

Asymptotic Analysis of Inpainting via Universal Shearlet Systems*

Martin Genzel[†] and Gitta Kutyniok[‡]

Abstract. Recently introduced inpainting algorithms using a combination of applied harmonic analysis and compressed sensing have turned out to be very successful. One key ingredient is a carefully chosen representation system which provides (optimally) sparse approximations of the original image. Due to the common assumption that images are typically governed by anisotropic features, directional representation systems have often been utilized. One prominent example of this class are *shearlets*, which have the additional benefit to allowing faithful implementations. Numerical results show that shearlets significantly outperform wavelets in inpainting tasks. One of those software packages, www.shearlab.org, even offers the flexibility of using a different parameter for each scale, which is not yet covered by shearlet theory.

In this paper, we first introduce *universal shearlet systems* which are associated with an arbitrary scaling sequence, thereby modeling the previously mentioned flexibility. In addition, this novel construction allows for a smooth transition between wavelets and shearlets and therefore enables us to analyze them in a uniform fashion. For a large class of such scaling sequences, we first prove that the associated universal shearlet systems form band-limited Parseval frames for $L^2(\mathbb{R}^2)$ consisting of Schwartz functions. Secondly, we analyze the performance for inpainting of this class of universal shearlet systems within a distributional model situation using an ℓ^1 -analysis minimization algorithm for reconstruction. Our main result in this part states that, provided the scaling sequence is comparable to the size of the (scale-dependent) gap, nearly-perfect inpainting is achieved at sufficiently fine scales.

Key words. Co-Sparsity, Compressed Sensing, Inpainting, ℓ_1 Minimization, Multiscale Representation Systems, Shearlets, Sparse Approximation, Wavelets

AMS subject classifications. 42C40, 42C15, 65J22, 65T60, 68U10

1. Introduction. In data processing, due to frequent failures of technologies, one of the most natural questions to ask is whether and how lost data can be recovered. In the context of imaging sciences, this is termed *inpainting* due to the according physical task when restoring a painting. Given a digital image which is corrupted, it is certainly impossible to reconstruct the missing content without any prior knowledge. For instance, one may think of an image of the sky where several birds can be seen, but with one bird now being deleted. Hence, each inpainting method has to make certain further assumptions on the given image signal.

Inpainting algorithms based on variational approaches propagate information from the boundaries and intend to guarantee smoothness; see, e.g., [1, 2, 6]. Those which are based on applied harmonic analysis—typically combined with ideas of compressed sensing—encode prior information by assuming that a carefully selected representation system provides sparse approximations of the original image; and we refer to the pioneering paper [13] as well as [4, 10, 22]. It should also be mentioned that most of these approaches assume that the position and shape of the area that requires inpainting is known. If this is not the case, “blind inpainting” has to be applied, which is, as one can easily imagine, an even much more difficult task; first approaches can be found in [3, 10].

In this paper, we will focus on inpainting via a combination of applied harmonic analysis and compressed sensing in the sense of ℓ^1 -analysis minimization and assuming that the missing area is known. In this setting, the directional representation system of shearlets has recently been shown to outperform not only wavelets, but also most other directional systems [34]. This raises the challenge to develop a theoretical framework which is particularly able to relate the structural difference between

*May 15, 2014

[†]Department of Mathematics, Technische Universität Berlin, 10623 Berlin, Germany (genzel@math.tu-berlin.de)

[‡]Department of Mathematics, Technische Universität Berlin, 10623 Berlin, Germany (kutyniok@math.tu-berlin.de)

wavelets and shearlets to the inpainting problem. A first study was performed in [26] which however could only show superiority of shearlets over wavelets for a basic thresholding algorithm.

In the sequel, we will pursue a different path and analyze a parametrized family of systems ranging from wavelets to shearlets, aiming to derive a deep understanding of the transition from isotropy to anisotropy for the considered inpainting approach. Along the way, with the introduction and analysis of *universal shearlet systems* as sparsifying systems, we will also provide a general framework—including wavelets and shearlets—which has the additional freedom to choose a different degree of anisotropy for each scale. Moreover, this approach has several impacts which also go beyond inpainting tasks (cf. Subsection 1.5).

1.1. Inpainting via Applied Harmonic Analysis and Compressed Sensing. Results of *compressed sensing* have shown that, assumed a signal is sparse within a frame, it can be recovered from highly underdetermined, non-adaptive linear measurements by ℓ^1 -minimization (cf. [9]). For this, recall that a *frame* for a Hilbert space \mathcal{H} is a sequence $\Phi = (\phi_i)_{i \in I}$ satisfying $A\|x\|^2 \leq \|(\langle x, \phi_i \rangle)_{i \in I}\|_{\ell^2(I)}^2 \leq B\|x\|^2$ for all $x \in \mathcal{H}$. If the frame bounds A, B can be chosen to be equal to 1, it is typically called *Parseval frame*.

This approach can be also utilized for solving the inverse problem of inpainting. Letting x^0 denote the original signal, we assume that x^0 is sparse within some frame Φ . Due to the redundancy of frames, there exist infinitely many expansions of x^0 in Φ . Therefore, it is typically assumed that the *analysis operator*

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

maps x^0 to a sparse sequence—this is nowadays referred to as *co-sparsity* [36]. In order to recover the original signal x^0 from its known part $P_K x^0$ (here P_K denotes the orthogonal projection onto the “known” subspace of \mathcal{H}), the algorithm in [13] suggests to solve the ℓ^1 -*analysis minimization problem* given by

$$\min_x \|(\langle x, \phi_i \rangle)_{i \in I}\|_{\ell^1(I)} \quad \text{subject to } P_K x = P_K x^0.$$

The key ingredient of this algorithmic approach is the choice of the sparsifying frame. In imaging science, this relates to the question of what an image actually is. Certainly, we cannot give a general answer to this question. However, one classical viewpoint states that images are governed by edge-like structures meaning that anisotropic (in the sense of directional) features are dominant. This point of view is also backed up by neurophysiology because it was shown that the human visual cortex is highly sensible to curvilinear structures [23, 8]. Images from applications such as seismic data might have the strongest bias towards such structures since they basically only consist of parabolic curves [21]. This is one of the reasons why directional representation systems are currently a focus of intense research.

1.2. From Classical Shearlets to Universal Shearlets. About 15 years ago, it was recognized that wavelet systems are not capable to optimally encode anisotropic features. Among the approaches for directional representation systems, that were suggested to cope with this problem, the framework of *curvelets* can be justifiably regarded as a breakthrough. Indeed, curvelet systems were the first representation system which have been shown to provide optimally sparse approximations for a certain class of model functions, called *cartoon-like functions*. These are functions of the form $f = f_0 + f_1 \chi_B$, where $B \subset [0, 1]^2$ is a set with ∂B being a closed C^2 -curve with bounded curvature, and $f_i \in C^2(\mathbb{R}^2)$ are functions with $\text{supp } f_i \subset [0, 1]^2$ and $\|f_i\|_{C^2} \leq 1$ for $i = 0, 1$ (cf. [5]).

Shearlet systems were originally introduced in [17] as an attempt to introduce a directional representation system with similar optimal approximation properties, but with the advantage of allowing for a unified treatment of the continuum and digital situation in the sense of faithful implementations. This could be achieved by using *parabolic scaling, shearing* as a means to change the orientation—curvelets

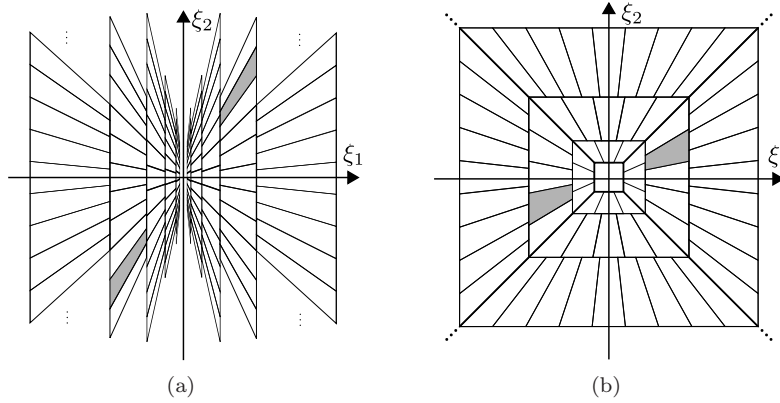


Figure 1. (a) Frequency tiling of a classical shearlet system. (b) Frequency tiling of a cone-adapted shearlet system.

employ rotation which destroys the digital grid—, and *translation* applied to very few generating functions. More precisely, these geometric transformations operate on some generator $\psi \in L^2(\mathbb{R}^2)$, leading to the system

$$(1.1) \quad \{\psi_{j,l,k} := 2^{3j/2} \psi(S^l A^j [\cdot] - k) \mid j \in \mathbb{Z}, l \in \mathbb{Z}, k \in \mathbb{Z}^2\},$$

where

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}, \quad S = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

denote the parabolic scaling matrix and shearing matrix, respectively. The tiling of frequency domain of such a simple shearlet system is illustrated in Figure 1(a). However, to avoid a directional bias and to promote more symmetry in the definition, this naive approach is usually enhanced by a *cone-adaptation*, see also Figure 1(b).

The original definition used band-limited generators, and the associated “classical” (*cone-adapted*) *shearlet systems* indeed provide optimally sparse approximations of cartoon-like functions [18], while constituting a Parseval frame for $L^2(\mathbb{R}^2)$ at the same time. However, due to an application of truncations in frequency domain, this frame has non-smooth Fourier transforms leading to bad spatial decay. But as suggested in [19], a slight modification of the seam elements can be used to obtain a band-limited Parseval frame for $L^2(\mathbb{R}^2)$ which consists of Schwartz functions. In order to achieve superior localization, compactly supported shearlets have then been introduced in [27], also providing similarly optimal sparsity (cf. [33]), but coming with the drawback that they do not form a Parseval frame.

It is certainly desirable to understand the structural difference between wavelets and shearlets, not only from a theoretical side to bridge the gap and to derive a deep understanding of the transition of properties, but also from a practical perspective to provide a parametrized family that allows optimal encoding of different anisotropies. A first approach in this direction were so-called α -*shearlets* [24, 32], which form compactly supported (non-Parseval) frames. But as mentioned before, for some applications band-limitedness or/and the Parseval property are crucial. A natural approach to tackle this problem is to transfer the construction from [19] to this setting.

Intriguingly—and this is by far not straightforward—, one problem is the complicated interaction of the scaling parameter with the overlapping of the seam elements at the boundaries of the cones (see Figure 1(b)). This requires the introduction of *scaling sequences* $(\alpha_j)_j \subseteq (-\infty, 2)$ with associated scaling matrices

$$A_{\alpha_j}^j = \begin{pmatrix} 4 & 0 \\ 0 & 2^{\alpha_j} \end{pmatrix},$$

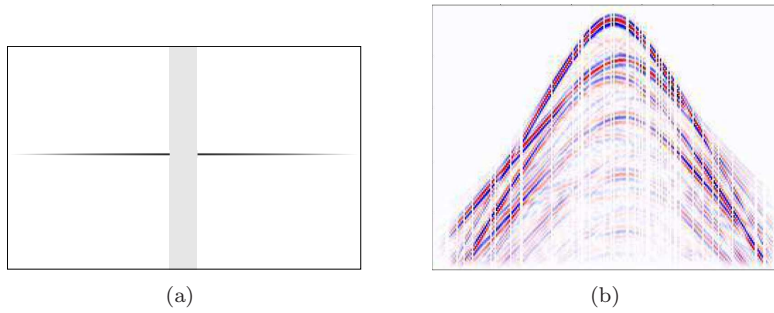


Figure 2. (a) Line singularity with a vertical strip removed. (b) Synthetic seismic data; some sensors failed to work resulting in randomly distributed missing traces (image from [21]).

finally offering a maximal flexibility in scaling, since the scaling parameter could vary independently for each j .

Based on this idea, we introduce *universal shearlet systems* in Definition 3.3. This notion now enables us for the first time to treat various band-limited systems—including classical cone-adapted shearlets, α -shearlets, and band-limited wavelets—in a uniform manner. Closing the gap between (band-limited) shearlets and wavelets in a continuous way does in particular allow for a common analysis. In addition, the relaxation in the directional scaling turns out to be even flexible enough to prove that, for a large class of scaling sequences, the associated universal shearlets form a band-limited Parseval frame for $L^2(\mathbb{R}^2)$ consisting of Schwartz functions (cf. Theorem 3.4).

The underlying construction of universal shearlet systems presented in Section 3 is self-contained. The constructed systems might be interesting also for other applications because of their desirable properties: smoothness, excellent space-frequency localization, and infinitely many vanishing moments. It should be mentioned that the shearlet toolbox, presented in [34] and available for download from www.shearlab.org, utilizes this novel construction technique and provides an implementation of compactly supported universal shearlets.

1.3. Our Setting. Universal shearlets are indeed perfectly suited to serve as sparsifying system in our analysis of the previously discussed inpainting algorithm, since they in particular allow for a unified view of wavelets and shearlets.

Concerning a suitable image model, remembering that we aim to model curvilinear structures, a discrete setting seems inappropriate because it would obscure the geometric content. Therefore, we choose a continuous model, more precisely, a *distributional model* which mimics a straight (horizontal) line. Since images are typically defined on bounded domains, this line is compactly supported and smoothly truncated at its endpoints to avoid any point singularities. The mask has the shape of the vertical strip and cuts out the middle part of the line (see Figure 2(a)). As simple as this approach seems, it even provides a quite accurate model for some applications; for example in seismology, where the absence of sensors produces white strips in the data, which basically only consist of smooth curves (see Figure 2(b)). Also, this model still enables us to relate the size of the gap with the length of the elements of the universal shearlet system, leading to very natural and intuitive necessary conditions for inpainting to succeed.

On the algorithmic side, we will choose the ℓ^1 -analysis minimization approach, which has been already discussed in Subsection 1.1.

1.4. Asymptotic Inpainting. Through a bandpass filtering process on the model, and simultaneously allowing the gap size to vary with the scale, we are able to perform an asymptotic analysis depending on the scale. Certainly, in practical applications we can only consider a finite number of scales and the gap size usually stays fixed, but the main goal of our analysis is to derive theoretical

insight into the interplay between the scaling behavior of the sparsifying system and the gap size.

Our main result concerning the recovery error of inpainting in the setting just described, Theorem 5.1, then proves that, assuming the gap size is asymptotically smaller than the length of the corresponding shearlet elements, *asymptotically perfect inpainting* is achieved. Indeed, the necessary condition provides a deep insight into the relation between the degree of anisotropy of the underlying system and the admissible gap size; and it particularly shows the structural difference between wavelets and shearlets.

1.5. Expected Impact. We anticipate our results to have the following impacts:

- *ShearLab*: A comprehensive software package for the 2D and 3D shearlet transform is available at www.shearlab.org, see also [34]. The provided algorithms offer the user to specify various parameters among which are also scaling parameters. However, the current shearlet theory does not incorporate this amount of flexibility. With the introduction and analysis of universal shearlets, now a first theoretical framework is available, giving also additional insight into convenient choices of scaling parameters.
- *Construction of compactly supported shearlet systems*: One main objective—which some people even coined the “holy grail” in applied harmonic analysis—is to construct a compactly supported directional representation system which forms a Parseval frame and provides optimally sparsifying approximations of cartoon-like functions. In shearlet theory, a certain class of band-limited shearlets which constitute a Parseval frame was shown to provide such optimal sparse approximations [18]. Some years later, compactly supported shearlets were introduced and under certain conditions similar approximation rates for cartoon-like functions could be achieved [33]; however those frames do not have the Parseval property. We believe that universal shearlets could be another attack point to this challenge due to their additional flexibility. It seems conceivable that a smart choice of scaling sequences could lead to optimally sparsifying, compactly supported shearlet systems whose frame bounds can be moved (arbitrarily) close to 1.
- *Structured dictionary learning*: Most dictionary learning algorithms suffer from the fact that the generated systems lack in structure such as good frame bounds or fast associated transforms. Some attempts have been made to equip the generated systems with structural properties such as, for instance, the double sparsity approach in [38]. The system of universal shearlets can be seen in this light, allowing an adaption to the data at hand by learning the scaling sequence, but still preserving their advantageous functional analytic and numerical properties. Thus, we expect this work to eventually initiate a new approach toward structured dictionary learning with choosable objectives such as, for instance, inpainting.
- *Interplay of systems and masks*: It is well-known that directional representation systems are superior to wavelets for inpainting via compressed sensing methodologies assumed that the image is governed by curvilinear structures. Rigorous results are however missing; only a first analysis of a simple thresholding algorithm was presented as one result of [26]. Deriving an analysis which enables to precisely show superiority of certain representation systems over others (depending on the size and structure of the mask) will require novel tools. But we strongly believe that universal shearlets and our inpainting results can serve as a first step. For instance, the notion of scaling sequences provides a means to compare different systems with respect to their anisotropy in a much more careful manner as possible before.

1.6. Outline. The paper is organized as follows. The abstract framework of inpainting and associated error estimates, in the setting of a general Hilbert space, are introduced and discussed in Section 2. The novel system of universal shearlets is introduced in Section 3 with the proof of the main result on its Parseval frame property being outsourced to Subsection 7.1. Section 4 is then devoted to the specific image model, that we exploit in our work, as well as its asymptotic decomposition. Having detailed the considered setting, our main result on the asymptotic behavior of shearlet inpainting and

its proof are presented in Section 5. Finally, some possible extensions are discussed (cf. Section 6), followed by some proofs in Section 7, which we decided to outsource due to their non-intuitive and more technical nature.

