eta_2 < 2\pi$ is analogous, since it follows from a simple observation of the identity $H_{\pi/2,2\pi} = 1 - H_{0,\pi/2}$.

In order to examine the decay rates of frame coefficients we will first look at simplified CPMs by assuming that the molecules are band-limited. Under this assumption the computations become considerably simpler, but we should still get an indication of what needs to be done in the general case.

4.3.1 Band Limited Molecules

Let $\Gamma = \{m_\lambda : \lambda \in \Lambda_\Gamma\}$ be CPM family and assume that for all λ the frequency support of $\varphi^{(\lambda)}$ satisfies

$$\operatorname{supp} \hat{\varphi}^{(\lambda)} \subseteq [A_1, A_2] \times [-B, B], \tag{4.9}$$

where $A_1, A_2, B > 0$ and $A_2 > A_1$. All known constructions of band-limited parabolic analysing systems satisfy support conditions of this type, see for example our earlier discussion regarding second generation curvelets or band-limited shearlets in Sections 1.4.3 and 1.4.4. We can now show the following.

Lemma 4.3.3. *Let $\mathbf{p} = \mathbf{0} \in \mathbb{R}^2$, and $\Gamma = \{m_\lambda : \lambda \in \Lambda_\Gamma\}$ be a family of band-limited continuous parabolic molecules that satisfy (4.9) for all $\lambda \in \Lambda_\Gamma$. For $\lambda = (a_\lambda, \theta_\lambda, \mathbf{p})$ we have*

$$\langle H_{\eta_1,\eta_2}, m_\lambda \rangle \lesssim a_\lambda^{5/4} \frac{\sin(\eta_2 - \eta_1)}{\cos(\eta_1 + \theta_\lambda)\cos(\eta_2 + \theta_\lambda)}$$

when $\cos(\theta_\lambda + \eta_i) \neq 0$ for all $i = 1, 2$. If on the other hand $\cos(\eta_j + \theta_\lambda) = 0$, then we have

$$\langle H_{\eta_1,\eta_2}, m_\lambda \rangle \lesssim a_\lambda^{3/4} \frac{\sin(\eta_2 - \eta_1)}{\cos(\eta_k + \theta_\lambda)},$$

where $k \in \{1, 2\} - j$.

Proof. Let us first examine the case when $\cos(\theta_\lambda + \eta_i) \neq 0$ for $i = 1, 2$. We can notice that the Dirac delta function contributions from (4.6), that is (4.8), vanish when the dilation parameter a is small enough since the supports of the Dirac delta contributions (the origin) and $\hat{\varphi}^{(\lambda)}$ do not intersect due to the assumptions on the support of $\hat{\varphi}^{(\lambda)}$. Therefore, we have

$$\begin{aligned}
\langle H_{\eta_1,\eta_2}, m_\lambda \rangle &= \langle \hat{H}_{\eta_1,\eta_2}, \hat{m}_\lambda \rangle = \langle \hat{H}_{0,\pi/2}, \hat{m}_\lambda(\mathrm{A}^\top \cdot) \rangle \\
&= c \cdot \int_{\mathbb{R}^2} \frac{\overline{\hat{m}_\lambda}(\mathrm{A}^\top \boldsymbol{\xi})}{\xi_1 \xi_2} d\boldsymbol{\xi} \\
&= C \cdot \int_{\mathbb{R}^2} \frac{\overline{\hat{\varphi}^{(\lambda)}}(\boldsymbol{\xi}) d\boldsymbol{\xi}}{(\xi_1 + \sqrt{a_\lambda}\xi_2 \tan(\eta_1 + \theta_\lambda))(\xi_1 + \sqrt{a_\lambda}\xi_2 \tan(\eta_2 + \theta_\lambda))},
\end{aligned}$$

where

$$C = a_\lambda^{5/4} \frac{\sin(\eta_2 - \eta_1)}{2\sqrt{2}\cos(\eta_1 + \theta_\lambda)\cos(\eta_2 + \theta_\lambda)}.$$

For $\boldsymbol{\xi} \in \operatorname{supp} \hat{\varphi}^{(\lambda)} \subseteq [A_1, A_2] \times [-B, B]$ we have that if the dilation parameter a is small enough, then $\xi_1 + \sqrt{a_\lambda}\xi_2 \tan(\eta_1 + \theta_\lambda)$ and $\xi_1 + \sqrt{a_\lambda}\xi_2 \tan(\eta_2 + \theta_\lambda)$ will both be uniformly away from 0, and moreover, will be uniformly bounded from above and from below, which gives

$$\int_{\mathbb{R}^2} \frac{\overline{\hat{\varphi}^{(\lambda)}}(\boldsymbol{\xi}) d\boldsymbol{\xi}}{(\xi_1 + \sqrt{a_\lambda}\xi_2 \tan(\eta_1 + \theta_\lambda))(\xi_1 + \sqrt{a_\lambda}\xi_2 \tan(\eta_2 + \theta_\lambda))} \lesssim \int_{\mathbb{R}^2} \overline{\hat{\varphi}^{(\lambda)}}(\boldsymbol{\xi}) d\boldsymbol{\xi} \overset{(2.3)}{\lesssim} 1.$$

Therefore,

$$\langle H_{\eta_1,\eta_2}, m_\lambda \rangle \lesssim a_\lambda^{5/4} \frac{\sin(\eta_2 - \eta_1)}{\cos(\eta_1 + \theta_\lambda)\cos(\eta_2 + \theta_\lambda)}.$$

Let us consider now the case when θ_λ is normal to $\mathcal{W}_{\eta_1,\eta_2}$ at $\mathbf{p}$. Without loss of generality, let us assume $\cos(\eta_1 + \theta_\lambda) = 0$. Here the Dirac delta contributions do not completely vanish. The contribution of terms that involve $\delta(\xi_2)$ does vanish using the same argumentation as earlier. What we are left with is essentially an integral along one of the lines that define the angular wedge $\mathcal{W}_{\eta_1,\eta_2}$. We will examine that expression first. We have

$$\begin{aligned}
\left\langle \frac{1}{\xi_2}\delta(\xi_1), \hat{m}_\lambda(\mathrm{A}^\top \cdot) \right\rangle &= c \cdot \int_{\mathbb{R}^2} \frac{\overline{\hat{m}_\lambda}(\mathrm{A}^\top \boldsymbol{\xi})}{\xi_2} \delta(\xi_1) d\boldsymbol{\xi} \\
&= \tilde{C} \cdot \int_{\mathbb{R}^2} \frac{\delta(\mathsf{u}_2)\overline{\hat{\varphi}^{(\lambda)}}(\mathbf{u})}{\mathsf{u}_1 + \sqrt{a_\lambda}\tan(\eta_2 + \theta_\lambda)\mathsf{u}_2} d\mathbf{u} \\
&\lesssim a_\lambda^{3/4} \frac{\sin(\eta_2 - \eta_1)}{\cos(\eta_2 + \theta_\lambda)},
\end{aligned}$$

where $\tilde{C} = a_\lambda^{3/4} \frac{\sin(\eta_2 - \eta_1)}{\cos(\eta_2 + \theta_\lambda)}$, and the assertion in the last line follows due to the assumptions on $\operatorname{supp} \hat{\varphi}^{(\lambda)}$. Analogous analysis gives

$$\left\langle \frac{1}{\xi_1\xi_2}, \hat{m}_\lambda(\mathrm{A}^\top \cdot) \right\rangle \lesssim a_\lambda^{3/4} \frac{\sin(\eta_2 - \eta_1)}{\cos(\eta_2 + \theta_\lambda)} \int_{\mathbb{R}^2} \frac{\overline{\hat{\varphi}^{(\lambda)}(\boldsymbol{\xi})}}{\xi_2} d\boldsymbol{\xi},$$

where the integral is finite due to the assumptions regarding the support and the smoothness of $\hat{\varphi}^{(\lambda)}$. Therefore, putting it all together we have

$$\langle H_{\eta_1,\eta_2}, m_\lambda \rangle \lesssim a_\lambda^{3/4} \frac{\sin(\eta_2 - \eta_1)}{\cos(\eta_2 + \theta_\lambda)},$$

as desired. □

For full disclosure, the rate $a^{3/4}$ can achieved, up to a multiplicative constant, by the same arguments as in (4.2), that is, by considering the L^∞ nature of the Heaviside function. Lemma 4.3.3 on the other hand gives us further information about the geometric interplay between the orientation of the molecule and the angles at the corner point.

4.3.2 General Parabolic Molecules

We will now show that the results of the previous section can be extended to more general scenarios, namely, for CPMs that are not necessarily band-limited. Even though the general principles remain the same, the computations will get significantly more delicate and we will need to impose some rather mild conditions on the CPM family we will be working with. In this section we will constantly work under the assumption that the underlying CPM family is of sufficiently high order (R, M, N_1, N_2), since our focus is not on deriving precise and quantifiable estimates. We hope that this will be clear in every proof and statement.

Let us first recall that the Fourier transform of H_{η_1,η_2} is given by

$$\hat{H}_{\eta_1,\eta_2}(\boldsymbol{\xi}) = \tan\left(\frac{\eta_2-\eta_1}{2}\right)\hat{H}_{0,\pi/2}(\mathrm{A}^\top\boldsymbol{\xi}),$$

where

$$\hat{H}_{0,\pi/2}(\boldsymbol{\xi}) = \frac{1}{4}\delta(\xi_1)\delta(\xi_2) - \frac{\imath}{4\pi}\left(\delta(\xi_2)\mathsf{PV}\frac{1}{\xi_1} + \delta(\xi_1)\mathsf{PV}\frac{1}{\xi_2}\right) - \frac{1}{4\pi^2}\mathsf{PV}\frac{1}{\xi_1}\mathsf{PV}\frac{1}{\xi_2},$$

with A as defined in (4.7).

To simplify the notation for the computations, let us split $\langle H_{\eta_1,\eta_2}, m_\lambda\rangle$ into three terms as follows,

$$\begin{aligned}\langle H_{\eta_1,\eta_2}, m_\lambda\rangle &= \langle \hat{H}_{\eta_1,\eta_2}, \hat{m}_\lambda\rangle \\ &= \langle \hat{H}_{0,\pi/2}, \hat{m}_\lambda(\mathrm{A}^\top\cdot)\rangle \\ &= \frac{1}{4}\langle H_1, \hat{m}_\lambda(\mathrm{A}^\top\cdot)\rangle - \frac{\imath}{4\pi}\langle H_2, \hat{m}_\lambda(\mathrm{A}^\top\cdot)\rangle - \frac{1}{4\pi^2}\langle H_3, \hat{m}_\lambda(\mathrm{A}^\top\cdot)\rangle,\end{aligned}$$

where (omitting PV) we define

$$\begin{aligned}H_1(\boldsymbol{\xi}) &= \delta(\xi_1)\delta(\xi_2),\\ H_2(\boldsymbol{\xi}) &= \frac{\delta(\xi_1)}{\xi_2} + \frac{\delta(\xi_2)}{\xi_1},\\ H_3(\boldsymbol{\xi}) &= \frac{1}{\xi_1\xi_2}.\end{aligned} \tag{4.10}$$

Let us first address the case when the molecule m_λ is not orthogonal to either sides of the wedge. In contrast with the discussion in Section 4.3.1, the action of the distribution $\hat{H}_{\eta_1,\eta_2}$ on a function $\hat{\varphi}^{(\lambda)}$, whose support is not necessarily compact, will generally include the contributions of all three terms, H_1, H_2 and H_3. In order to see this, we can start by observing the contribution of the first term.

Lemma 4.3.4. *Consider a CPM family $\Gamma = \{m_\lambda : \lambda \in \Lambda_\Gamma\}$ of order (R, M, N_1, N_2) and the distribution H_1 defined in (4.10). Provided the dilation parameter a_λ is small enough we have*

$$\langle H_1, \hat{m}_\lambda(\mathrm{A}^\top\cdot)\rangle \lesssim a_\lambda^{3/4+M}, \tag{4.11}$$

where the matrix A *is defined in (4.7).*

Proof. A simple computation gives

$$\langle \delta(\xi_1)\delta(\xi_2), \hat{m}_\lambda(\mathrm{A}^\top \cdot)\rangle \sim a_\lambda^{3/4}\hat{\varphi}^{(\lambda)}(\mathbf{0}) \leqslant a_\lambda^{M+3/4}, \tag{4.12}$$

where we used (2.3). □

Notice that for band-limited molecules that obey support assumptions (4.9) the left hand side of (4.11) would be equal to 0 since the origin is not in the support of $\hat{\varphi}^{(\lambda)}$.

Let us now look at the second term.

Lemma 4.3.5. *Consider a CPM family $\Gamma = \{m_\lambda : \lambda \in \Lambda_\Gamma\}$ and the distribution H_2 defined in (4.10). Provided the dilation parameter a_λ is small enough we have*

$$\langle H_2, \hat{m}_\lambda(\mathrm{A}^\top \cdot)\rangle \lesssim a_\lambda^{5/4+C_{M,N_1}}, \tag{4.13}$$

where $C_{M,N_1} > 0$ is a positive constant that depends on the order of family Γ, and becomes arbitrarily large provided M and N_1 are big enough, and matrix A is defined in (4.7).

Proof. Consider first the action of $\frac{1}{\xi_2}\delta(\xi_1)$ and omit the indices. The other term can be treated analogously. We have

$$\left\langle \frac{1}{\xi_2}\delta(\xi_1), \hat{m}(\mathrm{A}^\top \cdot)\right\rangle = a^{3/4}\int_{\mathbb{R}^2} \frac{1}{\xi_2}\delta(\xi_1)\overline{\hat{\varphi}}(\mathrm{T}\boldsymbol{\xi})d\boldsymbol{\xi},$$

where $\mathrm{T} = \mathrm{D}_a\mathrm{R}_\theta\mathrm{A}^\top$, with A. Denote now

$$f(\mathsf{y}) := \overline{\hat{\varphi}}\left(\mathsf{y}(a\cos(\theta+\eta_1), \sqrt{a}\sin(\theta+\eta_1)\right) = \hat{\varphi}(\tilde{\mathbf{y}}).$$

If we apply the Dirac delta and use the change of variables

$$\xi_2 \mapsto \frac{\mathsf{y}}{\sqrt{2}\cos\left(\frac{\eta_2-\eta_1}{2}\right)},$$

it follows

$$\left\langle \frac{1}{\xi_2}\delta(\xi_1), \hat{m}(\mathrm{A}^\top \cdot)\right\rangle \sim a^{3/4}\int_{\mathbb{R}} \frac{f(\mathsf{y})}{\mathsf{y}}d\mathsf{y}.$$

Following the standard proof of well-definedness of Cauchy's principal value, we can split up the integral as

$$\int_{|\mathsf{y}|>\epsilon} f(\mathsf{y})\frac{1}{\mathsf{y}}d\mathsf{y} = \int_{|\mathsf{y}|>\epsilon}^{C} f(\mathsf{y})\frac{1}{\mathsf{y}}d\mathsf{y} + \int_{|\mathsf{y}|>C}^{\infty} f(\mathsf{y})\frac{1}{\mathsf{y}}d\mathsf{y}, \tag{4.14}$$

where $C > 0$ is a constant that will be specified later. In order to bound the first term we observe

$$\int_{|\mathsf{y}|>\epsilon}^{C} \frac{f(\mathsf{y})}{\mathsf{y}}d\mathsf{y} = \int_{\mathsf{y}>\epsilon}^{C} \frac{f(\mathsf{y}) - f(-\mathsf{y})}{\mathsf{y}}d\mathsf{y} \leqslant 2C \cdot \sup_{\mathsf{y}\in[0,C]} |f'(\mathsf{y})|, \tag{4.15}$$

which follows from the continuity and the mean value theorem. To bound $f'(\mathsf{y})$ we have by the chain rule

$$|f'(\mathsf{y})| \leqslant \|\partial\hat{\varphi}(\tilde{\mathbf{y}})\|\|\nabla\tilde{\mathbf{y}}\| \leqslant a^{1/2}\|\partial\hat{\varphi}(\tilde{\mathbf{y}})\|.$$

Take $C > 0$ to be such that $1 + |\mathsf{y}| \leqslant a^{-\beta}$ holds for all $\mathsf{y} \in [0, C]$ and some $\beta > 0$ which will be discussed later. Notice that this implies $C \sim a^{-\beta}$. We now have

$$|f'(\mathsf{y})| \lesssim a^{1/2}\min(1, a(1+|\mathsf{y}|))^M \leqslant a^{1/2+M(1-\beta)}, \tag{4.16}$$

using the moment condition from (2.3). Plugging (4.16) into (4.15) it follows

$$\int_{|\mathsf{y}|>\epsilon}^{C} \frac{f(\mathsf{y})}{\mathsf{y}} d\mathsf{y} \lesssim a^{M(1-\beta)-\beta+1/2}. \tag{4.17}$$

On the other hand, we can write

$$\int_{|\mathsf{y}|>C} \frac{f(\mathsf{y})}{\mathsf{y}} d\mathsf{y} = \int_{|\mathsf{y}|>C} \frac{\mathsf{y}f(\mathsf{y})}{\mathsf{y}^2} d\mathsf{y} \lesssim C^{-1} \sup_{\mathsf{y}\in[C,\infty)} |\mathsf{y}f(\mathsf{y})|. \tag{4.18}$$

Furthermore, since we now have $1 + |\mathsf{y}| \geqslant a^{-\beta}$ it follows $1+|\mathsf{y}|^2 \geqslant \frac{a^{-2\beta}}{2}$. Thus, using the decay conditions (2.3) we have $|\hat{\varphi}(\tilde{\mathbf{y}})| \lesssim \langle\|\tilde{\mathbf{y}}\|\rangle^{-N_1}\langle\tilde{\mathbf{y}}_2\rangle^{-1}$, which yields

$$|\mathsf{y}f(\mathsf{y})| \leqslant a^{-1}\langle\tilde{\mathbf{y}}_2\rangle|\hat{\varphi}(\tilde{\mathbf{y}})| \leqslant a^{-1}\langle\|\tilde{\mathbf{y}}\|\rangle^{-N_2} \lesssim a^{N_1(2\beta-1)-1}. \tag{4.19}$$

Therefore, plugging (4.19) into (4.18) we have

$$\int_{|\mathsf{y}|>C} \frac{f(\mathsf{y})}{\mathsf{y}} d\mathsf{y} \lesssim a^{N_1(2\beta-1)+\beta-1}. \tag{4.20}$$

Looking at equations (4.17) and (4.20) we see that we need $1/2 < \beta < 1$. Therefore, provided M and N_1 are both large enough, the terms in (4.17) and (4.20) will have arbitrarily fast decay. To be more precise, in order to ensure that the decay rate is faster than $5/4$ we need

$$M \geqslant \frac{3/4+\beta}{1-\beta} \text{ and } N_1 \geqslant \frac{9/4-\beta}{2\beta-1}.$$

An entirely analogous argument yields the same decay order for the other term, $\langle\frac{1}{\xi_1}\delta(\xi_2), \hat{m}_\lambda(\mathrm{A}^\top\cdot)\rangle$. Hence, the conclusion follows. □

The estimates (4.11) and (4.13) confirm what we claimed from the beginning of this chapter; the terms $\langle H_1, m_\lambda\rangle$ and $\langle H_2, m_\lambda\rangle$ do not vanish, but rather exhibit fast decay with respect to the scaling parameter a, and the rate of decay depends on the smoothness of the molecule.

