

# Cartoon Approximation with $\alpha$ -Curvelets

Philipp Grohs, Sandra Keiper, Gitta Kutyniok, Martin Schäfer \*

April 3, 2014

## Abstract

It is well-known that curvelets provide optimal approximations for so-called cartoon images which are defined as piecewise  $C^2$ -functions, separated by a  $C^2$  singularity curve. In this paper, we consider the more general case of piecewise  $C^\beta$ -functions, separated by a  $C^\beta$  singularity curve for  $\beta \in (1, 2]$ . We first prove a benchmark result for the possibly achievable best  $N$ -term approximation rate for this more general signal model. Then we introduce what we call  $\alpha$ -curvelets, which are systems that interpolate between wavelet systems on the one hand ( $\alpha = 1$ ) and curvelet systems on the other hand ( $\alpha = \frac{1}{2}$ ). Our main result states that those frames achieve this optimal rate for  $\alpha = \frac{1}{\beta}$ , up to log-factors.

## 1 Introduction

In various applications in signal processing it has proven useful to decompose a given signal in a multiscale dictionary, for instance to achieve compression by coefficient thresholding or to solve inverse problems. The most popular family of such dictionaries are undoubtedly wavelets, which have had a tremendous impact in applied mathematics since Daubechies' construction of orthonormal wavelet bases with compact support in the 1980s. While wavelets are now a well-established tool in numerical signal processing (for instance the JPEG2000 coding standard is based on a wavelet transform), it has been recognized in the past decades that they also possess several shortcomings, in particular with respect to the treatment of multidimensional data where anisotropic structures such as edges in images are typically present. This deficiency of wavelets has given birth to the research area of geometric multiscale analysis where frame constructions which are optimally adapted to anisotropic structures are sought.

### 1.1 Geometric Multiscale Analysis

A milestone in this area has been the construction of curvelet and shearlet frames which are indeed capable of optimally resolving curved singularities in multidimensional data. To be more precise, the landmark paper [3] has shown that simple coefficient thresholding in a curvelet frame yields an, up to (negligible) log-factors, optimal  $N$ -term approximation rate for the class of so-called *cartoon images* which are roughly defined as compactly supported, bivariate, piecewise  $C^2$  functions, separated by a  $C^2$  discontinuity curve. After this breakthrough result new constructions of anisotropic frame systems have been introduced which achieve the same optimal approximation rates for cartoon images. Among those we would like to single out *shearlets* which are better suited for digital implementation than the original curvelets. They were first introduced in [12] and a comprehensive summary of their properties can be found in the survey [14]. The recent work [11] introduced the framework of *parabolic molecules* which subsumes all the earlier constructions mentioned above and which established a transference principle of approximation results between any two systems of parabolic molecules, provided that one of them constitutes a tight frame for  $L^2(\mathbb{R}^2)$ .

---

\*PG was supported in part by Swiss National Fund (SNF) Grant 146356. SK acknowledges support from the Berlin Mathematical School and the DFG Collaborative Research Center TRR 109 "Discretization in Geometry and Dynamics. GK was supported in part by the Einstein Foundation Berlin, by the Einstein Center for Mathematics Berlin (ECMath), by Deutsche Forschungsgemeinschaft (DFG) Grant KU 1446/14, by the DFG Collaborative Research Center TRR 109 "Discretization in Geometry and Dynamics", and by the DFG Research Center MATHEON "Mathematics for key technologies" in Berlin.

## 1.2 General Image Models

By now the model of such cartoon-images is widely recognized as a useful mathematical model for image data. However, for certain applications it might appear still too restrictive. For instance one could imagine signals which consist of piecewise smooth parts separated by a singularity curve which might not necessarily be exactly  $C^2$  but of higher or lower smoothness.

The approximation of such a generalized signal class by hybrid shearlets has been studied in [15] for three-dimensional data. In that paper the best  $N$ -term approximation results which could be established are suboptimal by a small margin. In addition, hybrid shearlets are compactly supported, but do not form a tight frame.

## 1.3 Our Contribution

In the present paper we study optimal approximation schemes for a more general model, namely bivariate piecewise  $C^\beta$  functions, separated by a  $C^\beta$  discontinuity curve, where  $\beta \in (1, 2]$ . We establish a benchmark result which states that one cannot in general achieve an  $N$ -term approximation rate higher than  $\frac{\beta}{2}$  for such functions. We then introduce the notion of  $\alpha$ -curvelet frames which generalize the construction of [3], where  $\alpha \in [0, 1]$  describes the degree of anisotropy in the scaling operation used for the frame construction. The parameter  $\alpha = 0$  corresponds to ridgelets [2], the case  $\alpha = \frac{1}{2}$  to second-generation curvelets as in [3], and the case  $\alpha = 1$  to wavelets.  $\alpha$ -curvelets form a tight frame for any parameter  $\alpha$  in contrast to the systems considered in [15]. Our main result then states that all intermediate cases  $\alpha \in [\frac{1}{2}, 1]$  provide optimal (up to log-factors)  $N$ -term approximation rates for the above mentioned signal class with  $\beta = \frac{1}{\alpha}$ . This is particularly surprising, since for the result in the 3D situation ([15]) for intermediate parameters only a suboptimal rate could be proven. Our proof techniques rely somehow on [3], however we wish to mention that the technical details are actually quite different from those used there. In particular, we have to deal with smoothness spaces of fractional orders which forces us to replace estimates for derivatives in [3] by estimates on certain moduli of smoothness.

Having such an approximation result for a tight frame system at hand will allow us to transform our approximation result to arbitrary systems of frames of  $\alpha$ -molecules (including compactly supported hybrid shearlets as special case), as first introduced in [9]. This will be the subject of the forthcoming work [10], generalizing the framework of parabolic molecules [11].

## 1.4 Outline

After introducing the general class of image models in Section 2 and deriving the benchmark result in Theorem 2.2, Section 3 is devoted to the construction of  $\alpha$ -curvelets. Our main result on the optimal approximation properties of  $\alpha$ -curvelets for any parameter  $\alpha \in [\frac{1}{2}, 1]$  is then stated in the next section as Theorem 4.1. This section as well as the appendix do contain the lengthy, quite technical proof of this result.

## 1.5 Preliminaries

Let us fix some notational conventions used throughout this work. For  $x \in \mathbb{R}^d$  we denote the Euclidean distance by  $|x|_2$  and the  $\ell_1$ -norm by  $|x|$ . Further, we put  $x^m = x_1^{m_1} \cdots x_d^{m_d}$  for  $m \in \mathbb{N}_0^d$ .

For a function  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  the forward difference operator  $\Delta_{(h_1, \dots, h_d)}$ , where  $h_1, \dots, h_d \in \mathbb{R}$ , is given by

$$\Delta_{(h_1, \dots, h_d)} f(x_1, \dots, x_d) := f(x_1 + h_1, \dots, x_d + h_d) - f(x_1, \dots, x_d).$$

In the one-dimensional case it takes the simple form  $\Delta_h f(t) := f(t + h) - f(t)$ . We will often apply the forward difference operator to a bivariate function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ . To simplify notation, the operator  $\Delta_h$  shall exclusively act on variables denoted  $t$  or  $\tau \in \mathbb{R}$ , e.g., the symbol  $\Delta_h f(t, u)$  denotes the function  $(t, u) \mapsto f(t + h, u) - f(t, u)$ .

We denote by  $\partial_i$  the derivative in the  $i$ -th coordinate direction,  $i \in \{1, \dots, d\}$ , and we let  $\partial^m = \partial_1^{m_1} \cdots \partial_d^{m_d}$  for  $m \in \mathbb{N}_0^d$ .

We write  $\|\cdot\|_p$  for the respective norms in the spaces  $L^p(\mathbb{R}^d)$  and  $\ell^p(\Lambda)$ , where  $\Lambda$  is a discrete set.

For  $f \in L^1(\mathbb{R}^d)$ ,  $d \in \mathbb{N}$ , the Fourier transform  $\mathcal{F}f$  is defined by

$$\mathcal{F}f(\xi) = \int_{\mathbb{R}^d} f(x) \exp(-2\pi i x \cdot \xi) dx.$$

The short-hand notation for  $\mathcal{F}f$  is  $\widehat{f}$ .

We shall also make extensive use of the Radon transform  $\mathcal{R}f$  given by

$$\mathcal{R}F(t, \eta) := \int_{\mathfrak{L}_{t, \eta}} F(x) dl(x),$$

with  $(t, \eta) \in \mathbb{R} \times (-\pi/2, \pi/2]$  and  $\mathfrak{L}_{t, \eta} := \{(x_1, x_2) \in \mathbb{R}^2 : \sin(\eta)x_1 + \cos(\eta)x_2 = t\}$ .

The Hölder spaces  $C^\gamma$  for  $\gamma \in (1, 2]$  are defined as follows:

$$C^\gamma(\mathbb{R}^d) := \left\{ f \in C(\mathbb{R}^d) : \max_{i=1}^d \text{Höl}(\partial_i f, \gamma - 1) < \infty \right\}$$

where for a function  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  we use the notation

$$\text{Höl}(f, \alpha) = \sup_{x, y \in \mathbb{R}^d} \frac{|f(x) - f(y)|}{|x - y|^\alpha}, \quad \alpha \in [0, 1],$$

for the Hölder constant. The notation  $C_0^\gamma$  refers to compactly supported functions in  $C^\gamma$ .

In the sequel, for two quantities  $A, B \in \mathbb{R}$ , which may depend on several parameters we shall write  $A \lesssim B$ , if there exists a constant  $C > 0$  such that  $A \leq CB$ , uniformly in the parameters. If the converse inequality holds true, we write  $A \gtrsim B$  and if both inequalities hold we shall write  $A \asymp B$ .

## 2 Cartoon Approximation

Many applications require efficient encoding of multivariate data in the sense of optimal (sparse) approximation rates by a suitable representation system. This is typically phrased as a problem of best  $N$ -term approximation as explained in Subsection 2.1. The performance of an approximation scheme is then analyzed with respect to certain subclasses of the Hilbert space  $L^2(\mathbb{R}^2)$ , which is the standard continuum domain model for 2-dimensional data, in particular in imaging science. The key feature of most multivariate data is the appearance of anisotropic phenomena. Hence such a subclass of  $L^2(\mathbb{R}^2)$  is required to provide a suitable model for this fact, which is fulfilled by the subclass of so-called cartoon-like images as defined in Subsection 2.2. The main result of this section, Theorem 2.8, will provide an upper bound for the maximal achievable approximation rate for this class of cartoon images. It serves as a benchmark for the approximation performance of concrete representation systems.

### 2.1 Sparse Approximation

Let us start with a short introduction to some aspects of approximation theory. The standard continuum model for 2-dimensional data, in particular in imaging science, is the Hilbert space  $L^2(\mathbb{R}^2)$ . From a practical standpoint however, a function  $f \in L^2(\mathbb{R}^2)$  is a rather intractable object. Therefore, in order to analyze  $f$ , the most common approach is to represent it with respect to some representation system  $\Phi = (\varphi_\lambda)_{\lambda \in \Lambda} \subset L^2(\mathbb{R}^2)$ , i.e., to expand  $f$  as

$$f = \sum_{\lambda \in \Lambda} c_\lambda \varphi_\lambda, \tag{1}$$

and then consider the coefficients  $c_\lambda \in \mathbb{R}$ .

Moreover, since in real world applications infinitely many coefficients are infeasible, the function  $f$  has to be approximated by a finite subset of this system. Letting  $N$  be the number of elements allowed in this approximation, we obtain what is called an  $N$ -term approximation for  $f$  with respect to  $\Phi$ . The *best  $N$ -term approximation*, usually denoted by  $f_N$ , is optimal among those in terms of a minimal approximation error and is defined by

$$f_N = \underset{g = \sum_{\lambda \in \Lambda_N} c_\lambda \varphi_\lambda}{\text{argmin}} \|f - g\|_2^2 \quad \text{s.t.} \quad \#\Lambda_N \leq N.$$

The approximation performance of a system  $\Phi$  for a subclass  $\mathfrak{F} \subset L^2(\mathbb{R}^2)$  is typically measured by the *asymptotic approximation rate*, i.e., the decay of the  $L^2$ -error of the best  $N$ -term approximation  $\|f - f_N\|_2$  as  $N \rightarrow \infty$ .

In a very general form the representation system  $\Phi$  can just be an arbitrary *dictionary*, i.e. we require only  $L^2(\mathbb{R}^2) = \overline{\text{span } \Phi}$ . General dictionaries  $\Phi$  can be rather pathological, and for  $f \in L^2(\mathbb{R}^2)$  there might not even exist a best  $N$ -term approximation. For example, since  $L^2(\mathbb{R}^2)$  is separable, we can choose  $\Phi$  to be a countable dense subset of  $L^2(\mathbb{R}^2)$ . This choice would yield arbitrarily good 1-term approximations since the elements of  $\Phi$  can come arbitrarily close to  $f$ .

Therefore, without reasonable restrictions on the search depth, the investigation of best  $N$ -term approximation with respect to a given dictionary may not even make sense. A commonly used and suitable restriction is to impose *polynomial depth search*, which requires that the terms of the  $N$ -term approximation have to come from the first  $\pi(N)$  elements of the dictionary, where  $\pi$  is some fixed polynomial [6].

### 2.1.1 Polynomial Depth Search in a Dictionary

Let us get more specific and assume that we have a countable dictionary  $\Phi = (\varphi_n)_{n \in \mathbb{N}}$  indexed by the natural numbers, and a fixed polynomial  $\pi$  specifying the search depth. A non-linear  $N$ -term approximation scheme can then be described by a set-valued selection function  $\mathcal{S}$ , which determines for given  $f \in L^2(\mathbb{R}^2)$  and  $N \in \mathbb{N}$  the selected dictionary elements, i.e.  $\mathcal{S}(f, N) \subset \Phi$  with  $\#\mathcal{S}(f, N) = N$ . The function  $\mathcal{S}$  shall obey the polynomial depth search constraint, i.e.  $\mathcal{S}(f, N) \subset \{\varphi_1, \dots, \varphi_{\pi(N)}\}$ , and the dependence on  $f$  allows for adaptive approximation schemes.

Let us recall a benchmark concerning the optimality rate of approximation in a dictionary with polynomial depth search, as derived in [6]. Before, we have to define what it means for a class of functions to contain a copy of  $\ell_0^p$ .

**Definition 2.1.** 1. A class of functions  $\mathfrak{F} \subset L^2(\mathbb{R}^2)$  is said to contain an embedded orthogonal hypercube of dimension  $m$  and sidelength  $\delta$  if there exist  $f_0 \in \mathfrak{F}$  and orthogonal functions  $\psi_i \in L^2(\mathbb{R}^2)$  for  $i = 1, \dots, m$  with  $\|\psi_i\|_2 = \delta$  such that the collection of hypercube vertices

$$\mathcal{H}(m; f_0, (\psi_i)_i) = \left\{ h = f_0 + \sum_{i=1}^m \varepsilon_i \psi_i : \varepsilon_i \in \{0, 1\} \right\}$$

is contained in  $\mathfrak{F}$ . It should be noted that  $\mathcal{H}$  just consists of its vertices.

2. A class of functions  $\mathfrak{F}$  is said to contain a copy of  $\ell_0^p$ ,  $p > 0$ , if there exists a sequence of orthogonal hypercubes  $(\mathcal{H}_k)_{k \in \mathbb{N}}$ , embedded in  $\mathfrak{F}$ , which have dimensions  $m_k$  and side-lengths  $\delta_k$ , such that  $\delta_k \rightarrow 0$  and for some constant  $C > 0$

$$m_k \geq C \delta_k^{-p} \quad \text{for all } k \in \mathbb{N}. \quad (2)$$

Note, that if  $\mathfrak{F}$  contains a copy of  $\ell_0^p$ , then it also contains a copy of  $\ell_0^q$  for all  $0 < q < p$ . It was shown in [6] that if a function class  $\mathfrak{F}$  contains a copy of  $\ell_0^p$  there exists an upper bound on the maximal achievable approximation rate via reconstruction in a fixed dictionary.

We state a reformulation of this landmark result, which in its original form [6] was stated in terms of the coefficient decay. The original proof however can be adapted to lead to the following formulation in terms of the best  $N$ -term approximation, which is more appropriate for our needs. We remark, that this theorem is also valid in a general Hilbert space setting.

**Theorem 2.2.** Suppose, that a class of functions  $\mathfrak{F} \subset L^2(\mathbb{R}^2)$  is uniformly  $L^2$ -bounded and contains a copy of  $\ell_0^p$ . Then, allowing only polynomial depth search in a given dictionary, there is a constant  $C > 0$  such that for every  $N_0 \in \mathbb{N}$  there is a function  $f \in \mathfrak{F}$  and an  $N \in \mathbb{N}$ ,  $N \geq N_0$  such that

$$\|f - f_N\|_2^2 \geq C (N \log_2(N))^{-(2-p)/p},$$

where  $f_N$  denotes the best  $N$ -term approximation under the polynomial depth search constraint.

*Proof.* Let  $\Phi = (\varphi_n)_{n \in \mathbb{N}}$  be a given dictionary and  $\pi$  the polynomial, specifying the search depth. The best  $N$ -term approximation of  $f \in L^2(\mathbb{R}^2)$  obtained in this setting shall be denoted by  $f_N$ , the corresponding selection rule, as described above, by  $\mathcal{S}$ .

Each system  $\mathcal{S}(f, N)$  can be orthonormalized by the Gram-Schmidt procedure (starting from lower indices to higher indices), giving rise to an orthonormal basis of  $\text{span } \mathcal{S}(f, N)$  (with the exception of some possible zero vectors). Therefore we can represent each  $f_N$  by the unique set of coefficients obtained from an expansion in this basis. (If a basis element is zero, the corresponding coefficient is chosen to be zero.)

In order to apply information theoretic arguments, we consider the following coding procedure. For  $f \in \mathfrak{F}$  we select the dictionary elements  $\mathcal{S}(f, N)$  and quantize the coefficients of  $f_N$  obtained as above by rounding to multiples of the quantity  $\mathfrak{q} = N^{-2/p}$ .

We need  $N \log_2(\pi(N))$  bits of information to encode the locations of the selected elements  $\mathcal{S}(f, N)$  and  $N \log_2(2T/\mathfrak{q})$  bits for the coefficients themselves, where  $T$  is the  $L^2$ -bound for the elements of  $\mathfrak{F}$ . Hence, in this procedure we are encoding with at most

$$R(N) = N(C_1 + C_2 \log_2(N)), \quad C_1, C_2 > 0,$$

bits, and for  $N \geq 2$  we have  $R(N) \leq C_3 N \log_2(N)$  for some constant  $C_3 > 0$ . To decode, we simply reconstruct the rounded values of the coefficients and then synthesize using the selected dictionary elements.

Let  $\mathcal{H}$  be a hypercube in  $\mathfrak{F}$  of dimension  $m$  and sidelength  $\delta$ . Starting with a vertex  $h \in \mathcal{H}$  the coding-decoding procedure (for some fixed  $N \in \mathbb{N}$ ) yields some  $\tilde{h} \in L^2(\mathbb{R}^2)$ . By passing to the  $L^2$ -closest vertex  $\hat{h}$ , we again obtain an element of the hypercube  $\mathcal{H}$ .

Every vertex  $h \in \mathcal{H}$  can be represented as a word of  $m$  bits, each bit corresponding to one side of the cube. Thus the above coding procedure gives a map of the  $m$  bits, which specify the vertex  $h \in \mathcal{H}$ , to  $R = R(N)$  bits. The decoding then reconstructs the  $m$  bits specifying the vertex  $\hat{h} \in \mathcal{H}$ . Since at the intermediate step we just have  $R$  bits of information we unavoidably lose information if  $R < m$ .

Now we can apply an information theoretic argument. By rate-distortion theory [6, 1] there must be at least one vertex  $h \in \mathcal{H}$ , where the number of false reconstructed bits is larger than  $D_m(R)$ . Here  $D_m(R)$  is the so-called  $m$ -letter distortion rate function. Since each bit determines a side of the cube, the error we make for this vertex  $h$  obeys

$$\|h - \hat{h}\|_2^2 \geq \delta^2 \cdot D_m(R).$$

Since by construction  $\|\tilde{h} - h\|_2 \geq \|\tilde{h} - \hat{h}\|_2$  we have  $\|\tilde{h} - h\|_2 \geq \frac{1}{2} \|\hat{h} - h\|_2$ . It follows

$$\|\tilde{h} - h\|_2^2 \geq \frac{1}{4} \delta^2 \cdot D_m(R).$$

By assumption,  $\mathfrak{F}$  contains a copy of  $\ell_0^p$ . Therefore we can find a sequence of hypercubes  $\mathcal{H}_k$  with sidelengths  $\delta_k \rightarrow 0$  as  $k \rightarrow \infty$  and dimensions  $m_k = m_k(\delta_k) \geq C \delta_k^{-p}$ . We pick  $N_k \in \mathbb{N}$  such that  $C_3 N_k \log_2(N_k) \leq \frac{1}{3} m_k$  and  $N_k \rightarrow \infty$  as  $k \rightarrow \infty$ . For large  $k$  we then obey the inequality  $R(N_k) \leq \frac{1}{3} m_k$ , in fact  $N_k \geq 2$  is sufficient.

Here we can apply another result from rate-distortion theory. If  $\frac{R}{m} \leq \rho$  for some  $\rho < \frac{1}{2}$  it holds  $D_m(R)/m \geq D_1(\rho)$ , where  $D_1$  is the so-called single-letter distortion-rate function. Hence, if  $\frac{R}{m} \leq \frac{1}{3}$ , we have

$$\|\tilde{h} - h\|_2^2 \geq \frac{1}{4} D_1\left(\frac{1}{3}\right) \delta^2 m.$$

Let  $h_k$  denote the vertices with maximal reconstruction error  $\|h_k - \tilde{h}_k\|_2$  at each hypercube  $\mathcal{H}_k$ . Then we can conclude for large  $k$

$$\|\tilde{h}_k - h_k\|_2^2 \geq \frac{1}{4} D_1\left(\frac{1}{3}\right) \delta_k^2 m_k \gtrsim \delta_k^2 m_k \gtrsim (N_k \log_2(N_k))^{-(2-p)/p}.$$

Finally we have to take care of the rounding errors. The Euclidean distance between the best  $N_k$ -term approximation  $h'_k$  differs from  $\tilde{h}_k$  by at most  $\mathfrak{q} \sqrt{N_k}$ , i.e.

$$\|\tilde{h}_k - h'_k\|_2 \leq \mathfrak{q} \sqrt{N_k},$$

since the coefficients belong to an orthonormal basis. It follows, with some constant  $C > 0$ ,

$$\|h_k - h'_k\|_2 \geq \|\tilde{h}_k - h_k\|_2 - \|\tilde{h}_k - h'_k\|_2 \geq C(N_k \log_2(N_k))^{\frac{1}{2} - \frac{1}{p}} - N_k^{1/2-2/p} \gtrsim (N_k \log_2(N_k))^{-(2-p)/(2p)}.$$

This finishes the proof.  $\square$

### 2.1.2 Frame Approximation

In practice one needs representations, which are robust and stable with respect to noise. This leads to the notion of a *frame* [4], which is a special kind of dictionary.

A system  $\Phi = (\varphi_\lambda)_{\lambda \in \Lambda} \subset L^2(\mathbb{R}^2)$  is called a frame for  $L^2(\mathbb{R}^2)$ , if there exist constants  $0 < C_1 \leq C_2 < \infty$  such that

$$C_1 \|f\|_2^2 \leq \sum_{\lambda \in \Lambda} |\langle f, \varphi_\lambda \rangle|^2 \leq C_2 \|f\|_2^2 \quad \text{for all } f \in L^2(\mathbb{R}^2).$$

A frame is called *tight*, if  $C_1 = C_2$  is possible, and *Parseval*, if  $C_1 = C_2 = 1$ . Since the *frame operator*  $S : L^2(\mathbb{R}^2) \rightarrow L^2(\mathbb{R}^2)$  defined by  $Sf = \sum_{\lambda \in \Lambda} \langle f, \varphi_\lambda \rangle \varphi_\lambda$  is always invertible, it follows that one particular sequence of coefficients in the expansion (1) can be computed as

$$c_\lambda = \langle f, S^{-1} \varphi_\lambda \rangle, \quad \lambda \in \Lambda, \quad (3)$$

where  $(S^{-1} \varphi_\lambda)_\lambda$  is usually referred to as the *canonical dual frame*. Note, that in general for a redundant system the expansion coefficients in (1) are not unique. The particular coefficients (3), however, are usually referred to as *the* frame coefficients. They have the distinct property that they minimize the  $\ell^2$ -norm.

Even in the frame setting, the computation of the best  $N$ -term approximation is not yet well-understood. The delicacy of this problem can for instance be seen in [7]. A typical approach to circumvent this problem is to consider instead the  $N$ -term approximation obtained by choosing the  $N$  largest frame coefficients. It is evident that the associated error then also provides a bound for the error of best  $N$ -term approximation.

There exists a close relation between the  $N$ -term approximation rate achieved by a frame, and the decay rate of the corresponding frame coefficients. A typical measure for the sparsity of a sequence  $(c_\lambda)_\lambda$  are the  $\ell^p$ -(quasi-)norms, for  $p > 0$  defined by

$$\|(c_\lambda)_\lambda\|_{\omega \ell^p} := \left( \sup_{\varepsilon > 0} \varepsilon^p \cdot \#\{\lambda : |c_\lambda| > \varepsilon\} \right)^{1/p}.$$

Equivalently, for a zero sequence  $(c_n)_{n \in \mathbb{N}}$  indexed by  $\mathbb{N}$ , these (quasi-)norms can be characterized by  $\|(c_n)_n\|_{\omega \ell^p} = \sup_{n > 0} n^{1/p} |c_n^*|$ , where  $(c_n^*)_n$  denotes the non-increasing rearrangement of  $(c_n)_n$ .