2. Abstract Inpainting Framework.

2.1. Modeling Spaces. We follow [26] and first tackle the problem of inpainting in the very general setting of a separable Hilbert space \mathcal{H} . As underlying and undamaged *signal*, which we ultimately want to recover, we choose an arbitrary element $x^0 \in \mathcal{H}$. To incorporate the fact that some “information” of x^0 got lost, we assume that \mathcal{H} decomposes into an orthogonal sum of two closed subspaces, namely a *known part* and a *missing part*,

$$\mathcal{H} = \mathcal{H}_K \oplus \mathcal{H}_M = P_K \mathcal{H} \oplus P_M \mathcal{H},$$

where P_K and P_M denote the orthogonal projections onto \mathcal{H}_K and \mathcal{H}_M , respectively. This leads to the following challenge: *Given a corrupt signal $P_K x^0$, recover the missing part $P_M x^0$.*

This includes in particular that the space \mathcal{H}_M is explicitly known to us. Moreover, the above task implies that we are rather looking for the “real” missing part $P_M x^0$ and not only for some completion of $P_K x^0$ which appears “reasonable” to the beholder (for example, if objects are removed from the background, cf. [28]). We finally wish to emphasize that the signal model and the space decomposition are defined independently, enabling us to analyze the x^0 and \mathcal{H}_M almost separately in Subsection 2.4.

In terms of image inpainting, we will consider a continuous image model situation where $\mathcal{H} = L^2(\mathbb{R}^2)$. Here, the missing space is going to be $\mathcal{H}_M = L^2(\mathcal{M})$ for some measurable set $\mathcal{M} \subset \mathbb{R}^2$, which should be seen as a *mask* covering the corrupted parts of the painting. Thus, it might be helpful to already think of this concrete setting, even when dealing with the abstract framework.

2.2. Inpainting via ℓ^1 -Minimization. In order to formulate an appropriate reconstruction method, it is indispensable to make further assumptions on the signal and missing space. Throughout this paper, we presume that x^0 can be “efficiently” represented by a certain Parseval frame $\Phi = (\phi_i)_{i \in I}$ for \mathcal{H} , which can be either chosen non-adaptively or adaptively, based on the given model. In the classical theory of sparse representations, this would ask for solving the ℓ^0 -minimization problem

$$(2.1) \quad \min_{c \in \ell^2(I)} \|c\|_{\ell^0(I)} \quad \text{subject to } x^0 = T_\Phi^* c = \sum_{i \in I} c_i \phi_i,$$

i.e., one tries to find a synthesis sequence of x^0 that has as few as possible non-zero elements measured by $\|c\|_{\ell^0} := \#\{c_i \mid c_i \neq 0\}$. Those coefficients may be later identified with the “significant” features of x^0 .

As in [11, 26], we will pursue a different strategy—the so-called *analysis approach*—which seems to be more appropriate for our purposes. Instead of regarding all synthesis sequences for x^0 , we merely consider the *analysis coefficients* $T_\Phi x^0 = (\langle x^0, \phi_i \rangle)_{i \in I}$ which, due to the Parseval property, allow for a reconstruction by $x^0 = T_\Phi^* (\langle x^0, \phi_i \rangle)_{i \in I} = \sum_{i \in I} \langle x^0, \phi_i \rangle \phi_i$. Since Φ does not necessarily need to constitute a basis, $T_\Phi x^0$ might not be a (sparse) solution of (2.1). It is therefore not completely clear what it means to provide a sparse representation in the sense of this analysis approach; a precise definition will follow in the next subsection. However, one major benefit of this strategy is the circumvention of certain problems that could occur when using synthesis coefficients, e.g., numerical instabilities. Moreover, we just aim for a good reconstruction in \mathcal{H} ; thus, knowing the sparsest representation might not give further profit [26].

Inspired by the ideas of *compressed sensing* (cf. [9, 14]), we now present a recovery algorithm which was already used in [25, 26]:

Algorithm 2.1 (Inpainting via ℓ^1 -minimization).

INPUT: *corrupted signal* $P_K x^0 \in \mathcal{H}_K$, *Parseval frame* $\Phi = (\phi_i)_{i \in I}$ for \mathcal{H}

COMPUTE:

$$(\ell^1\text{-INP}) \quad x^* = \operatorname{argmin}_{x \in \mathcal{H}} \|T_\Phi x\|_{\ell^1(I)} \quad \text{subject to } P_K x^0 = P_K x$$

OUTPUT: *recovered signal* $x^* \in \mathcal{H}$

Summarized in words, Algorithm 2.1 minimizes the ℓ^1 -norm among all possible reconstruction candidates, which is the set of all signals coinciding with x^0 on \mathcal{H}_K . If we assume that the undamaged signal x^0 is sufficiently sparsified by Φ —this particularly implies a small ℓ^1 -norm of $T_\Phi x^0$ —the completion x^* hopefully provides a good recovery respecting the significant features of x^0 .

The numerical realization of Algorithm 2.1 implicitly requires that the minimum in $(\ell^1\text{-INP})$ is really attained. However, the error estimate of Theorem 2.5 (see Subsection 2.4) still holds for all $x \in \mathcal{H}$ satisfying $P_K x^0 = P_K x$ and $\|T_\Phi x\|_{\ell^1} \leq \|T_\Phi x^0\|_{\ell^1}$. Assuming that $\|T_\Phi x^0\|_{\ell^1} < \infty$, these two conditions are trivially satisfied for $x = x^0$.

We wish to mention that for various applications, e.g., geometric separation [11], the analysis approach, combined with an ℓ^1 -optimization similar to $(\ell^1\text{-INP})$, has turned out to be successful for both empirical problems and theoretical analysis. Interestingly, Algorithm 2.1 is closely related to some other optimization methods. For instance, it can be viewed as a relaxation of the following *co-sparsity* problem:

$$x^* = \operatorname{argmin}_{x \in \mathcal{H}} \|T_\Phi x\|_{\ell^0(I)} \quad \text{subject to } P_K x^0 = P_K x.$$

General theoretical results on co-sparsity can be found in [35, 36]. Another class of reconstruction methods is the one of thresholding algorithms. In a certain sense, they can be seen as approximations of the ℓ^1 -problem given by $(\ell^1\text{-INP})$. A very basic version of thresholding, called “one-step thresholding”, was introduced in [30] and in an adapted form used in [26]. It surprisingly turned out that very similar convergence results can be proven for this approach (cf. [26]), although the algorithmic approach seems much cruder than $(\ell^1\text{-INP})$.

2.3. Analysis Tools.

2.3.1. δ -Clustered Sparsity. To prove appropriate error estimates for recovery by Algorithm 2.1, we first need to introduce several analysis tools that give us further insight into the structure of the abstract model. The following definition introduces the novel concept of *clustered sparsity*. It was originally used to study the geometric separation problem [11], and precisely specifies our notion of (approximate) sparsity:

Definition 2.2 ([26]). Let $\delta > 0$ and $\Lambda \subset I$. A signal $x \in \mathcal{H}$ is called δ -clustered sparse in Φ with respect to Λ , if

$$(2.2) \quad \|\mathbf{1}_{\Lambda^c} T_\Phi x\|_{\ell^1} \leq \delta.$$

In this case, Λ is said to be a δ -cluster for x in Φ .

Intuitively spoken, being δ -clustered sparse for a small δ means that all coefficients lying outside of Λ are small, or vice versa: The analysis coefficients are highly concentrated on Λ . However, the coefficients contained in Λ do not necessarily need to be large. In particular, successively enlarging Λ ensures (2.2) for smaller δ , i.e., better sparsity can be always achieved independent on the structure of x . According to this, one should regard Λ rather as an analysis tool. In fact, Definition 2.2 only makes sense when x is δ -clustered sparse with respect to a “small” cluster Λ , and δ is sufficiently small at the

same time. The concept of clustered sparsity is closely related to classical notions of (approximate) sparsity in this situation because $T_\Phi x = (\langle x^0, \phi_i \rangle)_{i \in I}$ contains only a few large elements (lying in Λ).

Concerning inpainting, one major task is therefore to design a dictionary Φ which provides sufficient (clustered) sparsity for the given benchmark signal x^0 . In this way, the significant features of x^0 get encoded in terms of an appropriate δ -cluster, giving us deeper insight into the “geometric” structure of x^0 .

2.3.2. Cluster Coherence. We did not touch any specific properties of \mathcal{H}_M up to now. Hence, there is still the need for a concept that enables us to analyze the missing part of x^0 :

Definition 2.3 ([26]). *Let $\Lambda \subset I$. The cluster coherence of Φ with respect to \mathcal{H}_M and Λ is given by*

$$\mu_c(\Lambda, P_M \Phi) := \max_{j \in I} \sum_{i \in \Lambda} |\langle P_M \phi_i, P_M \phi_j \rangle|,$$

where $P_M \Phi := (P_M \phi_i)_{i \in I}$.

Compared with other measures for the “size” of \mathcal{H}_M , e.g., the notion of *concentration* introduced in [26], $\mu_c(\Lambda, P_M \Phi)$ has a very explicit form and is rather easy to estimate. Note that the signal x^0 is not involved in this definition, and Λ does not need to be a δ -cluster; but later on, we will use one and the same cluster for both definitions.

The following aspects shall help to acquire a better understanding of why the cluster coherence can be viewed as an abstract measure for the gap size: Since we do not know $P_M x^0$, it is natural to ask for the “maximal amount” of missing information. More precisely, we consider the Parseval expansion $x^0 = \sum_{j \in I} \langle x^0, \phi_j \rangle \phi_j$ and estimate the analysis coefficients of its projection onto \mathcal{H}_M :

$$(2.3) \quad |\langle P_M x^0, \phi_i \rangle| \leq \sum_{i \in I} |\langle x^0, \phi_j \rangle| |\langle P_M \phi_j, \phi_i \rangle| = \sum_{i \in I} |\langle x^0, \phi_j \rangle| |\langle P_M \phi_j, P_M \phi_i \rangle|.$$

Similar to classical notions of coherence (cf. [12]), the scalar products $|\langle P_M \phi_j, P_M \phi_i \rangle|$ indicate to which extent measurements with ϕ_i and ϕ_j , respectively, are correlated on \mathcal{H}_M . Assuming that x^0 is also δ -clustered sparse (with respect to some cluster Λ), we may restrict in (2.3) to those frame elements ϕ_i which are contained in Λ . In this sense, the cluster coherence can be interpreted as a signal-independent magnitude for the lack of information; but note that the actual value of $\mu_c(\Lambda, P_M \Phi)$ results from a strong interaction between the system Φ and the “geometry” of \mathcal{H}_M .

2.4. Abstract Error Estimate.

2.4.1. The ℓ^1 -Analysis Space. First of all, we shall investigate the question of how to measure the recovery error. In the pioneer works of [11, 26], the Hilbert space norm was used to estimate the distance between x^0 and its ℓ^1 -recovery x^* . For an estimate from above, this indeed worked well; but to show an optimality result—that is, an estimate from below—measuring with $\|\cdot\|$ seems to be inappropriate. This drawback is essentially caused by the (Parseval) frame property of Φ implying $\|x^0 - x^*\| = \|T_\Phi(x^0 - x^*)\|_{\ell^2}$. Compared to ℓ^1 , the ℓ^2 -norm has an averaging effect on the coefficients, and therefore, sparsity features might be hidden when using the Hilbert space norm. For the concrete example of image inpainting (cf. Subsection 5.3), we will see that $\|T_\Phi(x^0 - x^*)\|_{\ell^2}$ is virtually independent of the scaling parameters α_j , i.e., any estimate from below would not take into account the important feature of anisotropy.

Hence, we propose a new measure which naturally arises from our analysis approach of Algorithm 2.1, and enables us to detect sparsity.

Definition 2.4. *Let $x \in \mathcal{H}$ and Φ be a Parseval frame for \mathcal{H} . Then we define the ℓ^1 -analysis norm, with respect to Φ , by*

$$\|x\|_{1, \Phi} := \|T_\Phi x\|_{\ell^1} = \|(\langle x, \phi_i \rangle)_{i \in I}\|_{\ell^1},$$

and the ℓ^1 -analysis space is given by $\mathcal{H}_{1,\Phi} := \{x \in \mathcal{H} \mid \|x\|_{1,\Phi} < \infty\}$, equipped with $\|\cdot\|_{1,\Phi}$.

The tuple $(\mathcal{H}_{1,\Phi}, \|\cdot\|_{1,\Phi})$ indeed defines a normed vector space, due to the injectivity of T_Φ . Moreover,

$$(2.4) \quad \|x\| = \|T_\Phi x\|_{\ell^2} \leq \|T_\Phi x\|_{\ell^1} = \|x\|_{1,\Phi}$$

yields the embedding $\mathcal{H}_{1,\Phi} \hookrightarrow \mathcal{H}$. In particular, by choosing $x = x^* - x^0$, we conclude that small recovery errors in $\mathcal{H}_{1,\Phi}$ always imply small ones in \mathcal{H} .

To give more intuition, we may think of Φ to be a shearlet system in $L^2(\mathbb{R}^2)$ for the moment. In this situation, $\|x\|_{1,\Phi}$ becomes particularly large when x is governed by curvilinear structures. It was already mentioned in the introductory part that there is a surprising connection to the human visual system, which is highly sensible to directional features [23, 8]. Therefore, the ℓ^1 -analysis norm evaluates the content of images in a similar way as the human brain does, and a small reconstruction error will probably give good empirical inpainting results.

2.4.2. Recovery Error Estimate. We conclude this section with an abstract estimation of the recovery error. A detailed proof of the following theorem can be found in [26] (actually, the corresponding statement in [26] is an estimate for $\|x^* - x^0\|$, but the proof makes use of (2.4) in the very beginning and (2.5) is then shown explicitly).

Theorem 2.5. *Let $\delta > 0$ and $\Lambda \subset I$ be a δ -cluster for x^0 in Φ . Moreover, assume that $\mu_c(\Lambda, P_M\Phi) < 1/2$. If $x^0 \in \mathcal{H}_{1,\Phi}$, then the ℓ^1 -minimizer x^* of Algorithm 2.1 is also contained in $\mathcal{H}_{1,\Phi}$ and satisfies*

$$(2.5) \quad \|x^* - x^0\|_{1,\Phi} \leq \frac{2\delta}{1 - 2\mu_c(\Lambda, P_M\Phi)}.$$

Theorem 2.5 provides intuitively good estimates if Φ sparsifies x^0 and the size of the missing space, measured in terms of cluster coherence, is not too large. More precisely, we intend to choose a cluster Λ such that x^0 becomes δ -clustered sparse for small δ and, simultaneously, μ_c shall not exceed the bound of $1/2$. While the notions of cluster sparsity and cluster coherence were introduced independently in the previous subsection, both terms are now coupled in terms of Λ . In particular, a variation of this cluster reveals the essential trade-off relation: Enlarging Λ makes (2.5) true for smaller δ , whereas $\mu_c(\Lambda, P_M\Phi)$ increases and could exceed $1/2$ at some point. Applying Theorem 2.5 therefore involves the fundamental task of choosing an appropriate cluster [26].

In the following sections, the above abstract framework is transferred to the specific case of a continuous image model in $\mathcal{H} = L^2(\mathbb{R}^2)$. Based on Theorem 2.5, we will finally present a convergence result in Section 5 which rigorously proves the success of image inpainting via (universal) shearlets.

3. Universal Shearlets. It was already emphasized in the introductory part that shearlets provide an excellent, non-adaptive dictionary to represent anisotropic structures. In this section—which is essentially self-contained, and might be of interest on its own—we present a novel shearlet construction which overcomes the limitations of parabolic scaling. The crucial feature of our new system is a generalized scaling matrix,¹

$$A_{\alpha_j, (h)} = \begin{bmatrix} 2^{2j} & 0 \\ 0 & 2^{\alpha_j j} \end{bmatrix},$$

where the *scaling parameters* α_j can be chosen independently for each scaling level j . Compared with usual cone-adapted shearlet systems, this offers much more flexibility, and allows us to achieve an almost arbitrary scaling behavior. The following definitions are based on the original construction of Guo and Labate in [19]. In particular, our adaption still constitutes a smooth Parseval frame for

¹The (h) stands for horizontal and its meaning will become clear in the sequel.

$L^2(\mathbb{R}^2)$ and preserves the band-limiting property, as well as rapid decay in the spacial domain (cf. Theorem 3.4). An implementation of universal shearlets is provided in www.shearlab.org, see also [34].

Before presenting more technical details, let us recall the notion of *rapidly decreasing functions* (or *Schwartz functions*)

$$\mathcal{S}(\mathbb{R}^d) := \left\{ \varphi \in C^\infty(\mathbb{R}^d) \mid \forall K, N \in \mathbb{N}_0: \sup_{x \in \mathbb{R}^d} (1 + |x|^2)^{-N/2} \sum_{|\alpha| \leq K} |D^\alpha \varphi(x)| < \infty \right\}.$$

For the sake of convenience, we will make use of the compact notation $\langle |x| \rangle := (1 + |x|^2)^{-N/2}$. The *Fourier transform*, given by

$$\mathcal{F}: \mathcal{S}(\mathbb{R}^d) \rightarrow \mathcal{S}(\mathbb{R}^d), \quad \widehat{\varphi}(\xi) := \int_{\mathbb{R}^d} \varphi(x) e^{-2\pi i \xi^\top x} dx,$$

is then a well-defined operator, which extends to an unitary operator from $L^2(\mathbb{R}^d)$ to $L^2(\mathbb{R}^d)$ (cf. [15]).

3.1. Construction and Basic Properties. The basic component of our construction is a function $\phi \in \mathcal{S}(\mathbb{R})$ satisfying $0 \leq \widehat{\phi} \leq 1$, $\widehat{\phi}(u) = 1$ for $u \in [-\frac{1}{16}, \frac{1}{16}]$, and $\text{supp } \widehat{\phi} \subset [-\frac{1}{8}, \frac{1}{8}]$. A function with these properties is commonly known as a *Meyer scaling function* (see Figure 3(a)). Now, we define the *corona scaling functions* for $j \in \mathbb{N}_0$ by $(\xi = (\xi_1, \xi_2) \in \mathbb{R}^2)$

$$\begin{aligned} \widehat{\Phi}(\xi) &:= \widehat{\phi}(\xi_1) \widehat{\phi}(\xi_2), \\ W(\xi) &:= \sqrt{\widehat{\Phi}^2(2^{-2}\xi) - \widehat{\Phi}^2(\xi)}, \\ W_j(\xi) &:= W(2^{-2j}\xi). \end{aligned}$$

The definition of ϕ implies that W_j is compactly supported in corona-shaped *scaling levels*, i.e.,

$$(3.1) \quad \text{supp } W_j \subset \mathcal{K}_j := [-2^{2j-1}, 2^{2j-1}]^2 \setminus (-2^{2j-4}, 2^{2j-4})^2,$$

and—this is the crucial feature—the Fourier domain is decomposed by the sequence of scaling functions:

$$(3.2) \quad \widehat{\Phi}^2(\xi) + \sum_{j \geq 0} W_j^2(\xi) = 1, \quad \xi \in \mathbb{R}^2.$$

For an illustration, we refer to Figure 3(b). The function W can be viewed as a wavelet, in the sense that it satisfies (3.2), which is a version of the *discrete Calderón condition* (cf. [7]). In fact, this property is key to ensuring that our shearlet system is capable to detect edge singularities.

