Lappeenranta
University of Technology

Jouni Sampo

ON CONVERGENCE OF TRANSFORMS BASED ON PARABOLIC SCALING

Thesis for the degree of Doctor of Technology to be presented with due permission for public examination and criticism in the Auditorium 1383 at Lappeenranta University of Technology, Lappeenranta, Finland, on the 21st of December 2010, at noon.

Acta Universitatis
Lappeenrantaensis 417

Supervisor	Professor, Ph.D. Heikki Haario Lappeenranta University of Technology Department of Mathematics and Physics Lappeenranta, Finland
Reviewer	Professor, D.Sc. Ole Christensen Technical University of Denmark Department of Mathematics Lyngby, Denmark
Reviewer	Docent, Ph.D. Mikko Salo University of Helsinki Department of Mathematics and Statistics, Helsinki, Finland
Opponent	Professor, D.Sc. Ole Christensen Technical University of Denmark Department of Mathematics Lyngby, Denmark

ISBN 978-952-265-025-2
ISBN 978-952-265-026-9 (PDF)
ISSN 1456-4491

Lappeenrannan teknillinen yliopisto
Digipaino 2010

Acknowledgements

There are many people that have influenced positively to this thesis, either directly or indirectly along the long process. I would like to thank you all. I would like to thank my supervisor Heikki Haario and the reviewers Mikko Salo and Ole Chirstensen for their valuable comments , even on last meters, and my family for being so patient while I have been so occupied with this project.

I would like to thank also all my friends, colleagues and collaborators for providing an excellent atmosphere, where thoughts can grow and fly free.

But the main thanks for the existence of this thesis belong to my dear collaborator Songkiat Sumetkijakan. Without you I probably would not have started with this line of research at all.

And last but not least, I would like to express my gratitude to the following instances for financial support: Finnish Craduate School in Inverse Problems, The Foundation of Lappeenranta University of Technology, Lappeenranta University of Technology, The Foundation of Jenny and Antti Wihuri, Finnish Foundation for Technology Promotion, The Foundation of Ulla Tuominen and The Foundation of Emil Aaltonen.

Lappeenranta, Finland — Jouni Sampo
December 7, 2010

Abstract

Jouni Sampo

ON CONVERGENCE OF TRANSFORMS BASED ON PARABOLIC SCALING

Lappeenranta, 2010
72 p.

Acta Universitatis Lappeenrantaensis 322
Diss. Lappeenranta University of Technology

ISBN 978-952-265-025-2, 978-952-265-026-9 (PDF)
ISSN 1456-4491

This thesis studies properties of transforms based on parabolic scaling, like Curvelet-, Contourlet-, Shearlet- and Hart-Smith-transform. Essentially, two different questions are considered: How these transforms can characterize Hölder regularity and how non-linear approximation of a piecewise smooth function converges.

In study of Hölder regularities, several theorems that relate regularity of a function $f : \mathbb{R}^2 \to \mathbb{R}$ to decay properties of its transform are presented. Of particular interest is the case where a function has lower regularity along some line segment than elsewhere. Theorems that give estimates for direction and location of this line, and regularity of the function are presented. Numerical demonstrations suggest also that similar theorems would hold for more general shape of segment of low regularity. Theorems related to uniform and pointwise Hölder regularity are presented as well. Although none of the theorems presented give full characterization of regularity, the sufficient and necessary conditions are very similar.

Another theme of the thesis is the study of convergence of non-linear $M-$term approximation of functions that have discontinuous on some curves and otherwise are smooth. With particular smoothness assumptions, it is well known that squared L^2 approximation error is $\mathcal{O}(M^{-2}(\log M)^3)$ for curvelet, shearlet or contourlet bases. Here it is shown that assuming higher smoothness properties, the log-factor can be removed, even if the function still is discontinuous.

Keywords: Curvelets, Shearlets, Hart Smith transform, Non-linear approximation, Hölder regularity.

UDC 517.44 : 517.9 : 519.65

Contents

1 Introduction

Some of the most concrete visual element in images are edges. Edges contain the information that machine vision systems build by man, as well as human brains, need to understand the geometry of the environment that they are working with. Therefore, efficient and robust methods to characterize and detect edges from images are needed. Moreover, because most images have edges, it would be beneficial to be able to store efficiently images that contain edges. Edge-like features are apparent also in some physical non-visual situations, where sudden changes happen on some boundary, for example conductivity changes inside an object. When these kind of features are studied numerically, it is often beneficial to use methods that can present them and find such boundaries efficiently.

In the beginning of this millennium, a new and mathematically very powerful tool for analysis of images with edges, the curvelet transform [4, 3, 39, 40, 2, 5], was introduced. Motivated by the success of curvelet transform, soon other similar transforms, like shearlets and contourlets, were also presented [11, 34, 15, 13, 17, 16]. In this thesis some mathematical details about the characterization of direction of edges and regularity properties with this kind of transforms are developed. Some interesting theoretical result for approximation of images by using these transforms are given.

The Hölder regularity of a function is related to the smoothness of the function and the derivatives of it, and therefore also to the efficiency of polynomial approximation. Hölder regularity properties can also be used as features in signal classification or pattern recognition [22, 41], as well as for signal de-noising or estimation [27]. The Hölder regularity of a function f can be explicitly analyzed for a relatively small class of functions. Sometimes there exists no formula, but we only have some discrete samples from f. In these cases, it still is possible to estimate the regularity by investigating decay properties of the wavelet transform [19, 29]. Since wavelets are isotropic, some other transform like curvelet or shearlet transform might be a more natural choice in the case where some directional information about function is needed. However, the author is not aware of any other research in characterization of Hölder regularities by using parabolic transforms than that presented in [24, 36, 33].

The thesis is organized as follows. The introduction is devoted to give readers unfamiliar with transforms, bases, and approximations some idea why they are used and how they work. Although some mathematical definitions are given in this section, statements given here are meant to be more descriptive than mathematically precise. Section 2 gives more mathematical details of transforms that are particularly discussed in this thesis. Section 3. is all about proving some new results for convergence of curvelet transform and non-linear approximation properties of curvelet basis. This is one of the main contributions of this thesis. The last section deals with a study of

characterization of functions and edges by curvelet and shearlet transform.

In this thesis, all theorems and lemmas which have proofs included are original contributions made by the author. Results in Section 3 have not been published before. Results in other sections are published in [36, 24]. In [36] the author of this thesis was responsible for technical development of theorems and lemmas. In [24] the original research idea was from the author, however collaborators, Lakhonchai and Sumetkijakan, had the main responsibility in technical translation of theorems from [36] to the shearlet case. All numerical demonstrations in [36, 24] were done by the author of this thesis.

1.1 Geometry of images and basis functions

Let us think what kind of images there are stored in computers, cameras or mobile phones. First, there is an object that we want to image. Often, this is done by taking a photo from one direction with camera. If camera would have an infinite resolution, a reasonable model for this image would be a function of two variables, $f : \mathbb{R}^2 \to \mathbb{R}$, where $f(x)$ would be associated to the color value in location x.

One of the main themes of this thesis is a theoretical investigation on how well certain functions $f : \mathbb{R}^2 \to \mathbb{R}$ can be approximated with a limited number M of coefficients $c_i \in \mathbb{R}$, if each of indexes i is associated to some fixed building block f_i. As a non-mathematical example from such approximation, one could think that these building blocks are like Lego$^{\text{TM}}$ blocks. Index i would define in what location on floor the Lego block f_i is, and what are its dimensions. Real numbers c_i could define how many blocks f_i we would put on top of each others. If f_i and f_j would have same location but different sizes, then smaller blocks would be placed on top of bigger ones. If there would be no limit on how small these blocks can be, we could approximate arbitrary well any smooth function $f : \mathbb{R}^2 \to \mathbb{R}^+$ by building it from basic Lego$^{\text{TM}}$ blocks! The geometry of building blocks f_i is of course essential for how easily some desired shapes can be constructed. For example, discontinuities of f along some line can be seen as walls in Lego land. With Lego blocks the vertical and horizontal walls are easier to build than diagonal walls. This would suggest that building blocks with more orientations would be nice. However, the more special building blocks are used, the more indexes i must be used. This makes the system more complicated, when trying to choose the best building blocks f_i and the associated coefficients c_i.

In this thesis we restrict to functions with finite energy. This restriction is common and, from a practical point of view, not very restrictive: most functions related to real signals, at least to images, tend to have finite energy. This class of functions is

commonly denoted as

$$L^2(\mathbb{R}^2) := \left\{ f : \mathbb{R}^2 \to \mathbb{R} \mid \|f\|_2^2 := \int_{\mathbb{R}^2} |f(x)|^2 \, dx < \infty \right\}, \tag{1}$$

where the integral is Lebesgue integral.

In mathematical language, the building blocks f_i are called as *basis functions* and the set $\{f_i\}_i$ is called *basis* if any function $f \in L^2(\mathbb{R}^2)$ can approximated arbitrary well with some linear combination of f_i, i.e. for all $\epsilon > 0$ exists coefficients c_i such that $\|f - \sum_i c_i f_i\|_2 < \epsilon$. In many applications, like estimation or compression, it is important that too many building blocks f_i are not needed when ϵ is fixed to some reasonable small number.

The function space $L^2(\mathbb{R}^2)$ is very large, containing many different kind of functions. Some functions might be very smooth, some very rough since the definition of $L^2(\mathbb{R}^2)$ does not consider derivatives. Therefore, it is quite clear that same basis can not be very efficient for approximation of all functions in $L^2(\mathbb{R}^2)$. Also, the domain of functions, $\mathbb{R}^2$, is unnecessary large for most of applications: often values $f(x)$ are interesting only when x belongs to the unit square $[0, 1]^2$, or to some other compact set $\Omega \subset \mathbb{R}^2$. In this case, the popular and simple method is to set $f(x) = 0$ if $x \notin \Omega$.

1.1.1 Wavelets in Lego$^{\text{TM}}$ land and some extensions

The wavelet bases are defined with little more mathematical details later, here the simple geometric explanation is given. Roughly speaking, wavelet basis can be seen as set of Lego$^{\text{TM}}$ blocks where all blocks are cubes with side the length l that scales in powers of two, i.e. $l = C2^j$, where $j \in \mathbb{Z}$. This immediately suggests that wavelet bases can be used to approximate functions that have horizontal or vertical edges, but wavelets would not work in best way with diagonal edges. Wavelets would be also very efficient in approximation of isotropic features like some sharp peaks. The main geometric difference between Legos and wavelets is that wavelets have been designed so that also smooth parts of functions can be presented efficiently by them. Mathematically, this is simply done by demanding them to be smooth enough and being orthogonal to polynomials of degree high enough, as well as each others. Even if this sounds simple, keeping them compactly supported, i.e. non-zero only in short interval, and providing nice digital implementation is not so simple but was a real break through in the late 80's [9, 32, 28].

If we could look for new basis every time when f changes, a good approximation for f by using few f_i would be easier to establish. The characteristics of a good basis depend essentially on properties of f. Exploring those properties might be difficult or at least a time consuming task. And the basis functions f_i should have some clear structure so that all information about their shapes would be easy to store. These

bases are called *"adaptive bases"*. For example if an edge is not horizontal, we can twist the image so that after twisting the edge is horizontal, and then do wavelet approximation for this twisted image. Or, we can do the other way round: we twist the wavelet basis functions so that they see the edge as a straight line. In both cases we have to first recognize the edge and save information about it somewhere. Saving the information of geometry is usually not the problem, but finding is. Bandelet basis implements this idea of twisting wavelets in a very efficient way [26, 25, 31, 30].

Normal wavelets have essentially only two different orientation (third, diagonal orientation, is very weak). One way to generalize wavelets is to increase the number of orientations. There exists several such approaches like complex wavelets and steerable pyramids [35]. A common feature for these constructions is that the number of orientations is fixed and the same at every scale. These variants of wavelet bases may work quite nicely in practice sometimes, but theoretically they are far from optimal in the sense of efficiency of approximation.

1.1.2 Scaling goes parabolic and number of orientations grows

If we want to stick to non-adaptive basis and avoid all problems of exploring geometry of image, what could be done? As said earlier, more orientations would be nice. Before adding those let us consider the geometry of smooth curves. Usually by "a smooth curve" S in $\mathbb{R}^2$ we mean a curve that has a bounded second derivative. In that case, there exists a constant $C < \infty$ such that a piece with length $\sqrt{a}$ from this curve will fit to a rectangle R that has length about $\sqrt{a}$ and width about a. This gives an idea that for efficient analysis or approximation of a function $f \in L^2(\mathbb{R}^2)$ that has singularities along this kind of a smooth curve, we should use basis functions f_i in $L^2(\mathbb{R}^2)$ such that most of its energy is concentrated to this rectangle.

Since orientation for each piece of S can vary, we should use also functions f_i with multiple orientations. Also, we should use multiple scales a since we get otherwise just rough approximation about S. Basis functions related to smaller scales a could adapt better to S and therefore the use of them would refine approximation.

The number of orientations also should increase when the scale a gets smaller. This is because the ratio between length and width of rectangles is about $a^{-1/2}$. About $a^{-1/2}$ rectangles are needed to cover a ball with radius $\sqrt{a}$, i.e. to cover all possibly interested orientations. So at least about $a^{-1/2}$ differently oriented functions f_i should be available, otherwise they cannot adapt to geometry of S. Emmanuel Candes and David L. Donoho introduced first this kind of basis functions, called *curvelets* for approximation purposes. After that, inspired by curvelets, other bases with similar properties were developed by other research groups. Best known are *contourlets* by Do and Vetterli [11, 34] and *shearlets* by Kanghui and Labate [15]. Each of these

transforms are discussed with more details in later sections.

1.2 Approximations of a function

When a photo is taken, we do not necessary get an optimal image (i.e. function f) but only some approximation because of limited resolution of camera, unperfect lenses, etc. When an image is saved to computer, camera or mobile, some more approximations are usually done, even if original image is taken by digital camera. This is because an image often is too big to save or transfer, and therefore must be compressed. So, even with an imagined infinite resolution camera, the final image would be just approximation.

Approximation or compression of a function $f \in L^2(\mathbb{R}^2)$ is often based on algorithms where f is first represented as a linear combination of basis functions $f_i \in L^2(\mathbb{R}^2)$, i.e.

$$f = \sum_i c_i f_i,$$

where $c_i \in \mathbb{R}$, and then coefficients c_i with small absolute values are rounded to zero. In this kind of process the orthonormality of basis functions f_i is often desired since it gives an easy way to minimize error due to this rounding: If I is the index set that contains those coefficients that are not rounded zero and $\tilde{f} := \sum_{i \in I} c_i f_i$, then $\left\| f - \tilde{f} \right\|_2^2 = \sum_{i \notin I} |c_i|^2$. In this case, if we want to use only M coefficients, the optimal approximation $\tilde{f}$ for f is the one where all but M biggest coefficients c_i are rounded to zero. Coefficients c_i are also simple to calculate in the case of orthonormal basis: $c_i = \langle f, f_i \rangle := \int_{\mathbb{R}^2} f(x) f_i(x) dx$.

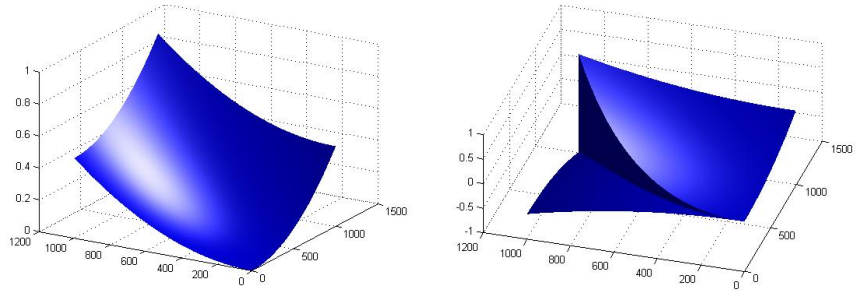
However, a slightly more general property called "*tight frame*" provides the same property [18]:

Definition 1 *The set of functions $B = \{f_i\}_i$ in $L^2(\mathbb{R}^2)$ is called a normalized tight frame for $L^2(\mathbb{R}^2)$ if for any $f \in L^2(\mathbb{R}^2)$ there holds*

$$\|f\|_2^2 = \sum_i |\langle f, f_i \rangle|^2 . \tag{2}$$

Corollary 2 *If the set $\{f_i\}_i$ in $L^2(\mathbb{R}^2)$ is a normalized tight frame, then any $f \in L^2(\mathbb{R}^2)$ has the representation $f = \sum_i \langle f, f_i \rangle f_i$*

An exact definition of best M-term non-linear approximation with tight frames can be given as follows



(a) (b)

Figure 1: (a) Graph of function f that is smooth; (b) Graph of function g that is smooth apart from a smooth curve.

Definition 3 *Let $f = \sum_i c_i f_i \in L^2(\mathbb{R}^2)$, $B = \{f_i\}_i$ be a normalized tight frame and $I_{M,f}$ be the index set that contains indexes i for M coefficients c_i with biggest absolute value $|c_i|$. Best M-term non-linear approximation of f in basis B is given then by*

$$f_{M,B} := \sum_{i \in I_{M,f}} c_i f_i, \tag{3}$$

where $c_i = \langle f, f_i \rangle$.

When discussing convergence rates, the notation of type $h(M) \leq \mathcal{O}(p(M))$ will be used frequently later. This means that for all $M < \infty$ exists coefficients C, independent of M such that $h(M) \leq Cp(M)$.

The rest of this section is devoted to clarify how different the behavior of bases might be for different kind of functions. While emphasizing differences between bases, all mathematical details are not mentioned here, some appropriate references are given in [11, 29, 5].

Figure 1 illustrates graphs of two different kind of functions.

We now describe how the best M-term approximation of these would behave in three different bases. Let

- F be Fourier-basis (i.e. functions f_i are modulated versions of sines and cosines).
- W be wavelet basis (i.e. functions f_i are translated and scaled versions of single "generating" function ψ). This kind of basis is used, for example, in the JPEG2000 image compression standard.
- C be curvelet, contourlet or shearlet basis (detailed definitions in later sections).

The function f in Figure 1 a) is at least two times differentiable with bounded derivatives. For this type of function we have

$$\|f - f_{M,F}\|_2^2 \leq \mathcal{O}(M^{-2})$$
$$\|f - f_{M,W}\|_2^2 \leq \mathcal{O}(M^{-2})$$
$$\|f - f_{M,C}\|_2^2 \leq \mathcal{O}(M^{-2})$$

Notice that in this case all transforms had the same approximation rate. Things change dramatically when the function is taken from a different class. Suppose that the function g has a discontinuity along a smooth curve and otherwise it has exactly the same differentiability properties as f had. Function g is illustrated in Figure 1 b). For functions like g, the approximation now reads as:

$$\|g - g_{M,F}\|_2^2 \leq \mathcal{O}(M^{-1/2})$$
$$\|g - g_{M,W}\|_2^2 \leq \mathcal{O}(M^{-1})$$
$$\|g - g_{M,C}\|_2^2 \leq \mathcal{O}(M^{-2}\log_2(M)^3)$$

The approximation rate with wavelet basis collapsed now far worse than what it was for f. The approximation rate with Fourier basis was even worse than with wavelet basis. Comparing Fourier and wavelet basis, the curvelet basis hardly noticed the discontinuity of g.

In the above examples, the approximation rates for f were actually optimal, meaning that there exist no other bases that could do better. For g, the curvelets achieved also almost optimal approximation rate: for functions that are smooth apart from (smooth) discontinuity curves the optimal rate would be $\mathcal{O}(M^{-2})$. Bandelets provide this optimal rate, but bandelet basis is an adaptive basis [26]. So, comparing optimal rate, the curvelet basis had only the extra factor $\log_2(M)^3$. However, in practical applications M rarely is less than 2^{10}, which means that despite of logarithm, this factor is still quite large. One of the main theme of this thesis is to investigate if this factor can be removed from curvelet approximations, also when the function has a discontinuity along a smooth curve. Theorem 17 states that under some extra assumptions for smoothness this can be done.

1.3 From approximations to estimates

Despite of the fact that transforms based on parabolic scaling have some excellent theoretical properties for approximation and compression, they have not been able to make breakthrough on these fields yet. However, in some noise removal application, these transforms are really superb compared to the wavelet transform, that has been a popular method during last decades. Although this thesis does not focus to noise removal or signal estimation, non-linear approximation can be powerful tool in solving

these problems. More precisely: approximations are good estimates. Therefore the approximation theory that is considered in this thesis has a direct relation to signal estimation and noise removal. In this section we shortly explain why.

Suppose that f is corrupted by noise from a measurement devise or some other source. If the magnitude of noise e is independent of values of f, the model between f and observed function m is given by

$$m = f + e.$$

The problem is to find an estimate for f when m is given (measured). Now m can be written as

$$m = \sum_i \langle m, f_i \rangle f_i = \sum_i (\langle e, f_i \rangle + \langle f, f_i \rangle) f_i.$$

From this form it can be seen that if f_i are designed so that $|\langle e, f_i \rangle|$ tend to be small as compared to the M largest values of $|\langle f, f_i \rangle|$, then largest values of $\langle m, f_i \rangle$ are close to values of $\langle f, f_i \rangle$. This is the usual case when the noise level $\|e\|_2^2 / \|f\|_2^2$ is not too high and the basis is such that best M-term approximation is efficient. To be a bit more precise, we can write

$$\begin{aligned} m_{M,B} &= \sum_{i \in I_{M,m}} \langle m, f_i \rangle f_i \\ &= \sum_{i \in I_{M,m}} \langle e, f_i \rangle f_i + \sum_{i \in I_{M,m}} \langle f, f_i \rangle f_i \\ &= \sum_{i \in I_{M,m}} \langle e, f_i \rangle f_i + f_{M,B} + h, \end{aligned}$$

where

$$h = \sum_{i \in I_{M,m} \setminus I_{M,f}} \langle f, f_i \rangle f_i - \sum_{i \in I_{M,f} \setminus I_{M,m}} \langle f, f_i \rangle f_i = \sum_{i \in J} \langle f, f_i \rangle f_i$$

If M is not too large compared to the noise level, the index set J is probably quite small, even empty. Also in this case if $k \in J$, then probably $|\langle f, f_k \rangle| \approx \min \{ |\langle f, f_i \rangle| \, : \, i \in I_{M,f} \}$. If M is too large, then all $|\langle f, f_k \rangle|$, $k \in J$, are relatively small (the level and pattern of the noise e essentially limit how small), but there may be a lot of them: the size of the set might start to grow in the worst case almost linearly with respect to M. For the estimation error we can write

$$\begin{aligned} \|f - m_{M,B}\| &\leq \|f - f_{M,B}\| + \|f_{M,B} - m_{M,B}\| \\ &\leq \|f - f_{M,B}\| + \left\| \sum_{i \in I_{M,m}} \langle e, f_i \rangle f_i \right\| + \|h\| \\ &= \|f - f_{M,B}\| + \sqrt{\sum_{i \in I_{M,m}} |\langle e, f_i \rangle|^2} + \|h\| \\ &= E_1 + E_2 + E_3 \end{aligned}$$

The term E_2^2 increases approximately linearly with respect to M. Therefore, small M is desired. Because E_1 is just a M-term approximation error of f, it is important that this error decreases as fast as possible, since we no not want to have a large M. Moreover, as explained above, E_3 tends to grow also when M grows too large. Because all of this, fast convergence of $\|f - f_{M,B}\|$ is important. It is worth to note that the faster is the decay rate of E_1, the more early E_3 start to grow with respect to M. This somehow tells that the noise can not be removed totally, no matter how efficient M-term approximation is. This is natural, since random noise causes always some uncertainty.

For functions that are smooth apart from a smooth discontinuity curve, curvelets are actually almost optimal basis for this kind of estimation [4, 39].

In addition to this simple thresholding method, the remaining coefficients also can be shrunk and this might visually improve the estimate [12]. However, even for this shrinking method, it is essential that a basis with good non-linear approximation properties is used.

