Journal of Mathematical Imaging and Vision manuscript No.
(will be inserted by the editor)

Analysis of Inpainting via Clustered Sparsity and Microlocal Analysis

Emily J. King · Gitta Kutyniok · Xiaosheng Zhuang

Received: date / Accepted: date

Abstract Recently, compressed sensing techniques in combination with both wavelet and directional representation systems have been very effectively applied to the problem of image inpainting. However, a mathematical analysis of these techniques which reveals the underlying geometrical content is completely missing. In this paper, we provide the first comprehensive analysis in the continuum domain utilizing the novel concept of clustered sparsity, which besides leading to asymptotic error bounds also makes the superior behavior of directional representation systems over wavelets precise. First, we propose an abstract model for problems of data recovery and derive error bounds for two different recovery schemes, namely ℓ_1 minimization and thresholding. Second, we set up a microlocal model for an image governed by edges with a particular focus on seismic data as well as a mask to model the missing data. Applying the abstract estimate in the case of wavelets and of shearlets we prove that – provided the size of the missing part is comparable to the size of the analyzing functions – asymptotically precise inpainting can be obtained. Finally, we show that shearlets can fill strictly larger gaps than wavelets.

Keywords ℓ_1 Minimization · Cluster Coherence · Inpainting · Parseval Frames · Sparse Representation · Data Recovery · Shearlets · Meyer Wavelets

1 Introduction

A common problem is that of missing data. The human visual system has an amazing ability to fill in the missing parts of images, but automating this process is not trivial. Also, depending on the type of data, the human senses may be unable to fill in the gaps. Conservators working to repair damaged paintings use the term *inpainting* to describe the process. This word now also means digitally recovering missing data in videos and images. The removal of overlaid text in images, the repair of scratched photos and audio recordings, and the recovery of missing blocks in a streamed video are all examples of inpainting. Seismic data are also commonly incomplete due to land development and bodies of water preventing optimal sensor placement [23, 24].

One very common approach to inpainting is using variational methods [2–4, 11]. However, recently the novel methodology of compressed sensing, namely exact recovery of sparse or sparsified data from highly incomplete linear nonadaptive measurements by ℓ_1 minimization or thresholding, has been very effectively applied to this problem. The pioneering paper is [18], which uses curvelets as sparsifying system for inpainting. Various intriguing successive empirical results have since then been obtained using applied harmonic analysis in combination with convex optimization [5, 13, 18]. Theoretical analysis of those types of methods typically use a discrete model which does not allow the geome-

Emily J. King is supported by a fellowship for postdoctoral researchers from the Alexander von Humboldt Foundation. Gitta Kutyniok would like to thank David Donoho for discussions on this and related topics. She is grateful to the Department of Statistics at Stanford University and the Department of Mathematics at Yale University for their hospitality and support during her visits. She also acknowledges support by the Einstein Foundation Berlin, by Deutsche Forschungsgemeinschaft (DFG) Heisenberg fellowship KU 1446/8, Grant SPP-1324 KU 1446/13 and DFG Grant KU 1446/14, and by the DFG Research Center MATHEON "Mathematics for key technologies" in Berlin. Xiaosheng Zhuang acknowledges support by DFG Grant KU 1446/14.

Department of Mathematics, Technische Universität Berlin, 10623 Berlin, Germany

E. J. King
E-mail: king@math.tu-berlin.de

G. Kutyniok
E-mail: kutyniok@math.tu-berlin.de

X. Zhuang
E-mail: xzhuang@math.tu-berlin.de

try of the problem to be taken into account. However, variational methods are built on continuous methods and may be analyzed using a continuous model, for example, [10]. Also, some work has been done to compare variational approaches with those built on ℓ_1 minimization [6,36]. It also prohibits a deep understanding of why directional representation systems such as shearlets outperform wavelets when inpainting images strongly governed by curvilinear structures such as seismic images.

The preliminary results presented in the *SPIE Proceedings* paper [26] combined with the theory in this paper provide the first comprehensive analysis of discrete dictionaries inpainting the continuum domain utilizing the novel concept of clustered sparsity, which besides leading to asymptotic error bounds also makes the superior behavior of directional representation systems over wavelets precise. Along the way, our abstract model and analysis even lay a common theoretical foundation for all data recovery problems when utilizing either ℓ_1 minimization or thresholding as recovery schemes.

1.1 A Continuum Model

One of the first practitioners of curvelet inpainting for applications was the seismologist Felix Herrmann, who achieved superior recovery results for images which consisted of curvilinear singularities in which vertical strips are missing due to missing sensors. These techniques were soon also exploited for astronomical imaging, etc., the common bracket being the governing by curvilinear singularities. It is evident, that no *discrete* model can appropriately capture such a geometrical content.

Thus a continuum domain model seems appropriate. In fact, in this paper we choose a distributional model which is a weighted distribution $w\mathscr{L}$ acting on "nice" functions by

$$\langle w\mathscr{L}, f\rangle = \int_{-\rho}^{\rho} w(x_1) f(x_1,0)dx_1,$$

the weight w and length 2ρ being specified in the main body of the paper. Mimicking the seismic imaging situation, we might then choose the shape of the missing part to be

$$\mathscr{M}_h = \mathbb{1}_{\{|x_1|\le h\}},$$

i.e., a vertical strip of width $2h$. Further, we let $P_{\mathscr{M}_h}$ and $P_{\mathbf{R}^2\setminus\mathscr{M}_h}$ denote the orthogonal projection of $L^2(\mathbf{R}^2)$ onto the missing part and the known part, respectively. One task can now be formulated mathematically precise in the following way. Given

$$f = P_{\mathbf{R}^2\setminus\mathscr{M}_h} w\mathscr{L},$$

recover $w\mathscr{L}$.

It should be mentioned that such a microlocal viewpoint was first introduced and studied in the situation of image separation [16].

1.2 Sparsifying Systems

It was recently made precise that the optimal sparsifying systems for such images governed by anisotropic structures are curvelets [7] and shearlets [30,34]. Of these two systems shearlets have the advantage that they provide a unified concept of the continuum and digital domain, which curvelets do not achieve. However, many inpainting algorithms even still use wavelets, and one might ask whether shearlets provably outperform wavelets. In fact, we will make the superior behavior of shearlets within our model situation precise.

For our analysis, we will use the following type of wavelets and of shearlets.

1.2.1 Wavelets

Meyer wavelets are some of the earliest known examples of orthonormal wavelets; they also happen to have high regularity [12,35]. For the construction, let $\nu \in C^\infty(\mathbf{R})$ satisfy $\nu(\cdot)+\nu(1-\cdot) = \mathbb{1}_{\mathbb{R}}(\cdot)$, where the *indicator function* $\mathbb{1}_A$ is defined to take the value 1 on A and 0 on A^c, and

$$\nu(x) = \begin{cases} 0 : x \le 0, \\ 1 : x \le 1. \end{cases}$$

Then the Fourier transform of the 1D Meyer is defined by

$$W(\xi) = \begin{cases} e^{-\pi i\xi}\sin\left[\frac{\pi}{2}\nu\left(3|\xi|-1\right)\right] & : \frac{1}{3}\le|\xi|\le\frac{2}{3}, \\ e^{-\pi i\xi}\cos\left[\frac{\pi}{2}\nu\left(\frac{3}{2}|\xi|-1\right)\right] & : \frac{2}{3}\le|\xi|\le\frac{4}{3}, \\ 0 & : \quad \text{else}, \end{cases}$$

and the Fourier transform of the 1D Meyer scaling function is

$$\hat{\phi}(\xi) = \begin{cases} 1 & : \quad |\xi|\le\frac{1}{3}, \\ \cos\left[\frac{\pi}{2}\nu\left(3|\xi|-1\right)\right] & : \frac{1}{3}\le|\xi|\le\frac{2}{3}, \\ 0 & : \quad \text{else}. \end{cases}$$