The following lemma shows that membership of the coefficient sequence to an  $\ell^p$ -space for small  $p$  implies good  $N$ -term approximation rates. For a proof, we refer to [5, 16].

**Lemma 2.3.** *Let  $f = \sum c_\lambda \varphi_\lambda$  be an expansion of  $f \in L^2(\mathbb{R}^2)$  with respect to a frame  $(\varphi_\lambda)_{\lambda \in \Lambda}$ . Further, assume that the coefficients satisfy  $(c_\lambda)_\lambda \in \omega \ell^{2/(2k+1)}$  for some  $k > 0$ . Then the best  $N$ -term approximation rate is at least of order  $N^{-k}$ , i.e.*

$$\|f - f_N\|_2 \lesssim N^{-k}.$$

Next, we apply Theorem 2.2 to obtain an upper bound on the approximation rate of cartoon-like images.

## 2.2 Cartoon-like Images

To model the fact that multivariate data appearing in applications is typically governed by anisotropic features – in the 2-dimensional case curvilinear structures –, the so-called *cartoon-like functions* were introduced in [6]. This class is by now a widely used standard model in particular for natural images. It mimics the fact that natural images often consist of nearly smooth parts separated by discontinuities.

The first rigorous mathematical definition was given in [6] and extensively employed starting from the work in [3]. It postulates that images consist of  $C^2$ -regions separated by piecewise smooth  $C^2$ -curves.

Since its introduction several extensions of this model have been introduced and studied, starting with the extended model in [15]. We consider images that contain two smooth  $C^\beta$ -regions separated by a piecewise smooth  $C^\gamma$ -curve, where  $\beta, \gamma \in (1, 2]$ .

Without loss of generality we will subsequently assume the smoothness of the edge curve. In addition, to avoid technicalities, we restrict our considerations to star-shaped discontinuity curves, which allow an easy parametrization. This is a classical simplification also used in [6, 3, 16, 15]. We remark however that this restriction to star-shaped curves is artificial. In fact, we could work with the class of regular  $C^\gamma$ -Jordan domains contained in  $[0, 1]^2$ , where  $\gamma \in (1, 2]$  and the Hölder constants in the canonical parametrization are bounded.

We begin by introducing the class of star-shaped sets  $STAR^\gamma(\nu)$  for  $\gamma \in (1, 2]$  and  $\nu > 0$ . For that we take a  $C^\gamma$ -function  $\rho : \mathbb{T} \rightarrow [0, \infty)$  defined on the torus  $\mathbb{T} = [0, 2\pi]$ , where the boundary points are identified. Additionally, we assume that there exists  $0 < \rho_0 < 1$  such that  $\rho(\eta) \leq \rho_0$  for all  $\eta \in \mathbb{T}$ . Then we define the subset  $\mathcal{B} \subset \mathbb{R}^2$  by

$$\mathcal{B} = \left\{ x \in \mathbb{R}^2 : x = (|x|_2, \eta) \text{ in polar coordinates with } \eta \in \mathbb{T}, |x|_2 \leq \rho(\eta) \right\}, \quad (4)$$

such that the boundary  $\Gamma = \partial\mathcal{B}$  of  $\mathcal{B}$  is a closed regular  $C^\gamma$ -Jordan curve in  $[0, 1]^2$  parameterized by

$$b(\eta) = \begin{pmatrix} \rho(\eta) \cos(\eta) \\ \rho(\eta) \sin(\eta) \end{pmatrix}, \quad \eta \in \mathbb{T}. \quad (5)$$

Furthermore, we require the Hölder constant of the radius function  $\rho$  to be bounded by  $\nu$ , i.e.,

$$\text{Höl}(\rho', \gamma - 1) = \sup_{\eta, \tilde{\eta} \in \mathbb{T}} \frac{|\rho'(\eta) - \rho'(\tilde{\eta})|}{|\eta - \tilde{\eta}|^{\gamma-1}} \leq \nu, \quad (6)$$

where the distance  $|\eta - \tilde{\eta}|$  is measured on the torus  $\mathbb{T}$ .

**Definition 2.4.** For  $\nu > 0$  the set  $STAR^\gamma(\nu)$  is defined as the collection of all subsets  $\mathcal{B} \subset [0, 1]^2$ , which are translates of sets of the form (4) with a boundary obeying (5) and (6).

The class of cartoon-like functions is then defined as follows.

**Definition 2.5.** 1. Let  $\nu > 0$ ,  $\beta, \gamma \in (1, 2]$ . The cartoon-like functions  $\mathcal{E}_\gamma^\beta(\mathbb{R}^2; \nu) \subset L^2(\mathbb{R}^2)$  are defined as the set of functions  $f : \mathbb{R}^2 \rightarrow \mathbb{C}$  of the form

$$f = f_0 + f_1 \chi_{\mathcal{B}},$$

where  $\mathcal{B} \in STAR^\gamma(\nu)$  and  $f_i \in C^\beta(\mathbb{R}^2)$  with  $\text{supp } f_0 \subset [0, 1]^2$  and  $\|f_i\|_{C^\beta} \leq \nu$  for each  $i = 0, 1$ . To simplify notation we let  $\mathcal{E}^\beta(\mathbb{R}^2) = \mathcal{E}_\beta^\beta(\mathbb{R}^2)$ .

2.  $\mathcal{E}_\gamma^{\text{bin}}(\mathbb{R}^2; \nu)$  denotes the class of binary cartoon-like images, i.e., functions  $f = f_0 + f_1 \chi_{\mathcal{B}} \in \mathcal{E}_\gamma^\beta(\mathbb{R}^2; \nu)$ , where  $f_0 = 0$  and  $f_1 = 1$ .

We aim for a benchmark for sparse approximation of functions belonging to  $\mathcal{E}_\gamma^\beta(\mathbb{R}^2)$ . For this we can use the results from the previous section. In order to apply these results to our model of cartoon-like functions we investigate for which  $p > 0$  the class of cartoon-like images contains a copy of  $\ell_0^p$  following [15, 13].

**Theorem 2.6.** Let  $\nu > 0$  be fixed, and  $\beta, \gamma \in (1, 2]$ .

1. The class of binary cartoon-like images  $\mathcal{E}_\gamma^{\text{bin}}(\mathbb{R}^2; \nu)$  contains a copy of  $\ell_0^p$  for  $p = 2/(\gamma + 1)$ .
2. The class  $C_0^\beta([0, 1]^2; \nu)$  of Hölder smooth functions  $f \in C_0^\beta([0, 1]^2)$  with  $\|f\|_{C^\beta} \leq \nu$  contains a copy of  $\ell_0^p$  for  $p = 2/(\beta + 1)$ .

*Proof.* ad 1) This was proved by Donoho in [6].

ad 2) To show that  $C_0^\beta([0, 1]^2; \nu)$  contains a copy of  $\ell_0^p$  we have to find a sequence of embedded orthogonal hypercubes  $(\mathcal{H}_k)_{k \in \mathbb{N}}$  of dimension  $m_k$  and size  $\delta_k$  such that  $\delta_k \rightarrow 0$  and (2) holds.

Let  $\phi \in C_0^\infty(\mathbb{R})$  with  $\text{supp } \phi \subset [0, 1]$  and  $\phi \geq 0$  and put  $\psi(t) = \phi(t_1)\phi(t_2)$  for  $t = (t_1, t_2) \in \mathbb{R}^2$ . Then  $\psi \in C^\beta(\mathbb{R}^2)$  with  $\text{supp } \psi \subset [0, 1]^2$ . By possibly rescaling we can ensure  $\|\psi\|_{C^\beta(\mathbb{R}^2)} \leq \nu$ . Further, we define for  $k \in \mathbb{N}$  and  $i = (i_1, i_2) \in \{0, \dots, k-1\}^2$  the functions

$$\psi_{i,k}(t) = k^{-\beta} \phi(kt_1 - i_1) \phi(kt_2 - i_2).$$

These functions  $\psi_{i,k} \in C^\beta(\mathbb{R}^2)$  are dilated and translated versions of  $\psi$  satisfying  $H\ddot{o}l(\psi_{i,k}, \beta) = H\ddot{o}l(\psi, \beta)$  and  $\|\psi_{i,k}\|_{C^\beta(\mathbb{R}^2)} \leq \|\psi\|_{C^\beta(\mathbb{R}^2)} \leq \nu$ . Moreover, it holds  $\|\psi_{i,k}\|_2^2 = k^{-2\beta-2} \|\psi\|_2^2$ , and since  $\text{supp } \psi_{i,k} \subset [\frac{i_1}{k}, \frac{i_1+1}{k}] \times [\frac{i_2}{k}, \frac{i_2+1}{k}]$  the functions  $\psi_{i,k}$  and  $\psi_{j,k}$  are orthogonal in  $L^2(\mathbb{R}^2)$  for  $i \neq j$ .

The hypercubes  $\mathcal{H}_k$  of dimensions  $m_k = k^2$  given by

$$\begin{aligned} \mathcal{H}_k &:= \mathcal{H}(k^2; 0, (\psi_{i,k})) = \left\{ h = \sum_{i_1=1}^k \sum_{i_2=1}^k \varepsilon_{i_1, i_2} \phi(k \cdot -i_1) \phi(k \cdot -i_2) \mid \varepsilon_{i_1, i_2} \in \{0, 1\} \right\} \\ &= \left\{ h = \sum_{i=1}^{m^2} \varepsilon_i \psi_{i,k} \mid \varepsilon_i \in \{0, 1\} \right\} \end{aligned}$$

with side-lengths  $\delta_k = \|\psi_{i,k}\|_2 = k^{-\beta-1} \|\psi\|_2$  are clearly contained in  $C_0^\beta([0, 1]^2; \nu)$ . We have to check (2). It holds

$$m_k = k^2 = \left( \frac{\delta_k}{\|\psi\|_2} \right)^{-\frac{2}{\beta+1}} = \|\psi\|_2^{\frac{2}{\beta+1}} \cdot (\delta_k)^{-\frac{2}{\beta+1}},$$

and the sequence  $(\delta_k)_{k \in \mathbb{N}}$  obeys  $\delta_k \rightarrow 0$ . This finishes the proof.  $\square$

An immediate corollary is the following result.

**Corollary 2.7.** *The function class  $\mathcal{E}_\gamma^\beta(\mathbb{R}^2; \nu)$  contains a copy of  $\ell_0^p$  for  $p = \max\{2/(\beta+1), 2/(\gamma+1)\}$ .*

As a direct consequence of Theorem 2.2 we now state the main result of this subsection, which has also been proved in [15].

**Theorem 2.8.** *Given an arbitrary dictionary and allowing only polynomial depth search, for the class of cartoon-like images  $f \in \mathcal{E}_\gamma^\beta(\mathbb{R}^2)$  the decay rate of the error  $\|f - f_N\|_2^2$  of the best  $N$ -term approximation cannot exceed  $N^{-\min\{\beta, \gamma\}}$ .*

We observe that the upper bound on the approximation rate for  $\mathcal{E}_\gamma^\beta(\mathbb{R}^2)$  is the same as for the larger class  $\mathcal{E}^{\min\{\beta, \gamma\}}(\mathbb{R}^2)$ . Therefore we restrict our investigation to classes of the form  $\mathcal{E}^\beta(\mathbb{R}^2)$ . Since we are able to establish a lower bound for  $\mathcal{E}^\beta(\mathbb{R}^2)$  differing only by a log-factor from this upper bound, we do not loose much.

### 3 Hybrid Curvelets

Second generation curvelets, which are nowadays simply referred to as curvelets, were introduced in 2004 by Candès and Donoho [3]. The construction involves a parabolic scaling law, which can be viewed as a natural compromise between isotropic scaling, as utilized for wavelets, and scaling in only one coordinate direction, as utilized for ridgelets [8].

A more general anisotropic scaling law is realized by so-called  $\alpha$ -scaling, where the parameter  $\alpha \in [0, 1]$  is used to control the degree of anisotropy in the scaling. The associated  $\alpha$ -scaling matrix is defined by

$$D_s = \begin{pmatrix} s & 0 \\ 0 & s^\alpha \end{pmatrix}, \quad s \in \mathbb{R}_+.$$

Isotropic scaling corresponds to  $\alpha = 1$ , and scaling in one coordinate direction to  $\alpha = 0$ . Parabolic scaling is the special case for  $\alpha = \frac{1}{2}$ . In this sense, curvelets can be viewed as lying in between ridgelets and wavelets.

Adapting the construction of curvelets to  $\alpha$ -scaling yields, what we will call hybrid curvelets, or more specifically  $\alpha$ -curvelets. In the following we will construct bandlimited tight frames of  $\alpha$ -curvelets for every  $\alpha \in [0, 1]$ . Thus, we obtain a whole scale of representation systems, which interpolate between wavelets for  $\alpha = 1$  on the one end and ridgelets for  $\alpha = 0$  on the other end.



The construction follows the same recipe used for the tight ridgelet frames in [8]. As mentioned above it can also be seen as a variation of the classical second generation curvelet frame from [3].

The construction is simplified by treating the radial and angular components separately. For the construction of the radial functions  $W^{(j)}$  at each scale  $j \in \mathbb{N}_0$  we start with  $C^\infty$ -functions  $\widetilde{W}^{(0)} : \mathbb{R}_+ \rightarrow [0, 1]$  and  $\widetilde{W} : \mathbb{R}_+ \rightarrow [0, 1]$  with the following properties

$$\begin{aligned} \text{supp } \widetilde{W}^{(0)} &\subset [0, 2], & \widetilde{W}^{(0)}(r) &= 1 \text{ for all } r \in [0, \frac{3}{2}], \\ \text{supp } \widetilde{W} &\subset [\frac{1}{2}, 2], & \widetilde{W}(r) &= 1 \text{ for all } r \in [\frac{3}{4}, \frac{3}{2}]. \end{aligned}$$

Then we define for  $j \in \mathbb{N}$  the functions  $\widetilde{W}^{(j)}(r) := \widetilde{W}(2^{-j}r)$  on  $\mathbb{R}_+$ . Finally, we rescale for every  $j \in \mathbb{N}_0$  (to obtain an integer grid later)

$$W^{(j)}(r) := \widetilde{W}^{(j)}(8\pi r), \quad r \in \mathbb{R}_+.$$

Notice, that it holds  $2 \geq \sum W^{(j)} \geq 1$ .

Next, we define the angular functions  $V^{(j,\ell)} : \mathbb{S}^1 \rightarrow [0, 1]$  on the unit circle  $\mathbb{S}^1 \subset \mathbb{R}^2$ , where  $j \in \mathbb{N}$  specifies the scale and  $\ell$  the orientation, running through  $\{0, \dots, L_j - 1\}$  with

$$L_j = 2^{\lfloor j(1-\alpha) \rfloor}, \quad j \in \mathbb{N}.$$

This time we start with a  $C^\infty$ -function  $V : \mathbb{R} \rightarrow [0, 1]$ , living on the whole of  $\mathbb{R}$ , satisfying

$$\text{supp } V \subset [-\frac{3}{4}\pi, \frac{3}{4}\pi] \quad \text{and} \quad V(t) = 1 \text{ for all } t \in [-\frac{\pi}{2}, \frac{\pi}{2}].$$

Since the interval  $[-\pi, \pi]$  can be identified via  $t \mapsto e^{it}$  with  $\mathbb{S}^1$ , we obtain  $C^\infty$ -functions  $\widetilde{V}^{(j,0)} : \mathbb{S}^1 \rightarrow [0, 1]$  for every  $j \in \mathbb{N}$  by restricting the scaled functions  $V(2^{\lfloor j(1-\alpha) \rfloor} \cdot)$  to  $[-\pi, \pi]$ . In order to end up with real-valued curvelets, we then symmetrize

$$V^{(j,0)}(\xi) := \widetilde{V}^{(j,0)}(\xi) + \widetilde{V}^{(j,0)}(-\xi) \quad , \xi \in \mathbb{S}^1.$$

At each scale  $j \in \mathbb{N}$  we define the characteristic angle  $\omega_j = \pi 2^{-\lfloor j(1-\alpha) \rfloor}$ , and for  $\ell \in \{0, 1, \dots, L_j - 1\}$  we let

$$R_{j,\ell} = \begin{pmatrix} \cos(\omega_{j,\ell}) & -\sin(\omega_{j,\ell}) \\ \sin(\omega_{j,\ell}) & \cos(\omega_{j,\ell}) \end{pmatrix} \quad (7)$$

be the rotation matrix by the angle  $\omega_{j,\ell} = \ell \omega_j$ . By rotating  $V^{(j,0)}$  we finally get  $V^{(j,\ell)} : \mathbb{S}^1 \rightarrow [0, 1]$ ,

$$V^{(j,\ell)}(\xi) := V^{(j,0)}(R_{j,\ell}\xi) \quad \text{for } \xi \in \mathbb{S}^1.$$

In order to secure the tightness of the frame we need to renormalize with the function

$$\Psi(\xi) := W^{(0)}(|\xi|)^2 + \sum_{j,\ell} W^{(j)}(|\xi|)^2 V^{(j,\ell)}\left(\frac{\xi}{|\xi|}\right)^2,$$

which satisfies  $1 \leq \Psi(\xi) \leq 8$  for all  $\xi \in \mathbb{R}^2$ . Bringing the radial and angular components together, we obtain the functions

$$\chi_0(\xi) := \frac{W^{(0)}(|\xi|)}{\sqrt{\Psi(\xi)}}, \quad \chi_{j,\ell}(\xi) = \frac{W^{(j)}(|\xi|) V^{(j,\ell)}\left(\frac{\xi}{|\xi|}\right)}{\sqrt{\Psi(\xi)}} \quad \text{for } \xi \in \mathbb{R}^2. \quad (8)$$

It is obvious that  $\chi_0, \chi_{j,\ell} \in C^\infty(\mathbb{R}^2)$ . Moreover, these functions are real-valued, non-negative, compactly supported, and  $L^\infty$ -bounded by 1. Let  $\mathbf{e}_1 = (1, 0) \in \mathbb{R}^2$  be the first unit vector and put

$$S_j := \left\{ x \in \mathbb{S}^1 : |\langle x, \mathbf{e}_1 \rangle| \geq \cos(\omega_j/2) \right\}.$$

Indeed, we have  $\text{supp } \chi_0 \subset \mathcal{W}_0 := \{\xi \in \mathbb{R}^2 : 8\pi|\xi| \leq 2\}$  and  $\text{supp } \chi_{j,\ell} \subset \mathcal{W}_{j,\ell}$ , where for  $j \in \mathbb{N}$ ,  $\ell \in \{0, \dots, L_j - 1\}$  the sets

$$\mathcal{W}_{j,\ell} := \left\{ \xi \in \mathbb{R}^2 : 2^{j-1} \leq 8\pi|\xi| \leq 2^{j+1}, \frac{\xi}{|\xi|} \in R_{j,\ell}^{-1} S_j \right\} \quad (9)$$

are antipodal pairs of symmetric wedges.

After this preparation, we are ready to define the functions  $\psi_0$  and  $\psi_{j,\ell}$  on the Fourier side by

$$\widehat{\psi}_0(\xi) := \chi_0(\xi) \quad \text{and} \quad \widehat{\psi}_{j,\ell}(\xi) := \chi_{j,\ell}(\xi) \quad \text{for } \xi \in \mathbb{R}^2.$$

They have the following important property.

**Lemma 3.1.** *For every  $f \in L^2(\mathbb{R}^2)$  it holds*

$$\|f\|_2^2 = \|f * \psi_0\|_2^2 + \sum_{j,\ell} \|f * \psi_{j,\ell}\|_2^2.$$

*Proof.* By construction the system satisfies the discrete Caldéron condition, i.e.,

$$|\widehat{\psi}_0(\xi)|^2 + \sum_{j,\ell} |\widehat{\psi}_{j,\ell}(\xi)|^2 = 1 \quad \text{for all } \xi \in \mathbb{R}^2.$$

This yields for all  $f \in L^2(\mathbb{R}^2)$

$$\begin{aligned} \|f\|_2^2 &= \int_{\mathbb{R}^2} |\widehat{f}(\xi)|^2 d\xi \\ &= \int_{\mathbb{R}^2} |\widehat{\psi}_0(\xi)|^2 |\widehat{f}(\xi)|^2 d\xi + \int_{\mathbb{R}^2} \sum_{j,\ell} |\widehat{\psi}_{j,\ell}(\xi)|^2 |\widehat{f}(\xi)|^2 d\xi \\ &= \|f * \psi_0\|_2^2 + \sum_{j,\ell} \|f * \psi_{j,\ell}\|_2^2. \end{aligned}$$

□

The full frame of  $\alpha$ -curvelets is obtained by taking translates of  $\psi_0$  and  $\psi_{j,\ell}$  in the spatial domain and  $L^2$ -normalizing afterwards. Accordingly, for  $j \in \mathbb{N}$ ,  $\ell \in \{0, 1, \dots, L_j - 1\}$ , and  $k \in \mathbb{Z}^2$  we define  $\psi_{0,k} := \psi_0(\cdot - k)$  and

$$\psi_{j,\ell,k} := 2^{-j(1+\alpha)/2} \cdot \psi_{j,\ell}(\cdot - x_{j,\ell,k}) \quad \text{with} \quad x_{j,\ell,k} = R_{j,\ell}^{-1} D_{2^j}^{-1} k.$$

The corresponding set of curvelet indices  $\mu = (j, \ell, k)$  will henceforth be denoted by  $M$ . Further, we will use the following notation for this system

$$\mathcal{C}_\alpha(W^{(0)}, W, V) = (\psi_\mu)_{\mu \in M} = \left\{ \psi_{0,k} : k \in \mathbb{Z}^2 \right\} \cup \left\{ \psi_{j,\ell,k} : j \in \mathbb{N}, k \in \mathbb{Z}^2, \ell \in \{0, 1, \dots, L_j - 1\} \right\}.$$

**Theorem 3.2.** *Let  $\alpha \in [0, 1]$  and  $W^{(0)}, W, V$  be defined as above. The  $\alpha$ -curvelet system  $\mathcal{C}_\alpha(W^{(0)}, W, V)$  constitutes a tight frame for  $L^2(\mathbb{R}^2)$ .*

*Proof.* For  $j \in \mathbb{N}$  the wedges  $\mathcal{W}_{j,0}$  are contained in the rectangles  $\Xi_{j,0}$ ,

$$\mathcal{W}_{j,0} \subset \frac{1}{8\pi} \left( [-2^{j+1}, 2^{j+1}] \times [-4\pi 2^{j\alpha}, 4\pi 2^{j\alpha}] \right) \subset [-2^{j-1}, 2^{j-1}] \times [-2^{j\alpha-1}, 2^{j\alpha-1}] =: \Xi_{j,0}, \quad (10)$$

and  $\mathcal{W}_0$  is contained in the cube  $[-2, 2]^2 / (8\pi) \subset [-\frac{1}{2}, \frac{1}{2}]^2 =: \Xi_0$ . Since  $\mathcal{W}_{j,\ell} = R_{j,\ell}^{-1} \mathcal{W}_{j,0}$  we can put  $\Xi_{j,\ell} := R_{j,\ell}^{-1} \Xi_{j,0}$  such that  $\mathcal{W}_{j,\ell} \subset \Xi_{j,\ell}$ .

For each fixed  $j \in \mathbb{N}$  the Fourier system  $(u_{j,k})_{k \in \mathbb{Z}^2}$  given by

$$u_{j,k}(\xi) = 2^{-j(1+\alpha)/2} \exp(2\pi i x_{j,0,k} \cdot \xi) \quad \text{with} \quad x_{j,0,k} = (2^{-j} k_1, 2^{-j\alpha} k_2) = D_{2^j}^{-1} k \quad (11)$$

constitutes an orthonormal basis for  $L^2(\Xi_{j,0})$ . Observe that  $x_{j,\ell,k} = R_{j,\ell}^{-1}x_{j,0,k}$ . Therefore,

$$u_{j,k}(R_{j,\ell}\xi) = 2^{-j(1+\alpha)/2} \exp(2\pi i x_{j,\ell,k} \cdot \xi),$$

and the system  $(u_{j,k}(R_{j,\ell}\cdot))_{k \in \mathbb{Z}^2}$  is an orthonormal basis for  $L^2(\Xi_{j,\ell})$ . By Lemma 3.1 it holds

$$\|f\|_2^2 = \|f * \psi_0\|_2^2 + \sum_{j,\ell} \|f * \psi_{j,\ell}\|_2^2.$$

By the above observation we have for  $j \in \mathbb{N}$

$$\begin{aligned} \|f * \psi_{j,\ell}\|_2^2 &= \|\widehat{f} \cdot \widehat{\psi}_{j,\ell}\|_2^2 = \int_{\Xi_{j,\ell}} |\widehat{f}(\xi) \widehat{\psi}_{j,\ell}(\xi)|^2 d\xi = \sum_{k \in \mathbb{Z}^2} \left| \int \widehat{f}(\xi) \widehat{\psi}_{j,\ell}(\xi) u_{j,k}(R_{j,\ell}\xi) d\xi \right|^2 \\ &= \sum_{k \in \mathbb{Z}^2} |\langle \widehat{f}, 2^{-j(1+\alpha)/2} \widehat{\psi}_{j,\ell} e^{2\pi i x_{j,\ell,k} \cdot} \rangle|^2 = \sum_{k \in \mathbb{Z}^2} |\langle f, \psi_{j,\ell,k} \rangle|^2. \end{aligned}$$

A similar argument for  $\|f * \psi_0\|_2^2$  finishes the proof.  $\square$

In order to simplify the notation we henceforth use the capital letter  $J$  to denote a scale-angle pair  $(j, \ell)$ . In this context  $|J|$  shall denote the corresponding scale variable  $j$ , i.e. for  $J = (j, \ell)$  it is  $|J| = j$ . Therefore, from now on we may write e.g.  $\omega_J, R_J$  to abbreviate  $\omega_{j,\ell}, R_{j,\ell}$ , etc.