Let us now address the last remaining term.

Lemma 4.3.6. *Consider a CPM family $\Gamma = \{m_\lambda : \lambda \in \Lambda_\Gamma\}$ such that the functions $\varphi^{(\lambda)}$ have one vanishing moment in the x_1 direction in the sense of (2.19), the distribution H_3 defined in (4.10), and consider the matrix A as defined in (4.7). Provided the dilation parameter a is small enough it follows*

$$\langle H_3, \hat{m}_\lambda(\mathrm{A}^\top\cdot)\rangle \lesssim a_\lambda^{5/4}\frac{\sin(\eta_2-\eta_1)}{\cos(\eta_1+\theta_\lambda)\cos(\eta_2+\theta_\lambda)}.$$

Proof. Let us omit the indices and notice that same as in Section 4.3.1 we have

$$\left\langle \frac{1}{\xi_1\xi_2}, \hat{m}(\mathrm{A}^\top) \right\rangle = C \int_{\mathbb{R}^2} \frac{\overline{\hat{\varphi}}(\boldsymbol{\xi}) d\boldsymbol{\xi}}{(\xi_1 + \sqrt{a}\xi_2 \tan(\eta_1 + \theta))(\xi_1 + \sqrt{a}\xi_2 \tan(\eta_2 + \theta))}, \tag{4.21}$$

where

$$C = a^{5/4} \frac{\sin(\eta_2 - \eta_1)}{2\sqrt{2}\cos(\eta_1 + \theta)\cos(\eta_2 + \theta)}.$$

The integral in (4.21) can be split in three parts. The first part is the integral over $\left|\frac{\xi_2}{\xi_1}\right| \leqslant a^{-\alpha}$, where we write $\alpha = \frac{1}{2} - \varepsilon > 0$. Denote $t_i = \tan(\eta_i + \theta)$. We have

$$1 - |t_i| a^{\varepsilon} \leqslant \left| 1 + t_i \sqrt{a} \frac{\xi_2}{\xi_1} \right| \leqslant 1 + |t_i| a^{\varepsilon}. \tag{4.22}$$

Hence, provided the dilation parameter a is small enough it follows

$$\frac{1}{\xi_1 + \sqrt{a}t_1\xi_2}, \frac{1}{\xi_1 + \sqrt{a}t_2\xi_2} \sim \frac{1}{\xi_1}.$$

Therefore,

$$\int_{\left|\frac{\xi_2}{\xi_1}\right| \leqslant a^{-\alpha}} \frac{\overline{\hat{\varphi}}(\boldsymbol{\xi}) d\boldsymbol{\xi}}{(\xi_1 + \sqrt{a}t_1\xi_2)(\xi_1 + \sqrt{a}t_2\xi_2)} \sim \int_{\left|\frac{\xi_2}{\xi_1}\right| \leqslant a^{-\alpha}} \frac{\overline{\hat{\varphi}}(\boldsymbol{\xi}) d\boldsymbol{\xi}}{\xi_1^2}, \tag{4.23}$$

which is finite since we assumed φ satisfies the vanishing moments condition.

Consider now the second contribution, that is, the integral over $\left|\frac{\xi_2}{\xi_1}\right| \geqslant a^{-\alpha}$. Let us first observe the change of variables under the linear transformation

$$\boldsymbol{\xi} \mapsto \mathsf{v} = \begin{pmatrix} 1 & \sqrt{a}t_1 \\ 1 & \sqrt{a}t_2 \end{pmatrix} \boldsymbol{\xi}. \tag{4.24}$$

Applying (4.24) gives

$$\int_{\left|\frac{\xi_2}{\xi_1}\right| \geqslant a^{-\alpha}} \frac{\overline{\hat{\varphi}}(\boldsymbol{\xi}) d\boldsymbol{\xi}}{(\xi_1 + \sqrt{a}t_1\xi_2)(\xi_1 + \sqrt{a}t_2\xi_2)} = \frac{a^{-1/2}}{|t_2 - t_1|} \int_{\left|\frac{\xi_2}{\xi_1}\right| \geqslant a^{-\alpha}} \frac{\hat{\varphi}^a_{t_2 t_1}(\mathsf{v})}{\mathsf{v}_1\mathsf{v}_2} d\mathsf{v}_1 d\mathsf{v}_2, \tag{4.25}$$

where $\hat{\varphi}^a_{t_2 t_1}$ is defined as

$$\hat{\varphi}^a_{t_2 t_1}(\mathsf{v}) = \overline{\hat{\varphi}}\left(\frac{1}{t_2 - t_1} (t_2\mathsf{v}_1 - t_1\mathsf{v}_2, a^{-1/2}(\mathsf{v}_2 - \mathsf{v}_1) \right).$$

We will now split the area of integration in the integral (4.25) in two pieces. The first piece is the box $I_{a,\gamma} = [-a^\gamma, a^\gamma]^2$, for $\gamma > 0$ that will be specified later. Applying the standard methods of Cauchy's principal value we have

$$\frac{a^{-1/2}}{|t_2 - t_1|} \int_{I_{a,\gamma}} \frac{\hat{\varphi}^a_{t_2 t_1}(\mathsf{v})}{\mathsf{v}_1\mathsf{v}_2} d\mathsf{v} \lesssim \frac{a^{-1/2}}{|t_2 - t_1|} \int_{I_{a,\gamma}} \sup_{\tilde{\mathsf{v}} \in I_{a,\gamma}} |\partial^{(1,1)} \left(\hat{\varphi}^a_{t_2 t_1}(\tilde{\mathsf{v}}) \right)| d\mathsf{v}. \tag{4.26}$$

The chain rule gives

$$\left|\partial^{(1,1)}\left(\hat{\varphi}^a_{t_2t_1}(\mathsf{v})\right)\right| \lesssim a^{-1}\max_{|\boldsymbol{\alpha}|=2}\left|\partial^{\boldsymbol{\alpha}}\hat{\varphi}\left(\frac{1}{t_2-t_1}(t_2\mathsf{v}_1-t_1\mathsf{v}_2, a^{-1/2}(\mathsf{v}_2-\mathsf{v}_1))\right)\right|,$$

where $\boldsymbol{\alpha}\in\mathbb{N}_0^2$ is a multi-index. This can be bounded using (2.3) as

$$\left|\partial^{\boldsymbol{\alpha}}\hat{\varphi}\Big(\frac{1}{t_2-t_1}(t_2\mathsf{v}_1-t_1\mathsf{v}_2, a^{-1/2}(\mathsf{v}_2-\mathsf{v}_1)\Big)\right| \lesssim \left(a+\frac{|t_2\mathsf{v}_1-t_1\mathsf{v}_2|+|\mathsf{v}_2-\mathsf{v}_1|}{|t_2-t_1|}\right)^M.$$

Therefore,

$$\left|\partial^{\boldsymbol{\alpha}}\hat{\varphi}\left(\frac{1}{t_2-t_1}(t_2\mathsf{v}_1-t_1\mathsf{v}_2, a^{-1/2}(\mathsf{v}_2-\mathsf{v}_1)\right)\right| \lesssim a^{M\gamma}, \quad \text{for all } \mathbf{v}\in I_{a,\gamma}.$$

Plugging it back into (4.26) we get

$$\frac{a^{-1/2}}{|t_2-t_1|}\int_{I_{a,\gamma}}\frac{\hat{\varphi}^a_{t_2t_1}(\mathbf{v})}{\mathsf{v}_1\mathsf{v}_2}d\mathbf{v} \lesssim a^{M\gamma+2\gamma-3/2}. \tag{4.27}$$

Consider now the image of the cone $\left|\frac{\xi_2}{\xi_1}\right| \geqslant a^{-\alpha}$ under the linear transformation defined by (4.24). The result is the cone $\mathcal{C}^a_{t_1,t_2}$ that is determined by the lines through the points $(\pm 1\pm a^\epsilon t_1, \pm 1\pm a^\epsilon t_2)$. Therefore, we can decompose it four equivalent pieces.

Let us assume now, without loss of generality, that $t_1>t_2$. It follows

$$\frac{a^{-1/2}}{|t_2-t_1|}\int_{\mathcal{C}^a_{t_1t_2}-I_{a,\gamma}}\frac{\hat{\varphi}^a_{t_2t_1}(\mathbf{v})}{\mathsf{v}_1\mathsf{v}_2}d\mathbf{v} \lesssim a^{-1/2-2\gamma}\int_{\mathsf{v}_2=a^\gamma}^{\infty}\int_{\mathsf{v}_1=-\mathsf{v}_2}^{(1-a^\epsilon)\mathsf{v}_2}|\hat{\varphi}^a_{t_1,t_2}(\mathbf{v})|d\mathbf{v}.$$

We can now use the decay estimate $|\partial^{\boldsymbol{\alpha}}\hat{\varphi}(\mathbf{v})|\lesssim|\mathsf{v}_2|^{-N_2}$, which yields

$$\begin{aligned}\frac{a^{-1/2}}{|t_2-t_1|}\int_{\mathcal{C}^a_{t_1t_2}-I_{a,\gamma}}\frac{\hat{\varphi}^a_{t_2t_1}(\mathbf{v})}{\mathsf{v}_1\mathsf{v}_2}d\mathbf{v} &\lesssim a^{-1/2+N_2/2-2\gamma}\int_{\mathcal{C}^a_{t_1t_2}-I_{a,\gamma}}|\mathsf{v}_2-\mathsf{v}_1|^{-N_2}d\mathbf{v}\\ &\lesssim a^{-1/2+N_2/2-2\gamma+2\gamma+\epsilon-N_2(\gamma+\epsilon)}\\ &= a^{N_2\left(\frac{1}{2}-(\gamma+\epsilon)\right)+\epsilon-1/2}.\end{aligned} \tag{4.28}$$

Let us recollect the requirements on γ and ϵ. First of all, we need $\gamma,\epsilon>0$, whereas (4.22) and (4.28) further impose $\epsilon<1/2$ and $\gamma+\epsilon<1/2$. Choosing γ and ϵ that satisfy these conditions, we have fast decay provided M and N_2 are big enough and the statement follows from (4.23), (4.27) and (4.28). □

We can combine the three preceding lemmas in the following theorem.

Theorem 4.3.7. *Assume $\Gamma=\{m_\lambda : \lambda\in\Lambda_\Gamma\}$ is a family of parabolic molecules. Let $\lambda=(a_\lambda,\theta_\lambda,\mathbf{p})$ be such that θ_λ is orthogonal to neither η_2 nor η_1. Then we have*

$$\langle H_{\eta_1,\eta_2}, m_\lambda\rangle \lesssim a_\lambda^{5/4}\frac{\sin(\eta_2-\eta_1)}{\cos(\eta_1+\theta_\lambda)\cos(\eta_2+\theta_\lambda)}+a^{C_{M,N_1,N_2}},$$

where $C_{M,N_1,N_2}>0$ is a positive constant that depends on M,N_1 and N_2, is greater than $5/4$ and becomes arbitrarily large for large M,N_1 and N_2.

Proof.

$$\langle H_{\eta_1,\eta_2}, m_\lambda\rangle = \frac{1}{4}\langle H_1, \hat{m}_\lambda(\mathrm{A}^\top\cdot)\rangle - \frac{\imath}{4\pi}\langle H_2, \hat{m}_\lambda(\mathrm{A}^\top\cdot)\rangle - \frac{1}{4\pi^2}\langle H_3, \hat{m}_\lambda(\mathrm{A}^\top\cdot)\rangle,$$

where H_1, H_2 and H_3 are as defined in (4.10). The statement then follows by using the triangle inequality and applying Lemmas 4.3.4, 4.3.5 and 4.3.6. □

An analogous analysis can be applied to the case when the parabolic molecule is aligned with one of the wedge lines. We will not provide the details here since the calculations are very similar to those we have already performed and since essentially the same decay rate can be again obtained by merely using the L^∞ nature of H_{η_1,η_2}.

Theorem 4.3.8. *Assume $\Gamma = \{m_\lambda : \lambda \in \Lambda_\Gamma\}$ is a family of parabolic molecules. Let $\lambda = (a_\lambda, \theta_\lambda, \mathbf{p})$ be such that $\cos(\eta_j + \theta_\lambda) = 0$, and consider $k \in \{1,2\} - j$. Then we have*

$$\langle H_{\eta_1,\eta_2}, m_\lambda\rangle \lesssim a_\lambda^{3/4}\frac{\sin(\eta_2 - \eta_1)}{\cos(\eta_k + \theta)} + a^{C_{M,N_1,N_2}},$$

where $C_{M,N_1,N_2} > 0$ depends on M, N_1 and N_2, and is greater than $3/4$ and becomes arbitrarily large for large M, N_1 and N_2.

4.3.3 Polygons

The tools we developed thus far allow us to identify the corner points of polygons. To begin, let us show that translating any given set does not affect the decay rates. Taking a set $\Omega \subseteq \mathbb{R}^2$ and a point $\mathbf{p} \in \mathbb{R}^2$, it follows

$$\chi_{\Omega+\mathbf{p}}(\mathbf{x}) = \chi_\Omega(\mathbf{x} - \mathbf{p}) = \mathsf{T}_\mathbf{p}\chi_\Omega(\mathbf{x}),$$

where $\mathsf{T}_\mathbf{p}$ denotes the translation operator. Since the Fourier transform maps translations into modulations, we have

$$\begin{aligned}\langle \chi_{\Omega+\mathbf{p}}, m_\lambda\rangle &= \langle \hat{\chi}_{\Omega+\mathbf{p}}, \hat{m}_\lambda\rangle \\ &= \int_{\mathbb{R}^2} e^{-2\pi\imath\boldsymbol{\xi}\cdot\mathbf{p}}\hat{\chi}_\Omega(\boldsymbol{\xi})e^{2\pi\imath\boldsymbol{\xi}\cdot\mathbf{p}}\overline{\hat{\varphi}^{(\lambda)}}(\mathrm{D}_{a_\lambda}\mathrm{R}_{\theta_\lambda}\boldsymbol{\xi})d\boldsymbol{\xi} \\ &= \int_{\mathbb{R}^2} \hat{\chi}_\Omega(\boldsymbol{\xi})\overline{\hat{\varphi}^{(\lambda)}}(\mathrm{D}_{a_\lambda}\mathrm{R}_{\theta_\lambda}\boldsymbol{\xi})d\boldsymbol{\xi}\end{aligned}$$

where $\lambda = (a_\lambda, \theta_\lambda, \mathbf{p})$. Therefore, as expected, translating the set does not affect the decay rates of the frame coefficients, and it is sufficient to restrict our attention to the study of the decay rates for $\mathbf{p} = \mathbf{0}$.

Consider now a polygon[1] $\mathcal{P}$ and a corner point $\mathbf{p}$ on its boundary and let η_1 and η_2 be the angles determined by the two lines of the polygon that meet at $\mathbf{p}$. Using the localisation Lemma 4.3.2 it follows that the decay rates of coefficients $\langle \chi_\mathcal{P}, m_\lambda\rangle$ are the same as those of $\langle H_{\eta_1,\eta_2}, m_\lambda\rangle$. Therefore, we can apply the results from Section 4.3.2. We can summarise these findings in the following theorem.