We will not detail the interpretation of a scaling function but refer the interested reader to [12,35]. Then we define the $C^\infty\cap L^2(\mathbf{R}^2)$-functions W^v, W^h, and W^d by

$$\begin{aligned} W^v(\xi) &= \hat{\phi}(\xi_1)W(\xi_2), \\ W^h(\xi) &= W(\xi_1)\hat{\phi}(\xi_2), \text{ and} \\ W^d(\xi) &= W(\xi_1)W(\xi_2). \end{aligned}$$

We denote

$$\hat{\psi}_\lambda(\xi) = 2^{-j}W^\iota(\xi/2^j)e^{-2\pi ik\xi/2^j}, \quad \lambda=(\iota,j,k).$$

Then the *orthonormal Meyer wavelet system* is given by

$$\{\psi_\lambda : \lambda=(\iota,j,k), \iota\in\{h,v,d\}, j\in\mathbf{Z}, k\in\mathbf{Z}^2\}$$

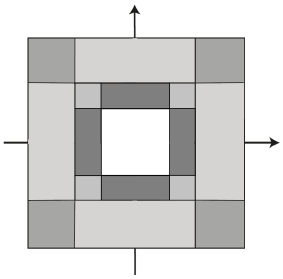

Fig. 1 Frequency tiling of Meyer wavelets.

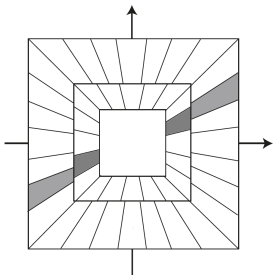

Fig. 2 Frequency tiling of cone-adapted shearlets.

1.2.2 Shearlets

Let the parabolic scaling matrices A_a^h and A_a^v and shearing matrices S_s^h and S_s^v be defined as

$$A_a^h = \begin{bmatrix} a & 0 \\ 0 & \sqrt{a} \end{bmatrix}, \quad A_a^v = \begin{bmatrix} \sqrt{a} & 0 \\ 0 & a \end{bmatrix},$$
$$S_s^h = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix}, \quad \text{and} \quad S_s^v = \begin{bmatrix} 1 & 0 \\ s & 1 \end{bmatrix}.$$

Let $V \in L^2(\mathbf{R})$ satisfy $\hat{V} \in C^\infty(\mathbf{R})$, supp $\hat{V} \subseteq [-1,1]$, and

$$\sum_{k=-1}^{1} |\hat{V}(\xi + \pi k)|^2 = 1, \quad \xi \in [-1,1].$$

The horizontal σ^h and vertical σ^v shearlets are defined using their Fourier transforms as the product of the Meyer window function W and of the "bump function" V as

$$\hat{\sigma}^h(\xi_1,\xi_2) = W(\xi_1)V(\frac{\xi_2}{\xi_1}), \quad \hat{\sigma}^v(\xi_1,\xi_2) = W(\xi_2)V(\frac{\xi_1}{\xi_2}).$$

We employ the notation

$$\hat{\sigma}_\eta = 2^{3j/4}\sigma^\iota(S_\ell^\iota A_{2^j}^\iota \cdot - k), \quad \eta = (\iota, j, k, \ell),$$

where $\iota \in \{h,v\}$, $j \in \mathbf{Z}$, $k \in \mathbf{Z}^2$, and $\ell \in \mathbf{Z}$. Then the *cone-adapted shearlet system* is

$$\{\phi(\cdot - k) : k \in \mathbf{Z}^2\}$$
$$\cup\{\sigma_\eta : \iota \in \{h,v\}, j \in \mathbf{Z}, j \geq 0, k \in \mathbf{Z}^2, \ell \in \mathbf{Z}, |\ell| \leq \lceil 2^{j/2} \rceil\},$$

where $\lceil x \rceil$ denotes the smallest integer $\geq x$. This system is a Parseval frame, which is a generalization of an orthonormal basis. This concept will be specifically defined later.

1.3 Recovery Algorithms

We next decide upon a recovery strategy. Compressed sensing offers a variety of such, the most common ones being ℓ_1 minimization and thresholding. We will also use these. However, for preparation purposes to derive an asymptotic scale dependent analysis – the fact that the energy of our model is arbitrary high frequencies requires this approach –, we first perform a band-pass filtering on $w\mathscr{L}$. The band-pass filter will be roughly speaking chosen according to the band given be wavelets and shearlets, see Figures 1 and 2, leading to the sequence

$$(f_j)_j = (P_{\mathbf{R}^2 \setminus \mathscr{M}_h} w\mathscr{L}_j)_j.$$

The ℓ_1 minimization problem we choose has the form

$$L_j = \text{argmin}_L \|\Phi^* L\|_1 \text{ subject to } f_j = P_{\mathbf{R}^2 \setminus \mathscr{M}_h}(L). \tag{1}$$

We emphasize that this variant minimizes the *analysis* coefficients and is hence related to the newly introduced cosparsity model [37,38].

The thresholding strategy we choose is brutally simple. We only perform one step of hard thresholding, namely, setting $\mathscr{T}_j = \{i : |\langle f_j, \phi_i \rangle| \geq \beta_j\}$ for some threshold β_j, the reconstructed image is

$$L_j = \Phi \mathbb{1}_{\mathscr{T}_j} \Phi^* w\mathscr{L}_j. \tag{2}$$

It is thus surprising that the geometry is strong enough to achieve the same asymptotic recovery results as for ℓ_1 minimization.

1.4 Clustered Sparsity

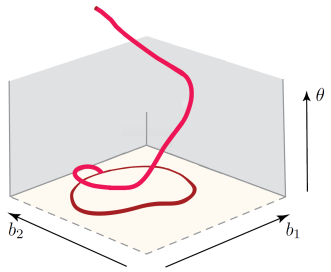

Fig. 3 Wavefront set of a curvilinear singularity $\mathscr{C}$.

One might ask where the geometry we mentioned before will come into play. This can best be explained and illustrated using microlocal analysis in phase space. For a more detailed explanation of the fundamentals of microlocal analysis, see [25], and for applications of microlocal analysis to shearlet and curvelet theory, see [8,20,29]. Phase space in

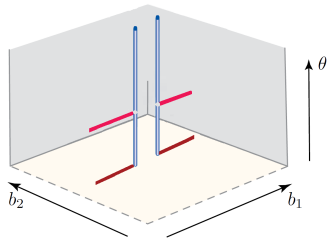

Fig. 4 Wavefront set of a masked linear singularity $\mathscr{M}_h w\mathscr{L}$.