## 4 Cartoon Approximation with $\alpha$ -Curvelets

In this main section we will use the tight frame of  $\alpha$ -curvelets  $\mathcal{C}_\alpha(W^{(0)}, W, V) = (\psi_\mu)_{\mu \in M}$ , constructed in the previous section, to approximate the cartoon image class  $\mathcal{E}^\beta(\mathbb{R}^2)$ . In this investigation we keep  $\alpha \in [\frac{1}{2}, 1)$  fixed in the range between wavelets and curvelets, and put  $\beta = \alpha^{-1}$ . Our main goal is to prove the following approximation result, which generalizes an analogous result in [3] for second generation curvelets.

**Theorem 4.1.** *Let  $\alpha \in [\frac{1}{2}, 1)$  and  $\beta = \alpha^{-1}$ . The tight frame of  $\alpha$ -curvelets  $\mathcal{C}_\alpha(W^{(0)}, W, V)$  constructed in Section 3 provides almost optimally sparse approximations for cartoon-like functions  $\mathcal{E}^\beta(\mathbb{R}^2)$ . More precisely, there exists some constant  $C$  such that for every  $f \in \mathcal{E}^\beta(\mathbb{R}^2)$*

$$\|f - f_N\|_2^2 \leq CN^{-\beta} \cdot (\log_2 N)^{\beta+1} \quad \text{as } N \rightarrow \infty,$$

where  $f_N$  is the  $N$ -term approximation of  $f$  obtained by choosing the  $N$  largest curvelet coefficients.

When we compare this theorem with the benchmark Theorem 2.8, we see that  $\mathcal{C}_\alpha(W^{(0)}, W, V) = (\psi_\mu)_{\mu \in M}$  attains the maximal achievable approximation rate up to a log-factor. This is remarkable since this rate is achieved by simply thresholding the coefficients, leading to an intrinsically non-adaptive approximation scheme.

Theorem 4.1 is proved in several steps, extending the techniques used in [3]. The basic idea is to study the decay of the sequence  $(\theta_\mu)_{\mu \in M}$  of curvelet coefficients  $\theta_\mu = \langle f, \psi_\mu \rangle$  for  $f \in \mathcal{E}^\beta(\mathbb{R}^2)$ . According to Lemma 2.3 this decay rate provides us with the desired information on the approximation rate.

### 4.1 Sparsity of Curvelet Coefficients for Cartoon-like Images

We first state a simple a-priori estimate for the size of the curvelet coefficients  $\theta_\mu = \langle f, \psi_\mu \rangle$  at scale  $j$ . The curvelets  $\psi_\mu$  are  $L^2$ -normalized and essentially supported in a box of length  $2^{-\alpha j}$  and width  $2^{-j}$ . Therefore we can estimate

$$\|\psi_\mu\|_1 \leq B 2^{-(1+\alpha)j/2} \|\psi_\mu\|_2 \leq B 2^{-(1+\alpha)j/2},$$

where the constant  $B$  is uniform over all scales. This argument can easily be made rigorous, and we obtain the following estimate for the coefficients at scale  $j$ :

$$|\theta_\mu| = |\langle f, \psi_\mu \rangle| \leq \|f\|_\infty \|\psi_\mu\|_1 \leq B \|f\|_\infty 2^{-(1+\alpha)j/2}. \quad (12)$$

Not surprisingly, this a-priori estimate is not yet sufficient to prove Theorem 4.1. A more sophisticated result is the following theorem.

**Theorem 4.2.** Let  $\theta_N^*$  denote the (in modulus)  $N$ -th largest curvelet coefficient. Then there exists some universal constant  $C$  such that

$$\sup_{f \in \mathcal{E}^\beta(\mathbb{R}^2)} |\theta_N^*| \leq C \cdot N^{-(1+\beta)/2} \cdot (\log_2 N)^{(1+\beta)/2}.$$

*Proof.* Let  $M_j \subset M$  denote the indices corresponding to curvelets at scale  $j$ , and for  $\varepsilon > 0$  put

$$M_{j,\varepsilon} = \left\{ \mu \in M_j, |\theta_\mu| > \varepsilon \right\}.$$

By Theorem 4.3, which is stated and proved below, we have for  $\varepsilon > 0$

$$\#M_{j,\varepsilon} = \#\left\{ \mu \in M_j, |\theta_\mu| > \varepsilon \right\} \lesssim \varepsilon^{-2/(1+\beta)}. \quad (13)$$

On the other hand, (12) shows that there is a constant  $B$ , independent of scale, such that

$$|\theta_\mu| \leq B \|f\|_\infty 2^{-(1+\alpha)j/2}.$$

It follows that for each  $\varepsilon > 0$  there is  $j_\varepsilon$  such that at scales  $j \geq j_\varepsilon$  the coefficients satisfy  $\theta_\mu < \varepsilon$ . Hence, for  $j \geq j_\varepsilon$

$$\#M_{j,\varepsilon} = \#\left\{ \mu \in M_j, |\theta_\mu| > \varepsilon \right\} = 0.$$

The number of scales at which  $M_{j,\varepsilon}$  is nonempty is therefore bounded by

$$\frac{2}{1+\alpha} \left( \log_2(B) + \log_2(\|f\|_\infty) + \log_2(\varepsilon^{-1}) \right) \lesssim \log_2(\varepsilon^{-1}). \quad (14)$$

It follows from (13) and (14) that there is a constant  $\tilde{C} \geq 1$  such that

$$\#\left\{ \mu \in M, |\theta_\mu| > \varepsilon \right\} = \sum_j \#\left\{ \mu \in M_j, |\theta_\mu| > \varepsilon \right\} \leq \tilde{C} \varepsilon^{-2/(1+\beta)} \log_2(\varepsilon^{-1}).$$

Let  $\theta_N^*$  be the  $N$ -th largest coefficient. Then for  $\varepsilon_N > \delta_N$ , where  $\delta_N$  satisfies  $N = \tilde{C} \delta_N^{-2/(1+\beta)} \log_2(\delta_N^{-1})$ , we have  $|\theta_N^*| \leq \varepsilon_N$ . If  $N \geq 2$  it holds  $\tilde{C} N^{4/(1+\beta)} \log_2(N^2) \geq N$ , because  $1 \leq \beta \leq 2$  and  $\tilde{C} \geq 1$ . For  $N \geq 2$  therefore

$$N^{4/(1+\beta)} \log_2(N^2) \geq \delta_N^{-2/(1+\beta)} \log_2(\delta_N^{-1}).$$

This implies  $\delta_N \geq N^{-2}$ , and we can conclude that  $\varepsilon_N > \delta_N$  if we choose  $\varepsilon_N$  as the solution of

$$N = \tilde{C} \varepsilon_N^{-2/(1+\beta)} \log_2(N^2).$$

This choice leads to

$$\varepsilon_N = (2\tilde{C})^{(1+\beta)/2} \cdot N^{-(1+\beta)/2} (\log_2 N)^{(1+\beta)/2},$$

which proves our claim with constant  $C = (2\tilde{C})^{(1+\beta)/2}$ .  $\square$

In fact, Theorem 4.2 is strong enough to deduce Theorem 4.1 via Lemma 2.3.

*Proof of Theorem 4.1.* Applying Lemma 2.3 and Theorem 4.2 we can estimate

$$\|f - f_N\|^2 \lesssim \sum_{m>N} |\theta_m^*|^2 \lesssim \sum_{m>N} m^{-(1+\beta)} \cdot (\log_2 m)^{(1+\beta)} \lesssim \int_N^\infty t^{-(1+\beta)} \cdot (\log_2 t)^{(1+\beta)} dt.$$

Using partial integration we obtain

$$\begin{aligned} \int_N^\infty t^{-(1+\beta)} \cdot (\log_2 t)^{(1+\beta)} dt &\lesssim [-t^{-\beta} (\log_2 t)^{(1+\beta)}]_N^\infty + \int_N^\infty t^{-(1+\beta)} \cdot (\log_2 t)^\beta dt \\ &\lesssim N^{-\beta} (\log_2 N)^{(1+\beta)} + \int_N^\infty t^{-(1+\beta)} \cdot (\log_2 t)^{\lceil \beta \rceil} dt \\ &\lesssim \dots \lesssim N^{-\beta} (\log_2 N)^{(1+\beta)} + \int_N^\infty t^{-(1+\beta)} dt \\ &\lesssim N^{-\beta} (\log_2 N)^{(1+\beta)}. \end{aligned}$$

$\square$

It remains to prove Theorem 4.3, which is the main building block in the proof of Theorem 4.2. By abuse of notation, we write  $\theta_j$  for the subsequence  $(\theta_\mu)_{\mu \in M_j}$  of curvelet coefficients, where  $M_j$  denotes the curvelet indices at scale  $j$ . Theorem 4.3 then reads as follows.

**Theorem 4.3.** *The sequence  $\theta_j$  obeys*

$$\|\theta_j\|_{\omega\ell_{2/(1+\beta)}} \leq C,$$

for some constant  $C$  independent of scale.

For the analysis of the coefficient sequence  $\theta_j$  we follow the technique in [3] and smoothly decompose  $f$  into so-called *fragments*, which can then be analyzed separately. For that we cover  $\mathbb{R}^2$  at each scale  $j \in \mathbb{N}_0$  with cubes

$$Q = [(k_1 - 1)2^{-j\alpha}, (k_1 + 1)2^{-j\alpha}] \times [(k_2 - 1)2^{-j\alpha}, (k_2 + 1)2^{-j\alpha}], \quad j \in \mathbb{N}, (k_1, k_2) \in \mathbb{Z}^2,$$

which we collect in the sets  $\mathcal{Q}_j$ . Further, we put  $\mathcal{Q} := \bigcup_{j \in \mathbb{N}} \mathcal{Q}_j$ . Note how the size of the squares depends upon the scale  $2^{-j}$ : The ‘width’ of the curvelets at scale  $j$  obeys  $\sim 2^{-j}$  and the ‘length’ of the curvelets is approximately  $\sim 2^{-\alpha j}$ . Thus, the size of the squares is about the length of the curvelets.

Next, we take a smooth partition of unity  $(\omega_Q)_{Q \in \mathcal{Q}_j}$ , where these squares are used as the index set and the functions  $\omega_Q$  are supported in the corresponding squares  $Q := (2^{-j\alpha}k_1, 2^{-j\alpha}k_2) + [-2^{-j\alpha}, 2^{-j\alpha}]^2$ . More precisely, for some fixed nonnegative  $C^\infty$ -function  $\omega$  vanishing outside the square  $[-1, 1]^2$ , we put  $\omega_Q = \omega(2^{j\alpha}x_1 - k_1, 2^{j\alpha}x_2 - k_2)$  and assume that  $\sum_{Q \in \mathcal{Q}_j} \omega_Q(x) \equiv 1$ . The function  $f$  can then at each scale  $j \in \mathbb{N}_0$  be smoothly localized into the fragments

$$f_Q := f\omega_Q, \quad Q \in \mathcal{Q}_j.$$

For  $Q \in \mathcal{Q}_j$  let  $\theta_Q$  denote the curvelet coefficient sequence of  $f_Q$  at scale  $j$ , i.e.,

$$\theta_Q = \left( \langle f_Q, \psi_\mu \rangle \right)_{\mu \in M_j}. \quad (15)$$

The strategy laid out in [3] is to analyze the sparsity of the sequences  $\theta_Q$  and combine these results to obtain Theorem 4.3. In this investigation we have to distinguish between two cases: Either the square  $Q \in \mathcal{Q}_j$  meets the edge curve  $\Gamma = \partial\mathcal{B}$  or not. Accordingly, we let  $\mathcal{Q}_j^0$  be the subset of  $\mathcal{Q}_j$  containing those cubes, which intersect the edge curve  $\Gamma$ . Among the remaining cubes of  $\mathcal{Q}_j$  we collect those, which intersect  $\text{supp } f$ , in  $\mathcal{Q}_j^1$ . The others can be neglected, because they lead to trivial sequences  $\theta_Q$ .

The following two propositions directly lead to Theorem 4.3.

**Proposition 4.4.** *Let  $Q$  be a square such that  $Q \in \mathcal{Q}_j^0$ . The curvelet coefficient sequence  $\theta_Q$  obeys*

$$\|\theta_Q\|_{\omega\ell_{2/(1+\beta)}} \leq C \cdot 2^{-(1+\alpha)j/2},$$

for some constant  $C$  independent of  $Q$  and  $j$ .

**Proposition 4.5.** *Let  $Q$  be a square such that  $Q \in \mathcal{Q}_j^1$ . The curvelet coefficient sequence  $\theta_Q$  obeys*

$$\|\theta_Q\|_{\omega\ell_{2/(1+\beta)}} \leq C \cdot 2^{-(1+\alpha)j},$$

for some constant  $C$  independent of  $Q$  and  $j$ .

Theorem 4.3 is an easy consequence of these two results and the observation, that there are constants  $A_0$  and  $A_1$ , independent of scale, such that

$$\#\mathcal{Q}_j^0 \leq A_0 2^{\alpha j} \quad \text{and} \quad \#\mathcal{Q}_j^1 \leq A_1 (2^{2\alpha j} + 4 \cdot 2^{\alpha j}). \quad (16)$$

The estimates (16) hold true since  $f$  is supported in  $[0, 1]^2 \subset [-1, 1]^2$ .

*Proof of Theorem 4.3.* For  $0 < p \leq 1$  we have the  $p$ -triangle inequality

$$\|a + b\|_{\omega\ell_p}^p \leq \|a\|_{\omega\ell_p}^p + \|b\|_{\omega\ell_p}^p, \quad a, b \in \omega\ell_p.$$

Since  $\theta_j = \sum_{Q \in \mathcal{Q}_j} \theta_Q$ , we can conclude

$$\|\theta_j\|_{\omega \ell_{2/(1+\beta)}}^{2/(1+\beta)} \leq \sum_{Q \in \mathcal{Q}_j} \|\theta_Q\|_{\omega \ell_{2/(1+\beta)}}^{2/(1+\beta)} \leq (\#\mathcal{Q}_j^0) \cdot \sup_{\mathcal{Q}_j^0} \|\theta_Q\|_{\omega \ell_{2/(1+\beta)}}^{2/(1+\beta)} + (\#\mathcal{Q}_j^1) \cdot \sup_{\mathcal{Q}_j^1} \|\theta_Q\|_{\omega \ell_{2/(1+\beta)}}^{2/(1+\beta)}.$$

The claim follows now from the above two propositions together with observation (16).  $\square$

It remains to prove Propositions 4.4 and 4.5. For that let  $Q \in \mathcal{Q}_j$  be a fixed cube at a fixed scale  $j \in \mathbb{N}_0$ , which nontrivially intersects  $\text{supp } f$ . We need to analyze the decay of the sequence  $\theta_Q = (\langle f_Q, \psi_\mu \rangle)_{\mu \in M_j}$ . Since the frame elements  $\psi_\mu$  are bandlimited, it is advantageous to turn to the Fourier side. The Plancherel identity yields

$$\langle f_Q, \psi_\mu \rangle = \langle \widehat{f}_Q, \widehat{\psi}_\mu \rangle.$$

These scalar products can be estimated, if we have knowledge about the localization of the functions  $\widehat{f}_Q$ . This investigation is carried out separately in Subsections 4.2 and 4.3 for the cases  $Q \in \mathcal{Q}_j^0$  and  $Q \in \mathcal{Q}_j^1$ , respectively.

## 4.2 Analysis of a Smooth Fragment

Let us first treat the case  $Q \in \mathcal{Q}_j^1$ , where the cube  $Q$  does not intersect the edge curve  $\Gamma = \partial\mathcal{B}$ . In this case we call  $f_Q = f\omega_Q$  a *smooth fragment*.

Before we begin, we briefly recall our setting. The parameters  $\alpha \in [\frac{1}{2}, 1)$  and  $\beta = \alpha^{-1} \in (1, 2]$  are fixed, as is  $f \in \mathcal{E}^\beta(\mathbb{R}^2)$ . Since  $Q$  does not intersect the edge curve, there is a function  $g \in C^\beta(\mathbb{R}^2)$  such that  $f_Q = g\omega_Q$ . By smoothly cutting  $g$  off outside the square  $[-1, 1]^2$ , we can even assume  $g \in C_0^\beta(\mathbb{R}^2)$ .

We want to analyze  $\widehat{f}_Q$  and for simplicity we look at the following model situation. Without loss of generality we assume that the cube  $Q$  is centered at the origin, by possibly translating the coordinate system. In this case the smooth fragment takes the simple form

$$f_Q = g\omega(2^{\alpha j} \cdot),$$

where  $g \in C_0^\beta(\mathbb{R}^2)$  and  $\omega \in C_0^\infty(\mathbb{R}^2)$ ,  $\text{supp } \omega \subset [-1, 1]^2$ , is the fixed window generating the partition of unity  $(\omega_Q)_Q$ . By rescaling we obtain for each scale  $j$  the functions

$$F(x) = g(2^{-\alpha j} x)\omega(x), \quad x \in \mathbb{R}^2, \quad (17)$$

with  $\text{supp } F \subset [-1, 1]^2$ . We put  $g_j(x) := g(2^{-\alpha j} x)$ , so that we can write  $F(x) = g_j(x)\omega(x)$ . It is important to note, that  $g_j$  and  $F$  depend on the scale, whereas  $\omega$  remains fixed.

Since the following investigation takes place in the model situation, the explicit reference to the cube  $Q$  is not necessary in this exposition. Therefore, by abuse of notation, we will write henceforth  $f$  for the standard segment  $f_Q$  and accordingly  $\theta$  for the sequence  $\theta_Q$ .

### 4.2.1 Fourier Analysis of a Smooth Fragment

We first analyze the localization of  $\widehat{F}$ . The key result in this direction will be Theorem 4.7. Its proof relies on Lemma 4.6 below.

**Lemma 4.6.** *Assume that  $h = C2^{-(1-\alpha)j}$  for some fixed constant  $C > 0$ . We then have*

$$\|\Delta_{(h,0)}^3 \partial_1 F\|_2^2 \lesssim h^{2\beta} 2^{-2j\alpha},$$

where the implicit constant is independent of the scale  $j$ . Notice that  $h$  is not independent and depends on the scale  $j$ .

*Proof.* Since  $\text{supp } F \subset [-1, 1]^2$  it suffices to prove

$$\|\Delta_{(h,0)}^3 \partial_1 F\|_\infty \lesssim h^\beta 2^{-j\alpha}. \quad (18)$$

By the product rule we have  $\partial_1 F = \partial_1 g_j \cdot \omega + g_j \cdot \partial_1 \omega$  and it holds

$$\begin{aligned}\Delta_{(h,0)}^3(\partial_1 g_j \cdot \omega) &= \sum_{k=0}^3 \Delta_{(h,0)}^k \partial_1 g_j \cdot \Delta_{(h,0)}^{3-k} \omega(\cdot + kh, \cdot), \\ \Delta_{(h,0)}^3(g_j \cdot \partial_1 \omega) &= \sum_{k=0}^3 \Delta_{(h,0)}^k g_j \cdot \Delta_{(h,0)}^{3-k} \partial_1 \omega(\cdot + kh, \cdot).\end{aligned}$$

Clearly, we have for every  $k \in \mathbb{N}$  the estimates  $\|\Delta_{(h,0)}^k \omega\|_\infty \lesssim h^k$  and  $\|\Delta_{(h,0)}^k \partial_1 \omega\|_\infty \lesssim h^k$ . Further, according to Lemma A.3, it holds  $\|\partial_1 g_j\|_\infty \lesssim 2^{-\alpha j}$ ,  $\|\Delta_{(h,0)} g_j\|_\infty \lesssim h 2^{-\alpha j}$ , and

$$\begin{aligned}\|\Delta_{(h,0)}^k g_j\|_\infty &\lesssim h^\beta 2^{-j} \quad \text{for } k \geq 2, \\ \|\Delta_{(h,0)}^k \partial_1 g_j\|_\infty &\lesssim h^\beta 2^{-\alpha j} \quad \text{for } k \geq 1.\end{aligned}$$

Since  $h^3 \lesssim h^\beta 2^{-j\alpha}$  for  $h = C 2^{-(1-\alpha)j}$  and  $\alpha \in [\frac{1}{2}, 1)$ , the assertion (18) follows.  $\square$

The previous lemma is key to the proof of the following theorem. Here we use the notation  $|\xi| \sim 2^{(1-\alpha)j}$  to indicate  $|\xi| \in [C_1 2^{(1-\alpha)j}, C_2 2^{(1-\alpha)j}]$  for some arbitrary but fixed constants  $0 < C_1 \leq C_2 < \infty$ . A typical choice would be  $C_1 = 1$  and  $C_2 = 2^{1-\alpha}$ .

**Theorem 4.7.** *It holds independent of the scale  $j$*

$$\int_{|\xi| \sim 2^{(1-\alpha)j}} |\widehat{F}(\xi)|^2 d\xi \lesssim 2^{-2\beta j}.$$

*Proof.* Let  $0 < C_1 \leq C_2 < \infty$  be fixed and choose  $C > 0$  such that  $C_2 C < 2\pi$ . Putting  $h := C 2^{-(1-\alpha)j}$ , there then exists  $c > 0$  such that  $|e^{i\xi_1 h} - 1|^2 \geq c$  for every  $\xi_1$  with  $|\xi_1| \in [C_1 2^{(1-\alpha)j}, C_2 2^{(1-\alpha)j}]$ . Using Lemma 4.6 we then estimate the integrals on the vertical strips:

$$\begin{aligned}2^{2(1-\alpha)j} \int_{|\xi_1| \sim 2^{(1-\alpha)j}} \int_{\xi_2} |\widehat{F}(\xi_1, \xi_2)|^2 d\xi_2 d\xi_1 &\asymp \int_{|\xi_1| \sim 2^{(1-\alpha)j}} \int_{\xi_2} |\xi_1|^2 |\widehat{F}(\xi_1, \xi_2)|^2 d\xi_2 d\xi_1 \\ &\lesssim \int_{|\xi_1| \sim 2^{(1-\alpha)j}} \int_{\xi_2} |e^{i\xi_1 h} - 1|^6 |\xi_1|^2 |\widehat{F}(\xi_1, \xi_2)|^2 d\xi_2 d\xi_1 \\ &= \int_{|\xi_1| \sim 2^{(1-\alpha)j}} \int_{\xi_2} |\widehat{\Delta_{(h,0)}^3 \partial_1 F}(\xi_1, \xi_2)|^2 d\xi_2 d\xi_1 \\ &\leq \int_{\mathbb{R}^2} |\widehat{\Delta_{(h,0)}^3 \partial_1 F}(\xi)|^2 d\xi = \|\Delta_{(h,0)}^3 \partial_1 F\|_2^2 \lesssim h^{2\beta} 2^{-2j\alpha}.\end{aligned}$$

Interchanging  $\xi_1$  and  $\xi_2$  yields analogous estimates for the horizontal strips. Altogether, we obtain

$$\int_{|\xi| \sim 2^{(1-\alpha)j}} |\widehat{F}(\xi)|^2 d\xi \lesssim 2^{-2j(1-\alpha)} h^{2\beta} 2^{-2j\alpha} \asymp 2^{-2\beta j}.$$

$\square$

As an immediate corollary, we deduce a corresponding estimate for the original smooth fragment  $f$ .

**Theorem 4.8.** *We have independent of scale  $j$*

$$\int_{|\xi| \sim 2^j} |\widehat{f}(\xi)|^2 d\xi \lesssim 2^{-2(\beta+\alpha)j}.$$

*Proof.* The statement follows from the relation  $\widehat{f}(\xi) = 2^{-2\alpha j} \widehat{F}(2^{-\alpha j} \xi)$ .  $\square$

Finally, we state a refinement of Theorem 4.8.

**Corollary 4.9.** *Let  $m = (m_1, m_2) \in \mathbb{N}_0^2$  and  $\partial^m = \partial_1^{m_1} \partial_2^{m_2}$ . We have*

$$\int_{|\xi| \sim 2^j} |\partial^m \widehat{f}(\xi)|^2 d\xi \lesssim 2^{-2j\alpha|m|} 2^{-2(\beta+\alpha)j}.$$

*Proof.* Recall that  $f = g\omega(2^{\alpha j} \cdot)$ . Let us define the window  $\omega_m(x) := x^m \omega(x)$  and the function  $f_m(x) := g(x)\omega_m(2^{\alpha j}x)$  for  $x \in \mathbb{R}^2$ . Then for every  $x \in \mathbb{R}^2$

$$x^m f(x) = g(x)x^m \omega(2^{\alpha j}x) = 2^{-j\alpha|m|} g(x)\omega_m(2^{\alpha j}x) = 2^{-j\alpha|m|} f_m(x).$$

We conclude with Theorem 4.8

$$\int_{|\xi| \sim 2^j} |\partial^m \widehat{f}(\xi)|^2 d\xi = \int_{|\xi| \sim 2^j} |\widehat{x^m f(x)}(\xi)|^2 d\xi = 2^{-2j\alpha|m|} \int_{|\xi| \sim 2^j} |\widehat{f_m}(\xi)|^2 d\xi \lesssim 2^{-2j\alpha|m|} 2^{-2(\beta+\alpha)j}.$$

□

#### 4.2.2 Curvelet Analysis of a Smooth Fragment

Let  $J$  be a scale-angle pair and  $\chi_0$  and  $\chi_J$  the functions from (8), used in the construction of the  $\alpha$ -curvelet frame  $\mathcal{C}_\alpha(W^{(0)}, W, V)$ . We remark, that  $\chi_J$  is generally not a characteristic function. However,  $\chi_J$  is a non-negative real-valued function, supported in the wedges  $\mathcal{W}_J$  given in (9) and satisfying  $\|\chi_J\|_\infty \leq 1$ .