Next, we consider a *bump-like* function $v \in C^\infty(\mathbb{R})$ which satisfies $\text{supp } v \subset [-1, 1]$ and

$$(3.3) \quad |v(u-1)|^2 + |v(u)|^2 + |v(u+1)|^2 = 1 \text{ for } u \in [-1, 1], \text{ and}$$

$$(3.4) \quad v(0) = 1 \text{ and } v^{(n)}(0) = 0 \text{ for } n \geq 1.$$

An explicit construction of v can be found in [18]. Compared to W , the actual role of v will be to produce the directional scaling feature of the system.

The simplicity of the shearlet system that was defined in the introductory part of this work (see (1.1)) causes a directional bias which is already recognizable in Figure 1(a): For large shearing parameters l , the corresponding elements do increasingly elongate along the ξ_2 -axis (in the Fourier domain). This drawback will lead to problems when approximating horizontal edges and renders the system ineffective for practical applications [19]. To circumvent this problem, one first restricts the number of shears such that the remaining system is contained in the *horizontal frequency cone*

$$\mathcal{C}_{(h)} := \left\{ (\xi_1, \xi_2) \in \mathbb{R}^2 \mid \left| \frac{\xi_2}{\xi_1} \right| \leq 1 \right\}.$$

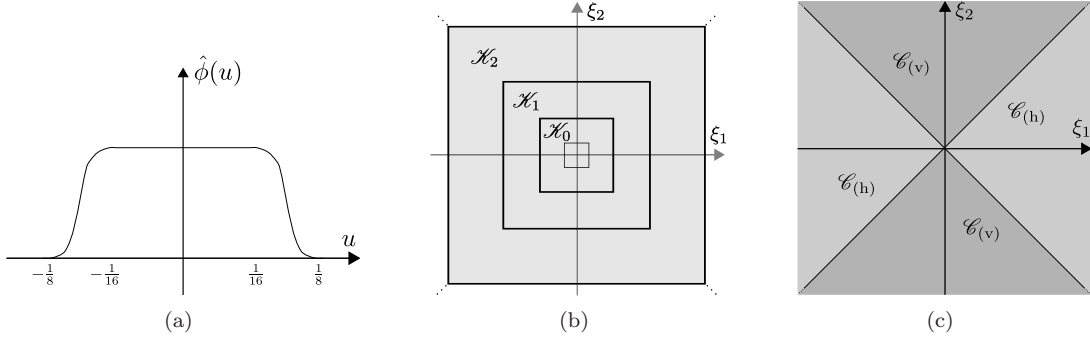


Figure 3. Ingredients of the construction: (a) Fourier transform of a Meyer scaling function. (b) Decomposition of the frequency plane by corona functions W_j and Φ . Note that the corona shapes \mathcal{K}_j could slightly overlap. (c) Symmetric frequency decomposition by cones.

Then, by interchanging the variables ξ_1 and ξ_2 , we can define a flipped system on the *vertical frequency cone*

$$\mathcal{C}_{(v)} := \left\{ (\xi_1, \xi_2) \in \mathbb{R}^2 \mid \left| \frac{\xi_1}{\xi_2} \right| \leq 1 \right\}.$$

This *cone-based approach*, initially introduced in [17], provides a much more symmetric frequency tiling (see also Figure 3(c)), and has turned out to be suited for both theoretical analysis and applications (cf. [31]).

To give a precise definition of a cone-adapted shearlet system, we still need to introduce adapted versions of the usual shearing and scaling matrix,

$$\begin{aligned} A_{\alpha, (h)} &:= \begin{bmatrix} 2^2 & 0 \\ 0 & 2^\alpha \end{bmatrix}, & S_{(h)} &:= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \\ A_{\alpha, (v)} &:= \begin{bmatrix} 2^\alpha & 0 \\ 0 & 2^2 \end{bmatrix}, & S_{(v)} &:= \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \end{aligned}$$

where $\alpha \in (-\infty, 2)$ is the *scaling parameter*. Similarly, the adapted *cone functions* are given by

$$V_{(h)}(\xi) := v\left(\frac{\xi_2}{\xi_1}\right), \quad V_{(v)}(\xi) := v\left(\frac{\xi_1}{\xi_2}\right), \quad \xi \in \mathbb{R}^2.$$

Now, we are able to define the single ingredients of a universal shearlet system; notice that the definition of a universal shearlet system will then follow in Definition 3.3. The respective parts of the following definition are visualized in Figure 4.

Definition 3.1. Let $\Phi, W, V_{(h)}, V_{(v)} \in L^2(\mathbb{R}^2)$ be defined as before.

1. Coarse scaling functions: For $k \in \mathbb{Z}^2$, we set

$$\psi_{-1, k}(x) := \Phi(x - k), \quad x \in \mathbb{R}^2.$$

2. Interior shearlets: Let $\alpha \in (-\infty, 2)$, $j \in \mathbb{N}_0$, $l \in \mathbb{Z}$ with $|l| < 2^{(2-\alpha)j}$, $k \in \mathbb{Z}^2$ and $\iota \in \{h, v\}$. Then we define $\psi_{j, l, k}^{\alpha, (\iota)}$ by its Fourier representation,

$$(3.5) \quad \widehat{\psi}_{j, l, k}^{\alpha, (\iota)}(\xi) := 2^{-(2+\alpha)j/2} W(2^{-2j}\xi) V_{(\iota)}\left(\xi^\top A_{\alpha, (\iota)}^{-j} S_{(\iota)}^{-l}\right) e^{-2\pi i \xi^\top A_{\alpha, (\iota)}^{-j} S_{(\iota)}^{-l} k}, \quad \xi \in \mathbb{R}^2.$$

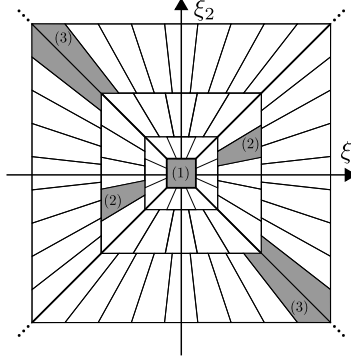


Figure 4. Illustration of the Fourier supports of the different components. (1) Coarse scaling part covering low frequencies. (2) Interior shearlets entirely contained in the interior of the cones. (3) Boundary shearlets “glued” together along the boundaries of the cones.

3. Boundary shearlets: For $\alpha \in (-\infty, 2)$, $j \geq 1$, $l = \pm \lceil 2^{(2-\alpha)j} \rceil$ and $k \in \mathbb{Z}^2$, we define

$$(3.6) \quad \widehat{\psi}_{j,l,k}^\alpha(\xi) := \begin{cases} 2^{-\frac{(2+\alpha)j}{2} - \frac{1}{2}} W(2^{-2j}\xi) V_{(h)} \left(\xi^\top A_{\alpha,(h)}^{-j} S_{(h)}^{-l} \right) e^{-\pi i \xi^\top A_{\alpha,(h)}^{-j} S_{(h)}^{-l} k}, & \xi \in \mathcal{C}_{(h)}, \\ 2^{-\frac{(2+\alpha)j}{2} - \frac{1}{2}} W(2^{-2j}\xi) V_{(v)} \left(\xi^\top A_{\alpha,(v)}^{-j} S_{(v)}^{-l} \right) e^{-\pi i \xi^\top A_{\alpha,(v)}^{-j} S_{(v)}^{-l} k}, & \xi \in \mathcal{C}_{(v)}, \end{cases}$$

and in the case $j = 0$, $l = \pm 1$, we put

$$\widehat{\psi}_{0,l,k}^\alpha(\xi) := \begin{cases} W(\xi) V_{(h)} \left(\xi^\top S_{(h)}^{-l} \right) e^{-2\pi i \xi^\top k}, & \xi \in \mathcal{C}_{(h)}, \\ W(\xi) V_{(v)} \left(\xi^\top S_{(v)}^{-l} \right) e^{-2\pi i \xi^\top k}, & \xi \in \mathcal{C}_{(v)}. \end{cases}$$

We remark that the case $\alpha = 1$ produces exactly the original parabolic-scaling construction of [19]. Compared to the Fourier transform of the “classical” shearlet system $\{\psi_{j,l,k} \mid j \in \mathbb{Z}, l \in \mathbb{Z}, k \in \mathbb{Z}^2\}$ in Subsection 1.2, the interior shearlets of equation (3.5) are not *shear-invariant*, meaning that W_j contains no shearing term. Hence, the system of Definition 3.1 has no chance to be singly generated. However, $\widehat{\psi}_{j,l,k}^{\alpha,(h)}$ has compact support in the trapezoidal region

$$(3.7) \quad \left\{ \xi \in \mathbb{R}^2 \mid \xi_1 \in [-2^{2j-1}, 2^{2j-1}] \setminus (-2^{2j-4}, 2^{2j-4}), \left| \frac{\xi_2}{\xi_1} - l 2^{-(2-\alpha)j} \right| \leq 2^{-(2-\alpha)j} \right\},$$

which essentially coincides with $\text{supp } \widehat{\psi}_{j,l,k}$ at least when $\alpha = 1$. Thus, $\psi_{j,l,k}^{\alpha,(h)}$ and $\psi_{j,l,k}$ have virtually the same space-frequency behavior. For a more detailed discussion see also [19].

Up to now, it is quite unclear if the elements of Definition 3.1 form a (Parseval) frame for $L^2(\mathbb{R}^2)$. A rather simple approach to satisfy the frame property would be to project the system onto the subspaces

$$(L^2(\mathcal{C}_{(i)}))^{\vee} = \{f \in L^2(\mathbb{R}^2) \mid \text{supp } \widehat{f} \subset \mathcal{C}_{(i)}\}, \quad i = h, v,$$

corresponding to a rough truncation of the Fourier supports. Then, one can show that these truncated systems constitute frames for $(L^2(\mathcal{C}_{(i)}))^{\vee}$, respectively (cf. [19]); their union indeed yields a frame for the whole $L^2(\mathbb{R}^2)$. However, cutting-off along the boundaries of the cones causes discontinuities which finally lead to bad decay behavior in space.

It is therefore much more favorable to consider the system on the entire space $L^2(\mathbb{R}^2)$. Due to the piecewise definition of the boundary shearlets, smoothness (of the Fourier representations) is not guaranteed for all $\alpha \in (-\infty, 2)$. It turns out that the exponent of $2^{(2-\alpha)j}$ needs necessarily to be

integer-valued when analyzing the first partial derivatives of (3.6)—this is explicitly done in the proof of Theorem 3.4 (cf. equation (7.2) and (7.3)). To satisfy this for every $j \in \mathbb{Z}$, we would have to restrict the set of admissible α to the set of integers; but this would completely burst the major goal of arbitrary scaling behavior.

The key idea to overcome this problem is to relax the condition of a globally fixed α and, instead of that, to introduce a *separate* scaling parameter α_j on *each* scale. Thus, the universal shearlet system will be associated with a whole sequence $(\alpha_j)_{j \in \mathbb{N}_0}$. In this looser setting, the set of admissible scaling parameters can be substantially enlarged because we have $(2 - \alpha_j)j \in \mathbb{Z}$ whenever α_j is a multiple of $1/j$. This approach of discretization naturally gives rise to the following notion:

Definition 3.2. A sequence $(\alpha_j)_{j \in \mathbb{N}_0} \subset \mathbb{R}$ is called a scaling sequence if

$$\alpha_j \in \mathbf{A}_j := \left\{ \frac{m}{j} \mid m \in \mathbb{Z}, m \leq 2j - 1 \right\} = \left\{ \dots, -\frac{2}{j}, -\frac{1}{j}, 0, \frac{1}{j}, \dots, 2 - \frac{1}{j} \right\}$$

for $j \geq 1$ and $\alpha_0 = 0$.

Now, we are able to give the main definition of this section:

Definition 3.3. Let $(\alpha_j)_{j \in \mathbb{N}_0}$ be a scaling sequence. Then we define the associated universal-scaling shearlet system, or shorter, universal shearlet system, by

$$\text{SH}(\phi, v, (\alpha_j)_j) := \text{SH}_{\text{Low}}(\phi) \cup \text{SH}_{\text{Int}}(\phi, v, (\alpha_j)_j) \cup \text{SH}_{\text{Bound}}(\phi, v, (\alpha_j)_j),$$

where

$$\begin{aligned} \text{SH}_{\text{Low}}(\phi) &:= \{ \psi_{-1,k} \mid k \in \mathbb{Z}^2 \}, \\ \text{SH}_{\text{Int}}(\phi, v, (\alpha_j)_j) &:= \left\{ \psi_{j,l,k}^{\alpha_j, (\iota)} \mid j \geq 0, |l| < 2^{(2-\alpha_j)j}, k \in \mathbb{Z}^2, \iota \in \{\text{h}, \text{v}\} \right\}, \\ \text{SH}_{\text{Bound}}(\phi, v, (\alpha_j)_j) &:= \left\{ \psi_{j,l,k}^{\alpha_j} \mid j \geq 0, |l| = \pm 2^{(2-\alpha_j)j}, k \in \mathbb{Z}^2 \right\}. \end{aligned}$$

Notice that the elements of a scaling sequence α_j can be chosen independently for each scaling level and, in particular, the sequence $(\alpha_j)_j$ does not need to converge. Interestingly, this additional degree of freedom allows for a unification of the formal shearlet definition and its numerical implementation. In practice, one usually determines the total number of shearings (on each scale) first, rather than selecting a fixed α . The scaling parameter is then implicitly given and can be computed by an easy relation (cf. [34]). But since the amount of shearings is necessarily integer-valued, the resulting α_j might not coincide for all j . According to this phenomenon, the formal concept of a scaling sequence matches precisely to what is done in practice.

The following theorem collects the most important properties of universal shearlets. In particular, all systems provide excellent localization in both Fourier and spatial domain. We postpone the proof to Subsection 7.1.

Theorem 3.4. Let $(\alpha_j)_j$ be a scaling sequence and $\text{SH}(\phi, v, (\alpha_j)_j)$ be an associated universal shearlet system. Then $\text{SH}(\phi, v, (\alpha_j)_j)$ constitutes a Parseval frame for $L^2(\mathbb{R}^2)$ consisting of band-limited Schwartz functions. Moreover, the interior and boundary shearlets have infinitely many vanishing moments.

To illustrate the flexibility of our new construction, we finally present two natural choices of $(\alpha_j)_j$ inducing well-known representation systems.

3.2. α -Shearlets. Let $\alpha \in (-\infty, 2)$, and recall the definition of \mathbf{A}_j from Definition 3.2. We then choose $(\alpha_j)_j$ such that it provides the best possible approximation of α , i.e.,

$$\alpha_j := \underset{\tilde{\alpha}_j \in \mathbf{A}_j}{\operatorname{argmin}} |\tilde{\alpha}_j - \alpha|, \quad j \geq 1.$$

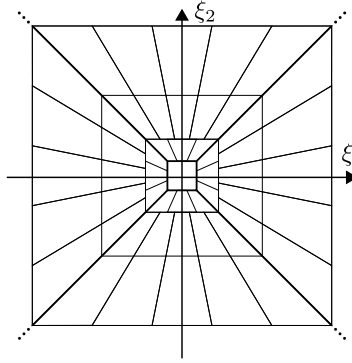


Figure 5. Frequency tiling obtained by $\text{SH}(\phi, v, (\alpha_j)_j)$ in the wavelet case.

One easily verifies $2^{\alpha_j j} \in \Theta(2^{\alpha j})$ as $j \rightarrow \infty$. Hence, the corresponding system $\text{SH}(\phi, v, (\alpha_j)_j)$ has asymptotically the same scaling behavior as α -shearlets, see [32, 24] and more generally [16]. If $\alpha = 1$, we exactly obtain the cone-adapted shearlet frame from [19]. On the other hand, the choice of a small (possibly negative) α produces more anisotropic elements; letting $\alpha_j \rightarrow -\infty$ this could even give rise to an approximation of the “ridgelet case”.

3.3. Wavelets. While the original scaling parameter α had to be strictly less than 2, it is now possible to reach this limit case. Concretely, we choose $(\alpha_j)_j$ as large as possible, namely $\alpha_j := 2 - \frac{1}{j}$ for $j \geq 1$. Then we have $2^{\alpha_j j} \in \Theta(2^{2j})$ implying that the elements of $\text{SH}(\phi, v, (\alpha_j)_j)$ scale in an *isotropic* fashion. Furthermore, the number of shears is constant (equal to 10) on each scaling level (see Figure 5). Therefore, $\text{SH}(\phi, v, (\alpha_j)_j)$ can be viewed as a special kind of a wavelet frame. This particularly allows us to treat wavelets and shearlets in a uniform fashion and to compare their respective (inpainting) analysis directly.

4. Image Model on $L^2(\mathbb{R}^2)$.

4.1. Line Distribution and Mask. As already motivated in the first section, we intend to analyze the model of a corrupted line segment. From a theoretical point of view, such a line object shall not have any spatial widening (i.e., no thickness) and its support shall be parametrizable by a curve in \mathbb{R}^2 . It would not be clear how to distinguish surfaces and curvilinear structures otherwise. However, one-dimensional objects have measure zero in \mathbb{R}^2 and cannot be directly represented by L^2 -functions. It is therefore convenient to work in the space of tempered distributions $\mathcal{S}'(\mathbb{R}^2)$, which is the set of all linear and continuous functionals on $\mathcal{S}(\mathbb{R}^2)$. To adapt the example of seismic image data, we follow [26] again, and consider a compactly supported edge singularity along the x_1 -axis.