2 Multiscale transforms based of parabolic scaling

2.1 Mathematical notations

Let us start with some notations that will be used frequently later. The dilatation (or scaling) matrix $\mathrm{D}_{\frac{1}{a}}$ that provides a parabolic scaling is given as

$$\mathrm{D}_{\frac{1}{a}} := \begin{pmatrix} \frac{1}{a} & 0 \\ 0 & \frac{1}{\sqrt{a}} \end{pmatrix}. \tag{4}$$

Denote by $\mathrm{R}_{-\theta}$ the matrix affecting planar rotation of θ radians in clockwise direction, i.e.

$$\mathrm{R}_{-\theta}^{-1} := \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \tag{5}$$

Inverses of these matrices are therefore $\mathrm{D}_{\frac{1}{a}}^{-1} = \mathrm{D}_a$ and $\mathrm{R}_{-\theta}^{-1} = \mathrm{R}_\theta$.

For general (non-index) vectors $x = (x_1, x_2) \in \mathbb{R}^2$, we denote $x^\nu := x_1^{\nu_1} x_2^{\nu_2}$ and norm is the usual Euclidean norm $\|x\| := \sqrt{x_1^2 + x_2^2}$. For each scale a and direction θ, let us define the norm

$$\|v\|_{a,\theta} := \left\| \mathrm{D}_{\frac{1}{a}} \mathrm{R}_{-\theta} v \right\| \quad \text{for } v \in \mathbb{R}^2. \tag{6}$$

An immediate consequence of the above definitions is that $\frac{\|v\|}{\sqrt{a}} \leq \|v\|_{a,\theta} \leq \frac{\|v\|}{a}$ when $0 < a \leq 1$.

The vector v_θ is defined as the unit vector parallel to the major axis of the ellipsoidal $\|v\|_{a,\theta} = 1$, i.e.

$$v_\theta := \mathrm{R}_\theta(0,1)^T. \tag{7}$$

The length of the major axis of this ellipsoidal is $\sqrt{a}$ and the length of the minor axis is a, i.e. ellipsoidal becomes more and more needle like when scale a goes to zero.

Partial derivatives of a function $f : \mathbb{R}^2 \to \mathbb{R}$ are denoted by $\partial^\nu f = \partial_1^{\nu_1}\partial_2^{\nu_2} f$ where ∂_i means the partial derivative with respect to the i^{th}-variable and the index vector $\nu = (\nu_1, \nu_2)$ is in $\mathbb{N}_0^2$, with $|\nu| = \nu_1 + \nu_2$.

The Fourier transform is defined by

$$\hat{f}(\xi) := \int_{\mathbb{R}^2} f(x) e^{-2\pi i x\cdot\xi}\, dx, \quad \text{for } \xi \in \mathbb{R}^2, \tag{8}$$

so that the Plancherel's formula becomes

$$\langle f, g\rangle = \langle \hat{f}, \hat{g}\rangle, \tag{9}$$

and the Fourier inversion formula reads

$$f(x) = \int_{\mathbb{R}^2} \hat{f}(\xi) e^{2\pi i \xi\cdot x}\, d\xi. \tag{10}$$

Although for general $f \in L^2(\mathbb{R}^2)$ the Fourier transform should be defined in a bit more complicated way (see the definition for example from [29]). However, Fourier transforms are applied in proofs only to functions for which the above definitions are equivalent.

We will often use the notation

$$\theta = \frac{\pi}{2} k a^{1/2},$$

where $0 \leq |k| \leq, a^{-1/2}$. The advantage of this notation will make sense in later sections, when we investigate convergence of non-linear approximations.

Characteric function $\chi_A : \mathbb{R}^2 \to \{0, 1\}$ of a set A is defined as

$$\chi_A(x) := \begin{cases} 1 & , \quad x \in A \\ 0, & \quad x \notin A \end{cases} \tag{11}$$

2.2 Wavelet transform

To emphasize new ideas of curvelets, shearlets and contourlets compared to wavelets, some basics about wavelet bases and transforms are first revised.

2.2.1 Continuous wavelet transform

Like the Fourier-transform defined in the previous section, Continuous Wavelet Transform (CWT) is an integral transform defined by the formula

$$\mathcal{W}_f(a,b) := \int_{\mathbb{R}^2} f(x)\psi_{ab}(x)dx,$$

where the

$$\psi_{ab}(x) := a^{-1}\psi(\mathrm{diag}(a^{-1},a^{-1})(x-b)),$$

$a > 0$, $b \in \mathbb{R}^2$ and $\psi \in L^2(\mathbb{R}^2)$. *Mother wavelet* ψ should also satisfy the so called admissability conditions that ensure the existence of inverse transform. Some details can be found in [10]. The most interesting property that follows in practical situations from admissability conditions is that

$$\int_{\mathbb{R}^2} \psi(x)dx = 0. \tag{12}$$

This means that a wavelet should have a zero-order vanishing moment. In practice, also higher order vanishing moments are demanded. These vanishing moments are the key for approximation properties. In comparison, curvelets, contourlets and shearlets have directional vanishing moments, defined in section 2.7.1, that wavelets do not necessary have. Especially, the scaling law of wavelets means that even if ψ has directional vanishing moments to some directions, the number of directions do not grow when the scale a decreases, as it does with parabolic scaling.

A direction parameter can be easily added to wavelet transform, just simply add the rotation

$$\mathcal{W}_f(a,b,\theta) := \int_{\mathbb{R}^2} f(x)\psi_{ab\theta}(x)dx,$$
$$\psi_{ab\theta}(x) = a^{-1}\psi(\mathrm{diag}(a^{-1},a^{-1})\mathrm{R}_\theta(\mathrm{x}-\mathrm{b})).$$

However, this rotation does not add any more directional vanishing moments.

2.2.2 Wavelet basis

When constructing wavelet bases for $L^2(\mathbb{R}^2)$, the usual approach that leads to efficient digital implementation is to construct three different mother wavelets $\psi^{(i)} : \mathbb{R}^2 \to \mathbb{R}$, $i = 1,2,3$ by tensor products applied to a mother wavelet and a so called scaling function (or father wavelet) that are related to multiresolution analysis. Basis functions for $L^2(\mathbb{R}^2)$ are found then by properly sampling parameters a and b for functions $\psi^{(i)}_{ab}(x)$. However, usually digital implementation is actually made with finite impulse response (FIR) filters. This approach gives wavelets with three different orientations. Details can be found for example in [29]. Using more than one generating function is also the idea used with shearlets and contourlets, which also leads to more straightforward digital implementation than what curvelets have.

2.3 Hart Smith transform

The first transform that is considered is the Hart Smith transform. According to our knowledge, only continuous Hart Smith transform is introduced in literature and a related basis for $L^2(\mathbb{R}^2)$ has been never presented.

Originally this transform was defined in Fourier domain and L^1-normalization was used[38]. In this thesis we use the variation that was described in [6, 7] which adopts L^2-normalization.

For a given $\varphi \in L^2(\mathbb{R}^2)$, a family of functions $\varphi_{ab\theta}$ is defined by

$$\varphi_{ab\theta}(x) = a^{-\frac{3}{4}} \varphi \left(\mathrm{D}_{\frac{1}{a}} \mathrm{R}_{-\theta} \left(x - b \right) \right),$$

for $\theta \in [0, 2\pi)$, $b \in \mathbb{R}^2$, and $0 < a < a_0$, where a_0 is a fixed coarsest scale. The Hart Smith's directional wavelet transform is defined as an integral transform

$$\overline{\Gamma}_f \left(a, b, \theta \right) := \left\langle \varphi_{ab\theta}, f \right\rangle .$$

This is a true affine transform based on parabolic scaling, i.e., the family $\{\varphi_{ab\theta}\}$ is generated by translating, rotating, and parabolically dilating a single generating function φ.

Because of this, we can think that the energy of $\varphi_{ab\theta}$ is essentially concentrated on the ellipse centered at b with minor axis of length "about" a and major axis of length "about" $\sqrt{a}$ pointing in the direction v_θ. This means that the kernel function $\varphi_{ab\theta}$ becomes "needle shaped" when the scale goes small. The only visible difference to directional wavelet transform is the use of parabolic scaling instead of isotropic scaling.

High-frequency function is a square integrable function whose Fourier transform is supported outside a ball of some fixed radius. For high-frequency functions the following reconstruction formula is proposed in [6]. See also [38, 7].

Theorem 4 *There exists a Fourier multiplier M of order 0 so that whenever $f \in L^2(\mathbb{R}^2)$ is a high-frequency function supported in frequency space $\|\xi\| > \frac{2}{a_0}$, then*

$$f = \int_0^{a_0} \int_0^{2\pi} \int_{\mathbb{R}^2} \left\langle \varphi_{ab\theta}, Mf \right\rangle \varphi_{ab\theta} \, db \, d\theta \, \frac{da}{a^3}. \tag{13}$$

The function Mf is defined in the frequency domain by a multiplier formula $\widehat{Mf}(\xi) = m(\|\xi\|)\hat{f}(\xi)$, where m is a standard Fourier multiplier of order 0 (that is, for each $k \geq 0$, there is a constant C_k such that for all $t \in \mathbb{R}$, $|m^{(k)}(t)| \leq C_k \left(1 + |t|^2\right)^{-k/2}$).

The restriction to high frequency functions makes the reconstruction formula a bit useless in some situations. In the study of characterization of regularity, the problems

do not arise: a general $f \in L^2$ can be first divided to low and high frequency parts f_L and f_H, i.e. produce f_L first by lowpass filtering and then set $f_H = f - f_L$. Because f_L is always infinitely differentiable, regularity properties of $f = f_L + f_H$ depend only on f_H. Therefore tools for working with just high frequency functions are enough for this problem. However, this kind of preprocessing with lowpass filter is not necessary (see for example the proof of Theorem 24).

Another inconvenience in using Hart Smith transform is that we have to use a Fourier multiplier M. Generally, by Plancherel's formula we get

$$\langle \varphi_{ab\theta}, Mf \rangle = \left\langle \widehat{\varphi_{ab\theta}}, m(\|\cdot\|)\hat{f} \right\rangle = \left\langle m(\|\cdot\|)\widehat{\varphi_{ab\theta}}, \hat{f} \right\rangle = \langle M\varphi_{ab\theta}, f \rangle .$$

Because of this and the fact that $\varphi_{ab\theta}$ and $M\varphi_{ab\theta}$ are duals [6, 7], the reconstruction formula can be written also in the form

$$\begin{aligned} f &= \int_0^{a_0} \int_0^{2\pi} \int_{\mathbb{R}^2} \langle M\varphi_{ab\theta}, f \rangle \, \varphi_{ab\theta} \, db \, d\theta \, \frac{da}{a^3} \\ &= \int_0^{a_0} \int_0^{2\pi} \int_{\mathbb{R}^2} \langle \varphi_{ab\theta}, f \rangle \, M\varphi_{ab\theta} \, db \, d\theta \, \frac{da}{a^3}. \end{aligned}$$

The geometrical properties of the functions $\varphi_{ab\theta}$ and $M\varphi_{ab\theta}$ are very similar. First, because translation in spatial domain means modulation in frequency domain, the parameter b will just translate $M\varphi_{ab\theta}$. Secondly, since m is purely a radial function in Fourier domain, it does not affect rotation properties, i.e. θ act as a rotation parameter also for $M\varphi_{ab\theta}$. True parabolic scaling is the only property that is lost because of Fourier multiplier. In Section 2.7.3 it is shown that if φ is properly chosen then $\varphi_{ab\theta}$ and $M\varphi_{ab\theta}$ have essentially the same decay properties (see Lemma 11). This will lead to the fact that $\varphi_{ab\theta}$ and $M\varphi_{ab\theta}$ can be used in exactly the same way when investigating regularity properties with techniques presented in this thesis.

2.4 Curvelet transform

Since curvelets are the most used basis and curvelet transform the most used transform in this thesis, they are reviewed in details in this section. The continuous curvelet transform (CCT) is used in regularity analysis in Section 4.2 and the discrete curvelet transform, i.e. curvelet basis, is utilized in theorems of Section 3.

2.4.1 Continuous Curvelet transform

There exists different constructions of curvelets, the one used in [6, 7] is reviewed in this section. This continuous curvelet transform (CCT) has a simpler inversion formula than that of Hart Smith transform. Still, it has essentially similar decay properties as $\varphi_{ab\theta}$, just like $M\varphi_{ab\theta}$ have. See Lemma 11.

CCT is defined in the polar coordinates (r, ω) of the Fourier/frequency domain. Let W be a positive real-valued function supported inside $\left(\frac{1}{2}, 2\right)$, called a *radial window*, and let V be a real-valued function supported on $[-1, 1]$, called an *angular window*. Windows W and V should meet the following admissibility conditions:

$$\int_0^\infty W(r)^2 \frac{dr}{r} = 1 \qquad \text{and} \qquad \int_{-1}^1 V(\omega)^2 \, d\omega = 1. \tag{14}$$

At each scale a, $0 < a < a_0$, γ_{a00} is defined by

$$\widehat{\gamma_{a00}}\left(r\cos(\omega), r\sin(\omega)\right) = a^{\frac{3}{4}} W\left(ar\right) V\left(\omega/\sqrt{a}\right) \quad \text{for } r \geq 0 \text{ and } \omega \in [0, 2\pi).$$

For each $0 < a < a_0$, $b \in \mathbb{R}^2$, and $\theta \in [0, 2\pi)$, a *curvelet* $\gamma_{ab\theta}$ is defined by

$$\gamma_{ab\theta}(x) = \gamma_{a00}\left(\mathrm{R}_\theta\left(x - b\right)\right), \quad \text{for } x \in \mathbb{R}^2. \tag{15}$$

Notice that now curvelets have a little bit different generating function γ for each scale. This is different to Hart Smith or wavelet transform. Also, because of definition of radial window, functions $\gamma_{ab\theta}$ are high frequency functions.

The continuous curvelet transform is, just like Hart Smith or wavelet transform, an integral transform

$$\Gamma_f\left(a, b, \theta\right) := \langle \gamma_{ab\theta}, f \rangle \quad \text{for all } 0 < a < a_0, b \in \mathbb{R}^2, \text{ and } \theta \in [0, 2\pi).$$

In [6] it was assumed that V and W are C^∞, which implies that the curvelets and their derivatives have rapid decay. For purposes of this thesis, it would be enough to assume only that V and W are C^N for N large enough. This assumption would then ensure that the resulting curvelets and their derivatives up to desired order (in regularity analysis, largest α of interest), would decay fast enough. Anyway, to keep notations a bit more nice, the assumption that $V, W \in C^\infty$ is used also in this thesis. Lemma 10 will state that this construction will lead to directional vanishing moments with increasing number of directions when a degreases.

The following reconstruction formula exists for curvelet transform, for all $f \in L^2(\mathbb{R}^2)$ [7].

Theorem 5 *There exists a bandlimited purely radial function Φ such that for all $f \in L^2(\mathbb{R}^2)$,*

$$f = \int_{\mathbb{R}^2} \langle \Phi_b, f \rangle \, \Phi_b \, db + \int_0^{a_0} \int_0^{2\pi} \int_{\mathbb{R}^2} \langle \gamma_{ab\theta}, f \rangle \, \gamma_{ab\theta} \, db \, d\theta \, \frac{da}{a^3}, \tag{16}$$

where $\Phi_b(x) = \Phi(x - b)$.

Just like with Hart Smith transform the low frequency part $\int_{\mathbb{R}^2} \langle \Phi_b, f \rangle \, \Phi_b \, db$ of function f is uninteresting in this thesis since continuous curvelet transform is used here only in regularity analysis.

2.4.2 Discrete Curvelet transform

Candes and Donoho have presented a slightly different way to construct curvelet tight frames [7, 4], both leading to essentially similar basis functions. We give a fast review on how the construction used in [5] is related to CCT defined above. The aim is not to be very rigorous mathematically, but clarify that the construction can be done essentially so that basis functions have the same geometric properties as $\gamma_{ab\theta}$ related to CCT. Details can be found in [5].

First, remember that $\widehat{\gamma_{a00}}$ in CCT was defined as product of a (smooth) radial and angular window. In the discrete curvelet transform we have a very same starting point: compactly supported and smooth angular and radial windows are first defined, and then a window w_1 is defined as product of those windows. Next this window is moved to left side of the ξ_1-axis by reflecting it around ξ_2 axis, let us denote this by w_2. The final window is then $\widehat{\gamma_{a00}} := w_1 + w_2$.

We would like to note that other conditions for window functions are little different from the admissability conditions presented in the continuous case. These so called "partition of unity" properties, are explained in [5] with details.

Next the translation parameter is constructed. First define rectangles r_1 and r_2 s.t. $\text{supp}(w_a) \subset r_1 \cup r_2$. For $L^2(r_1 \cup r_2)$ define now orthonormal basis by using orthonormal basis of complex exponentials in both directions ξ_1 and ξ_2. For this basis $\{f_{a,b}\}_b$ the index $b = (b_1, b_2)$ describes coefficients in exponents of basis function. Curvelet basis functions γ_{ab0} for fixed scale a and orientation $\theta = 0$ are defined now in Fourier domain by

$$\widehat{\gamma_{ab0}} = \widehat{\gamma_{a00}} \cdot f_{a,b}.$$

Notice, essentially, that multiplying in Fourier domain by complex exponential means the same as translation of function in spatial domain. Therefore, by taking inverse Fourier transform of $\widehat{\gamma_{ab0}}$, the term $f_{a,b}$ affects essentially as a translation while the term γ_{a00} describes the shape of γ_{ab0}. Basis functions $f_{a,b}$ are chosen so that when b_1 changes by one, translation in x_1 direction is about length a, and if b_2 changes by one, translation in x_2 direction is about $a^{1/2}$.

Note: to be specific, $f_{a,b}$ does not affect only as an translation for ψ, but that part of it that would affect the shape can be included in the original radial window. Details can be found again from [5].

The rotation parameter is now easy to add. Just define $\gamma_{ab\theta}(x) := \gamma_{ab0}(\text{R}_{-\theta}x)$.

It is still unclear how parameters a and θ should be discretized. As discussed in the Section 1.1.2, from the approximation point of view it would be good if there are about $a^{-1/2}$ different angles. Therefore, the only question is how to discretize a. For example, $a = 2^{-j}$ would work otherwise fine, but then the number of orientations is

not always integer, and therefore it should be rounded on every other scale. So, from implementation point of view, the choice $a = 4^{-j} = 2^{-2j}$ might be better. Generally any discretization of type $a = c^{-j}$, where $c > 1$ is constants, would lead to same results [5].

The last thing would be to introduce some basis functions Φ_b that are concentrated to low frequencies, just as in the continuous case. However, from a theoretical point, these are not very interesting: in regularity analysis the low frequencies do not matter, and for the convergence of the M-term approximation the fixed single scale can not contain too many basis functions with with large coefficients, as long as these basis functions have a rapid decay. Once again, details of construction of Φ_b are given in [5]. The set $\{\gamma_{ab\theta}\} \cup \{\Phi_b\}$ forms now the normalized tight frame for $L^2(\mathbb{R}^2)$ [5].

2.5 Shearlet transform

In this section only a short revision on shearlet transform is given. The aim is to emphasize its differences to curvelet and contourlet transforms. In later sections some theorems related to shearlets are revised from [24], and numerical examples are given. All this will demonstrate that shearlets are very similar to curvelets and contourlets.

2.5.1 Continouous Shearlet transform

This section follows the definitions and notations in G. Kutyniok and D. Labate[23]. Previously the function ψ was related to wavelets. Wavelets are not considered in later sections and ψ is from now on related to shearlets. While other transforms had rotation parameter θ, corresponding parameter for shearlets is shearing s, which emphasize the difference.

Given ψ_1 and $\psi_2 \in L^2(\mathbb{R})$, let $\psi \in L^2(\mathbb{R}^2)$ be defined by

$$\hat{\psi}(\xi) := \hat{\psi}_1(\xi_1)\hat{\psi}_2\left(\frac{\xi_2}{\xi_1}\right) \tag{17}$$

when $\xi = (\xi_1, \xi_2) \in \mathbb{R}^2$ with $\xi_1 \neq 0$.

Definition 6 *Let $\psi \in L^2(\mathbb{R}^2)$ be given by (17) where*

i. *$\psi_1 \in L^2(\mathbb{R})$ satisfies the admissibility condition, and $\hat{\psi}_1 \in C^\infty(\mathbb{R})$ with $\operatorname{supp}\hat{\psi}_1 \subset [-2, -\frac{1}{2}] \cup [\frac{1}{2}, 2]$;*

ii. *$\|\psi\|_2 = 1$, and $\hat{\psi}_2 \in C^\infty(\mathbb{R})$ with $\operatorname{supp}\hat{\psi}_2 \subset [-1, 1]$ and $\hat{\psi}_2 > 0$ on $(-1, 1)$.*

Such a function ψ is called a *shearlet function*. Notice that the roles of ψ_1 and ψ_2 are somewhat similar to the roles of angular and radial window functions in CCT, although polar coordinates are not used. The use of polar coordinates is the key point that make a digital implementation of curvelet transform a bit problematic and shearlets try to avoid that.

A *continuous shearlet system* is the set of functions generated by ψ, namely,

$$\left\{\psi_{ast} = a^{-\frac{3}{4}}\psi\left(\mathrm{M}_{as}^{-1}(\cdot - t)\right) : \, a \in \mathrm{I} \subset \mathbb{R}^+, \, s \in \mathrm{S} \subset \mathbb{R}, \, t \in \mathbb{R}^2\right\},$$

where $\mathrm{M}_{as} = \mathrm{B}_s\mathrm{D}_a$, B_s is the *shear matrix* $\begin{pmatrix} 1 & -s \\ 0 & 1 \end{pmatrix}$, and D_a is the diagonal matrix $\begin{pmatrix} a & 0 \\ 0 & \sqrt{a} \end{pmatrix}$. The *continuous shearlet transform* (CST) of f is then defined by

$$\mathcal{SH}_\psi f(a, s, t) = \langle f, \psi_{ast} \rangle \, , \, a \in \mathrm{I} \subset \mathbb{R}^+, \, s \in \mathrm{S} \subset \mathbb{R}, \, t \in \mathbb{R}^2.$$

The most striking difference between shearlet and curvelet transform is that the rotation matrix is now replaced by shearing matrix. This shearing matrix changes the geometry while the rotation matrix kept the geometry. The bigger s is, the more the geometry of ψ_{ast} differs from the geometry of ψ_{a0t}. This might make a direct interpretation of the values of $\mathcal{SH}_\psi f(a, s, t)$ a bit complicated. For shearlets, the proper choice for set I of scales is a bit more delicate question, that will be specified a little later, as well as a set S of shear parameters.

Directly from the definition of ψ we have

$$\widehat{\psi_{ast}}(\xi) = a^{\frac{3}{4}} e^{-2\pi i \xi t} \hat{\psi}_1(a\xi_1) \hat{\psi}_2 \left(\frac{1}{\sqrt{a}} \left(\frac{\xi_2}{\xi_1} - s \right) \right).$$

The support for each function $\hat{\psi}_{ast}$ satisfies therefore

$$\operatorname{supp} \widehat{\psi_{ast}} \subseteq \left\{ (\xi_1, \xi_2) \colon \xi_1 \in \left[-\frac{2}{a}, -\frac{1}{2a} \right] \cup \left[\frac{1}{2a}, \frac{2}{a} \right], \left| \frac{\xi_2}{\xi_1} - s \right| \leq \sqrt{a} \right\}.$$

For this shearlet system the following reconstruction formula holds.

Theorem 7 *Let* $\mathrm{I} = \mathbb{R}^+$, $\mathrm{S} = \mathbb{R}$, *and* $\psi \in L^2(\mathbb{R}^2)$ *be a shearlet function. Then, for all* $f \in L^2(\mathbb{R}^2)$,

$$f = \int_{\mathbb{R}^2} \int_{\mathbb{R}} \int_{\mathbb{R}^+} \langle \psi_{ast}, f \rangle \, \psi_{ast} \, \frac{da}{a^3} \, ds \, dt \, . \tag{18}$$

A drawback in the above reconstruction formula is that the interval for shearing parameter is huge comparing to interval $[0, 2\pi)$ for θ in 16.