[1]By a polygon we refer to a plane figure whose boundary consists of a finite series of connected straight lines closed in a loop

Theorem 4.3.9. *Let $\mathcal{P} \subseteq \mathbb{R}^2$ be a polygon and $\Gamma = \{m_\lambda : \lambda \in \Lambda_\Gamma\}$ a family of continuous parabolic molecules. Consider $\mathbf{p} \in \partial\mathcal{P}$. Provided $\mathbf{p}$ is a corner point of $\mathcal{P}$ and $\lambda = (a_\lambda, \theta_\lambda, \mathbf{p})$ then if θ_λ is not orthogonal to $\partial\mathcal{P}$ at $\mathbf{p}$, we have*

$$\langle \chi_{\mathcal{P}}, m_\lambda \rangle \lesssim a_\lambda^{5/4} \frac{\sin(\eta_2 - \eta_1)}{\cos(\eta_1 + \theta_\lambda)\cos(\eta_2 + \theta_\lambda)},$$

and otherwise if $\cos(\eta_j + \theta_\lambda) = 0$, then for $k \in \{1, 2\} - j$ we have

$$\langle \chi_{\mathcal{P}}, m_\lambda \rangle \lesssim a_\lambda^{3/4} \frac{\sin(\eta_2 - \eta_1)}{\cos(\eta_k + \theta_\lambda)}.$$

Proof. Since $H_{\eta_1,\eta_2} = \chi_{\mathcal{P}}$ in the neighbourhood of $\mathbf{p}$, by the localisation Lemma 4.3.2 we have that $\langle \chi_{\mathcal{P}}, m_\lambda \rangle$ is of the same decay order as $\langle H_{\eta_1,\eta_2}, m_\lambda \rangle$. The result now follows directly from Theorems 4.3.7 and 4.3.8. □

4.4 Analysis of Corner Points of General Sets

In order to detect edges and corner points of general sets we have to work around the obvious limitation of the work presented thus far; boundaries of general sets do not consist of straight lines. As it turns out, this issue can be addressed by a fairly straightforward approximation argument. In the following, we shall locally approximate the boundary of a given corner point of $\partial\Omega$ using straight lines whose orientations are determined by the tangents at the corner point. It will then follow that such an approximating procedure preserves the decay rates. Therefore, decay rates that we have computed for angular wedges will give decay rates on domains with more general boundaries.

Remark 4.4.1. *The procedure we are about to describe cannot be extended to more general classes of corner points, such as when*

$$\alpha_\Omega^{(k)}(t_0^+) = \alpha_\Omega^{(k)}(t_0^-) \text{ for all } k < j, \text{ and } \alpha_\Omega^{(j)}(t_0^+) \neq \pm\alpha_\Omega^{(j)}(t_0^-).$$

The problem lies in the fact that there is a direct correlation between the smoothness of the boundary around a given point, and the decay rates of coefficients with a given (continuous) parabolic frame; the higher the smoothness the faster the decay with respect to the dilation parameter. On the other hand, since the nature of our approach means that we are assessing the approximation rather than assessing the set itself, higher decay rates are rendered undetectable due to the fact that we are using a first order approximation.

Assume again that $\Omega \subseteq \mathbb{R}^2$ is a bounded and open set with continuous and piecewise-smooth boundary that has non-vanishing and bounded curvature, and is parametrised by $\alpha_\Omega : [0, 1] \to \mathbb{R}^2$. Consider a corner point $\mathbf{p} \in \partial\Omega$ where $\mathbf{p} = \alpha_\Omega(t_0)$ with $\alpha'_\Omega(t_0^+) \neq \pm\alpha'_\Omega(t_0^-)$. Let $\epsilon > 0$ be small enough so that $\mathcal{B}_\epsilon(\mathbf{p})$ intersects $\partial\Omega$ at exactly 2 points (the same then holds for all smaller ϵ). Then let $\mathcal{R}$ denote the set that approximates Ω. For the points outside of $\mathcal{B}_\epsilon(\mathbf{p})$, the set $\mathcal{R}$ is equal to Ω, that is

$$\mathcal{R} - \mathcal{B}_\epsilon(\mathbf{p}) = \Omega - \mathcal{B}_\epsilon(\mathbf{p}),$$

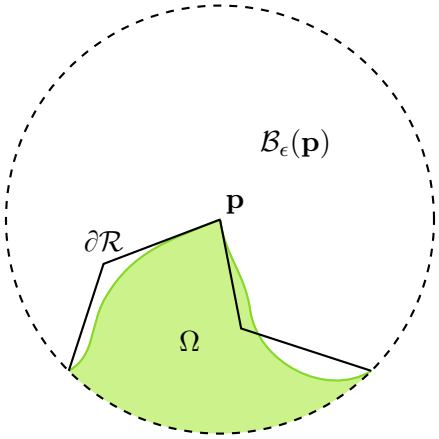


Figure 4.2: Local approximation of the boundary around a corner point

and in the part of the boundary $\partial\mathcal{R}$ within $\mathcal{B}_\epsilon(\mathbf{p})$ is a linear approximation of α_Ω obtained by locally replacing the boundary of $\partial\Omega$ with tangent lines emanating from $\mathbf{p}$ (look at Figure 4.4). That is, from the left we use $\alpha'_\Omega(t_0^-)$ and from the right $\alpha'_\Omega(t_0^+)$, to be the slopes of the respective linear interpolants through $\mathbf{p}$ (we assume α_Ω takes a positive orientation). We will denote the parametrisation of the $\partial\mathcal{R}$ as $\alpha_\mathcal{R}$. What is important is that in $\mathcal{B}_\epsilon(\mathbf{p})$ [8] we have

$$\|\alpha_\Omega(t) - \alpha_\mathcal{R}(t)\| \leqslant C(t - t_0)^2.$$

The first step in our approximation procedure is to show that the linear approximation $\chi_\mathcal{R}$ exhibits the same decay rates at $\mathbf{p}$ as its corresponding translated angular wedge $\mathcal{W}_{\eta_1,\eta_2} + \mathbf{p}$, where η_1 and η_2 denote the angles defined by the tangents $\alpha'_\Omega(t_0^+)$ and $\alpha'_\Omega(t_0^-)$. By the discussion at the beginning of Section 4.3.3, we can without loss of generality assume $\mathbf{p} = \mathbf{0}$. Furthermore, in the following we will assume that we are working with a CPM family $\Gamma = \{m_\lambda : \lambda \in \Lambda_\Gamma\}$ which is of sufficiently high order.

Lemma 4.4.2. *Let $\chi_\mathcal{R}$ be the local approximation of the set Ω that we just described, and let the angles η_1 and η_2 correspond to the tangents $\alpha'_\Omega(t_0^+)$ and $\alpha'_\Omega(t_0^-)$. Then the following holds*

$$\text{if for } \rho > 0 \quad \langle H_{\eta_1,\eta_2}, m_\lambda\rangle \lesssim a_\lambda^\rho, \text{ as } a_\lambda \to 0, \text{ then } \langle \chi_\mathcal{R}, m_\lambda\rangle \lesssim a_\lambda^\rho, \text{ as } a_\lambda \to 0,$$

where $\lambda = (a_\lambda, \theta_\lambda, \mathbf{p})$.

Proof. We have

$$\langle \chi_\mathcal{R}, m_\lambda\rangle = \langle H_{\eta_1,\eta_2}, m_\lambda\rangle + \langle \chi_\mathcal{R} - H_{\eta_1,\eta_2}, m_\lambda\rangle.$$

Therefore, since $H_{\eta_1,\eta_2} = \chi_\mathcal{R}$ holds on $\mathcal{B}_{\epsilon/2}(\mathbf{p})$, the localisation Lemma 4.3.2 gives

$$\langle \chi_\mathcal{R} - H_{\eta_1,\eta_2}, m_\lambda\rangle \lesssim a^N,$$

for every $N \in \mathbb{N}$ for which $\Delta^N\hat{\varphi}$ exists and has a finite L^1 norm. Since we by assumption have $\langle H_{\eta_1,\eta_2}, m_\lambda\rangle \leqslant \tilde{C}a^\rho$, the statement follows (provided $N \geqslant \rho$). □

Notice now

$$\langle \chi_\Omega, m_\lambda \rangle = \langle \chi_\mathcal{R}, m_\lambda \rangle + \langle \chi_\Omega - \chi_\mathcal{R}, m_\lambda \rangle. \tag{4.29}$$

Let us focus on the second term.

Lemma 4.4.3. *Let Ω and $\mathcal{R}$ be the previously described sets. Then for a CPM family $\Gamma = \{m_\lambda : \lambda \in \Lambda_\Gamma\}$, we have*

$$\langle \chi_\Omega - \chi_\mathcal{R}, m_\lambda \rangle \lesssim a_\lambda^{C_N},$$

where $C_N > \frac{5}{4}$ and $\lambda = (a_\lambda, \theta_\lambda, \mathbf{p})$.

Proof. Let us omit the λ indices. We have

$$\begin{aligned}\langle \chi_\Omega - \chi_\mathcal{R}, m \rangle &= \int_{\mathbb{R}^2} m(\mathbf{x}) \left(\chi_\Omega - \chi_\mathcal{R}\right)(\mathbf{x}) d\mathbf{x} \\ &= \underbrace{\int_{\mathcal{B}_{a^\gamma}(\mathbf{p})} m(\mathbf{x}) \left(\chi_\Omega - \chi_\mathcal{R}\right)(\mathbf{x}) d\mathbf{x}}_{I_1} + \underbrace{\int_{\mathcal{B}^c_{a^\gamma}(\mathbf{p})} m(\mathbf{x}) \left(\chi_\Omega - \chi_\mathcal{R}\right)(\mathbf{x}) d\mathbf{x}}_{I_2}.\end{aligned}$$

where $\gamma > 0$, and will be specified later, and a is taken small enough so that $a^\gamma < \epsilon$ is ensured. We will approach the discussion in slightly broader generality by allowing $\|\alpha_\Omega(t) - \alpha_\mathcal{R}(t)\| \leqslant C|t - t_0|^k$ for some $k \in \mathbb{N}$, and not just for $k = 2$. Furthermore, we will use q to denote the decay rate, instead of $5/4$. We would like to see how does the interaction between k and q play out.

Estimating I_1 we have

$$\begin{aligned}|I_1| &\lesssim a^{-3/4} \int_{\mathcal{B}_{a^\gamma}(\mathbf{p})} |\chi_\Omega - \chi_\mathcal{R}|(\mathbf{x}) d\mathbf{x} \\ &\lesssim a^{-3/4} \int_{t_0 - a^\gamma}^{t_0 + a^\gamma} \|\alpha_\Omega(t) - \alpha_\mathcal{R}(t)\| dt \\ &\lesssim a^{-3/4} \int_{t_0 - a^\gamma}^{t_0 + a^\gamma} |t_0 - t|^k dt \\ &\lesssim a^{-3/4 + \gamma(k+1)}.\end{aligned}$$

Hence, we need

$$-3/4 + \gamma(k+1) > q.$$

Let us now estimate I_2. Since

$$\Omega - \mathcal{B}_\epsilon(\mathbf{p}) = \mathcal{R} - \mathcal{B}_\epsilon(\mathbf{p}),$$

we have

$$\begin{aligned}|I_2| &\lesssim \int_{\mathcal{B}^c_{a^\gamma}(\mathbf{p})} |\varphi(\mathrm{M}(\mathbf{x} - \mathbf{p}))| |\chi_\Omega - \chi_\mathcal{R}|(\mathbf{x}) d\mathbf{x} \\ &\lesssim a^{-3/4} \int_{\mathcal{B}_\epsilon(\mathbf{p}) - \mathcal{B}_{a^\gamma}(\mathbf{p})} |\varphi(\mathrm{M}(\mathbf{x} - \mathbf{p}))| d\mathbf{x}\end{aligned}$$

$$\lesssim a^{-3/4} \int_{\mathcal{B}_\epsilon(\mathbf{p})-\mathcal{B}_{a^\gamma}(\mathbf{p})} |\varphi(\mathrm{D}_{1/a}(\mathbf{x}-\mathbf{p}))| d\mathbf{x},$$

where $\mathrm{M} = \mathrm{D}_{1/a}\mathrm{R}_\theta$. Using the estimate (4.4), we now have

$$\begin{aligned}
|I_2| &\lesssim a^{-3/4} \int_{\mathcal{B}_\epsilon(\mathbf{p})-\mathcal{B}_{a^\gamma}(\mathbf{p})} \left(1 + a^{-2}\mathsf{x}_1^2 + a^{-1}\mathsf{x}_2^2\right)^{-N} d\mathbf{x} \\
&\lesssim a^{-3/4} \int_{\mathcal{B}_\epsilon(\mathbf{p})-\mathcal{B}_{a^\gamma}(\mathbf{p})} \left(a^{-2}\mathsf{x}_1^2\right)^{-N} d\mathbf{x} \\
&\lesssim \epsilon a^{2N-3/4} \int_{a^\gamma}^{\infty} \mathsf{x}_1^{-2N} d\mathsf{x}_1 \\
&\lesssim \epsilon a^{2N(1-\gamma)-3/4+\gamma}.
\end{aligned}$$

Therefore, for the statement to hold we need

$$2N(1-\gamma) - 3/4 + \gamma > q,$$

which is the case provided $\gamma < 1$ and provided N is big enough.

Getting back to the statement of the lemma, for $q = 5/4$, we would need $\gamma > \frac{2}{k+1}$ and $\gamma < 1 - \frac{1}{2N-1}$. Furthermore, since we are using linear interpolation through $\mathcal{R}$ in order to locally approximate Ω, we have $k = 2$. It follows that $\gamma > \frac{2}{3}$. Therefore, such a γ will exist provided $N \geqslant 3$. □

We can now combine the previous lemmas into the following theorem.

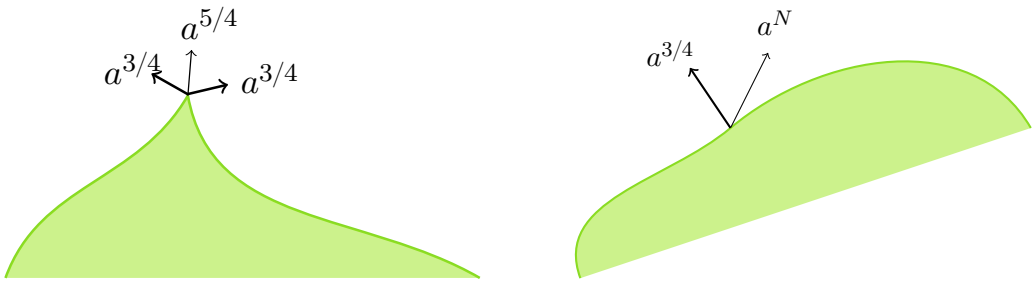
Theorem 4.4.4. *Let $\Omega \subseteq \mathbb{R}^2$ be a bounded and open set with continuous and piecewise-smooth boundary that has non-vanishing and bounded curvature. Assume $\Gamma = \{m_\lambda : \lambda \in \Lambda_\Gamma\}$ is a family of continuous parabolic molecules which satisfy the assumptions of Lemmas 4.4.2 and 4.4.3. Then if $\mathbf{p}$ is a corner point of Ω, we have*

$$\langle \chi_\Omega, m_\lambda \rangle \lesssim a_\lambda^{5/4},$$

when $\lambda = (a_\lambda, \theta_\lambda, \mathbf{p})$ and θ_λ is orthogonal to neither $\alpha'_\Omega(t_0^+)$ nor $\alpha'_\Omega(t_0^-)$, where $\alpha_\Omega(t_0) = \mathbf{p}$. Otherwise, if θ_λ is orthogonal to $\partial\Omega$ at $\mathbf{p}$ we have

$$\langle \chi_\Omega, m_\lambda \rangle \lesssim a_\lambda^{3/4}.$$

Proof. Let η_1 and η_2 be the angles associated with the tangents, $\alpha'(t_0^+)$ and $\alpha'(t_0^-)$, at $\mathbf{p} \in \partial\Omega$, and assume without loss of generality that $\mathbf{p} = \mathbf{0}$. Since $\mathbf{p}$ is a corner point we have $\eta_2 - \eta_2 \notin \pi\mathbb{Z}$ by assumption. Consider first the case when θ_λ is orthogonal to neither η_1 nor η_2. By Lemma 2.2 we have that $\langle H_{\eta_1,\eta_2}, m_\lambda \rangle \lesssim a_\lambda^{5/4}$, for $\lambda = (a_\lambda, \theta_\lambda, \mathbf{p})$. Lemma 2.3 then says that it follows $\langle \chi_\mathcal{R}, m_\lambda \rangle \lesssim a_\lambda^{5/4}$, where with $\mathcal{R}$ we denote the locally linear (around $\mathbf{p}$) approximation of Ω. Since Lemma 2.4 tells us that $\langle \chi_\Omega - \chi_\mathcal{R}, m_\lambda \rangle \leqslant a_\lambda^{C_N}$, where $C_N > \frac{5}{4}$ provided $\hat{m}_\lambda$ is sufficiently smooth, it follows by equation (4.29) that $\langle \chi_\Omega, m_\lambda \rangle \lesssim a_\lambda^{5/4}$, where the higher order terms can be readily neglected. Analogous analysis gives that $\langle \chi_\Omega, m_\lambda \rangle \lesssim a_\lambda^{3/4}$ holds when θ_λ is orthogonal to either η_1 or η_2. □



(a) Decay rates at a corner point. The rates $a^{3/4}$ correspond to orthogonal directions while $a^{5/4}$ correspond to the non-orthongal directions

(b) On smooth parts of the boundary we have the decay rate $a^{3/4}$ in the orthogonal direction and otherwise a^N

Figure 4.3: Decay rates of the points along the edge of a domain.