Let $0 \neq w \in C^\infty(\mathbb{R})$ be a *weighting function* satisfying $0 \leq w \leq 1$ and $\text{supp } w \subset [-\rho, \rho]$ for some fixed $\rho > 0$. The *weighted edge singularity* $w\mathfrak{L} \in \mathcal{S}'(\mathbb{R}^2)$ is then given by the product of w and a line distribution along the x_1 -axis,

$$\langle w\mathfrak{L}, \varphi \rangle := \langle \mathfrak{L}, w\varphi \rangle := \int_{-\rho}^{\rho} w(x_1)\varphi(x_1, 0) dx_1, \quad \varphi \in \mathcal{S}(\mathbb{R}^2).$$

From this definition, it immediately follows that $w\mathfrak{L}$ has compact support on $[-\rho, \rho] \times \{0\}$. The “graph” of $w\mathfrak{L}$ is sketched in Figure 6(b).

We will frequently work with Fourier representations in the forthcoming section. Thus, we already compute the Fourier transform of $w\mathfrak{L}$ at this point:

$$\langle \widehat{w\mathfrak{L}}, \varphi \rangle = \langle w\mathfrak{L}, \widehat{\varphi} \rangle = \int_{-\rho}^{\rho} w(x_1) \left(\int_{\mathbb{R}^2} \varphi(\xi) e^{-2\pi i \xi_1 x_1} d\xi \right) dx_1 = \int_{\mathbb{R}^2} \widehat{w}(\xi_1) \varphi(\xi) d\xi, \quad \varphi \in \mathcal{S}(\mathbb{R}^2).$$

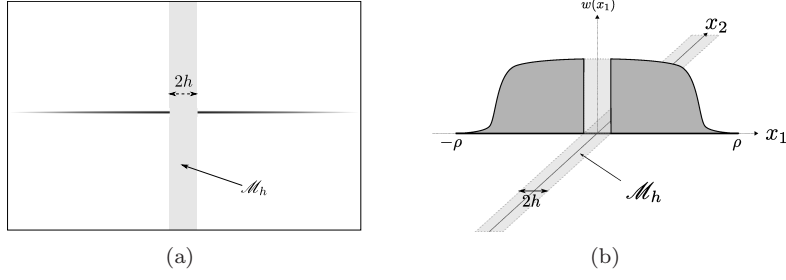


Figure 6. (a) Sketch of the corrupted modeling image. Note that the line singularity has no “thickness” but some (gray-scale) intensity. (b) “Graph” of the corrupted line distribution $P_K w_{\mathcal{L}}$, which is compactly supported on the x_1 -axis.

This particularly shows that $\widehat{w_{\mathcal{L}}}$ is a regular distribution with $\widehat{w_{\mathcal{L}}}(\xi) = \widehat{w}(\xi_1)$, and $\widehat{w_{\mathcal{L}}} \in C_b^\infty(\mathbb{R}^2)$. However, it has no decay in the ξ_2 -direction and $w_{\mathcal{L}}$ “carries” therefore arbitrarily high frequencies.

To specify the destroyed region of our image model, we let us be inspired by the motivating example of seismic data once more: We cut off a small vertical strip around the x_2 -axis which is given by the symmetric *mask*

$$\mathcal{M}_h := \{(x_1, x_2) \in \mathbb{R}^2 \mid |x_1| \leq h\}$$

of diameter $2h > 0$ (see Figure 6 for visualization). The gap space is then set to $\mathcal{H}_M = L^2(\mathcal{M}_h)$. Note that, for this choice, the orthogonal projections can be written in terms of characteristic functions:

$$P_M f = \chi_{\mathcal{M}_h} f, \quad P_K f = \chi_{\mathcal{M}_h^c} f, \quad f \in L^2(\mathbb{R}^2).$$

4.2. Filter Decomposition. Due to the fact that the energy of $w_{\mathcal{L}}$ is infinite, we need to break this model down to $L^2(\mathbb{R}^2)$ in a reasonable manner. For the sake of convenience, this decomposition should be also compatible with the universal shearlet systems, which will serve as underlying Parseval frames for ℓ^1 -inpainting (cf. Algorithm 2.1). According to that, we introduce a sequence of *frequency filters* $F_j \in \mathcal{S}(\mathbb{R}^2)$, $j \geq 0$, which are defined by the inverse Fourier transform of the corona functions from the shearlet construction (cf. Subsection 3.1), i.e.,

$$\widehat{F}_j(\xi) := W_j(\xi) = W(2^{-2j}\xi), \quad \xi \in \mathbb{R}^2, \quad j \geq 0.$$

So we obtain a whole sequence of (scale-dependent) *benchmark models* $(w_{\mathcal{L}_j})_{j \geq 0}$ given by

$$w_{\mathcal{L}_j}(x) := w_{\mathcal{L}} * F_j(x) = \langle w_{\mathcal{L}}, F_j(x - [\cdot]) \rangle, \quad x \in \mathbb{R}^2.$$

Note that—although the definition of universal shearlets is involved here—there is no dependence on the scaling sequence $(\alpha_j)_j$. Hence, the models are comparable with different choices of α_j . Using relation (3.2), the original image model can be recovered from $(w_{\mathcal{L}_j})_{j \geq 0}$ by

$$w_{\mathcal{L}} = \Phi * (\Phi * w_{\mathcal{L}}) + \sum_{j \geq 0} F_j * w_{\mathcal{L}_j},$$

where the first term of the sum corresponds to the low-frequency part of $w_{\mathcal{L}}$.

A computation of the Fourier transforms,

$$(4.1) \quad \widehat{w_{\mathcal{L}_j}}(\xi) = \widehat{w_{\mathcal{L}}}(\xi) \cdot \widehat{F}_j(\xi) = \widehat{w}(\xi_1) \cdot W_j(\xi), \quad \xi \in \mathbb{R}^2,$$

shows that the $w_{\mathcal{L}_j}$ are band-limited Schwartz functions and have therefore finite energy, that is $w_{\mathcal{L}_j} \in L^2(\mathbb{R}^2)$. Furthermore, (4.1) allows an important conclusion on the filter decomposition: Due

to the multiplication with W_j , the frequency support of $w\mathfrak{L}$ gets truncated to the corona \mathcal{K}_j (see Figure 3(b)). Recalling the geometry of \mathcal{K}_j , this implies that $w\mathfrak{L}_j$ covers increasingly higher frequency parts of $w\mathfrak{L}$ (as j grows). But at the same time, the energy of $w\mathfrak{L}_j$ strongly concentrates around the horizontal axis:

$$|w\mathfrak{L}_j(x_1, x_2)| \leq C_N \langle |2^{2j}x_2| \rangle^{-N}, \quad \text{for every } N \in \mathbb{N},$$

see also computation (7.7) for more details. Roughly spoken, we approximate a sharp edge singularity when sending j to infinity. This also coincides with our intuition that paintings with high frequency content are often governed by curvilinear structures. Hence, the following inpainting analysis will mainly focus on the asymptotic case $j \rightarrow \infty$.

Contemplating, at the first sight, the above frequency decomposition appears to be a rather artificial and technical approach to represent an edge model in $L^2(\mathbb{R}^2)$, but this is actually a crucial step toward a theoretical result on inpainting: If we would analyze only one single image model, it would be essentially unclear how to evaluate the quality of the recovery. On the contrary, an asymptotic analysis of a whole sequence of models—more precisely, we consider the inpainting behavior for $(w\mathfrak{L}_j)_{j \geq 0}$ as $j \rightarrow \infty$ —allows us to make a statement about convergence, independently from any concrete error quantities.

5. A Convergence Theorem for Image Inpainting. In the previous two sections, we have specified the ingredients of the image inpainting problem formulated by Algorithm 2.1 in case of $\mathcal{H} = L^2(\mathbb{R}^2)$. Now, we will apply the abstract error estimate of Theorem 2.5 to show the following convergence theorem, which is the main result of this paper and formally proves the success of image inpainting via universal shearlets. Note that in particular the inpainting results from [26] for shearlets and wavelets included as special cases.

Theorem 5.1. *Let $(\alpha_j)_j$ be a scaling sequence and let $\Psi := \text{SH}(\phi, v, (\alpha_j)_j)$ be an associated universal shearlet system. We choose a gap width $h_j > 0$ for every $j \geq 0$ and consider the models $P_K w\mathfrak{L}_j = \chi_{\mathcal{M}_{h_j}^c} w\mathfrak{L}_j$ of the corrupted image. The respective recoveries provided by Algorithm 2.1, using Ψ , are denoted by $w\mathfrak{L}_j^*$. Furthermore, we assume that the following two conditions are satisfied:*

(INP1) *There exist $\delta > 0$ and $j_0 \in \mathbb{N}$ such that $\alpha_j > \delta$ for all $j \geq j_0$, i.e., $\liminf_{j \rightarrow \infty} \alpha_j > 0$.*

(INP2) *For a fixed $\varepsilon > 0$, the sequence of gap widths satisfies $(h_j)_j \in o(2^{-(\alpha_j + \varepsilon)j})$.*

Then, the (relative) recovery error decays rapidly in the ℓ^1 -analysis norm, i.e.,

$$(5.1) \quad \left(\frac{\|w\mathfrak{L}_j^* - w\mathfrak{L}_j\|_{1, \Psi}}{\|w\mathfrak{L}_j\|_{1, \Psi}} \right)_j \in o(2^{-Nj}), \quad \text{as } j \rightarrow \infty$$

for every $N \in \mathbb{N}_0$. (Note that the asymptotic constants in (5.1) may depend on N .)

Theorem 5.1 will be a direct consequence of the Propositions 5.2 and 5.6—their proof will be the major challenge of this section—applied to Theorem 2.5. Note that the normalization in (5.1) has only a minor influence on the actually statement, as Corollary 5.5 will show.

Before giving a complete proof, we wish to point out an important perspective of how to read this result. As already discussed at the end of the previous section, Theorem 5.1 provides an asymptotical statement about the inpainting performance for $P_K w\mathfrak{L}_j$ when the scaling level becomes sufficiently high. In fact, each model is reconstructed separately, and also the corresponding gap width h_j depends on the scale j ; this presumption is essential because the Fourier support of $w\mathfrak{L}_j$ is scale-dependent.

Under the assumption of “adequate control of gap size” in terms of (INP2), we indeed obtain rapid error decay. As supposed in [26], we call the underlying recovery process *asymptotically perfect* whenever (5.1) holds. In the sense of a theoretical analysis, an asymptotically perfect recovery of a model sequence should be rather seen as a *benchmark criterion* for the success of inpainting. The actual question that one should ask is therefore:

Under which (sufficient and/or necessary) conditions do we obtain (5.1)?

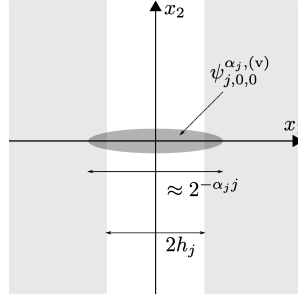


Figure 7. Visualization of condition (INP2). The gap width h_j shall not exceed the essential length of the horizontally oriented shearlet elements (on scale j).

Form this point of view, the most interesting aspects of Theorem 5.1 are the (sufficient) conditions (INP1) and (INP2). The first one excludes the artificial case of negative scaling: When the elements α_j of the scaling sequence are all non-positive, the essential length of the shearlet element becomes constant or even blows up to infinity. In fact, this restriction is quite natural because otherwise, we could fill in gaps of arbitrary size for a finitely supported model.

The second condition forms the “heart” of Theorem 5.1. Intuitively spoken, (INP2) restricts the allowed gap widths so that they can be covered by one single shearlet element (see also Figure 7). In this way, Theorem 5.1 relates the degree of anisotropic scaling to the admissible gap sizes: When α_j becomes smaller, the asymptotical condition $(h_j)_j \in o(2^{-(\alpha_j + \varepsilon)j})$ is satisfied for larger h_j . For the two sample systems from Subsection 3.2 and 3.3, we particularly conclude that, in terms of image inpainting, (α) -shearlet frames are superior to isotropic wavelet systems; more general, increasing the degree of anisotropy (by making α smaller) “continuously” improves the above convergence result.

To summarize, the essence of our formal inpainting analysis is a *comparison* of several representation frames that vary in their degree of directional scaling. However, the theorem gives only sufficient conditions for asymptotically perfect inpainting; a rigorous proof that, for instance, shearlets are superior to wavelets would also require a necessary condition. This (still open) problem is discussed in Subsection 5.3.

5.1. Clustered Sparsity. At first, we shall investigate the Fourier support properties of the modeling functions $w\mathfrak{L}_j$. For the sake of convenience, we introduce some indexing sets for universal shearlet frames:

$$\begin{aligned} \Gamma &:= \{(j, l, k; \alpha_j, \iota) \mid j \geq 0, |l| < 2^{(2-\alpha_j)j}, k \in \mathbb{Z}^2, \iota \in \{h, v\}\} \\ &\quad \cup \{(j, l, k; \alpha_j, \emptyset) \mid j \geq 0, |l| = 2^{(2-\alpha_j)j}, k \in \mathbb{Z}^2\}, \\ \Gamma_j &:= \{(j', l, k; \alpha_j, \iota) \in \Gamma \mid j' = j\}, \quad j \geq 0. \end{aligned}$$

Using the notation $\psi_\gamma := \psi_{j,l,k}^{\alpha_j, (\iota)}$ for $\gamma = (j, l, k; \alpha_j, \iota) \in \Gamma$ (here, $\iota = \emptyset$ means that this index is omitted), we can write in short

$$\bigcup_{\gamma \in \Gamma} \psi_\gamma = \text{SH}_{\text{Int}}(\phi, v, (\alpha_j)_j) \cup \text{SH}_{\text{Bound}}(\phi, v, (\alpha_j)_j).$$

By definition, the support of $\widehat{w\mathfrak{L}}_j$ is contained in the frequency corona \mathcal{K}_j . Due to (3.1), this region also overlaps with the neighboring coronas \mathcal{K}_{j-1} and \mathcal{K}_{j+1} , but is disjoint with all the other ones. Hence, for any $j \geq 1$, we easily conclude that

$$(5.2) \quad \begin{aligned} \langle w\mathfrak{L}_j, \psi_{-1,k} \rangle &= 0, \quad k \in \mathbb{Z}^2, \\ \langle w\mathfrak{L}_j, \psi_{j',l,k}^{\alpha_j, (\iota)} \rangle &= 0, \quad (j', l, k; \alpha_j, \iota) \in \Gamma_{j'}, \quad |j' - j| > 1. \end{aligned}$$

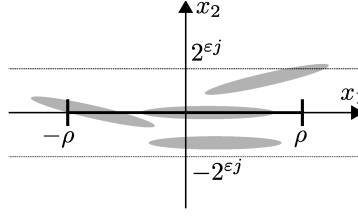


Figure 8. Some shearlet elements associated with the cluster Λ_j .

This shows that all non-zero coefficients are actually contained in the subsequence $(\langle w\mathfrak{L}_j, \psi_\gamma \rangle)_{\gamma \in \Gamma_j^{\pm 1}}$, where we used a comfortable notation from [11]:

$$A_j^{\pm 1} := A_{j-1} \cup A_j \cup A_{j+1}, \quad j \geq 0, \quad (A_j)_{j \in \mathbb{N}_0} \text{ are arbitrary sets and } A_{-1} := \emptyset.$$

At this point, it also should become more clear why we have chosen the functions $F_j = \check{W}_j$ as frequency filters in Section 4.

Now, we fix some $\varepsilon > 0$ (this will be the same as in (INP2)) and define the clusters

$$\Lambda_j := \{(j, l, k; \alpha_j, v) \mid |l| \leq 1, k \in \mathbb{Z}^2, |k_2 - lk_1| \leq 2^{\varepsilon j}\} \subset \Gamma_j, \quad j \geq 0.$$

Figure 8 visualizes the geometric shape of Λ_j : We consider only those shearlet elements which are almost horizontally oriented and concentrated in a tube of width $2^{\varepsilon j}$ around the x_1 -axis (recall that “v” means a vertical direction in frequency which corresponds to a horizontal orientation in the space). This choice is clearly influenced by the geometry of $w\mathfrak{L}$ whose support ($\text{supp } w\mathfrak{L} = [-\rho, \rho] \times \{0\}$) is entirely covered by every Λ_j . But note that, although the clusters are carefully chosen, there may occur very small shearlet coefficients belonging to Λ_j as well; this is due the band-limiting property of $\text{SH}(\phi, v, (\alpha_j)_j)$ which produces strong oscillations in the spatial domain. The proofs of the main Propositions 5.2 and 5.6 will show that the term $2^{\varepsilon j}$ plays a fundamental role in our analysis. In fact, ε controls the trade-off relation between clustered sparsity and cluster coherence that was already mentioned in the course of Theorem 2.5.

Taking the technicality of “neighboring coronas” into account, we finally put

$$\delta_j^{j'-j} := \|\mathbb{1}_{\Gamma_{j'} \setminus \Lambda_{j'}}(T_\Psi w\mathfrak{L}_j)\|_{\ell^1} = \sum_{\gamma \in \Gamma_{j'} \setminus \Lambda_{j'}} |\langle w\mathfrak{L}_j, \psi_\gamma \rangle|, \quad \text{for } j, j' \geq 1, |j' - j| \leq 1,$$

and then sum up (recall that $\Psi = \text{SH}(\phi, v, (\alpha_j)_j)$),

$$\delta_j := \delta_j^{-1} + \delta_j^0 + \delta_j^{+1} \stackrel{(5.2)}{=} \|\mathbb{1}_{(\Lambda_j^{\pm 1})^c}(T_\Psi w\mathfrak{L}_j)\|_{\ell^1} = \sum_{\gamma \in (\Lambda_j^{\pm 1})^c} |\langle w\mathfrak{L}_j, \psi_\gamma \rangle|, \quad \text{for } j \geq 1.$$

By definition, $w\mathfrak{L}_j$ is δ_j -clustered sparse in $\text{SH}(\phi, v, (\alpha_j)_j)$ with respect to the cluster $\Lambda_j^{\pm 1}$. In the following, we will show that $\Lambda_j^{\pm 1}$ indeed captures the sparsity behavior:

Proposition 5.2. *Assuming that $\liminf_{j \rightarrow \infty} \alpha_j > 0$, the sequence $(\delta_j)_{j \in \mathbb{N}}$ decays rapidly, i.e.,*

$$(5.3) \quad \delta_j \in o(2^{-Nj}), \quad j \rightarrow \infty,$$

for every $N \in \mathbb{N}$.

where $\tilde{w}_{N,j} := |w| * \langle |2^{2j}[\cdot]| \rangle^{-N}$.