If the set S is not all of $\mathbb{R}$, then two different type of shearlets are needed: horizontally and vertically oriented.

The original shearlet do well as long as the shearing parameter s is not large. Consider the subspace of $L^2(\mathbb{R}^2)$ given by $L^2(\mathrm{C})\check{} = \{f \in L^2(\mathbb{R}^2) \colon \operatorname{supp} \hat{f} \subset C\}$ where C is given by

$$C = \left\{ (\xi_1, \xi_2) \in \mathbb{R}^2 \colon |\xi_1| \geq 2 \text{ and } \left| \frac{\xi_2}{\xi_1} \right| \leq 1 \right\}.$$

Theorem 8 *Let* $\mathrm{I} = \{a \colon 0 < a < 1\}$, $\mathrm{S} = \{s \colon |s| \leq 2\}$ *and* $\psi \in L^2(\mathbb{R}^2)$ *be a shearlet function. Then, for all* $f \in L^2(\check{C})$,

$$f = \int_{\mathbb{R}^2} \int_{-2}^{2} \int_{0}^{1} \langle \psi_{ast}, f \rangle \, \psi_{ast} \, \frac{da}{a^3} \, ds \, dt. \tag{19}$$

From now on $\mathrm{S} = \{s \colon |s| \leq 2\}$ and $\mathrm{I} = \{a \colon 0 < a < 1\}$. Theorem 8 takes care of f that are supported on horizontal cones in Fourier domain.

Next define "vertical" shearlets that can reproduce the functions that are supported on vertical cones in Fourier domain. Most easily this is done by changing roles of ξ_1 and ξ_2 in every formula above:

$$\widehat{\psi^{(v)}}(\xi) = \widehat{\psi^{(v)}}(\xi_1, \xi_2) = \hat{\psi}_1(\xi_2) \hat{\psi}_2 \left(\frac{\xi_1}{\xi_2} \right) \tag{20}$$

where $\hat{\psi}_1, \hat{\psi}_2$ are defined as in Definition 6. The shearlets $\psi^{(v)}_{ast}$ are defined by $\psi^{(v)}_{ast} = a^{-\frac{3}{4}} \psi \left((\mathrm{M}^{(v)}_{as})^{-1} (\cdot - t) \right)$, where $\mathrm{M}^{(v)}_{as} = \mathrm{B}^{(v)}_s \mathrm{D}^{(v)}_a$ with $\mathrm{B}^{(v)}_s = \mathrm{B}^T_s$ and $\mathrm{D}^{(v)}_a = \begin{pmatrix} \sqrt{a} & 0 \\ 0 & a \end{pmatrix}$. Then $\left\{ \psi^{(v)}_{ast} \right\}$ is a continuous shearlet system for $L^2(C^{(v)})\check{}$ where $C^{(v)}$ is the vertical cone;

$$C^{(v)} = \left\{ (\xi_1, \xi_2) \in \mathbb{R}^2 \colon |\xi_2| \geq 2 \text{ and } \left| \frac{\xi_2}{\xi_1} \right| > 1 \right\}.$$

"Vertical" continuous shearlet transform is then

$$\mathcal{SH}^{(v)}_{\psi} f(a, s, t) = \left\langle f, \psi^{(v)}_{ast} \right\rangle,$$

and we have $\widehat{\psi^{(v)}_{ast}}(\xi) = a^{\frac{3}{4}} e^{-2\pi i \xi t} \hat{\psi}_1(a\xi_2) \hat{\psi}_2 \left(\frac{1}{\sqrt{a}} \left(\frac{\xi_1}{\xi_2} - s \right) \right)$, hence

$$\operatorname{supp} \widehat{\psi^{(v)}_{ast}} \subseteq \left\{ (\xi_1, \xi_2) \colon \xi_2 \in \left[-\frac{2}{a}, -\frac{1}{2a} \right] \cup \left[\frac{1}{2a}, \frac{2}{a} \right], \left| \frac{\xi_1}{\xi_2} - s \right| \leq \sqrt{a} \right\}.$$

Now the part $C^{(v)} \cup C$ of Fourier plane is taken care and only the low frequency part $\mathbb{R}^2 \setminus (C^{(v)} \cup C)$ is left. As argued before, this might not be very interesting in context of this thesis but it can be done: Let $W(x)$ be such that $\widehat{W}(\xi) \in C^\infty(\mathbb{R}^2)$ and

$$|\widehat{W}(\xi)|^2 + \chi_{C_1}(\xi) \int_0^1 |\hat{\psi}_1(a\xi_1)|^2 \frac{da}{a} + \chi_{C_2}(\xi) \int_0^1 |\hat{\psi}_1(a\xi_2)|^2 \frac{da}{a} = 1, \text{ for a.e. } \xi \in \mathbb{R}^2,$$

where $C_1 = \left\{(\xi_1, \xi_2) \in \mathbb{R}^2 \colon |\frac{\xi_2}{\xi_1}| \leq 1\right\}$ and $C_2 = \left\{(\xi_1, \xi_2) \in \mathbb{R}^2 \colon |\frac{\xi_2}{\xi_1}| > 1\right\}$.

Then it follows that W is a C^∞-window function on $\mathbb{R}^2$ with $\widehat{W}(\xi) = 1$ for $\xi \in [-1/2, 1/2]^2$ and $\widehat{W} = 0$ outside the box $[-2, 2]^2$. Finally, let $\widehat{P_{C_1} f} = \hat{f}\chi_{C_1}$ and $\widehat{P_{C_2} f} = \hat{f}\chi_{C_2}$. Then, for each $f \in L^2(\mathbb{R}^2)$ we have

$$f(x) = \int_{\mathbb{R}^2} \langle W(\cdot - t), f\rangle\, W(x-t)\, dt + \int_{\mathbb{R}^2}\int_{-2}^{2}\int_{0}^{1} \langle \psi_{ast}, P_{C_1} f\rangle\, \psi_{ast}(x)\, \frac{da}{a^3}\, ds\, dt$$
$$+ \int_{\mathbb{R}^2}\int_{-2}^{2}\int_{0}^{1} \left\langle \psi^{(v)}_{ast}, P_{C_2} f\right\rangle \psi^{(v)}_{ast}(x)\, \frac{da}{a^3}\, ds\, dt.$$

It is worth to notice that functions ψ_{ast} and $\psi^{(v)}_{ast}$ are actually very similar. Their supports go away from origin of the Fourier plane essentially similarly as what happen for curvelets when scale a degreases.

2.5.2 Discrete Shearlet transform

Moving from continuous transform to discrete transform happens with shearlets in a quite similar way as it happened with curvelets: sampling parameters a, t and s of continuous shearlets. This must be done separeately for horizontal and vertical shearlets. As an example, in horizontal case the samples are of form $a_j = 2^{-2j}$ for $j \geq 0$, $s_{jk} = k\sqrt{a_j} = k2^{-j}$ for $-2^j \leq k \leq 2^j$ and $t_{jkm} = \mathrm{B}_{s_{jk}}\mathrm{D}_{a_j} m$ for $m \in \mathbb{Z}^2$.

With some extra conditions for ψ_1 and ψ_2, this kind of straightforward construction leads to tight frames for $L^2(\mathbb{R}^2)$. Details can be founded in [14].

2.6 Contourlet transform

The contourlet transform is not a main theme of this thesis, since none of theorems in the thesis is written for contourlets. However, there is a good reason to believe that similar theorems might exist also for contourlets, since this transform is so similar to curvelet and shearlet transform. Here we discuss similarities and differences of the contourlet transform as compared to transforms presented in earlier sections.

The major difference is that the contourlet transform is defined with iterated (directional) filter bank structure. So, the nature of transform is discrete and it has an efficient implementation with digital filters. Contourlet basis functions are defined via these filters, but there is no real need to construct basis functions since all numerical calculations are done directly by using filters.

Essentially, two different kind of approaches are proposed to design these filters. First is the method that tries to maximize the number of directions for directional vanishing moments. The second method proposes to use filters with (approximately)

ideal frequency selectivity. This method leads to a very similar frequency support for basis functions as with shearlets. If these filters are designed so that they lead to rapidly decaying basis functions, then contourlet basis functions would have essentially similar properties as those of shearlet or curvelet basis functions. In that case, theorems similar to the ones presented in Section 3 for curvelets might hold also for contourlets. However, the author of the thesis is not aware of existence of this kind of filters. If filters with finite length are used in design of contourletes, then related contourlet basis functions have compact support but ideal frequency selectivity, that was mentioned before, is not possible. For more information about contourlets and filter desing, see [11, 34, 8].

2.7 General properties of transforms

In this section it is shown that all functions $\varphi_{ab\theta}$, $M\varphi_{ab\theta}$, and $\gamma_{ab\theta}$ do have the same essential properties that are needed for proofs of theorems in later Sections 3 and 4. This will make all of them almost equivalent when investigating Hölder regularities in the Section 4. For approximation of function, that is discussed in Section 3, only properties of curvelets $\gamma_{ab\theta}$ are important since those theorems are related only to curvelets. Although proofs for theorems in this section are written only for $\gamma_{ab\theta}$ that is defined as in CCT, they would be almost identical for basis functions of curvelet tight frames since, discussed in Section , the construction is essentially similar. Lemmas as presented in this section are from [36]. Similar lemmas hold also for shearlets [24].

2.7.1 Directional vanishing moments

For any nonzero vectors v and v' in $\mathbb{R}^2$, let us denote the angle from v to v' in clockwise direction by $\angle(v, v')$.

Definition 9 *A function f of two variables is said to have a L-*order directional vanishing moments along a direction $v = (v_1, v_2)^T$ *(suppose that $v_1 \neq 0$; if $v_1 = 0$ then $v_2 \neq 0$ and we can swap the two dimensions) if*

$$\int_{\mathbb{R}} t^n f(t, tv_2/v_1 - c)dt = 0, \quad \forall c \in \mathbb{R},\ 0 \leq n \leq L.$$

Essentially, the above definition means that any 1-D slices of the function have vanishing moments of order L. Notice from the definition that f has directional vanishing moment along direction v if and only if the same holds along direction $-v$. The Lemma 10 below essentially relates directions of vanishing moments to scale a.

Lemma 10 *There exists $C < \infty$ (independent of a, b and θ) such that the curvelet functions $\gamma_{ab\theta}$ have directional vanishing moments of any order $L < \infty$ along all*

directions v that satisfy $\pi/2 \geq |\angle(v_\theta, v)| \geq Ca^{1/2}$. Moreover, if there exists finite and strictly positive constants C_1, C_1' and C_2 such that $\operatorname{supp}(\hat{\varphi}) \subset [C_1, C_1'] \times [-C_2, C_2]$, then the above is true also for functions $\varphi_{ab\theta}$ and $M\varphi_{ab\theta}$.

Proof. Because m is smooth and bounded, the Fourier transforms $\widehat{\varphi_{ab\theta}}(\xi)$ and $m(\|\xi\|)\widehat{\varphi_{ab\theta}}(\xi)$ of functions $\varphi_{ab\theta}$ and $M\varphi_{ab\theta}$ have the same compact support (and the same order of differentiability). This makes them equivalent in the sense that both satisfy conditions of the Lemma and therefore it is enough to produce a proof for $\varphi_{ab\theta}$. Moreover, compact support in frequency domain implies that $\varphi_{ab\theta}$ and $M\varphi_{ab\theta}$ are in C^∞.

Let $a > 0$, $b \in \mathbb{R}^2$ and $\theta = 0$, so $v_\theta = (0,1)^T$. Let $v = (v_1, v_2)^T = (\cos\theta', \sin\theta')^T$ be a unit vector such that $|\angle(v_0, v)| \geq Ca^{1/2}$. We make now a restriction to values $\theta' \in [0, \pi/2)$, i.e. $0 < \angle(v_0, v) \leq \pi/2$. On this interval we always have $\angle(v_0, v) < \tan(\angle(v_0, v))$. Choose now $C = C_2/C_1$. Because

$$\frac{v_1}{v_2} = \tan\left(\angle(v_0, v)\right) > \angle(v_0, v) \geq Ca^{1/2} = \frac{C_2 a^{-1/2}}{C_1 a^{-1}},$$

it follows that the lines passing through the origin with slopes $\pm\frac{v_1}{v_2}$ do not intersect the rectangle $[a^{-1}C_1, a^{-1}C_1'] \times [-a^{-1/2}C_2, a^{-1/2}C_2]$. In particular, since $\operatorname{supp}(\widehat{\varphi_{ab0}})$ is a subset of the rectangle, it does not intersect with the line $\xi_2 = -\frac{v_1}{v_2}\xi_1$. Now, since rotation commutes with Fourier transform, we have $\operatorname{supp}(\widehat{\varphi_{ab\sigma}}) = \mathrm{R}_\sigma \operatorname{supp}(\widehat{\varphi_{ab0}})$ and so

$$\begin{aligned}
&\operatorname{supp}(\widehat{\varphi_{ab0}}) \cap \{(\xi_1, \xi_2) \colon \xi_2 = -\frac{v_1}{v_2}\xi_1\} = \emptyset \\
\iff &\operatorname{supp}(\widehat{\varphi_{ab(-\theta')}}) \cap \mathrm{R}_{-\theta'}\{(\xi_1, \xi_2) \colon \xi_2 = -\frac{v_1}{v_2}\xi_1\} = \emptyset \\
\iff &\operatorname{supp}(\widehat{\varphi_{ab(-\theta')}}) \cap \{(\xi_1, \xi_2) \colon \xi_1 = 0\} = \emptyset
\end{aligned}$$

Consequently, all partial derivatives of $\widehat{\varphi_{ab(-\theta')}}$ vanish on the ξ_2-axis. This property of $\varphi_{ab(-\theta')}$ implies that it has directional vanishing moments along the direction of x_1-axis of any order L. Indeed, if we denote $g(x_2) := \int x_1^n \varphi_{ab(-\theta')}(x_1, x_2)dx_1$ then differentiating under the integral sign gives

$$\begin{aligned}
0 &= \partial_1^n \widehat{\varphi_{ab(-\theta')}}(0, \xi_2) \\
&= (-2\pi i)^n \int \left(\int x_1^n \varphi_{ab(-\theta')}(x_1, x_2)dx_1 \right) e^{-2\pi i x_2 \xi_2} dx_2 \\
&= (-2\pi i)^n \int g(x_2) e^{-2\pi i x_2 \xi_2} dx_2 \\
&= (-2\pi i)^n \hat{g}(\xi_2),
\end{aligned}$$

for all $\xi_2 \in \hat{\mathbb{R}}$, which implies that $g(x_2) \equiv 0$. Therefore, φ_{ab0} has vanishing moments along the direction v. For angles $\theta' \in (\pi/2, \pi]$ one can make similar derivations since

$[a^{-1}C_1, a^{-1}C_1'] \times [-a^{-1/2}C_2, a^{-1/2}C_2]$ is symmetric about the ξ_1-axis. And proof for the case $\theta = 0$ is complete.

The general case for $\theta \in [0, 2\pi)$ follows easily from the fact that θ just rotates the function $\varphi_{ab\theta}$. Because $\varphi_{ab\theta}$ has directional vanishing moments along the direction v if and only if φ_{ab0} have directional vanishing moments along direction $\mathrm{R}_{-\theta}v$ and

$$|\angle(v_\theta, v)| = |\angle(\mathrm{R}_{-\theta}v_\theta, \mathrm{R}_{-\theta}v)| = |\angle(v_0, \mathrm{R}_{-\theta}v)|\,,$$

we can denote $\mathrm{R}_{-\theta}v := (\cos\theta', \sin\theta')^T$ and make exactly same calculations as in the case $\theta = 0$. For curvelets, one can use exactly the same proof because clearly $\mathrm{supp}(\widehat{\gamma_{a00}}) \subseteq [a^{-1}C_1, a^{-1}C_1'] \times [-a^{-1/2}C_2, a^{-1/2}C_2]$ for some strictly positive C_1, C_1' and C_2. ■

2.7.2 Smoothness of kernel and basis functions

Smoothness of basis functions is an important subject in approximation and compression applications. If basis functions are very rough, the visual quality might be bad even if the error $\|f - f_{,M,B}\|_2$ is fairly small. With basis functions used in this thesis this is not an issue since all basis functions had compact support in Fourier domain, that makes them to be C^∞ functions.

2.7.3 Decay of kernel and basis functions

A result on the decay of $\partial^\nu\varphi_{ab\theta}$, $\partial^\nu M\varphi_{ab\theta}$, and $\partial^\nu\gamma_{ab\theta}$ is given below.

Lemma 11 *Suppose that the windows V and W in the definition of CCT are C^∞ and have compact supports. Then for each $N = 1, 2, ...$ there is a constant C_N such that*

$$\forall x \in \mathbb{R}^2 \quad |\partial^\nu\gamma_{ab\theta}(x)| \leq \frac{C_N a^{-3/4-|\nu|}}{1 + \|x - b\|_{a,\theta}^{2N}}. \tag{21}$$

Moreover, if $\hat{\varphi} \in C^\infty$ and if there exist finite and strictly positive constants C_1, C_1', and C_2 such that $\mathrm{supp}(\hat{\varphi}) \subset [C_1, C_1'] \times [-C_2, C_2]$, then (21) also holds for functions $\varphi_{ab\theta}$ and $M\varphi_{ab\theta}$.

Proof. We produce a proof only for $M\varphi_{ab\theta}$. The proof is identical for $\varphi_{ab\theta}$ and, at the end of the proof, we will point out differences of proof for curvelets.

We first recall the basic properties of the Fourier transform:

$$f(x) = (-2\pi ix)^\nu g(x) \quad \Leftrightarrow \quad \hat{f}(\xi) = \partial^\nu\hat{g}(\xi)$$

$$f(x) = \partial^\nu g(x) \quad \Leftrightarrow \quad \hat{f}(\xi) = (2\pi i\xi)^\nu\hat{g}(\xi).$$

Therefore we have

$$\int_{\mathbb{R}^2} \Delta^k \hat{g}(\xi) e^{2\pi i x\cdot\xi} d\xi = (-4\pi^2 \|x\|^2)^k g(x), \tag{22}$$

where Δ is the Laplacian i.e. $\Delta := \partial_1^2 + \partial_2^2$. We restrict first ourself to the case $\theta = 0$ and $b = 0$. Fix an index vector $\nu := (\nu_1, \nu_2)$ and define

$$g_a(x) := M\varphi_{a00}(\mathrm{D}_a x) \quad \text{and} \quad h_a(x) := \partial^\nu g_a(x) = a^{\nu_1 + \frac{\nu_2}{2}} (\partial^\nu M\varphi_{a00})(\mathrm{D}_a x).$$

A straightforward computation yields

$$\begin{aligned}
\hat{h}_a(\xi) &= (i2\pi\xi)^\nu \hat{g}_a(\xi) \\
&= (2\pi i\xi)^\nu a^{-3/2} \widehat{M\varphi}_{a00}(\mathrm{D}_{1/a}\xi) \\
&= (2\pi i\xi)^\nu a^{-3/2} m(\left\|\mathrm{D}_{1/a}\xi\right\|) \hat{\varphi}_{a00}(\mathrm{D}_{1/a}\xi) \\
&= (2\pi i\xi)^\nu a^{-3/2} m(\left\|\mathrm{D}_{1/a}\xi\right\|) a^{-3/4} \hat{\varphi}(\xi) a^{3/2} \\
&= (2\pi i\xi)^\nu a^{-3/4} m(\left\|\mathrm{D}_{1/a}\xi\right\|) \hat{\varphi}(\xi).
\end{aligned} \tag{23}$$

If we now replace $\hat{g}$ by $\hat{h}_a$ and x by $\mathrm{D}_{1/a}x$ in equation (22) we get

$$\begin{aligned}
&\left|(-4\pi^2 \left\|\mathrm{D}_{1/a}x\right\|^2)^k a^{\nu_1+\nu_2/2} (\partial^\nu M\varphi_{a00})(x)\right| \\
&= \left|(-4\pi^2 \left\|\mathrm{D}_{1/a}x\right\|^2)^k h_a(\mathrm{D}_{1/a}x)\right| \\
&= \left|\int_{\mathbb{R}^2} (\Delta^k \hat{h}_a)(\xi) e^{2\pi i \mathrm{D}_{1/a}x\cdot\xi} d\xi\right| \\
&\le \int_{\mathbb{R}^2} \left|(\Delta^k \hat{h}_a)(\xi)\right| \left|e^{2\pi i \mathrm{D}_{1/a}x\cdot\xi}\right| d\xi \\
&= \int_{\mathbb{R}^2} \left|(\Delta^k \hat{h}_a)(\xi)\right| d\xi \\
&= a^{-3/4} \int_{\mathbb{R}^2} \left|\Delta^k ((2\pi i\xi)^\nu m(\left\|\mathrm{D}_{1/a}\xi\right\|)\hat{\varphi}(\xi))\right| d\xi \\
&\le C a^{-3/4}.
\end{aligned}$$

In the last step we used the observation that $\left|\Delta^k ((2\pi i\xi)^\nu m(\left\|\mathrm{D}_{1/a}\xi\right\|)\hat{\varphi}(\xi))\right| \le C$ where C is independent of a. This is a direct consequence of the fact that if $\xi \in \mathrm{supp}(\hat{\varphi})$ then

$$\left|m^{(n)}(\left\|\mathrm{D}_{1/a}\xi\right\|)\right| \le C \left\|\mathrm{D}_{1/a}\xi\right\|^{-n} \le C \left\|(a^{-1}C_1, 0)^T\right\|^{-n} = Ca^n. \tag{24}$$

When $k = 0$ this reduces to

$$\left|a^{\nu_1+\nu_2/2} (\partial^\nu M\varphi_{a00})(x)\right| \le C a^{-3/4}.$$

By combining the above estimates, we get

$$\begin{aligned}
&a^{\nu_1+\nu_2/2}(1+4^k\pi^{2k}\left\|\mathrm{D}_{1/a}x\right\|^{2k})\left|\partial^\nu M\varphi_{a00})(x)\right| \\
&= \left|a^{\nu_1+\nu_2/2}(\partial^\nu M\varphi_{a00})(x)\right| + \left|(-4\pi^2\left\|\mathrm{D}_{1/a}x\right\|^2)^k a^{\nu_1+\nu_2/2}(\partial^\nu M\varphi_{a00})(x)\right| \\
&\leq Ca^{-3/4},
\end{aligned}$$

which finally gives the inequality

$$|\partial^\nu M\varphi_{a00})(x)| \leq \frac{Ca^{-3/4-\nu_1-\nu_2/2}}{1+(2\pi)^{2k}\left\|\mathrm{D}_{1/a}x\right\|^{2k}} \leq \frac{Ca^{-3/4-\nu_1-\nu_2/2}}{1+\left\|\mathrm{D}_{1/a}x\right\|^{2k}}.$$

Now we give brief arguments how we get estimate for general θ and b. Because translation do not change regularity properties, it is clear that all above would hold also for a general b. Rotation will clearly change properties of partial derivatives of a function. However, taking partial derivative of a smooth function $M\varphi_{ab\theta}$ with respect to x_1 is the same as taking directional derivative of $M\varphi_{ab0}$ to direction $\mathrm{R}_{-\theta}(1,0)^T$. Since a directional derivative is a linear combination of partial derivatives, and partial derivatives with respect to x_1 and x_2 in the above calculations produced factors $a^{-\nu_1-\nu_2/2}$, in the worst case we will get factor $a^{-\nu_1-\nu_2} = a^{-|\nu|}$ instead of $a^{-\nu_1-\nu_2/2}$. For curvelets the only essential difference is that we choose $g_a(x) := \gamma_{a00}(\mathrm{D}_a x)$, which leads to formula

$$\begin{aligned}
\hat{h}_a(\xi) &= (2\pi i\xi)^\nu\, a^{-3/4}\hat{\varphi}_{a00}(\mathrm{D}_{1/a}\xi) \\
&= (2\pi i\xi)^\nu\, a^{-3/4} W\left(\sqrt{\xi_1^2+a\xi_2^2}\right) V\left(a^{-1/2}\arctan\frac{a^{1/2}\xi_2}{\xi_1}\right).
\end{aligned} \tag{25}$$

From compact supports of V and W we easily see that there exists positive constants C' and C'' s.t. $\xi \in \mathrm{supp}(\hat{h}_a)$ implicates $C' \leq \xi_1 \leq C''$ and $|\xi_2| \leq C''$ for all sufficiently small a. Therefore it is also easy to see from (25) that partial derivatives of $\hat{h}_a(\xi)$ are bounded by $Ca^{-3/4}$, and so we get the same estimate as before. ■

We would like to remark that, by geometric arguments, the above lemmas would be valid also when $\mathrm{supp}(\hat{\varphi}) \subset (([-C_1', C_1] \cup [C_1, C_1']) \times [-C_2, C_2])$. This is important if one prefers a real valued function φ.