Theorem 4.8 directly leads to a central result, namely that it holds

$$\int_{\mathbb{R}^2} \sum_{|J|=j} |(\widehat{f\chi_J})(\xi)|^2 d\xi \lesssim 2^{-2j(\beta+\alpha)}. \quad (19)$$

Our next goal, is to refine this result. Let us first record a basic fact.

**Lemma 4.10.** *Let  $m \in \mathbb{N}_0^2$ . It holds for all  $\xi \in \mathbb{R}^2$*

$$\sum_{|J|=j} |\partial^m \chi_J(\xi)|^2 \lesssim 2^{-2j\alpha|m|}.$$

*Proof.* From the definition it follows that  $\chi_J$  scales with  $2^{-\alpha j}$  in one direction and with  $2^{-j}$  in the orthogonal direction. No matter what direction, we always do better than  $|\partial^m \chi_J(\xi)|^2 \lesssim 2^{-2j\alpha|m|}$ . For fixed  $\xi$  only a fixed number of summands are not zero, uniformly for all  $\xi$ . The claim follows. □

Next we prove an auxiliary lemma. Here  $\Delta = \partial_1^2 + \partial_2^2$  denotes the standard Laplacian.

**Lemma 4.11.** *It holds for  $m \in \mathbb{N}_0$*

$$\int \sum_{|J|=j} |\Delta^m (\widehat{f\chi_J})(\xi)|^2 d\xi \lesssim 2^{-2j(\beta+\alpha)} \cdot 2^{-4m\alpha j}.$$

*Proof.* For  $m = 0$  this is just (19), a direct consequence of Theorem 4.8. Now let  $m > 0$ . It holds with  $a, b \in \mathbb{N}_0^2$  and certain coefficients  $c_{a,b} \in \mathbb{N}_0$

$$\Delta^m (\widehat{f\chi_J})(\xi) = \sum_{|a|+|b|=2m} c_{a,b} \partial^a \widehat{f}(\xi) \partial^b \chi_J(\xi).$$

Let  $a, b \in \mathbb{N}_0^2$  such that  $|a| + |b| = 2m$ . Then with Lemma 4.10 and Corollary 4.9

$$\begin{aligned} \int \sum_{|J|=j} |\partial^a \widehat{f}(\xi)|^2 |\partial^b \chi_J(\xi)|^2 d\xi &\lesssim 2^{-2j\alpha|b|} \int_{|\xi| \sim 2^j} |\partial^a \widehat{f}(\xi)|^2 d\xi \\ &\lesssim 2^{-2j\alpha|b|} 2^{-2j\alpha|a|} 2^{-2(\beta+\alpha)j} \\ &= 2^{-2j\alpha(|a|+|b|)} 2^{-2(\beta+\alpha)j} = 2^{-4j\alpha m} 2^{-2(\beta+\alpha)j}. \end{aligned}$$

□



Now we come to the refinement of (19). For that, we need the differential operator

$$\mathcal{L} = \mathcal{I} - 2^{2\alpha j} \Delta,$$

where  $\mathcal{I}$  is the identity and  $\Delta$  the standard Laplacian. The theorem below shows that  $\mathcal{L}^2(\widehat{f}_{\chi_J})$  obeys the same estimate (19) as  $\widehat{f}_{\chi_J}$ .

**Theorem 4.12.** *It holds*

$$\int_{\mathbb{R}^2} \sum_{|J|=j} |\mathcal{L}^2(\widehat{f}_{\chi_J})(\xi)|^2 d\xi \lesssim 2^{-2j(\beta+\alpha)}$$

*Proof.* It holds

$$\mathcal{L}^2 = \mathcal{I} - 2 \cdot 2^{2\alpha j} \Delta + 2^{4\alpha j} \Delta^2.$$

Applying (19) and Lemma 4.11 yields the desired result.  $\square$

Finally we can give the proof of Proposition 4.5.

*Proof of Proposition 4.5.* Recall the curvelet frame  $\mathcal{C}_\alpha(W^{(0)}, W, V) = (\psi_\mu)_{\mu \in M}$ . On the Fourier side

$$\widehat{\psi}_{j,\ell,k} = \chi_J u_{j,k}(R_J \cdot),$$

with rotation matrix  $R_J$  given as in (7) and functions

$$u_{j,k}(\xi) = 2^{-j(1+\alpha)/2} e^{2\pi i(2^{-j}k_1, 2^{-\alpha j}k_2) \cdot \xi}, \quad \xi \in \mathbb{R}^2.$$

The curvelet coefficients  $(\theta_\mu)_{\mu \in M}$  of  $f = f_Q$  are therefore given by the formula

$$\theta_\mu = \langle f, \psi_{j,\ell,k} \rangle = \int_{\mathbb{R}^2} \widehat{f}_{\chi_J}(\xi) \overline{u_{j,k}(R_J \xi)} d\xi.$$

We have to study the decay of the subsequence  $\theta = \theta_Q$  defined in (15). We observe

$$\mathcal{L}u_{j,k} = (1 + 2^{-2j(1-\alpha)}k_1^2 + k_2^2)u_{j,k},$$

which also holds for the rotated versions  $u_{j,k}(R_J \cdot)$ . Partial integration thus yields

$$\theta_\mu = \int_{\mathbb{R}^2} \widehat{f}(\xi) \chi_J(\xi) \overline{u_{j,k}(R_J \xi)} d\xi = (1 + 2^{-2j(1-\alpha)}k_1^2 + k_2^2)^{-2} \int_{\mathbb{R}^2} \mathcal{L}^2(\widehat{f}_{\chi_J}) \overline{u_{j,k}(R_J \xi)} d\xi.$$

For  $K = (K_1, K_2) \in \mathbb{Z}^2$  we define the set

$$\mathfrak{Z}_K := \left\{ (k_1, k_2) \in \mathbb{Z}^2 : k_1 2^{-j(1-\alpha)} \in [K_1, K_1 + 1), k_2 = K_2 \right\}.$$

Further, we put

$$M_{j,K} := \left\{ \mu = (j, \ell, k) \in M_j : k \in \mathfrak{Z}_K \right\},$$

where  $M_j$  denotes the curvelet indices at scale  $j$ . It follows from the orthogonality properties of the Fourier system  $(u_{j,k})_{k \in \mathbb{Z}^2}$  that

$$\sum_{k \in \mathfrak{Z}_K} |\theta_\mu|^2 \leq (1 + |K|^2)^{-4} \int_{\mathbb{R}^2} |\mathcal{L}^2(\widehat{f}_{\chi_J})(\xi)|^2 d\xi,$$

where  $|K|^2 = K_1^2 + K_2^2$ . We further conclude

$$\sum_{M_{j,K}} |\theta_\mu|^2 = \sum_{|J|=j} \sum_{k \in \mathfrak{Z}_K} |\theta_\mu|^2 \leq (1 + |K|^2)^{-4} \int_{\mathbb{R}^2} \sum_{|J|=j} |\mathcal{L}^2(\widehat{f}_{\chi_J})(\xi)|^2 d\xi.$$

Now we apply Theorem 4.12 and obtain

$$\sum_{M_{j,K}} |\theta_\mu|^2 \lesssim 2^{-2j(\beta+\alpha)} (1 + |K|^2)^{-4},$$

which directly implies

$$\|(\theta_\mu)_{\mu \in M_{j,K}}\|_{\ell^2} \lesssim 2^{-j(\beta+\alpha)} (1 + |K|^2)^{-2}. \quad (20)$$

It holds  $\#\mathfrak{Z}_K \leq 1 + 2^{j(1-\alpha)}$  and therefore, since  $L_j = 2^{\lfloor j(1-\alpha) \rfloor}$ , the estimate  $\#M_{j,K} \leq 2 \cdot 2^{2j(1-\alpha)}$ . Now we recall the interpolation inequality  $\|(c_\lambda)\|_{\ell^p} \leq n^{1/p-1/2} \|(c_\lambda)\|_{\ell_2}$  for a finite sequence  $(c_\lambda)_\lambda$  with  $n$  nonzero entries. Applying this inequality with  $p = 2/(1+\beta)$  we get from (20)

$$\|(\theta_\mu)_{\mu \in M_{j,K}}\|_{\ell^{2/(1+\beta)}} \lesssim 2^{j(\beta-1)} \|(\theta_\mu)_{\mu \in M_{j,K}}\|_{\ell^2} \leq 2^{-j(1+\alpha)} (1 + |K|^2)^{-2}.$$

It follows

$$\sum_{\mu \in M_j} |\theta_\mu|^{2/(1+\beta)} \lesssim 2^{-j(1+\alpha)2/(1+\beta)} \cdot (1 + |K|^2)^{-4/(1+\beta)} = 2^{-2\alpha j} (1 + |K|^2)^{-4/(1+\beta)}.$$

Finally, we have

$$\sum_{\mu \in M_j} |\theta_\mu|^{2/(1+\beta)} = \sum_{K \in \mathbb{Z}^2} \sum_{\mu \in M_{j,K}} |\theta_\mu|^{2/(1+\beta)} \leq 2^{-2\alpha j} \sum_{K \in \mathbb{Z}^2} (1 + |K|^2)^{-4/(1+\beta)} \lesssim 2^{-2\alpha j}.$$

The desired estimate for  $\theta = \theta_Q$  follows, i.e.

$$\|\theta_Q\|_{\ell_{2/(1+\beta)}} \lesssim 2^{-j(\alpha+1)}.$$

□

The following subsection is devoted to the proof of Proposition 4.4.

### 4.3 Analysis of an Edge Fragment

Let us turn to the more complicated case  $Q \in \mathcal{Q}_j^0$  and the proof of Proposition 4.4. In this case the cube  $Q$  intersects the edge curve  $\Gamma$  and  $f_Q = f\omega_Q$  is accordingly called an *edge fragment*.

In order to prove Proposition 4.4 we need to analyze the decay of the sequence  $\theta_Q = (\langle f_Q, \psi_\mu \rangle)_{\mu \in M_j} = (\langle \widehat{f}_Q, \widehat{\psi}_\mu \rangle)_{\mu \in M_j}$ . To estimate these scalar products, we again study the localization of the function  $\widehat{f}_Q$ . As in the treatment of the smooth fragments, our investigation starts with some simplifying reductions.

Firstly, we note that it suffices to prove Proposition 4.4 for a function  $f \in \mathcal{E}^\beta(\mathbb{R}^2)$  of the form  $f = g\chi_{\mathcal{B}}$  with  $g \in C_0^\beta(\mathbb{R}^2)$ . In fact, the curvelet coefficient sequence of a general cartoon  $f = f^{(0)} + f^{(1)} = g_0 \cdot \chi_{\mathcal{B}} + g_1$  can be decomposed into  $\theta_Q^{(0)} = (\langle f_Q^{(0)}, \psi_\mu \rangle)_{\mu \in M_j}$  and  $\theta_Q^{(1)} = (\langle f_Q^{(1)}, \psi_\mu \rangle)_{\mu \in M_j}$ . From Proposition 4.5 we already know

$$\|\theta_Q^{(1)}\|_{\omega \ell_{2/(1+\beta)}} \lesssim 2^{-(1+\alpha)j} \lesssim 2^{-(1+\alpha)j/2}.$$

Therefore it only remains to show  $\|\theta_Q^{(0)}\|_{\omega \ell_{2/(1+\beta)}} \lesssim 2^{-(1+\alpha)j/2}$ . Since  $\mathcal{B} \subset [-1, 1]^2$ , we can further smoothly cut off  $g_0$  outside of  $[-1, 1]^2$  to obtain a function  $g \in C_0^\beta(\mathbb{R}^2)$  such that  $f^{(0)} = g\chi_{\mathcal{B}}$ .

Secondly, without loss of generality we restrict to the following model situation. The cube  $Q$  is centered at the origin and the edge curve  $\Gamma$  is the graph of a function  $E : [-2^{-j\alpha}, 2^{-j\alpha}] \rightarrow [-2^{-j\alpha}, 2^{-j\alpha}]$  belonging to  $C^\beta(\mathbb{R})$ , with  $x_1 = E(x_2)$ . Further, it shall hold  $E(0) = E'(0) = 0$ , so that  $\Gamma$  approximates a vertical line through the origin. If the scale  $j$  is big enough, say bigger than some fixed base scale  $j_0 \in \mathbb{N}_0$ , it is always possible to arrive at this setting by possibly translating or rotating the coordinate axes. Henceforth, we assume  $j \in \mathbb{N}$  and  $j \geq j_0$  which clearly poses no loss of generality.

In this simplified model situation the edge fragment  $f = f_Q$  can be written in the form

$$f(x) = \omega(2^{\alpha j} x) g(x) \chi_{\{x_1 \geq E(x_2)\}}, \quad x = (x_1, x_2) \in \mathbb{R}^2,$$

where  $g \in C_0^\beta(\mathbb{R}^2)$ , and  $\omega \in C_0^\infty(\mathbb{R}^2)$  is the nonnegative window with  $\text{supp } \omega \subset [-1, 1]^2$ , generating the partition of unity  $(\omega_Q)_Q$ . For our investigation it is again more convenient to work with a rescaled version  $F$  of the edge fragment. Therefore we put  $g_j = g(2^{-\alpha j} \cdot)$  and define

$$F(x) := \omega(x)g_j(x)\chi_{\{x_1 \geq E_j(x_2)\}}, \quad x = (x_1, x_2) \in \mathbb{R}^2, \quad (21)$$

with the rescaled edge function

$$E_j : [-1, 1] \rightarrow [-1, 1], \quad E_j(x_2) = 2^{\alpha j} E(2^{-\alpha j} x_2).$$

It holds  $E_j \in C^\beta([-1, 1])$  with  $E'_j = E'(2^{-\alpha j} \cdot)$  and  $\text{Höl}(E'_j, \beta - 1) \leq \delta_j$ , where

$$\delta_j = 2^{-j(1-\alpha)} \cdot \text{Höl}(E', \beta - 1).$$

Observe that  $\text{Höl}(E', \beta - 1)$  is a constant independent of the scale  $j$ . Together with  $E_j(0) = E'_j(0) = 0$  this implies for all  $u \in [-1, 1]$  that

$$|E_j(u)| \leq \delta_j \quad \text{and} \quad |E'_j(u)| \leq \delta_j. \quad (22)$$

For convenience, we continuously extend the function  $E_j$  to the whole of  $\mathbb{R}$  by attaching straight lines on the left and on the right, with constant slopes  $E'_j(1)$  and  $E'_j(-1)$  respectively. Since this extension occurs outside of the square  $[-1, 1]^2$ , it does not change the representation (21) of the edge fragment. Furthermore, it also does not alter the regularity and the Hölder constant.

#### 4.3.1 Fourier Analysis of an Edge Fragment

Our first goal is to analyze the Fourier transform  $\widehat{F}$  (and thus also  $\widehat{f}$ ) along radial lines, whose orientations are specified by angles  $\eta \in [-\pi/2, \pi/2]$  with respect to the  $x_1$ -axis. If the angle  $\eta$  satisfies  $|\sin \eta| > \delta_j$ , it is possible because of (22) to define a function  $u = u_j(\cdot, \eta) : \mathbb{R} \rightarrow \mathbb{R}$  implicitly by

$$E_j(u(t)) \cos \eta + u(t) \sin \eta = t. \quad (23)$$

The value  $u(t)$  is the  $x_2$ -coordinate of the intersection point of the (extended) edge curve  $\Gamma$  and the line  $\mathfrak{L}_{t,\eta}$  defined by

$$\mathfrak{L}_{t,\eta} := \left\{ x = (x_1, x_2) \in \mathbb{R}^2 : x_1 \cos \eta + x_2 \sin \eta = t \right\}. \quad (24)$$

Further, we can define the function  $a = a_j(\cdot, \eta) : \mathbb{R} \rightarrow \mathbb{R}$  by

$$a(t) := -E_j(u(t)) \sin \eta + u(t) \cos \eta. \quad (25)$$

The value  $a(t)$  is the  $x_2$ -coordinate of the point  $(E_j(u), u)^T \in \Gamma$  in the coordinate system rotated by the angle  $\eta$ . For an illustration we refer to Figure 4.3.1.

The functions  $u$  and  $a$  are strictly monotone, increasing if  $\eta > 0$  and decreasing if  $\eta < 0$ . Note, that we suppressed the dependence of  $u$  and  $a$  on  $j$  in the notation. The following lemma studies the regularity of  $u$  under the assumption  $|\sin \eta| \geq 2\delta_j$ .

**Lemma 4.13.** *Assume  $|\sin \eta| \geq 2\delta_j$ . Then the function  $u : \mathbb{R} \rightarrow \mathbb{R}$  defined implicitly by (23) belongs to  $C^\beta(\mathbb{R})$ . Moreover, we have  $\|u'\|_\infty \lesssim |\sin \eta|^{-1}$  and*

$$\|\Delta_h u'\|_\infty \lesssim \delta_j h^{\beta-1} |\sin \eta|^{-1-\beta},$$

where the implicit constants are independent of the scale  $j$ , the angle  $\eta$ , and  $h \geq 0$ .

*Proof.* First of all it is not difficult to show that  $u = u_j(\cdot, \eta) \in C^1(\mathbb{R})$  with

$$u'(t) = (\sin \eta + E'_j(u(t)) \cos \eta)^{-1}.$$

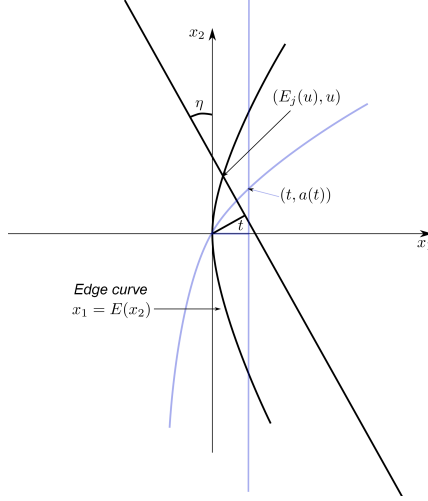


Figure 1: Illustration of standard edge fragment.

Under the assumption  $|\sin \eta| \geq 2\delta_j$  it follows  $\|u'\|_\infty \lesssim |\sin \eta|^{-1}$  because of  $|E'_j(u)| \leq \delta_j \leq \frac{1}{2}|\sin \eta|$  for all  $u \in [-1, 1]$ . Finally, we examine  $\Delta_h u'$ . For  $t \in \mathbb{R}$

$$\begin{aligned} \Delta_h u'(t) &= u'(t+h) - u'(t) = u'(t+h)u'(t)(u'(t)^{-1} - u'(t+h)^{-1}) \\ &= u'(t+h)u'(t) \cos \eta (E'_j(u(t)) - E'_j(u(t+h))). \end{aligned}$$

Using Höl( $E'_j, \beta - 1$ )  $\leq \delta_j$  and the mean value theorem leads to

$$\|\Delta_h u'\|_\infty \leq \|u'\|_\infty^2 \delta_j \|\Delta_h u\|_\infty^{\beta-1} \leq \|u'\|_\infty^{\beta+1} \delta_j h^{\beta-1} \lesssim \delta_j h^{\beta-1} |\sin \eta|^{-1-\beta}.$$

□

The following lemma collects some properties of the function  $a : \mathbb{R} \rightarrow \mathbb{R}$  defined in (25).

**Lemma 4.14.** *Assume  $|\sin \eta| \geq 2\delta_j$ . It holds  $a \in C^\beta(\mathbb{R})$  with*

$$\|a'\|_\infty \lesssim |\sin \eta|^{-1}, \quad \|\Delta_h a\|_\infty \lesssim h |\sin \eta|^{-1}, \quad \|\Delta_h a'\|_\infty \lesssim \delta_j h^{\beta-1} |\sin \eta|^{-1-\beta},$$

with implicit constants independent of  $j, \eta$ , and  $h \geq 0$ .

*Proof.* This is an easy consequence of the properties of  $u$  proved in the previous lemma. □

Next, we introduce the scale-dependent interval

$$I(\eta) := I_j(\eta) := [a_j(\eta), b_j(\eta)], \tag{26}$$

where  $a_j(\eta) = E_j(-1) \cos \eta - \sin \eta$  and  $b_j(\eta) = E_j(1) \cos \eta + \sin \eta$ . The restrictions of  $u$  and  $a$  to  $I(\eta)$  correspond precisely to that part of the edge curve  $\Gamma$  lying inside the square  $[-1, 1]^2$ . In particular, we have a bijection  $u : I(\eta) \rightarrow [-1, 1]$ . In the sequel it is more convenient to work with an extension of  $I(\eta)$ , given by

$$\tilde{I}(\eta) := \tilde{I}_j(\eta) := [a_j(\eta) - C\delta_j, b_j(\eta) + C\delta_j],$$

for some suitable fixed constant  $C > 0$ .

**Lemma 4.15.** *For  $|\sin \eta| > \delta_j$  we have*

$$|I(\eta)| \lesssim |\sin \eta| \quad \text{and} \quad |\tilde{I}(\eta)| \lesssim |\sin \eta|.$$

*Proof.* In view of  $|\sin \eta| > \delta_j$  and (22) we can estimate

$$|I(\eta)| \leq |E_j(1) - E_j(-1)| |\cos \eta| + 2|\sin \eta| \leq 2\delta_j + 2|\sin \eta| \lesssim |\sin \eta|.$$

The estimate for  $\tilde{I}(\eta)$  then follows directly from  $|\sin \eta| > \delta_j$ .  $\square$

We want to analyze  $\hat{F}$  along lines through the origin with orientation  $\eta \in [-\pi/2, \pi/2]$ . The central tool in this investigation is the Fourier slice theorem. In view of this theorem it makes sense to first study the Radon transform  $\mathcal{R}F$ , in particular its regularity. By Paley-Wiener type arguments we can then later extract information about the decay of  $\hat{F}$ . This is the same approach taken in [3]. Due to the lack of regularity in our case, however, we have to use a more refined technique in this investigation. The main idea is to use finite differences instead of derivatives.

The value  $\mathcal{R}F(t, \eta)$  of the Radon transform is obtained by integrating  $F$  along the line  $\mathfrak{L}_{t, \eta}$  defined in (24). Rotating  $F$  by the angle  $\eta$  yields the function  $F^\eta$  and we can write

$$(\mathcal{R}F)(t, \eta) = \int_{\mathbb{R}} F^\eta(t, u) du.$$

The rescaled edge fragment  $F$  can be rewritten as the product  $F = G\chi_{\{x_1 \geq E_j(x_2)\}}$  with the function

$$G := \omega g(2^{-\alpha j} \cdot) = \omega g_j.$$

Then we have

$$(\mathcal{R}F)(t, \eta) = \int_{-\infty}^{a(t, \eta)} G^\eta(t, u) du, \quad (27)$$

where  $G^\eta$  is the function obtained by rotating  $G$  by the angle  $\eta$ . Using the notation  $g_j^\eta$  and  $\omega^\eta$  for the rotated versions of  $g_j$  and  $\omega$ , the integrand of (27) takes the form

$$G^\eta = g_j^\eta \omega^\eta.$$

We see, that the component  $g_j^\eta = g^\eta(2^{-\alpha j} \cdot) \in C_0^\beta(\mathbb{R}^2)$  of  $G^\eta$  is scaled and the window  $\omega^\eta \in C_0^\infty(\mathbb{R}^2)$  remains fixed. During the following investigation the angle  $\eta$  remains constant. Therefore, we can simplify the notation by omitting the index  $\eta$ .

The central lemma of this subsection is given below. Its proof relies on estimates of the functions  $g_j^\eta$  and  $\omega^\eta$ , outsourced to the appendix.

**Lemma 4.16.** *Assume that  $|\sin \eta| \geq 2\delta_j$ . For  $h = C2^{-(1-\alpha)j}$ , where  $C > 0$  is some fixed constant, we then have*

$$\Delta_h \partial_1 \mathcal{R}F(t, \eta) = S_1(t, \eta) + S_2(t, \eta)$$

with

$$\begin{aligned} \|S_1(\cdot, \eta)\|_2^2 &\lesssim \delta_j^2 h^{2(\beta-1)} |\sin \eta|^{-1-2\beta}, \\ \|\Delta_{(h,0)} S_2(\cdot, \eta)\|_2^2 &\lesssim h^{2\beta} |\sin \eta|^{-1-2\beta}, \end{aligned}$$

where the implicit constants are independent of the scale  $j$  and the angle  $\eta$ .