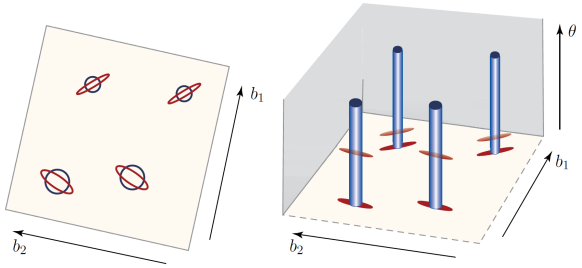

Fig. 5 Left: Effective supports of wavelets (disks) and shearlets (ellipses). Right: Phase space portrait of the same wavelets (spikes) and shearlets (ellipses).

this context is indexed by position-orientation pairs (b,θ) which describe the singular behavior of a distribution. The orientation component θ is an element of real projective space, which for simplicity's sake we shall identify in what follows with $[0,\pi)$. The wavefront set $WF(f)$ of a distribution f is roughly the set of elements in the phase space at which f is nonsmooth. First consider a curvilinear singularity $\mathscr{C}$ along a closed curve $\tau:[0,1]\to\mathbf{R}^2$:

$$\mathscr{C}=\int \delta_{\tau(t)}(\cdot)dt,$$

where δ_x is the usual Dirac delta distribution located at x. As illustrated in Figure 3, the wavefront set of $\mathscr{C}$ is

$$WF(\mathscr{C})=\{(\tau(t),\theta(t)):t\in[0,1]\},$$

where $\theta(t)$ is the normal direction of $\mathscr{C}$ at $\tau(t)$. Now consider the model from Section 1.1,

$$f=P_{\mathbf{R}^2\setminus\mathscr{M}_h}w\mathscr{L}.$$

As can be seen in Figure 4 the wavefront set of f almost looks like f itself except that the wavefront set fills all possible angles (i.e., forms a spike) at the end points of the missing mask. This is because at the end points, the distribution is singular in all but the parallel direction. The difference between the wavefront sets of shearlets and wavelets is demonstrated in Figure 5. Furthermore, in Figure 6 (Left) the small overlap of the wavefront set of a cluster of shearlets with a spike in the phase space, which represents an end point of the mask of missing information $\mathscr{M}_h$, can be clearly seen. Thus shearlet clusters are incoherent with the end points. On the other hand, it is easy to see how easily a cluster of shearlets can span a gap of missing data (Figure 6 (Right)).

1.5 Asymptotical Analysis

The width of the area to be inpainted plays a key role, even when using other inpainting techniques. In [9], variational inpainting methods are analyzed theoretically, showing that the local thickness of the area to be inpainted affects the success of the inpainting more than the overall size of the area to be inpainted.

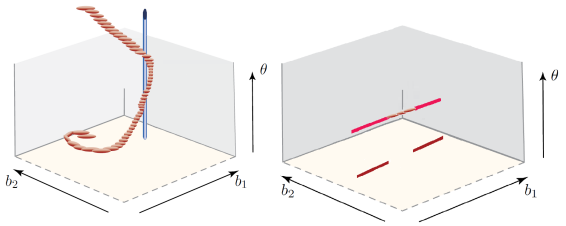

Fig. 6 Left: Phase space portrait of a cluster of shearlets and one single wavelet. Right: Phase space portrait of shearlets tiling a gap.

Thus our analysis shall also take this into account. We accomplish this by also making the gap size h dependent on the scale j. This leads to the problem of recovering $w\mathscr{L}_j$ from knowledge of

$$f_j=P_{\mathbf{R}^2\setminus\mathscr{M}_{h_j}}w\mathscr{L}_j,$$

for each scale j. Letting L_j denote the recovered image by either one of the proposed algorithms, we will show that asymptotically precise inpainting, i.e.,

$$\frac{\|L_j-w\mathscr{L}_j\|_2}{\|w\mathscr{L}_j\|_2}\to 0,\qquad j\to\infty,$$

is achieved for wavelets provided that $h_j=o(2^{-j})$ (Theorems 1 and 2) as $j\to\infty$ and for shearlets provided that $h_j=o(2^{-j/2})$ (Theorems 4 and 5) as $j\to\infty$. In fact, this is exactly what one would imagine. Inpainting succeeds provided that the gap size is comparable to the size of the analyzing elements.

1.6 Wavelets versus Shearlets

This observation seems to indicate that shearlets indeed perform better than wavelets. However, the previously mentioned theorems just state positive results. In order to show that shearlets outperform wavelets in the model situation

which we consider, we require a negative result of the following type: If $h_j > O(2^{-j/2})$ as $j \to \infty$ and L_j is recovered by wavelets, then

$$\frac{\|L_j - w\mathscr{L}_j\|_2}{\|w\mathscr{L}_j\|_2} \not\to 0, \qquad j \to \infty.$$

And in fact, this is what we will prove in Theorem 7. In this sense, we now have a mathematically precise statement showing that shearlets are strictly better for inpainting provided that the image is governed by curvilinear structures.

The only slight disappointment is the fact that this statement will only be proven for thresholding as the recovery scheme. We strongly suspect that this result also holds for ℓ_1 minimization. However, we are not aware of any analysis tools strong enough to derive these results also in this situation.

1.7 Extensions

Our analysis has focused primarily on revealing the fundamental mathematical concepts which lead to successful image inpainting using wavelets or shearlets. The viewpoint we take, however, is that this is just the "tip of the iceberg," and the main results are susceptible of very extensive generalizations and extensions.

- *More General Models*. Our results can be generalized to a much broader setting. For example, using bending techniques explored in [28], the line singularities can be extended to curvilinear singularities.
- *Different Masks*. In this paper, we focus on a vertical strip as mask. One might also think of a ball with radius h as mask, in which case similar results can be obtained. Other imaginable shapes could be horizontal strips flat ellipsoids.
- *Different Recovery Techniques*. Both hard and soft iterative thresholding techniques are quite common and usually produce convincing results. The results in this paper concern one-step-(hard)-thresholding rather than iterative thresholding. Although extending these results to iterative thresholding would not be straightforward, we strongly believe that a similar abstract analysis can be derived leading to asymptotically precise inpainting results.
- *Other Dictionaries*. It should also be pointed out that the results in this paper hold not only for the Meyer orthonormal wavelets and shearlets, but also, for instance, for radial wavelets – or any types of wavelets with isotropic feature at each scale similar to the radial wavelets – and other directional multiscale representation systems such as curvelets. The necessary changes in the proofs are foreseeable. Also, the novel framework of parabolic molecules advocated in [21] could be applied. Furthermore given the construction of 3-dimensional shearlets in [22,31–33], it seems likely that the proofs in Sections 5 and 7 will easily generalize to the 3-dimensional case.
- *Noise*. Data is typically affected by noise, a situation we considered in the abstract setting. This analysis can be directly applied also for the wavelet and shearlet inpainting results, leading to the same asymptotical behavior, provided that the noise n is small comparing to the signal; i.e., the ℓ_1 norm of $\Phi^* n$ is of order smaller than the ℓ_2 norm of filtered signal.

1.8 Contents

We begin in Section 2 with an abstract analysis of data recovery via ℓ_1 minimization introducing clustered sparsity and concentration in a Hilbert space as tools. We then apply the results in Section 2 to a particular class of inpainting problems which are described in Section 3. In Sections 4 and 5, we prove that both wavelets and shearlets, respectively, are able to inpaint a missing band but that shearlets can handle wider gaps. It is shown in Section 6 that the inpainting result for wavelets in Section 4 is tight; i.e., shearlets strictly outperform wavelets in the considered model situation. Finally, Section 7 is an appendix that contains auxiliary results concerning shearlets needed for Section 5.

2 Abstract Analysis of Data Recovery

We start by analyzing missing data recovery via ℓ_1 minimization and thresholding in an abstract model situation. The error estimates we will derive can be applied in a variety of situations. In this paper, – as discussed before – we aim to utilize them to analyze inpainting via wavelets and shearlets following a continuum domain model. In fact, these error estimates will later on be applied to each scale while driving an asymptotic analysis.