3 Approximation of piecewise smooth functions by Curvelets

In later sections we will constantly build estimates for expressions that contain parameters a, b and k or θ (recall the notation $\theta = \frac{\pi}{2}ka^{1/2}$). Although some of these estimates hold for all a, b and k, for our purposes it does not matter if estimates hold only for $a < a_0$, for some constant a_0. This is because at coarse scale the Φ_b are used.

Many estimates consider how the transform start to decay when the angle parameter $|k|$ exceeds some constant value k_0, that depends only on f and γ. The optimal value for k_0 is out of interest here since it would just affect to a constant in front of final estimates. Therefore, we will define that the notation

$$\text{expression1}(a,b,k) \approx \text{expression2}(a,b,k)$$

is equivalent to the existence of a constant $C \in \mathbb{R}$, independent from a, b and k (i.e. θ) such that

$$\text{expression1}(a,b,k) = C \cdot \text{expression2}(a,b,k)$$

for all $b \in \mathbb{R}^2$, $a < a_0$ and $k > k_0$. Similarly

$$\text{expression1}(a,b,k) \lesssim \text{expression2}(a,b,k)$$

is equivalent to that there exists a constant $C \in \mathbb{R}$, independent from a, b and k such that

$$\text{expression1}(a,b,k) \leq C \cdot \text{expression2}(a,b,k)$$

for all $b \in \mathbb{R}^2$, $a < a_0$ and $k > k_0$. With help of these notations we do not have to update constant coefficients all the time.

3.1 Piecewise smooth functions

In the introduction, the class of piecewise smooth functions was discussed without mathematical details. Here we will give a bit more rigorous definitions. We restrict to functions $f \in L^2(\mathbb{R}^2)$ that have singularities only along smooth curves $S \in \mathbb{R}^2$. We start by giving rigorous definition for a smooth curve. In [5] where the convergence rates of $\|f - f_M\|_2^2$ are studied, S is defined in polar coordinates. However, inside the proofs the boundary is often considered locally as function of one variable. This is close to the approach used in design of bandelets [26], that are adaptive orthogonal basis which have optimal convergence rate of $\|f - f_M\|_2^2$. We choose to define the boundary in this way since we can use then smoothness properties more directly in proofs.

Definition 12 *Assume that a plane curve $S \in \mathbb{R}^2$ has tangent in all points and $s_{p,r}$ is a square centered at p and having side length r. Inside this square the coordinates are defined so that x_1-axis is parallel to the tangent of S at p, x_2-axis orthogonal to x_1-axis and origin in p. If there exists $r_0 > 0$ s.t. $\forall r < r_0$ inside $s_{p,r}$ the curve S is a function $g_{p,r} : \mathbb{R} \to \mathbb{R}$, $g_{p,r} \in C^n$, and all first n derivatives of $g_{p,r}$ are bounded by a constant C (independent from p and r), then we say that S is C^n smooth with bounded derivatives.*

Next we define the class $F_{\alpha,n}$ of functions for which the convergence of $\|f - f_{M,B}\|_2^2$ is studied.

Definition 13 *We say that $f \in F_{\alpha,n}$ if it is of the form $f = \chi_{\mathbb{R}^2\setminus A} f_1 + \chi_A f_2$, where $f_1, f_2 \in C^\alpha(\mathbb{R}^2)$ and have bounded derivatives and compact supports, and A is a compact set with boundary curve S that is C^n smooth with bounded derivatives.*

Now, for example, if $f \in F_{2,2}$ then curvelets have the approximation rate $\|f - f_{M,C}\|_2^2 \leq CM^{-2}(\log(M))^3$ but wavelets still have only the rate $\|f - f_{M,W}\|_2^2 \leq CM^{-1}$.

We make some remarks on Definition 13. In practice, supports are often restricted to the square $[0,1]^2$ (or more generally a rectangle) since this is quite standard in real applications, as in analysis of images. Secondly, $F_{\alpha,n}$ does not contain functions where S could have corners, which is usually the starting point. In some publications, such as [5] a careful analysis of the effect of corner points to the approximation rate is performed, but often it is omitted, see [11, 15]. Third, the author is not aware of any research with $n, \alpha > 2$ when transforms based on parabolic scaling are studied. With adaptive basis like bandelets [25] this kind of research is done. The last note is that there could be more than one region and boundary curve if those would not intersect each others. However in literature only this simple case of one curve is usually considered since techniques for proofs of theorems are the same.

3.2 Decay of transforms of piecewise smooth functions

The next theorem shows that by assuming more smoothness for f and S, the transform will decay faster when orientation turns away from orientation of S.

Theorem 14 *Let $f \in F_{n,n}$, $b \in S$ and $\theta' \approx ka^{1/2}$ be angle between major axis of $\gamma_{ab\theta}$ and tangents of S in point b, then*

$$\left| \int_{\mathbb{R}^2} f(x)\gamma_{ab\theta}(x)dx \right| \lesssim \begin{cases} a^{3/4} & , \quad |\theta'| \lesssim a^{1/2} \\ \frac{a^{3/4}}{|k|^3} & , \quad n = 2,\ |\theta'| \gtrsim a^{1/2} \\ \frac{a^{3/4}}{|k|^{3+\varepsilon}} & , \quad n \geq 3,\ |\theta'| \gtrsim a^{1/2} \end{cases} \tag{26}$$

for any fixed $\varepsilon < 2$.

Proof. From the parabolic scaling law its clear that

$$\left| \int_{\mathbb{R}^2} f(x)\gamma_{ab\theta}(x)dx \right| \lesssim a^{3/4}. \tag{27}$$

We use this estimate for "small" angles $|\theta'| \lesssim a^{1/2}$. The real challenge is to find good estimate when $|\theta'| \gtrsim a^{1/2}$. The proof is somewhat long, so we divide it in five different steps:

1. Localization of function and definition of appropriate coordinate systems
2. Partitioning of $\mathbb{R}^2$
3. Performing local twisting of function
4. Technical calculation of estimates of integrals in restricted domain
5. Extending estimates to whole domain.

1. Localization of function and defining appropriate coordinate systems For simplicity, we define x_1-axis as in the definition 12. In that case $\theta' = \theta$. We also choose $b = 0$, although it would be sufficient that $\|b\|_{a,\theta} \lesssim 1$.

The main ingredient of the proof is "twisting" the discontinuity curve to a straight line to gain maximum advantage from vanishing moments of $\gamma_{ab\theta}$. We actually make that twisting by a change of variable in integration. To avoid the error between twisted and real integral being too large, we do not make this change in the whole domain but only in a small region.

Next, we define $w_r \in C^\infty(\mathbb{R}^2)$ to be a smooth window function s.t. $w_r(x) = 0$ when $\|x\| > r/2$ and $w_r(x) = 1$ when $\|x\| < r/4$. Therefore

$$\left|\int_{\mathbb{R}^2} f\gamma_{ab\theta}dx\right| \leq \left|\int_{\mathbb{R}^2} (1-w_r)f\gamma_{ab\theta}dx\right| + \left|\int_{\mathbb{R}^2} w_r f\gamma_{ab\theta}dx\right|. \tag{28}$$

Because the size of window w_r is independent of a and γ has rapid decay, it is clear that for all $N < \infty$ there exists C_N s.t.

$$\left|\int_{\mathbb{R}^2} (1-w_r)f\gamma_{ab\theta}dx\right| \leq C_N a^N. \tag{29}$$

Now fix r be so small that S can be presented as a function $g(x_1)$ inside the support of w_r. By definition of $F_{n,n}$ it follows that if $n \geq 2$, then

$$g(0) = g'(0) = 0. \tag{30}$$

2. Partitioning of $\mathbb{R}^2$. We will divide $\mathbb{R}^2$ to small parallelograms by dividing it first by horizontal lines and then by lines parallel to major axis of $\gamma_{ab\theta}$.

First we divide $\mathbb{R}^2$ with horizontal lines to slices $R_{i,a}$, $i \in \mathbb{Z}$, that have same heights h_R which obey

$$h_R = a^{1/2}\left|\sin(\theta)\right| \lesssim a^{1/2}\left|\theta\right| \approx a\left|k\right|. \tag{31}$$

We can then write

$$\begin{aligned}\left|\int_{\mathbb{R}^2} w_r f \gamma_{ab\theta} dx\right| &= \left|\sum_{|i|=1}^{\infty} \int_{\mathbb{R}^2} \chi_{R_{i,a}} w_r f \gamma_{ab\theta} dx\right| \\ &\leq \sum_{|i|=1}^{\infty} \left|\int_{\mathbb{R}^2} \chi_{R_{i,a}} w_r f \gamma_{ab\theta} dx\right|\end{aligned} \tag{32}$$

Next, make a division with lines parallel to the main axis of $\gamma_{ab\theta}$. We set distance between each line being the same as length of translations caused to curvelet $\gamma_{ab\theta}$ when b_1 increase or decrease by one, i.e. this length is about a. This divides each $R_{i,a}$ to parallelograms $R_{i,a,l}$ that are aligned with major axis of $\gamma_{ab\theta}$. Notice that each of these parallelogram fit inside some rectangle that have side lengths about a and $a^{1/2}$ and orientation angel equal to θ. Also it is essential to notice that now each slice $R_{i,a}$ is constructed from these disjoint parallelograms $R_{i,a,l}$.

We will first investigate the slice $R_{i,a}$ that contains x_1-axis and is therefore most "corrupted" by the singularity curve. Without loss of generality, the partition $\mathbb{R}^2 = \cup_{|i|=1}^{\infty} R_{i,a}$ can be chosen so that the index i that refer to this slice is $i = i' = 0$.

Let us investigate a point $u = (u_1, u_2) \in S$. Because $g'(0) = 0$ and g'' is bounded, $|u_2| \leq |u_1|^2$. Therefore with relatively small values of $|l|$, inside in each $R_{i',a,l}$ there exists a much smaller region $A_{i',a,l}$ (another parallelogram with sides oriented similarly to sides of $R_{i',a,l}$) such that $u \in R_{i',a,l}$ only if $u \in A_{i',a,l}$. By the construction, the side of $A_{i',a,l}$ that is aligned with the x_1-axis has always the same length d,

$$d \approx \frac{a}{\sin(\theta)} \lesssim \frac{a}{a^{1/2}\,|k|} \approx \frac{a^{1/2}}{|k|}, \tag{33}$$

but the heights can vary. However, techniques we use later require that this height, let denote it by h_A, is the same for all $A_{i',a,l}$. We set this height to be

$$h_A \approx \frac{a}{|k|^{\varepsilon}}, \tag{34}$$

where $\varepsilon \geq 0$. The particular choice $\varepsilon = 0$ will lead to the decay estimate that provides well known non-linear approximation rate $O(M^{-2}(\log(M))^3$. However, by choosing $\varepsilon > 0$ and assuming f and S being a bit smoother, we will gain a decay estimate that will provide the approximation rate $O(M^{-2})$.

Next we derive the bound l_{max} such that

$$x \in S,\ x \in R_{i',a,l},\ |l| < l_{max} \quad \Rightarrow \quad x \in A_{i',a,l} \tag{35}$$

Since

$$\sup_{x \in A_{i',a,l}} |x_1| \approx \frac{h_A}{|\tan(\theta)|} + |l|\, d, \tag{36}$$

we get

$$\begin{aligned}\sup_{x\in A_{i',a,l}} |g(x_1)| &\lesssim \sup_{x\in A_{i',a,l}} |x_1|^2 \\ &\approx \left(\frac{h_A}{\tan(\theta)} + |l|\, d\right)^2 \\ &\lesssim \left(\frac{a^{1/2}}{|k|^{1+\varepsilon}} + \frac{|l|\, a^{1/2}}{|k|}\right)^2 \\ &\lesssim \left(\frac{|l|\, a^{1/2}}{|k|}\right)^2 \\ &= \frac{l^2 a}{k^2}.\end{aligned} \tag{37}$$

Because we need $\sup_{x\in A_{i',a,l}} |g(x_1)| \le h_A$, we will put a condition for l:

$$\begin{aligned}&\frac{l^2 a}{k^2} \lesssim h_A \\ \Leftrightarrow\quad &|l| \lesssim |k|^{1-\varepsilon/2}\end{aligned} \tag{38}$$

We denote now the bound found as

$$l_{max} \approx |k|^{1-\varepsilon/2}\,. \tag{39}$$

By the construction, we can set regions $A_{i',a,l}$ so that $\cup_{|l|=0}^{\infty} A_{i',a,l}$ is a horizontal slice of $\mathbb{R}^2$ and $\cup_{|l|=0}^{l_{max}} A_{i',a,l}$ is parallelogram. We add to this parallelogram two "small" right-angled triangles, A_1 and A_2, so that

$$R_T := \left(\cup_{l=1}^{l_{max}} A_{i',l,a}\right) \cup A_1 \cup A_2 \tag{40}$$

is rectangle. Notice that area of A_1 and A_2 is really negligible compared to the area of $\cup_{|l|=0}^{l_{max}} A_{i',a,l}$.

3. Performing local twisting of function Inside R_T we will perform a "twisting", i.e., make a change of variable that will allow us to consider f as smooth in direction of x_1-axis.

Let us define the twisting operator $T_g : \mathbb{R}^2 \to \mathbb{R}^2$ by formula

$$T_g(x_1, x_2) := \begin{cases} (x_1, x_2 - \frac{h_A - |x_2|}{h_A} g(x_1)) & , \quad (x_1, x_2) \in R_T \\ (x_1, x_2) & , \quad (x_1, x_2) \notin R_T \end{cases} \tag{41}$$

For a subset R of $\mathbb{R}^2$ operator T_g is defined similarly

$$T_g R := \{T_g(x_1, x_2) \mid (x_1, x_2) \in R\}\,. \tag{42}$$

It is essential to notice that

$$T_g R_T = R_T. \tag{43}$$

Even more vital is to notice that if apart from S, f is n times continuously differentiable (with bounded derivatives) and first n derivatives of g are continuous and bounded, then first n derivatives of the function

$$\tilde{f}(x) := \chi_{R_{i',a}} w_r f(T_g x), \tag{44}$$

in direction of x_1-axis are bounded and continuous inside R_T, and elsewhere in slide $R_{i',a}$ function $\tilde{f}(x)$ behaves essentially like f.

We perform now the change of variable

$$x = T_g y, \tag{45}$$

$$\begin{aligned} dx &= \begin{cases} \det(J(y))dy & , \quad y \in R_T \\ dy & , \quad y \notin R_T \end{cases} \\ &= \begin{cases} \frac{\partial x_2}{\partial y_2} dy & , \quad y \in R_T \\ dy & , \quad y \notin R_T \end{cases} \\ &= \begin{cases} (1 + sqn(y_2)\frac{g(y_1)}{h_A})dy & , \quad y \in R_T \\ dy & , \quad y \notin R_T \end{cases}, \end{aligned} \tag{46}$$

which gives

$$\left| \int_{\mathbb{R}^2} (1_{R_{i',a}} w_r f)(x) \gamma_{ab\theta}(x) dx \right| = \left| \int_{\mathbb{R}^2} \tilde{f}(y) \det(J(y)) \tilde{\gamma}_{ab\theta}(y) dy \right|, \tag{47}$$

where

$$h(y) := \tilde{f}(y) \det(J(y)) = \begin{cases} \tilde{f}(y) \det(J(y)) & , \quad y \in R_T \\ f(y) & , \quad y \notin R_T \end{cases} \tag{48}$$

and

$$\tilde{\gamma}_{ab\theta}(y) := \gamma_{ab\theta}(T_g y). \tag{49}$$

Notice that with this kind of change of variables the integration must be done separately inside and outside of region R_T, if regularity properties are used. This is because the mapping T_g might destroy all regularity properties on the border of R_T. We observe now some properties of h and $\tilde{\gamma}_{ab\theta}$. Since

$$|g(y_1)| \lesssim y_1^2, \quad |g'(y_1)| \lesssim |y_1|, \quad 0 \leq \frac{h_A - |y_2|}{h_A} \leq 1, \tag{50}$$

it is quite clear that

$$|h(y)| \lesssim 1, \quad y \in \mathbb{R}^2, \tag{51}$$

$$\left|\frac{\partial h(y)}{\partial y_1}\right| \lesssim \begin{cases} \frac{|y_1|}{h_A} & , \quad y \in R_T \\ 1 & , \quad y \notin R_T \end{cases} \tag{52}$$

and for $2 \leq m \leq n$

$$\left|\frac{\partial^m h(y)}{\partial y_1^m}\right| \lesssim \begin{cases} \frac{1}{h_A} & , \quad y \in R_T \\ 1 & , \quad y \notin R_T \end{cases} \tag{53}$$

On the border of R_T the function $\tilde{\gamma}_{ab\theta}$ is discontinuous but all decay properties of $\tilde{\gamma}_{ab\theta}$ (and it's derivatives) remain. Also

$$\tilde{\gamma}(y) = \gamma(y), \forall y \notin R_T. \tag{54}$$

However, unlike $\gamma_{ab\theta}$, the function $\tilde{\gamma}_{ab\theta}$ does not have directional vanishing moments. Because of that we will "recreate" function $\gamma_{ab\theta}$:

$$\begin{aligned} &\left|\int_{\mathbb{R}^2} (1_{R_{i',a}} w_r f)(x)\gamma_{ab\theta}(x)dx\right| \\ &= \left|\int_{\mathbb{R}^2} h(y)\tilde{\gamma}_{ab\theta}(y)dy\right| \\ = &\left|\int_{\mathbb{R}^2} h(y)(\tilde{\gamma}_{ab\theta}(y) - \gamma_{ab\theta}(y) + \gamma_{ab\theta}(y))dy\right| \\ \leq &\left|\int_{\mathbb{R}^2} h(y)(\tilde{\gamma}_{ab\theta}(y) - \gamma_{ab\theta}(y))dy\right| + \left|\int_{\mathbb{R}^2} h\gamma_{ab\theta}(y)dy\right|. \end{aligned} \tag{55}$$

4. Technical calculation of estimates of integrals in restricted domain Estimates for the integrals

$$I_1 = \left|\int_{\mathbb{R}^2} h(y)(\tilde{\gamma}_{ab\theta}(y) - \gamma_{ab\theta}(y))dy\right| = \left|\int_{R_T} h(y)(\tilde{\gamma}_{ab\theta}(y) - \gamma_{ab\theta}(y))dy\right| \tag{56}$$

and

$$I_2 = \left|\int_{\mathbb{R}^2} h\gamma_{ab\theta}(y)dy\right| \tag{57}$$

can be made separately. We first consider the integral I_2.

Remember, that in domain $R_{i',a}$ function h is smooth in direction of y_1-axis and therefore we can use efficiently directional vanishing moments of γ. For that we construct function P so that it is good polynomial approximation for h in y_1-direction. First we take a 1-D slice $h_{y_2}(y_1)$ of h in direction of y_1, i.e.

$$h_{y_2}(y_1) := h(y_1, y_2). \tag{58}$$

Then we define

$$P(y) := P_{y_2}(y_1), \tag{59}$$

where $P_{y_2}(y_1)$ is $n-1$:th order Taylor polynomial of $h_{y_2}(y_1)$. Next we use a basic trick to divide integral I_2 to two different integrals

$$\begin{aligned} I_2 &= \left| \int_{\mathbb{R}^2} h\gamma_{ab\theta} dy \right| \\ &= \left| \int_{\mathbb{R}^2} (h - P + P)\gamma_{ab\theta} dy \right| \\ &\leq \left| \int_{\mathbb{R}^2} (h - P)\gamma_{ab\theta} dy \right| + \left| \int_{\mathbb{R}^2} P\gamma_{ab\theta} dy \right| \\ &= I_{21} + I_{22} \end{aligned} \tag{60}$$

Because of directional vanishing moments of γ,

$$\int_{\mathbb{R}} P_{y_2}(y_1)\gamma_{ab\theta}(y_1, y_2) dy_1 = 0. \tag{61}$$

Since $P\gamma_{ab\theta}$ is absolutely integrable (because of rapid decay of γ), we can use the theorem of Fubini to make integration first in direction of y_1 and get

$$I_{22} = \left| \int_{\mathbb{R}^2} P\gamma_{ab\theta} dy \right| = 0. \tag{62}$$

From now on let us denote

$$R_I := \left(\cup_{|l| \leq l_{max}} R_{i',l,a} \right) \cup A_1 \cup A_2. \tag{63}$$

Integral I_{21} is first divided to three parts

$$\begin{aligned} I_{21} &= \left| \int_{\mathbb{R}^2} (h - P)\gamma_{ab\theta} dy \right| \\ &= \left| \int_{R_{i',a}} (h - P)\gamma_{ab\theta} dy \right| \\ &\leq \left| \int_{R_{i'a} \setminus R_I} (h - P)\gamma_{ab\theta} dy \right| + \left| \int_{R_I \setminus R_T} (h - P)\gamma_{ab\theta} dy \right| + \left| \int_{R_T} (h - P)\gamma_{ab\theta} dy \right| \\ &= I_{211} + I_{212} + I_{213}. \end{aligned} \tag{64}$$

The key for creating an estimate for I_{211} is rapid decay of γ:

$$\begin{aligned}
I_{211} &= \left| \int_{R_{i'a} \setminus R_I} (h - P) \gamma_{ab\theta} dy \right| \\
&\leq \int_{R_{i',a} \setminus R_I} |h - P| \, |\gamma_{ab\theta}| \, dy \\
&\lesssim \int_{R_{i',a} \setminus R_I} \max\{1, \|y_1\|^n\} \frac{a^{-3/4}}{1 + \|y\|_{a,\theta}^K} dy \\
&\lesssim \sum_{|l|=l_{max}}^{\infty} \int_{R_{i',a,l}} \max\{1, |dl|^n\} \frac{a^{-3/4}}{1 + |l|^K} dy \\
&\lesssim \sum_{|l|=l_{max}}^{\infty} h_R d \max\{1, |dl|^n\} \frac{a^{-3/4}}{1 + |l|^K} \\
&\lesssim \sum_{|l|=l_{max}}^{\infty} h_R d \frac{a^{-3/4}}{1 + |l|^{K-1-n}} \\
&\lesssim h_R d a^{-3/4} \, |l_{max}|^{-K+2+n} \\
&\lesssim a \, |k| \, \frac{a^{1/2}}{|k|} a^{-3/4} \, |k|^{(-K+2+n)(1-\varepsilon/2)} \\
&\approx \frac{a^{3/4}}{|k|^{K(1-\varepsilon/2)-(1-\varepsilon/2)(2+n)}}
\end{aligned} \tag{65}$$

.

When $\varepsilon < 2$ and K is chosen to be large enough this estimate is clearly acceptable.