Now, we are able to give a proof of the main result of this section:

Proof of Proposition 5.2.

(STEP1) We start with an estimation of T_2^0 . For this, we fix $j \geq 0$ and $|l| > 1$ first, and consider some index $(j, l, k; \alpha_j, \nu) \in \Gamma_j$. Then, for $N \geq 2$, we have

$$\begin{aligned} \sum_{\substack{k \in \mathbb{Z}^2, \\ t_1^{(\nu)} \neq 0, t_2^{(\nu)} \neq 0}} \left| t_1^{(\nu)} \right|^{-N} \left| t_2^{(\nu)} \right|^{-N} &= \sum_{\substack{k \in \mathbb{Z}^2, \\ k_1 \neq 0, k_2 \neq lk_1}} |2^{-\alpha_j j} k_1|^{-N} |2^{-2j} (k_2 - lk_1)|^{-N} \\ &= 2^{(\alpha_j N + 2N)j} \sum_{\substack{k \in \mathbb{Z}^2, \\ k_1 \neq 0, k_2 \neq 0}} |k_1|^{-N} |k_2|^{-N} \leq C_N 2^{N(\alpha_j + 2)j} \int_{|x_1| \geq 1, |x_2| \geq 1} |x_1|^{-N} |x_2|^{-N} dx \leq C_N 2^{N(\alpha_j + 2)j}. \end{aligned}$$

Similar estimates for the cases $t_1^{(\nu)} = 0$ and/or $t_2^{(\nu)} = 0$ yield altogether

$$\sum_{k \in \mathbb{Z}^2} \left| t_1^{(\nu)} \right|^{-N^{(\nu)}} \left| t_2^{(\nu)} \right|^{-N^{(\nu)}} \leq C_N 2^{N^{(\nu)}(\alpha_j + 2)j}.$$

Finally, by Lemma 5.3(i),

$$\begin{aligned} T_2^0 &= \sum_{\substack{k \in \mathbb{Z}^2, \\ 1 < |l| \leq 2^{(2-\alpha_j)j}}} \left| \langle w \mathfrak{L}_j, \psi_{j,l,k}^{\alpha_j, (\nu)} \rangle \right| \leq C_{N,M} \sum_{\substack{k \in \mathbb{Z}^2, \\ 1 < |l| \leq 2^{(2-\alpha_j)j}}} \left| t_1^{(\nu)} \right|^{-N^{(\nu)}} \left| t_2^{(\nu)} \right|^{-N^{(\nu)}} 2^{(2-\alpha_j)j/2} 2^{-N^{(\nu)}\alpha_j j} \langle |2^{\alpha_j j}| \rangle^{-M} \\ &\leq C_{N,M} 2^{3(2-\alpha_j)j/2} 2^{-N^{(\nu)}\alpha_j j} \langle |2^{\alpha_j j}| \rangle^{-M} \sum_{k \in \mathbb{Z}^2} \left| t_1^{(\nu)} \right|^{-N^{(\nu)}} \left| t_2^{(\nu)} \right|^{-N^{(\nu)}} \\ &\leq C_{N,M} 2^{3(2-\alpha_j)j/2} 2^{N^{(\nu)}j} \langle |2^{\alpha_j j}| \rangle^{-M}. \end{aligned}$$

Using the assumption $\liminf_{j \rightarrow \infty} \alpha_j > 0$ and choosing M sufficiently large, we obtain the desired decay rates of (5.3). The estimates for T_3^0 and T_4^0 are done analogously and are therefore omitted at this point.

(STEP2) To show the respective decay rate for T_1^0 , we are going to apply Lemma 5.4 for $N \geq 2$ to obtain

$$\begin{aligned} T_1^0 &= \sum_{\substack{k \in \mathbb{Z}^2, |l| \leq 1 \\ |k_2 - lk_1| > 2^{\varepsilon_j}}} \left| \langle w \mathfrak{L}_j, \psi_{j,l,k}^{\alpha_j, (\nu)} \rangle \right| \\ &\leq C_N 2^{(5-\alpha_j)j/2} \sum_{\substack{k \in \mathbb{Z}^2, |l| \leq 1 \\ |k_2 - lk_1| > 2^{\varepsilon_j}}} \int_{\mathbb{R}} \tilde{w}_{N,j}(2^{-\alpha_j j}(x_1 + k_1)) \langle |x_1| \rangle^{-N} \langle |lx_1 + lk_1 - k_2| \rangle^{-N} dx_1 \\ &= C_N 2^{(5-\alpha_j)j/2} \sum_{\substack{k \in \mathbb{Z}^2, |l| \leq 1 \\ |k_2| > 2^{\varepsilon_j}}} \int_{\mathbb{R}} \tilde{w}_{N,j}(2^{-\alpha_j j}(x_1 + k_1)) \langle |x_1| \rangle^{-N} \langle |lx_1 - k_2| \rangle^{-N} dx_1 \\ &= C_N 2^{(5-\alpha_j)j/2} \sum_{\substack{|l| \leq 1 \\ k_2 \in \mathbb{Z}, |k_2| > 2^{\varepsilon_j}}} \int_{\mathbb{R}} \left(\sum_{k_1 \in \mathbb{Z}} \tilde{w}_{N,j}(2^{-\alpha_j j}(x_1 + k_1)) \right) \langle |x_1| \rangle^{-N} \langle |lx_1 - k_2| \rangle^{-N} dx_1. \end{aligned}$$

The first factor of the integrand can be estimated by

$$\begin{aligned} \sum_{k_1 \in \mathbb{Z}} \tilde{w}_{N,j}(2^{-\alpha_j j}(x_1 + k_1)) &= \sum_{k_1 \in \mathbb{Z}} \int_{\mathbb{R}} |w(y_1)| \langle |2^{2j}(y_1 - (2^{-\alpha_j j}(x_1 + k_1)))| \rangle^{-N} dy_1 \\ &= \sum_{k_1 \in \mathbb{Z}} \int_{\mathbb{R}} |w(y_1)| \langle |2^{(2-\alpha_j)j}(k_1 + x_1 - 2^{\alpha_j j} y_1)| \rangle^{-N} dy_1 \\ &\leq \int_{\mathbb{R}} |w(y_1)| \underbrace{\left(\sum_{k_1 \in \mathbb{Z}} \langle |k_1 + x_1 - 2^{\alpha_j j} y_1| \rangle^{-N} \right)}_{\leq C_N} dy_1 \leq C_N. \end{aligned}$$

For the last inequality recall that $2^{(2-\alpha_j)j} \geq 1$. Finally, by a similar argumentation as in (7.9), we obtain

$$\begin{aligned} T_1^0 &\leq C_N 2^{(5-\alpha_j)j/2} \sum_{\substack{|l| \leq 1 \\ k_2 \in \mathbb{Z}, |k_2| > 2^{\varepsilon j}}} \int_{\mathbb{R}} \langle |x_1| \rangle^{-N} \langle |lx_1 - k_2| \rangle^{-N} dx_1 \\ &\leq C_N 2^{(5-\alpha_j)j/2} \sum_{k_2 \in \mathbb{Z}, |k_2| > 2^{\varepsilon j}} \langle |k_2| \rangle^{-N} \leq C_N 2^{(5-\alpha_j)j/2} \int_{|x_2| > 2^{\varepsilon j}} \langle |x_2| \rangle^{-N} dx_2 \\ &\leq C_N 2^{(5-\alpha_j)j/2} 2^{-(N-1)\varepsilon j}. \end{aligned}$$

This proves the claim since we can choose N arbitrarily large. \blacksquare

The necessity of the assumption $\liminf_j \alpha_j > 0$ in Proposition 5.2 becomes clear within (STEP1) of the proof; the factor $\langle |2^{\alpha_j j}| \rangle^{-M}$ would not produce the required decay otherwise. Moreover, the calculations of (STEP2) emphasize the importance of the condition $|k_2 - lk_1| \leq 2^{\varepsilon j}$ in the definition of Λ_j : To obtain a sufficient coefficient decay, we require a certain distance between the model singularity and the essential shearlet support. However, one should be aware of the fact that—although rapid decay is achieved for arbitrary small $\varepsilon > 0$ —the constants in (5.3) will blow up as ε approaches 0.

The following corollary, clarifies the behavior of the normalization factor in (5.1):

Corollary 5.5. *There exists an $N \geq 0$ such that $\|w\mathfrak{L}_j\|_{1,\Psi} \in o(2^{Nj})$ as $j \rightarrow \infty$.*

Proof. Due to Proposition 5.2 is sufficient to investigate what happens “inside” of $\Lambda_j^{\pm 1}$. For this, one can make use of Lemma 5.4 again, and proceed similar to (STEP2) in the proof of Proposition 5.2. \blacksquare

5.2. Cluster Coherence. We now analyze the cluster coherence. At this point, the condition on the gap size (INP2), which has not occurred yet, comes into play. Conversely, the image models $w\mathfrak{L}_j$ are only implicitly involved, namely by the geometric shape of the clusters $\Lambda_j^{\pm 1}$.

Proposition 5.6. *Assuming that $(h_j)_j \in o(2^{-(\alpha_j + \varepsilon)j})$, we have (recall that $\Psi = \text{SH}(\phi, v, (\alpha_j)_j)$)*

$$\mu_c(\Lambda_j^{\pm 1}, \chi_{\mathcal{M}_{h_j}} \Psi) \rightarrow 0, \quad j \rightarrow \infty.$$

Proof. We first separate the neighboring scales,

$$\mu_c(\Lambda_j^{\pm 1}, \chi_{\mathcal{M}_{h_j}} \Psi) \leq \mu_c(\Lambda_{j-1}, \chi_{\mathcal{M}_{h_j}} \Psi) + \mu_c(\Lambda_j, \chi_{\mathcal{M}_{h_j}} \Psi) + \mu_c(\Lambda_{j+1}, \chi_{\mathcal{M}_{h_j}} \Psi).$$

In the following, we will only consider the main-scale-term (the respective estimates of neighboring-scale-terms will then follow from slight modifications):

$$\begin{aligned} \mu_c(\Lambda_j, \chi_{\mathcal{M}_{h_j}} \Psi) &= \max_{\gamma_2 \in \Gamma} \sum_{\gamma_1 \in \Lambda_j} |\langle \chi_{\mathcal{M}_{h_j}} \psi_{\gamma_1}, \psi_{\gamma_2} \rangle| \\ &= \underbrace{\max_{\substack{\gamma_2 \in \Gamma, \\ i=v}} \sum_{\gamma_1 \in \Lambda_j} |\langle \chi_{\mathcal{M}_{h_j}} \psi_{\gamma_1}, \psi_{\gamma_2} \rangle|}_{=: T_v} + \underbrace{\max_{\substack{\gamma_2 \in \Gamma, \\ i=h}} \sum_{\gamma_1 \in \Lambda_j} |\langle \chi_{\mathcal{M}_{h_j}} \psi_{\gamma_1}, \psi_{\gamma_2} \rangle|}_{=: T_h} + \underbrace{\max_{\substack{\gamma_2 \in \Gamma, \\ i=\emptyset}} \sum_{\gamma_1 \in \Lambda_j} |\langle \chi_{\mathcal{M}_{h_j}} \psi_{\gamma_1}, \psi_{\gamma_2} \rangle|}_{=: T_\emptyset}. \end{aligned}$$

To derive a better intuition for the following computations, it might be helpful to consider the visualizations of Figure 9.

(CASE1) We start by estimating T_v . For this, let $\gamma_1 = (j, l, k; \alpha_j, v) \in \Lambda_j$ and $\gamma_2 = (j, l', k'; \alpha_j, v) \in \Gamma_j$. Using the rapid decay properties of the universal shearlets, we obtain ($N \geq 2$)

$$\begin{aligned} & \left| \langle \chi_{\mathcal{M}_{h_j}} \psi_{j,l,k}^{\alpha_j, (v)}, \psi_{j,l',k'}^{\alpha_j, (v)} \rangle \right| \leq \int_{-h_j}^{h_j} \int_{\mathbb{R}} \left| \psi_{j,l,k}^{\alpha_j, (v)}(x) \right| \left| \psi_{j,l',k'}^{\alpha_j, (v)}(x) \right| dx_2 dx_1 \\ & \leq C_N \int_{-h_j}^{h_j} \int_{\mathbb{R}} 2^{(2+\alpha_j)j} \langle |S_{(v)}^l A_{\alpha_j, (v)}^j x - k| \rangle^{-N} \langle |S_{(v)}^{l'} A_{\alpha_j, (v)}^j x - k'| \rangle^{-N} dx_2 dx_1 \\ & \leq C_N \int_{-h_j}^{h_j} \int_{\mathbb{R}} 2^{(2+\alpha_j)j} \langle |2^{\alpha_j j} x_1 - k_1| \rangle^{-N} \langle |l 2^{\alpha_j j} x_1 + 2^{2j} x_2 - k_2| \rangle^{-N} \\ & \quad \times \langle |2^{\alpha_j j} x_1 - k'_1| \rangle^{-N} \langle |l' 2^{\alpha_j j} x_1 + 2^{2j} x_2 - k'_2| \rangle^{-N} dx_2 dx_1. \\ & = C_N \int_{-2^{\alpha_j j} h_j}^{2^{\alpha_j j} h_j} \int_{\mathbb{R}} \langle |x_1 - k_1| \rangle^{-N} \langle |l x_1 + x_2 - k_2| \rangle^{-N} \underbrace{\langle |x_1 - k'_1| \rangle^{-N} \langle |l' x_1 + x_2 - k'_2| \rangle^{-N}}_{\leq 1} dx_2 dx_1. \end{aligned}$$

Without loss of generality, we may assume that the maximum in T_v is attained for some $\gamma_2 \in \Gamma_j$. This leads to

$$\begin{aligned} T_v &= \max_{\substack{\gamma_2 \in \Gamma, \\ i=v}} \sum_{\gamma_1 \in \Lambda_j} |\langle \chi_{\mathcal{M}_{h_j}} \psi_{\gamma_1}, \psi_{\gamma_2} \rangle| \leq \max_{(j,l',k'; \alpha_j, v) \in \Gamma_j} \sum_{|l| \leq 1} \sum_{k \in \mathbb{Z}^2} \left| \langle \chi_{\mathcal{M}_{h_j}} \psi_{j,l,k}^{\alpha_j, (v)}, \psi_{j,l',k'}^{\alpha_j, (v)} \rangle \right| \\ &\leq C_N \max_{(j,l',k'; \alpha_j, v) \in \Gamma_j} \sum_{\substack{|l| \leq 1 \\ k \in \mathbb{Z}^2}} \int_{-2^{\alpha_j j} h_j}^{2^{\alpha_j j} h_j} \int_{\mathbb{R}} \langle |x_1 - k_1| \rangle^{-N} \langle |l x_1 + x_2 - k_2| \rangle^{-N} \langle |l' x_1 + x_2 - k'_2| \rangle^{-N} dx_2 dx_1 \\ &\leq C_N \max_{(j,l',k'; \alpha_j, v) \in \Gamma_j} \sum_{|l| \leq 1} \int_{-2^{\alpha_j j} h_j}^{2^{\alpha_j j} h_j} \int_{\mathbb{R}} \langle |l' x_1 + x_2 - k'_2| \rangle^{-N} \\ & \quad \times \underbrace{\left(\sum_{k \in \mathbb{Z}^2} \langle |x_1 - k_1| \rangle^{-N} \langle |l x_1 + x_2 - k_2| \rangle^{-N} \right)}_{\leq \tilde{C}_N} dx_2 dx_1 \\ &\leq C_N \max_{(j,l',k'; \alpha_j, v) \in \Gamma_j} \underbrace{\int_{-2^{\alpha_j j} h_j}^{2^{\alpha_j j} h_j} \int_{\mathbb{R}} \langle |x_2 + l' x_1 - k'_1| \rangle^{-N} dx_2 dx_1}_{\leq \tilde{C}_N} \leq C_N 2^{\alpha_j j} h_j \rightarrow 0, \quad j \rightarrow \infty. \end{aligned}$$

(CASE2) For the term T_h , let $\gamma_1 = (j, l, k; \alpha_j, v) \in \Lambda_j$ and $\gamma_2 = (j, l', k'; \alpha_j, h) \in \Gamma_j$. For $N \geq 2$ we

then have

$$\begin{aligned}
& \left| \left\langle \chi_{\mathcal{M}_{h_j}} \psi_{j,l,k}^{\alpha_j,(\mathbf{v})}, \psi_{j,l',k'}^{\alpha_j,(\mathbf{h})} \right\rangle \right| \leq \int_{-h_j}^{h_j} \int_{\mathbb{R}} \left| \psi_{j,l,k}^{\alpha_j,(\mathbf{v})}(x) \right| \left| \psi_{j,l',k'}^{\alpha_j,(\mathbf{h})}(x) \right| dx_2 dx_1 \\
& \leq C_N \int_{-h_j}^{h_j} \int_{\mathbb{R}} 2^{(2+\alpha_j)j} \langle |S_{(\mathbf{v})}^l A_{\alpha_j,(\mathbf{v})}^j x - k| \rangle^{-N} \langle |S_{(\mathbf{h})}^{l'} A_{\alpha_j,(\mathbf{h})}^j x - k'| \rangle^{-N} dx_2 dx_1 \\
& \leq C_N 2^{(2+\alpha_j)j} \int_{-h_j}^{h_j} \int_{\mathbb{R}} \langle |2^{\alpha_j j} x_1 - k_1| \rangle^{-N} \langle |l 2^{\alpha_j j} x_1 + 2^{2j} x_2 - k_2| \rangle^{-N} \\
& \quad \times \langle |l' 2^{\alpha_j j} x_2 + 2^{2j} x_1 - k'_1| \rangle^{-N} \langle |2^{\alpha_j j} x_2 - k'_2| \rangle^{-N} dx_2 dx_1 \\
& \leq C_N 2^{\alpha_j j} \int_{-h_j}^{h_j} \int_{\mathbb{R}} \langle |2^{\alpha_j j} x_1 - k_1| \rangle^{-N} \langle |l 2^{\alpha_j j} x_1 + x_2 - k_2| \rangle^{-N} \\
& \quad \times \underbrace{\langle |l' 2^{(\alpha_j-2)j} x_2 + 2^{2j} x_1 - k'_1| \rangle^{-N} \langle |2^{(\alpha_j-2)j} x_2 - k'_2| \rangle^{-N}}_{\leq 1} dx_2 dx_1 \\
& \leq C_N 2^{\alpha_j j} \int_{-h_j}^{h_j} \langle |2^{\alpha_j j} x_1 - k_1| \rangle^{-N} dx_1 \underbrace{\int_{\mathbb{R}} \langle |x_2 + l 2^{\alpha_j j} x_1 - k_2| \rangle^{-N} dx_2}_{\leq \tilde{C}_N}.
\end{aligned}$$