2.1 Abstract Model

Let $x^0 \in \mathscr{H}$ be a signal in a Hilbert space $\mathscr{H}$. To model the data recovery problem correctly, we assume that $\mathscr{H}$ can be decomposed into a direct sum

$$\mathscr{H} = \mathscr{H}_M \oplus \mathscr{H}_K$$

of a subspace $\mathscr{H}_M$ which is associated with the *missing* part of x^0 and a subspace $\mathscr{H}_K$ which relates to the *known* part of the signal. Further, let P_M and P_K denote the orthogonal projections onto those subspaces, respectively. The problem of data recovery can then be formulated as follows: Assuming that $P_K x^0$ is known to us, recover x^0.

We now present the concept of Parseval frames, which generalize orthonormal bases in a manner which will be useful in the sequel.

Definition 1 A collection of vectors $\Phi = \{\varphi_i\}_{i\in I}$ in a separable Hilbert space $\mathscr{H}$ forms a *Parseval frame* for $\mathscr{H}$ if for all $x \in \mathscr{H}$,

$$\sum_{i\in I} |\langle x, \varphi_i\rangle|^2 = \|x\|^2.$$

With a slight abuse of notation, given a Parseval frame Φ, we also use Φ to denote the *synthesis operator*

$$\Phi : \ell_2(I) \to \mathscr{H}, \quad \Phi(\{c_i\}_{i\in I}) = \sum_{i\in I} c_i \varphi_i.$$

With this notation, Φ^* is called the *analysis operator*.

Following the philosophy of compressed sensing, suppose that there exists a Parseval frame Φ which – in a way yet to be made precise – sparsifies the original signal x^0. Either Φ can be selected non-adaptively such as choosing a wavelet or shearlet system which will be our avenue in the sequel, or Φ can be chosen adaptively using dictionary learning algorithms such as [1,19,39].

To already draw the connection to the special situation of inpainting at this point, assume that $\mathscr{H} = L^2(\mathbf{R}^2)$. If the measurable subset $B \subseteq \mathbf{R}^2$ is the missing area of the image, we set $\mathscr{H}_K = L^2(\mathbf{R}^2 \setminus B)$ and $\mathscr{H}_M = L^2(B)$.

2.2 Inpainting via ℓ_1 Minimization

A first methodology from compressed sensing to achieve recovery is ℓ_1 minimization, which recovers the original signal by solving

$$(\text{INP}) \qquad x^\star = \operatorname{argmin}_x \|\Phi^* x\|_1 \text{ subject to } P_K x = P_K x^0.$$

We wish to remark that in this problem, the norm is placed on the *analysis* coefficients rather than on the *synthesis* coefficients as in [14,17] and other papers on basis pursuit. This is necessary due to the fact that we intend to also apply this optimization problem in the situation when Φ does not form a basis, but merely a frame. To avoid numerical instabilities which are expected to occur since, for each $x \in \mathscr{H}$, the linear system of equations $x = \Phi c$ has infinitely many solutions, only the specific solution $\Phi^* x$ is analyzed. A similar strategy was pursued in [26] and [28]. Various inpainting algorithms which are based on the core idea of (INP) combined with geometric separation are heuristically shown to be successful in [5,13,18].

Interestingly, this minimization problem can be also regarded as a mixed ℓ_1-ℓ_2 problem [27], since the analysis coefficient sequence $\Phi^* x$ is exactly the minimizer of

$$\min\{\|c\|_2 : c \in \ell_2, x = \Phi c\},$$

that is, the coefficient sequence which is minimal in the ℓ_2 norm. The optimization problem in (INP) may also be thought of a relaxation of the *cosparsity* problem

$$x^\star = \operatorname{argmin}_x \|\Phi^* x\|_0 \text{ subject to } P_K x = P_K x^0.$$

Theoretical results concerning cosparsity may be found in [37,38].

We also consider the noisy case. Assume now that we know $\tilde{x} = P_K x^0 + n$, where x^0 and n are unknown, but n is assumed to be small in the sense of $\|\Phi^* n\|_1 \le \varepsilon$ for small ε. Also, clearly $n = P_K n$. Then we solve

$$(\text{INPNOISE}) \quad \tilde{x}^\star = \operatorname{argmin}_x \|\Phi^* x\|_1 \text{ subject to } P_K x = \tilde{x}.$$

To analyze this optimization problem, we require the following notion, which intuitively measures the maximal fraction of the total ℓ_1 norm which can be concentrated to the index set Λ restricted to functions in $\mathscr{H}_M$. In this sense, the geometric relation between the missing part $\mathscr{H}_M$ and expansions in Φ is encoded.

Definition 2 Let Φ be a Parseval frame, and let Λ be a index set of coefficients. We then define the *concentration on* $\mathscr{H}_M$ by

$$\kappa = \kappa(\Lambda, \mathscr{H}_M) = \sup_{f\in\mathscr{H}_M} \frac{\|\mathbb{1}_\Lambda \Phi^* f\|_1}{\|\Phi^* f\|_1}.$$

This notion allows us to formulate our first estimate concerning the ℓ_2 error of the reconstruction via (INP). The reader should notice that the considered error $\|x^\star - x^0\|_2$ is solely measured on $\mathscr{H}_M$, the masked space, since $P_K x^\star = P_K x^0$ due to the constraint in (INP). Another important notion is that of *relative sparsity*.

Definition 3 Fix $\delta > 0$. Given a Hilbert space $\mathscr{H}$ with a Parseval frame Φ, $x \in \mathscr{H}$ is *δ-relatively sparse in Φ (with respect to Λ* if

$$\|\mathbb{1}_{\Lambda^c} \Phi^* x\|_1 \le \delta,$$

where given a space X and a subset $A \subseteq X$, A^c denotes $X \setminus A$.

We now present a pair of lemmas which were first published in [26] without proof.

Lemma 1 *Fix $\delta > 0$ and suppose that x^0 is δ-relatively sparse in Φ. Let $x^\star$ solve (*INP*). Then*

$$\|x^\star - x^0\|_2 \le \frac{2\delta}{1-2\kappa}.$$

The noiseless case Lemma 1 holds as a corollary to the case with noise, which follows.

Lemma 2 *Fix $\delta > 0$ and suppose that x^0 is δ-relatively sparse in Φ. Let $\tilde{x}^\star$ solve (*INPNOISE*). Also assume that the noise satisfies $\|\Phi^* n\|_1 \le \varepsilon$. Then*

$$\|\tilde{x}^\star - x^0\|_2 \le \frac{2\delta + (3+\kappa^2)\varepsilon}{1-2\kappa}.$$

Proof Since Φ is Parseval,

$$\|\tilde{x}^\star - x^0\|_2 \leq \|\Phi^*(\tilde{x}^\star - x^0)\|_1. \tag{3}$$

We invoke the relation $P_K\tilde{x}^\star = P_K x^0 + n$, which implies that $P_K(\tilde{x}^\star - x^0) = n$. Using the definition of κ, we obtain

$$\begin{aligned}\|\mathbb{1}_\Lambda \Phi^*(\tilde{x}^\star - x^0)\|_1 &\leq \|\mathbb{1}_\Lambda \Phi^* P_M(\tilde{x}^\star - x^0)\|_1 + \|\mathbb{1}_\Lambda \Phi^* n\|_1 \\ &\leq \kappa\|\Phi^* P_M(\tilde{x}^\star - x^0)\|_1 + \|\Phi^* n\|_1 \\ &\leq \kappa\|\Phi^*(\tilde{x}^\star - x^0)\|_1 + (1+\kappa)\|\Phi^* n\|_1 \\ &\leq \kappa\|\Phi^*(\tilde{x}^\star - x^0)\|_1 + (1+\kappa)\varepsilon. \end{aligned} \tag{4}$$