*Proof.* First we differentiate  $\mathcal{R}F(t, \eta)$  with respect to  $t$  and obtain from (27)

$$\partial_1(\mathcal{R}F)(t, \eta) = a'(t)G(t, a(t)) + \int_{-\infty}^{a(t)} \partial_1 G(t, u) du =: T(t)$$

where on the right-hand side the dependence on  $\eta$  is omitted in the notation. Applying  $\Delta_h$  then yields for  $t \in \mathbb{R}$

$$\begin{aligned} \Delta_h T(t) &= \Delta_h a'(t)G(t+h, a(t+h)) + a'(t)\Delta_h G(t, a(t)) + \int_{a(t)}^{a(t+h)} \partial_1 G(t+h, u) du + \int_{-\infty}^{a(t)} \Delta_{(h,0)} \partial_1 G(t, u) du \\ &=: T_1(t) + T_2(t) + T_3(t) + T_4(t). \end{aligned}$$

Next, we estimate the  $L^\infty$ -norms of the functions  $T_i$  for  $i \in \{1, 2, 3, 4\}$ . Let us begin with  $T_1$ . Applying Lemma 4.14 we obtain

$$\|T_1\|_\infty \leq \|\Delta_h a'\|_\infty \|G\|_\infty \lesssim \|\Delta_h a'\|_\infty \lesssim \delta_j h^{\beta-1} |\sin \eta|^{-1-\beta} \lesssim h^\beta |\sin \eta|^{-1-\beta}.$$

The estimate of  $T_2$  takes some more effort. The product rule yields for  $t \in \mathbb{R}$

$$T_2(t) = a'(t) \Delta_h G(t, a(t)) = a'(t) \Delta_h g_j(t, a(t)) \omega(t+h, a(t+h)) + a'(t) g_j(t, a(t)) \Delta_h \omega(t, a(t)) =: T_{21}(t) + T_{22}(t).$$

Using the mean value theorem and Lemmas 4.14 and A.4 yields

$$\|T_{21}\|_\infty \leq \|a'\|_\infty \sup_{t \in \mathbb{R}} |\Delta_h g_j(t, a(t))| \|\omega\|_\infty \lesssim h |\sin \eta|^{-2} 2^{-\alpha j} \lesssim h^\beta |\sin \eta|^{-1-\beta}.$$

We take another forward difference of the component  $T_{22}$  and obtain

$$\begin{aligned} \Delta_h T_{22}(t) &= \Delta_h a'(t) g_j(t+h, a(t+h)) \Delta_h \omega(t+h, a(t+h)) \\ &\quad + a'(t) \Delta_h g_j(t, a(t)) \Delta_h \omega(t+h, a(t+h)) + a'(t) g(t, a(t)) \Delta_h^2 \omega(t, a(t)) \\ &=: T_{22}^1(t) + T_{22}^2(t) + T_{22}^3(t). \end{aligned}$$

These terms allow the following estimates, where we use Lemmas 4.14, A.6 and A.4. Also note  $h \lesssim |\sin \eta|$ .

$$\begin{aligned} \|T_{22}^1\|_\infty &\leq h^{\beta+1} |\sin \eta|^{-2-\beta} \lesssim h^\beta |\sin \eta|^{-1-\beta}, \\ \|T_{22}^2\|_\infty &\leq h^2 |\sin \eta|^{-3} 2^{-\alpha j} \lesssim h^\beta |\sin \eta|^{-1-\beta}, \\ \|T_{22}^3\|_\infty &\leq h^2 |\sin \eta|^{-3} + h^{\beta+1} |\sin \eta|^{-2-\beta} \lesssim h^\beta |\sin \eta|^{-1-\beta}. \end{aligned}$$

By substitution the term  $T_3$  transforms to

$$\begin{aligned} T_3(t) &= \int_t^{t+h} \partial_1 G(t+h, a(u)) a'(u) du \\ &= \int_t^{t+h} \partial_1 g_j(t+h, a(u)) \omega(t+h, a(u)) a'(u) du + \int_t^{t+h} g_j(t+h, a(u)) \partial_1 \omega(t+h, a(u)) a'(u) du \\ &=: T_{31}(t) + T_{32}(t) \end{aligned}$$

We apply  $\Delta_h$  to  $T_{31}$ . Here  $\Delta_h$  acts exclusively on  $t$  and  $\tau$ . We obtain

$$\begin{aligned} \Delta_h T_{31}(t) &= \int_t^{t+h} \Delta_h [\partial_1 g_j(t+h, a(\tau)) \omega(t+h, a(\tau)) a'(\tau)] d\tau \\ &= \int_t^{t+h} \Delta_h \partial_1 g_j(t+h, a(\tau)) \omega(t+2h, a(\tau+h)) a'(\tau+h) d\tau \\ &\quad + \int_t^{t+h} \partial_1 g_j(t+h, a(\tau)) \Delta_h \omega(t+h, a(\tau)) a'(\tau+h) d\tau \\ &\quad + \int_t^{t+h} \partial_1 g_j(t+h, a(\tau)) \omega(t+h, a(\tau)) \Delta_h a'(\tau) d\tau \\ &=: T_{31}^1(t) + T_{31}^2(t) + T_{31}^3(t). \end{aligned}$$

Analogously, we decompose

$$\Delta_h T_{32}(t) = \int_t^{t+h} \Delta_h [g_j(t+h, a(\tau)) \partial_1 \omega(t+h, a(\tau)) a'(\tau)] d\tau =: T_{32}^1(t) + T_{32}^2(t) + T_{32}^3(t).$$

Then we estimate with the results from the appendix

$$\begin{aligned} \|T_{31}^1\|_\infty &\lesssim h |\sin \eta|^{-1} 2^{-j} (h^{\beta-1} + h^{\beta-1} |\sin \eta|^{1-\beta}) \lesssim h^\beta |\sin \eta|^{-1-\beta}, \\ \|T_{31}^2\|_\infty &\lesssim h |\sin \eta|^{-1} 2^{-\alpha j} (h + h |\sin \eta|^{-1}) \lesssim h^\beta |\sin \eta|^{-1-\beta}, \\ \|T_{31}^3\|_\infty &\lesssim h 2^{-\alpha j} \|\Delta_h a'\|_\infty \lesssim h^\beta |\sin \eta|^{-1-\beta}, \end{aligned}$$

and

$$\begin{aligned}\|T_{32}^1\|_\infty &\lesssim h|\sin\eta|^{-1}2^{-\alpha j}(h+h|\sin\eta|^{-1}) \lesssim h^\beta|\sin\eta|^{-1-\beta}, \\ \|T_{32}^2\|_\infty &\lesssim h|\sin\eta|^{-1}(h+h|\sin\eta|^{-1}) \lesssim h^\beta|\sin\eta|^{-1-\beta}, \\ \|T_{32}^3\|_\infty &\lesssim h\|\Delta_h a'\|_\infty \lesssim h^\beta|\sin\eta|^{-1-\beta}.\end{aligned}$$

Finally, we treat the term  $T_4$ ,

$$\begin{aligned}T_4(t) &= \int_{-\infty}^{a(t)} \Delta_h \partial_1 G(t, u) du = \int_{-\infty}^{a(t)} \Delta_h (\partial_1 g_j(t, u)\omega(t, u) + g_j(t, u)\partial_1 \omega(t, u)) du \\ &= \int_{-\infty}^{a(t)} \Delta_h \partial_1 g_j(t, u)\omega(t+h, u) du + \int_{-\infty}^{a(t)} (\partial_1 g_j(t, u)\Delta_h \omega(t, u) + \Delta_h g_j(t, u)\partial_1 \omega(t+h, u)) du \\ &\quad + \int_{-\infty}^{a(t)} g_j(t, u)\Delta_h \partial_1 \omega(t, u) du =: T_{41}(t) + T_{42}(t) + T_{43}(t).\end{aligned}$$

The terms  $T_{41}$  and  $T_{42}$  can be estimated directly,

$$\begin{aligned}\|T_{41}\|_\infty &\lesssim h^{\beta-1} \cdot 2^{-j} \leq h^\beta, \\ \|T_{42}\|_\infty &\lesssim h \cdot 2^{-\alpha j} \asymp 2^{-j} \leq 2^{-j(\beta-1)} \asymp h^\beta.\end{aligned}$$

The term  $T_{43}$  again needs some further preparation,

$$\Delta_h T_{43}(t) = \int_{a(t)}^{a(t+h)} g_j(t+h, u)\Delta_h \partial_1 \omega(t+h, u) du + \int_{-\infty}^{a(t)} \Delta_h (g_j(t, u)\Delta_h \partial_1 \omega(t, u)) du =: T_{43}^1(t) + T_{43}^2(t).$$

In the end we arrive at

$$\begin{aligned}\|T_{43}^1\|_\infty &\lesssim h^2|\sin\eta|^{-1} \lesssim h^\beta|\sin\eta|^{-1}, \\ \|T_{43}^2\|_\infty &\lesssim h^2 \lesssim h^\beta.\end{aligned}$$

Now we collect the appropriate terms and add them up to obtain  $S_1$  and  $S_2$ . In a last step, we use our  $L^\infty$ -estimates to obtain the desired  $L^2$ -estimates. Here we use that  $|\text{supp } T_i| \lesssim |I(\eta)| \lesssim |\sin\eta|$  according to Lemma 4.15 for  $i \in \{1, 2, 3\}$  and  $|\text{supp } T_4| \lesssim 1$ . This finishes the proof.  $\square$

The previous lemma is the key to the following theorem. The notation  $|\lambda| \sim 2^{(1-\alpha)j}$  indicates that  $|\lambda| \in [C_1 2^{(1-\alpha)j}, C_2 2^{(1-\alpha)j}]$  for fixed constants  $0 < C_1 \leq C_2 < \infty$ . A typical choice would be  $C_1 = 1$  and  $C_2 = 2^{1-\alpha}$ .

**Theorem 4.17.** *It holds*

$$\int_{|\lambda| \sim 2^{(1-\alpha)j}} |\widehat{F}(\lambda \cos \eta, \lambda \sin \eta)|^2 d\lambda \lesssim 2^{-(1-\alpha)j} \left(1 + 2^{(1-\alpha)j} |\sin \eta|\right)^{-1-2\beta}$$

with an implicit constant independent of  $j$  and  $\eta$ .

*Proof.* First we assume  $|\sin \eta| \geq 2\delta_j$ . The integration domain is given by  $[C_1 2^{(1-\alpha)j}, C_2 2^{(1-\alpha)j}]$  for fixed constants  $0 < C_1 \leq C_2 < \infty$ . Let us fix  $h := C 2^{-j(1-\alpha)}$ , where  $C > 0$  is chosen such that  $C_2 C < 2\pi$ . For this choice of  $h$  there is  $c > 0$  such that  $|e^{i\lambda h} - 1|^2 \geq c$  for all  $|\lambda| \in [C_1 2^{(1-\alpha)j}, C_2 2^{(1-\alpha)j}]$  at all scales.

We conclude, where  $S_i$ ,  $i = 1, 2$ , denote the entities from Lemma 4.16:

$$\begin{aligned}
\int_{|\lambda| \sim 2^{j(1-\alpha)}} |\lambda|^2 |\widehat{F}(\lambda \cos \eta, \lambda \sin \eta)|^2 d\lambda &\lesssim \int_{|\lambda| \sim 2^{j(1-\alpha)}} |e^{i\lambda h} - 1|^2 |\lambda|^2 |\widehat{F}(\lambda \cos \eta, \lambda \sin \eta)|^2 d\lambda \\
&= \int_{|\lambda| \sim 2^{j(1-\alpha)}} |\mathcal{F}[\Delta_h \partial_1 \mathcal{R}F(\cdot, \eta)](\lambda)|^2 d\lambda \\
&\lesssim \sum_{i=1}^2 \int_{|\lambda| \sim 2^{j(1-\alpha)}} |\widehat{S_i(\cdot, \eta)}(\lambda)|^2 d\lambda \\
&\lesssim \int_{|\lambda| \sim 2^{j(1-\alpha)}} |e^{i\lambda h} - 1|^2 |\widehat{S_2(\cdot, \eta)}(\lambda)|^2 d\lambda + \int_{|\lambda| \sim 2^{j(1-\alpha)}} |\widehat{S_1(\cdot, \eta)}(\lambda)|^2 d\lambda \\
&\leq \int_{\mathbb{R}} |e^{i\lambda h} - 1|^2 |\widehat{S_2(\cdot, \eta)}(\lambda)|^2 d\lambda + \int_{\mathbb{R}} |\widehat{S_1(\cdot, \eta)}(\lambda)|^2 d\lambda \\
&= \|S_1(\cdot, \eta)\|_2^2 + \|\Delta_h S_2(\cdot, \eta)\|_2^2 \lesssim h^{2\beta} |\sin \eta|^{-1-2\beta}.
\end{aligned}$$

It follows

$$\int_{|\lambda| \sim 2^{j(1-\alpha)}} |\widehat{F}(\lambda \cos \eta, \lambda \sin \eta)|^2 d\lambda \lesssim 2^{-j(1-\alpha)} (2^{j(1-\alpha)} |\sin \eta|)^{-1-2\beta}.$$

Next, we handle the case  $|\sin \eta| < 2\delta_j$ . We want to show

$$\int_{|\lambda| \sim 2^{j(1-\alpha)}} |\widehat{F}(\lambda \cos \eta, \lambda \sin \eta)|^2 d\lambda \lesssim 2^{-j(1-\alpha)}. \tag{28}$$

Altogether we then obtain the desired estimate

$$\int_{|\lambda| \sim 2^{(1-\alpha)j}} |\widehat{F}(\lambda \cos \eta, \lambda \sin \eta)|^2 d\lambda \lesssim 2^{-(1-\alpha)j} \left(1 + 2^{(1-\alpha)j} |\sin \eta|\right)^{-1-2\beta},$$

since  $2^{(1-\alpha)j} |\sin \eta| \gtrsim 1$  if  $|\sin \eta| \geq 2\delta_j$  and  $2^{(1-\alpha)j} |\sin \eta| \lesssim 1$  if  $|\sin \eta| < 2\delta_j$ ,

It remains to show (28). For this we write the edge fragment as a sum  $F = F_0 + D$ , where

$$F_0(x) = g(2^{-j\alpha} x) \omega(x) \chi_{\{x_1 \geq \delta_j\}}, \quad x \in \mathbb{R}^2,$$

is a fragment with a straight edge and  $D(x) = F(x) - F_0(x)$  is the deviation.

The function  $D$  is supported in a vertical strip around the  $x_2$ -axis of width  $2\delta_j$ . For  $\eta$  satisfying  $|\sin \eta| < 2\delta_j$  the Radon transform  $\mathcal{R}D(\cdot, \eta)$  is  $L^\infty$ -bounded and supported in an interval of length

$$2(\delta_j \cos \eta + \sin \eta) \lesssim \delta_j.$$

It follows  $\|\mathcal{R}D(\cdot, \eta)\|_2^2 \lesssim \delta_j \lesssim 2^{-j(1-\alpha)}$ , and therefore

$$\int_{|\lambda| \sim 2^{j(1-\alpha)}} |\widehat{D}(\lambda \cos \eta, \lambda \sin \eta)|^2 d\lambda \leq \int_{\mathbb{R}} |\widehat{\mathcal{R}D(\cdot, \eta)}(\lambda)|^2 d\lambda \lesssim 2^{-j(1-\alpha)}.$$

Finally, the estimate

$$\int_{|\lambda| \sim 2^{j(1-\alpha)}} |\widehat{F_0}(\lambda \cos \eta, \lambda \sin \eta)|^2 d\lambda \lesssim 2^{-j(1-\alpha)}$$

follows from the fact, that we have decay  $|\widehat{F_0}(\lambda, 0)| \sim |\lambda|^{-1/2}$  normal to the straight singularity curve, and that the second argument  $(\lambda \sin \eta)$  remains bounded due to the condition  $|\sin \eta| < 2\delta_j$ . This finishes the proof.  $\square$



A direct consequence is the following corollary.

**Corollary 4.18.** *We have*

$$\int_{|\lambda| \sim 2^j} |\widehat{f}(\lambda \cos \eta, \lambda \sin \eta)|^2 d\lambda \lesssim 2^{-(1+2\alpha)j} \left(1 + 2^{(1-\alpha)j} |\sin \eta|\right)^{-1-2\beta}.$$

*Proof.* The statement follows directly from the relation  $\widehat{f}(\xi) = 2^{-2\alpha j} \widehat{F}(2^{-\alpha j} \xi)$ .  $\square$

### 4.3.2 Refinement of the Previous Discussion

In this subsection we will prove a refinement of Corollary 4.18. For that we need to analyze the modified edge fragment  $\widetilde{F}$ , given for fixed  $m \in \mathbb{N}_0$  by

$$\widetilde{F}(x) = r(x)^m F(x), \quad x \in \mathbb{R}^2, \quad (29)$$

where the function  $r : \mathbb{R}^2 \rightarrow \mathbb{R}$  maps  $x \in \mathbb{R}^2$  to its first component  $x_1 \in \mathbb{R}$ . It can be written as the product  $\widetilde{F}(x) = \widetilde{G}(x) \chi_{\{x_1 \geq E_j(x_2)\}}$  with the function

$$\widetilde{G}(x) := r(x)^m G(x) = r(x)^m \omega(x) g_j(x), \quad x \in \mathbb{R}^2. \quad (30)$$

Rotating by the angle  $\eta$  yields  $\widetilde{G}^\eta(x) = (r^\eta(x))^m G^\eta(x)$ , where  $G^\eta$  and  $r^\eta$  are the functions obtained by rotating  $G$  and  $r$ , respectively. The function  $r^\eta : \mathbb{R}^2 \rightarrow \mathbb{R}$  then has the form

$$r^\eta(t, a) := t \cos \eta - a \sin \eta, \quad (t, a) \in \mathbb{R}^2.$$

Note, that since  $G^\eta(x) = g_j^\eta(x) \omega^\eta(x)$  we have  $\widetilde{G}^\eta(x) = (r^\eta(x))^m g_j^\eta(x) \omega^\eta(x)$ . Some important properties of  $r^\eta$  and  $\widetilde{G}^\eta$  are collected in the appendix.

Similar to the investigation of  $F$  in the previous subsection, the angle  $\eta$  remains fixed in the remainder. We can therefore again simplify the notation by omitting this index. The following result generalizes Lemma 4.16.

**Lemma 4.19.** *For  $m \in \mathbb{N}_0$  let  $\widetilde{F}$  be the modified edge fragment (29). Further, assume that  $|\sin \eta| \geq 2\delta_j$  and  $h \asymp 2^{-(1-\alpha)j}$ . Then the function  $S := \Delta_h \partial_1 \mathcal{R} \widetilde{F}(\cdot, \eta)$  admits a decomposition*

$$\begin{aligned} S &= S_1^0 + S_2^0, \\ \text{such that} \quad \Delta_h S_2^0 &= S_1^1 + S_2^1, \\ \Delta_h S_2^1 &= S_1^2 + S_2^2, \\ &\vdots \\ \Delta_h S_2^{m-1} &= S_1^m + S_2^m, \\ \Delta_h S_2^m &= S_1^{m+1}, \end{aligned}$$

with the estimates

$$\|S_1^k\|_2^2 \lesssim 2^{-2jm(1-\alpha)} h^{2\beta} |\sin \eta|^{-1-2\beta} + 2^{-j(1-\alpha)(2\beta+1)}, \quad k = 0, 1, \dots, m+1.$$

For convenience we set  $S_2^{m+1} = 0$ .

We introduce the following language and say, that the function  $S$  admits a decomposition  $(S_1^k, S_2^k)_k$  of the form (\*) of length  $m+1$  with the estimates

$$\|S_1^k\|_2^2 \lesssim 2^{-2jm(1-\alpha)} h^{2\beta} |\sin \eta|^{-1-2\beta} + 2^{-j(1-\alpha)(2\beta+1)}, \quad k = 0, 1, \dots, m+1.$$

*Proof.* We have  $\widetilde{G}^\eta(x) = r^\eta(x)^m G^\eta(x)$  for  $x \in \mathbb{R}^2$  and

$$\mathcal{R} \widetilde{F}(t, \eta) = \int_{-\infty}^{a(t, \eta)} \widetilde{G}^\eta(t, u) du.$$

For simplicity we omit the superindex  $\eta$  subsequently. Similar to the proof of Lemma 4.16 we obtain

$$\begin{aligned} S(t) &= \Delta_h a'(t) \tilde{G}(t+h, a(t+h)) + a'(t) \Delta_h \tilde{G}(t, a(t)) + \int_{a(t)}^{a(t+h)} \partial_1 \tilde{G}(t+h, u) du + \int_{-\infty}^{a(t)} \Delta_{(h,0)} \partial_1 \tilde{G}(t, u) du \\ &=: \tilde{T}_1(t) + \tilde{T}_2(t) + \tilde{T}_3(t) + \tilde{T}_4(t). \end{aligned}$$

We will show the assertion for each of these terms separately. Moreover, it suffices to prove  $L^\infty$ -estimates, which can be transformed to the desired  $L^2$ -estimates via the corresponding support properties. Note that  $|\text{supp } \tilde{T}_i| \lesssim |\tilde{I}(\eta)| \lesssim |\sin \eta|$  according to Lemma 4.15 for  $i \in \{1, 2, 3\}$  and that  $|\text{supp } \tilde{T}_4| \lesssim 1$ .

For  $\tilde{T}_1$  the estimate

$$\begin{aligned} \|\tilde{T}_1\|_\infty &\leq \|\Delta_h a'\|_\infty \sup_{t \in \mathbb{R}} |\tilde{G}(t, a(t))| \lesssim \|\Delta_h a'\|_\infty \sup_{t \in I(\eta)} |r(t, a(t))|^m \\ &\lesssim \delta_j^{m+1} h^{\beta-1} |\sin \eta|^{-1-\beta} \lesssim h^m h^\beta |\sin \eta|^{-1-\beta} \end{aligned}$$

is sufficient. Next, we show that  $\tilde{T}_2$  and  $\tilde{T}_3$  admit decompositions  $(\tilde{T}_1^k, \tilde{T}_2^k)_k$  of the form (\*) of length  $m+1$  with  $\text{supp } \tilde{T}_i^k \subset \tilde{I}(\eta)$ ,  $i \in \{1, 2\}$ , and the estimates

$$\begin{aligned} \|\tilde{T}_1^k\|_\infty &\lesssim h^m h^\beta |\sin \eta|^{-1-\beta}, \quad k = 0, \dots, m+1, \\ \|\tilde{T}_2^k\|_\infty &\lesssim h^m |\sin \eta|^{-1}, \quad k = 0, \dots, m. \end{aligned}$$

The decomposition of the component  $\tilde{T}_2$  is provided by Lemma A.10. Let us turn to  $\tilde{T}_3$ . By substitution this term transforms to

$$\tilde{T}_3(t) = \int_t^{t+h} \partial_1 \tilde{G}(t+h, a(u)) a'(u) du.$$

We put  $\tilde{T}_1^0 = 0$  and  $\tilde{T}_2^0 = \tilde{T}_3$ . These terms clearly satisfy the assertions. Next we take the forward difference of  $\tilde{T}_2^0$ . Here  $\Delta_h$  acts on both  $t$  and  $\tau$ . We obtain

$$\begin{aligned} \Delta_h \tilde{T}_2^0(t) = \Delta_h \tilde{T}_3(t) &= \int_t^{t+h} \Delta_h (\partial_1 \tilde{G}(t+h, a(\tau)) a'(\tau)) d\tau \\ &= \int_t^{t+h} \Delta_h \partial_1 \tilde{G}(t+h, a(\tau)) a'(\tau) d\tau \\ &\quad + \int_t^{t+h} \partial_1 \tilde{G}(t+2h, a(\tau+h)) \Delta_h a'(\tau) d\tau \\ &=: \tilde{T}_{31}(t) + \tilde{T}_{32}(t). \end{aligned}$$

Lemma A.12 then yields a decomposition  $(\tilde{S}_1^k, \tilde{S}_2^k)_k$ , such that we can write

$$\Delta_h \partial_1 \tilde{G}(t+h, a(\tau)) a'(\tau) = \tilde{S}_1^0(t, \tau) + \tilde{S}_2^0(t, \tau).$$

This leads to

$$\tilde{T}_{31}(t) = \int_t^{t+h} \Delta_h \partial_1 \tilde{G}(t+h, a(\tau)) a'(\tau) d\tau = \int_t^{t+h} \tilde{S}_1^0(t, \tau) d\tau + \int_t^{t+h} \tilde{S}_2^0(t, \tau) d\tau.$$

We put  $\tilde{T}_1^1(t) := \tilde{T}_{32}(t) + \int_t^{t+h} \tilde{S}_1^0(t, \tau) d\tau$  and  $\tilde{T}_2^1(t) := \int_t^{t+h} \tilde{S}_2^0(t, \tau) d\tau$ . These terms  $\tilde{T}_1^1$  and  $\tilde{T}_2^1$  then satisfy the requirements, i.e.,

$$\begin{aligned} \|\tilde{T}_1^1\|_\infty &\lesssim \|\tilde{T}_{32}\|_\infty + h \cdot \sup_{t \in \mathbb{R}} \sup_{\tau \in [t, t+h]} |\tilde{S}_1^0(t, \tau)| \lesssim h \cdot \delta_j^{m-1} \cdot h^{\beta-1} \delta_j |\sin \eta|^{-1-\beta}, \\ \|\tilde{T}_2^1\|_\infty &\lesssim h \sup_{t \in \mathbb{R}} \sup_{\tau \in [t, t+h]} |\tilde{S}_2^0(t, \tau)| \lesssim h^m |\sin \eta|^{-1}. \end{aligned}$$

Taking another forward difference of  $\tilde{T}_2^1$  yields

$$\Delta_h \tilde{T}_2^1(t) = \int_t^{t+h} \Delta_h S_2^0(t, \tau) d\tau.$$

Proceeding inductively from here with Lemma A.12 settles the claim for the component  $\tilde{T}_3$ .