Summarizing, we obtain

$$\begin{aligned}
T_h &= \max_{\substack{\gamma_2 \in \Gamma \\ i=h}} \sum_{\gamma_1 \in \Lambda_j} |\langle \chi_{\mathcal{M}_{h_j}} \psi_{\gamma_1}, \psi_{\gamma_2} \rangle| \leq \max_{(j,l',k';\alpha_j,\mathbf{h}) \in \Gamma_j} \sum_{|l| \leq 1} \sum_{\substack{k \in \mathbb{Z}^2 \\ |k_2 - lk_1| \leq 2^{\varepsilon j}}} \left| \left\langle \chi_{\mathcal{M}_{h_j}} \psi_{j,l,k}^{\alpha_j,(\mathbf{v})}, \psi_{j,l',k'}^{\alpha_j,(\mathbf{h})} \right\rangle \right| \\
&\leq C_N 2^{\alpha_j j} \max_{(j,l',k';\alpha_j,\mathbf{h}) \in \Gamma_j} 2^{\varepsilon j} \sum_{k_1 \in \mathbb{Z}} \int_{-h_j}^{h_j} \langle |2^{\alpha_j j} x_1 - k_1| \rangle^{-N} dx_1 \\
&= C_N 2^{\varepsilon j} 2^{\alpha_j j} \max_{(j,l',k';\alpha_j,\mathbf{h}) \in \Gamma_j} \underbrace{\int_{-h_j}^{h_j} \sum_{k_1 \in \mathbb{Z}} \langle |2^{\alpha_j j} x_1 - k_1| \rangle^{-N} dx_1}_{\leq \tilde{C}_N} \leq C_N 2^{(\alpha_j + \varepsilon)j} h_j \rightarrow 0, \quad j \rightarrow \infty.
\end{aligned}$$

(CASE3) Since boundary shearlets are a combination of the horizontal and vertical elements, it is easy to show that there exists a constant $C > 0$ such that $T_\emptyset \leq C(T_h + T_v)$. Therefore, the claim follows from the first two cases. \blacksquare

Intuitively, the scalar products $|\langle \chi_{\mathcal{M}_{h_j}} \psi_{\gamma_1}, \psi_{\gamma_2} \rangle|$ measure how strong the corresponding frame elements overlap (or interact) inside of \mathcal{M}_{h_j} (cf. Figure 9). The definition of Λ_j forces ψ_{γ_1} to be almost horizontally oriented, whereas ψ_{γ_2} can have arbitrary direction.

Also notice that on the one hand, in (CASE1) of the proof, it was sufficient to assume that h_j decays faster than the essential length of the shearlets elements. On the other hand, the essential support of the vertically oriented shearlets in (CASE2) could be entirely contained in \mathcal{M}_{h_j} (cf. Figure 9(b)). Hence, we shall not allow too many vertical translations in Λ_j . However, the “tube width” has to be at least $2^{\varepsilon j}$ to guarantee sufficient sparsity (see final remark of previous subsection). These observations show the fundamental trade-off relation between clustered sparsity and cluster coherence. We have already described this interaction subsequent to Theorem 2.5, and this can now be made precise for the specific setting of image inpainting. In particular, we have derived an intuition of how the cluster coherence is related to the amount of missing information: In the proof of Proposition 5.6, we observed that $\mu_c(\Lambda_j^{\pm 1}, \chi_{\mathcal{M}_{h_j}} \Psi)$ can be estimated by $2^{(\alpha_j + \varepsilon)j} h_j$ implying that the coherence will decrease whenever the gap width does.

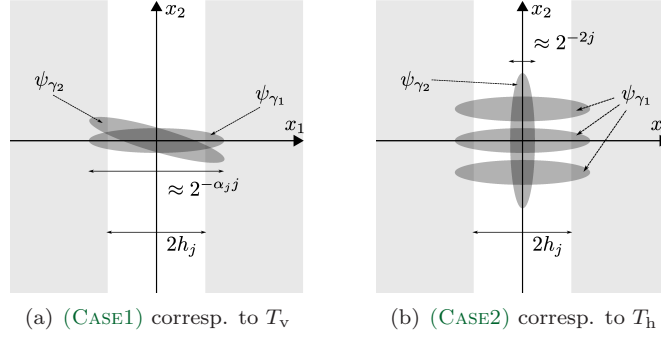


Figure 9. This illustration shows how strong the essential shearlet supports may overlap (at worst) in the proof of Proposition 5.6.

5.3. Optimal Gap Sizes. As already mentioned at the beginning of this section, the condition (INP2) on the gap width in Theorem 5.1 is only a sufficient criterion for asymptotically perfect inpainting. It is therefore an essential (and still open) question, whether this assumption is also necessary, i.e., the choice of $h_j \approx 2^{-\alpha_j j}$ would be *optimal*. In fact, this would, for instance, formally prove that shearlets are superior to wavelets when recovering curvilinear structures and, more general, that directional scaling has an improving effect on the process of inpainting. Such a result however seems to be unreachable to us at the moment, in particular, because of the vastly unknown ℓ^1 -recoveries $w\mathfrak{L}_j^*$. Without the introduction of novel analysis techniques, it appears almost impossible to find an appropriate lower estimate for $\|w\mathfrak{L}_j^* - w\mathfrak{L}_j\|_{1,\Psi}$.

In order to tackle this problem, one could use another inpainting algorithm that returns more explicit recoveries. A very simple (but prominent) example is *one-step thresholding* which was also analyzed in [26]. In this method, the damaged signal $P_K w\mathfrak{L}_j$ is first analyzed with Ψ and small coefficients are truncated afterward. A final synthesis yields the recovery

$$(5.4) \quad w\mathfrak{L}_j^* := T_\Psi^*(\mathbb{1}_{\Lambda_{\beta_j}} T_\Psi(P_K w\mathfrak{L}_j)),$$

where $\Lambda_{\beta_j} := \{(j, l, k; \alpha_j, v) \mid |\langle P_K w\mathfrak{L}_j, \psi_{j,l,k}^{\alpha_j, (v)} \rangle| \geq \beta_j\}$ for certain thresholds β_j . Compared to ℓ^1 -minimization, $w\mathfrak{L}_j^*$ is given by the explicit expression (5.4); but conversely, we have virtually no information about the geometric form of the thresholding-clusters Λ_{β_j} . Due to the band-limiting of $\text{SH}(\phi, v, (\alpha_j)_j)$, the shearlet elements are strongly oscillating in spatial domain, which makes it extremely difficult to determine the large coefficients. It is particularly impossible to choose the thresholds β_j such that Λ_{β_j} has a “tube shape” as in Subsection 5.1. Although the drawbacks of simple thresholding and ℓ^1 -optimization are quite different—in both cases, a certain part of the recovery is only implicitly given—a statement on optimality seems to be currently out of our reach.

However, we strongly believe that at least a weaker version of $(h_j)_j \in o(2^{-(\alpha_j + \varepsilon)j})$ is necessary to achieve asymptotically perfect inpainting. To give an intuitive argument for this conjecture, we may consider some cluster index $(j, l, k; \alpha_j, v) \in \Lambda_j$. Recalling the notation $t^{(v)} = (2^{-\alpha_j j} k_1, 2^{-2j}(k_2 - lk_1))$ from Subsection 5.1, we observe that $\psi_{j,l,k}^{\alpha_j, (v)}$ is “concentrated” inside of \mathcal{M}_{h_j} provided that $|k_1| \leq 2^{\alpha_j j} h_j$. Hence—as a heuristic argument—the “content” of the missing part might be estimated by

$$\begin{aligned} \|P_M w\mathfrak{L}_j\|_{1,\Psi} &= \sum_{\psi \in \Psi} |\langle P_M w\mathfrak{L}_j, \psi \rangle| \gtrsim \sum_{\substack{(j,l,k;\alpha_j,v) \in \Lambda_j \\ l=0, |k_1| \leq 2^{\alpha_j j} h_j}} \underbrace{|\langle w\mathfrak{L}_j, \psi_{j,l,k}^{\alpha_j, (v)} \rangle|}_{\approx 2^{(5-\alpha_j)j/2}} \\ &\approx 2^{\varepsilon j} 2^{\alpha_j j} h_j 2^{(5-\alpha_j)j/2} = 2^{5(1+2\varepsilon)j/2} 2^{\alpha_j j/2} h_j. \end{aligned}$$

This at least indicates that excessively enlarging h_j produces a lack of “information” and will cause recoveries deviating strongly from the original image. Interestingly, a similar estimate for the energy of $P_M w \mathfrak{L}_j$ gives

$$\begin{aligned} \|P_M w \mathfrak{L}_j\|_{L^2}^2 &= \sum_{\psi \in \Psi} |\langle P_M w \mathfrak{L}_j, \psi \rangle|^2 \gtrsim \sum_{\substack{(j,l,k;\alpha_j,v) \in \Lambda_j \\ l=0, |k_1| \leq 2^{\alpha_j j} h_j}} |\langle w \mathfrak{L}_j, \psi_{j,l,k}^{\alpha_j,v} \rangle|^2 \\ &\approx 2^{\varepsilon j} 2^{\alpha_j j} h_j 2^{(5-\alpha_j)j} = 2^{(5+\varepsilon)j} h_j. \end{aligned}$$

Since the term $2^{\alpha_j j}$ cancels out here, our estimate from below does not involve the crucial feature of anisotropy. Therefore, measuring the error in terms of energy seems to be inappropriate to prove an optimality result that compares shearlets and wavelets (see also Subsection 2.4).

6. Extensions and Further Directions. The two main Theorems 3.4 and 5.1 are amenable to several extensions. In the following, we present a brief listing of various directions in which our modeling situation could be generalized to:

- *α -molecules:* We anticipate that the notion of universal shearlets, and in particular the idea of scaling sequences, can be extended to the setting of α -molecules [16], which is a novel framework to unify numerous representation systems—including ridgelets, wavelets, shearlets, and curvelets—as well as their sparse approximation properties.
- *Non-asymptotic analysis:* The statement of Theorem 5.1 is of purely asymptotical nature and we did not investigate the decay constants in (5.1). In particular, the constants will blow up when α_j is closed to 0. Therefore, the non-asymptotical case, i.e., when only a finite number of scales is considered, might be also interesting. This directly leads to the optimization problem of finding the “best possible” α_j in the range of $0 \leq \alpha_j \leq 2$ (the second bound corresponds to the wavelet case). In practice, this challenge could be tackled by a combination of ℓ^1 -minimization and an appropriate *dictionary learning* algorithm such as K-SVD [37], which tries to learn the optimal scaling sequence $(\alpha_j)_j$.
- *Curvilinear singularities:* By using the *tubular neighborhood theorem*, it is possible—though very technical—to extend our results toward a curve model instead of the straight-line distribution $w \mathfrak{L}$. This can be done similarly to the approach in [11], where this technique was applied to the problem of geometric separation.
- *Other mask models:* Other models for the mask such as ball or box shaped structures can be imagined. The general strategy of our work should be a conceivable path for an analysis also in those cases, but certainly, the technical details will highly depend on the chosen model for the mask.
- *Other image models:* In this paper, we focused on the recovery of anisotropic features. However, one might also ask about inpainting of other structures such as texture, or even a superposition of different structural components. This certainly requires a careful adaption of the arguments, but we believe that our general strategy is applicable. In the case of texture, one could use for instance *Gabor systems*; but in order to perform an asymptotic analysis, an additional parameter needs then to be inserted as it was done in [29].
- *Higher dimensions:* The construction of Guo and Labate in [19] also contains higher dimensional versions of shearlets. The respective modifications of universal shearlets are straightforward, hence we decided to not include those. Of special interest is then the three-dimensional case, which corresponds to the problem of *video-inpainting*.
- *Noise:* A very common question is to ask for the stability of a certain algorithm when the input signal is affected by noise. For the problem of (abstract) inpainting, theoretical results are proven in [26] and can be directly applied to our analysis, showing that it is indeed stable toward noise.

7. Proofs.

7.1. Proof of Theorem 3.4.

Band-limiting and vanishing moment property: Observing that $(0, 0) \notin \text{supp } W_j \subset \mathcal{K}_j$, this immediately follows from Definition 3.1.

Smoothness: Due to the band-limiting of $\text{SH}(\phi, v, (\alpha_j)_j)$, it suffices to show that the Fourier transforms of the elements are smooth. This then automatically implies that $\text{SH}(\phi, v, (\alpha_j)_j)$ is contained in $\mathcal{S}(\mathbb{R}^2)$.

The smoothness of $\text{SH}_{\text{Low}}(\phi)$ and $\text{SH}_{\text{Int}}(\phi, v, (\alpha_j)_j)$ are induced by their smooth defining functions ϕ and v . Hence it remains to consider some element $\psi_{j,l,k}^{\alpha_j} \in \text{SH}_{\text{Bound}}(\phi, v, (\alpha_j)_j)$. In the interior of $\mathcal{C}_{(h)}$ and $\mathcal{C}_{(v)}$, the smoothness is again obvious. Thus, we only need to analyze the boundary lines of the cones which are given by $\{|\xi_1| = |\xi_2|\}$.

For this, let us use the shortcut $l_j := 2^{(2-\alpha_j)j}$ for the maximal shearing number (on scale j) and simplify the definition of $\widehat{\psi}_{j,l_j,k}^{\alpha_j}$ (for $j \geq 1$):

$$(7.1) \quad \widehat{\psi}_{j,l_j,k}^{\alpha_j}(\xi) = 2^{-\frac{(2+\alpha_j)j}{2} - \frac{1}{2}} \cdot \begin{cases} W(2^{-2j}\xi)v(l_j(\xi_2/\xi_1 - 1))e^{-\pi i[2^{-2j}\xi_1 k_1 + 2^{-\alpha_j j}(\xi_2 - \xi_1)k_2]}, & \xi \in \mathcal{C}_{(h)}, \\ W(2^{-2j}\xi)v(l_j(\xi_1/\xi_2 - 1))e^{-\pi i[2^{-2j}\xi_1 k_1 + 2^{-\alpha_j j}(\xi_2 - \xi_1)k_2]}, & \xi \in \mathcal{C}_{(v)}. \end{cases}$$

In case of $\xi_1 = \pm \xi_2$, both terms coincide implying the continuity of $\widehat{\psi}_{j,l_j,k}^{\alpha_j}$. By the same argument, the continuity of $\widehat{\psi}_{j,-l_j,k}^{\alpha_j}$ is verified. Next, we compute the partial derivatives of both terms in (7.1) which are given by

$$(7.2) \quad \begin{aligned} & \left. \frac{\partial}{\partial \xi_1} \right|_{\xi_1 = \xi_2} \left[W(2^{-2j}\xi)v\left(l_j\left(\frac{\xi_2}{\xi_1} - 1\right)\right)e^{-\pi i[2^{-2j}\xi_1 k_1 + 2^{-\alpha_j j}(\xi_2 - \xi_1)k_2]} \right] \\ &= 2^{-2j} \frac{\partial W}{\partial \xi_1}(2^{-2j}\xi_1, 2^{-2j}\xi_1)v(0)e^{-2^{-2j}\pi i \xi_1 k_1} - \frac{l_j}{\xi_1} W(2^{-2j}\xi_1, 2^{-2j}\xi_1)v'(0)e^{-2^{-2j}\pi i \xi_1 k_1} \\ & \quad - \pi i(2^{-2j}k_1 - 2^{-\alpha_j j}k_2)W(2^{-2j}\xi_1, 2^{-2j}\xi_1)v(0)e^{-2^{-2j}\pi i \xi_1 k_1} \end{aligned}$$

and

$$(7.3) \quad \begin{aligned} & \left. \frac{\partial}{\partial \xi_1} \right|_{\xi_1 = \xi_2} \left[W(2^{-2j}\xi)v\left(l_j\left(\frac{\xi_1}{\xi_2} - 1\right)\right)e^{-\pi i[2^{-2j}\xi_1 k_1 + 2^{-\alpha_j j}(\xi_2 - \xi_1)k_2]} \right] \\ &= 2^{-2j} \frac{\partial W}{\partial \xi_1}(2^{-2j}\xi_1, 2^{-2j}\xi_1)v(0)e^{-2^{-2j}\pi i \xi_1 k_1} + \frac{l_j}{\xi_1} W(2^{-2j}\xi_1, 2^{-2j}\xi_1)v'(0)e^{-2^{-2j}\pi i \xi_1 k_1} \\ & \quad - \pi i(2^{-2j}k_1 - 2^{-\alpha_j j}k_2)W(2^{-2j}\xi_1, 2^{-2j}\xi_1)v(0)e^{-2^{-2j}\pi i \xi_1 k_1}. \end{aligned}$$

These two expressions indeed coincide, since $v'(0) = 0$. The same can be done in a similar fashion for the partial derivative with respect to ξ_2 , and by obvious modifications, one verifies the smoothness for $\widehat{\psi}_{j,-l_j,k}^{\alpha_j}$ as well as for the case $j = 0$. Finally, the differentiability of higher order is proven by induction and successive use of (3.4).

Parseval frame property: In order to prove this property, the general strategy is to first make use of Parseval's identity to eliminate the translations (modulations in the Fourier representation, respectively), followed by application of the covering properties (3.2) and (3.3).