4.5 Multiplication with a Smooth Function

Theorem 4.4.4 admits a very simple and immediate generalisation. In the following we will try to address the decay rates for coefficients of the form $\langle f\chi_\Omega, m_\lambda\rangle$, where f is a (locally) smooth function. To begin the analysis we will look at monomials and build up from there. For a multi-index $\boldsymbol{\alpha} \in \mathbb{N}_0^2$ we compute

$$\begin{aligned}\langle \mathbf{x}^{\boldsymbol{\alpha}}\chi_\Omega, m_\lambda\rangle &= \langle \chi_\Omega, \mathbf{x}^{\boldsymbol{\alpha}} m_\lambda\rangle \\ &= C_{\boldsymbol{\alpha}}\langle \hat{\chi}_\Omega, \partial^{\boldsymbol{\alpha}}(\hat{m}_\lambda)\rangle \\ &= C_{\boldsymbol{\alpha}}\left\langle \hat{\chi}_\Omega, \sum_{|\boldsymbol{\beta}|=|\boldsymbol{\alpha}|} C_{\theta,\boldsymbol{\beta}} a_\lambda^{\beta_1+\frac{\beta_2}{2}} \hat{m}_{\lambda,\boldsymbol{\beta}}\right\rangle \\ &= C_{\boldsymbol{\alpha}} \sum_{|\boldsymbol{\beta}|=|\boldsymbol{\alpha}|} C_{\theta,\boldsymbol{\beta}} a_\lambda^{\beta_1+\frac{\beta_2}{2}} \langle \hat{\chi}_\Omega, \hat{m}_{\lambda,\boldsymbol{\beta}}\rangle,\end{aligned} \tag{4.30}$$

where

$$\hat{m}_{\lambda,\boldsymbol{\beta}} = a_\lambda^{3/4}\varphi_{\boldsymbol{\beta}}^{(\lambda)}\left(\mathrm{D}_{a_\lambda}\mathrm{R}_{\theta_\lambda}\boldsymbol{\xi}\right) \text{ and } \varphi_{\boldsymbol{\beta}}^{(\lambda)}(\boldsymbol{\xi}) = \partial^{\boldsymbol{\beta}}\varphi^{(\lambda)}(\boldsymbol{\xi}). \tag{4.31}$$

Extracting the highest common power of the dilation parameter from the expression (4.30), we have

$$\langle \mathbf{x}^{\boldsymbol{\alpha}}\chi_\Omega, m_\lambda\rangle \lesssim a_\lambda^{\frac{|\boldsymbol{\alpha}|}{2}} \sum_{|\boldsymbol{\beta}|=|\boldsymbol{\alpha}|} \langle \chi_\Omega, m_{\lambda,\boldsymbol{\beta}}\rangle. \tag{4.32}$$

The analysis developed in the preceding sections can now be readily applied to each of the coefficients $\langle \chi_\Omega, m_{\lambda,\boldsymbol{\beta}}\rangle$. A careful examination of the arguments in previous sections reveals that the computations, and the subsequent decay rates, came about by applying the time-frequency localisation of functions m_λ to show boundedness of various integrals, that is, solely by using the decay conditions (2.3) and the smoothness of $\hat{m}_\lambda$. However, it is clear from (4.31) that the functions $m_{\lambda,\beta}$ satisfy the decay conditions (2.3), provided $\hat{m}_\lambda$ is smooth enough. To be more precise, $m_{\lambda,\beta}$ is of order $(R-|\boldsymbol{\beta}|, M, N_1, N_2)$, assuming

m_λ is a molecule of order (R, M, N_1, N_2). Therefore, applying the results of Theorem 4.4.4, for a corner point $\mathbf{p}$ and $\lambda = (a_\lambda, \theta_\lambda, \mathbf{p})$, we have

$$\langle \mathbf{x}^{\boldsymbol{\alpha}} \chi_\Omega, m_\lambda \rangle \lesssim a_\lambda^{5/4+\frac{|\boldsymbol{\alpha}|}{2}},$$

when m_λ is not orthogonal to the tangents at $\mathbf{p}$ and

$$\langle \mathbf{x}^{\boldsymbol{\alpha}} \chi_\Omega, m_\lambda \rangle \lesssim a_\lambda^{3/4+\frac{|\boldsymbol{\alpha}|}{2}}$$

when it is.

The same line of argument can now be extended to general polynomials. Consider a polynomial $P_{\boldsymbol{\alpha}}(\mathbf{x})$, of finite degree, such that $\partial^{\boldsymbol{\beta}} P_{\boldsymbol{\alpha}}(\mathbf{0}) = 0$, for all multi-indices $\boldsymbol{\beta} \in \mathbb{N}_0^2$ such that$|\boldsymbol{\beta}| \leqslant \boldsymbol{\alpha}$, where $\boldsymbol{\alpha}$ is the smallest element of $\mathbb{N}_0^2$ with that property. In other words, $P_{\boldsymbol{\alpha}}$ satisfies

$$P_{\boldsymbol{\alpha}}(\mathbf{x}) = \sum_{\boldsymbol{\beta} \in \mathbb{J}} C_{\boldsymbol{\beta}} \mathbf{x}^{\boldsymbol{\beta}}.$$

where $\mathbb{J} \subseteq \mathbb{N}_0^2$, with $0 < |\mathbb{J}| < \infty$, and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2) \in \mathbb{N}_0^2$ is defined by

$$\alpha_i = \min_{\boldsymbol{\beta} \in \mathbb{I}} \beta_i, \text{ for } i = 1, 2.$$

If $\mathbf{p}$ is a corner point of Ω it follows by linearity

$$\begin{aligned} \langle P_{\boldsymbol{\alpha}} \chi_\Omega, m_\lambda \rangle &= \sum_{\boldsymbol{\beta} \in \mathbb{J}} C_{\boldsymbol{\beta}} \langle \mathbf{x}^\beta H_{\eta_1, \eta_2}, m_\lambda \rangle \\ &\lesssim a_\lambda^{\frac{|\boldsymbol{\alpha}|}{2}} \sum_{\boldsymbol{\beta} \in \mathbb{J}} \sum_{|\boldsymbol{\gamma}| = |\boldsymbol{\beta}|} \langle \chi_\Omega, m_{\lambda, \boldsymbol{\gamma}} \rangle \\ &\lesssim a_\lambda^{5/4+\frac{|\boldsymbol{\alpha}|}{2}}, \end{aligned} \tag{4.33}$$

when the molecule m_λ is not orthogonal to $\partial\Omega$ at $\mathbf{p}$. An analogous argument shows that $\langle P_{\boldsymbol{\alpha}} \chi_\Omega, m_\lambda \rangle \lesssim a_\lambda^{3/4+\frac{|\boldsymbol{\alpha}|}{2}}$ when m_λ is not orthogonal to $\partial\Omega$ at $\mathbf{p}$.

Remark 4.5.1. *It would be possible to produce a more precise assessment of the decay rates had we picked the underlying parabolic molecule with more scrutiny. To argue that this is the case it is sufficient to re-examine (4.30). The fundamental issue that hinders a better analysis lies in the fact that a general parabolic molecule $\hat{m}_\lambda$ does not separate the variables ξ_1 and ξ_2. Consequently, any partial derivative of $\hat{m}_\lambda$ will necessarily involve all partial derivatives of the same order and the dilations imposed by the matrix D_a cannot be decoupled. A much finer analysis could be obtained by considering parabolic families that can achieve this decoupling, such as the classical shearlets 1.14. But we shall not pursue this line of argument since imposing such restrictions would be at odds with what we are trying to achieve. Interested reader is directed to [74] for further discussion on this topic.*

We can now extend our analysis to more general functions by locally approximating a given function with its Taylor polynomial. Take $f \in C^m\left(\mathbb{R}^2\right)$ and denote by P_k its Taylor polynomial of degree k around $\mathbf{p}$, that is

$$f(\mathbf{x}) = \sum_{|\boldsymbol{\alpha}| \leqslant k} \partial^{\boldsymbol{\alpha}} f(\mathbf{p}) \, (\mathbf{x} - \mathbf{p})^{\boldsymbol{\alpha}} + R_k(\mathbf{x}) = P_k(\mathbf{x}) + R_k(\mathbf{x}). \tag{4.34}$$

Let $\mathbf{p}$ be a corner point of Ω and assume without loss of generality that $\mathbf{p} = \mathbf{0}$. Let f be such that $\partial^{\boldsymbol{\beta}} f(\mathbf{0}) = 0$, holds for all $\boldsymbol{\beta} \in \mathbb{N}_0^2$ such that $|\boldsymbol{\beta}| \leqslant l$, where $l < k$. Then it follows

$$\langle f\chi_\Omega, m_\lambda \rangle = \langle f\chi_\Omega - P_k\chi_\Omega, m_\lambda \rangle + \langle P_k\chi_\Omega, m_\lambda \rangle. \tag{4.35}$$

Therefore, provided

$$\langle f\,\chi_\Omega - P_k\chi_\Omega, m_\lambda \rangle \lesssim a_\lambda^N, \tag{4.36}$$

holds for a big enough $N \in \mathbb{N}$, it will follow

$$\langle f\,\chi_\Omega, m_\lambda \rangle \lesssim a_\lambda^{q+\frac{|\boldsymbol{\alpha}|}{2}}, \tag{4.37}$$

where $q = \frac{3}{4}$ when m_λ is orthogonal to $\partial\Omega$ at $\mathbf{p}$, and $q = \frac{5}{4}$ otherwise.

The proof of (4.36) is analogous to that of Lemma 4.4.3, and it is the content of Lemma A.2.1 which can be found in the Appendix. Furthermore, (4.37) clearly holds not just for globally smooth functions but for more general classes of functions by adhering to the localisation Lemma 4.3.2.

Looking at (4.37) it is clear that the resulting rate of decay comes from two contributions. The first contribution, a^q for $q = 3/4$ or $q = 5/4$, comes from coefficients of the form $\langle \chi_\Omega, \varphi_\lambda \rangle$, where φ_λ is a partial derivative of a CPM and satisfies a condition of the form (2.3). The second contribution, $a^{\frac{|\boldsymbol{\alpha}|}{2}}$, comes by using the property of the Fourier transform to draw a relationship between polynomial multiplication in the space domain with partial derivatives in the frequency domain. In other words, if a corner point $\mathbf{p}$ is a root of f, then the strength of singularity of $f\chi_\Omega$ at $\mathbf{p}$ is counteracted by the multiplicity of the root, and this can be observed through the increased rate of decay of the frame coefficients. The rate of decay will increase relative to the multiplicity of the root.

We can now use the approximation strategy developed in Section 4.4 to extended the results for any set Ω satisfying the assumptions of Theorem 4.4.4.

Theorem 4.5.2. *Let $f \in C^m(\mathbb{R}^2)$ be a function such that $\partial^{\boldsymbol{\beta}} f(\mathbf{p}) = 0$, for all $\boldsymbol{\beta} \in \mathbb{N}_0^2$ such that $|\boldsymbol{\beta}| \leqslant l$, and let $P_k(\mathbf{x})$ be its corresponding Taylor polynomial around $\mathbf{p}$ of degree k where $k > l$. Assume a family of continuous parabolic molecule $\Gamma = \{m_\lambda : \lambda \in \Lambda_\Gamma\}$ and a set Ω satisfy the conditions of Theorem 4.4.4. Consider a corner point $\mathbf{p}$ of Ω, and take $\lambda = (a_\lambda, \theta_\lambda, \mathbf{p})$. If θ_λ is orthogonal to neither $\alpha'_\Omega(t_0^+)$ nor $\alpha'_\Omega(t_0^-)$, where $\alpha_\Omega(t_0) = \mathbf{p}$ then*

$$\langle f\chi_\Omega, m_\lambda \rangle \lesssim a_\lambda^{5/4+\frac{|\boldsymbol{\alpha}|}{2}},$$

and otherwise

$$\langle f\chi_\Omega, m_\lambda \rangle \lesssim a_\lambda^{3/4+\frac{|\boldsymbol{\alpha}|}{2}}.$$

Proof. Consider (4.35), and estimate the first term by (4.36) using Lemma A.2.1. The statement then follows by (4.33). □

4.6 α-Molecules

Throughout the preceding analysis we used the parabolic scaling of the variable, as dictated by Definition 2.2.2. A natural question we might ask ourselves is how, and indeed if, would the results change if we scaled the variables differently. Since the parabolic scaling is propagated through the parabolic dilation matrices, the most immediate workaround would be to generalise the dilation matrices by $\mathrm{D}_{a,\alpha} = \mathrm{diag}(a, a^{\alpha})$ where $\alpha \in [0, 1]$, and revise the definition of continuous molecules accordingly. Constructions of this type are typically called *α-molecules* [45]. Notice that the parabolic molecules correspond to the case $\alpha = 1/2$.

We will not be concerned at this point with questions regarding existence or convergence of integrals and such things, but rather just with quantitative changes in terms of the decay with respect to the scaling parameter a. It is immediately clear that we ought to replace $a^{5/4}$ with $a^{\frac{3-\alpha}{2}}$, and $a^{3/4}$ with $a^{\frac{1+\alpha}{2}}$, in the statement of Lemma 4.3.3, and in the statements of Theorems 4.3.7 and 4.3.8 and other theorems and lemmas. Therefore, taking $\alpha = 1$ (which corresponds to the case when we treat both axes equally) we get that the decay rates satisfy $\frac{3-\alpha}{2} = \frac{1+\alpha}{2}$. Therefore, we cannot distinguish between edge and corner points, nor can we distinguish between different orientations associated to a corner point. This means that that we indeed need an unequal treatment of the axes to conduct analysis of this type.

What might be a bit more of surprise is that changing the dilation matrices does not change the conditions of Lemma 2.4. That is, in order for the lemma to hold we again get two conditions on γ, namely

$$-\frac{1+\alpha}{2} + \gamma(k+1) > \frac{3-\alpha}{2}, \quad \text{and} \quad 2N(1-\gamma) - \frac{1+\alpha}{2} + \gamma > \frac{3-\gamma}{2},$$

which are readily reduced to

$$\gamma > \frac{2}{k+1}, \quad \text{and} \quad \gamma < 1 - \frac{1}{2N-1}.$$

But those are exactly the same conditions we got with parabolic molecules. Therefore, the conclusion of the revised version of Theorem 4.4.4 would remain the same, which means that from the theoretical perspective, in terms of the detection of geometric features such as corner points, the particular choice of a dilation of the axes does not make a tangible effect as long as some sort of directional bias is implemented.

4.7 Relationship with other Work and a Simple Numerical Study

As we have mentioned on more than one occasion, the main contribution of our work is the level of generality. This generality comes with the obvious drawback, which is that we can only get the upper bounds of the decay rates, and we make no claims on the

real-world applicability of our work. We will now put our results on edge and corner detection in the context of contemporary research.

Apart from the initial studies in [15] and [66], we are aware of two approaches that are currently on the market. The common feature across most of the existing methods is the usage of an approximation procedure and a localisation argument to obtain the results on edge and corner detection.

The first approach, studied in for example [72] and [52], relies on a couple of features. Firstly, the underlying dictionary is assumed (or rather constructed) to be band-limited, where the mother function separates the angular and radial influences in one way or another. The next step uses an ingenious trick, taken from [75] and [76] to compute the Fourier transform of χ_Ω. Let $\Omega \subseteq \mathbb{R}^2$ again be an open and bounded set whose parametrisation is denoted by α_Ω and define

$$F(\boldsymbol{\xi}, \mathbf{x}) = (F_1(\boldsymbol{\xi}, \mathbf{x}), F_2(\boldsymbol{\xi}, \mathbf{x})) = -2\pi\imath \|\boldsymbol{\xi}\|^{-2} e^{-2\pi\imath \boldsymbol{\xi}\cdot\mathbf{x}} \boldsymbol{\xi}^\perp, \text{ where } \boldsymbol{\xi}^\perp = (-\xi_2, \xi_1)^\top.$$

Using the Green's theorem (or the divergence theorem) we then have

$$\begin{aligned}
\hat{\chi}_\Omega(\boldsymbol{\xi}) &= \int_\Omega e^{2\pi\imath\boldsymbol{\xi}\cdot\mathbf{x}} d\mathbf{x} = \int_\Omega \left(\partial^{(1,0)} F_2(\boldsymbol{\xi}, \mathbf{x}) - \partial^{(0,1)} F_1(\boldsymbol{\xi}, \mathbf{x}) \right) d\mathbf{x} \\
&= \int_{\partial\Omega} F(\boldsymbol{\xi}, \alpha_\Omega(t)) \alpha_\Omega'(t) dt \\
&= -\frac{1}{2\pi\imath\|\boldsymbol{\xi}\|^2} \int_{\partial\Omega} e^{-2\pi\imath\boldsymbol{\xi}\cdot\alpha_\Omega(t)} \boldsymbol{\xi}^\perp \cdot \alpha_\Omega'(t) dt
\end{aligned}$$

The phase $\boldsymbol{\xi} \cdot \alpha_\Omega(t)$ becomes stationary when $\boldsymbol{\xi} \cdot \alpha_\Omega'(t) = 0$, which corresponds to the case when $\boldsymbol{\xi}$ is normal to the boundary curve. Going to polar coordinates, one then studies the behaviour of $\hat{\chi}_\Omega$ as $\|\boldsymbol{\xi}\|$ tends to infinity and uses the method of stationary phase. Studying the coefficients $\langle \hat{\chi}_\Omega, m_\lambda \rangle$ with respect to a band-limited family of functions allows one to precisely isolate the orientations of the normals at edge and corner points.

The second approach, observed in [73], studies the case when the underlying dictionary is compactly supported. This assumption poses a unique set of challenges since a limited smoothness in the spatial domain gives a comparatively limited decay in the frequency domain, but this study is grounded in the understanding that the singularities of two-dimensional objects are essentially local properties in the spatial domain. Hence, the computations are conducted in the spatial domain, using the so-called *detector shearlets*. The results are promising and some aspects of their functionality can be extended to three dimensions.

The results in [72], [73], and our results agree in several respects. In all three studies, frame coefficients at an edge point have the decay rate of $a^{3/4}$ when the molecule is orthogonal to $\partial\Omega$ at a given point, and a^N in all other directions. These results are also in accordance with earlier studies in [72] and [52]. Furthermore, the decay rates for a corner point $\mathbf{p}$, when the molecule is orthogonal to $\partial\Omega$ at $\mathbf{p}$ are in all three cases given as $a^{3/4}$. Where the results differ is the case when the given molecules is not orthogonal to $\partial\Omega$ at a corner point. The results in [73] and results from Chapter 4 suggest that the rate is $a^{5/4}$ whereas the results of [72] suggest the rate $a^{9/4}$, which is claimed to be the

result of cancellations of certain integrals due to the properties of the construction of the underlying shearlets. Notice that this is not an immediate contradiction with our results since we can claim only upper bounds.