It follows that

$$\begin{aligned}&\|\Phi^*(\tilde{x}^\star - x^0)\|_1 = \|\mathbb{1}_\Lambda \Phi^*(\tilde{x}^\star - x^0)\|_1 + \|\mathbb{1}_{\Lambda^c} \Phi^*(\tilde{x}^\star - x^0)\|_1 \\ &\leq \kappa\|\Phi^*(\tilde{x}^\star - x^0)\|_1 + \|\mathbb{1}_{\Lambda^c} \Phi^*(\tilde{x}^\star - x^0)\|_1 + (1+\kappa)\varepsilon.\end{aligned}$$

The relative sparsity of x^0 now implies

$$\begin{aligned}&\|\Phi^*(\tilde{x}^\star - x^0)\|_1 \\ &\leq \frac{1}{1-\kappa}\left(\|\mathbb{1}_{\Lambda^c} \Phi^*(\tilde{x}^\star - x^0)\|_1 + (1+\kappa)\varepsilon\right) \\ &\leq \frac{1}{1-\kappa}\left(\|\mathbb{1}_{\Lambda^c} \Phi^* \tilde{x}^\star\|_1 + \delta + (1+\kappa)\varepsilon\right).\end{aligned} \tag{5}$$

Applying the sparsity of x^0 again and the minimality of $\tilde{x}^\star$, we have

$$\begin{aligned}&\|\mathbb{1}_{\Lambda^c} \Phi^* \tilde{x}^\star\|_1 = \|\Phi^* \tilde{x}^\star\|_1 - \|\mathbb{1}_\Lambda \Phi^* \tilde{x}^\star\|_1 \\ &\leq \|\Phi^*(x^0 + n)\|_1 - \|\mathbb{1}_\Lambda \Phi^* \tilde{x}^\star\|_1 \\ &\leq \|\Phi^* x^0\|_1 - \|\mathbb{1}_\Lambda \Phi^* \tilde{x}^\star\|_1 + \varepsilon \\ &\leq \|\Phi^* x^0\|_1 + \|\mathbb{1}_\Lambda \Phi^*(\tilde{x}^\star - x^0)\|_1 - \|\mathbb{1}_\Lambda \Phi^* x^0\|_1 + \varepsilon \\ &\leq \|\mathbb{1}_\Lambda \Phi^*(\tilde{x}^\star - x^0)\|_1 + \delta + \varepsilon.\end{aligned}$$

Using (4) and (5), this leads to

$$\begin{aligned}&\|\Phi^*(\tilde{x}^\star - x^0)\|_1 \leq \frac{1}{1-\kappa}\left(\|\mathbb{1}_{\Lambda^c} \Phi^* \tilde{x}^\star\|_1 + \delta + (1+\kappa)\varepsilon\right) \\ &\leq \frac{1}{1-\kappa}\left(\|\mathbb{1}_\Lambda \Phi^*(\tilde{x}^\star - x^0)\|_1 + 2\delta\right) + \frac{(2+\kappa)\varepsilon}{1-\kappa} \\ &\leq \frac{1}{1-\kappa}\left(\kappa\|\Phi^*(\tilde{x}^\star - x^0)\|_1 + 2\delta\right) + \frac{(3+2\kappa)\varepsilon}{1-\kappa}.\end{aligned}$$

Combining this with (3), we finally obtain

$$\begin{aligned}\|\tilde{x}^\star - x^0\|_2 &\leq \left(1 - \frac{\kappa}{1-\kappa}\right)^{-1} \frac{2\delta + (3+2\kappa)\varepsilon}{1-\kappa} \\ &= \frac{2\delta + (3+2\kappa)\varepsilon}{1-2\kappa}\end{aligned}$$

□

We now establish a relation between the concentration $\kappa(\Lambda, \mathscr{H}_M)$ on $\mathscr{H}_M$ and the notion of cluster coherence μ_c first introduced in [16]. For this, by abusing notation, we will write $P_M\Phi$ and $P_K\Phi$ for the projected frame elements.

To first introduce the notion of cluster coherence, recall that in many studies of ℓ_1 optimization, one utilizes the *mutual coherence*

$$\mu(\Phi_1, \Phi_2) = \max_j \max_i |\langle \varphi_{1i}, \varphi_{2j}\rangle|,$$

whose importance was shown by [15]. This may be called the *singleton coherence.* We modify the definition to take into account clustering of the coefficients arising from the geometry of the situation.

Definition 4 Let $\Phi_1 = \{\varphi_{1i}\}_{i\in I}$ and $\Phi_2 = \{\varphi_{2j}\}_{j\in J}$ lie in a Hilbert space $\mathscr{H}$ and let $\Lambda \subseteq I$. Then the *cluster coherence* $\mu_c(\Lambda, \Phi_1; \Phi_2)$ of Φ_1 and Φ_2 with respect to Λ is defined by

$$\mu_c(\Lambda, \Phi_1; \Phi_2) = \max_{j\in J} \sum_{i\in\Lambda} |\langle \varphi_{1i}, \varphi_{2j}\rangle|.$$

The following relation is a specific case of Proposition 3 in [26]. We include a proof for completeness.

Lemma 3 *We have*

$$\kappa(\Lambda, \mathscr{H}_M) \leq \mu_c(\Lambda, P_M\Phi; P_M\Phi) = \mu_c(\Lambda, P_M\Phi; \Phi).$$

Proof For each $f \in \mathscr{H}_M$, we choose a coefficient sequence α such that $f = \Phi\alpha$ and $\|\alpha\|_1 \leq \|\beta\|_1$ for all β satisfying $f = \Phi\beta$. Invoking the fact that Φ is a tight frame, hence $f = \Phi\Phi^T\Phi\alpha$, and the fact that $f = (P_M\Phi)\alpha$, we obtain

$$\begin{aligned}\|1_\Lambda \Phi^T f\|_1 &= \|1_\Lambda (P_M\Phi)^T f\|_1 \\ &= \|1_\Lambda (P_M\Phi)^T (P_M\Phi)\alpha\|_1 \\ &\leq \sum_{i\in\Lambda}\left(\sum_j |\langle P_M\phi_i, P_M\phi_j\rangle||\alpha_j|\right) \\ &= \sum_j\left(\sum_{i\in\Lambda} |\langle P_M\phi_i, P_M\phi_j\rangle|\right)|\alpha_j| \\ &\leq \mu_c(\Lambda, P_M\Phi; P_M\Phi)\|\alpha\|_1 \\ &\leq \mu_c(\Lambda, P_M\Phi; P_M\Phi)\|\Phi^T\Phi\alpha\|_1 \\ &= \mu_c(\Lambda, P_M\Phi; P_M\Phi)\|\Phi^T f\|_1.\end{aligned}$$

□

Combining Lemmata 1 and 3 proves the final noiseless estimate and combining Lemmata 2 and 3 proves the final estimate with noise:

Proposition 1 *Fix $\delta > 0$ and suppose that x^0 is δ-relatively sparse in Φ. Let $x^\star$ solve (*INP*). Then*

$$\|x^\star - x^0\|_2 \leq \frac{2\delta}{1 - 2\mu_c(\Lambda, P_M\Phi; \Phi)}.$$

Proposition 2 *Fix $\delta > 0$ and suppose that x^0 is δ-relatively sparse in Φ. Let $\tilde{x}^\star$ solve (*INPNOISE*). Also assume that the noise satisfies $\|\Phi^* n\|_1 \leq \varepsilon$. Then*

$$\|\tilde{x}^\star - x^0\|_2 \leq \frac{2\delta + (3+2\kappa)\varepsilon}{1 - 2\mu_c(\Lambda, P_M\Phi; \Phi)}.$$

Let us briefly interpret this estimate, first focusing on the noiseless case. As expected the error decreases linearly with the relative sparsity. It should also be emphasized that both relative sparsity and cluster coherence depend on the

chosen "geometric set of indices" Λ. Thus this set is crucial for determining whether Φ is a good dictionary for inpainting. This will be illustrated in the sequel when considering a particular situation; however, Λ is merely an analysis tool and explicit knowledge of it is not necessary to recover data. Note that in general, the larger the set Λ is, the smaller $\|\mathbb{1}_{\Lambda^c}\Phi^*x^0\|_1$ is (i.e., x^0 is δ-relatively sparse for a smaller δ) and the larger the cluster coherence is. This seems to be a contradiction, but if Φ sparsifies x^0 well, then a small set Λ can be chosen which keeps $\|\mathbb{1}_{\Lambda^c}\Phi^*x^0\|_1$ small. Finally, considering the noisy case, as also expected the error estimate depends linearly on the ℓ_2 bound for the noise.

2.3 Inpainting via Thresholding

Another fundamental methodology from compressed sensing for sparse recovery is thresholding. The beauty of this approach lies in its simplicity and its associated fast algorithms. Typically, it is also possible to prove success of recovery in similar situations as in which ℓ_1 minimization succeeds.

Various thresholding strategies are available such as iterative thresholding, etc. It is thus surprising that the most simple imaginable strategy, which is to perform just *one step* of hard thresholding, already allows for error estimates as strong of for ℓ_1 minimization. We start by presenting this thresholding strategy. For technical reasons, – note also that this is no restriction at all – we now assume that the Parseval frame $\Phi = (\phi_i)_i$ consists of frame vectors with equal norm, i.e., $\|\phi_i\| = c$ for all i.

ONE-STEP-THRESHOLDING

Parameters:

- Incomplete signal $\tilde{x} = P_K x^0$ (noiseless) or $P_K x^0 + n$ (with noise).
- Thresholding parameter β.

Algorithm:

1) *Threshold Coefficients with Respect to Frame Φ:*
   a) Compute $\langle \tilde{x}, \phi_i \rangle$ for all i.
   b) Apply threshold and set $\mathscr{T} = \{i : |\langle \tilde{x}, \phi_i \rangle| \geq \beta\}$.
2) *Reconstruct Original Signal:*
   a) Compute $x^\star = \Phi \mathbb{1}_{\mathscr{T}} \Phi^* \tilde{x}$.

Output:

- Significant thresholding coefficients: $\mathscr{T}$.
- Approximation to x^0: $x^\star$.

Fig. 7 ONE-STEP-THRESHOLDING Algorithm to reconstruct x^0 from noiseless $P_K x^0$ or noisy $P_K x^0 + n$.

The following result provides us with an estimate for the ℓ_2 error of the synthesized signal $x^\star$ computed via ONE-STEP-THRESHOLDING.

Proposition 3 *Let $\mathscr{T}$ and $x^\star$ be computed via the algorithm* ONE-STEP-THRESHOLDING *(Figure 7) for noiseless data, and for $\delta > 0$ assume that x^0 is relatively sparse in Φ with respect to $\mathscr{T}$. Then*

$$\|x^\star - x^0\|_2 \leq c\left[\delta + \|\mathbb{1}_{\mathscr{T}}\Phi^* P_M x^0\|_1\right].$$

As before, Proposition 3 follows as a corollary to the case with noise:

Proposition 4 *Let $\mathscr{T}$ and $x^\star$ be computed via the algorithm* ONE-STEP-THRESHOLDING *for data with noise, and for $\delta > 0$ assume that x^0 is relatively sparse in Φ with respect to $\mathscr{T}$. Also assume that the noise satisfies $\|\Phi^* n\|_1 \leq \varepsilon$. Then*

$$\|x^\star - x^0\|_2 \leq c\left(\|\mathbb{1}_{\mathscr{T}}\Phi^* P_M x^0\|_1 + \delta + \varepsilon\right).$$

Proof Invoking the decomposition of $\mathscr{H}$ and the fact that Φ is Parseval,

$$\begin{aligned}\|x^\star - x^0\|_2 &= \|\Phi\mathbb{1}_{\mathscr{T}}\Phi^*(P_K x^0 + n) - \Phi\Phi^* P_K x^0 - P_M x^0\|_2 \\ &= \|\Phi\mathbb{1}_{\mathscr{T}^c}\Phi^* P_K x^0 + \Phi\mathbb{1}_{\mathscr{T}}\Phi^* n - P_M x^0\|_2.\end{aligned}$$

Since

$$P_M x^0 = \Phi\mathbb{1}_{\mathscr{T}}\Phi^* P_M x^0 + \Phi\mathbb{1}_{\mathscr{T}^c}\Phi^* P_M x^0$$

and $P_K x^0 + P_M x^0 = x^0$, it follows that

$$\begin{aligned}\|x^\star - x^0\|_2 \leq\ & \|\Phi\mathbb{1}_{\mathscr{T}^c}\Phi^* x^0\|_2 + \\ & + \|\Phi\mathbb{1}_{\mathscr{T}}\Phi^* P_M x^0\|_2 + \|\Phi\mathbb{1}_{\mathscr{T}}\Phi^* n\|_2.\end{aligned}$$

It follows from the equal-norm condition on the frame Φ that for any ℓ_1 sequence x,

$$\|\Phi x\|_2 \leq c\|x\|_1.$$

Applying the relative sparsity of x^0 we obtain

$$\|x^\star - x^0\|_2 \leq c\left(\|\mathbb{1}_{\mathscr{T}}\Phi^* P_M x^0\|_1 + \delta + \varepsilon\right),$$

which is what we intended to prove.

□

As before, let us also interpret this estimate. Now the situation is slightly different from the estimate for the ℓ_1 approach. Again, the estimate depends linearly on the relative sparsity and the noise. The difference now is the appearance of the term $\|\mathbb{1}_{\mathscr{T}}\Phi^* P_M x^0\|_1$ in the numerator instead of the cluster coherence in the denominator. Note, however, that

$$\begin{aligned}\|\mathbb{1}_{\mathscr{T}}\Phi^* P_M x^0\|_1 &\leq \kappa\|\Phi^* P_M x^0\|_1 \\ &\leq \mu_c(\mathscr{T}, P_M\Phi; \Phi)\|\Phi^* P_M x^0\|_1.\end{aligned}$$

Thus both in the ℓ_1 minimization case Proposition 1 and in the thresholding case Proposition 3, the bound on the error is lower when the cluster coherence is lower. Furthermore, $\|\Phi^* P_M x^0\|_1$ is a quantification of how much of the signal is missing, which clearly can not be too high.