For this, let $f \in L^2(\mathbb{R}^2)$ be arbitrary. We now consider the respective parts of $\text{SH}(\phi, v, (\alpha_j)_j)$ separately:

(CASE1) Boundary shearlets with $j \geq 1$: By Plancherel's Theorem, we obtain

$$(7.4) \quad \sum_{k \in \mathbb{Z}^2} |\langle f, \psi_{j,l_j,k}^{\alpha_j} \rangle|^2 = \sum_{k \in \mathbb{Z}^2} |\langle \widehat{f}, \widehat{\psi}_{j,l_j,k}^{\alpha_j} \rangle|^2 = \sum_{k \in \mathbb{Z}^2} \left| \int_{\mathbb{R}^2} 2^{-\frac{(2+\alpha_j)j}{2}} \widehat{f}(\xi) W(2^{-2j}\xi) e^{\pi i \xi^\top A_{\alpha_j,(\text{h})}^{-j} S_{(\text{h})}^{-l_j} k} \right. \\ \left. \times \left[\chi_{\mathcal{E}_{(\text{h})}}(\xi) V_{(\text{h})}(\xi^\top A_{\alpha_j,(\text{h})}^{-j} S_{(\text{h})}^{-l_j}) + \chi_{\mathcal{E}_{(\text{v})}}(\xi) V_{(\text{v})}(\xi^\top A_{\alpha_j,(\text{v})}^{-j} S_{(\text{v})}^{-l_j}) \right] d\xi \right|^2.$$

To apply Parseval's identity, we perform a substitution

$$\eta^\top := 2^{-1} \xi^\top A_{\alpha_j,(\text{h})}^{-j} S_{(\text{h})}^{-l_j} \Leftrightarrow \xi = \xi(\eta) = (2^{2j+1}\eta_1, 2^{2j+1}\eta_1 + 2^{\alpha_j j+1}\eta_2).$$

The single expressions of (7.4) then have the form

$$V_{(\text{h})}(\xi^\top A_{\alpha_j,(\text{h})}^{-j} S_{(\text{h})}^{-l_j}) = v\left(\frac{\eta_2}{\eta_1}\right), \\ V_{(\text{v})}(\xi^\top A_{\alpha_j,(\text{v})}^{-j} S_{(\text{v})}^{-l_j}) = v\left(2^{(2-\alpha_j)j} (\xi_1/\xi_2 - 1)\right) = v\left(-\frac{\eta_2}{\eta_1 + 2^{(\alpha_j-2)j}\eta_2}\right), \\ W(2^{-2j}\xi) = W\left(2\eta_1, 2[\eta_1 + 2^{(\alpha_j-2)j}\eta_2]\right).$$

By (3.1), the mapping $(\eta_1, \eta_2) \mapsto W(2\eta_1, 2[\eta_1 + 2^{(\alpha_j-2)j}\eta_2])$ is supported in the "strip" $\{|\eta_1| \leq 1/4\}$. Since $\text{supp } v \subset [-1, 1]$, we can conclude that the mappings

$$(\eta_1, \eta_2) \mapsto U_{(\text{h}),j}(\eta) := W\left(2\eta_1, 2[\eta_1 + 2^{(\alpha_j-2)j}\eta_2]\right) v\left(\frac{\eta_2}{\eta_1}\right), \\ (\eta_1, \eta_2) \mapsto U_{(\text{v}),j}(\eta) := W\left(2\eta_1, 2[\eta_1 + 2^{(\alpha_j-2)j}\eta_2]\right) v\left(-\frac{\eta_2}{\eta_1 + 2^{(\alpha_j-2)j}\eta_2}\right)$$

are both supported in the square $Q^2 := [-\frac{1}{2}, \frac{1}{2}]^2$. Here, we used

$$\left| \frac{-\eta_2}{\eta_1 + 2^{(\alpha_j-2)j}\eta_2} \right| \leq 1 \Rightarrow \left| \frac{\eta_2}{\eta_1} \right| \leq \left| 1 + 2^{(\alpha_j-2)j} \frac{\eta_2}{\eta_1} \right| \leq 1 + 2^{(\alpha_j-2)j} \left| \frac{\eta_2}{\eta_1} \right| \Rightarrow \left| \frac{\eta_2}{\eta_1} \right| \leq \frac{1}{1 - 2^{(\alpha_j-2)j}} \leq 2,$$

where in the last estimate is due to $\alpha_j \leq 2 - 1/j$. With this observation, we can proceed in (7.4) by

$$\sum_{k \in \mathbb{Z}^2} |\langle f, \psi_{j,l_j,k}^{\alpha_j} \rangle|^2 \\ = \sum_{k \in \mathbb{Z}^2} \left| \int_{Q^2} 2^{\frac{(2+\alpha_j)j}{2} + \frac{1}{2}} \widehat{f}(\xi(\eta)) \left[\chi_{\mathcal{E}_{(\text{h})}}(\xi(\eta)) U_{(\text{h}),j}(\eta) + \chi_{\mathcal{E}_{(\text{v})}}(\xi(\eta)) U_{(\text{v}),j}(\eta) \right] e^{2\pi i \eta^\top k} d\eta \right|^2 \\ = \int_{Q^2} 2^{(2+\alpha_j)j+1} |\widehat{f}(\xi(\eta))|^2 \left| \chi_{\mathcal{E}_{(\text{h})}}(\xi(\eta)) U_{(\text{h}),j}(\eta) + \chi_{\mathcal{E}_{(\text{v})}}(\xi(\eta)) U_{(\text{v}),j}(\eta) \right|^2 d\eta \\ = \int_{\mathcal{E}_{(\text{h})}} |\widehat{f}(\xi)|^2 |W(2^{-2j}\xi)|^2 |v(l_j(\xi_2/\xi_1 - 1))|^2 d\xi + \int_{\mathcal{E}_{(\text{v})}} |\widehat{f}(\xi)|^2 |W(2^{-2j}\xi)|^2 |v(l_j(\xi_1/\xi_2 - 1))|^2 d\xi.$$

A very similar computation gives

$$\sum_{k \in \mathbb{Z}^2} |\langle f, \psi_{j,-l_j,k}^{\alpha_j} \rangle|^2 = \int_{\mathcal{E}_{(\text{h})}} |\widehat{f}(\xi)|^2 |W(2^{-2j}\xi)|^2 |v(l_j(\xi_2/\xi_1 + 1))|^2 d\xi \\ + \int_{\mathcal{E}_{(\text{v})}} |\widehat{f}(\xi)|^2 |W(2^{-2j}\xi)|^2 |v(l_j(\xi_1/\xi_2 + 1))|^2 d\xi.$$

(CASE2) Boundary shearlets with $j = 0$: Observing that $\text{supp } W \subset Q^2$, we obtain

$$\begin{aligned} & \sum_{k \in \mathbb{Z}^2} |\langle f, \psi_{0, \pm 1, k}^{\alpha_j} \rangle|^2 = \sum_{k \in \mathbb{Z}^2} |\langle \widehat{f}, \widehat{\psi}_{0, \pm 1, k}^{\alpha_j} \rangle|^2 \\ & = \sum_{k \in \mathbb{Z}^2} \left| \int_{Q^2} \widehat{f}(\xi) W(\xi) \left[\chi_{\mathcal{C}(\text{h})}(\xi) v\left(\frac{\xi_2}{\xi_1} \mp 1\right) + \chi_{\mathcal{C}(\text{v})}(\xi) v\left(\frac{\xi_1}{\xi_2} \mp 1\right) \right] e^{2\pi i \xi^\top k} d\xi \right|^2 \\ & = \int_{\mathcal{C}(\text{h})} |\widehat{f}(\xi)|^2 |W(\xi)|^2 \left| v\left(\frac{\xi_2}{\xi_1} \mp 1\right) \right|^2 d\xi + \int_{\mathcal{C}(\text{v})} |\widehat{f}(\xi)|^2 |W(\xi)|^2 \left| v\left(\frac{\xi_1}{\xi_2} \mp 1\right) \right|^2 d\xi. \end{aligned}$$

(CASE3) Interior shearlets: The substitution $\eta^\top = \xi^\top A_{\alpha_j, (\iota)}^{-j} S_{(\iota)}^{-l}$ for $|l| < l_j$, $\iota \in \{\text{h}, \text{v}\}$ yields

$$\begin{aligned} & \sum_{k \in \mathbb{Z}^2} |\langle f, \psi_{j, l, k}^{\alpha_j, (\text{h})} \rangle|^2 = \int_{\mathbb{R}^2} |\widehat{f}(\xi)|^2 |W(2^{-2j}\xi)|^2 \left| v\left(l_j \frac{\xi_2}{\xi_1} - l\right) \right|^2 d\xi, \\ & \sum_{k \in \mathbb{Z}^2} |\langle f, \psi_{j, l, k}^{\alpha_j, (\text{v})} \rangle|^2 = \int_{\mathbb{R}^2} |\widehat{f}(\xi)|^2 |W(2^{-2j}\xi)|^2 \left| v\left(l_j \frac{\xi_1}{\xi_2} - l\right) \right|^2 d\xi. \end{aligned}$$

(CASE4) Coarse scaling functions: Since $\text{supp } \Phi \subset Q^2$, we have

$$\sum_{k \in \mathbb{Z}^2} |\langle f, \psi_{-1, k} \rangle|^2 = \sum_{k \in \mathbb{Z}^2} \left| \int_{Q^2} \widehat{f}(\xi) \widehat{\Phi}(\xi) e^{-2\pi i \xi^\top k} d\xi \right|^2 = \int_{\mathbb{R}^2} |\widehat{f}(\xi)|^2 |\widehat{\Phi}(\xi)|^2 d\xi.$$

Finally, by using (CASE1)–(CASE4), we can conclude that

$$\begin{aligned} & \sum_{\psi \in \text{SH}(\phi, v, (\alpha_j)_j)} |\langle f, \psi \rangle|^2 \\ & = \sum_{k \in \mathbb{Z}^2} \left(|\langle f, \psi_{-1, k} \rangle|^2 + \sum_{\iota \in \{\text{h}, \text{v}\}} \sum_{j \in \mathbb{N}_0} \sum_{|l| < l_j} |\langle f, \psi_{j, l, k}^{\alpha_j, (\iota)} \rangle|^2 + \sum_{j \in \mathbb{N}_0} \sum_{l = \pm l_j} |\langle f, \psi_{j, l, k}^{\alpha_j} \rangle|^2 \right) \\ & = \int_{\mathbb{R}^2} |\widehat{f}(\xi)|^2 |\widehat{\Phi}(\xi)|^2 d\xi \\ & \quad + \int_{\mathbb{R}^2} |\widehat{f}(\xi)|^2 \sum_{j \in \mathbb{N}_0} |W(2^{-2j}\xi)|^2 \left[\sum_{|l| < l_j} \left| v\left(l_j \frac{\xi_2}{\xi_1} - l\right) \right|^2 + \sum_{|l| < l_j} \left| v\left(l_j \frac{\xi_1}{\xi_2} - l\right) \right|^2 \right] d\xi \\ & \quad + \int_{\mathcal{C}(\text{h})} |\widehat{f}(\xi)|^2 \sum_{j \in \mathbb{N}_0} |W(2^{-2j}\xi)|^2 |v(l_j (\xi_2/\xi_1 - 1))|^2 d\xi \\ & \quad + \int_{\mathcal{C}(\text{v})} |\widehat{f}(\xi)|^2 \sum_{j \in \mathbb{N}_0} |W(2^{-2j}\xi)|^2 |v(l_j (\xi_1/\xi_2 - 1))|^2 d\xi \\ & \quad + \int_{\mathcal{C}(\text{h})} |\widehat{f}(\xi)|^2 \sum_{j \in \mathbb{N}_0} |W(2^{-2j}\xi)|^2 |v(l_j (\xi_2/\xi_1 + 1))|^2 d\xi \\ & \quad + \int_{\mathcal{C}(\text{v})} |\widehat{f}(\xi)|^2 \sum_{j \in \mathbb{N}_0} |W(2^{-2j}\xi)|^2 |v(l_j (\xi_1/\xi_2 + 1))|^2 d\xi \\ & = \int_{\mathbb{R}^2} |\widehat{f}(\xi)|^2 |\widehat{\Phi}(\xi)|^2 d\xi + \int_{\mathbb{R}^2} |\widehat{f}(\xi)|^2 \sum_{j \in \mathbb{N}_0} |W(2^{-2j}\xi)|^2 \times \\ & \quad \times \left[\underbrace{\chi_{\mathcal{C}(\text{h})}(\xi) \sum_{|l| \leq l_j} \left| v\left(l_j \frac{\xi_2}{\xi_1} - l\right) \right|^2}_{\stackrel{(3.1)}{=} 1} + \underbrace{\chi_{\mathcal{C}(\text{v})}(\xi) \sum_{|l| \leq l_j} \left| v\left(l_j \frac{\xi_1}{\xi_2} - l\right) \right|^2}_{\stackrel{(3.1)}{=} 1} \right] d\xi \end{aligned}$$

$$= \int_{\mathbb{R}^2} |\widehat{f}(\xi)|^2 \underbrace{\left[|\widehat{\Phi}(\xi)|^2 + \sum_{j \in \mathbb{N}_0} |W(2^{-2j}\xi)|^2 \right]}_{\stackrel{(3.2)}{=} 1} d\xi = \|\widehat{f}\|_{L^2}^2 = \|f\|_{L^2}^2.$$

This finishes the proof. \blacksquare

7.2. Proofs of Decay Lemmas from Subsection 5.1. The approaches of the following two proofs are significantly different from each other: The first result, which is Lemma 5.3, covers the cases of shearlet elements which “touch” or “overlap” the singularity $w_{\mathfrak{L}_j}$. This requires application of the precise definitions (which means working with the shearlet Fourier representations) and utilizing specific properties such as vanishing moments conditions. In contrast to this, the statement of Lemma 5.4 is a rather general geometric insight: Due to the rapid decay of the models and shearlet functions, the coefficients $|\langle w_{\mathfrak{L}_j}, \psi_{j,l,k}^{\alpha_j,(\nu)} \rangle|$ will become small whenever the “distance” of $w_{\mathfrak{L}_j}$ and $\psi_{j,l,k}^{\alpha_j,(\nu)}$ is sufficiently large. Its proof is therefore essentially based on spatial estimates, and avoids any Fourier techniques.

Proof of Lemma 5.3.

(i) Plancherel’s Theorem and the definition of universal shearlets imply

$$\begin{aligned} \langle w_{\mathfrak{L}_j}, \psi_{j,l,k}^{\alpha_j,(\nu)} \rangle &= \langle \widehat{w}_{\mathfrak{L}_j}, \widehat{\psi}_{j,l,k}^{\alpha_j,(\nu)} \rangle = \int_{\mathbb{R}^2} \widehat{w}(\xi_1) W(2^{-2j}\xi) \overline{\widehat{\psi}_{j,l,k}^{\alpha_j,(\nu)}(\xi)} d\xi \\ &= \int_{\mathbb{R}} e^{2\pi i t_2^{(\nu)} \xi_2} \int_{\mathbb{R}} \widehat{w}(\xi_1) \underbrace{W(2^{-2j}\xi) \widehat{\psi}_{j,l,0}^{\alpha_j,(\nu)}(\xi)}_{=:\sigma_{j,l}(\xi)} e^{2\pi i t_1^{(\nu)} \xi_1} d\xi_1 d\xi_2. \end{aligned}$$

Next, we perform partial integrations, where $\xi \mapsto \widehat{w}(\xi_1)\sigma_{j,l}(\xi)$ is differentiated and the modulation terms are integrated, respectively. Repeating this $N_i^{(\nu)}$ -times for $i = 1, 2$ (recall that $N_i^{(\nu)} = 0$ when $t_i^{(\nu)} = 0$), we obtain

$$(7.5) \quad \left| \langle w_{\mathfrak{L}_j}, \psi_{j,l,k}^{\alpha_j,(\nu)} \rangle \right| \leq C_{N_1, N_2} \left| t_1^{(\nu)} \right|^{-N_1^{(\nu)}} \left| t_2^{(\nu)} \right|^{-N_2^{(\nu)}} \underbrace{\int_{\mathbb{R}} \int_{\mathbb{R}} \left| D^{(N_1^{(\nu)}, N_2^{(\nu)})} [\widehat{w}(\xi_1)\sigma_{j,l}(\xi)] \right| d\xi_1 d\xi_2}_{=: h_{N_1, N_2}(\xi_2)},$$

where the boundary terms vanish due to the compact support of $\xi \mapsto \widehat{w}(\xi_1)\sigma_{j,l}(\xi)$. From (3.7) (the roles of ξ_1 and ξ_2 of course need to be interchanged), we conclude for any $\xi \in \text{supp } \sigma_{j,l}$ that

$$(7.6) \quad (l-1)2^{(\alpha_j-2)j} \leq \frac{\xi_1}{\xi_2} \leq (l+1)2^{(\alpha_j-2)j} \quad \text{and} \quad 2^{2j-4} \leq |\xi_2| \leq 2^{2j-1}.$$

The assumption $|l| > 1$ now implies that there exist constants $C_1, C_2 > 0$ such that

$$\xi_1 \in I_{j,l} := [-C_2 l 2^{\alpha_j j}, -C_1 l 2^{\alpha_j j}] \cup [C_1 l 2^{\alpha_j j}, C_2 l 2^{\alpha_j j}].$$

The Leibniz rule with respect to ξ_1 yields

$$D^{(N_1^{(\nu)}, N_2^{(\nu)})} [\widehat{w}(\xi_1)\sigma_{j,l}(\xi)] = \sum_{n_1=0}^{N_1^{(\nu)}} \binom{N_1^{(\nu)}}{n_1} \widehat{w}^{(n_1)}(\xi_1) D^{(N_1^{(\nu)}-n_1, N_2^{(\nu)})} \sigma_{j,l}(\xi), \quad \xi \in \mathbb{R}^2,$$

and, combined with Hölder’s inequality, we obtain (recall the definition of h_{N_1, N_2} from (7.5))

$$\begin{aligned} h_{N_1, N_2}(\xi_2) &\leq \sum_{n_1=0}^{N_1^{(\nu)}} \binom{N_1^{(\nu)}}{n_1} \int_{\mathbb{R}} \left| \widehat{w}^{(n_1)}(\xi_1) D^{(N_1^{(\nu)}-n_1, N_2^{(\nu)})} \sigma_{j,l}(\xi) \right| d\xi_1 \\ &\leq \sum_{n_1=0}^{N_1^{(\nu)}} \binom{N_1^{(\nu)}}{n_1} \|\widehat{w}^{(n_1)}\|_{L^1(I_{j,l})} \|D^{(N_1^{(\nu)}-n_1, N_2^{(\nu)})} \sigma_{j,l}\|_{L^\infty(\mathbb{R}^2)}. \end{aligned}$$