The follow-up studies to those [72] can be found in [74], where the authors studied the case similar to our studies in Section 4.5, as we previously mentioned. There the authors use a somewhat unorthodox construction of shearlets where the mother shearlet is defined by

$$\hat{\psi}(\boldsymbol{\xi}) = w(\xi_1)v\left(\frac{\xi_2}{\xi_1}\right), \quad \text{where } \operatorname{supp} w, \operatorname{supp} v \subseteq [-1,1].$$

Therefore, the mother shearlet is a classical shearlet, satisfying (1.14), but its support does not abide by the standard paradigms of the shearlet transform, such as those we discussed in (1.13), since the support of w is centred around the origin. The results that follow are analogous to those in the earlier paper [72], and it is claimed that the same type of cancellations still occur, including for coefficients of the form $\langle f\chi_\Omega, \psi_\lambda\rangle$.

Due to this unorthodox nature of the underlying shearlet constructions, its disagreement with other work, and a couple of other peculiarities that a diligent examination of the proofs seems to uncover, we decided to conduct a very simple numerical study of the asymptotics of decay rates for shearlets constructed in the initial paper [72], since the work in [74] is based on those results. To conduct this small experiment we used MATLAB to compute the values of the coefficients $\langle H_{\eta_1,\eta_2}, m_\lambda\rangle$, for $\theta_\lambda = 0$ and a various selection of η_1 and η_2 which correspond to the case when the molecule is not orthogonal to the wedge $\mathcal{W}_{\eta_1,\eta_2}$ at $\mathbf{p} = \mathbf{0}$, and dilation parameters $a = 2^{-5}, \ldots, 2^{-10}$, and then extrapolated the behaviour of the decay rate.

The chosen family of shearlets was such that it satisfies the assumptions of the initial work in [72], where the authors considered a mother shearlet obeying the classical construction $\hat{\psi}(\boldsymbol{\xi}) = \psi_1(\xi_1)\psi_2\left(\frac{\xi_2}{\xi_1}\right)$ such that

$$\operatorname{supp}\hat{\psi}_1 \subseteq \left[-2, -\frac{1}{2}\right] \cup \left[\frac{1}{2}, 2\right] \text{ and } \operatorname{supp}\hat{\psi}_2 \subseteq [-1,1],$$

and where $\hat{\psi}_1$ is a smooth odd function, and $\hat{\psi}_2$ is an even function that decreases on $[0,1)$ and obeys $\|\psi_2\|_2 = 1$.

The results are summarised in Figure 4.4 and Table 4.1.

	$\eta_1 = \pi/6$	$\eta_1 = 2\pi/6$	$\eta_1 = \pi/30$	$\eta_1 = -\pi/6$
Angles	$\eta_2 = 5\pi/6$	$\eta_2 = 7\pi/6$	$\eta_2 = 2\pi/6$	$\eta_2 = 3\pi/6$
Extrapolations	1.2504	1.2557	1.2542	1.2536

Table 4.1: Extrapolated decay rates for the results collected in Figure 4.4

Our elementary numerical study seem to confirm that the rate $a^{5/4}$ is the actual rate in this context. We make no claims about the robustness or conclusiveness of our numerical experiment or its results, for which a more detailed study would be needed,

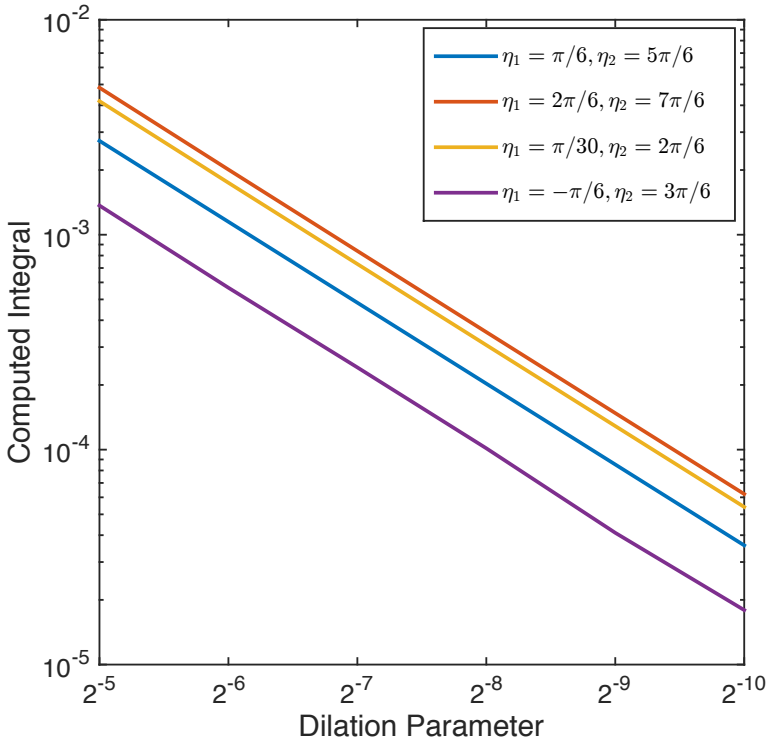


Figure 4.4: Logarithmic plot of the asymptotic behaviour of the decay rates

rather it is merely meant as an indication. We should add that we believe the results in [72] and [74] are still essentially valid, though perhaps with some minor tweaks.

4.8 Concluding Remarks

In this chapter we presented arguments that yield upper bounds on the decay rate of frame coefficients at corner points. While it might be possible to also obtain lower bounds, this does not seem likely since the existing upper bounds hold only in very specific cases, that is, only for delicately constructed families of shearlets, look at the discussion in [69, 73]. Therefore, since our approach is based in generality and level of abstractness, lower bounds seem out of reach. On the other hand, it might be possible to get a microlocal characterisation of corner points by perhaps using the geometrical dependence

of the angle at the corner point, and the orientation parameter of the molecule.

Another possible avenue to explore is the analysis of corner points of the j^{th} order, that is, instead of considering only $\alpha'(t_0^+) \neq \alpha'(t_0^-)$, we could look at $\alpha^{(j)}(t_0^+) \neq \pm\alpha^{(j)}(t_0^-)$. Methods used in the present chapter clearly cannot be adapted to such cases directly, since as we discussed earlier, the approximation procedure described here comes with limitations. Thus, a different approach would be needed. At the moment there are some results for band-limited molecules, but thus far they are not in the generality we would want them to be.

Chapter 5

Discussion and Outlook

In the introduction we said that the main goals of this thesis were to create a broad and comprehensive framework for parabolic dictionaries in the continuous realm that can answer topics in microlocal analysis. This work was built on top of the pre-existing notion of discrete parabolic molecules, but adapted to the continuous setting, and grounded in many other studies such as [4, 5, 15, 45, 66].

The framework of continuous parabolic molecules was established in Chapter 2, were we introduced a number of related notions needed for its study. We continued by showing that this framework encompasses shearlet and curvelet based constructions of many different types, and concluded by establishing a result on almost-orthogonality of pairs of continuous parabolic molecules. This entire chapter is founded and inspired by the work in [5, 9], which is reflected in a number of notions and proof techniques.

In Chapter 3 we set out to prove the effectiveness of our framework, echoing the principles of [9]. To do that we looked at topics in microlocal analysis and showed that there are fundamental notions relating to microlocal properties of functions and distributions which can be reached with our methods. We hope to have pleaded a convincing case that these results are universal and examined ways of converting the results from one parabolic dictionary to another.

In Chapter 4 we turned our attention to elements of microlocal analysis which are of particular interest, edges of regions in $\mathbb{R}^2$. We argued that corner points and edge points of piecewise-smooth regions can be detected by considering the decay of the corresponding frame coefficients with respect to families of continuous parabolic molecules that exhibit sufficient localisation in time-frequency. This also brought out some disadvantages of our framework into view. Due to its inherent nature, we could only derive upper bounds on frame coefficients and a full assessment might be out of reach at this level of generality. We examined the results of the related work [8, 73] which assume a more select and precise choice of the underlying dictionary and tried to argue that a certain consensus can be reached.

We hope the work presented in this thesis proves the validity of our assumptions.

The future looks promising and there are many questions that merit our attention. Let us spare a few words to comment on a couple of interesting topics. As we mentioned

in Section 2.2.1, an interesting question to look at is concerning the properties of a discrete parabolic frame that was sampled from a continuous one. Specifically, we would like to know when is the dual frame of a sampled discrete parabolic frame also a family of discrete parabolic molecules. This is often called the intrinsic localisation of a frame and is important for a variety of topics. The present state of art gives satisfactory answers only for simple geometries [77] while the work on more complex settings [56] leaves some open questions, and does not consider discrete frames that were directly sampled from the continuous ones.

Furthermore, we would want to simplify many of the assumptions needed for the proofs of results such as Theorem 3.2.3. Namely, it would be desirable to show that the dual frame of a CPM is automatically a CPM family.

Work on edge detection can also be improved. The first and obvious point of interest is regarding the lower bounds on the decay rates. For that, we might try to reexamine the approaches developed in [52, 72, 74], which we briefly discussed in Section 4.7. Our first couple of attempts were to no avail. We also hope that adapting those methods would also allow for a look into the case of generalised corner points that we described in Section 4.8.

On a different note, throughout this thesis we discussed the anisotropic systems that exist on the plane. In a great many applications though, the data naturally inhabit complex (non-Euclidean) manifolds. For example, the study of cosmic microwave background is often studied by methods of astronomical imaging [78]. While there are wavelets based schemes that try to address theses topics, the data is inherently anisotropic and there is a lack of methods that adopt this point of view. Problems of the same type exist in other fields such as fracture lines on various materials, seismic imaging, and so on.

Appendix A

Various Proofs

A.1 Additional Proofs For Chapter 2

Lemma A.1.1. *Let $d(\lambda,\nu)$ be as in Definition 2.2.3. Then for all $q>0$ it holds*

$$\int_{[0,2\pi)\times\mathbb{R}^2}\left[1+q^{-1}d(\lambda,\nu)\right]^{-k}d\theta d\mathbf{b}\lesssim q^2.$$

Proof. Denote $\delta\theta=\theta_\lambda-\theta_\nu$ and $\delta\mathbf{b}=\mathbf{b}_\lambda-\mathbf{b}_\nu$. For a $q\in\mathbb{R}^+$ we have

$$\begin{aligned}
&\int_{[0,2\pi)\times\mathbb{R}^2}\left[1+q^{-1}d(\lambda,\nu)\right]^{-k}d\mathbf{b}d\theta\\
&\qquad=\int_{[0,2\pi)\times\mathbb{R}^2}\left[1+q^{-1}\left(|\delta\theta|^2+|\delta\mathbf{b}|^2+|\langle\mathbf{e}_\lambda,\delta\mathbf{b}\rangle|\right)\right]^{-k}d\mathbf{b}d\theta\\
&\qquad=\int_{[0,2\pi)\times\mathbb{R}^2}\left[1+q^{-1}\left(|\delta\theta|^2+|\mathbf{b}|^2+|\langle\mathbf{e}_\lambda,\mathbf{b}\rangle|\right)\right]^{-k}\mathbf{b}d\theta\\
&\qquad=\int_{[\theta_\nu,2\pi+\theta_\nu)\times\mathbb{R}^2}\left[1+q^{-1}\left(\theta^2+|\mathbf{b}|^2+|\langle\mathrm{R}_{\theta_\lambda+\theta_\nu}^{\top}\,\mathbf{e}_\lambda,\mathbf{b}\rangle|\right)\right]^{-k}d\mathbf{b}d\theta\\
&\qquad=\int_{[\theta_\nu,2\pi+\theta_\nu)\times\mathbb{R}^2}\left[1+q^{-1}\left(\theta^2+|\mathbf{b}|^2+|\mathsf{b}_1|\right)\right]^{-k}d\mathbf{b}d\theta\\
&\qquad\leqslant\int_{\mathbb{R}\times\mathbb{R}^2}\left[1+q^{-1}\left(\theta^2+\mathsf{b}_2^2+|\mathsf{b}_1|\right)\right]^{-k}d\mathbf{b}d\theta\\
&\qquad\leqslant\int_{\mathbb{R}\times\mathbb{R}^2}\left[1+q^{-1}\left(q\theta^2+q\mathsf{b}_2^2+q|\mathsf{b}_1|\right)\right]^{-k}q^2d\mathbf{b}d\theta\\
&\qquad\leqslant q^2\int_{\mathbb{R}\times\mathbb{R}^2}\left[1+\left(\theta^2+\mathsf{b}_2^2+|\mathsf{b}_1|\right)\right]^{-k}d\mathbf{b}d\theta\\
&\qquad\lesssim q^2.
\end{aligned}$$

□

A.1.1 Lemmas Needed for the Proof of Theorem 2.4.1

Lemma A.1.2. *Consider the differential operator $\mathcal{L}_{\lambda,\nu}$ as defined in Equation (2.33). The following holds*

$$\mathcal{L}_{\lambda,\nu}\left(e^{-2\pi\imath\boldsymbol{\xi}\cdot\delta\mathbf{x}}\right) = \alpha e^{-2\pi\imath\boldsymbol{\xi}\cdot\delta\mathbf{x}},$$

where

$$\alpha = 1 + 4\pi^2 a_M^{-1}|\delta\mathbf{x}|^2 + 4\pi^2 \frac{a_M^{-2}}{1 + a_M^{-1}|\delta\theta|^2}\langle\mathbf{e}_\lambda, \delta\mathbf{x}\rangle^2.$$

Consequently, for $k \in \mathbb{N}$ we have

$$\mathcal{L}_{\lambda,\nu}^{-k}\left(e^{-2\pi\imath\boldsymbol{\xi}\cdot\delta\mathbf{x}}\right) = \alpha^{-k}\left(e^{-2\pi\imath\boldsymbol{\xi}\cdot\delta\mathbf{x}}\right).$$

Proof. Define once again $\delta\theta = \theta_\lambda - \theta_\nu$ and $\delta\mathbf{b} = \mathbf{b}_\lambda - \mathbf{b}_\nu$, where $\delta\mathsf{x}_i$ are components of $\delta\mathbf{x}$, for $i = 1, 2$. As a reminder, the differential operator $\mathcal{L}_{\lambda,\nu}$ is defined for frequency variables $\boldsymbol{\xi}$ by

$$\mathcal{L}_{\lambda,\nu} = \mathcal{I}_d - a_M^{-1}\Delta - \frac{a_M^{-2}}{1 + a_M^{-1}|\theta_\lambda - \theta_\nu|^2}\frac{\partial^2}{\partial\mathbf{e}_\lambda^2}.$$

Hence, we can decompose $\mathcal{L}_{\lambda,\nu}$ as a sum of three operators and deal with them one at a time. The identity operator is trivial. We have

$$\mathcal{I}_d\left(e^{-2\pi\imath\boldsymbol{\xi}\cdot\delta\mathbf{x}}\right) = e^{-2\pi\imath\boldsymbol{\xi}\cdot\delta\mathbf{x}}.$$

The Laplace operator yields

$$\begin{aligned}
\Delta\left(e^{-2\pi\imath\boldsymbol{\xi}\cdot\delta\mathbf{x}}\right) &= \left(\frac{\partial^2}{\partial\xi_1^2} + \frac{\partial^2}{\partial\xi_2^2}\right)\left(e^{-2\pi\imath\boldsymbol{\xi}\cdot\delta\mathbf{x}}\right) \\
&= \frac{\partial}{\partial\xi_1}\left((-2\pi\imath\delta\mathsf{x}_1)e^{-2\pi\imath\boldsymbol{\xi}\cdot\delta\mathbf{x}}\right) + \frac{\partial}{\partial\xi_2}\left((-2\pi\imath\delta\mathsf{x}_2)e^{-2\pi\imath\boldsymbol{\xi}\cdot\delta\mathbf{x}}\right) \\
&= \left((-2\pi\imath\delta\mathsf{x}_1)^2 + (-2\pi\imath\delta\mathsf{x}_2)^2\right)e^{-2\pi\imath\boldsymbol{\xi}\cdot\delta\mathbf{x}} \\
&= -4\pi^2|\delta\mathbf{x}|^2 e^{-2\pi\imath\boldsymbol{\xi}\cdot\delta\mathbf{x}}.
\end{aligned}$$

Lastly, computing for the third component gives

$$\begin{aligned}
\frac{\partial^2}{\partial\mathbf{e}_\lambda}e^{-2\pi\imath\boldsymbol{\xi}\cdot\delta\mathbf{x}} &= \frac{\partial}{\partial\mathbf{e}_\lambda}\left(\nabla e^{-2\pi\imath\boldsymbol{\xi}\cdot\delta\mathbf{x}}\cdot\mathbf{e}_\lambda\right) = \frac{\partial}{\partial\mathbf{e}_\lambda}\left(-2\pi\imath\delta\mathbf{x}\cdot\mathbf{e}_\lambda e^{-2\pi\imath\boldsymbol{\xi}\cdot\delta\mathbf{x}}\right) \\
&= \left(-2\pi\delta\mathbf{x}\cdot\mathbf{e}_\lambda\right)^2 e^{-2\pi\imath\boldsymbol{\xi}\cdot\delta\mathbf{x}} \\
&= -4\pi^2\langle\mathbf{e}_\lambda, \delta\mathbf{x}\rangle^2 e^{-2\pi\imath\boldsymbol{\xi}\cdot\delta\mathbf{x}}.
\end{aligned}$$

Thus, combining all of the above, we have

$$\mathcal{L}_{\lambda,\nu}\left(e^{-2\pi\imath\boldsymbol{\xi}\cdot\delta\mathbf{x}}\right) = \left(1 + \frac{4\pi^2|\delta\mathbf{x}|^2}{a_M} + 4\pi^2\frac{a_M^{-2}}{1 + a_M^{-1}|\delta\theta|^2}\langle\mathbf{e}_\lambda, \delta\mathbf{x}\rangle^2\right)e^{-2\pi\imath\boldsymbol{\xi}\cdot\delta\mathbf{x}},$$

as desired. □

Lemma A.1.3. *Define*

$$S_{\lambda,M,N_1,N_2}(r,\phi) = \min\left(1, a_\lambda(1+r)\right)^M (1+a_\lambda r)^{-N_1} \left(1 + a_\lambda^{-1/2}|\sin(\phi+\theta_\lambda)|\right)^{-N_2}.$$

The following holds

$$\left| \mathcal{L}_{\lambda,\nu}^k \left(\hat{\psi}_\lambda \left(\mathrm{D}_{a_\lambda}\mathrm{R}_{\theta_\lambda}\boldsymbol{\xi}\right) \overline{\hat{\phi}_\nu \left(\mathrm{D}_{a_\nu}\mathrm{R}_{\theta_\nu}\boldsymbol{\xi}\right)} \right) \right| \lesssim S_{\lambda,M-N_2,N_1,N_2}(\boldsymbol{\xi}) S_{\nu,M-N_2,N_1,N_2}(\boldsymbol{\xi}),$$

where $\mathcal{L}_{\lambda,\nu}$ is the differential operator defined in Equation (2.33).