Finally, we turn to the function  $\tilde{T}_4(t) = \int_{-\infty}^{a(t)} \Delta_{(h,0)} \partial_1 \tilde{G}(t, u) du$ . First, we calculate

$$\Delta_h \tilde{T}_4(t) = \int_{a(t)}^{a(t+h)} \Delta_{(h,0)} \partial_1 \tilde{G}(t+h, u) du + \int_{-\infty}^{a(t)} \Delta_{(h,0)}^2 \partial_1 \tilde{G}(t, u) du =: \tilde{T}_{41}(t) + \tilde{T}_{42}(t)$$

$$\text{and } \Delta_h \tilde{T}_{42}(t) = \int_{a(t)}^{a(t+h)} \Delta_{(h,0)}^2 \partial_1 \tilde{G}(t+h, u) du + \int_{-\infty}^{a(t)} \Delta_{(h,0)}^3 \partial_1 \tilde{G}(t, u) du =: \tilde{T}_{43}(t) + \tilde{T}_{44}(t).$$

Next, we show  $\|\tilde{T}_{44}\|_\infty \lesssim h^{\beta+\frac{1}{2}}$  because then, in view of  $|\text{supp } \tilde{T}_{44}| \asymp 1$ ,

$$\|\tilde{T}_{44}\|_2^2 \lesssim h^{2\beta+1} \asymp 2^{-j(1-\alpha)(2\beta+1)}.$$

The  $L^\infty$ -estimate of the term  $\tilde{T}_{44}$  relies on the fact, that for  $h \asymp 2^{-j(1-\alpha)}$

$$\|\Delta_{(h,0)}^3 \partial_1 \tilde{G}\|_\infty \lesssim h^\beta 2^{-\alpha j} \lesssim h^{\beta+1}.$$

This estimate is a consequence of Lemmas A.3, A.5, and A.7 and is analogous to (18). Essential is the observation that since  $\alpha \geq \frac{1}{2}$  Lemma A.3 yields

$$\|\Delta_{(h,0)} \partial_1 g_j\|_\infty \lesssim 2^{-\alpha j} h^\beta \lesssim h^{\beta+1}.$$

Finally, we take care of the remaining terms  $\tilde{T}_{41}$  and  $\tilde{T}_{43}$ . First we note that  $|\text{supp } \tilde{T}_{41}| \lesssim |\tilde{I}(\eta)| \lesssim |\sin \eta|$  and also  $|\text{supp } \tilde{T}_{43}| \lesssim |\tilde{I}(\eta)| \lesssim |\sin \eta|$ . Hence, it suffices to prove  $\|\tilde{T}_{41}\|_\infty \lesssim h^m h^\beta |\sin \eta|^{-1-\beta}$  and  $\|\tilde{T}_{43}\|_\infty \lesssim h^m h^\beta |\sin \eta|^{-1-\beta}$ . It holds

$$\|\tilde{T}_{41}\|_\infty \leq \sup_{t \in \mathbb{R}} \left| \int_{a(t)}^{a(t+h)} \partial_1 \tilde{G}(t+2h, u) du \right| + \sup_{t \in \mathbb{R}} \left| \int_{a(t)}^{a(t+h)} \partial_1 \tilde{G}(t+h, u) du \right|.$$

Analogously, we have

$$\begin{aligned} \|\tilde{T}_{43}\|_\infty &\leq \sup_{t \in \mathbb{R}} \left| \int_{a(t)}^{a(t+h)} \partial_1 \tilde{G}(t+3h, u) du \right| + 2 \sup_{t \in \mathbb{R}} \left| \int_{a(t)}^{a(t+h)} \partial_1 \tilde{G}(t+2h, u) du \right| \\ &\quad + \sup_{t \in \mathbb{R}} \left| \int_{a(t)}^{a(t+h)} \partial_1 \tilde{G}(t+h, u) du \right|. \end{aligned}$$

All these terms on the right hand side can be estimated in the same way as

$$\tilde{T}_3(t) = \int_t^{t+h} \partial_1 \tilde{G}(t+h, a(u)) a'(u) du.$$

This finishes the proof.  $\square$

The following theorem generalizes Theorem 4.17.

**Theorem 4.20.** *We have for  $m = (m_1, m_2) \in \mathbb{N}_0^2$  the estimate*

$$\int_{|\lambda| \sim 2^{j(1-\alpha)}} |\partial^m \hat{F}(\lambda \cos \eta, \lambda \sin \eta)|^2 d\lambda \lesssim 2^{-2(1-\alpha)jm_1} 2^{-(1-\alpha)j} (1 + 2^{(1-\alpha)j} |\sin \eta|)^{-1-2\beta} + 2^{3\alpha j} 2^{(-1-2\beta)j}$$

*Proof.* Note that  $\partial^m \widehat{F} = (x^m F)^\wedge$ . We define the edge fragment  $F := x_2^{m_2} F$ , with a little abuse of notation. Further, we put  $\widetilde{F} := x_1^{m_1} F$ . Analogous to Theorem 4.17 we distinguish between the cases  $|\sin \eta| < 2\delta_j$  and  $|\sin \eta| \geq 2\delta_j$ .

In case  $|\sin \eta| < 2\delta_j$  we write  $\widetilde{F} = \widetilde{F}_0 + \widetilde{D}$  with

$$\widetilde{F}_0(x) = x_1^{m_1} F_0(x) = x_1^{m_1} g(2^{-j\alpha} x) \omega(x) \chi_{\{x_1 \geq \delta_j\}} = x_1^{m_1} F(x) \chi_{\{x_1 \geq \delta_j\}}$$

and  $\widetilde{D}(x) = \widetilde{F}(x) - \widetilde{F}_0(x)$  the deviation. We also let  $D(x) = F(x) - F_0(x)$ . Note that  $\widetilde{F}_0$  is a fragment with a straight edge of height about  $\delta_j^{m_1}$ . We want to show

$$\int_{|\lambda| \sim 2^{j(1-\alpha)}} |\widehat{\widetilde{F}}(\lambda \cos \eta, \lambda \sin \eta)|^2 d\lambda \lesssim 2^{-2jm_1(1-\alpha)} 2^{-j(1-\alpha)}.$$

The function  $\widetilde{D}$  is supported in a vertical strip of width  $2\delta_j$ . For  $\eta$  satisfying  $|\sin \eta| < 2\delta_j$  the Radon transform  $\mathcal{R}\widetilde{D}(\cdot, \eta)$  is  $L^\infty$ -bounded with  $\|\mathcal{R}\widetilde{D}(\cdot, \eta)\|_\infty \lesssim \delta_j^{m_1} \|\mathcal{R}D(\cdot, \eta)\|_\infty$  and it is supported in an interval of length

$$2(\delta_j \cos \eta + \sin \eta) \lesssim \delta_j.$$

It follows  $\|\mathcal{R}D(\cdot, \eta)\|_2^2 \lesssim \delta_j^{2m_1} \delta_j \lesssim \delta_j^{2m_1} 2^{-j(1-\alpha)}$ . Therefore

$$\int_{|\lambda| \sim 2^{j(1-\alpha)}} |\widehat{\widetilde{D}}(\lambda \cos \eta, \lambda \sin \eta)|^2 d\lambda \leq \int_{\mathbb{R}} |\widehat{\mathcal{R}D(\cdot, \eta)}(\lambda)|^2 d\lambda \lesssim \delta_j^{2m_1} 2^{-j(1-\alpha)}.$$

It remains to show

$$\int_{|\lambda| \sim 2^{j(1-\alpha)}} |\widehat{\widetilde{F}_0}(\lambda \cos \eta, \lambda \sin \eta)|^2 d\lambda \lesssim \delta_j^{2m_1} 2^{-j(1-\alpha)}.$$

This follows from the fact, that we have decay  $|\widehat{\widetilde{F}_0}(\lambda, 0)| \lesssim \delta_j^{m_1} |\lambda|^{-1/2}$  normal to the straight singularity curve, since the height of the jump is  $\delta_j^{m_1}$ . Further, the second argument  $(\lambda \sin \eta)$  remains bounded due to the condition  $|\sin \eta| < 2\delta_j$ .

In case  $|\sin \eta| \geq 2\delta_j$  we conclude as follows. Let  $C_1, C_2 > 0$  be the constants specifying the integration domain  $[C_1 2^{j(1-\alpha)}, C_2 2^{j(1-\alpha)}]$ . We choose  $C > 0$  such that  $C_2 C < 2\pi$  and fix  $h := C 2^{-j(1-\alpha)}$ . Then there is  $c > 0$  such that  $|e^{i\lambda h} - 1|^{m_1} \geq c$  for  $|\lambda| \in [C_1 2^{j(1-\alpha)}, C_2 2^{j(1-\alpha)}]$  at all scales. We obtain

$$\begin{aligned} \int_{|\lambda| \sim 2^{j(1-\alpha)}} |\lambda|^2 |\partial^m \widehat{F}(\lambda \cos \eta, \lambda \sin \eta)|^2 d\lambda &\lesssim \int_{|\lambda| \sim 2^{j(1-\alpha)}} |e^{i\lambda h} - 1|^2 |\lambda|^2 |\widehat{x^m F}(\lambda \cos \eta, \lambda \sin \eta)|^2 d\lambda \\ &\lesssim \int_{|\lambda| \sim 2^{j(1-\alpha)}} |e^{i\lambda h} - 1|^2 |\lambda|^2 |(\mathcal{R}(x^m F)(\cdot, \eta))^\wedge(\lambda)|^2 d\lambda \\ &\lesssim \int_{|\lambda| \sim 2^{j(1-\alpha)}} |[\Delta_h \partial_1 \mathcal{R}\widetilde{F}(\cdot, \eta)]^\wedge(\lambda)|^2 d\lambda \end{aligned}$$

From Lemma 4.19 we know, that  $S = \Delta_h \partial_1 \mathcal{R}\widetilde{F}(\cdot, \eta)$  admits a decomposition  $(S_1^k, S_2^k)_k$  of length  $m_1$  with estimates

$$\|S_1^k\|_2^2 \lesssim 2^{-2jm_1(1-\alpha)} h^{2\beta} |\sin \eta|^{-1-2\beta} + 2^{-j(1-\alpha)(2\beta+1)}, \quad k = 0, 1, \dots, m_1 + 1.$$

Using the same trick as in Theorem 4.17 we can then conclude

$$\begin{aligned} \int_{|\lambda| \sim 2^{j(1-\alpha)}} |\lambda|^2 |\partial^m \widehat{F}(\lambda \cos \eta, \lambda \sin \eta)|^2 d\lambda &\lesssim \int_{|\lambda| \sim 2^{j(1-\alpha)}} |[\Delta_h \partial_1 \mathcal{R}\widetilde{F}(\cdot, \eta)]^\wedge(\lambda)|^2 d\lambda \\ &\lesssim \sum_{k=0}^{m_1+1} \int_{|\lambda| \sim 2^{j(1-\alpha)}} |\widehat{S_1^k}|^2 d\lambda \\ &\lesssim \sum_{k=0}^{m_1+1} \|S_1^k\|_2^2 \\ &\lesssim 2^{-2jm_1(1-\alpha)} h^{2\beta} |\sin \eta|^{-1-2\beta} + 2^{-j(1-\alpha)(2\beta+1)}. \end{aligned}$$

It follows

$$\begin{aligned} \int_{|\lambda| \sim 2^{j(1-\alpha)}} |\partial^m \widehat{F}(\lambda \cos \eta, \lambda \sin \eta)|^2 d\lambda &\lesssim 2^{-2j(1-\alpha)(m_1+1)} h^{2\beta} |\sin \eta|^{-1-2\beta} + 2^{-j(1-\alpha)(2\beta+3)} \\ &\lesssim 2^{-2(1-\alpha)jm_1} 2^{-(1-\alpha)j} (2^{(1-\alpha)j} |\sin \eta|)^{-1-2\beta} + 2^{3\alpha j} 2^{(-1-2\beta)j}. \end{aligned}$$

This finishes the proof.  $\square$

Rescaling  $F$  to the original edge fragment  $f$  yields the following corollary.

**Corollary 4.21.** *We have for  $m = (m_1, m_2) \in \mathbb{N}_0^2$  the estimate*

$$\int_{|\lambda| \sim 2^j} |\partial^m \widehat{f}(\lambda \cos \eta, \lambda \sin \eta)|^2 d\lambda \lesssim 2^{-j2\alpha|m|} \left( 2^{-j2(1-\alpha)m_1} 2^{-(1+2\alpha)j} \left( 1 + 2^{(1-\alpha)j} |\sin \eta| \right)^{-1-2\beta} + 2^{(-1-2\beta)j} \right).$$

*Proof.* The statement follows directly from the relation  $\widehat{f}(\xi) = 2^{-2\alpha j} \widehat{F}(2^{-\alpha j} \xi)$ .  $\square$

### 4.3.3 Curvelet Analysis of an Edge Fragment

In the following  $J = (j, \ell)$  shall denote a scale-angle pair with  $j \in \mathbb{N}_0$ ,  $\ell \in \{0, \dots, L_j - 1\}$ , and  $\chi_J$  and  $\chi_0$  shall be the functions from (8). Recall also the characteristic angle  $\omega_j = \pi 2^{-\lfloor j(1-\alpha) \rfloor}$  at scale  $j$  and the corresponding orientations  $\omega_J = \ell \omega_j$ , ranging between 0 and  $\pi$ . From Corollary 4.18 we can directly conclude the following result.

**Theorem 4.22.** *We have*

$$\int_{\mathbb{R}^2} |\widehat{f} \chi_J(\xi)|^2 d\xi \lesssim 2^{-(1+\alpha)j} \left( 1 + 2^{(1-\alpha)j} |\sin \omega_J| \right)^{-1-2\beta}. \quad (31)$$

*Proof.* It holds  $\|\chi_J\|_\infty \leq 1$  and  $\text{supp } \chi_J \subset \mathcal{W}_J$ , where  $\mathcal{W}_J$  is the wedge defined in (9). Let us define the intervals  $\mathcal{K}_j = \frac{1}{8\pi} [2^{j-1}, 2^{j+1}]$  and  $\mathcal{A}_J = [\omega_J - \omega_j/2, \omega_J + \omega_j/2]$ . Using Corollary 4.18 we calculate

$$\begin{aligned} \int_{\mathbb{R}^2} |\widehat{f} \chi_J(\xi)|^2 d\xi &\leq \int_{\mathcal{W}_J} |\widehat{f}(\xi)|^2 d\xi \\ &= \int_{\mathcal{K}_j} \left( \int_{\mathcal{A}_J} + \int_{\pi + \mathcal{A}_J} \right) |\widehat{f}(\lambda, \eta)|^2 \lambda d\eta d\lambda \\ &\lesssim \left( \int_{\mathcal{A}_J} + \int_{\pi + \mathcal{A}_J} \right) 2^{-(1+2\alpha)j} \left( 1 + 2^{(1-\alpha)j} |\sin \eta| \right)^{-1-2\beta} 2^j d\eta \\ &\lesssim 2^{-(1+\alpha)j} \left( 1 + 2^{(1-\alpha)j} |\sin \omega_J| \right)^{-1-2\beta}. \end{aligned}$$

$\square$

For a scale-angle pair  $J = (j, \ell)$  let us define

$$\ell_J = 1 + 2^{(1-\alpha)j} |\sin \omega_J|$$

and the differential operator

$$\mathcal{L} = (\mathcal{I} - (2^j/\ell_J)^2 \mathcal{D}_1^2)(\mathcal{I} - 2^{2\alpha j} \mathcal{D}_2^2) = \mathcal{I} - 2^{2j} \ell_J^{-2} \mathcal{D}_1^2 - 2^{2\alpha j} \mathcal{D}_2^2 + 2^{2(1+\alpha)j} \ell_J^{-2} \mathcal{D}_1^2 \mathcal{D}_2^2,$$

with identity  $\mathcal{I}$  and partial derivatives

$$\mathcal{D}_1 = \cos \omega_J \cdot \partial_1 + \sin \omega_J \cdot \partial_2 \quad \text{and} \quad \mathcal{D}_2 = -\sin \omega_J \cdot \partial_1 + \cos \omega_J \cdot \partial_2. \quad (32)$$

We will show that  $\mathcal{L}(\widehat{f} \chi_J)$  obeys the same estimate (31) as  $\widehat{f} \chi_J$ .

**Theorem 4.23.** *We have*

$$\int_{\mathbb{R}^2} |\mathcal{L}(\widehat{f}\chi_J)(\xi)|^2 d\xi \lesssim 2^{-(1+\alpha)j} \left(1 + 2^{(1-\alpha)j} |\sin \omega_J|\right)^{-1-2\beta}.$$

*Proof.* First, observe that for each pair  $m = (m_1, m_2) \in \mathbb{N}_0^2$  the mixed derivative of  $\chi_J$  obeys

$$\|\mathcal{D}_1^{m_1} \mathcal{D}_2^{m_2} \chi_J\|_\infty = \mathcal{O}(2^{-jm_1} \cdot 2^{-j\alpha m_2}). \quad (33)$$

This follows from the fact, that the functions  $\chi_J$  from (8) scale with their support wedges  $\mathcal{W}_J$ , which are of length  $\sim 2^j$  and width  $\sim 2^{\alpha j}$ .

Next, from the definition (32) of the operators  $\mathcal{D}_1$  and  $\mathcal{D}_2$  we deduce for  $m_1 \in \mathbb{N}_0$

$$\mathcal{D}_1^{m_1} \widehat{f} = \sum_{a+b=m_1} c_{a,b} (\cos \omega_J)^a (\sin \omega_J)^b \partial^{(a,b)} \widehat{f}$$

with binomial coefficients  $c_{a,b} \in \mathbb{N}$ . A similar formula holds for  $\mathcal{D}_2^{m_2} \widehat{f}$  and  $m_2 \in \mathbb{N}_0$ . Using  $|\sin \omega_J| \leq 2^{-(1-\alpha)j} \ell_J$  we obtain the estimate

$$\|\mathcal{D}_1^{m_1} \widehat{f}\|_{L^2(\mathcal{W}_J)}^2 \lesssim \sum_{a+b=m_1} |\sin \omega_J|^{2b} \cdot \|\partial^{(a,b)} \widehat{f}\|_{L^2(\mathcal{W}_J)}^2 \leq \sum_{a+b=m_1} (2^{-(1-\alpha)j} \ell_J)^{2b} \cdot \|\partial^{(a,b)} \widehat{f}\|_{L^2(\mathcal{W}_J)}^2.$$

Analogously, we obtain

$$\|\mathcal{D}_2^{m_2} \widehat{f}\|_{L^2(\mathcal{W}_J)}^2 \lesssim \sum_{a+b=m_2} \|\partial^{(a,b)} \widehat{f}\|_{L^2(\mathcal{W}_J)}^2.$$

Taking into account the width  $\sim 2^{\alpha j}$  of the wedges  $\mathcal{W}_J$ , Corollary 4.21 gives for  $(a, b) \in \mathbb{N}_0^2$  the bound

$$\|\partial^{(a,b)} \widehat{f}\|_{L^2(\mathcal{W}_J)}^2 \leq C_{a,b} 2^{\alpha j} \cdot 2^{-j2\alpha(a+b)} (2^{-2j(1-\alpha)a} 2^{-(1+2\alpha)j} \ell_J^{-1-2\beta} + 2^{(-1-2\beta)j}),$$

with some constant  $C_{a,b} > 0$  independent of scale. Therefore we can estimate for  $m_1 \in \mathbb{N}_0$

$$\|\mathcal{D}_1^{m_1} \widehat{f}\|_{L^2(\mathcal{W}_J)}^2 \lesssim 2^{-j2\alpha m_1} (2^{-2j(1-\alpha)m_1} 2^{-(1+\alpha)j} \ell_J^{2m_1-1-2\beta} + 2^{\alpha j} 2^{(-1-2\beta)j}).$$

If  $m_1 \leq 2$  this further simplifies to

$$\|\mathcal{D}_1^{m_1} \widehat{f}\|_{L^2(\mathcal{W}_J)}^2 \lesssim 2^{-2jm_1} 2^{-(1+\alpha)j} \ell_J^{2m_1-1-2\beta}, \quad (34)$$

since for every  $m_1 \leq (1+\alpha)/\alpha$  we have

$$2^{\alpha j} 2^{(-1-2\beta)j} \lesssim 2^{-2j(1-\alpha)m_1} 2^{-(1+\alpha)j} \ell_J^{2m_1-1-2\beta}.$$

Similar calculations lead to

$$\|\mathcal{D}_2^{m_2} \widehat{f}\|_{L^2(\mathcal{W}_J)}^2 \lesssim 2^{-2\alpha j m_2} 2^{-(1+\alpha)j} \ell_J^{-1-2\beta}. \quad (35)$$

Indeed, if  $a+b = m_2$  we have

$$\|\partial_1^a \partial_2^b \widehat{f}\|_{L^2(\mathcal{W}_J)}^2 \leq C_{a,b} \cdot 2^{\alpha j} \cdot 2^{-2\alpha j m_2} (2^{-(1+2\alpha)j} \ell_J^{-1-2\beta} + 2^{(-1-2\beta)j}).$$

Since  $1 \leq \ell_J \leq 2 \cdot 2^{(1-\alpha)j}$  it holds

$$2^{j(1+\alpha)} 2^{-2\beta j} = 2^{-(1-\alpha)(1+2\beta)j} \lesssim \ell_J^{-1-2\beta}.$$

Therefore, taking into account  $1 \leq 2^{(1-\alpha)j}$ , we can conclude

$$2^{(-1-2\beta)j} \leq 2^{j(1+\alpha)} 2^{-2\beta j} 2^{-(1+2\alpha)j} \lesssim 2^{-(1+2\alpha)j} \ell_J^{-1-2\beta}.$$

Altogether we obtain the desired estimate (35).

After this preliminary work we can finally prove the statement of Theorem 4.23. We have

$$\mathcal{L}(\widehat{f}\chi_J) = \widehat{f}\chi_J - 2^{2j}\ell_J^{-2}\mathcal{D}_1^2(\widehat{f}\chi_J) - 2^{2\alpha j}\mathcal{D}_2^2(\widehat{f}\chi_J) + 2^{2(1+\alpha)j}\ell_J^{-2}\mathcal{D}_1^2\mathcal{D}_2^2(\widehat{f}\chi_J),$$

which allows us to show the desired estimate for each term separately. For  $\widehat{f}\chi_J$  the estimate holds true by Theorem 4.22.

Let us turn to the second term. The product rule yields

$$\mathcal{D}_1^2(\widehat{f}\chi_J) = (\mathcal{D}_1^2\widehat{f})\chi_J + 2(\mathcal{D}_1\widehat{f})(\mathcal{D}_1\chi_J) + \widehat{f}(\mathcal{D}_1^2\chi_J).$$

The previous estimates together with the Hölder inequality then lead to

$$\|\mathcal{D}_1^2(\widehat{f}\chi_J)\|_2^2 \lesssim 2^{-4j}\ell_J^4 2^{-(1+\alpha)j}\ell_J^{-1-2\beta}.$$

Here (34) was used, and that  $\|\mathcal{D}_1^{m_1}\chi_J\|_\infty^2 \leq 2^{-2jm_1} \leq 2^{-2jm_1}\ell_J^{2m_1}$  by (33). This settles the claim for the second term.

Analogously, we can deduce

$$\|\mathcal{D}_2^2(\widehat{f}\chi_J)\|_2^2 \lesssim 2^{-4\alpha j} 2^{-(1+\alpha)j}\ell_J^{-1-2\beta},$$

using (35) and that  $\|\mathcal{D}_2^{m_2}\chi_J\|_\infty^2 \leq 2^{-2\alpha jm_2}$  by (33). This gives the estimate for the third term.

Finally, it also holds

$$\|\mathcal{D}_1^2\mathcal{D}_2^2(\widehat{f}\chi_J)\|_2^2 \lesssim 2^{-4j(1+\alpha)}\ell_J^4 2^{-(1+\alpha)j}\ell_J^{-1-2\beta},$$

which establishes the result for the fourth term. □

At last we are ready to give the proof of Proposition 4.4. The essential tool is Theorem 4.23.

*Proof of Proposition 4.4.* Recall the curvelet frame  $\mathcal{C}_\alpha(W^{(0)}, W, V) = (\psi_\mu)_{\mu \in M}$ . On the Fourier side we have

$$\widehat{\psi}_{j,\ell,k} = \chi_J u_{j,k}(R_J \cdot),$$

with rotation matrix  $R_J$  given as in (7) and functions

$$u_{j,k}(\xi) = 2^{-j(1+\alpha)/2} e^{2\pi i(2^{-j}k_1, 2^{-\alpha j}k_2) \cdot \xi}, \quad \xi \in \mathbb{R}^2.$$

The curvelet coefficients  $(\theta_\mu)_{\mu \in M}$  are therefore given by the formula

$$\theta_\mu = \langle f, \psi_{j,\ell,k} \rangle = \int_{\mathbb{R}^2} \widehat{f}\chi_J(\xi) \overline{u_{j,k}(R_J\xi)} d\xi.$$

Since

$$\mathcal{L}(u_{j,k}) = (1 + \ell_J^{-2}k_1^2)(1 + k_2^2)u_{j,k},$$

integration by parts yields

$$\theta_\mu = (1 + \ell_J^{-2}k_1^2)^{-1}(1 + k_2^2)^{-1} \int_{\mathbb{R}^2} \mathcal{L}(\widehat{f}\chi_J)(\xi) \overline{u_{j,k}(R_J\xi)} d\xi.$$

Let  $K = (K_1, K_2) \in \mathbb{Z}^2$  and define

$$\mathfrak{K} := \left\{ (k_1, k_2) \in \mathbb{Z}^2 : \ell_J^{-1}k_1 \in [K_1, K_1 + 1), k_2 = K_2 \right\}.$$

For fixed  $J = (j, \ell)$  the Fourier system  $(u_{j,k}(R_J \cdot))_{k \in \mathbb{Z}^2}$  is an orthonormal basis for  $L^2(\Xi_J)$ , where  $\Xi_J$  is the rectangle defined in (10) containing the support of  $\chi_J$ . Therefore,

$$\sum_{k \in \mathfrak{K}} |\theta_\mu|^2 \lesssim (1 + K_1^2)^{-2}(1 + K_2^2)^{-2} \int_{\mathbb{R}^2} |\mathcal{L}(\widehat{f}\chi_J)(\xi)|^2 d\xi.$$

The integral on the right-hand side is bounded by Theorem 4.23, and we thus arrive at

$$\sum_{k \in \mathfrak{Z}_K} |\theta_\mu|^2 \lesssim (L_K)^{-2} 2^{-(1+\alpha)j} \ell_J^{-1-2\beta} \quad (36)$$

with  $L_K := (1 + K_1^2)(1 + K_2^2)$ .