3 Mathematical Model

We next provide a specific mathematical model which is motivated by the fact that images are typically governed by edges, which can most prominently be seen in, for example, seismic imaging (Figure 8). Following this line of thought,

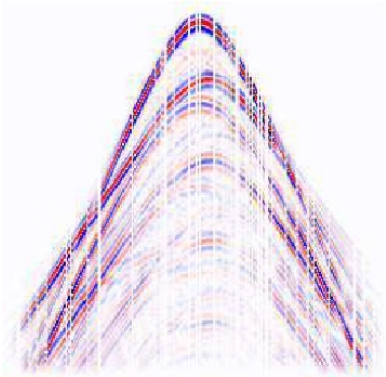

Fig. 8 Incomplete seismic data [24]

our model is based on line singularities – which can as explained later be extended to curvilinear singularities – with missing data of the forms as gaps or holes.

In this section, such a model for the original image and the mask will be introduced. Since the analysis we derive later is based on the behavior in Fourier domain, the Fourier content of the models is another focus.

3.1 Image Model

Let $w : \mathbf{R} \mapsto [0,1]$ be a smooth function that is supported in $[-\rho,\rho]$, where we always assume that ρ is sufficiently large, in particular, much larger than h (a measure of the missing data which will be defined in the next subsection). For now, we consider as a prototype of a line singularity the weighted distribution $w\mathscr{L}$ acting on "nice" functions by

$$\langle w\mathscr{L}, f\rangle = \int_{-\rho}^{\rho} w(x_1) f(x_1,0) dx_1.$$

Notice that this distribution is supported on the segment

$$[-\rho,\rho]\times\{0\}$$

of the x-axis, hence can be employed as a model for a horizontal linear singularity. The weighting was chosen to ensure that we are dealing with an L_2-function after filtering. The Fourier transform of $w\mathscr{L}$ can be computed to be

$$\langle \widehat{w\mathscr{L}}, f\rangle = \langle w\mathscr{L}, \hat{f}\rangle = \int_{\mathbf{R}} \hat{w}(\xi_1) \int_{\mathbf{R}} f(\xi_1,\xi_2) d\xi_2 d\xi_1,$$

where we use the following Fourier transform definition for $f \in L^1(\mathbf{R}^n)$

$$\mathscr{F} f := \hat{f} = \int_{\mathbf{R}^n} f(x) e^{-2\pi i\langle x,\cdot\rangle} dx,$$

(where $\langle\cdot,\cdot\rangle$ is the standard Euclidean inner product) which can be naturally extended to functions in $L^2(\mathbf{R}^n)$. The inverse Fourier transform is given by

$$\mathscr{F}^{-1} f := \check{f} = \int_{\mathbf{R}^n} f(\xi) e^{2\pi i\langle\cdot,\xi\rangle} d\xi.$$

Let now $\check{F}_j$ be a filter corresponding to the frequency tiling of 2D Meyer wavelets or shearlets at level j defined by its Fourier transform F_j, which is defined to be

$$F_j = \sum_{\iota\in\{h,v,d\}} W^\iota(2^{-j}\xi)$$

We consider the filtered version of $w\mathscr{L}$ which we denote by $w\mathscr{L}_j$, i.e.,

$$w\mathscr{L}_j = w\mathscr{L} \star \check{F}_j = \int_{\mathbf{R}^2} w\mathscr{L}(\cdot - t)\check{F}_j(t) dt.$$

The next result provides us with an estimate of the norm of $w\mathscr{L}_j$.

Lemma 4 *For some $c > 0$,*

$$\|w\mathscr{L}_j\|_2 \geq c 2^{j/2}, \qquad j \to \infty.$$

Proof We have

$$\begin{aligned}\|w\mathscr{L}_j\|_2 &\geq \left(\int_{\xi_1\in\mathbf{R}} |\hat{w}(\xi_1)|^2 d\xi_1 \int_{|\xi_2|\in[2^j c_0, 2^{j+1}c_0]} d\xi_2\right)^{1/2} \\ &\approx c 2^{j/2}.\end{aligned}$$

□

3.2 Masks

Inspired by the missing sensor scenario in seismic data we will define the mask of a missing piece of the image as follows. The mask $\mathscr{M}_h$ is a vertical strip of diameter $2h$ and is given by

$$\mathscr{M}_h = \mathbb{1}_{\{|x_1|\leq h\}}.$$

For an illustration, we refer to Figure 9.

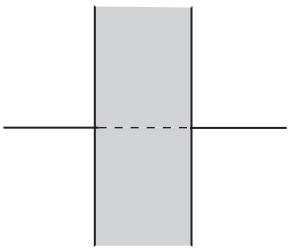

Fig. 9 Mask $\mathscr{M}_h$, together with the linear singularity $w\mathscr{L}$.

For later use, we compute the associated Fourier transforms, where as usual we set $\mathrm{sinc}(y) = \sin(\pi y)/(\pi y)$ for $y \in \mathbf{R}$.

Lemma 5 *We have*

$$\hat{\mathscr{M}}_h = 2h\mathrm{sinc}(2h\xi_1)\hat{\mathscr{L}}_y,$$

where $\mathscr{L}_y$ is the distribution acting as

$$\langle \mathscr{L}_y, f\rangle = \int f(0,y)dy$$

and $\langle \hat{\mathscr{L}}_y, f\rangle = \int f(x,0)dx$.

Proof Define the planar Heaviside by $H(x) = \mathbb{1}_{\{x_1\geq 0\}}$. Since $\mathscr{L}_y = \frac{\partial}{\partial x_1}H$, we have $\hat{H}(\xi) = (2\pi i\xi_1)^{-1}\hat{\mathscr{L}}_y$. We now express $\mathscr{M}_h$ in terms of H by

$$\mathscr{M}_h = H(x+(h,0)) - H(x-(h,0)).$$

This leads to

$$\begin{aligned}\hat{\mathscr{M}}_h &= (e^{2\pi ih\xi_1} - e^{-2\pi ih\xi_1})(2\pi i\xi_1)^{-1}\hat{\mathscr{L}}_y \\ &= 2\sin(2\pi h\xi_1)/(2\pi\xi_1)\hat{\mathscr{L}}_y = 2h\mathrm{sinc}(2h\xi_1)\hat{\mathscr{L}}_y.\end{aligned}$$

The proof is finished. □

3.3 Transfer of Abstract Setting

All of the main proofs in Sections 4 and 5 will follow a particular pattern. Either Proposition 1 (in the case of ℓ_1 minimization) or Proposition 3 (in the case of thresholding) is applied to the situation in which x^0 is chosen to be the filtered linear singularity $w\mathscr{L}_j$, the Hilbert space $\mathscr{H}_M$ is defined by $\{f\mathscr{M}_h : f \in L^2(\mathbf{R}^2)\}$, and Φ is either the system of Meyer orthonormal wavelets or of shearlets at scale j.

In the analysis that follows, δ_j will denote the optimal δ-sparsity for filtered coefficients. That is, for ℓ_1 minimization with a fixed filter level j, we will fix a set Λ_j of significant coefficients of $\Phi = \{\psi_\lambda\}$ and set

$$\delta_j = \sum_{\lambda\in\Lambda_j^c} |\langle w\mathscr{L}_j, \psi_\lambda\rangle|.$$

Similarly, we will analyze thresholding schemes by setting

$$\delta_j = \sum_{\lambda\in\mathscr{T}_j^c} |\langle w\mathscr{L}_j, \psi_\lambda\rangle|.$$

The inpainting accomplished (i.e., the solution in Proposition 1 or Proposition 3) on the filtered levels j will be denoted L_j. $w\mathscr{L}_j$ will denote the filtered real image; that is, $w\mathscr{L}\star\check{F}_j$, where $\mathscr{L}$ is the original, complete image. Thus, the main theorems in Sections 4 and 5 will show that

$$\frac{\|L_j - w\mathscr{L}_j\|_2}{\|w\mathscr{L}_j\|_2} \to 0, \qquad j\to\infty.$$

The results will specifically depend on the asymptotic behavior of the gap h_j.