For integral I_{212} the approach is quite similar, now we just use Taylor's theorem to bound $|h - P|$;

$$\begin{aligned}
I_{212} &= \left| \int_{R_I \setminus R_T} (h - P) \gamma_{ab\theta} dy \right| \\
&\lesssim \sum_{|l|=0}^{l_{max}} \int_{R_{i',a,l} \setminus A_{i,a,l}} |h - P| \frac{a^{-3/4}}{1 + \|y\|_{a,\theta}^K} dy \\
&\lesssim \sum_{|l|=0}^{l_{max}} \int_{R_{i',a,l} \setminus A_{i,a,l}} |d(|l| + 1)|^n \frac{a^{-3/4}}{1 + |l|^K} dy \\
&\lesssim \sum_{|l|=0}^{l_{max}} h_R d \, |d|^n \frac{a^{-3/4} (|l| + 1)^n}{1 + |l|^K} \\
&\lesssim h_R d \, |d|^n \, a^{-3/4} \\
&\lesssim a \, |k| \left(\frac{a^{1/2}}{|k|} \right)^{n+1} a^{-3/4}.
\end{aligned} \tag{66}$$

Estimate used for $\|y\|_{a,\theta}$ follows from the fact how we divided $\mathbb{R}^2$ to "grid" of parallelograms $R_{i',a,l}$: just have to count how many grid lines has to be crossed when move from origin to point y. From above we especially get

$$n = 2 \quad \Rightarrow \quad I_{212} \leq a^{7/4} |k|^{-2} \leq a^{3/4} |k|^{-4}, \tag{67}$$

$$n \geq 3 \quad \Rightarrow \quad I_{212} \leq a^{9/4} |k|^{-3} \leq a^{3/4} |k|^{-6}, \tag{68}$$

that are acceptable. For integral I_{213} the approach is identical to I_{212}

$$\begin{aligned} I_{213} &= \left| \int_{R_T} (h - P) \gamma_{ab\theta} dy \right| \\ &\lesssim \sum_{|l|=0}^{l_{max}} \int_{A_{i,l,a}} |h - P| \frac{a^{-3/4}}{1 + |y|_{a,\theta}^K} dy \\ &\lesssim \sum_{|l|=0}^{l_{max}} \int_{A_{i,l,a}} \frac{|d(|l| + 1)|^n}{h_A} \frac{a^{-3/4}}{1 + |l|^K} dy \\ &\lesssim \sum_{|l|=0}^{l_{max}} d |d|^n \frac{a^{-3/4} (|l| + 1)^n}{1 + |l|^K} \\ &\lesssim d |d|^n a^{-3/4} \\ &\lesssim \left(\frac{a^{1/2}}{|k|} \right)^{n+1} a^{-3/4}. \end{aligned} \tag{69}$$

Notice that when $n = 2$ this bound is $a^{3/4} |k|^{-3}$ and when $n = 3$ the bound is $\frac{a^2}{|k|^4} a^{-3/4} \leq a^{3/4} k^{-5}$ i.e. it is acceptable.

As a summary, if we wish to get I_2 bounded by $a^{3/4} |k|^{-3}$, then $n = 2$ is enough. If we wish to get bound I_2 by $a^{3/4} |k|^{-3-\varepsilon}$ for $2 > \varepsilon > 0$, then $n \geq 3$ is a sufficient condition.

Now we turn to investigate I_1. The key point is to make an estimate for the difference $|\tilde{\gamma}_{ab\theta}(y) - \gamma_{ab\theta}(y)|$. We first remind that

$$|\tilde{\gamma}_{ab\theta}(y) - \gamma_{ab\theta}(y)| = |\gamma_{ab\theta}(T_g y) - \gamma_{ab\theta}(y)|. \tag{70}$$

The difference between the arguments is therefore

$$\begin{aligned} &|T_g(y) - y| \\ &= \left| (y_1, y_2) - (y_1, y_2 - \frac{h_A - |y_2|}{h_A} g(y_1)) \right| \\ &= \left| (0, \frac{h_A - |y_2|}{h_A} g(y_1)) \right| \\ &\lesssim \frac{h_A - |y_2|}{h_A} y_1^2. \end{aligned} \tag{71}$$

Notice, especially, that in y_1 direction there is no difference in arguments. Now by remembering that derivatives of $\gamma_{ab\theta}$ have rapid decay (11), we can write that inside $A_{i',l,a}$ holds

$$\begin{aligned}
&|\gamma_{ab\theta}(T_g y) - \gamma_{ab\theta}(y)| \\
&\lesssim \frac{h_A - |y_2|}{h_A} y_1^2 \sup_{y \in A_{i',l,a}} \left|\partial^{(0,1)} \gamma_{ab\theta}(y)\right| \\
&\lesssim \frac{h_A - |y_2|}{h_A} y_1^2 \frac{a^{-7/4}}{1 + |l|^K} \\
&\leq y_1^2 \frac{a^{-7/4}}{1 + |l|^K}
\end{aligned} \quad . \tag{72}$$

Another important fact to notice is that

$$y \in A_{i,l,a} \quad \Rightarrow \quad |y_1| \lesssim h_A \left|\tan(\theta)\right| + d\,|l| \lesssim d\,|l|\,. \tag{73}$$

With these observations, and since h is bounded, we finally get

$$\begin{aligned}
&\left| \int_{\mathbb{R}} h(y)(\tilde{\gamma}_{ab\theta}(y) - \gamma_{ab\theta}(y)) dy \right| \\
\lesssim &\int_{\mathbb{R}} |\tilde{\gamma}_{ab\theta}(y) - \gamma_{ab\theta}(y)|\, dy \\
= &\int_{R_T} |\tilde{\gamma}_{ab\theta}(y) - \gamma_{ab\theta}(y)|\, dy \\
\lesssim &\sum_{|l|=0}^{l_{max}} \int_{A_{i',l,a}} |\tilde{\gamma}_{ab\theta}(y) - \gamma_{ab\theta}(y)|\, dy \\
\lesssim &\sum_{|l|=1}^{l_{max}} \int_{A_{i',l,a}} (dl)^2 \frac{a^{-7/4}}{1 + |l|^K} dy \\
\lesssim &\sum_{|l|=1}^{l_{max}} h_A d (dl)^2 \frac{a^{-7/4}}{1 + |l|^K} \\
\lesssim &h_A d^3 a^{-7/4} \\
\approx &\frac{a^{3/4}}{|k|^{3+\varepsilon}}
\end{aligned} \tag{74}$$

Since we got $I_1, I_2 \lesssim \frac{a^{3/4}}{|k|^{3+\varepsilon}}$, the estimate is now ready for the slice $R_{i',a} = R_{0,a}$.

5. Extending estimates to the whole domain For other slices $R_{i,a}$ we use similar approach. The key is to notice how in all estimates we used rapid decay property of

$\gamma_{ab\theta}$, in form $\frac{1}{1+|l|^K}$. Although we never wrote it before, we divide the summation over l as

$$\sum_{|l|=0}^{l_{max}} c_l \frac{1}{1+|l|^K} \lesssim \sum_{|l|=1}^{l_{max}} c_l \frac{1}{|l|^K} \tag{75}$$

when calculating any estimates. With this trick we do not have to worry about $l = 0$. Now for slices $R_{i,a}$, $i \neq 0$, the essential decay estimate is

$$\frac{1}{|i+l|^K} \leq \frac{1}{|i|^K + |il|^{K/2} + |l|^K} \leq \frac{1}{|i|^{K/2}} \frac{1}{|l|^{K/2}}. \tag{76}$$

This means that in all calculations compared to case $i = 0$, we get an extra factor $\frac{1}{|i|^{K/2}}$. We divide investigation to "small" and "large" values of $|i|$. Let investigate first the case $|i| \gtrsim |k|^\kappa$, where $0 < \kappa < 1/2$.

$$\begin{aligned}
&\sum_{|i|=|k|^\kappa}^{\infty} \left| \int_{\mathbb{R}^2} 1_{R_{i,a}} w_r f \gamma_{ab\theta} dx \right| \\
\lesssim &\sum_{|i|=|k|^\kappa}^{\infty} \sum_{|l|=0}^{\infty} a^{3/2} \frac{a^{-3/4}}{1+|i+l|^K} \\
\lesssim &\sum_{|i|=|k|^\kappa}^{\infty} \left(\frac{a^{3/4}}{|i|^K} + \sum_{l=1}^{\infty} a^{3/2} \frac{a^{-3/4}}{1+|i+l|^K} \right) \\
\lesssim &\sum_{|i|=|k|^\kappa}^{\infty} \left(\frac{a^{3/4}}{|i|^K} + \frac{1}{|i|^{K/2}} \sum_{l=1}^{\infty} a^{3/2} \frac{a^{-3/4}}{|l|^{K/2}} \right) \\
\lesssim &\sum_{|i|=|k|^\kappa}^{\infty} \frac{a^{3/4}}{|i|^{K/2}} \\
\lesssim &\frac{a^{3/4}}{|k|^{(K/2-1)\kappa}}.
\end{aligned} \tag{77}$$

With a fixed κ we just have to choose K large enough.

Next consider values $i \lesssim |k|^\kappa$. We note that "twisting" is not necessary if we do not let $|l|$ grow too large, i.e. if region R_I do not intersect with S. In the case $i = 0$ the bound $|l| < l_{max}$ was used only when estimating I_{22} to take care of large values of $|l|$. The value for l_{max} was chosen so that "twisting error I_{21} does not grow too large. If $R_I \cap S = \emptyset$, we do not need any twisting, so generally in this case we could consider some other (smaller) bound for $|l|$. However, it turns out that $|l| \lesssim |k|^{1-\varepsilon/2}$ will provide $R_I \cap S = \emptyset$ when $0 < |i| \lesssim |k|^\kappa$. We check this now.

For every region $R_{i,a,l}$

$$y \in R_{i,a,l} \quad \Rightarrow \quad \begin{cases} |y_1| \lesssim |i|\, a^{1/2} + |l|\, d \lesssim (|i| + |l/k|) a^{1/2} \\ |y_2| \approx |ik|\, a \end{cases} \tag{78}$$

and

$$y \in R_{i,a,l} \quad \Rightarrow \quad |g(y_1)| \lesssim |y_1|^2 \lesssim (|i| + |l/k|)^2 a. \tag{79}$$

Therefore

$$|ik|\, a \gtrsim (|i| + |l/k|)^2 a \quad \Rightarrow \quad R_{i,a,l} \cap S = \emptyset. \tag{80}$$

Now we just have to check that the left hand side of (80) holds for all $|i| \lesssim |k|^\kappa$. This can be done simply by noticing that

$$(|i| + |l/k|)^2 \lesssim (|i| + |k|^{-\varepsilon/2})^2 \lesssim |i|^2 \lesssim |k|^{2\kappa}, \tag{81}$$

and $|k| \gtrsim |k|^{2\kappa}$ we get

$$|ik|\, a \gtrsim |k|^{2\kappa}\, a \gtrsim (|i| + |l/k|)^2 a. \tag{82}$$

We have now shown that

$$0 < |i| \lesssim |k|^\kappa,\ |l| \le l_{max} \quad \Rightarrow \quad R_{i,a,l} \cap S = \emptyset \tag{83}$$

which means that for all slices $R_{i,a}$ we can use exactly same techniques (in case $i \neq 0$ twisting step can be omitted). By remembering that in the case $i = 0$ we used in all non-zero estimates the rapid decay property of γ, it's clear that with general i we get again therefore an extra factor $\frac{1}{|i|^{K/2}}$. Detailed steps to create $\frac{1}{|i|^{K/2}}$ are the same as in case $|i| \gtrsim |k|^\kappa$ and we omit them. The estimate for with small values of $|i|$ is therefore

$$\begin{aligned} &\sum_{|i|=0}^{|k|^\kappa} \left| \int_{\mathbb{R}^2} \chi_{R_{i,a}} w_r f \gamma_{ab\theta} dx \right| \\ \le &\sum_{|i|=0}^{|k|^\kappa} \frac{a^{3/4}}{|k|^{3+\varepsilon}\, |i|^{-K/2}} \\ \le &\frac{a^{3/4}}{|k|^{3+\varepsilon}}, \end{aligned} \tag{84}$$

and the final estimate

$$\left| \int_{\mathbb{R}^2} f(x) \gamma_{ab\theta}(x) dx \right| \lesssim \frac{a^{3/4}}{|k|^{3+\varepsilon}} \tag{85}$$

comes now by putting (77) and (84) together. ■

The previous theorem considered only the case when $b \in S$. However, this is hardly ever the case, if we do not consider a continuous transform but coefficients of a discrete curvelet transform. If γ would have compact support, then this would not be so restrictive since we could do exactly the same calculations for b such that $supp(\gamma_{ab\theta}) \cap S \neq \emptyset$. Unfortunately, this is not possible since γ has compact support in Fourier domain. All calculations in the proof of theorem still work same way

for some finite number of b that that are closest to point $p \in S$. However, since $supp(\gamma_{ab\theta}) \cap S \neq \emptyset$ for all b, we are interested also how $\left|\int_{\mathbb{R}^2} f(x)\gamma_{ab\theta}(x)dx\right|$ decay as function of b when the distance between b and S grows. The next theorem give some estimate for this.

Theorem 15 *Let $f \in F_{N,n}$, $N \geq n$, $p \in S$ be point that minimizes $L = |D_a R_\theta(b-p)|$ and $\theta' \approx ka^{1/2}$ be the angle between major axis of $\gamma_{ab\theta}$ and tangent of S at point p. Then for any $K > 0$ and $0 < \varepsilon < 2$*

$$\begin{aligned} &\left|\int_{\mathbb{R}^2} f(x)\gamma_{ab\theta}(x)dx\right| \\ &\lesssim \begin{cases} \max\left\{a^{3/4+N}, \frac{a^{3/4}}{L^K}\right\} & , \quad |\theta'| \lesssim a^{1/2} \\ \max\left\{a^{3/4+N/2}, \frac{a^{3/4}}{|k|^3 L^K}\right\} & , \quad n = 2,\ |\theta'| \gtrsim a^{1/2} \\ \max\left\{a^{3/4+N/2}, \frac{a^{3/4}}{|k|^{3+\varepsilon} L^K}\right\} & , \quad n \geq 3,\ |\theta'| \gtrsim a^{1/2} \end{cases} \end{aligned} \tag{86}$$

Proof. We divide the investigation in two parts: $|k|^{1-\varepsilon/2} \lesssim L$ and $|k|^{1-\varepsilon/2} \gtrsim L$.

The case $|k|^{1-\varepsilon/2} \lesssim L$ is quite simple to check. Let A be a rectangle with side lengths La and $La^{1/2}$, center in $D_a R_\theta b$ and orientation the same as with $\gamma_{ab\theta}$.

We can use similar tricks as in the proof of Theorem 14. Let us think the coordinate system where x_1 axis is parallel to major axis of $\gamma_{ab\theta}$, x_2 axis is parallel to minor axis of $\gamma_{ab\theta}$ and b is in origin. First we take slices from f in direction of x_2 axis and construct Taylor polynomials $P_{x_1}(x_2)$ of one variable for those slices. Then we define a polynomial $P : \mathbb{R}^2 \to \mathbb{R}$ by $P(x) := P_{x_1}(x_2)$. We have

$$\begin{aligned} &\left|\int_{\mathbb{R}^2} f(x)\gamma_{ab\theta}(x)dx\right| \\ &\leq \left|\int_{\mathbb{R}^2} (f(x) - P(x))\gamma_{ab\theta}(x)dx\right| + \left|\int_{\mathbb{R}^2} P(x)\gamma_{ab\theta}(x)dx\right| \\ &\leq \left|\int_{\mathbb{R}^2} |f(x) - P(x)|\,|\gamma_{ab\theta}(x)|\,dx\right| \\ &= \left|\int_{\mathbb{R}^2\setminus A} |f(x) - P(x)|\,|\gamma_{ab\theta}(x)|\,dx\right| + \left|\int_{A} |f(x) - P(x)|\,|\gamma_{ab\theta}(x)|\,dx\right| \\ &= I_1 + I_2. \end{aligned} \tag{87}$$

Because of rapid decay of γ, by using very similar calculations as for I_{211} in the proof of Theorem 14, we get

$$I_1 \lesssim \frac{a^{3/4}}{L^{2K}} \lesssim \frac{a^{3/4}}{|k|^{(1-\varepsilon/2)K}\, L^K}, \tag{88}$$

that is clearly acceptable since K can be arbitrary large. Because $A \cap S = \emptyset$, the integral I_2 can be handled similarly to I_{212} in (66) by using Taylor's approximation

theorem. This gives

$$I_2 \lesssim a^{3/4+N}. \tag{89}$$

In calculations of I_1 and I_2, instead of parallelograms $R_{i,a,l}$ that were used in proof of Theorem 14, we use rectangulars that are oriented like $\gamma_{ab\theta}$ and take a double summation over both i and l, so that the whole domain $\mathbb{R}^2$ is handled at one time. The advantage of the use of rectangulars instead of parallelograms is that we get the factor a^N instead of $(a^{1/2}/|k|)^N$. By combining I_1 and I_2 it follows that

$$\left|\int_{\mathbb{R}^2} f(x)\gamma_{ab\theta}(x)dx\right| \lesssim \max\left\{a^{3/4+N}, \frac{a^{3/4}}{|k|^K L^K}\right\}, \quad \forall k^{1-\varepsilon/2} \lesssim L. \tag{90}$$

Now we turn to case $|k|^{1-\varepsilon/2} \gtrsim L$.

First we investigate that the smoothness of S implies that point p must belong to x_1 axis always if $|k| > k_0$, for some $k_0 < \infty$, and $|k|^{1-\varepsilon/2} \gtrsim L$.

Let assume that $p \approx (a^{1/2}i, aj) \notin x_1$-axis. With this notation, $|i|+|j| \approx L \lesssim |k|^{1-\varepsilon/2}$. Since in this coordinate system $\left|g'(a^{1/2}i)\right| \approx a^{1/2}|k|$, it follows that $\left|g(a^{1/2}i + a^{1/2}) - g(a^{1/2}i)\right| \gtrsim a|k|$. However, if u is the intersection point of S and x_1-axis, then is $|u_1 - p_1| \approx a^{1/2}j/|k|$. Therefore there exists k_0 s.t. for all $k > k_0$ holds $\|b-u\|_{\theta,a} < \|b-p\|_{\theta,a}$. This contradict the definition of that p should be closest point and therefore the assumption $p \notin x_1$-axis must be wrong.

By the same arguments that were used in the case $|k|^{1-\varepsilon/2} \lesssim L$, for those $|k| < k_0$ that satisfy $\|b-u\|_{\theta,a} > \|b-p\|_{\theta,a}$ we give estimate

$$\left|\int_{\mathbb{R}^2} f(x)\gamma_{ab\theta}(x)dx\right| \lesssim \max\left\{a^{3/4+N}, \frac{a^{3/4}}{L^{K'}}\right\}. \tag{91}$$

Now we can concentrate only to the situation where x_1 axis intersects with S in point p. The calculations for estimate follow the proof of Theorem 14. Without loss of generality, we can restrict $b \in R_{L,a}$. Essentially in the Theorem 14 was $p = b$. Now, since $L = |D_a R_\theta(b-p)|$, rapid decay of $\gamma_{ab\theta}$ give an extra factor L^{-K} to estimate (74) that restricted to slice $R_{0,a}$, i.e.,

$$\left|\int_{R_{0,a}} f\gamma_{ab\theta}dx\right| \lesssim \begin{cases} \frac{a^{3/4}}{|k|^3 L^K}, & n = 2 \\ \frac{a^{3/4}}{|k|^{3+\varepsilon} L^K}, & n \geq 3. \end{cases} \tag{92}$$

Let us investigate other slices $R_{i,a}$ a bit more careful. With similar calculations as in the proof of Theorem 14 we get

$$\begin{aligned}
&\sum_{|i|=1}^{\infty}\sum_{|l|=|k|^{1-\varepsilon/2}}^{\infty}\left|\int_{R_{i,a,l}} f\gamma_{ab\theta}dx\right| \\
&\leq \sum_{|i|=1}^{\infty} C\frac{a^{3/4}}{(1+|L-i|^{K})\,|k|^{(1-\varepsilon/2)K}} \\
&\leq C\frac{a^{3/4}}{|k|^{(1-\varepsilon/2)K}} \\
&\leq C\frac{a^{3/4}}{|k|^{K'}\,L^{K'}},
\end{aligned} \tag{93}$$

where K' can be arbitrary big since K can be chosen to be arbitrary large. The last step comes from the assumption $L \lesssim |k|^{1-\varepsilon/2}$. When $|l| < |k|^{1-\varepsilon/2}$ and $i \neq 0$, f can be assumed to be in $C^N(\cup_{|l|<|k|^{1-\varepsilon/2}} R_{i,a,l})$ and therefore again, with similar calculations as in proof of Theorem 14, we get

$$\begin{aligned}
&\sum_{|i|=0}^{\infty}\sum_{|l|=0}^{l_{max}}\left|\int_{R_{i,a,l}} f\gamma_{ab\theta}dx\right| \\
&\lesssim \sum_{|i|=1}^{\infty}\frac{a^{3/4+N/2}}{1+|L-i|^{K}} \\
&\lesssim a^{3/4+N/2}.
\end{aligned} \tag{94}$$

The final estimate comes noticing that from estimates (90), (92), (91), (93) and (94) the estimates (92) and (94) dominate over other estimates. ■

3.3 Convergence rate of non-linear approximation

It has been proved that the following theorem holds for curvelets, contourlets and shearlets [5, 11, 15].

Theorem 16 *Let $f_{M,B}$ be M-term non-linear approximation of f by using curvelets, contourlets or shearlets. If $f \in F_{2,2}$, then $\|f - f_{M,B}\|_2^2 \leq \mathcal{O}(M^{-2}(\log_2(M))^3)$.*

The next theorem states that by assuming more regularity, we can eliminate the factor $(\log_2(M))^3$. We note that Theorem 17 and the proof are written for curvelets but they would be probably similar for contourlets and shearlets (with compact support in Fourier domain). For compactly supported contourlets and shearlets some of techniques used in the proof may not work.

Theorem 17 *Let $f_{M,C}$ be M-term non-linear approximation of f by using curvelets. If $f \in F_{5,3}$, then $\|f - f_{M,C}\|_2^2 \leq \mathcal{O}(M^{-2})$.*

Proof. Slightly different discretizations and notations may be used. We choose $a = 2^{-j}$ and $\theta' = k2^{-j/2}$.

Estimate (86) gets now the form

$$\left| \int_{\mathbb{R}^2} f \gamma_{ab\theta} \right| \lesssim \max \left\{ 2^{(-5/2-3/4)j}, 2^{-3j/4} \left|k\right|^{-3-\varepsilon} L^{-K} \right\}. \tag{95}$$

We will look how many of the coefficients $\int_{\mathbb{R}} f \gamma_{ab\theta}$ have magnitude over ϵ, i.e., we demand that

$$\epsilon < \left| \int_{\mathbb{R}^2} f \gamma_{ab\theta} \right|. \tag{96}$$

We will investigate separately "distant domain", "smooth domain" and "rough domain" curvelets.