To estimate the first factor, we use the rapid decay of \widehat{w} as well as the specific form of $I_{j,l}$:

$$\begin{aligned} \|\widehat{w}^{(n_1)}\|_{L^1(I_{j,l})} &\leq \text{vol}(I_{j,l}) \sup_{\xi_1 \in I_{j,l}} |\widehat{w}^{(n_1)}(\xi_1)| \leq C_{M,n_1} \cdot |l2^{\alpha_j j}| \cdot \langle |l2^{\alpha_j j}| \rangle^{-(M+1)} \\ &\leq C_{M,n_1} \langle |l2^{\alpha_j j}| \rangle^{-M} \stackrel{|l|>1}{\leq} C_{M,n_1} \langle |2^{\alpha_j j}| \rangle^{-M}. \end{aligned}$$

For the other term, we proceed similar to (7.2) und (7.3) and additionally use (7.6):

$$\begin{aligned} 2^{(2+\alpha_j)j/2} \left| \frac{\partial \sigma_{j,l}}{\partial \xi_1}(\xi) \right| &= \left| \frac{\partial}{\partial \xi_1} \left[W^2(2^{-2j}\xi) v \left(2^{(2-\alpha_j)j} \frac{\xi_1}{\xi_2} - l \right) \right] \right| \\ &\leq C_1 2^{-2j} + C_2 \frac{2^{(2-\alpha_j)j}}{|\xi_2|} \stackrel{\alpha_j \leq 2}{\leq} C 2^{-\alpha_j j}, \\ 2^{(2+\alpha_j)j/2} \left| \frac{\partial \sigma_{j,l}}{\partial \xi_2}(\xi) \right| &= \left| \frac{\partial}{\partial \xi_2} \left[W^2(2^{-2j}\xi) v \left(2^{(2-\alpha_j)j} \frac{\xi_1}{\xi_2} - l \right) \right] \right| \\ &\leq C_1 2^{-2j} + C_2 2^{(2-\alpha_j)j} \frac{|\xi_1|}{|\xi_2|^2} \leq C 2^{-\alpha_j j}. \end{aligned}$$

By induction, we obtain

$$\|D^{(N_1^{(v)} - n_1, N_2^{(v)})} \sigma_{j,l}\|_{L^\infty(\mathbb{R}^2)} \leq C_{N_1, N_2, n_1} 2^{-(2+\alpha_j)j/2} 2^{-(N_1^{(v)} - n_1)\alpha_j j} 2^{-N_2^{(v)}\alpha_j j}.$$

This implies

$$\begin{aligned} h_{N_1, N_2}(\xi_2) &\leq \sum_{n_1=0}^{N_1^{(v)}} \binom{N_1^{(v)}}{n_1} C_{N_1, N_2, n_1, M} \langle |2^{\alpha_j j}| \rangle^{-M} 2^{-(2+\alpha_j)j/2} 2^{-(N_1^{(v)} - n_1)\alpha_j j} 2^{-N_2^{(v)}\alpha_j j} \\ &\leq C_{N_1, N_2, M} \langle |2^{\alpha_j j}| \rangle^{-M} 2^{-(2+\alpha_j)j/2} 2^{-N_2^{(v)}\alpha_j j} \underbrace{\sum_{n_1=0}^{N_1^{(v)}} \binom{N_1^{(v)}}{n_1} (2^{-\alpha_j j})^{N_1^{(v)} - n_1}}_{=(1+2^{-\alpha_j j})^{N_1^{(v)}} \leq 2^{N_1^{(v)}}} \\ &\leq C_{N_1, N_2, M} \langle |2^{\alpha_j j}| \rangle^{-M} 2^{-(2+\alpha_j)j/2} 2^{-N_2^{(v)}\alpha_j j}, \end{aligned}$$

where we made use of the assumption $\alpha_j \geq 0$ to estimate $(1 + 2^{-\alpha_j j})^{N_1^{(v)}}$. Since $\text{vol}(\text{supp } h_{N_1, N_2}) \leq C 2^{2j}$, a final estimate gives

$$\begin{aligned} \left| \left\langle w \mathfrak{L}_j, \psi_{j,l,k}^{\alpha_j, (v)} \right\rangle \right| &\leq C_{N_1, N_2} \left| t_1^{(v)} \right|^{-N_1^{(v)}} \left| t_2^{(v)} \right|^{-N_2^{(v)}} \int_{\mathbb{R}} h_{N_1, N_2}(\xi_2) d\xi_2 \\ &\leq C_{N_1, N_2, M} \left| t_1^{(v)} \right|^{-N_1^{(v)}} \left| t_2^{(v)} \right|^{-N_2^{(v)}} \langle |2^{\alpha_j j}| \rangle^{-M} 2^{(2-\alpha_j)j/2} 2^{-N_2^{(v)}\alpha_j j}. \end{aligned}$$

(ii) This proof works analogously to the one of (i). Basically, only the intervals $I_{j,l}$ need to be modified slightly. Due to $|\xi_1| \in \Theta(2^{2j})$ for any $\xi \in \text{supp } \widehat{\psi}_{j,l,k}^{\alpha_j, (h)}$, they have the form

$$[-C_2 2^{2j}, -C_1 2^{2j}] \cup [C_1 2^{2j}, C_2 2^{2j}], \quad C_1, C_2 > 0.$$

This leads to the different factor $\langle |2^{2j}| \rangle^{-M}$ in (ii), compared with (i).

(iii) Recalling the definition of the boundary shearlets, one can easily verify this estimate by a combination of part (i) and (ii). ■

It is worth noting that the crucial point of the proof above is the specific form of the intervals $I_{j,l}$, which is due to the “geometry” of $\text{supp } \widehat{\psi}_{j,l,k}^{\alpha_j,(\nu)}$. Interestingly, the fact that 0 is not contained in $I_{j,l}$, precisely corresponds to the fact that all moments of $\psi_{j,l,k}^{\alpha_j,(\nu)}$ vanish with respect to x_2 (at least when $|l| > 1$).

Proof of Lemma 5.4. At first, we provide appropriate decay estimates of our edge model,

$$(7.7) \quad \begin{aligned} |w\mathfrak{L}_j(x)| &= |[w\mathfrak{L} * F_j](x)| = \left| \int_{\mathbb{R}} w(y_1) F_j(x - (y_1, 0)) dy_1 \right| \\ &\leq \int_{\mathbb{R}} |w(y_1)| 2^{4j} \underbrace{|\check{W}(2^{2j}(x - (y_1, 0)))|}_{\leq C_N \langle |2^{2j}x_2| \rangle^{-N} \langle |2^{2j}(y_1 - x_1)| \rangle^{-N}} dy_1 = C_N 2^{4j} \langle |2^{2j}x_2| \rangle^{-N} \underbrace{[|w| * \langle |2^{2j}[\cdot]| \rangle^{-N}]}_{=\tilde{w}_{N,j}(x_1)}(x_1) \\ &= C_N 2^{4j} \langle |2^{2j}x_2| \rangle^{-N} \tilde{w}_{N,j}(x_1), \quad x = (x_1, x_2) \in \mathbb{R}^2, \end{aligned}$$

as well as of the shearlet elements,

$$\begin{aligned} |\psi_{j,l,k}^{\alpha_j,(\nu)}(x)| &\leq C_N 2^{(\alpha_j+2)j/2} \langle |S_{(\nu)}^l A_{\alpha_j,(\nu)}^j x - k| \rangle^{-N} \\ &\leq C_N 2^{(\alpha_j+2)j/2} \langle |2^{\alpha_j j} x_1 - k_1| \rangle^{-N} \langle |l 2^{\alpha_j j} x_1 + 2^{2j} x_2 - k_2| \rangle^{-N}, \quad x = (x_1, x_2) \in \mathbb{R}^2. \end{aligned}$$

With this, we can estimate the analysis coefficients:

$$(7.8) \quad \begin{aligned} &\left| \left\langle w\mathfrak{L}_j, \psi_{j,l,k}^{\alpha_j,(\nu)} \right\rangle \right| \\ &\leq C_N \int_{\mathbb{R}^2} 2^{4j} \langle |2^{2j}x_2| \rangle^{-N} \tilde{w}_{N,j}(x_1) 2^{(\alpha_j+2)j/2} \langle |2^{\alpha_j j} x_1 - k_1| \rangle^{-N} \langle |l 2^{\alpha_j j} x_1 + 2^{2j} x_2 - k_2| \rangle^{-N} dx \\ &= C_N 2^{(5-\alpha_j)j/2} \int_{\mathbb{R}^2} \langle |x_2| \rangle^{-N} \tilde{w}_{N,j}(2^{-\alpha_j j}(x_1 + k_1)) \langle |x_1| \rangle^{-N} \langle |lx_1 + x_2 + lk_1 - k_2| \rangle^{-N} dx. \end{aligned}$$

Now, let us consider the product $\langle |x_2| \rangle^{-N} \langle |x_2 + (lx_1 + lk_1 - k_2)| \rangle^{-N}$. One of the two factors (which one it is might depend on x_2) clearly has to be smaller than $\langle |\frac{1}{2}(lx_1 + lk_1 - k_2)| \rangle^{-N}$. This yields ($K := lx_1 + lk_1 - k_2$)

$$(7.9) \quad \begin{aligned} &\int_{\mathbb{R}} \langle |x_2| \rangle^{-N} \langle |x_2 + K| \rangle^{-N} dx_2 \\ &= \int_{\mathbb{R}} \underbrace{\max\{\langle |x_2| \rangle^{-N}, \langle |x_2 + K| \rangle^{-N}\}}_{\leq \langle |x_2| \rangle^{-N} + \langle |x_2 + K| \rangle^{-N}} \underbrace{\min\{\langle |x_2| \rangle^{-N}, \langle |x_2 + K| \rangle^{-N}\}}_{\leq \langle |\frac{K}{2}| \rangle^{-N}} dx_2 \\ &\leq C_N \langle |K| \rangle^{-N} = C_N \langle |lx_1 + lk_1 - k_2| \rangle^{-N}. \end{aligned}$$

Putting this into (7.8), we finally obtain

$$\left| \left\langle w\mathfrak{L}_j, \psi_{j,l,k}^{\alpha_j,(\nu)} \right\rangle \right| \leq C_N 2^{(5-\alpha_j)j/2} \int_{\mathbb{R}} \tilde{w}_{N,j}(2^{-\alpha_j j}(x_1 + k_1)) \langle |x_1| \rangle^{-N} \langle |lx_1 + lk_1 - k_2| \rangle^{-N} dx_1. \quad \blacksquare$$

Acknowledgements. The second author acknowledges support 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.

REFERENCES

- [1] C. BALLESTER, M. BERTALMIO, V. CASELLES, G. SAPIRO, AND J. VERDERA, *Filling-in by joint interpolation of vector fields and gray levels*, IEEE Trans. Image Process., 10 (2001), pp. 1200–1211.
- [2] M. BERTALMIO, A. L. BERTOZZI, AND G. SAPIRO, *Navier-stokes, fluid dynamics, and image and video inpainting*, in Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, vol. 1, IEEE, 2001, pp. I-355–I-362.
- [3] J. BOBIN, J.-L. STARCK, J. FADILI, AND Y. MOUDDEN, *Sparsity and morphological diversity in blind source separation*, IEEE Trans. Image Process., 16 (2007), pp. 2662–2674.
- [4] J.-F. CAI, R. H. CHAN, AND Z. SHEN, *Simultaneous cartoon and texture inpainting*, Inverse Probl. Imaging, 4 (2010), pp. 379–395.
- [5] E. CANDÈS AND D. L. DONOHO, *New tight frames of curvelets and optimal representation of objects with C^2 singularities*, Comm. Pure. Appl. Math., 57 (2004), pp. 219–266.
- [6] T. F. CHAN AND J. SHEN, *Mathematical models for local nontexture inpaintings*, SIAM J. Appl. Math., 62 (2001/02), pp. 1019–1043.
- [7] I. DAUBECHIES, *Ten lectures on wavelets*, vol. 61 of CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.
- [8] J. G. DAUGMAN, *Complete discrete 2-d gabor transforms by neural networks for image analysis and compression*, IEEE Trans. Acoust., Speech, Signal Processing, 36 (1988), pp. 1169–1179.
- [9] M. DAVENPORT, M. DUARTE, Y. ELДАР, AND G. KUTYNIOK, *Compressed Sensing: Theory and Applications*, Cambridge University Press, 2012, ch. Introduction to Compressed Sensing, pp. 1–68.
- [10] B. DONG, H. JI, J. LI, Z. SHEN, AND Y. XU, *Wavelet frame based blind image inpainting*, Appl. Comput. Harmon. Anal., 32 (2012), pp. 268–279.
- [11] D. DONOHO AND G. KUTYNIOK, *Microlocal analysis of the geometric separation problem*, Comm. Pure Appl. Math., 66 (2013), pp. 1–47.
- [12] D. L. DONOHO AND X. HUO, *Uncertainty principles and ideal atomic decomposition*, IEEE Trans. Inform. Theory, 47 (2001), pp. 2845–2862.
- [13] M. ELAD, J.-L. STARCK, P. QUERRE, AND D. L. DONOHO, *Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)*, Appl. Comput. Harmon. Anal., 19 (2005), pp. 340–358.
- [14] S. FOUCART AND H. RAUHUT, *A Mathematical Introduction to Compressive Sensing*, Birkhäuser, 2013.
- [15] L. GRAFAKOS, *Classical fourier analysis*, vol. 86, Springer, 2008.
- [16] P. GROHS, S. KEIPER, G. KUTYNIOK, AND M. SCHÄFER, *α -Molecules*. preprint, 2014.
- [17] 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 (2006), pp. 189–201.
- [18] K. GUO AND D. LABATE, *Optimally sparse multidimensional representation using shearlets*, SIAM J. Math. Anal., 39 (2007), pp. 298–318.
- [19] ———, *The construction of smooth Parseval frames of shearlets*, Math. Model. Nat. Phenom., 8 (2013), pp. 82–105.
- [20] G. HENNENFENT, L. FENELON, AND F. J. HERRMANN, *Nonequispaced curvelet transform for seismic data reconstruction: A sparsity-promoting approach*, Geophysics, 75 (2010), pp. WB203–WB210.
- [21] G. HENNENFENT AND F. J. HERRMANN, *Application of stable signal recovery to seismic data interpolation*, in SEG Technical Program Expanded Abstracts 2006, Society of Exploration Geophysicists, 2006, pp. 2797–2801.
- [22] F. J. HERRMANN AND G. HENNENFENT, *Non-parametric seismic data recovery with curvelet frames*, Geophys. J. Int., 173 (2008), pp. 233–248.
- [23] D. H. HUBEL AND T. N. WIESEL, *Receptive fields of single neurones in the cat's striate cortex*, J. Physiol., 148 (1959), pp. 574–591.
- [24] S. KEIPER, *A flexible shearlet transform - sparse approximations and dictionary learning*, bachelor's thesis, Technische Universität Berlin, 2012.
- [25] E. J. KING, G. KUTYNIOK, AND X. ZHUANG, *Analysis of data separation and recovery problems using clustered sparsity*, in Proceedings of the SPIE, Wavelets and Sparsity XIV, vol. 8138, 2011, pp. 813818–813818–11.
- [26] E. J. KING, G. KUTYNIOK, AND X. ZHUANG, *Analysis of inpainting via clustered sparsity and microlocal analysis.*, J. Math. Imaging Vis., 48 (2014), pp. 205–234.
- [27] P. KITTIPOOM, G. KUTYNIOK, AND W.-Q LIM, *Construction of compactly supported shearlet frames*, Constr. Approx., 35 (2012), pp. 21–72.
- [28] N. KOMODAKIS AND G. TZIRITAS, *Image completion using efficient belief propagation via priority scheduling and dynamic pruning*, IEEE Trans. Image Process., 16 (2007), pp. 2649–2661.
- [29] G. KUTYNIOK, *Clustered sparsity and separation of cartoon and texture*, SIAM J. Imaging Sci., 6 (2013), pp. 848–874.
- [30] ———, *Geometric separation by single-pass alternating thresholding*, Appl. Comput. Harmon. Anal., 36 (2014), pp. 23–50.
- [31] G. KUTYNIOK AND D. LABATE, *Shearlets: Multiscale Analysis for Multivariate Data*, Birkhäuser Boston, 2012, ch. Introduction to Shearlets, pp. 1–38.
- [32] G. KUTYNIOK, J. LEMVIG, AND W.-Q LIM, *Optimally sparse approximations of 3D functions by compactly supported*

- shearlet frames*, SIAM J. Math. Anal., 44 (2012), pp. 2962–3017.
- [33] G. KUTYNIOK AND W.-Q LIM, *Compactly supported shearlets are optimally sparse*, J. Approx. Theory, 163 (2011), pp. 1564–1589.
- [34] G. KUTYNIOK, W.-Q LIM, AND R. REISENHOFER, *Shearlab 3D: Faithful digital shearlet transforms based on compactly supported shearlets*. preprint, 2014.
- [35] S. NAM, M. E. DAVIES, M. ELAD, AND R. GRIBONVAL, *Cospase analysis modeling-uniqueness and algorithms*, in Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on, IEEE, 2011, pp. 5804–5807.
- [36] ———, *The cospase analysis model and algorithms*, Appl. Comput. Harmon. Anal., 34 (2013), pp. 30–56.
- [37] R. RUBINSTEIN, T. PELEG, AND M. ELAD, *Analysis K -SVD: a dictionary-learning algorithm for the analysis sparse model*, IEEE Trans. Signal Process., 61 (2013), pp. 661–677.
- [38] R. RUBINSTEIN, M. ZIBULEVSKY, AND M. ELAD, *Double sparsity: Learning sparse dictionaries for sparse signal approximation*, IEEE Trans. Signal Proc., 58 (2010), pp. 1553–1564.