Proof. Let us denote

$$A_{\lambda,M,N_1,N_2}(r,\phi) = \min\left(1, a_\lambda(1+r)\right)^M (1+a_\lambda r)^{-N_1} \left(1 + a_\lambda^{1/2}|\sin(\phi+\theta_\lambda)|\right)^{-N_2}. \quad \text{(A.1)}$$

Notice that the difference in the definitions of A_{λ,M,N_1,N_2} and S_{λ,M,N_1,N_2} is in the power of a_λ in the third factor.

By Lemma A.1.4 it is sufficient to show

$$\left| \mathcal{L}_{\lambda,\nu}^k \left(\hat{\psi}_\lambda \left(\mathrm{D}_{a_\lambda}\mathrm{R}_{\theta_\lambda}\boldsymbol{\xi}\right) \overline{\hat{\phi}_\nu \left(\mathrm{D}_{a_\nu}\mathrm{R}_{\theta_\nu}\boldsymbol{\xi}\right)} \right) \right| \lesssim A_{\lambda,M,N_1,N_2}(r,\varphi) A_{\nu,M,N_1,N_2}(r,\varphi).$$

We will construct our argument by induction in k. By Lemma A.1.7 the expression $\mathcal{L}_{\lambda,\nu} \left(\hat{\psi}_\lambda \left(\mathrm{D}_{a_\lambda}\mathrm{R}_{\theta_\lambda}\boldsymbol{\xi}\right) \overline{\hat{\phi}_\nu \left(\mathrm{D}_{a_\nu}\mathrm{R}_{\theta_\nu}\boldsymbol{\xi}\right)} \right)$ can be written as a finite linear combination of terms of the form $\hat{c}_\lambda \left(\mathrm{D}_{a_\lambda}\mathrm{R}_{\theta_\lambda}\boldsymbol{\xi}\right) \overline{\hat{d}_\nu \left(\mathrm{D}_{a_\nu}\mathrm{R}_{\theta_\nu}\boldsymbol{\xi}\right)}$ where the functions c_λ and d_ν satisfy the decay conditions (2.3) for $(R-2, M, N_1, N_2)$, that is, in the process we lose two degrees of smoothness. Therefore, by the same argument it follows that $\mathcal{L}_{\lambda,\nu}^2 \left(\hat{\psi}_\lambda \left(\mathrm{D}_{a_\lambda}\mathrm{R}_{\theta_\lambda}\boldsymbol{\xi}\right) \overline{\hat{\phi}_\nu \left(\mathrm{D}_{a_\nu}\mathrm{R}_{\theta_\nu}\boldsymbol{\xi}\right)} \right))$ simply means that we ought to apply $\mathcal{L}_{\lambda,\nu}$ to each of the terms $\hat{c}_\lambda \left(\mathrm{D}_{a_\lambda}\mathrm{R}_{\theta_\lambda}\boldsymbol{\xi}\right) \overline{\hat{d}_\nu \left(\mathrm{D}_{a_\nu}\mathrm{R}_{\theta_\nu}\boldsymbol{\xi}\right)}$, where in the process we lose two further degrees of smoothness.

This process can be repeated k times, for $k \leqslant R/2$ and we it follows that the expression $\mathcal{L}_{\lambda,\nu}^k \left(\hat{\psi}_\lambda \left(\mathrm{D}_{a_\lambda}\mathrm{R}_{\theta_\lambda}\boldsymbol{\xi}\right) \overline{\hat{\phi}_\nu \left(\mathrm{D}_{a_\nu}\mathrm{R}_{\theta_\nu}\boldsymbol{\xi}\right)} \right)$ can be written as a (finite) linear combination of terms of the form $\hat{c}_\lambda \left(\mathrm{D}_{a_\lambda}\mathrm{R}_{\theta_\lambda}\boldsymbol{\xi}\right) \overline{\hat{d}_\nu \left(\mathrm{D}_{a_\nu}\mathrm{R}_{\theta_\nu}\boldsymbol{\xi}\right)}$, where c_λ and d_ν satisfy bounds of the form

$$|\hat{c}_\lambda(\boldsymbol{\xi})| \lesssim \min\left(1, a_\lambda + |\xi_1| + a_\lambda^{1/2}|\xi_2|\right)^M \langle|\boldsymbol{\xi}|\rangle^{-N_1} \langle\xi_2\rangle^{-N_2}.$$

Now, denote

$$B_{\lambda,M,N_1,N_2}(\boldsymbol{\xi}) = \left(1, a_\lambda + |\xi_1| + a_\lambda^{1/2}|\xi_2|\right)^M \langle|\boldsymbol{\xi}|\rangle^{-N_1} \langle\xi_2\rangle^{-N_2}.$$

Since Lemma A.1.7 states that the coefficients of this linear combination are uniformly bounded, it follows

$$\left| \mathcal{L}_{\lambda,\nu}^k \left(\hat{\psi}_\lambda (\mathrm{D}_{a_\lambda}\mathrm{R}_{\theta_\lambda}\boldsymbol{\xi}) \overline{\hat{\phi}_\nu (\mathrm{D}_{a_\nu}\mathrm{R}_{\theta_\nu}\boldsymbol{\xi})} \right) \right| \quad \text{(A.2)}$$

$$\lesssim B_{\lambda,M,N_1,N_2}\left(\mathrm{D}_{a_\lambda}\mathrm{R}_{\theta_\lambda}\boldsymbol{\xi}\right) B_{\nu,M,N_1,N_2}\left(\mathrm{D}_{a_\nu}\mathrm{R}_{\theta_\nu}\boldsymbol{\xi}\right)$$

Since $\mathrm{D}_a\mathrm{R}_\theta\boldsymbol{\xi} = \big(a(\cos(\theta)\xi_1 - \sin(\theta)\xi_2), a^{1/2}(\sin(\theta)\xi_1 + \cos(\theta)\xi_2)\big)^\top$, we can write $\boldsymbol{\xi}$ in polar coordinates as $\boldsymbol{\xi} = (r\cos(\varphi), r\sin(\varphi))$ and get

$$\mathrm{D}_a\mathrm{R}_\theta\boldsymbol{\xi} = \begin{pmatrix} ar\cos(\theta+\varphi) \\ a^{1/2}r\sin(\theta+\varphi)\end{pmatrix} \tag{A.3}$$

Plugging (A.3) into (A.2) yields the desired statement.

□

Lemma A.1.4. *Consider A_{λ,M,N_1,N_2} as defined in Equation (A.1). For every $0 \leqslant L \leqslant N_2$ we have*

$$A_{\lambda,M,N_1,N_2}(r,\phi) \lesssim S_{\lambda,M-L,N,L}(r,\phi).$$

Proof. We will omit indices λ throughout the proof. Since $L \leqslant N_2$ it follows

$$\Big(1 + a^{1/2}r|\sin(\phi+\theta)|\Big)^{-N_2} \leqslant \Big(1 + a^{1/2}r|\sin(\phi+\theta)|\Big)^{-L}.$$

Therefore,

$$\frac{\min(1,a(1+r))^M}{(1+ar)^{N_1}\left(1+a^{1/2}r|\sin(\phi+\theta)|\right)^{N_2}} \leqslant \frac{\min\left(1,a(1+r)\right)^{M-L}}{(1+ar)^{N_1}}\left(\frac{\min\left(1,a(1+r)\right)}{1+a^{1/2}r|\sin(\phi+\theta)|}\right)^{L}.$$

What is left for us to do is to show the inequality

$$\frac{\min\left(1,a(1+r)\right)}{1+a^{1/2}r|\sin(\phi+\theta)|} \lesssim \frac{1}{1+a^{-1/2}|\sin(\phi+\theta)|}.$$

We have the following cases

(1) $r \geqslant a^{-1}$. Then $\min\left(1,a(1+r)\right) = 1$ and $a^{1/2}r \geqslant a^{-1/2}$. Thus

$$\frac{\min\left(1,a(1+r)\right)}{1+a^{1/2}r|\sin(\phi+\theta)|} \leqslant \frac{1}{1+a^{-1/2}|\sin(\phi+\theta)|}.$$

(2) $a_0^{-1} \leqslant r \leqslant a^{-1}$. We distinguish between two further cases.

(A) $\min\left(1,a(1+r)\right) = 1$.

This implies $a \geqslant \frac{1}{1+r}$. Therefore, because $(1+r)^{-1} \geqslant \frac{\min(1,a_0^{-1})}{2}r^{-1}$, we have $1 \lesssim ar$ and consequently $a^{-1/2} \lesssim a^{1/2}r$ which gives

$$\frac{\min\left(1,a(1+r)\right)}{1+a^{1/2}r|\sin(\phi+\theta)|} = \frac{1}{1+a^{1/2}r|\sin(\phi+\theta)|} \lesssim \frac{1}{1+a^{-1/2}|\sin(\phi+\theta)|}.$$

(B) $\min(1, a(1+r)) = a(1+r)$. We have

$$\begin{aligned}\frac{\min(1, a(1+r))}{1 + a^{1/2} r|\sin(\phi+\theta)|} &= \frac{a(1+r)}{1 + a^{1/2} r|\sin(\phi+\theta)|} \\ &= \frac{1+r}{r} \frac{1}{a^{-1}/r + a^{-1/2}|\sin(\phi+\theta)|} \\ &\lesssim \frac{1}{1 + a^{-1/2}|\sin(\phi+\theta)|},\end{aligned}$$

since $r^{-1} + 1 \leqslant (a_0 + 1)$.

(3) $r \leqslant a_0^{-1}$. Here we have

$$\frac{\min(1, a(1+r))}{1 + a^{1/2} r|\sin(\phi+\theta)|} \leqslant \min(1, a(1+r)) \leqslant a(1 + a_0^{-1}) \lesssim a.$$

Therefore, we conclusion will follow provided we can show

$$a \lesssim \frac{1}{1 + a^{-1/2} r|\sin(\phi+\theta)|}.$$

To this end, let us define $\mathbf{u} = (1, \sin(\phi+\theta))^\top$ and $\mathbf{v} = (1, a^{-1/2})^\top$. Using the Cauchy-Schwarz inequality and inequality of arithmetic and geometric means, we have

$$1 + a^{-1/2}|\sin(\phi+\theta)| = |\langle \mathbf{u}, \mathbf{v} \rangle| \leqslant \|\mathbf{u}\|\|\mathbf{v}\| \leqslant \sqrt{2(1+a^{-1})} \leqslant \frac{3 + a^{-1}}{2} \lesssim a^{-1}$$

That is, we have $1 + a^{-1/2}|\sin(\phi+\theta)| \lesssim a^{-1}$, which is what we wanted to show.

□

Lemma A.1.5. *Assume $M > A - 5/4$, $N_2 \geqslant B$ and $N_1 \geqslant A + 3/4$ hold for positive constants A and B. We then have*

$$(a_\lambda a_\nu)^{3/4} \int_{\mathbb{R}^2} S_{\lambda,M,N_1,N_2}(\boldsymbol{\xi}) S_{\nu,M,N_1,N_2}(\boldsymbol{\xi}) d\boldsymbol{\xi} \lesssim \left(\frac{a_M}{a_m}\right)^{-A} \left(1 + a_M^{-1/2}|\theta_\lambda - \theta_\nu|\right)^{-B}. \quad \text{(A.4)}$$

Proof. Without loss of generality we can assume $a_\lambda \leqslant a_\nu$, and denote

$$I_{\lambda,\nu} = \left[\int_{[0,2\pi)} \left(1 + a_\lambda^{-1/2}|\sin(\phi+\theta_\lambda)|\right)^{-N_2} \left(1 + a_\nu^{-1/2}|\sin(\phi+\theta_\nu)|\right)^{-N_2} d\phi\right], \quad \text{(A.5)}$$

and

$$\mathcal{S} = a_\nu^2 \int_{\mathbb{R}^+} \min(1, a_\lambda(1+r))^M \min(1, a_\nu(1+r))^M (1 + a_\lambda r)^{-N_1} (1 + a_\nu r)^{-N_1} r dr.$$

In polar coordinates we have

$$\int_{\mathbb{R}^2} S_{\lambda,M,N_1,N_2}(\boldsymbol{\xi}) S_{\nu,M,N_1,N_2}(\boldsymbol{\xi}) d\boldsymbol{\xi} = a_\nu^{-2}\, I_\phi\, \mathcal{S}.$$

With the help of Lemma A.1.6 we have

$$I_{\lambda,\nu} \lesssim a_\lambda^{1/2} \left(1 + a_\nu^{-1/2}|\delta\theta|\right)^{-N_2}.$$

Therefore,

$$(a_\lambda a_\nu)^{3/4} \int_{\mathbb{R}^2} S_{\lambda,M,N_1,N_2}(\boldsymbol{\xi}) S_{\nu,M,N_1,N_2}(\boldsymbol{\xi}) d\boldsymbol{\xi} \lesssim \mathcal{S}(a_\nu/a_\lambda)^{3/4} \left(1 + a_\nu^{-1/2}|\delta\theta|\right)^{-N_2}$$

What is left is to show $\mathcal{S} \lesssim (a_\nu/a_\lambda)^{-A-3/4}$. To that end, let us split $\mathbb{R}^+$ into three pieces

$$\left(0, \max\left(0, a_\nu^{-1} - 1\right)\right) \cup \left(\max\left(0, a_\nu^{-1} - 1\right), \max\left(0, a_\lambda^{-1} - 1\right)\right) \cup \left(\max\left(0, a_\lambda^{-1} - 1\right), \infty\right)$$

and rewrite the integral in the definition of $\mathcal{S}$ accordingly. In other words, we now write $\mathcal{S} = {a_\nu}^2(I_1 + I_2 + I_3)$.

Without loss of generality we can assume

$$\max(0, a_\nu^{-1} - 1) = a_\nu^{-1} - 1, \text{ and } \max(0, a_\lambda^{-1} - 1) = a_\lambda^{-1} - 1.$$

It follows

$$\begin{aligned}
&\int_0^{a_\nu^{-1}-1} \min(1, a_\lambda(1+r))^M \min(1, a_\nu(1+r))^M \underbrace{(1+a_\lambda r)^{-N_1}}_{\leqslant 1} \underbrace{(1+a_\nu r)^{-N_1}}_{\leqslant 1} r dr \\
&\quad \leqslant \int_0^{a_\nu^{-1}-1} \min(1, a_\lambda(1+r))^M \min(1, a_\nu(1+r))^M r dr \\
&\quad \leqslant (a_\nu a_\lambda)^M \int_0^{a_\nu^{-1}-1} (1+r)^{2M+1} dr \leqslant (a_\lambda a_\nu)^M \frac{(1+r)^{2M+2}}{2M+2}\Big|_0^{a_\nu^{-1}-1} \\
&\quad \lesssim a_\lambda^M a_\nu^{-M-2}
\end{aligned}$$

For I_2 we have

$$\begin{aligned}
&\int_{a_\nu^{-1}-1}^{a_\lambda^{-1}-1} \min(1, a_\lambda(1+r))^M \min(1, a_\nu(1+r))^M (1+a_\lambda r)^{-N_1} \underbrace{(1+a_\nu r)^{-N_1}}_{\leqslant 1} r dr \\
&\qquad \leqslant \int_{a_\nu^{-1}-1}^{a_\lambda^{-1}-1} (a_\lambda(1+r))^M (1+a_\lambda r)^{-N_1} (1+r) dr
\end{aligned}$$

Since $a_\lambda \lesssim 1$ we have $a_\lambda(1+r) \lesssim 1 + a_\lambda r$, thus

$$\begin{aligned} I_2 &\lesssim a_\lambda^M a_\nu^{-N_1} \int_{a_\nu^{-1}-1}^{a_\lambda^{-1}-1} (1+r)^{M-N_1+1} dr \lesssim a_\lambda^M a_\nu^{-N_1} \frac{(1+r)^{M+2-N_1}}{M+2-N_1}\Big|_{a_\nu^{-1}-1}^{a_\lambda^{-1}-1} \\ &\lesssim a_\lambda^M a_\nu^{-N_1} \left(a_\lambda^{N_1-2-M} - a_\nu^{N_1-2-M}\right) \lesssim a_\lambda^{N_1-2} a_\nu^{-N_1} \end{aligned}$$

Lastly, for I_3 we compute

$$\begin{aligned} I_3 &= \int_{a_\lambda^{-1}-1}^{\infty} (1+a_\lambda r)^{-N_1}(1+a_\nu r)^{-N_1} r dr \leqslant (a_\lambda a_\nu)^{-N_1} \int_{a_\lambda^{-1}-1}^{\infty} (1+r)^{1-2N_1} dr \\ &\leqslant (a_\lambda a_\nu)^{-N_1} \frac{(1+r)^{2-2N_1}}{2-2N_1}\Big|_{a_\lambda^{-1}-1}^{\infty} \lesssim a_\lambda^{N_1-2} a_\nu^{-N_1} \end{aligned}$$