First we look at the distant domain. Let us take a ball B that has diameter about double the size of diameter of support the of f. Let us assume a while that boundary of this ball would be a discontinuity curve. In that case we can use for all $b \notin B$ the estimate $\left| \int_{\mathbb{R}^2} f \gamma_{ab\theta} \right| \lesssim 2^{-3j/4} \left|k\right|^{-3-\varepsilon} L^{-K}$, without any problem. By applying this to (96) we can calculate the number of remaining "distant domain coefficients", denoted by M_d. Calculations are exactly similar to those we will do later for M_r, the number of remaining rough domain coefficients, so we omit the details here. That way calculating the estimate will be

$$M_d \lesssim \epsilon^{-2/3}. \tag{97}$$

Next we move on to smooth domains. On every level j there exists $N_j \lesssim 2^j 2^{j/2} 2^{j/2} \approx 2^{2j}$ curvelets that are centered inside ball B. Curvelets on "smooth domain" are those that are centered so far away from discontinuity curve S that

$$2^{(-5/2-3/4)j} \gtrsim 2^{-3j/4} \left|k\right|^{-3-\varepsilon} L^{-K}.$$

In that case $\left| \int_{\mathbb{R}^2} f \gamma_{ab\theta} \right| \lesssim 2^{-13j/4}$. Therefore (96) gives a bound for scale $j \leq \log_2 \left(\epsilon^{-4/13} / C \right)$. With higher scales (96) can not be true. This means that the total number $M_s(\epsilon)$ of smooth domain coefficients that obey (96) is bounded by

$$M_s(\epsilon) \lesssim \sum_{j=1}^{\log_2\left(\epsilon^{-4/13}/C\right)} 2^{2j} \lesssim \epsilon^{-8/13} \lesssim \epsilon^{-2/3}. \tag{98}$$

Finally, we move on to rough domain. Here curvelets are so close to the discontinuity curve that

$$2^{(-5/2-3/4)j} \lesssim 2^{-3j/4} \left|k\right|^{-3-\varepsilon} L^{-K}.$$

In this case

$$\epsilon < \left| \int_{\mathbb{R}^2} f \gamma_{ab\theta} \right| \lesssim 2^{-3j/4} \left|k\right|^{-3-\varepsilon} L^{-K}. \tag{99}$$

From (99) we get a bound L_m for L:

$$L_m \approx \epsilon^{-1/K} 2^{-3j/(4K)} \left|k\right|^{(-3-\varepsilon)/K}. \tag{100}$$

By substituting $L = 1$ into (99), we get a similar bound k_m for $|k|$,

$$k_m \approx \epsilon^{-1/(3+\varepsilon)} 2^{-3j/(4(3+\varepsilon))}, \tag{101}$$

and finally by substituting $L = 1$, $k = 1$ we get bound j_m for j:

$$j_m := \log_2 \left(\epsilon^{-4/3} / C \right). \tag{102}$$

Since the discontinuity curve S has finite length, there exists about $1/(a/\left|\sin(\theta)\right|) \lesssim 1/(2^{-j}/(\left|k\right| 2^{-j/2})) \approx 2^{j/2} \left|k\right|$ basis functions $\gamma_{ab\theta}$ s.t. angle between the major axis of $\gamma_{ab\theta}$ and tangent of S is about $\left|k\right| 2^{-j/2}$ and $\left\|b-p\right\|_{a,\theta} \approx 1$. Moreover, with fixed p and k, the number of parameters b that satisfy $\left\|b-p\right\|_{a,\theta} \approx L$ is clearly bounded by L^2. Therefore, the upper bound for total number of essentially similarly oriented basis functions is about $\left|k\right| 2^{-j/2} L^2$. This bound is of course all but sharp, but it does not matter because K in (95) can be arbitrary big, as we will see. Notice that if Fourier domain window functions in definition of curvelets would not be C^∞ we would have to be much more careful with this bound. We are now ready to derive a bound for M_r, the number of rough domain curvelets that satisfy (96).

$$\begin{aligned}
M_r &\lesssim \sum_{j=1}^{j_m} \sum_{k=1}^{k_m} 2^{j/2} \left|k\right| \sum_{L=1}^{L_m} L^2 \\
&\approx \sum_{j=1}^{j_m} \sum_{k=1}^{k_m} 2^{j/2} \left|k\right| L_m^3 \\
&\approx \epsilon^{-3/K} \sum_{j=1}^{j_m} 2^{j(1/2-9/4K)} \sum_{k=1}^{k_m} \left|k\right|^{3(-3-\varepsilon)/K+1} \\
&\approx \epsilon^{-3/K} \sum_{j=1}^{j_m} 2^{j(1/2-9/4K)} k_m^{3(-3-\varepsilon)/K+2} \\
&\approx \epsilon^{-2/(3+\varepsilon)} \sum_{j=1}^{j_m} 2^{j(1/2-6/(4(3+\varepsilon)))} \\
&\approx \epsilon^{-2/(3+\varepsilon)} \epsilon^{-4/3(1/2-6/(4(3+\varepsilon)))} \\
&\approx \epsilon^{-2/3}.
\end{aligned} \tag{103}$$

The last step follows just by simplifying the exponent and using standard summation formula. Notice that above the exponent $(1/2 - 6/(4(3+\varepsilon)))$ would be zero if $\varepsilon = 0$. In that case, taking sum over range of j we would get log factor to final estimate, and this would lead to the approximation error $O(M^{-2}(\log(M))^3)$.

The total number M of coefficients above threshold ϵ (as a function of ϵ) is

$$M = M_d + M_s + M_r \lesssim \epsilon^{-2/3}, \tag{104}$$

and it follows that ϵ (as function of M) has the bound

$$\epsilon(M) \lesssim M^{-3/2}. \tag{105}$$

Now, because the functions $\gamma_{ab\theta}$ form a tight frame, we can write an error estimate for a non-linear approximation where only M terms that obey (96) are kept. Let p be permutation operator that rearranges a sequence to descending order. Then

$$\begin{aligned} \|f - f_{M,C}\|_2^2 &= \sum_{m=M}^{\infty} p(\{|\langle f, \gamma_{ab\theta}\rangle^2|\}_{ab\theta}) \\ &\leq \sum_{m=M}^{\infty} (\epsilon(m))^2 \\ &\lesssim \sum_{m=M}^{\infty} (m^{-3/2})^2 \\ &\approx \sum_{m=M}^{\infty} m^{-3} \\ &\approx M^{-2}. \end{aligned} \tag{106}$$

■

4 Characterization of Hölder regularities

4.1 Definitions of Hölder regularities

Hölder regularity of functions of two variables is defined as follows.

Definition 18 *Let $\alpha > 0$ and $\alpha \notin \mathbb{N}$. A function $f\colon \mathbb{R}^2 \to \mathbb{R}$ is said to be* pointwise Hölder regular with exponent α at u, *denoted by $f \in C^\alpha(u)$, if there exists a polynomial P_u of degree less than α and a constant C_u such that for all x in a neighborhood of u*

$$|f(x) - P_u(x-u)| \leq C_u \|x-u\|^\alpha. \tag{107}$$

Let Ω be an open subset of $\mathbb{R}^2$. If (107) holds for all $x, u \in \Omega$ with C_u being a uniform constant independent of u, then we say that f is uniformly Hölder regular with exponent α on Ω *or shortly, $f \in C^\alpha(\Omega)$.*

The *uniform* and *pointwise Hölder exponents* of f on Ω and at u are then defined as

$$\alpha_l(\Omega) := \sup\{\alpha \colon f \in C^\alpha(\Omega)\}$$

and

$$\alpha_p(u) := \sup\{\alpha \colon f \in C^\alpha(u)\}.$$

Following [37], we define the local Hölder exponent as follows.

Definition 19 *Let $(I_n)_{n\in\mathbb{N}}$ be a family of nested open sets in $\mathbb{R}^2$, i.e. $I_{n+1} \subset I_n$, with intersection $\cap_n I_n = \{u\}$.* The local Hölder exponent *of a function f at u, denoted by $\alpha_l(u)$, is*

$$\alpha_l(u) = \lim_{n\to\infty} \alpha_l(I_n).$$

In many situations, local and pointwise Hölder exponents coincide, e.g., if $f(x) = |x|^\beta$ then $\alpha_p(0) = \alpha_l(0) = \beta$. However, local Hölder exponents $\alpha_l(u)$ is also sensitive to oscillating behavior of f near the point u. A simple example is $f(x) = |x|^\beta \sin(1/|x|^\gamma)$ for which $\alpha_p(0) = \beta$ but $\alpha_l(0) = \frac{\beta}{1+\gamma}$, i.e., α_l is influenced by the wild oscillatory behavior of f near 0. For more on the nature of α_l and α_p see [37, 21]

Finally, we give a definition of directional regularity.

Definition 20 *Let $v \in \mathbb{R}^2$ be a fixed unit vector and $u \in \mathbb{R}^2$. A function $f \colon \mathbb{R}^2 \to \mathbb{R}$ is* pointwise Hölder regular with exponent α at u in the direction v, *denoted by $f \in C^\alpha(u; v)$, if there exist a constant $C_{u,v}$ and a polynomial $P_{u,v}$ of degree less than α such that*

$$|f(u + \lambda v) - P_{u,v}(\lambda)| \le C_{u,v}|\lambda|^\alpha$$

holds for all λ in a neighborhood of $0 \in \mathbb{R}$.

If one can choose $C_{u,v}$ so that it is independent of u for all $u \in \Omega \subseteq \mathbb{R}^2$ and the inequality holds for all $\lambda \in \mathbb{R}$ such that $u + \lambda v \in \Omega$, then we say that f is uniformly Hölder regular with exponent α on Ω in direction v *or $f \in C^\alpha(\Omega; v)$.*

A simple example of functions with varying directional Hölder regularities at the origin is $f(x) = r^{\alpha(\theta)}$, where (r, θ) are the polar coordinates at x and $\alpha(\theta)$ is any continuous positive function. It is easy to see that $f \in C^{\alpha(\theta)}(0; (\cos\theta, \sin\theta))$. In applications, the problem of determining directional smoothness arises naturally in medical imaging such as the X-ray pictures of bones [1]. Whether or not it is possible to prescribe arbitrary directional smoothness at different points is still open, see [20].

4.2 Characterizations of Hölder requlariities by CCT and Hart Smith transform

4.2.1 Characterization of uniform regularity by CCT and Hart Smith transform

Let us first note here that in the proofs of Theorems 21 and 23 we do not need vanishing moments in more than one direction (for example in the direction of x_1-axis). In fact,

$$\int_{\mathbb{R}^2} \phi_{ab\theta}(x) P_u(x-u)dx = \int_{\mathbb{R}^2} \phi_{a00}(x) P_u(R_\theta(x-u)+b)dx = 0,$$

because $P_u(R_\theta(x-u)+b)$ is a polynomial.

The following theorem gives a necessary condition for Hölder regularity in terms of decay of Hart Smith or curvelet transforms.

Theorem 21 *If a bounded function $f \in C^\alpha(\mathbb{R}^2)$, then there exist a constant C and a fixed coarsest scale a_0 for which*

$$|\langle \phi_{ab\theta}, f\rangle| \leq C a^{\alpha+\frac{3}{4}}$$

for all $0 < a < a_0$, $b \in \mathbb{R}^2$, and $\theta \in [0, 2\pi)$.

Proof. Without loss of generality, we can assume that $b = 0$ and $\theta = 0$. The general case follows by simple translation and rotation of f because uniform regularity is invariant under those operations.

We first recall that uniform regularity of f means that there exist a constant C independent of $u \in \mathbb{R}^2$ and, for each u, a polynomial P_u of degree less than α such that

$$|f(x) - P_u(x-u)| \leq C\|x-u\|^\alpha$$

for all $x \in \mathbb{R}^2$. Therefore, for each $x_2 \in \mathbb{R}$, there exists a polynomial $P_{(0,x_2)}$ such that, for all $x_1 \in \mathbb{R}$,

$$|f(x_1,x_2) - P_{(0,x_2)}(x_1,0)| \leq C\|(x_1,x_2)-(0,x_2)\|^\alpha = C|x_1|^\alpha. \tag{108}$$

Also, since ϕ_{a00} has rapid decay, the integral $\int_{\mathbb{R}^2} \left|P_{(0,x_2)}(x_1,0)\phi_{a00}(x)\right| dx$ is finite which allows us to apply Fubini's theorem and write

$$\begin{aligned}\int_{\mathbb{R}^2} P_{(0,x_2)}(x_1,0)\phi_{a00}(x)\,dx &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} P_{(0,x_2)}(x_1,0)\phi_{a00}(x_1,x_2)dx_1\right) dx_2 \\ &= \int_{\mathbb{R}} 0\,dx_2 = 0,\end{aligned} \tag{109}$$

where we have used the assumption that ϕ_{a00} has directional vanishing moments of any order along the x_2-axis for a sufficiently small. By using (108), (109) and (21) we get

$$\begin{aligned}
|\langle \phi_{ab\theta}, f\rangle| &= \left| \int_{\mathbb{R}^2} \big(f(x) - P_{(0,x_2)}(x_1,0)\big)\, \phi_{a00}(x) dx \right| \\
&\leq \int_{\mathbb{R}^2} \big|f(x) - P_{(0,x_2)}(x_1,0)\big|\, |\phi_{a00}(x)|\, dx \\
&\lesssim \int_{\mathbb{R}^2} |x_1|^\alpha\, |\phi_{a00}(x)|\, dx \\
&\lesssim \int_{\mathbb{R}^2} |x_1|^\alpha \left| \frac{a^{-3/4}}{1 + \|x\|_{a,0}^{2N}} . \right| dx \\
&\lesssim \int_{\mathbb{R}^2} |a y_1|^\alpha \left| \frac{a^{3/4}}{1 + \|y\|^{2N}} . \right| dy \\
&\lesssim a^{\alpha + 3/4}.
\end{aligned}$$

■

A sufficient condition for a function f to be C^α is given in the next theorem. Unfortunately, the condition here is not the same to the necessary condition presented above, due to the effect of parabolic scaling.

Theorem 22 *Let $f \in L^2(\mathbb{R}^2)$ and $\alpha > 0$ a non-integer. If there is a constant $C < \infty$ such that*

$$|\langle \phi_{ab\theta}, f\rangle| \leq C a^{\alpha + \frac{5}{4}},$$

for all $0 < a < a_0$, $b \in \mathbb{R}^2$, and $\theta \in [0, 2\pi)$, then $f \in C^\alpha(\mathbb{R}^2)$.

We omit the proof of this theorem since it is essentially the same as that for the pointwise case, see Theorem 24.

4.2.2 Characterization of pointwise regularity by CCT and Hart Smith transform

Pointwise regularity estimates are harder to obtain than those for uniform regularity. Necessary conditions and sufficient conditions derived here will differ even more than in the uniform case.

Theorem 23 *If a bounded function $f \in C^\alpha(u)$, then there exists $C < \infty$ such that*

$$|\langle \phi_{ab\theta}, f\rangle| \leq C a^{\frac{\alpha}{2} + \frac{3}{4}} \left(1 + \left\| \frac{b-u}{a^{1/2}} \right\|^\alpha \right) \tag{110}$$

for all $0 < a < a_0$, $b \in \mathbb{R}^2$, and $\theta \in [0, 2\pi)$.

Proof. Because f is bounded, the polynomial approximation property (107) holds in all $\mathbb{R}^2$ although by definition it holds only in some neighborhood of point u. Therefore the proof is similar to that for uniform regularity, but we obviously do not have varying polynomials for different x_2. We do have

$$\begin{aligned}
|\langle \phi_{ab\theta}, f\rangle| &\leq \int_{\mathbb{R}^2} |\phi_{ab\theta}(x)|\,|f(x) - P_u(x-u)|\,dx \\
&\lesssim a^{-\frac{3}{4}} \int_{\mathbb{R}^2} \frac{\|x-u\|^\alpha}{1+\|x-b\|_{a,\theta}^{2N}}\,dx \\
&\approx a^{-\frac{3}{4}} \int_{\mathbb{R}^2} \frac{\|x-u\|^\alpha}{1+\|\mathrm{D}_{1/a}\mathrm{R}_{-\theta}(x-b)\|^{2N}}\,dx \\
&\approx a^{-\frac{3}{4}+\frac{3}{2}} \int_{\mathbb{R}^2} \frac{\|\mathrm{R}_\theta \mathrm{D}_a y + b - u\|^\alpha}{1+\|y\|^{2N}}\,dy \\
&\lesssim C a^{-\frac{3}{4}+\frac{3}{2}} \int_{\mathbb{R}^2} \frac{\|\mathrm{R}_\theta \mathrm{D}_a y\|^\alpha + \|b-u\|^\alpha}{1+\|y\|^{2N}}\,dy \\
&\lesssim C a^{\frac{3}{4}+\frac{\alpha}{2}} \left(1 + \left\| \frac{b-u}{a^{1/2}} \right\|^\alpha\right),
\end{aligned}$$

since we can choose N large enough so that the last integral is finite. We have also used the fact that $\mathrm{R}_\theta \mathrm{D}_a$ is a bounded linear operator with norm $\|\mathrm{R}_\theta \mathrm{D}_a\| = a^{1/2}$. ■

Theorem 24 *Let $f \in L^2(\mathbb{R}^2)$ and α be a non-integer positive number. If there exist $C < \infty$ and $\alpha' < 2\alpha$ such that*

$$|\langle \phi_{ab\theta}, f\rangle| \leq C a^{\alpha+\frac{5}{4}} \left(1 + \left\| \frac{b-u}{a^{1/2}} \right\|^{\alpha'}\right), \tag{111}$$

for all $0 < a < a_0$, $b \in \mathbb{R}^2$, and $\theta \in [0, 2\pi)$, then $f \in C^\alpha(u)$.

Proof. First we divide $f \in L^2(\mathbb{R})$ to low and high frequency parts f_L and f_H. As discussed before, regularity of the function f depends only on regularity of high frequency part f_H, and therefore it's enough to calculate estimates only for f_H. It is clear that, for a small enough, $\left\langle \hat{\phi}_{ab\theta}, \hat{f}_L \right\rangle = 0$ as the frequency support of $\phi_{ab\theta}$ is moving farther away from the origin. Because of that and by Plancherel's formula, we have

$$\langle \phi_{ab\theta}, f\rangle = \langle \phi_{ab\theta}, f_L\rangle + \langle \phi_{ab\theta}, f_H\rangle = \langle \phi_{ab\theta}, f_H\rangle .$$

Therefore the assumption (111) gives

$$|\langle \phi_{ab\theta}, f_H\rangle| \lesssim a^{\alpha+\frac{5}{4}} \left(1 + \left\| \frac{b-u}{a^{1/2}} \right\|^{\alpha'}\right). \tag{112}$$

Without loss of generality, we can choose $a_0 = 1$ in (13). We use the notation $x = u + \lambda v$, where $\|v\| = 1$. Here we can set $u = 0$, the general case follows by simple translation. The reconstruction formula (13) can be rewritten

$$f_H(x) = \sum_{j=-\infty}^{-1} \Delta_j(x), \tag{113}$$

where

$$\Delta_j(x) := \int_{2^j}^{2^{j+1}} \int_0^{2\pi} \int_{\mathbb{R}^2} \langle \phi_{ab\theta}, f_H \rangle \, \phi^\sharp_{ab\theta}(x) \, db \, d\theta \, \frac{da}{a^3}.$$

We try to approximate f_H by polynomial

$$P_0(\lambda v) := \sum_{k=0}^{\lfloor \alpha \rfloor} \frac{\lambda^k}{k!} \sum_{j=-\infty}^{-1} (v \cdot \nabla)^k \Delta_j(0).$$

Next we investigate how fast the terms $(\lambda v \cdot \nabla)^k \Delta_j(\lambda v)$ go to zero when $j \to -\infty$. First, by using assumptions of the theorem, implication (112) and the decay estimate (21), we can derive

$$\begin{aligned}
&\left|\partial_1^{k-l} \partial_2^l \Delta_j(x)\right| \\
&= \left| \int_{2^j}^{2^{j+1}} \int_0^{2\pi} \int_{\mathbb{R}^2} \langle \phi_{ab\theta}, f_H \rangle \, \partial_1^{k-l} \partial_2^l \phi^\sharp_{ab\theta}(x) \, db \, d\theta \, \frac{da}{a^3} \right| \\
&\lesssim \int_{2^j}^{2^{j+1}} \int_0^{2\pi} \int_{\mathbb{R}^2} a^{\alpha+5/4} \left(1 + \left\| \frac{b}{a^{1/2}} \right\|^{\alpha'}\right) \frac{a^{-3/4-k}}{1 + \|x-b\|_{a,\theta}^{2N}} \, db \, d\theta \, \frac{da}{a^3} \\
&\approx \int_{2^j}^{2^{j+1}} \int_0^{2\pi} \int_{\mathbb{R}^2} \left(1 + \left\| \frac{x - \mathrm{R}_\theta \mathrm{D}_a y}{a^{1/2}} \right\|^{\alpha'}\right) \frac{a^{\alpha+5/4-k+3/2-3/4}}{1 + \|y\|^{2N}} \, dy \, d\theta \, \frac{da}{a^3} \\
&\lesssim \int_{2^j}^{2^{j+1}} \int_0^{2\pi} \int_{\mathbb{R}^2} \frac{a^{\alpha-k+2} \left(1 + a^{-\alpha'/2} \left(\|x\|^{\alpha'} + \|\mathrm{R}_\theta \mathrm{D}_a y\|^{\alpha'} \right)\right)}{1 + \|y\|^{2N}} \, dy \, d\theta \, \frac{da}{a^3} \\
&\lesssim \int_{2^j}^{2^{j+1}} \int_0^{2\pi} \int_{\mathbb{R}^2} \frac{a^{\alpha-k+2} \left(1 + \|y\|^{\alpha'} + a^{-\alpha'/2} \|x\|^{\alpha'}\right)}{1 + \|y\|^{2N}} \, dy \, d\theta \, \frac{da}{a^3} \\
&\lesssim \int_{2^j}^{2^{j+1}} a^{\alpha-k-1} \left(C + a^{-\alpha'/2} \|x\|^{\alpha'} \right) da \\
&\lesssim 2^{j(\alpha-k)} + C 2^{j(\alpha-k-\alpha'/2)} \|x\|^{\alpha'}.
\end{aligned} \tag{114}$$

Therefore

$$\begin{aligned}\left|(v\cdot\nabla)^k\Delta_j(\lambda v)\right| &= \left|\left(v_1\partial_1+v_2\partial_2\right)^k\Delta_j(\lambda v)\right| \\ &= \left|\sum_{l=0}^{k}\binom{k}{l}v_1^{k-l}v_2^l\partial_1^{k-l}\partial_2^l\Delta_j(\lambda v)\right| \\ &\lesssim \sum_{l=0}^{k}\binom{k}{l}\left|v_1^{k-l}v_2^l\right|\left(2^{j(\alpha-k)}+2^{j(\alpha-k-\alpha'/2)}|\lambda|^{\alpha'}\right) \\ &\lesssim \left(2^{j(\alpha-k)}+2^{j(\alpha-k-\alpha'/2)}|\lambda|^{\alpha'}\right). \end{aligned} \tag{115}$$

Then, by the triangle inequality,

$$\begin{aligned}|f_H(\lambda v)-P_0(\lambda v)| &= \left|\sum_{j=-\infty}^{-1}\Delta_j(\lambda v)-\sum_{k=0}^{\lfloor\alpha\rfloor}\frac{\lambda^k}{k!}\sum_{j=-\infty}^{-1}(v\cdot\nabla)^k\Delta_j(0)\right| \\ &\leq \sum_{j=-\infty}^{-1}\left|\Delta_j(\lambda v)-\sum_{k=0}^{\lfloor\alpha\rfloor}\frac{\lambda^k}{k!}(v\cdot\nabla)^k\Delta_j(0)\right|.\end{aligned}$$

We investigate now coarse and fine scales separately and therefore we choose J such that $2^{J-1}\leq|\lambda|\leq 2^J$. It's essential to notice that our generic constant C can remain independent of J in all calculations. By noticing that each summand is the absolute error of the approximation of $\Delta_j(\lambda v)$ by its Taylor polynomial of degree $\lfloor\alpha\rfloor$, we get, for coarse scales,

$$\begin{aligned}&\sum_{j=J}^{-1}\left|\Delta_j(\lambda v)-\sum_{k=0}^{\lfloor\alpha\rfloor}\frac{\lambda^k}{k!}(v\cdot\nabla)^k\Delta_j(0)\right| \\ &\quad\leq \sum_{j=J}^{-1}\frac{|\lambda|^{\lfloor\alpha\rfloor+1}}{(\lfloor\alpha\rfloor+1)!}\sup_{h\in[0,\lambda]}\left|(v\cdot\nabla)^{\lfloor\alpha\rfloor+1}\Delta_j(hv)\right| \\ &\quad\lesssim \sum_{j=J}^{-1}|\lambda|^{\lfloor\alpha\rfloor+1}\left(2^{j(\alpha-\lfloor\alpha\rfloor-1)}+2^{j(\alpha-\lfloor\alpha\rfloor-1-\alpha'/2)}|\lambda|^{\alpha'}\right) \\ &\quad\lesssim |\lambda|^{\lfloor\alpha\rfloor+1}\left(\left(2^{J+1}\right)^{(\alpha-\lfloor\alpha\rfloor-1)}+\left(2^{J+1}\right)^{(\alpha-\lfloor\alpha\rfloor-1-\alpha'/2)}|\lambda|^{\alpha'}\right) \\ &\quad\lesssim |\lambda|^{\lfloor\alpha\rfloor+1}\left(|\lambda|^{(\alpha-\lfloor\alpha\rfloor-1)}+|\lambda|^{(\alpha-\lfloor\alpha\rfloor-1-\alpha'/2)}|\lambda|^{\alpha'}\right) \\ &\quad\lesssim |\lambda|^{\alpha},\end{aligned}$$

where the estimate (115) was used in the second inequality. At fine scales we use directly the estimate (115) and get

$$\begin{aligned}
&\sum_{j=-\infty}^{J-1} \left| \Delta_j(\lambda v) - \sum_{k=0}^{\lfloor \alpha \rfloor} \frac{\lambda^k}{k!} (v \cdot \nabla)^k \Delta_j(0) \right| \\
&\qquad \lesssim \sum_{j=-\infty}^{J-1} \left[\left(2^{j\alpha} + 2^{j(\alpha-\alpha'/2)} |\lambda|^{\alpha'} \right) + \sum_{k=0}^{\lfloor \alpha \rfloor} \frac{|\lambda|^k}{k!} 2^{j(\alpha-k)} \right] \\
&\qquad \lesssim |\lambda|^\alpha .
\end{aligned}$$

The assumption $\alpha' < 2\alpha$ was needed to make infinite summations to converge. Because $\|x\| = |\lambda|$, the theorem is proved. ■

4.2.3 Characterization of singularity lines by CCT and Hart Smith transform

In this section we study what kind of directional information the decay of the transforms can provide. Comparing to previous theorems, we now approve a weaker rate of decay of $|\langle \phi_{ab\theta}, f \rangle|$ for some angle θ_0. Loosely speaking, this angle can define direction where f has low regularity, for example 1-dimensional singularity. We first consider new sufficient conditions for regularity.