Let  $M_J$  denote the subset of curvelet coefficients associated with a fixed scale-angle pair  $J = (j, \ell)$ . Further, let  $N_{J,K}(\varepsilon)$  be the number of indices  $\mu \in M_J$  such that  $k \in \mathfrak{Z}_K$  and  $|\theta_\mu| > \varepsilon$ .

Since  $\#\mathfrak{Z}_K \leq \ell_J$  and because of (36) we can conclude

$$N_{J,K}(\varepsilon) \lesssim \min \left\{ \ell_J, (\varepsilon L_K)^{-2} 2^{-(1+\alpha)j} \ell_J^{-1-2\beta} \right\}. \quad (37)$$

For  $\omega_J = \pi \ell 2^{-\lfloor j(1-\alpha) \rfloor} \in [0, \pi)$  let  $\langle \omega_J \rangle$  denote the equivalent angle modulo  $\pi$  in the interval  $(-\pi/2, \pi/2]$ . The corresponding indices in the range  $\{[-L_j/2 + 1], \dots, \lfloor L_j/2 \rfloor\}$  shall be denoted by  $\langle \ell \rangle$ . Since it holds  $|\sin \eta| \asymp |\eta|$  for  $\eta \in [-\pi/2, \pi/2]$ , it follows

$$\ell_J = 1 + 2^{(1-\alpha)j} |\sin \omega_J| = 1 + 2^{(1-\alpha)j} |\sin \langle \omega_J \rangle| \asymp 1 + |\langle \ell \rangle|. \quad (38)$$

Let  $\ell_*$  be the solution of  $\ell_* = (\varepsilon L_K)^{-2} 2^{-(1+\alpha)j} \ell_*^{-1-2\beta}$  and put  $L^* = \lfloor \ell_* \rfloor$ . Utilizing (37) and (38) yields

$$\begin{aligned} \sum_{|J|=j} N_{J,K}(\varepsilon) &\lesssim \sum_{\substack{\ell \in \{0, \dots, L_j-1\} \\ |\langle \ell \rangle| \leq L^*-1}} (1 + |\langle \ell \rangle|) + \sum_{\substack{\ell \in \{0, \dots, L_j-1\} \\ |\langle \ell \rangle| \geq L^*}} (\varepsilon L_K)^{-2} 2^{-(1+\alpha)j} (1 + |\langle \ell \rangle|)^{-1-2\beta} \\ &\lesssim \sum_{\ell=0}^{L^*-1} (1 + |\ell|) + \sum_{\ell=L^*}^{\infty} (\varepsilon L_K)^{-2} 2^{-(1+\alpha)j} (1 + |\ell|)^{-1-2\beta} \\ &\lesssim (L^*)^2 + (\varepsilon L_K)^{-2} 2^{-(1+\alpha)j} (L^*)^{-2\beta}. \end{aligned}$$

This translates to

$$\sum_{|J|=j} N_{J,K}(\varepsilon) \lesssim \varepsilon^{-2/(1+\beta)} \cdot L_K^{-2/(1+\beta)} \cdot 2^{-(1+\alpha)j/(1+\beta)}.$$

Since  $\beta < 3$  we have  $\sum_{K \in \mathbb{Z}^2} L_K^{-2/(1+\beta)} < \infty$ . Hence

$$\#\left\{ \mu \in M_j, |\theta_\mu| > \varepsilon \right\} = \sum_{K \in \mathbb{Z}^2} \sum_{|J|=j} N_{J,K}(\varepsilon) \lesssim 2^{-(1+\alpha)j/(1+\beta)} \varepsilon^{-2/(1+\beta)}.$$

This finishes the proof.  $\square$

## References

- [1] T. Berger. *Rate-Distortion Theory*. Wiley Online Library, 1971.
- [2] E. Candes. Ridgelets: Theory and applications. PhD thesis, Stanford University, 1998.
- [3] E. J. Candès and D. L. Donoho. New tight frames of curvelets and optimal representations of objects with  $C^2$  singularities. *Comm. Pure Appl. Math.*, 56:219–266, 2004.
- [4] O. Christensen. *An Introduction to Frames and Riesz Bases*. Birkhäuser, 2003.
- [5] R. DeVore. Nonlinear approximation. *Acta Numerica*, 7:51–150, 1998.
- [6] D. L. Donoho. Sparse components of images and optimal atomic decomposition. *Constr. Approx.*, 17:353–382, 2001.



- [7] R. Gribonval and M. Nielsen. Non-linear approximation with dictionaries. I. Direct estimates. *J. Fourier Anal. Appl.*, 10:51–71, 2004.
- [8] P. Grohs. Ridgelet-type frame decompositions for sobolev spaces related to linear transport. *J. Fourier Anal. Appl.*, 18(2):309 – 325, 2012.
- [9] P. Grohs, S. Keiper, G. Kutyniok, and M. Schaefer.  $\alpha$ -molecules: Curvelets, shearlets, ridgelets, and beyond. In *Proceedings SPIE 2013*, 2013.
- [10] P. Grohs, S. Keiper, G. Kutyniok, and M. Schäfer.  $\alpha$ -Molecules. 2013. preprint.
- [11] P. Grohs and G. Kutyniok. Parabolic molecules. *Foundations of Computational Mathematics*, 14(2):299–337, 2014.
- [12] K. Guo, G. Kutyniok, and D. Labate. Sparse multidimensional representations using anisotropic dilation and shear operators. *Wavelets and Splines (Athens, GA, 2005)*, Nashboro Press., 14:189–201, 2006.
- [13] S. Keiper. A flexible shearlet transform - sparse approximation and dictionary learning. Bachelor’s thesis, TU Berlin, 2013.
- [14] G. Kutyniok and D. Labate. *Shearlets: Multiscale Analysis for Multivariate Data*, chapter Introduction to Shearlets, pages 1–38. Birkhäuser, 2012.
- [15] G. Kutyniok, J. Lemvig, and W.-Q. Lim. Compactly supported shearlet frames and optimally sparse approximations of functions in  $L^2(\mathbb{R}^3)$  with piecewise  $C^\alpha$  singularities. *SIAM J. Math. Anal.*, 44:2962–3017, 2012.
- [16] G. Kutyniok and W.-Q. Lim. Compactly supported shearlets are optimally sparse. *J. Approx. Theory*, 163(11):1564–1589, 2011.

## A Additional Technical Estimates

This appendix contains some technical estimates, which are needed for the proofs of the main results. They were outsourced from the main exposition to enhance the readability. We first state a lemma, which shows how scaling affects the Hölder constant in an abstract setting.

**Lemma A.1.** *Let  $f \in C^\alpha(\mathbb{R})$  and  $0 < \alpha < 1$ . Then for  $t, s > 0$*

$$\text{Höl}(sf(t), \alpha) = st^\alpha \cdot \text{Höl}(f, \alpha).$$

Subsequently we establish some estimates for the ‘elementary functions’  $g^\eta$ ,  $\omega^\eta$ , and  $r^\eta$ , which occur as components of the edge fragments  $F$  and  $\tilde{F}$ . These estimates provide the basis for the more complex estimates needed throughout this paper. To enhance readability we omit the superindex  $\eta$  during this discussion.

### A.1 Estimates for $g^\eta$

The functions  $g_j^\eta$  are defined for  $j \in \mathbb{N}_0$  by  $g_j^\eta := g^\eta(2^{-\alpha j} \cdot)$ , where  $g^\eta \in C_0^\beta(\mathbb{R}^2)$  is a fixed function and  $\beta \in (1, 2]$ . For simplicity we write henceforth  $g_j$  and  $g$ .

Clearly, still  $g_j \in C_0^\beta(\mathbb{R}^2)$ . However, the parameters of the regularity change. It is obvious that  $\|g_j\|_\infty \lesssim 1$ . Further, the chain rule yields

$$\|\partial_1 g_j\|_\infty \lesssim 2^{-\alpha j} \quad \text{and} \quad \|\partial_2 g_j\|_\infty \lesssim 2^{-\alpha j}.$$

Applying Lemma A.1 yields the following result.

**Lemma A.2.** *Let  $\beta \in (1, 2]$  and  $g \in C_0^\beta(\mathbb{R}^2)$ . Then for  $g_j = g(2^{-\alpha j} \cdot)$*

$$\text{Höl}(\partial_1 g_j, \beta - 1) = 2^{-j} \text{Höl}(\partial_1 g, \beta - 1).$$

*Proof.* In view of Lemma A.1 we have

$$\text{Höl}(\partial_1 g_j, \beta - 1) = \text{Höl}(2^{-\alpha j} \partial_1 g(2^{-\alpha j} \cdot), \beta - 1) = 2^{-j} \text{Höl}(\partial_1 g, \beta - 1).$$

□

Some more estimates for  $g_j$  are collected in the following two lemmas.

**Lemma A.3.** *The following estimates hold true for  $g_j$ :*

$$\begin{aligned} \|\Delta_{(h,0)} g_j\|_\infty &\lesssim 2^{-\alpha j} h, \\ \|\Delta_{(h,0)} \partial_1 g_j\|_\infty, \|\Delta_{(h,0)} \partial_2 g_j\|_\infty &\lesssim 2^{-j} h^{\beta-1} = 2^{-\alpha j} h^\beta, \\ \|\Delta_{(h,0)}^2 g_j\|_\infty &\lesssim 2^{-j} h^\beta, \end{aligned}$$

with implicit constants, that do not depend on  $j \in \mathbb{N}_0$  and  $h \geq 0$ .

*Proof.* Applying the mean value theorem yields

$$\|\Delta_{(h,0)} g_j\|_\infty \leq h \|\partial_1 g_j\|_\infty \lesssim 2^{-\alpha j} h.$$

Considering Lemma A.2 we obtain

$$\|\Delta_{(h,0)} \partial_1 g_j\|_\infty \lesssim 2^{-j} h^{\beta-1} = 2^{-\alpha j} h^\beta.$$

Noting the commutativity  $\partial_1 \Delta_{(h,0)} = \Delta_{(h,0)} \partial_1$ , we obtain

$$\|\Delta_{(h,0)}^2 g_j\|_\infty \lesssim h \|\Delta_{(h,0)} \partial_1 g_j\|_\infty \lesssim 2^{-j} h^\beta.$$

□

The next lemma gives estimates for  $g_j$  along the edge curve. Here the function  $a \in C^\beta(\mathbb{R})$  comes into play, which was defined in (25). The following estimates also depend on the properties of  $a$ , which are summarized in Lemma 4.14.

**Lemma A.4.** *Assume  $|\sin \eta| \geq 2\delta_j$ . The following estimates hold true for  $g_j$ :*

$$\begin{aligned} \sup_{t \in \mathbb{R}} |\Delta_h g_j(t, a(t))| &\lesssim h |\sin \eta|^{-1} 2^{-\alpha j}, \\ \sup_{t \in \mathbb{R}} |\Delta_h \partial_1 g_j(t, a(t))|, \sup_{t \in \mathbb{R}} |\Delta_h \partial_2 g_j(t, a(t))| &\lesssim h^{\beta-1} |\sin \eta|^{1-\beta} 2^{-j}, \\ \sup_{t \in \mathbb{R}} |\Delta_h^2 g_j(t, a(t))| &\lesssim h^\beta |\sin \eta|^{-1-\beta} 2^{-j}, \end{aligned}$$

where the implicit constants are independent of  $j \in \mathbb{N}_0$  and  $h \geq 0$ .

*Proof.* In view of Lemma 4.14 it holds

$$\begin{aligned} \sup_{t \in \mathbb{R}} |\Delta_h g_j(t, a(t))| &\lesssim h \cdot \sup_{t \in \mathbb{R}} \left| \frac{d}{dt} g_j(t, a(t)) \right| \\ &\lesssim h \cdot \left( \sup_{t \in \mathbb{R}} |\partial_1 g_j(t, a(t))| + \sup_{t \in \mathbb{R}} |\partial_2 g_j(t, a(t)) a'(t)| \right) \\ &\lesssim h \cdot |\sin \eta|^{-1} 2^{-\alpha j}. \end{aligned}$$

Considering the transformation behavior of the Hölder constant we obtain with Lemma 4.14

$$\begin{aligned} \sup_{t \in \mathbb{R}} |\Delta_h \partial_1 g_j(t, a(t))| &\lesssim 2^{-j} \sup_{t \in \mathbb{R}} |(h, a(t+h) - a(t))|_2^{\beta-1} \\ &\lesssim 2^{-j} (h^{\beta-1} + \sup_{t \in \mathbb{R}} |a(t+h) - a(t)|^{\beta-1}) \\ &\lesssim 2^{-j} h^{\beta-1} |\sin \eta|^{1-\beta}. \end{aligned}$$

Applying Lemma 4.14, the mean value theorem and  $\frac{d}{dt} \Delta_h = \Delta_h \frac{d}{dt}$  yields

$$\begin{aligned} \sup_{t \in \mathbb{R}} |\Delta_h^2 g_j(t, a(t))| &\lesssim h \cdot \sup_{t \in \mathbb{R}} \left| \Delta_h \frac{d}{dt} g_j(t, a(t)) \right| \\ &= h \cdot \sup_{t \in \mathbb{R}} |\Delta_h (\partial_1 g_j(t, a(t)) + \partial_2 g_j(t, a(t)) a'(t))| \\ &= h \cdot \sup_{t \in \mathbb{R}} |\Delta_h \partial_1 g_j(t, a(t)) + \Delta_h \partial_2 g_j(t, a(t)) a'(t+h) + \partial_2 g_j(t, a(t)) \Delta_h a'(t)| \\ &\lesssim h^\beta |\sin \eta|^{1-\beta} 2^{-j} + h^\beta |\sin \eta|^{-\beta} 2^{-j} + \delta_j h^\beta |\sin \eta|^{-1-\beta} 2^{-\alpha j}. \end{aligned}$$

□

## A.2 Estimates for $\omega^\eta$

Similarly, we obtain estimates for the window function  $\omega^\eta \in C_0^\infty(\mathbb{R}^2)$ , which in contrast to the functions  $g_j$  remains fixed at all scales. This fact and the smoothness of  $\omega$  result in different estimates.

First we state the trivial estimates  $\|\omega\|_\infty \lesssim 1$ ,  $\|\partial_1 \omega\|_\infty \lesssim 1$ , and  $\|\partial_2 \omega\|_\infty \lesssim 1$ . Next, we apply the forward difference operator  $\Delta_{(h,0)}$  to  $\omega$ .

**Lemma A.5.** *Let  $k \in \mathbb{N}_0$ . It holds with implicit constants independent of  $h \geq 0$*

$$\|\Delta_{(h,0)}^k \omega\|_\infty \lesssim h^k \quad \text{and} \quad \|\Delta_{(h,0)}^k \partial_1 \omega\|_\infty \lesssim h^k.$$

Analogous to Lemma A.4 we establish estimates along the edge curve.

**Lemma A.6.** *Assume  $|\sin \eta| \geq 2\delta_j$ . It holds*

$$\begin{aligned} \sup_{t \in \mathbb{R}} |\Delta_h \omega(t, a(t))| &\lesssim h |\sin \eta|^{-1}, \\ \sup_{t \in \mathbb{R}} |\Delta_h \partial_1 \omega(t, a(t))|, \sup_{t \in \mathbb{R}} |\Delta_h \partial_2 \omega(t, a(t))| &\lesssim h |\sin \eta|^{-1}, \\ \sup_{t \in \mathbb{R}} |\Delta_h^2 \omega(t, a(t))| &\lesssim h^2 |\sin \eta|^{-2} + \delta_j h^\beta |\sin \eta|^{-1-\beta}. \end{aligned}$$

*Proof.* This proof is analogous to the proof of Lemma A.4. □

### A.3 Estimates for $r^\eta$

Next we analyze the function  $r^\eta : \mathbb{R}^2 \rightarrow \mathbb{R}$  given by

$$r^\eta(t, a) := t \cos \eta - a \sin \eta, \quad (t, a) \in \mathbb{R}^2,$$

which is a component of the modified edge fragment  $\tilde{F}$ . Clearly  $r \in C^\infty(\mathbb{R}^2)$ . Note however, that  $r$  is not compactly supported. Since  $r$  only occurs as a factor in products with the window  $\omega$  this does not cause any problems.

Thanks to the smoothness of  $r$  we have the following result.

**Lemma A.7.** *Let  $k, m \in \mathbb{N}_0$  and  $K \subset \mathbb{R}^2$  a compact set. Then we have*

$$\|\Delta_{(h,0)}^k r^m\|_{L^\infty(K)} \lesssim h^k.$$

Along the edge curve the following estimates hold. Here  $\tilde{I}(\eta)$  denotes the interval defined in (26).

**Lemma A.8.** *Let  $|\sin \eta| \geq 2\delta_j$ . Then we have  $\sup_{t \in \tilde{I}(\eta)} |r(t, a(t))| \lesssim \delta_j$ . Moreover, for  $h \geq 0$  it holds*

$$\sup_{t \in \mathbb{R}} |\Delta_h r(t, a(t))| \lesssim h \quad \text{and} \quad \sup_{t \in \mathbb{R}} |\Delta_h^2 r(t, a(t))| \lesssim h^\beta \delta_j |\sin \eta|^{-\beta}.$$

*Proof.* For every  $t \in \mathbb{R}$  the point  $(t, a(t)) \in \mathbb{R}^2$  in rotated coordinates lies on the (extended) edge curve  $\Gamma$ . We know that the function  $E_j$  deviates little from zero and obeys  $\sup_{|x_2| \leq 1} |E_j(x_2)| \leq \delta_j \lesssim 2^{-j(1-\alpha)}$  according to (22). Furthermore, the slope of  $E_j$  outside of  $[-1, 1]$  is constant and bounded by  $\delta_j$ . This yields the estimate  $\sup_{t \in \tilde{I}(\eta)} |r(t, a(t))| \lesssim \delta_j$ .

The other estimates follow from Lemma 4.14. In view of this lemma we conclude

$$\sup_{t \in \mathbb{R}} |\Delta_h r(t, a(t))| \leq h \cdot \sup_{t \in \mathbb{R}} |\cos \eta - a'(t) \sin \eta| \lesssim h |\sin \eta|^{-1} |\sin \eta| = h,$$

and

$$\sup_{t \in \mathbb{R}} |\Delta_h^2 r(t, a(t))| \leq h |\sin \eta| \|\Delta_h a'\|_\infty \lesssim h^\beta \delta_j |\sin \eta|^{-\beta}.$$

□

### A.4 Estimates for $\tilde{G}^\eta$

The function  $\tilde{G}^\eta$  is the rotated version of the function

$$\tilde{G}(t, a) = r(t, a)^m G(t, a) = r(t, a)^m g_j(t, a) \omega(t, a), \quad (t, a) \in \mathbb{R}^2,$$

which is a composition of the ‘elementary functions’ discussed before. Hence we can apply the previous estimates to obtain estimates for  $\tilde{G}^\eta$ .

**Lemma A.9.** *Let  $|\sin \eta| \geq 2\delta_j$ . Let  $\tilde{G}^\eta(t, a) = r^\eta(t, a)^m G^\eta(t, a)$  for  $(t, a) \in \mathbb{R}^2$ ,  $m \in \mathbb{N}$ ,  $m \neq 0$ . Then there are the estimates*

$$\begin{aligned} \sup_{t \in \mathbb{R}} |\tilde{G}(t, a(t))| &\lesssim \delta_j^m, & \sup_{t \in \mathbb{R}} |\Delta_h \tilde{G}(t, a(t))| &\lesssim \delta_j^{m-1} h, \\ \sup_{t \in \mathbb{R}} |\partial_1 \tilde{G}(t, a(t))| &\lesssim \delta_j^{m-1}, & \sup_{t \in \mathbb{R}} |\partial_2 \tilde{G}(t, a(t))| &\lesssim \delta_j^{m-1} |\sin \eta|. \end{aligned}$$

*Proof.* We calculate for  $(t, a) \in \mathbb{R}^2$

$$\partial_1 \tilde{G}(t, a) = \partial_1 (r(t, a)^m G(t, a)) = (\cos \eta) m r(t, a)^{m-1} G(t, a) + r(t, a)^m \partial_1 G(t, a),$$

$$\text{and} \quad \partial_2 \tilde{G}(t, a) = \partial_2 (r(t, a)^m G(t, a)) = -(\sin \eta) m r(t, a)^{m-1} G(t, a) + r(t, a)^m \partial_2 G(t, a).$$

The assertion is then a consequence of the following facts. It holds  $\|G\|_\infty \lesssim 1$  and  $|r(t, a(t))| \leq \delta_j$  for all  $t \in I(\eta)$ . Further, for  $t \notin I(\eta)$  the expressions  $G(t, a(t))$ ,  $\partial_1 G(t, a(t))$ , and  $\partial_2 G(t, a(t))$  vanish. □

## A.5 Auxiliary Lemmas

In this last subsection we prove three important technical lemmas, which are the foundation of the proof of Lemma 4.19. We refer to Lemma 4.19 for an explanation of the terminology.

**Lemma A.10.** *Let  $\tilde{G}(x) = r(x)^m G(x)$  for  $x \in \mathbb{R}^2$  and  $m \in \mathbb{N}_0$ . Further, let  $h \asymp 2^{-j(1-\alpha)}$ . The function  $T : \mathbb{R} \rightarrow \mathbb{R}$  defined by  $T(t) = a'(t)\Delta_h \tilde{G}(t, a(t))$  then admits a decomposition  $(T_1^k, T_2^k)_k$  of the form (\*) of length  $(m+1)$  with the estimates*

$$\begin{aligned} \|T_1^k\|_\infty &\lesssim h^m h^\beta |\sin \eta|^{-1-\beta}, \quad k = 0, \dots, m+1, \\ \|T_2^k\|_\infty &\lesssim h^m |\sin \eta|^{-1}, \quad k = 0, \dots, m, \end{aligned}$$

and subject to the condition  $\text{supp } T_i^k \subset \tilde{I}(\eta)$ .

An explanation of the terminology is given in Lemma 4.19.

*Proof.* We prove this by induction on  $m$ . If  $m = 0$  we put  $T_1^0 = T_{21}$ ,  $T_2^0 = T_{22}$ ,  $T_1^1 = \Delta_h T_{22}$ , and  $T_2^1 = 0$ , with entities  $T_{21}$  and  $T_{22}$  as defined in the proof of Lemma 4.16. The estimates for  $T_{21}$  and  $\Delta_h T_{22}$  have been carried out there. In view of  $h \lesssim \sin \eta$  we can further estimate

$$\|T_2^0\|_\infty = \|T_{22}\|_\infty \lesssim h |\sin \eta|^{-2} \lesssim h^0 |\sin \eta|^{-1}.$$

This proves the case  $m = 0$ .

We proceed with the induction and assume that the lemma is true for  $T$ , where  $m \in \mathbb{N}_0$  is fixed but arbitrary. The associated decomposition of length  $m+1$  shall be denoted by  $(T_1^k, T_2^k)_k$ . We will show that under this hypothesis also the function  $\tilde{T}(t) := a'(t)\Delta_h \tilde{G}_+(t, a(t))$ , where  $\tilde{G}_+(x) = r(x)^{m+1}G(x)$  for  $x \in \mathbb{R}^2$ , admits a decomposition  $(\tilde{T}_1^k, \tilde{T}_2^k)_k$  of the form (\*) of length  $(m+2)$  with the desired properties.

First we can decompose as follows,

$$\begin{aligned} \tilde{T}(t) &= a'(t)\Delta_h(r(t, a(t))\tilde{G}(t, a(t))) \\ &= a'(t)\Delta_h r(t, a(t))\tilde{G}(t+h, a(t+h)) + r(t, a(t))a'(t)\Delta_h \tilde{G}(t, a(t)) \\ &= [r(t, a(t))T_1^0(t)] + [a'(t)\Delta_h r(t, a(t))\tilde{G}(t+h, a(t+h)) + r(t, a(t))T_2^0(t)] \\ &=: \tilde{T}_1^0(t) + \tilde{T}_2^0(t). \end{aligned}$$

In view of the properties of  $T_1^0$  and Lemma A.8 we see that the function  $\tilde{T}_1^0$  satisfies the assertion. The estimate

$$\|\tilde{T}_1^0\|_\infty \lesssim \|a'\|_\infty \sup_{t \in \mathbb{R}} |\Delta_h r(t, a(t))| \sup_{t \in \mathbb{R}} |\tilde{G}(t, a(t))| \lesssim |\sin \eta|^{-1} \cdot h \cdot \delta_j^m,$$

where Lemmas 4.14, A.8 and A.9 were used, shows the claim also for  $\tilde{T}_2^0$ .