Combining the bounds for I_1, I_2 and I_3, and using the assumptions stated in the wording of the lemma, we have

$$\mathcal{S} \lesssim a_\nu^2 \left(a_\lambda^M a_\nu^{M-2} + a_\lambda^{N_1-2} a_\nu^{-N_1}\right) = (a_\lambda/a_\nu)^M + (a_\lambda/a_\nu)^{N_1-2} \lesssim (a_\lambda/a_\nu)^{-A-3/4}$$

which is what we wanted to show. □

Lemma A.1.6. *For $a_\lambda \leqslant a_\nu$ and a $N \in \mathbb{N}$ the following holds*

$$\begin{aligned} \int_{[0,2\pi)} \left(1 + a_\lambda^{-1/2}|\sin(\phi+\theta_\lambda)|\right)^{-N} \left(1 + a_\nu^{-1/2}|\sin(\phi+\theta_\nu)|\right)^{-N} d\phi \\ \lesssim a_\lambda^{1/2} \left(1 + a_\nu^{-1/2}|\delta\theta|\right)^{-N_2}. \end{aligned}$$

Proof. Application of Lemma 5.2 from [9], where we replace ϕ with $\phi + \theta_\lambda$ and θ with $\delta\theta$, yields

$$I_{\lambda,\nu} \lesssim \max(a_\lambda^{-1/2}, a_\nu^{-1/2})^{-1} \left(1 + \min(a_\lambda^{-1/2}, a_\nu^{-1/2})|\delta\theta|\right)^{-N},$$

For $I_{\lambda,\nu}$ defined in (A.5). Since, $a_\lambda \leqslant a_\nu$ we have $a_\lambda^{-1/2} \geqslant a_\nu^{-1/2}$ and it follows

$$I_\phi \lesssim a_\lambda^{1/2} \left(1 + a_\nu^{-1/2}|\delta\theta|\right)^{-N},$$

giving the desired statement. □

Lemma A.1.7. *Let ψ_λ and ϕ_ν be functions that satisfy the decay conditions (2.3) for (R, M, N_1, N_2). Then the expression $\mathcal{L}_{\lambda,\nu}\left(\hat{\psi}_\lambda(\mathrm{D}_{a_\lambda}\mathrm{R}_{\theta_\lambda}\boldsymbol{\xi})\hat{\phi}_\nu(\mathrm{D}_{a_\nu}\mathrm{R}_{\theta_\nu}\boldsymbol{\xi})\right)$ can be written as a finite linear combination of terms of the form*

$$\hat{c}_\lambda(\mathrm{D}_{a_\lambda}\mathrm{R}_{\theta_\lambda}\boldsymbol{\xi})\overline{\hat{d}_\nu(\mathrm{D}_{a_\nu}\mathrm{R}_{\theta_\nu}\boldsymbol{\xi})}$$

such that $\hat{c}_\lambda$ and $\hat{d}_\nu$ satisfy (2.3) with $(R-2, M, N_1, N_2)$.

Proof. This is again essentially just a simple modification of proofs in [5,9]. Let us again reexamine the differential operator we are dealing with. We have

$$\mathcal{L}_{\lambda,\nu} = \mathcal{I}_d - a_M^{-1}\Delta - \frac{a_M^{-2}}{1 + a_M^{-1}|\theta_\lambda - \theta_\nu|^2}\frac{\partial^2}{\partial \mathbf{e}_\lambda^2}. \tag{A.6}$$

Therefore, we can again analyse through its three component parts.

It trivially follows that the first term,

$$\mathcal{I}_d\left(\hat{\psi}_\lambda(\mathrm{D}_{a_\lambda}\mathrm{R}_{\theta_\lambda}\boldsymbol{\xi})\hat{\phi}_\nu(\mathrm{D}_{a_\nu}\mathrm{R}_{\theta_\nu}\boldsymbol{\xi})\right),$$

satisfies the claim of the lemma.

Regarding the second term, $a_M^{-1}\Delta$, we can first observe that the Laplacian of a product of two functions is given as

$$\Delta(fg) = (\Delta f)\, g + f\,(\Delta g) + 2\left(\sum_{|\boldsymbol{\alpha}|=1} \partial^{\boldsymbol{\alpha}} f \partial^{\boldsymbol{\alpha}} g\right).$$

We can observe that it will follows easily that the result is a finite linear combination of the terms of the same form but with reduced smoothness, since if for example f is in C^R then Δf is in C^{R-2}. Thus, the only element that is not clear is whether the coefficients of the linear combination are uniformly bounded for all a, since we have to multiply the Laplacian by a_M^{-1}, as in (A.6) Let us assess the term $(\Delta f)\, g$ in a bit more detail for the sake of clarity and so that we give the idea how will the rest of the proof work. Plugging in for

$$f(\boldsymbol{\xi}) = \hat{\psi}_\lambda(\mathrm{D}_{a_\lambda}\mathrm{R}_{\theta_\lambda}\boldsymbol{\xi}), \text{ and } g(\boldsymbol{\xi}) = \hat{\phi}_\nu(\mathrm{D}_{a_\nu}\mathrm{R}_{\theta_\nu}\boldsymbol{\xi}),$$

we get

$$(\Delta f)\, g = \left(a_\lambda \partial^{(0,2)} f(\boldsymbol{\xi}) + a_\lambda^2 \partial^{(2,0)} f(\boldsymbol{\xi})\right) g(\boldsymbol{\xi}).$$

Therefore,

$$a_M^{-1}\,(\Delta f)\, g \leqslant \partial^{(0,2)} f(\boldsymbol{\xi}) g(\boldsymbol{\xi}) + a_\lambda \partial^{(2,0)} f(\boldsymbol{\xi}) g(\boldsymbol{\xi}),$$

which is a linear combination of terms of the form

$$\hat{c}_\lambda(\mathrm{D}_{a_\lambda}\mathrm{R}_{\theta_\lambda}\boldsymbol{\xi})\overline{\hat{d}_\nu(\mathrm{D}_{a_\nu}\mathrm{R}_{\theta_\nu}\boldsymbol{\xi})}$$

where $\hat{c}_\lambda$ and $\hat{d}_\nu$ satisfy (2.3) with $(R-2, M, N_1, N_2)$, as desired. The term $f\,(\Delta g)$ can be dealt with analogously. Regarding the last term, let $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)\mathbb{N}_0^2$ be a multi-index such that $|\boldsymbol{\alpha}| = 1$. A simple computation yields

$$\sum_{|\boldsymbol{\alpha}|=1} \partial^{\boldsymbol{\alpha}} f(\boldsymbol{\xi}) \partial^{\boldsymbol{\alpha}} g(\boldsymbol{\xi}) = \sum_{|\boldsymbol{\alpha}|=1, |\boldsymbol{\beta}|=1} a_\lambda^{\alpha_1 + \frac{\alpha_2}{2}} a_\nu^{\beta_1 + \frac{\beta_2}{2}} c_{\theta_\lambda,\theta_\nu} \partial^{\boldsymbol{\alpha}} f(\boldsymbol{\xi}) \partial^{\boldsymbol{\beta}} g(\boldsymbol{\xi}),$$

where $c_{\theta_\lambda,\theta_\nu}$ is a trigonometric function and $c_{\theta_\lambda,\theta_\nu} \leqslant 1$. Therefore, since

$$a_M^{-1} a_\lambda^{\alpha_1 + \frac{\alpha_2}{2}} a_\nu^{\beta_1 + \frac{\beta_2}{2}} \lesssim 1, \text{ for } |\boldsymbol{\alpha}| = 1 \text{ and } |\boldsymbol{\beta}| = 1,$$

it follows that $a_M^{-1}\sum_{|\boldsymbol{\alpha}|=1}\partial^{\boldsymbol{\alpha}}f\partial^{\boldsymbol{\alpha}}g$ satisfies the claim of the lemma.

Let us now address the last remaining term, $\frac{a_M^{-2}}{1+a_M^{-1}|\theta_\lambda-\theta_\nu|^2}\frac{\partial^2}{\partial\mathbf{e}_\lambda^2}$. First of all, we have

$$\frac{\partial^2}{\partial\mathbf{e}_\lambda^2}(fg)(\boldsymbol{\xi})=\sum_{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|=2}C_{\boldsymbol{\alpha},\boldsymbol{\beta},\theta_\lambda,\theta_\nu}\partial^{\boldsymbol{\alpha}}f(\boldsymbol{\xi})\partial^{\boldsymbol{\beta}}g(\boldsymbol{\xi}),$$

and

$$C_{\boldsymbol{\alpha},\boldsymbol{\beta},\theta_\lambda,\theta_\nu}=a_\lambda^{\alpha_1+\frac{\alpha_2}{2}}a_\nu^{\beta_1+\frac{\beta_2}{2}}C_{\theta_\lambda,\theta_\nu},$$

where $C_{\theta_\lambda,\theta_\nu}\lesssim 1$. Therefore, we have to show that $\frac{a_M^{-2}}{1+a_M^{-1}|\theta_\lambda-\theta_\nu|^2}C_{\boldsymbol{\alpha},\boldsymbol{\beta},\theta_\lambda,\theta_\nu}\lesssim 1$. Let us without loss of generality take $\theta_\lambda=0$, as that simplifies some of the expressions.

Given the restriction $|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|=2$, the exponent $\alpha_1+\beta_1+\frac{\alpha_2+\beta_2}{2}$ assumes only three values, $1,\frac{3}{2}$ and 2. Since $\frac{1}{1+a_M^{-1}|\theta_\nu|^2}\leqslant 1$, it follows that whenever $\alpha_1+\beta_1+\frac{\alpha_2+\beta_2}{2}=2$ we have

$$\frac{a_M^{-2}}{1+a_M^{-1}|\theta_\nu|^2}a_\lambda^{\alpha_1+\frac{\alpha_2}{2}}a_\nu^{\beta_1+\frac{\beta_2}{2}}\lesssim 1.$$

Terms corresponding to $\alpha_1+\beta_1+\frac{\alpha_2+\beta_2}{2}=1$ are

$$a_\nu\sin^2(\theta_\nu)f(\boldsymbol{\xi})\partial^{(0,2)}g(\boldsymbol{\xi}).$$

Using the fact $\sin(\theta)\leqslant\theta$ we have

$$\frac{a_M^{-2}}{1+a_M^{-1}|\theta_\nu|^2}a_\nu\sin^2(\theta_\nu)\lesssim\frac{a_M^{-1}}{|\theta_\nu|^2}a_\nu|\theta_\nu|^2\leqslant 1.$$

On the other hand, the terms corresponding to $\alpha_1+\beta_1+\frac{\alpha_2+\beta_2}{2}=\frac{3}{2}$ are

$$a_\lambda a_\nu^{1/2}\sin(\theta_\nu)\partial^{(1,0)}f(\boldsymbol{\xi})\partial^{(0,1)}g(\boldsymbol{\xi})+2a_\nu^{3/2}\sin(\theta_\nu)\cos(\theta_\nu)f(\boldsymbol{\xi})\partial^{(1,1)}g(\boldsymbol{\xi}).$$

Therefore, it is sufficient to show

$$\frac{a_M^{-1/2}\sin(\theta_\nu)}{1+a_M^{-1}|\theta_\nu|^2}=a_M^{-1/2}\frac{a_M^{2/2}\sin(\theta_\nu)}{a_M+|\theta_\nu|^2}\lesssim 1.$$

On one hand, if $|\theta_\nu|\leqslant a_M^{1/2}$ we have

$$\frac{a_M^{1/2}\sin(\theta_\nu)}{a_M+|\theta_\nu|^2}\leqslant a_M^{1/2}\sin(\theta_\nu)\lesssim 1,$$

since $\sin(\theta_\nu)\leqslant\theta_\nu$. On the other hand, for $|\theta_\nu|\geqslant a_M^{1/2}s$ we have

$$\frac{a_M^{1/2}\sin(\theta_\nu)}{a_M+|\theta_\nu|^2}\leqslant\frac{\sin(\theta_\nu)}{a_M^{1/2}+|\theta_\nu|}\lesssim 1.$$

□

A.2 Additional Proofs For Chapter 4

Lemma A.2.1. *Let $f \in C^m$ and let $P_k(\mathbf{x})$ be its corresponding Taylor polynomial around $\mathbf{p}$ of degree k where $k > l$. Assume that $\Gamma = \{m_\lambda : \lambda \in \Lambda_\Gamma\}$ is a family of continuous parabolic molecules of high-enough order and that Ω is a set that satisfies the conditions of Theorem 4.4.4. Then*

$$\langle f\,\chi_\Omega - P_{\boldsymbol{\alpha}}\chi_\Omega, m_\lambda\rangle \lesssim a_\lambda^K, \tag{A.7}$$

holds for $K > \frac{5}{4}$.

Proof. Without loss of generality let us assume $\mathbf{p} = \mathbf{0}$ and drop the λ indices. We have

$$\begin{aligned}
|\langle f\chi_\Omega - P_k\chi_\Omega, m\rangle| &\leqslant \int_{\mathbb{R}^2} |f(\mathbf{x}) - P_k(\mathbf{x})||\chi_\Omega(\mathbf{x})||m_\lambda(\mathbf{x})|d\mathbf{x} \\
&= \underbrace{\int_{\mathcal{B}_{a^\gamma}(\mathbf{0})} |f(\mathbf{x}) - P_k(\mathbf{x})||\chi_\Omega(\mathbf{x})||m(\mathbf{x})|d\mathbf{x}}_{I_1} \\
&\qquad + \underbrace{\int_{\mathcal{B}^c_{a^\gamma}(\mathbf{0})} |f(\mathbf{x}) - P_k(\mathbf{x})||\chi_\Omega(\mathbf{x})||m(\mathbf{x})|d\mathbf{x}}_{I_2}.
\end{aligned}$$

Estimating I_1 gives

$$\begin{aligned}
I_1 &= \int_{\mathcal{B}_{a^\gamma}(\mathbf{0})} |f(\mathbf{x}) - P_k(\mathbf{x})||\chi_\Omega(\mathbf{x})||m(\mathbf{x})|d\mathbf{x} \\
&\lesssim a^{-3/4} \int_{\mathcal{B}_{a^\gamma}(\mathbf{0})} |f(\mathbf{x}) - P_k(\mathbf{x})||\chi_\Omega(\mathbf{x})|d\mathbf{x} \\
&\lesssim \int_{\mathcal{B}_{a^\gamma}(\mathbf{0})} |\mathbf{x}|^k d\mathbf{x} \\
&\lesssim a^{\gamma(k+2)},
\end{aligned}$$

where $\gamma > 0$. On the other hand, for I_2 we have

$$\begin{aligned}
I_2 &= \int_{\mathcal{B}^c_{a^\gamma}(\mathbf{0})} |f(\mathbf{x}) - P_k(\mathbf{x})||\chi_\Omega(\mathbf{x})||m(\mathbf{x})|d\mathbf{x} \\
&\lesssim \int_{\mathcal{B}^c_{a^\gamma}(\mathbf{0})} |m(\mathbf{x})|d\mathbf{x} \lesssim \int_{\mathcal{B}^c_{a^\gamma}(\mathbf{0})} |\varphi(\mathbf{x})|\,d\mathbf{x} \\
&\lesssim \int_{\mathcal{B}^c_{a^\gamma}(\mathbf{0})} \left(1 + a^{-2}\mathsf{x}_1^2 + a^{-1}\mathsf{x}_2^2\right)^{-N} d\mathbf{x} \\
&\lesssim \int_{\mathcal{B}^c_{a^\gamma}(\mathbf{0})} a^N \left(\mathsf{x}_1^2 + \mathsf{x}_2^2\right)^{-N} d\mathbf{x} \\
&\lesssim a^{N-3/4} \int_{a^\gamma}^\infty r^{-1-2N} dr
\end{aligned}$$

$$\lesssim a^{N(1-2\gamma)-3/4}$$

Therefore, the statement holds for $0 < \gamma < \frac{1}{2}$ such that and $\gamma(k+2) > \frac{5}{4}$ and provided N is big enough. $\square$

Appendix B

Notation

List of Symbols

Symbol	Meaning
Miscellaneous	
x, ξ, etc.	Scalars
$\mathbf{b}, \mathbf{p}, \mathbf{x}$, etc.	d-dimensional vectors, that is, elements of $\mathbb{R}^d$ for $d \geqslant 1$
A, M, etc	Matrices
D_a	Parabolic dilation matrix $\mathrm{D}_a = \operatorname{diag}(a, a^{1/2})$, for $a > 0$
R_θ	Rotation matrix $\mathrm{R}_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$
$\boldsymbol{\alpha}$	Multi-index $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}_0^d$
$\mathbf{x}^{\boldsymbol{\alpha}}$	Defined as $\mathbf{x}^{\boldsymbol{\alpha}} = \prod_{i=1}^d \mathsf{x}_i$, for $\boldsymbol{\alpha} \in \mathbb{N}_0^d$
$\lesssim$	$A \lesssim B$ means $A \leqslant C \cdot B$ with $C > 0$ independent of A, B
$\sim$	Equivalence; $A \sim B$ if $A \lesssim B$ and $B \lesssim A$
$\langle \cdot, \cdot \rangle$	Inner product, or the action of a distribution
$\|f\|_p$, $p \in [1, \infty]$	L^p norm of the function f
$\|\mathbf{x}\|_2$	Euclidean norm of a vector $\mathbf{x} \in \mathbb{R}^d$
$\lvert\boldsymbol{\alpha}\rvert$	For $\boldsymbol{\alpha} \in \mathbb{N}_0^d$ defeind as $\lvert\boldsymbol{\alpha}\rvert = \alpha_1 + \cdots \alpha_d$
$\langle \cdot \rangle$	Regularised 1-norm, $\langle \mathsf{x} \rangle = \left(1 + \mathsf{x}^2\right)^{1/2}$
$\|\mathrm{A}\|_2$	Operator norm of a matrix
Sets	
$\mathbb{N}, \mathbb{Z}, \mathbb{R}, \dots$	Standard sets of (positive) integers, real numbers, etc.
$\mathbb{P}$	Parameter space $\mathbb{P} = \mathbb{R}^+ \times [0, 2\pi) \times \mathbb{R}^2$
$\Lambda_0, \Lambda_\Sigma, \Lambda_\Gamma$	Subsets of $\mathbb{P}$
$\operatorname{sing\,supp} f$	Singular support of f. Defined in (3.2)
$\mathrm{WF}(f)$	Wavefront set of f. Defined in (3.4)
$\mathcal{B}_r(\mathbf{x})$	$\{\mathbf{x} : \|\mathbf{x}\| < r\}$, open ball of radius r centred at $\mathbf{x}$