Theorem 25 *Let $f \in L^2(\mathbb{R}^2)$, $u \in \mathbb{R}^2$, and assume that $\alpha > 0$ is not an integer. If there exist $\alpha' < 2\alpha$, $\theta_0 \in [0, 2\pi]$, and $C < \infty$ such that*

$$|\langle \phi_{ab\theta}, f \rangle| \leq \begin{cases} Ca^{\alpha+\frac{5}{4}} \left(1 + \left\| \frac{b-u}{a^{1/2}} \right\|^{\alpha'} \right), & \text{if } |\theta - \theta_0| \gtrsim a^{1/2} \\ Ca^{\alpha+\frac{3}{4}} \left(1 + \left\| \frac{b-u}{a^{1/2}} \right\|^{\alpha'} \right), & \text{if } |\theta - \theta_0| \lesssim a^{1/2} \end{cases}$$

for all $0 < a < a_0$, $b \in \mathbb{R}^2$, and $\theta \in [0, 2\pi)$, then $f \in C^\alpha(u)$.

Proof. Let us denote $I_{\theta_0,a} := \theta_0 + Ca^{1/2}[-1, 1]$. The only difference from the proof of Theorem 24 is that, in the derivation of (114), we split the integral with respect to the angle θ into

$$\int_{[0,2\pi]} d\theta = \int_{[0,2\pi]\setminus I_{\theta_0,a}} d\theta + \int_{I_{\theta_0,a}} d\theta.$$

The first integral obviously gives the same estimate as in (114), while the second yields an extra factor of $2Ca^{1/2}$ instead of 2π. ■

We now give a simple example. Let $f(x_1, x_2) = g(x_1, x_2)|x_2|^\alpha$, where g is a bounded positive $C^\infty(\mathbb{R}^2)$ function. Then $f \in C^\alpha(\{(x_1, 0)|x_1 \in \mathbb{R}\}, v_0) \cap C^\infty(\mathbb{R}^2, v_{\pi/2})$. First

we recall that $\phi_{ab\theta}$ has vanishing moments of any order to all directions v such that $|\angle(v, v_\theta)| \gtrsim a^{1/2}$. Therefore it is clear that

$$|\langle \phi_{ab\theta}, f\rangle| \lesssim a^N \quad \text{whenever} \quad \left|\angle(v_{\pi/2}, v_\theta)\right| \gtrsim a^{1/2}$$

for any $N < \infty$. With similar methods as in the uniform regularity case, we can derive

$$|\langle \phi_{ab\theta}, f\rangle| \lesssim a^{\alpha+3/4} \quad \text{whenever} \quad \left|\angle(v_{\pi/2}, v_\theta)\right| \lesssim a.$$

For angles $a \lesssim \left|\angle(v_{\pi/2}, v_\theta)\right| \lesssim a^{1/2}$, the same estimate cannot be so easily obtained. The next theorem gives more precise necessary decay condition for special type of functions that includes our example. Essentially these functions might have aligned singularity lines. Because local Hölder exponent is used in the assumption of this theorem, we actually consider not only traditional singularities but also wild oscillatory behavior.

Theorem 26 *Let f be bounded with local Hölder exponent $\alpha \in (0, 1]$ at point u and $f \in C^{2\alpha+1+\varepsilon}(\mathbb{R}^2, v_{\theta_0})$ for some $\theta_0 \in [0, 2\pi)$ with any fixed $\varepsilon > 0$. Then there exist $\alpha' \in [\alpha - \varepsilon, \alpha]$ such that for $a > 0$ and $b \in \mathbb{R}^2$,*

$$|\langle \phi_{ab\theta}, f\rangle| \lesssim \begin{cases} a^{\alpha+\frac{5}{4}}, & \text{if } |\theta - \theta_0| \gtrsim a^{1/2}, \\ a^{\alpha'+\frac{3}{4}} \left(1 + \left\|\dfrac{b-u}{a}\right\|^{\alpha'}\right), & \text{if } |\theta - \theta_0| \lesssim a^{1/2}. \end{cases}$$

Proof. From Definition 19, for all $\varepsilon > 0$, there exists a ball $B(u, r_\varepsilon)$ such that $\alpha_l(B(u, r_\varepsilon)) \geq \alpha - \varepsilon$. This means that

$$|f(x) - f(y)| \leq C_\varepsilon \|x - y\|^{\alpha_l(B(u,r_\varepsilon))} \leq C_\varepsilon \|x - y\|^{\alpha-\varepsilon}, \quad \forall x, y \in B(u, r_\varepsilon)$$

For simplicity we assume again that $u = 0$ and $\theta_0 = 0$, the general case follows by simple translation and rotation. For angles θ such that $|\angle(v_0, v_\theta)| \leq Ca^{1/2}$ we can write that

$$\begin{aligned} |\langle \phi_{ab\theta}, f\rangle| &= \left| \int_{\mathbb{R}^2} (f(x) - f(0, x_2))\, \phi_{ab\theta}(x) dx \right| \\ &\lesssim \int_{\mathbb{R}^2} \left\|(x_1, 0)^T\right\|^{\alpha_l(B(0,r_\varepsilon))} \left(\frac{a^{-3/4}}{1 + \|x - b\|_{a,\theta}^{2N}}\right) dx \\ &\lesssim \int_{\mathbb{R}^2} \|x\|^{\alpha_l(B(0,r_\varepsilon))} \left(\frac{a^{-3/4}}{1 + \|x - b\|_{a,\theta}^{2N}}\right) dx \\ &\lesssim \int_{\mathbb{R}^2} \left((a\|y\|)^{\alpha_l(B(0,r_\varepsilon))} + \|b\|^{\alpha_l(B(0,r_\varepsilon))}\right) \left(\frac{a^{3/4}}{1 + \|y\|^{2N}}\right) dy \\ &\approx a^{\alpha_l(B(0,r_\varepsilon))+3/4} \left(1 + \left\|\frac{b}{a}\right\|^{\alpha_l(B(0,r_\varepsilon))}\right). \end{aligned}$$

So we choose $\alpha' = \alpha_l\left(B(0,r_\varepsilon)\right)$. We then turn to the investigation of angles θ such that $|\angle(v_0,v_\theta)| \gtrsim a^{1/2}$. Our method is similar as making the integration of $\langle \phi_{ab\theta}, f\rangle$ by taking line integrals to direction of x_2. This idea arises from the fact that f has the high uniform regularity in that direction. Since the regularity property that we will use now is invariant under translation, the estimate we will get here is the same for all b and we can restrict to the case $b = 0$. We start by taking the rectangle $A_a := [-a^{-c}, a^{-c}]^2$ for some $0 < c < 1/2$, to be determined later. We notice that $\mathrm{R}_\theta \mathrm{D}_a A_a$ is oriented similarly to the essential support of $\phi_{ab\theta}$ and $A_a \to \mathbb{R}^2$ and $\mathrm{R}_\theta \mathrm{D}_a A_a \to 0$ when $a \to 0$. We will also use here the notation $v(x) := x_1 v_\theta / |\cos(\theta)|$, i.e. $v(x)$ lies on major axis of $\mathrm{R}_\theta \mathrm{D}_a A_a$ and $v(x) - x$ is always parallel to x_2-axis.

$$\begin{aligned}
|\langle \phi_{ab\theta}, f\rangle| &= \left| \int_{\mathbb{R}^2} \big(f(x) - P_{v(x)}(x - v(x))\big)\, \phi_{a0\theta}(x) dx \right| \\
&\leq \left| \int_{\mathbb{R}^2 \setminus \mathrm{R}_\theta \mathrm{D}_a A_a} \big(f(x) - P_{v(x)}(x - v(x))\big)\, \phi_{a0\theta}(x) dx \right| \\
&\quad + \left| \int_{\mathrm{R}_\theta \mathrm{D}_a A_a} \big(f(x) - P_{v(x)}(x - v(x))\big)\, \phi_{a0\theta}(x) dx \right|.
\end{aligned}$$

For the first integral we can write

$$\begin{aligned}
&\left| \int_{\mathbb{R}^2 \setminus \mathrm{R}_\theta \mathrm{D}_a A_a} \big(f(x) - P_{v(x)}(x - v(x)\big)\, \phi_{a0\theta}(x) dx \right| \\
&\quad \lesssim a^{-\frac{3}{4}} \int_{\mathbb{R}^2 \setminus \mathrm{R}_\theta \mathrm{D}_a A_a} \frac{\big|f(x) - P_{v(x)}(x - v(x))\big|}{1 + \|x\|_{a,\theta}^{2N}}\, dx \\
&\quad \approx a^{\frac{3}{4}} \int_{\mathbb{R}^2 \setminus A_a} \frac{|f(\mathrm{R}_\theta \mathrm{D}_a y) - P_{y'}(\mathrm{R}_\theta \mathrm{D}_a y - y')|}{1 + \|y\|^{2N}} dy \leq C a^K,
\end{aligned}$$

where $y' = v(\mathrm{R}_\theta \mathrm{D}_a y)$ and K can be chosen arbitrarily large. That is because *i)* $1 + \|y\|^{2N}$ will clearly dominate the integration since it grows much faster than $f(\mathrm{R}_\theta \mathrm{D}_a y) - P_{y'}(\mathrm{R}_\theta \mathrm{D}_a y - y')$ when we choose N large enough (since f is bounded) and *ii)* $\|y\| \geq a^{-c}$, i.e., when c is first fixed, we just choose N large enough to make K as large as necessary.

The length l of the part of the line parallel to the x_2-axis lying inside the rectangle $\mathrm{R}_\theta \mathrm{D}_a A_a$ is at most $|l| \leq \sqrt{2} a^{1/2-c}$. Let us now assume that $f \in C^{\alpha_s}(\mathbb{R}^2, v_{\theta_0})$ i.e. for every $y \in \mathbb{R}^2$ there exists a polynomial P_y such that

$$|f(x) - P_y(x - y)| \leq C \|x - y\|^{\alpha_s}, \quad \text{when } (x - y) \parallel v_{\theta_0}$$

for all x in some neighborhood of y. Now $\mathrm{R}_\theta \mathrm{D}_a A_a \subset B(0,r)$ clearly for any fixed r when a is just small enough. Therefore, for $x \in \mathrm{R}_\theta \mathrm{D}_a A_a$,

$$\big|f(x) - P_{v(x)}(x - v(x))\big| \lesssim |l|^{\alpha_s} \lesssim a^{(1/2-c)\alpha_s}.$$

With these, we obtain an estimate for the second integral

$$\left|\int_{\mathrm{R}_\theta \mathrm{D}_a A_a} \big(f(x) - P_{v(x)}(x - v(x))\big)\, \phi_{ab\theta}(x) dx\right| \lesssim \int_{A_a} \frac{a^{\left(\frac{1}{2}-c\right)\alpha_s} a^{\frac{3}{4}}}{1 + \|y\|^{2N}}\, dy \lesssim a^{\alpha_s\left(\frac{1}{2}-c\right)+\frac{3}{4}}.$$

Therefore we have the condition

$$\alpha_s(1/2 - c) + 3/4 \geq \alpha_l(0) + 5/4.$$

Now we remember that we can choose any $c \in (0, 1/2)$ and, therefore, for any small ε we can choose $c = \dfrac{\varepsilon}{4\alpha_l + 2 + 2\varepsilon}$. With this the above condition is clearly true when

$$\alpha_s \geq 2\alpha_l(0) + 1 + \varepsilon.$$

■

4.3 Characterization of Hölder regularities by CST

Similar theorems as those presented in the previous section hold also for the continuous shearlet transform. Proofs are quite similar to the case of CCT and Hart Smith transform, see [24].

Theorem 27 *If a bounded function $f \in C^\alpha(\mathbb{R}^2)$, then there exists a constant C such that*

$$|\langle \psi_{ast}, f\rangle| \leq C a^{\alpha+\frac{3}{4}} \text{ and } \left|\left\langle \psi^{(v)}_{ast}, f \right\rangle\right| \leq C a^{\alpha+\frac{3}{4}}$$

for all $0 < a < 1$, $s \in [-2, 2]$ and $t \in \mathbb{R}^2$.

Notice, especially, how similar the above theorem is to Theorem 21 that was related to curvelet and Hart Smith transform. Next three theorems are shearlet versions from Theorems 22, 23 and 24.

Let us recall that $(P_E f)\widehat{\ } = \hat{f}\chi_E$ where E is either the cones C_1, C_2 defined in Section 2.5.1, or $T = \{(\xi_1, \xi_2) : |\xi_1| < 2 \text{ and } |\xi_2| < 2\}$.

Theorem 28 *Let $f \in L^2(\mathbb{R}^2)$ and $\alpha > 0$ be a non-integer. If there is a constant $C < \infty$ such that, for each $0 < a < 1$, $s \in [-2, 2]$ and $t \in \mathbb{R}^2$,*

$$|\langle \psi_{ast}, P_{C_1} f\rangle| \leq C a^{\alpha+\frac{5}{4}} \text{ and } \left|\left\langle \psi^{(v)}_{ast}, P_{C_2} f \right\rangle\right| \leq C a^{\alpha+\frac{5}{4}},$$

then $f \in C^\alpha(\mathbb{R}^2)$.

Theorem 29 *If a bounded function $f \in C^\alpha(u)$ then there exists $C < \infty$ such that*

$$|\langle \psi_{ast}, f \rangle| \le C a^{\frac{\alpha}{2}+\frac{3}{4}} \left(1 + \left\| \frac{t-u}{a^{1/2}} \right\|^\alpha\right) \text{ and}$$
$$\left|\left\langle \psi^{(v)}_{ast}, f \right\rangle\right| \le C a^{\frac{\alpha}{2}+\frac{3}{4}} \left(1 + \left\| \frac{t-u}{a^{1/2}} \right\|^\alpha\right) \tag{116}$$

for all $0 < a < 1$, $s \in [-2, 2]$ and $t \in \mathbb{R}^2$.

Theorem 30 *Let $f \in L^2(\mathbb{R}^2)$, $u \in \mathbb{R}^2$, and α be a non-integer positive number. Suppose there exist $C < \infty$ and $\alpha' < 2\alpha$ such that, for all $0 < a < 1$, $s \in [-2, 2]$ and $t \in \mathbb{R}^2$,*

$$|\langle \psi_{ast}, P_{C_1} f \rangle| \le C a^{\alpha+\frac{5}{4}} \left(1 + \left\| \frac{t-u}{a^{1/2}} \right\|^{\alpha'}\right) \tag{117}$$

and

$$\left|\left\langle \psi^{(v)}_{ast}, P_{C_2} f \right\rangle\right| \le C a^{\alpha+\frac{5}{4}} \left(1 + \left\| \frac{t-u}{a^{1/2}} \right\|^{\alpha'}\right). \tag{118}$$

Then $f \in C^\alpha(u)$.

Let Γ_u denote the vertical line passing through u and Γ_{u,s_0} denote the line passing through u with slope $-\frac{1}{s_0}$. Observe that we may write $\Gamma_u = \Gamma_{u,0}$ so that $(x_1, x_2) \in \Gamma_{u,s_0}$ if and only if $x_1 = -s_0(x_2 - u_2) + u_1$. The next theorem characterizes again a direction and regularity of a singularity. It is written in a bit more general form than the corresponding Theorem 26 for curvelets, since assumptions are a bit weaker.

Theorem 31 *Let f be bounded with $f \in C^\alpha(\Gamma_{u,s_0}, \mathbb{R}^2; (1,0))$ when $\alpha \in (0, 1]$ and $f \in C^{2\alpha+1+\varepsilon}(\mathbb{R}^2; \mathrm{B}_{s_0}(0,1))$ for some $s_0 \in [-2, 2]$ with any fixed $\varepsilon > 0$ and $u = (u_1, u_2) \in \mathbb{R}$. Then there exists $C < \infty$ such that for $0 < a < 1$, $t = (t_1, t_2) \in \mathbb{R}^2$, and $s \in [-2, 2]$,*

$$|\langle \psi_{ast}, f \rangle| \le \begin{cases} C a^{\alpha+\frac{5}{4}}, & \text{if } |s - s_0| \gtrsim \sqrt{a}, \\ C a^{\alpha+\frac{3}{4}} \left(1 + \left| \dfrac{t_1 + s_0 t_2 - u_1 - s_0 u_2}{a} \right|^\alpha\right), & \text{if } |s - s_0| \lesssim \sqrt{a}. \end{cases}$$

4.4 Regularity estimates by discrete curvelet and shearlet transforms

All theorems that state necessary conditions would be straightforward to state also in discrete case by demanding that values for a, θ, b, t and s are taken from some discrete sets. In theorems that state sufficient conditions, the reconstruction formula is needed in proofs. Transformation of proofs for discrete case would go again quite painless, instead triple integrals we would have just triple summations whenever reconstruction formulas would be used. Of course then we could not use change of variable in estimates, but we would use decay lemmas in similar way as in proofs in section 3.

4.5 Numerical demonstrations of convergence rates

In this section it is demonstrated by simple example how parameters s_0, θ_0 and α from Theorems 25, 26, 30 and 31 can be estimated numerically. Demonstrations are done for continuous shearlet transform and Hart Smith transforms.

4.5.1 Test settings

As generating kernel function ϕ the tensor product of the Meyer scaling function $\phi_M : \mathbb{R} \to \mathbb{R}$ and wavelet function $\psi_M : \mathbb{R} \to \mathbb{R}$ i.e., $\phi(x_1, x_2) = \psi_M(x_1)\phi_M(x_2)$ is used. Support of this function in Fourier domain meets the conditions in Lemmas 10 and 11, hence is suitable for our analysis. For more details about Meyer scaling function and wavelet, see for example [10].

We first choose the test set S of functions

$$S = \left\{ f(R_{\theta_0} x) \ : \ f(x) = e^{-\|x\|}|x_1|^\alpha, \alpha \in (0, 1], \theta_0 \in [0, \pi] \right\}.$$

All functions $f \in S$ clearly meets the conditions in Theorem 26. Notice also that for a particular $f \in S$, the local and pointwise Hölder exponents at the origin are the same.

As a test case we pick $f(x) = e^{-\|x\|}|x_1|^{0.25} \in S$. From the formula of f we clearly see that $\alpha = 0.25$ and $\theta_0 = 0$. The question is: Can we explore these values from some graphs if we only know that $f \in S$? It turns out that one can read these values from the behavior of the Hart Smith transform or shearlet transform.

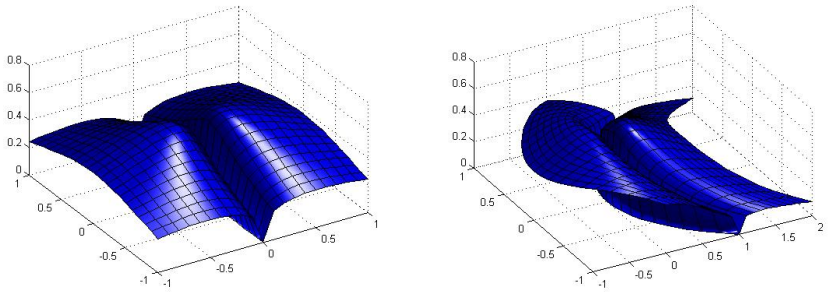
Theorems related to Hölder regularity in this thesis considered only line shaped singularity curves. However, it is interesting to try if numerical calculations would suggest that similar theorems may hold also for more general shapes of singularity curves. For this, we numerically estimate shearlet transform of function $h(x) = e^{-\|x\|}|x_1 - x_2^2|^{0.25}$. Notice that h is otherwise similar to f but has singularity along parabola instead of line. Graphs of f and h are shown in Figure 2.

4.5.2 Numerical calculations with Hart Smith transform

Numerically the transform can be calculated just for some finite discrete set of scales in finite time. We investigate scales $a = 2^{-i}$, $i \in I \subset \mathbb{N}$, i.e., we use similar scales that are often used for discrete shearlet, curvelet or contourlet transform.

Let us define the function

$$I_{\theta b}(i) := \log_2 \left(|\langle \varphi_{2^{-i}b\theta}, f \rangle| \right)$$



(a) (b)

Figure 2: (a) Graph of f; (b) Graph of h .

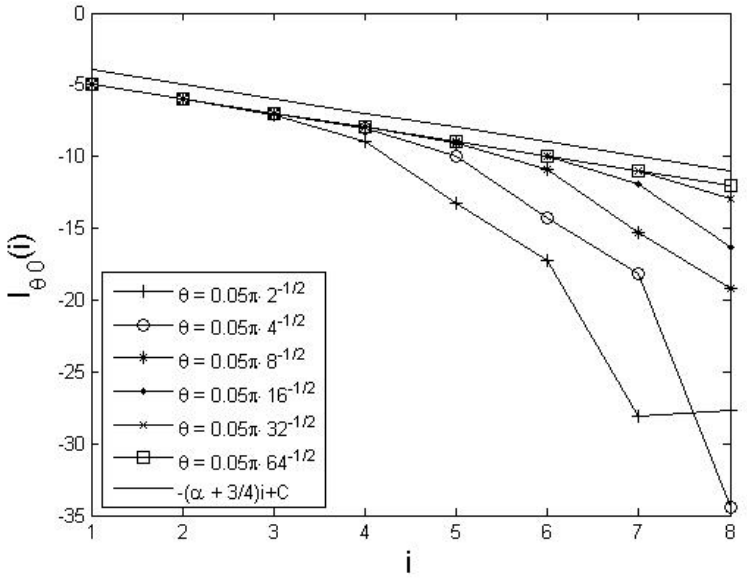


Figure 3: Decay behavior of $I_{\theta 0}(s)$ across scales $a = 2^{-i}$ at various angles θ for the function $f(x) = e^{-\|x\|}|x_1|^{0.25}$

and numerically evaluate its values when $b = 0$. This is illustrated in Figure 3.