We take another forward difference of the component  $\tilde{T}_2^0$  and obtain

$$\begin{aligned} \Delta_h \tilde{T}_2^0(t) &= \Delta_h a'(t)\Delta_h r(t+h, a(t+h))\tilde{G}(t+2h, a(t+2h)) + a'(t)\Delta_h^2 r(t, a(t))\tilde{G}(t+2h, a(t+2h)) \\ &\quad + \Delta_h r(t, a(t))a'(t)\Delta_h \tilde{G}(t+h, a(t+h)) + \Delta_h r(t, a(t))T_2^0(t+h) + r(t, a(t))\Delta_h T_2^0(t) \\ &= \Delta_h a'(t)\Delta_h r(t+h, a(t+h))\tilde{G}(t+2h, a(t+2h)) + a'(t)\Delta_h^2 r(t, a(t))\tilde{G}(t+2h, a(t+2h)) \\ &\quad + \Delta_h r(t, a(t))(T_1^0(t+h) + T_2^0(t+h)) - \Delta_h r(t, a(t))\Delta_h a'(t)\Delta_h \tilde{G}(t+h, a(t+h)) \\ &\quad + \Delta_h r(t, a(t))T_2^0(t+h) + r(t, a(t))T_1^1(t) + r(t, a(t))T_2^1(t) \\ &= [\Delta_h a'(t)\Delta_h r(t+h, a(t+h))\tilde{G}(t+2h, a(t+2h)) + a'(t)\Delta_h^2 r(t, a(t))\tilde{G}(t+2h, a(t+2h)) \\ &\quad - \Delta_h r(t, a(t))\Delta_h a'(t)\Delta_h \tilde{G}(t+h, a(t+h)) + r(t, a(t))T_1^1(t) + \Delta_h r(t, a(t))T_1^0(t+h)] \\ &\quad + [2\Delta_h r(t, a(t))T_2^0(t+h) + r(t, a(t))T_2^1(t)] \\ &=: \tilde{T}_1^1(t) + \tilde{T}_2^1(t). \end{aligned}$$

For  $\tilde{T}_1^1$  we check directly

$$\begin{aligned} \sup_{t \in \mathbb{R}} |\Delta_h a'(t) \Delta_h r(t, a(t)) \tilde{G}(t+h, a(t+h))| &\lesssim h^{\beta-1} \delta_j |\sin \eta|^{-1-\beta} h h^m = h^{m+1} h^\beta |\sin \eta|^{-1-\beta}, \\ \sup_{t \in \mathbb{R}} |a'(t) \Delta_h^2 r(t, a(t)) \tilde{G}(t+2h, a(t+2h))| &\lesssim |\sin \eta|^{-1} h^\beta \delta_j |\sin \eta|^{-\beta} h^m = h^{m+1} h^\beta |\sin \eta|^{-1-\beta}, \\ \sup_{t \in \mathbb{R}} |\Delta_h r(t, a(t)) \Delta_h a'(t) \Delta_h \tilde{G}(t+h, a(t+h))| &\lesssim \delta_j^m h^\beta h |\sin \eta|^{-1-\beta} \lesssim h^{m+1} h^\beta |\sin \eta|^{-1-\beta}. \end{aligned}$$

The estimates for the remaining two terms are obvious. Hence  $\tilde{T}_1^1$  fulfills the desired properties.

For  $\tilde{T}_2^1$  we use the induction hypothesis and Lemma A.8 to obtain

$$\|\tilde{T}_2^1\|_\infty \lesssim \sup_{t \in \mathbb{R}} |r(t, a(t)) T_2^1(t)| + \sup_{t \in \mathbb{R}} |\Delta_h r(t, a(t)) T_2^0(t+h)| \lesssim h^{m+1} |\sin \eta|^{-1}.$$

Moving forward, this procedure yields terms for  $k = 1, \dots, m+1$ ,

$$\begin{aligned} \tilde{T}_1^{k+1}(t) &= r(t, a(t)) T_1^{k+1}(t) + (k+1) \Delta_h r(t+h, a(t+h)) T_1^k(t+h) \\ &\quad + (k+1) \Delta_h^2 r(t, a(t)) T_2^{k-1}(t+h) + (k+1) \Delta_h^2 r(t, a(t)) T_2^k(t+h), \\ \tilde{T}_2^k(t) &= r(t, a(t)) T_2^k(t) + (k+1) \Delta_h r(t, a(t)) T_2^{k-1}(t+h), \end{aligned}$$

which satisfy the desired estimates. Here we put  $T_1^{m+2} = T_2^{m+2} = 0$  for convenience. Indeed, using the induction assumptions, we obtain

$$\begin{aligned} \|\tilde{T}_1^{k+1}\|_\infty &\lesssim \sup_{t \in \mathbb{R}} |r(t, a(t)) T_1^{k+1}(t)| + \sup_{t \in \mathbb{R}} |\Delta_h r(t+h, a(t+h)) T_1^k(t+h)| \\ &\quad + \sup_{t \in \mathbb{R}} |\Delta_h^2 r(t, a(t)) T_2^{k-1}(t+h)| + \sup_{t \in \mathbb{R}} |\Delta_h^2 r(t, a(t)) T_2^k(t+h)| \lesssim h^{m+1} h^\beta |\sin \eta|^{-1-\beta}, \\ \|\tilde{T}_2^k\|_\infty &\lesssim \sup_{t \in \mathbb{R}} |r(t, a(t)) T_2^k(t)| + \sup_{t \in \mathbb{R}} |\Delta_h r(t, a(t)) T_2^{k-1}(t+h)| \lesssim h^{m+1} |\sin \eta|^{-1}. \end{aligned}$$

Note, that  $T_2^{m+1} = T_1^{m+2} = T_2^{m+2} = 0$ . Hence, for  $k = m+1$  these expressions read

$$\begin{aligned} \tilde{T}_1^{m+2}(t) &= (m+2) \Delta_h r(t+h, a(t+h)) T_1^{m+1}(t+h) + (m+2) \Delta_h^2 r(t, a(t)) T_2^m(t+h), \\ \tilde{T}_2^{m+1}(t) &= (m+2) \Delta_h r(t, a(t)) T_2^m(t+h). \end{aligned}$$

Since  $\Delta_h \tilde{T}_2^{m+1} = \tilde{T}_1^{m+2}$  we have  $\tilde{T}_2^{m+2} = 0$  and the proof is finished.  $\square$

The following Lemma A.11 is in the same spirit as Lemma A.10.

**Lemma A.11.** *Let  $\tilde{G}(x) = r(x)^m G(x)$  for  $x \in \mathbb{R}^2$ ,  $m \in \mathbb{N}_0$ , and  $h \asymp 2^{-j(1-\alpha)}$ . Then the function  $S : \mathbb{R} \rightarrow \mathbb{R}$  defined by  $S(t) = a'(t) \Delta_h \partial_1 \tilde{G}(t, a(t))$  admits a decomposition  $(S_1^k, S_2^k)_k$  of the form (\*) of length  $m+1$  with estimates*

$$\begin{aligned} \|S_1^k\|_\infty &\lesssim h^{m-1} h^\beta |\sin \eta|^{-1-\beta}, \quad k = 0, \dots, m+1, \\ \|S_2^k\|_\infty &\lesssim h^{m-1} |\sin \eta|^{-1}, \quad k = 0, \dots, m. \end{aligned}$$

Moreover, these functions can be chosen such that  $\text{supp } S_i^k \subset \tilde{I}(\eta)$ .

*Proof.* The proof is by induction on  $m$ . The assumptions are clearly true for  $m = 0$ .

For the induction we let  $m \in \mathbb{N}_0$  be fixed and let  $S$  be the function defined in the setting. Further, let us assume that we have a decomposition  $(S_1^k, S_2^k)_k$  of length  $m+1$  with the desired properties for  $S$ . We put  $S_2^{m+1} = 0$  and for convenience we also define  $S_1^{m+2} = S_2^{m+2} = 0$ . We will show that under these assumptions the function  $\tilde{S} : \mathbb{R} \rightarrow \mathbb{R}$  given by  $\tilde{S}(t) := a'(t) \Delta_h \partial_1 \tilde{G}_+(t, a(t))$ , where  $\tilde{G}_+(x) = r(x)^{m+1} G(x)$  for  $x \in \mathbb{R}^2$ , admits a decomposition  $(\tilde{S}_1^k, \tilde{S}_2^k)_k$  of length  $m+2$  of the same form. First we calculate

$$\begin{aligned} \tilde{S}(t) &= a'(t) \Delta_h \partial_1 \tilde{G}_+(t, a(t)) = a'(t) \Delta_h (\cos \eta \tilde{G}(t, a(t)) + r(t, a(t)) \partial_1 \tilde{G}(t, a(t))) \\ &= a'(t) \cos \eta \Delta_h \tilde{G}(t, a(t)) + a'(t) \Delta_h r(t, a(t)) \partial_1 \tilde{G}(t, a(t)) + r(t+h, a(t+h)) a'(t) \Delta_h \partial_1 \tilde{G}(t, a(t)). \end{aligned}$$

Using the induction hypothesis we can proceed,

$$\begin{aligned}\tilde{S}(t) &= [r(t+h, a(t+h))S_1^0(t)] \\ &\quad + [r(t+h, a(t+h))S_2^0(t) + a'(t) \cos \eta \Delta_h \tilde{G}(t, a(t)) + a'(t) \Delta_h r(t, a(t)) \partial_1 \tilde{G}(t, a(t))] \\ &=: \tilde{S}_1^0(t) + \tilde{S}_2^0(t).\end{aligned}$$

The terms  $\tilde{S}_1^0$  and  $\tilde{S}_2^0$  have the desired properties, which follows from the estimates

$$\begin{aligned}\sup_{t \in \mathbb{R}} |r(t+h, a(t+h))S_1^0(t)| &\lesssim hh^{m-1}h^\beta |\sin \eta|^{-1-\beta}, \\ \sup_{t \in \mathbb{R}} |r(t+h, a(t+h))S_2^0(t)| &\lesssim hh^{m-1} |\sin \eta|^{-1}, \\ \sup_{t \in \mathbb{R}} |a'(t) \cos \eta \Delta_h \tilde{G}(t, a(t))| &\lesssim |\sin \eta|^{-1} \delta_j^{m-1} h, \\ \sup_{t \in \mathbb{R}} |a'(t) \Delta_h r(t, a(t)) \partial_1 \tilde{G}(t, a(t))| &\lesssim |\sin \eta|^{-1} h \delta_j^{m-1}.\end{aligned}$$

Taking another forward difference of  $\tilde{S}_2^0$  yields

$$\begin{aligned}\Delta_h \tilde{S}_2^0(t) &= \Delta_h r(t+h, a(t+h))S_2^0(t) + r(t+2h, a(t+2h))\Delta_h S_2^0(t) + \Delta_h a'(t) \cos \eta \Delta_h \tilde{G}(t, a(t)) \\ &\quad + a'(t+h) \cos \eta \Delta_h^2 \tilde{G}(t, a(t)) + \Delta_h a'(t) \Delta_h r(t, a(t)) \partial_1 \tilde{G}(t, a(t)) \\ &\quad + a'(t+h) \Delta_h^2 r(t, a(t)) \partial_1 \tilde{G}(t, a(t)) + a'(t+h) \Delta_h r(t+h, a(t+h)) \Delta_h \partial_1 \tilde{G}(t, a(t)).\end{aligned}$$

Let  $T$  denote the function from Lemma A.10. We observe,

$$\begin{aligned}a'(t+h) \Delta_h^2 \tilde{G}(t, a(t)) &= a'(t+h) (\Delta_h \tilde{G}(t+h, a(t+h)) - \Delta_h \tilde{G}(t, a(t))) \\ &= a'(t+h) \Delta_h \tilde{G}(t+h, a(t+h)) - a'(t) \Delta_h \tilde{G}(t, a(t)) + (a'(t) - a'(t+h)) \Delta_h \tilde{G}(t, a(t)) \\ &= a'(t+h) \Delta_h \tilde{G}(t+h, a(t+h)) - a'(t) \Delta_h \tilde{G}(t, a(t)) - \Delta_h a'(t) \Delta_h \tilde{G}(t, a(t)) \\ &= T(t+h) - T(t) - \Delta_h a'(t) \Delta_h \tilde{G}(t, a(t)) = \Delta_h T(t) - \Delta_h a'(t) \Delta_h \tilde{G}(t, a(t)).\end{aligned}$$

Now we know by Lemma A.10 that there is a decomposition  $(T_1^k, T_2^k)_k$  of  $T$  of length  $m+1$  with the specific properties given there. This allows to decompose  $\Delta_h T = \Delta_h T_1^0 + T_1^1 + T_2^1$  and we obtain

$$a'(t+h) \Delta_h^2 \tilde{G}(t, a(t)) = \Delta_h T_1^0(t) + T_1^1(t) + T_2^1(t) - \Delta_h a'(t) \Delta_h \tilde{G}(t, a(t)).$$

Using this observation we obtain

$$\begin{aligned}\Delta_h \tilde{S}_2^0(t) &= \Delta_h r(t+h, a(t+h))S_2^0(t) + r(t+2h, a(t+2h))S_1^1(t) + r(t+2h, a(t+2h))S_2^1(t) \\ &\quad + \Delta_h a'(t) \cos \eta \Delta_h \tilde{G}(t, a(t)) + \cos \eta \Delta_h T_1^0(t) + \cos \eta (T_1^1(t) + T_2^1(t)) - \cos \eta \Delta_h a'(t) \Delta_h \tilde{G}(t, a(t)) \\ &\quad + \Delta_h a'(t) \Delta_h r(t, a(t)) \partial_1 \tilde{G}(t, a(t)) + a'(t+h) \Delta_h^2 r(t, a(t)) \partial_1 \tilde{G}(t, a(t)) \\ &\quad + a'(t) \Delta_h r(t+h, a(t+h)) \Delta_h \partial_1 \tilde{G}(t, a(t)) + (a'(t+h) - a'(t)) \Delta_h r(t+h, a(t+h)) \Delta_h \partial_1 \tilde{G}(t, a(t))\end{aligned}$$

and further

$$\begin{aligned}\Delta_h \tilde{S}_2^0(t) &= [r(t+2h, a(t+2h))S_1^1(t) + \Delta_h a'(t) \cos \eta \Delta_h \tilde{G}(t, a(t)) \\ &\quad + \cos \eta \Delta_h T_1^0(t) - \cos \eta \Delta_h a'(t) \Delta_h \tilde{G}(t, a(t)) + \cos \eta T_1^1(t) \\ &\quad + \Delta_h a'(t) \Delta_h r(t, a(t)) \partial_1 \tilde{G}(t, a(t)) + a'(t+h) \Delta_h^2 r(t, a(t)) \partial_1 \tilde{G}(t, a(t)) \\ &\quad + \Delta_h a'(t) \Delta_h r(t+h, a(t+h)) \Delta_h \partial_1 \tilde{G}(t, a(t)) + \Delta_h r(t+h, a(t+h))S_1^0(t)] \\ &\quad + [r(t+2h, a(t+2h))S_2^1(t) + \cos \eta T_2^1(t) + 2\Delta_h r(t+h, a(t+h))S_2^0(t)] \\ &=: \tilde{S}_1^1(t) + \tilde{S}_2^1(t).\end{aligned}$$

Now we can split  $\Delta_h \tilde{S}_2^0 = \tilde{S}_1^1 + \tilde{S}_2^1$  with

$$\begin{aligned}\tilde{S}_1^1(t) &= r(t+2h, a(t+2h))S_1^1(t) + \cos \eta \Delta_h a'(t) \Delta_h \tilde{G}(t, a(t)) + \Delta_h a'(t) \Delta_h r(t, a(t)) \partial_1 \tilde{G}(t, a(t)) \\ &\quad + a'(t+h) \Delta_h^2 r(t, a(t)) \partial_1 \tilde{G}(t, a(t)) + \Delta_h r(t+h, a(t+h))S_1^0(t) + \cos \eta \Delta_h T_1^0(t) \\ &\quad + \Delta_h a'(t) \Delta_h r(t+h, a(t+h)) \Delta_h \partial_1 \tilde{G}(t, a(t)) - \cos \eta \Delta_h a'(t) \Delta_h \tilde{G}(t, a(t)) + \cos \eta T_1^1(t), \\ \tilde{S}_2^1(t) &= 2\Delta_h r(t+h, a(t+h))S_2^0(t) + r(t+2h, a(t+2h))S_2^1(t) + \cos \eta T_2^1(t).\end{aligned}$$

These terms have the desired properties. To see this, we calculate

$$\begin{aligned}
\sup_{t \in \mathbb{R}} |r(t+2h, a(t+2h))S_1^1(t)| &\lesssim hh^{m-1}h^\beta |\sin \eta|^{-1-\beta}, \\
\sup_{t \in \mathbb{R}} |\Delta_h a'(t)\Delta_h \tilde{G}(t, a(t))| &\lesssim \delta_j h^{\beta-1} |\sin \eta|^{-1-\beta} \cdot \delta_j^{m-1} h, \\
\sup_{t \in \mathbb{R}} |\Delta_h a'(t)\Delta_h r(t, a(t))\partial_1 \tilde{G}(t, a(t))| &\lesssim \delta_j h^{\beta-1} |\sin \eta|^{-1-\beta} \cdot h \cdot \delta_j^{m-1}, \\
\sup_{t \in \mathbb{R}} |a'(t+h)\Delta_h^2 r(t, a(t))\partial_1 \tilde{G}(t, a(t))| &\lesssim |\sin \eta|^{-1} \cdot h^\beta \delta_j |\sin \eta|^{-\beta} \cdot \delta_j^{m-1}, \\
\sup_{t \in \mathbb{R}} |\Delta_h r(t+h, a(t+h))S_1^0(t)| &\lesssim hh^{m-1}h^\beta |\sin \eta|^{-1-\beta}, \\
\sup_{t \in \mathbb{R}} |\Delta_h T_1^0(t)| &\lesssim h^m h^\beta |\sin \eta|^{-1-\beta}, \\
\sup_{t \in \mathbb{R}} |\Delta_h a'(t)\Delta_h \tilde{G}(t, a(t))| &\lesssim \delta_j h^{\beta-1} |\sin \eta|^{-1-\beta} \cdot \delta_j^{m-1} h, \\
\sup_{t \in \mathbb{R}} |\Delta_h a'(t)\Delta_h r(t+h, a(t+h))\Delta_h \partial_1 \tilde{G}(t, a(t))| &\lesssim \delta_j h^{\beta-1} |\sin \eta|^{-1-\beta} \cdot h \cdot \delta_j^{m-1}, \\
\sup_{t \in \mathbb{R}} |T_1^1(t)| &\lesssim \delta_j^m h^\beta |\sin \eta|^{-1-\beta},
\end{aligned}$$

and

$$\begin{aligned}
\sup_{t \in \mathbb{R}} |2\Delta_h r(t+h, a(t+h))S_2^0(t)| &\lesssim hh^{m-1} |\sin \eta|^{-1}, \\
\sup_{t \in \mathbb{R}} |r(t+2h, a(t+2h))S_2^1(t)| &\lesssim hh^{m-1} |\sin \eta|^{-1}, \\
\sup_{t \in \mathbb{R}} |T_2^1(t)| &\lesssim hh^{m-1} |\sin \eta|^{-1}.
\end{aligned}$$

We proceed with

$$\begin{aligned}
\Delta_h \tilde{S}_2^1(t) &= [\cos \eta T_1^2(t) + r(t+3h, a(t+3h))S_1^2(t) + 2\Delta_h r(t+2h, a(t+2h))S_1^1(t) \\
&\quad + 2\Delta_h^2 r(t+h, a(t+h))S_2^0(t)] + [\cos \eta T_2^2(t) + r(t+3h, a(t+3h))S_2^2(t) \\
&\quad + 3\Delta_h r(t+2h, a(t+2h))S_2^1(t)] =: \tilde{S}_1^2(t) + \tilde{S}_2^2(t).
\end{aligned}$$

Inductively, we put for  $k = 1, \dots, m+1$ , where for convenience  $T_1^{m+2} = 0$ ,

$$\begin{aligned}
\tilde{S}_1^{k+1}(t) &:= \cos \eta T_1^{k+1}(t) + r(t+(k+2)h, a(t+(k+2)h))S_1^{k+1}(t) \\
&\quad + (k+1)\Delta_h r(t+(k+1)h, a(t+(k+1)h))S_1^k(t) + (k+1)\Delta_h^2 r(t+kh, a(t+kh))S_2^{k-1}(t), \\
\tilde{S}_2^k(t) &:= \cos \eta T_2^k(t) + r(t+(k+1)h, a(t+(k+1)h))S_2^k(t) + (k+1)\Delta_h r(t+kh, a(t+kh))S_2^{k-1}(t).
\end{aligned}$$

These terms clearly satisfy  $\Delta_h \tilde{S}_2^k = \tilde{S}_1^{k+1} + \tilde{S}_2^{k+1}$ . They also have the desired properties since

$$\begin{aligned}
\sup_{t \in \mathbb{R}} |T_1^{k+1}(t)| &\lesssim h^m h^\beta |\sin \eta|^{-1-\beta}, \\
\sup_{t \in \mathbb{R}} |r(t+(k+2)h, a(t+(k+2)h))S_1^{k+1}(t)| &\lesssim h \cdot h^{m-1} h^\beta |\sin \eta|^{-1-\beta}, \\
\sup_{t \in \mathbb{R}} |\Delta_h r(t+(k+1)h, a(t+(k+1)h))S_1^k(t)| &\lesssim h \cdot h^{m-1} h^\beta |\sin \eta|^{-1-\beta}, \\
\sup_{t \in \mathbb{R}} |\Delta_h^2 r(t+kh, a(t+kh))S_2^{k-1}(t)| &\lesssim h^\beta \delta_j |\sin \eta|^{-\beta} \cdot h^{m-1} |\sin \eta|^{-1},
\end{aligned}$$

and

$$\begin{aligned}
\sup_{t \in \mathbb{R}} |T_2^k(t)| &\lesssim h^m |\sin \eta|^{-1}, \\
\sup_{t \in \mathbb{R}} |r(t+(k+1)h, a(t+(k+1)h))S_2^k(t)| &\lesssim h \cdot h^{m-1} |\sin \eta|^{-1}, \\
\sup_{t \in \mathbb{R}} |\Delta_h r(t+kh, a(t+kh))S_2^{k-1}(t)| &\lesssim h \cdot h^{m-1} |\sin \eta|^{-1}.
\end{aligned}$$



Since  $S_2^{m+1} = S_1^{m+2} = S_2^{m+2} = T_2^{m+1} = T_1^{m+2} = 0$ , for  $k = m + 1$  these expressions read

$$\begin{aligned}\tilde{S}_1^{m+2}(t) &= (m+2)\Delta_h r(t+(m+2)h, a(t+(m+2)h))S_1^{m+1}(t) \\ &\quad + (m+2)\Delta_h^2 r(t+(m+1)h, a(t+(m+1)h))S_2^m(t), \\ \tilde{S}_2^{m+1}(t) &= (m+2)\Delta_h r(t+(m+1)h, a(t+(m+1)h))S_2^m(t).\end{aligned}$$

We see that  $\Delta_h \tilde{S}_2^{m+1} = \tilde{S}_1^{m+2}$ . Therefore  $\tilde{S}_2^{m+2} = 0$  and the proof is finished.  $\square$

A slight modification of the previous proof leads to the following lemma.

**Lemma A.12.** *Let  $\tilde{G}(x) = r(x)^m G(x)$  for  $x \in \mathbb{R}^2$  and  $m \in \mathbb{N}_0$  and  $h \asymp 2^{-j(1-\alpha)}$ . The function  $\tilde{S} : \mathbb{R}^2 \rightarrow \mathbb{R}$  given by  $\tilde{S}(t, \tau) = a'(\tau)\Delta_h \partial_1 \tilde{G}(t, a(\tau))$  for  $(t, \tau) \in \mathbb{R}^2$  admits a decomposition  $(\tilde{S}_1^k, \tilde{S}_2^k)_k$  of the form (\*) of length  $m+1$  with estimates*

$$\begin{aligned}\sup_{t \in \mathbb{R}} \sup_{\tau \in [t-h, t+h]} |\tilde{S}_1^k(t, \tau)| &\lesssim h^{m-1} h^\beta |\sin \eta|^{-1-\beta}, \quad k = 0, \dots, m+1, \\ \sup_{t \in \mathbb{R}} \sup_{\tau \in [t-h, t+h]} |\tilde{S}_2^k(t, \tau)| &\lesssim h^{m-1} |\sin \eta|^{-1}, \quad k = 0, \dots, m.\end{aligned}$$

*Proof.* A small adaption of the previous proof is required to account for the little deviation of  $\tau$  from  $t$ . We just make the following remark. For  $t, \tau \in \mathbb{R}$  we have  $r(t, a(\tau)) = r(t, a(t)) + (a(t) - a(\tau)) \sin \eta$ . It follows for  $h \in \mathbb{R}$

$$\sup_{\tau \in [t-h, t+h]} |r(t, a(\tau))| \leq |r(t, a(t))| + |h \sin \eta| \|a'\|_\infty \lesssim |r(t, a(t))| + |h|.$$

Since  $h \asymp 2^{-j(1-\alpha)}$  this additional term poses no problem in the estimations.  $\square$