$\lvert\mathcal{J}\rvert$	Number of elements in a set $\mathcal{J}$
$\mathcal{W}_{\eta_1,\eta_2}$	Angular wedge with angles η_1, η_2. Defined in (4.5)

Function Spaces

C^m	m times continuously differentiable functions
C^∞	Infinitely differentiable functions
$L^p(\Omega)$, $p \in [1, \infty)$	Lebesgue measurable functions such that $\int_\Omega \lvert f(\mathbf{x})\rvert^p d\mathbf{x} < \infty$
$L^\infty(\Omega)$	Lebesgue measurable functions such that $\lvert f(\mathbf{x})\rvert < \infty$ for almost every $\mathbf{x} \in \Omega$
$\mathcal{S}(\Omega)$	Schwartz space of functions
$H^k(\theta_0, \mathbf{x}_0)$	Microlocal Sobolev space. Defined in (3.1.2)

Operators

$\mathcal{I}_d$	Identity
$f'(\mathsf{x})$	Derivative of a one-dimensional function
$\partial^{\boldsymbol{\alpha}} f(\mathbf{x})$	Partial derivative $\partial^{\boldsymbol{\alpha}} f = \frac{\partial^{\lvert\boldsymbol{\alpha}\rvert} f}{\partial \mathsf{x}_1^{\alpha_1} \cdots \partial \mathsf{x}_d^{\alpha_d}}$, with $\boldsymbol{\alpha} \in \mathbb{N}^d$
Δf	Laplacian $\Delta f = \sum_{i=1}^d \frac{\partial^2 f}{\partial \mathsf{x}_i^2}$
$f \star g$	Convolution $(f \star g)(\mathbf{x}) = \int_{\mathbb{R}^d} f(\mathbf{x} - \mathbf{y}) g(\mathbf{y}) d\mathbf{y}$
$\mathsf{T}_\mathbf{p}$	Translation operator $\mathsf{T}_\mathbf{p} f(\mathbf{x}) = f(\mathbf{x} - \mathbf{p})$
χ_Ω	Indicator function of a set $\Omega \in \mathbb{R}^d$. Defined as $\chi_\Omega(\mathbf{x}) = 1$ if $\mathbf{x} \in \Omega$, and 0 otherwise
δ	Dirac delta distribution. Defined in 1.16
PV	Cauchy principal value. Defined in 1.18
H	Heaviside function. Defined in 1.17
H_{η_1,η_2}	Indicator function of the wedge $\mathcal{W}_{\eta_1,\eta_2}$

Glossary of Acronyms

Acronym	**Meaning**
CPM	Continuous Parabolic Molecule
FIO	Fourier Integral Operator
FT	Fourier Transform
STFT	Short Time Fourier Transform

Bibliography

[1] E. J. Candès and D. L. Donoho. New Tight Frames of Curvelets and Optimal Representations of Objects with Piecewise C^2 Singularities. *Comm. Pure Appl. Math.*, 57(2):219--266, 2004.

[2] D. L. Labate, W-Q. Lim, G. Kutyniok, and G. Weiss. Sparse Multidimensional Representation Using Shearlets. In *Wavelets XI (San Diego, CA)*, volume 5914. SPIE, SPIE Proc., 2005.

[3] M. N. Do and M. Vetterli. The Contourlet Transform: an Efficient Directional Multiresolution Image Representation. *IEEE Trans. Image Process.*, 14(12):2091--2106, 2015.

[4] H. F. Smith. A Parametrix Construction for Wave Equations with $C^{1,1}$ Coefficients. *Ann. de l'Institut Fourier*, pages 797--835, 1998.

[5] E. J. Candès and L. Demanet. The Curvelet Representation of Wave Propagators is Optimally Sparse. *Comm. Pure Appl. Math.*, 58:1472--1528, 2004.

[6] E. J. Candès, D. L. Donoho, and J-L. Starck. The Curvelet Transform for Image Denoising. *IEEE Trans. Image Process.*, 11(6):131--141, 2002.

[7] E. J. King, G. Kutyniok, and X. Zhuang. Analysis of Inpainting via Clustered Sparsity and Microlocal Analysis. *J. Math. Imaging Vision*, 48:205--234, 2014.

[8] K. Guo and D. Labate. Representation of Fourier Integral Operators Using Shearlets. *J. Fourier Anal. Appl.*, 14:327--371, 2008.

[9] P. Grohs and G. Kutyniok. Parabolic Molecules. *Found. Comput. Math.*, pages 299--337, 2013.

[10] A. P. Calderón. Intermediate Spaces and Interpolation, the Complex Method. *Studia Math.*, 24:113--190, 1964.

[11] S. Mallat. *A Wavelet Tour of Signal Processing.* Academic Press, third edition, 2009.

[12] A. B. Watson. The Cortex Transform: Rapid Computation of Simulated Neural Images. *Comp. Vision, Graph. and Image Proc.*, 39(3):311--327, 1987.

[13] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger. Shiftable Transforms. *IEEE Trans. Inform. Theory*, 38(2):587--607, 1992.

[14] Environmental Resources Management Australia. Parameters of Human Vision Parameters of Human Vision and Viewshed Definition. `http://www.stockyardhillwindfarm.com.au/pdf/PPAR_Annexes/ATS/Annexes/Annex_J/AnnexJ-LVA_PART_12.pdf`, 2009.

[15] E. J. Candès and D. L. Donoho. Continuous Curvelet Transform: I. Resolution of the Wavefront Set. *Appl. Comput. Harmon. Anal.*, 19:162--197, 2003.

[16] S. Mallat and G. Peyré. A Review of Bandlet Methods for Geometrical Image Representation. *Numer. Algorithms*, 44(3):205--234, 2007.

[17] I. M. Gelfand and G. E. Shilov. *Generalized Functions: Volume I - II.* Academic Press, 1964.

[18] R. S. Strichartz. *A Guide to Distribution Theory and Fourier Transforms.* World Scientific, 2003.

[19] L. Hörmander. *The Analysis of Linear Partial Differential Operators I.* Springer, second edition, 1990.

[20] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic Decomposition by Basis Pursuit. *SIAM J. Sci. Comp.*, 20(1):33--61, 1998.

[21] W. K. Allard, G. Chen, and M. Maggioni. Multi-Scale Geometric Methods for Data Sets II: Geometric Multi-Resolution Analysis. *Appl. Comput. Harmon. Anal.*, 32(3):435--462, 2011.

[22] K. Gröchenig. *Foundations of Time-Frequency Analysis.* Birkhäuser, 2001.

[23] D. Gabor. Theory of Communication. *J. IEEE (London)*, 93(3):429--457, 1946.

[24] I. Daubechies, A. Grossman, and Y. Meyer. Painless Nonorthogonal Expansions. *J. Math. Phys.*, 27(5):1271--1283, 1986.

[25] M. Fornasier and H. Rauhut. Continuous Frames, Function Spaces, and the Discretisation Problem. *J. Fourier Anal. Appl.*, 11(3):245--287, 2005.

[26] G. Weiss and E. N. Wilson. *Twentieth Century Harmonic Analysis --- A Celebration*, chapter The Mathematical Theory of Wavelets, pages 329--366. Springer Netherlands, 2001.

[27] A. Grossman and J. Morlet. Decomposition of Hardy Functions into Square Integrable Wavelets of Constant Shapeof Hardy Functions into Square Integrable Wavelets of Constant Shape. *SIAM J. Mat. Anal.*, 15(4):723--736, 1984.

[28] F. W. King. *Hilbert Transforms: Volume I - II.* Cambridge University Press, 2009.

[29] S. Mallat and W. L. Hwang. Singularity Detection and Processing with Wavelets. *IEEE Trans. Inform. Theory*, 38(2):617--643, 1992.

[30] D. L. Donoho. Sparse Components of Images and Optimal Atomic Decompositions. *Constr. Approx.*, 17(3):353--382, 2001.

[31] E. J. Candès. *Ridgelets: Theory and Applications*. PhD thesis, Stanford University, 1998.

[32] E. J. Candès and D. L. Donoho. Ridgelets: a Key to Higher-Dimensional Intermittency? *Phil. Trans. R. Soc. Lond. A.*, 357:2495--2509, 1999.

[33] J. D. McEwen. Ridgelet Transform on the Sphere. IEEE Trans. Sig. Proc., submitted, 2015.

[34] J. Fadili and J-L. Starck. *Encyclopedia of Complexity and Systems Science*, chapter Curvelets and Ridgelets, pages 1718--1738. Springer New York, New York, NY, 2009.

[35] E. J. Candès and D. L. Donoho. Curvelets and Curvilinear Integrals. *J. Approx. Theory*, 113(1):59--90, 2001.

[36] J. Fadili and J-L. Starck. Curvelets and Ridgelets, 2007.

[37] E. J. Candès and D. L. Donoho. Continuous Curvelet Transform: Ii. Discretization and Frames. *Appl. Comput. Harmon. Anal*, 19:198--222, 2003.

[38] S. Dahlke, G. Kutyniok, G. Steidl, and G. Teschke. Shearlet Coorbit Spaces and Associated Banach Frames. *Appl. Comput. Harmon. Anal.*, 27(2):195--214, 2009.

[39] P. Kittipoom, G. Kutyniok, and W-Q. Lim. Construction of Compactly Supported Shearlet Frames. *Constr. Approx.*, 35:21--72, 2012.

[40] G. Kutyniok, D. Labate, and Editors. *Shearlets: Multiscale Analysis for Multivariate Data*. Birkhäuser, 2012.

[41] P. Grohs. Continuous Shearlet Frames and Resolution of the Wavefront Set. *Monatsh. Math.*, 164:393--426, 2010.

[42] G. Kutyniok and D. Labate. The Construction of Regular and Irregular Shearlet Frames. *J. Wavelet Theory Appl.*, 1:1--10, 2007.

[43] K. Guo and D. Labate. Optimally Sparse Multidimensional Representation using Shearlets. *SIAM J. Math. Anal.*, 39:298--318, 2007.

[44] L. Demanet. *Curvelets, Wave Atoms and Wave Equations*. PhD thesis, California Institute of Technology, 2006.

[45] P. Grohs, G. Kutyniok, S. Keiper, and M. Schäfer. α-molecules. *Appl. Comput. Harmon. Anal*, 42:297--336, 2016.

[46] D. L. Donoho, M. Elad, and J-L. Starck. Image Decomposition via the Combination of Sparse Representation and a Variational Approach. *IEEE Trans. Image Process.*, 14(10):1570--1582, 2005.

[47] C. Durasanti, J. D. McEwen, and Y. Wiaux. Localisation of Directional Scale-Discretised Wavelets on the Sphere. *Appl. Comput. Harmon. Anal.*, 2016.

[48] G. Hennenfent and F. J. Herrmann. Seismic Denoising with Nonuniformly Sampled Curvelets. *IEEE Comput. Sci. Eng.*, 8(3):906--916, 2006.

[49] C. Fefferman. A Note on Spherical Summation Multipliers. *Israel J. Math.*, 15:44--52, 1973.

[50] A. Córdoba and C. Fefferman. Wave packets and fourier integral operators. *Comm. PDE*, 3(11):979--1005, 1978.

[51] H. F. Smith. A Hardy Space for Fourier Integral Operators. *J. Geom. Anal.*, 8(4):629--653, 1998.

[52] L. Greengard and C. Stucchio. Spectral Edge Detection in Two Dimensions Using Wavefronts. *Appl. Comput. Harmon. Anal.*, 30:69--95, 2011.

[53] Y. Meyer and R. R. Coifman. *Wavelets, Calderón-Zygmund, and Multilinear Operators*. Cambridge University Press, 1997.

[54] P. Lax. Asymptotic Solutions of Oscillatory Initial Value Problems. *Duke Math J.*, 24:627--646, 1957.

[55] S. Dahlke, G. Kutyniok, G. Steidl, and G. Teschke. The Uncertainty Principle Associated with the Continuous Shearlet Transform. *Int. J. Wavelets Multiresolution. Inf. Process*, 6:157--181, 2008.

[56] P. Grohs. Intrinsic Localisation of Anisotropic Frames. *Appl. Comput. Harmon. Anal*, 35:264--283, 2013.

[57] R. Balan, P. G. Casazza, C. Heil, and Z. Landau. Density, Overcompleteness, and Localization of Frames. I. Theory. *J. Fourier Anal. Appl.*, 12(2):105--143, 2006.

[58] F. Futamura. Localizable Operators and the Construction of Localized Frames. *Proc. Amer. Math. Soc.*, 137(12):4187--4197, 2009.

[59] P. Grohs. Continuous Shearlet Tight Frames. *J. Fourier Anal. and Appl.*, pages 506--518, 2011.

[60] P. Grohs. Bandlimited Shearlet-Type Frames with Nice Duals. *J. Comput. Appl. Math.*, pages 139--151, 2013.

[61] R. Stevenson. Adaptive Solution of Operator Equations Using Wavelet Frames. *SIAM J. Numer. Anal.*, 41(3):1074--1100, 2003.

[62] M. E. Taylor. *Pseudodifferential Operators*. Princeton University Press, 1981.

[63] R. B. Melrose and G. Uhlmann. An Introduction to Microlocal Analysis. `http://www-math.mit.edu/~rbm/books/imaast.pdf`, 2000.

[64] A. Grigis and J. Sjöstrand. *Microlocal Analysis for Differential Operators: An Introduction*. Cambridge University Press, 1994.

[65] C. Brouder, N. V. Dang, and F. Helein. A Smooth Introduction to the Wavefront Set. `http://arxiv.org/abs/1404.1778`, 2014.

[66] G. Kutyniok and D. Labate. Resolution of the Wavefront Set Using Continuous Shearlets. *Trans. Amer. Math. Soc.*, 361(5):2719--2754, 2009.

[67] Y. Meyer. *Ondelettes et Operateurs: I-III*. Hermann, 1990.

[68] T. Tao. Lecture Notes for Fourier Analysis at UCLA. `http://www.math.ucla.edu/~tao/247a.1.06f/notes2.pdf`.

[69] K. Guo and D. Labate. The Construction of Smooth Parseval Frames of Shearlets. *Math. Model. Nat. Phenom.*, 8(1):82--105, 2013.

[70] A. Gelb and D. Cates. Segmentation of Images from Fourier Spectral Data. *Commun. Comput. Phys.*, 5(2-4):326--349, 2009.

[71] T. A. Gallagher, A. J. Nemeth, and L. Hacein-Bey. An Introduction to the Fourier Transform: Relationship to MRI. *Amer. J. of Roentgen.*, 180(5):1396--1405, 2008.

[72] K. Guo and D. Labate. Characterization and Analysis of Edges Using the Continuous Shearlet Transform. *SIAM J. Imaging Sci.*, 2(3):959--986, 2009.

[73] G. Kutyniok and P. Petersen. Classification of Edges Using Compactly Supported Shearlets. *Appl. Comput. Harmon. Anal.*, 2015.

[74] K. Guo and D. Labate. Characterization and Analysis of Edges in Piecewise Smooth Functions. *Appl. Comput. Harmon. Anal.*, 2015.

[75] C. S. Herz. Fourier Transforms Related to Convex Sets. *Ann. of Math.*, 2(75):81--92, 1962.

[76] E. Sorets. Fast Fourier Transforms of Piecewise Constant Functions. *J. Comp. Phys.*, 116(2):395--379, 1995.

[77] M. Fornasier and K. Gröchenig. Intrinsic Localisation of Frames. *Constr. Approx.*, 22(3):395--415, 2005.

[78] J. D. McEwen, P. Vielva, Y. Wiaux, R. B. Barreiro, L. Cayón, M. P. Hobson, A. N. Lasenby, E. Martínez-González, and J. L. Sanz. Cosmological Applications of a Wavelet Analysis on the Sphere. *J. Fourier Anal. Appl.*, 13(4):495--510, 2007.

[79] W-Q. Lim. The Discrete Shearlet Transform: A New Directional Transform and Compactly Supported Shearlet Frames. *IEEE Trans. Image Process.*, 19(5):1166--1180, 2010.

[80] C. D. Sogge. *Fourier Integrals in Classical Analysis.* Cambridge University Press, 1993.

[81] G. B. Folland. *Introduction to Partial Differential Equations.* Princeton University Press, second edition, 1995.

[82] G. B. Folland. *Real Analysis.* John Wiley, second edition, 1999.

[83] R. B. Melrose and P. Piazza. Analytic k-Theory on Manifolds with Corners. *Adv. Math.*, 92:1--26, 1992.

[84] S. T. Ali, J-P. Antoine, and J-P. Gazeau. *Coherent States, Wavelets, and Their Generalizations.* Springer, second edition, 2014.

[85] P. G. Casazza, G. Kutyniok, and Editors. *Finite Frames.* Birkhäuser, 2013.

[86] J. J. Benedetto. *Harmonic Analysis and Applications.* CRC Press, 1997.