The fact that $Ca^{1/2} = C2^{-i/2}$ is the critical angle, i.e. $\theta_0 = 0$, is more clearly emphasized once we define

$$g(i, j) := I_{\theta 0}(i) + (\alpha + 3/4)i, \quad \text{where } j := -2 \log_2 |\theta|,$$

whose graph is shown in Figure 4.

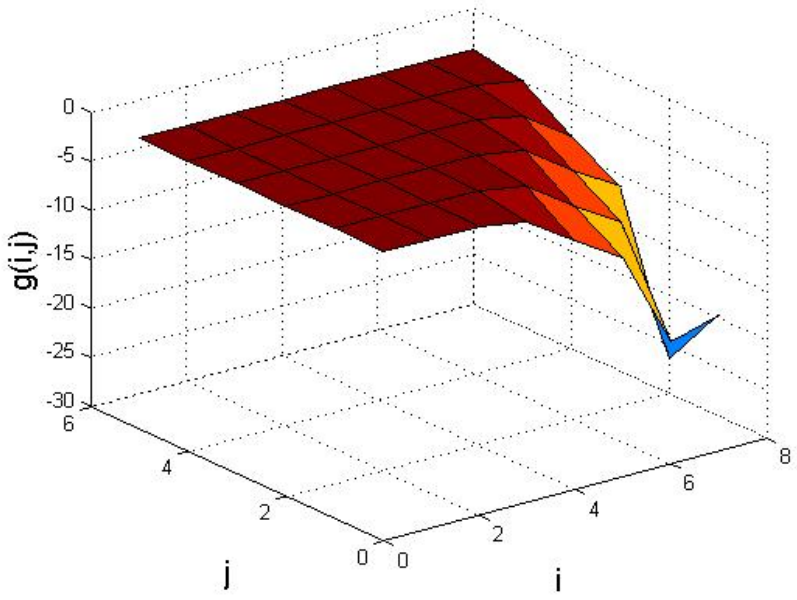


Figure 4: Graph of $g(i, j)$ with different scales $a = 2^i$ and angles $C = 2^{-j/2}$ for the function $f(x) = e^{-\|x\|}|x_1|^{0.25}$

There clearly exists a line $i = j + c$, where c is a constant, that divides the domain of g into two parts, the one where g is approximately constant (these are "small angles" where $I_{\theta 0}(i)$ decays slower) and another one where g still decays (these are "large angles" where $I_{\theta 0}(i)$ decays faster).

From Figures 3 and 4 we can see that $I_{\theta 0}(i) \approx -(0.25 + 3/4)i + C$ for angles small enough compared to the scale ($|\theta| < C\pi 2^{-j/2}$), i.e., it satisfies a sufficient condition on small angles for f being in $C^{0.25}(0)$. For $|\theta|$ larger than $C\pi 2^{-j/2}$, $I_{\theta 0}(i) \leq -(0.25 + 5/4)i + C$. So sufficient condition on large angles is also clearly satisfied. Therefore Theorem 25 suggest that $\alpha \geq 0.25$. To be precise, similar calculations should be done also for $b \neq 0$. Since $I_{\theta 0}(i) \approx -(0.25 + 3/4)i + C$ for $|\theta| < C2^{-j/2}$, Theorem 26 says

that the Hölder exponent cannot be higher than 0.25. Therefore, the estimate is that $\alpha \approx 0.25$.

In this simple example, we actually get a pretty good estimate α_e for α from the investigation of a function $\alpha_e(i, \theta) := I_{\theta 0}(i+1) - I_{\theta 0}(i)$. Figure 5 shows how the error $E(i, j) := \log_{10} |\alpha - \alpha_e(i, \theta)|$ behaves.

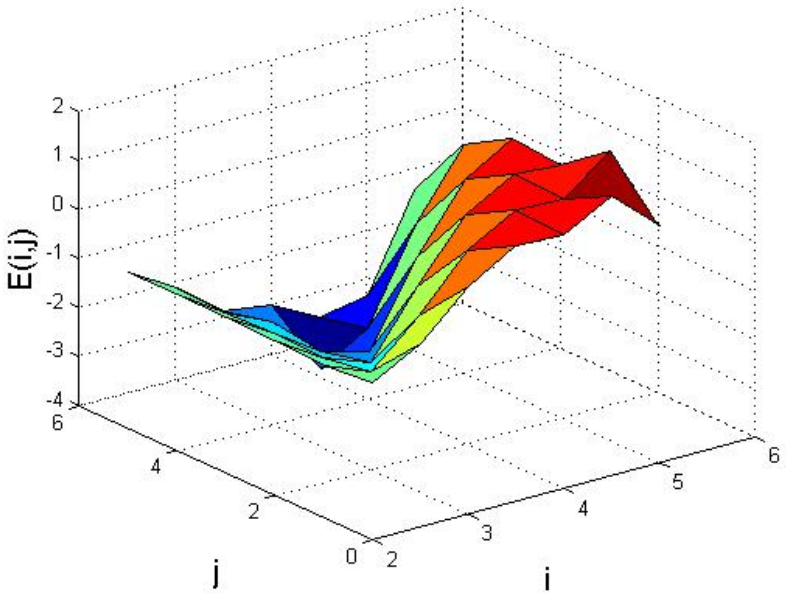


Figure 5: Estimation error $E(i, j)$ with different scales $a = 2^{-i}$ and angles $\theta = C2^{-j/2}$ for the function $f(x) = e^{-\|x\|}|x_1|^{0.25}$

Naturally $E(i, j)$ behaves very similarly to $g(i, j)$ and small estimation errors occur only on the domain where g is constant. Moreover, these errors seem to decrease along the lines $i = j + c$, i.e., the finer scales (and the smaller angles) we use, the better estimation accuracy we get. However, our implementation limits the accuracy of the estimation errors to be about 10^{-3} (this accuracy is achieved in many (i, j) couples), the peak value being $10^{-3.8}$ at scale $i = 5$.

In the examples, values of inner products $\langle f, \varphi_{ab\theta} \rangle$ are estimated by approximating integrals by sum of samples. Since lengths of supports of Φ_M and Ψ are infinite, tails must be cut off. For both functions the support is chosen to be $[-10, 10]$. This causes significant relative error for values when scale is small and true values of inner products should be small. Strange behavior in some graphs at fine scales and angles, where the transform is too small in value, are caused by errors numerical estimation

of inner products. All implementations are made with MATLAB.

4.5.3 Numerical calculations with shearlet transform

Regularity analysis of f is done in the same way with shearlets as with Hart Smith transform. The only small difference is in implementation: shearlet transform uses shearing matrix M_s while Hart Smith transform uses rotation matrix R_θ.

Similarly to Figure 3, Figure 6 illustrates the logarithmic decay rate of the continuous shearlet transform

$$I_{st}(i) := \log_2 \left(|\langle \psi_{2^{-i}st}, f \rangle| \right).$$

Moreover, similarly to Figure 4, also shearlet transform can be represented in a more illustrative form if the shear parameter s is written in the form $s = C2^{-j/2}$ and the function

$$g(j,i) := I_{(C2^{-j/2})0}(i) + (\alpha + 3/4)i$$

is considered. Figure 7 shows this graph for functions f and h, whose graphs were shown in Figure 2. It is interesting to notice that even if h does not posses a singularity line but singularity along parabola, the decay behavior is very similar to f. The transition from slow decay to fast decay just seems to take a few more scales for h than for f.

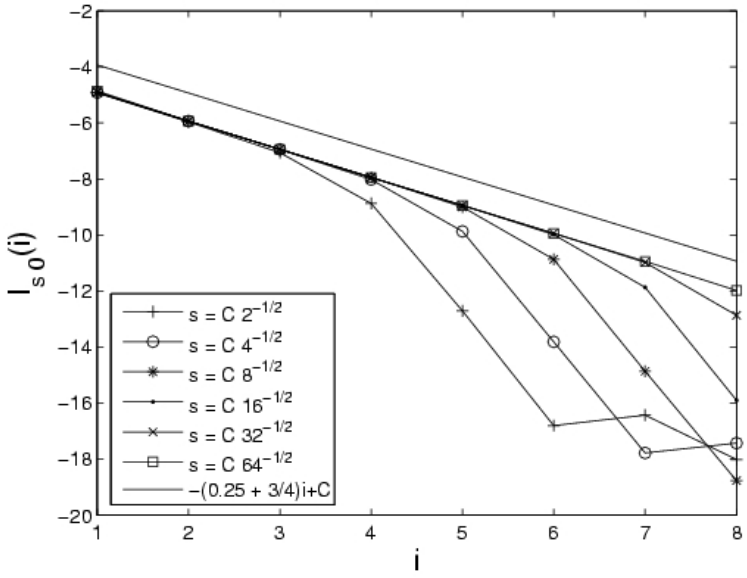
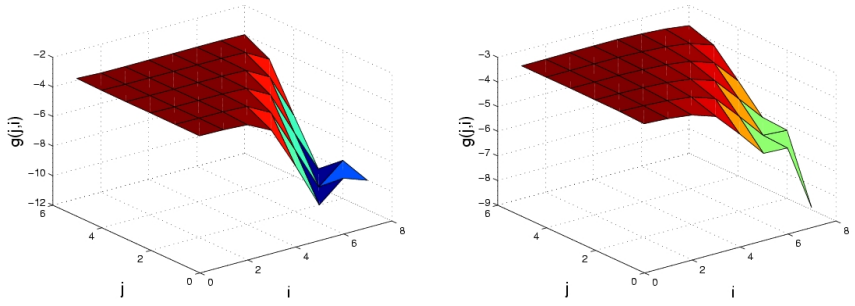


Figure 6: Decay behavior of shearlet transform across scales $a = 2^{-i}$ at various shear parameter s for the function $f(x) = e^{-\|x\|}|x_1|^{0.25}$

.



(a) (b)

Figure 7: Illustration of decay behavior of shearlet transform of (a) f; (b) h .

References

[1] A. Bonami and A. Estrade, *Anisotropic analysis of some gaussian models*, J. Fourier Anal. Appl. **9** (2003), no. 3, 215–236.

[2] Emmanuel Candès and Laurent Demanet, *Curvelets and Fourier integral operators*, C. R. Math. Acad. Sci. Paris **336** (2003), no. 5, 395–398. MR 1979352 (2004d:42059)

[3] Emmanuel J. Candès and David L. Donoho, *Curvelets and curvilinear integrals*, J. Approx. Theory **113** (2001), no. 1, 59–90. MR 1866248 (2002j:41012)

[4] ———, *Recovering edges in ill-posed inverse problems: optimality of curvelet frames*, Ann. Statist. **30** (2002), no. 3, 784–842, Dedicated to the memory of Lucien Le Cam. MR 1922542 (2003i:62085)

[5] ———, *New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities*, Comm. Pure Appl. Math. **57** (2004), no. 2, 219–266. MR 2012649 (2004k:42052)

[6] ———, *Continuous curvelet transform. I: Resolution of the wavefront set*, Appl. Comput. Harmon. Anal. **19** (2005), no. 2, 162–197.

[7] ———, *Continuous curvelet transform. II: Discretization and frames*, Appl. Comput. Harmon. Anal. **19** (2005), no. 2, 198–222.

[8] Arthur L. da Cunha and Minh N. Do, *On two-channel filter banks with directional vanishing moments*, IEEE Trans. Image Process. **16** (2007), no. 5, 1207–1219. MR 2464960 (2010e:94029)

[9] Ingrid Daubechies, *Orthonormal bases of compactly supported wavelets*, Comm. Pure Appl. Math. **41** (1988), no. 7, 909–996. MR 951745 (90m:42039)

[10] ———, *Ten lectures on wavelets*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.

[11] Minh N. Do and Martin Vetterli, *The contourlet transform: An efficient directional multiresolution image representation*, IEEE Trans. Image Process. **14** (2005), no. 12, 2091–2106.

[12] David L. Donoho, *De-noising by soft-thresholding*, IEEE Trans. Inform. Theory **41** (1995), no. 3, 613–627. MR 1331258 (96b:94002)

[13] Glenn Easley, Demetrio Labate, and Wang-Q Lim, *Sparse directional image representations using the discrete shearlet transform*, Appl. Comput. Harmon. Anal. **25** (2008), no. 1, 25–46. MR 2419703 (2009f:94006)

[14] Kanghui Guo, Gitta Kutyniok, and Demetrio Labate, *Sparse multidimensional representations using anisotropic dilation and shear operators*, Wavelets and splines: Athens 2005, Mod. Methods Math., Nashboro Press, Brentwood, TN, 2006, pp. 189–201. MR 2233452 (2007c:42050)

[15] Kanghui Guo and Demetrio Labate, *Optimally sparse multidimensional representation using shearlets*, SIAM J. Math. Anal. **39** (2007), no. 1, 298–318. MR 2318387 (2008k:42097)

[16] ———, *Characterization and analysis of edges using the continuous shearlet transform*, SIAM J. Imaging Sci. **2** (2009), no. 3, 959–986. MR 2551249

[17] Kanghui Guo, Demetrio Labate, and Wang-Q Lim, *Edge analysis and identification using the continuous shearlet transform*, Appl. Comput. Harmon. Anal. **27** (2009), no. 1, 24–46. MR 2526885 (2010d:42063)

[18] Deguang Han and David R. Larson, *Frames, bases and group representations*, Mem. Amer. Math. Soc. **147** (2000), no. 697, x+94. MR 1686653 (2001a:47013)

[19] Stéphane Jaffard, *Pointwise smoothness, two-microlocalization and wavelet coefficients*, Publ. Mat., Barc. **35** (1991), no. 1, 155–168.

[20] ———, *Multifractal functions: recent advances and open problems*, October 2004.

[21] K. M. Kolwankar and J. Lévy Véhel, *A time domain characterization of the fine local regularity of functions*, J. Fourier Anal. Appl. **8** (2002), no. 4, 319–334.

[22] Onkar C. Kulkarni, R. Vigneshwar, Valadi K. Jayaraman, and Bhaskar D. Kulkarni, *Identification of coding and non-coding sequences using local hölder exponent formalism*, Bioinformatics **21** (2005), no. 20, 3818–3823.

[23] Gitta Kutyniok and Demetrio Labate, *Resolution of the wavefront set using continuous shearlets*, Trans. Amer. Math. Soc. **361** (2009), no. 5, 2719–2754. MR 2471937 (2010b:42043)

[24] Panuvuth Lakhonchai, Jouni Sampo, and Songkiat Sumetkijakan, *Shearlet transforms and directional regularities*, Int. J. Wavelets, Multiresolution and Information processing **8** (2010), no. 5, 743–771.

[25] E. Le Pennec and S. Mallat, *Bandelet image approximation and compression*, Multiscale Model. Simul. **4** (2005), no. 3, 992–1039 (electronic). MR 2203949 (2006k:42082)

[26] Erwan Le Pennec and Stéphane Mallat, *Sparse geometric image representations with bandelets*, IEEE Trans. Image Process. **14** (2005), no. 4, 423–438. MR 2128287 (2006b:94004)

[27] P. Legrand and J.L. Vehel, *Local regularity-based image denoising*, Proceedings of International Conference on Image Processing (ICIP 2003) (2003).

[28] Stephane Mallat, *Multifrequency channel decomposition of images and wavelet models*, IEEE Trans. on Acoustics, Speech and Signal Processing **34** (1989), 2091–2110.

[29] Stéphane Mallat, *A wavelet tour of signal processing*, Academic Press, San Diego, CA, 1997.

[30] ———, *Geometrical grouplets*, Appl. Comput. Harmon. Anal. **26** (2009), no. 2, 161–180. MR 2490212 (2010a:42148)

[31] Stéphane Mallat and Gabriel Peyré, *A review of bandlet methods for geometrical image representation*, Numer. Algorithms **44** (2007), no. 3, 205–234. MR 2339522 (2008e:94007)

[32] Stephane G. Mallat, *Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbf{R})$*, Trans. Amer. Math. Soc. **315** (1989), no. 1, 69–87. MR 1008470 (90e:42046)

[33] K. Nualtong and S. Sumetkijakan, *Analysis of Hölder regularity by wavelet-like transforms with parabolic scaling*, Thai J. Math. **3** (2005), no. 2, 275–283.

[34] Duncan D.-Y. Po and Minh N. Do, *Directional multiscale modeling of images using the contourlet transform*, IEEE Trans. Image Process. **15** (2006), no. 6, 1610–1620. MR 2478998

[35] Javier Portilla and Eero P. Simoncelli, *A parametric texture model based on joint statistics of complex wavelet coefficients*, International Journal of Computer Vision **40** (2000), no. 1, 49–71.

[36] Jouni Sampo and Songkiat Sumetkijakan, *Estimations of Hölder regularities and direction of singularity by Hart Smith and curvelet transforms*, J. Fourier Anal. Appl. **15** (2009), no. 1, 58–79. MR 2491026 (2010h:26014)

[37] S. Seuret and J. Lévy Véhel, *The local Hölder function of a continuous function*, Appl. Comput. Harmon. Anal. **13** (2002), no. 3, 263–276.

[38] Hart F. Smith, *A Hardy space for Fourier integral operators*, J. Geom. Anal. **8** (1998), no. 4, 629–653.

[39] Jean-Luc Starck, Emmanuel J. Candès, and David L. Donoho, *The curvelet transform for image denoising*, IEEE Trans. Image Process. **11** (2002), no. 6, 670–684. MR 1929403 (2003j:94013)

[40] Jean-Luc Starck, Fionn Murtagh, Emmanuel J. Candès, and David L. Donoho, *Gray and color image contrast enhancement by the curvelet transform*, IEEE Trans. Image Process. **12** (2003), no. 6, 706–717. MR 1988273 (2004g:94010)

[41] Leonardo Trujillo, Gustavo Olague, Pierrick Legrand, and Evelyne Lutton, *Regularity based descriptor computed from local image oscillations*, Optics Express **15** (2007), no. 10, 6140–6145.

ACTA UNIVERSITATIS LAPPEENRANTAENSIS

374. JUSSILA, HANNE. Concentrated winding multiphase permanent magnet machine design and electromagnetic properties – Case axial flux machine. 2009. Diss.

375. AUVINEN, HARRI. Inversion and assimilation methods with applications in geophysical remote sensing. 2009. Diss.

376. KINDSIGO, MERIT. Wet oxidation of recalcitrant lignin waters: Experimental and kinetic studies. 2009. Diss.

377. PESSI, PEKKA. Novel robot solutions for carrying out field joint welding and machining in the assembly of the vacuum vessel of ITER. 2009. Diss.

378. STRÖM, JUHA-PEKKA. Actived*u*/d*t* filtering for variable-speed AC drives. 2009. Diss.

379. NURMI, SIMO A. Computational and experimental investigation of the grooved roll in paper machine environment. 2009. Diss.

380. HÄKKINEN, ANTTI. The influence of crystallization conditions on the filtration characteristics of sulphathiazole suspensions. 2009. Diss.

381. SYRJÄ, PASI. Pienten osakeyhtiöiden verosuunnittelu – empiirinen tutkimus. 2010. Diss.

382. KERKKÄNEN, ANNASTIINA. Improving demand forecasting practices in the industrial context. 2010. Diss.

383. TAHVANAINEN, KAISA. Managing regulatory risks when outsourcing network-related services in the electricity distribution sector. 2010. Diss.

384. RITALA, PAAVO. Coopetitive advantage – How firms create and appropriate value by collaborating with their competitors. 2010. Diss.

385. RAUVANTO, IRINA. The intrinsic mechanisms of softwood fiber damage in brown stock fiber line unit operations. 2010. Diss.

386. NAUMANEN, VILLE. Multilevel converter modulation: implementation and analysis. 2010. Diss.

387. IKÄVALKO, MARKKU. Contextuality in SME ownership – Studies on owner-managers´ ownership behavior. 2010. Diss.

388. SALOJÄRVI, HANNA. Customer knowledge processing in key account management. 2010. Diss.

389. ITKONEN, TONI. Parallel-operating three-phase voltage source inverters – Circulating current modeling, analysis and mitigation. 2010. Diss.

390. EEROLA, TUOMAS. Computational visual quality of digitally printed images. 2010. Diss.

391. TIAINEN, RISTO. Utilization of a time domain simulator in the technical and economic analysis of a wind turbine electric drive train. 2010. Diss.

392. GRÖNMAN AKI. Numerical modelling of small supersonic axial flow turbines. 2010. Diss.

393. KÄHKÖNEN, ANNI-KAISA. The role of power relations in strategic supply management – A value net approach. 2010. Diss.

394. VIROLAINEN, ILKKA. Johdon coaching: Rajanvetoja, taustateorioita ja prosesseja. 2010. Diss.

395. HONG, JIANZHONG. Cultural aspects of university-industry knowledge interaction. 2010. Diss.

396. AARNIOVUORI, LASSI. Induction motor drive energy efficiency – Simulation and analysis. 2010. Diss.

397. SALMINEN, KRISTIAN. The effects of some furnish and paper structure related factors on wet web tensile and relaxation characteristics. 2010. Diss.

398. WANDERA, CATHERINE. Performance of high power fiber laser cutting of thick-section steel and medium-section aluminium. 2010. Diss.

399. ALATALO, HANNU. Supersaturation-controlled crystallization. 2010. Diss.

400. RUNGI, MAIT. Management of interdependency in project portfolio management. 2010. Diss.

401. PITKÄNEN, HEIKKI. First principles modeling of metallic alloys and alloy surfaces. 2010. Diss.

402. VAHTERISTO, KARI. Kinetic modeling of mechanisms of industrially important organic reactions in gas and liquid phase. 2010. Diss.

403. LAAKKONEN, TOMMI. Distributed control architecture of power electronics building-block-based frequency converters. 2010. Diss.

404. PELTONIEMI, PASI. Phase voltage control and filtering in a converter-fed single-phase customer-end system of the LVDC distribution network. 2010. Diss.

405. TANSKANEN, ANNA. Analysis of electricity distribution network operation business models and capitalization of control room functions with DMS. 2010. Diss.

406. PIIRAINEN, KALLE A. IDEAS for strategic technology management: Design of an electronically mediated scenario process. 2010. Diss.

407. JOKINEN, MARKKU. Centralized motion control of a linear tooth belt drive: Analysis of the performance and limitations. 2010. Diss.

408. KÄMÄRI, VESA. Kumppanuusohjelman strateginen johtaminen – Monitapaustutkimus puolustushallinnossa. 2010. Diss.

409. KARJALAINEN, AHTI. Online ultrasound measurements of membrane compaction. 2010. Diss.

410. LOHTANDER, MIKA. On the development of object functions and restrictions for shapes made with a turret punch press. 2010. Diss.

411. SIHVO, VILLE. Insulated system in an integrated motor compressor. 2010. Diss.

412. SADOVNIKOV, ALBERT. Computational evaluation of print unevenness according to human vision. 2010. Diss.

413. SJÖGREN, HELENA. Osingonjakopäätökset pienissä osakeyhtiöissä. Empiirinen tutkimus osakeyhtiölain varojenjakosäännösten toteutumisesta. 2010. Diss.

414. KAUPPI, TOMI. Eye fundus image analysis for automatic detection of diabetic retinopathy. 2010. Diss.

415. ZAKHVALINSKII, VASILII. Magnetic and transport properties of $LaMnO_{3+\delta}$, $La_{1-x}Ca_xMnO_3$, $La_{1-x}Ca_xMn_{1-y}Fe_yO_3$ and $La_{1-x}Sr_xMn_{1-y}Fe_yO_3$. 2010. Diss.

416. HATAKKA, HENRY. Effect of hydrodynamics on modelling, monitoring and control of crystallization. 2010. Diss.