4 Positive Results for Wavelet Inpainting

We begin by proving theoretically for the first time what has been known heuristically; namely, that wavelets can successfully inpaint an edge as long as too much is not missing. In Subsection 4.1, we investigate the inpainting results of ℓ_1 minimization by estimating the δ-relative sparsity δ_j and cluster coherence μ_c with respect to $\Phi = \{\psi_\lambda : \lambda = (\iota, j, k), \iota = h, v, d; k \in \mathbf{Z}^2\}$ and a proper chosen index set Λ_j. In Subsection 4.2, we similarly give the estimation of δ_j and μ_c for inpainting using thresholding.

4.1 ℓ_1 Minimization

In what follows, we use the compact notation $\langle a\rangle := (1+|a|^2)^{1/2}$. We first need to choose the set of significant coefficients appropriately. We do this by setting

$$\Lambda_j = \{(\iota; j, k) : |k_1| \leq \rho n_j 2^j, |k_2| \leq n_j, \iota = h, v, d\},$$

where $n_j = 2^{\varepsilon j}$. This choice of Λ_j forces the relative sparsity to grow slower than the growth rate of $\|w\mathscr{L}_j\|_2$:

Lemma 6 $\delta_j = o(1) = o(\|w\mathscr{L}_j\|_2), \quad j\to\infty.$

Proof By definition, we have

$$\delta_j = \sum_{\lambda\in\Lambda_j^c} |\langle w\mathscr{L}_j, \psi_\lambda\rangle| = \sum_{\lambda\in\Lambda_j^c} |\langle \widehat{w\mathscr{L}_j}, \hat{\psi}_\lambda\rangle|;$$

that is,

$$\begin{aligned}\delta_j &= \sum_{\lambda\in\Lambda_j^c} \left| \int_{\mathbf{R}^2} 2^{-j}\hat{w}(\xi_1)\hat{F}_j(\xi)W^\iota(\xi/2^j)e^{-2\pi i\langle k,\xi/2^j\rangle}d\xi \right| \\ &:= \sum_{\lambda\in\Lambda_j^c} \left| \int_{\mathbf{R}^2} \hat{G}_j(\xi)e^{-2\pi i\langle k,\xi/2^j\rangle}d\xi \right|,\end{aligned}$$

where $\hat{G}_j(\xi) = 2^{-j}\hat{w}(\xi_1)\hat{F}_j(\xi)W^\iota(\xi/2^j)$ is a smooth and compactly supported function that is essentially supported on

$$[-1/\rho, 1/\rho]\times[-2^j c_0, 2^j c_0].$$

Applying the change of variable $(\xi_1,\xi_2)\mapsto(\rho^{-1}\xi_1, 2^j\xi_2)$ ensures that $\hat{G}_j(\rho^{-1}\xi_1, 2^j\xi_2)$ is smooth and compactly supported independent of j. Then

$$\begin{aligned}&\left| \int_{\mathbf{R}^2} \hat{G}_j(\xi)e^{-2\pi i\langle k,\xi/2^j\rangle}d\xi \right| \\ &\leq \tilde{c}_N\|\hat{G}_j\|_\infty(\rho^{-1}2^j)\langle|(\rho^{-1}k_1/2^j, k_2)|\rangle^{-N} \\ &\leq c_N(\rho^{-1}2^j)\langle|(\rho^{-1}k_1/2^j, k_2)|\rangle^{-N}.\end{aligned}$$

Consequently, δ_j/c_N is bounded above by

$$\rho^{-1}2^j \sum_{\lambda\in\Lambda_j^c} \langle|(\rho^{-1}k_1/2^j, k_2)|\rangle^{-N}$$

$$\leq \rho^{-1}2^j\Bigg(\sum_{|k_1|\geq\rho n_j 2^j,k_2}\langle|(\frac{\rho^{-1}k_1}{2^j},k_2)|\rangle^{-N}+$$
$$+\sum_{|k_1|\leq\rho n_j 2^j,|k_2|\geq n_j}\langle|(\frac{\rho^{-1}k_1}{2^j},k_2)|\rangle^{-N}\Bigg)$$
$$\leq \rho^{-1}2^j\Bigg(\int_{\rho n_j 2^j}^{\infty}\int_{\mathbf{R}}\langle|(\frac{\rho^{-1}x_1}{2^j},x_2)|\rangle^{-N}dx_2dx_1+$$
$$+\int_0^{\rho n_j 2^j}\int_{n_j}^{\infty}\langle|(\frac{\rho^{-1}x_1}{2^j},x_2)|\rangle^{-N}dx_2dx_1\Bigg)$$
$$\leq \int_{n_j}^{\infty}\int_{\mathbf{R}}\langle|(x_1,x_2)|\rangle^{-N}dx_2dx_1$$
$$+\int_0^{n_j}\int_{n_j}^{\infty}\langle|(x_1,x_2)|\rangle^{-N}dx_2dx_1$$
$$\leq 2^{j\varepsilon(1-N)}.$$

Thus for N large enough, $\delta_j\to 0$ as $j\to\infty$.

□

On the other hand, the choice of Λ_j offers low cluster coherence as well:

Lemma 7 *For $h_j = o(2^{-j})$ as $j\to\infty$, we have*

$$\mu_c(\Lambda_j,\{\mathscr{M}_{h_j}\psi_\lambda\};\{\psi_\lambda\})\to 0,\quad j\to\infty.$$

Proof By definition, we have

$$\mu_c(\Lambda_j,\{\mathscr{M}_{h_j}\psi_\lambda\};\{\psi_\lambda\}) = \max_{\lambda'}\sum_{\lambda\in\Lambda_j}\left|\langle\mathscr{M}_{h_j}\psi_\lambda,\psi_{\lambda'}\rangle\right|$$
$$= \max_{\lambda'}\sum_{\lambda\in\Lambda_j}\left|\langle\hat{\mathscr{M}}_{h_j}\star\hat{\psi}_\lambda,\hat{\psi}_{\lambda'}\rangle\right|.$$

Note that for $\lambda=(\iota,j,k)$, we can choose $\lambda'=(\iota',j,0)$.

$$\langle\hat{\mathscr{M}}_{h_j}\star\hat{\psi}_\lambda,\hat{\psi}_{\lambda'}\rangle = \int_{\mathbf{R}^2}\int_{\mathbf{R}^2}\hat{\mathscr{M}}_{h_j}(\xi)\hat{\psi}_\lambda(\tau-\xi)d\xi\,\overline{\hat{\psi}_{\lambda'}(\tau)}d\tau$$
$$= \int_{\mathbf{R}^2}\int_{\mathbf{R}}2h_j\mathrm{sinc}(2h_j\xi_1)\hat{\psi}_\lambda(\tau-(\xi_1,0))d\xi_1\overline{\hat{\psi}_{\lambda'}(\tau)}d\tau$$
$$= \int_{\mathbf{R}}2h_j\mathrm{sinc}(2h_j\xi_1)\Bigg[\int_{\mathbf{R}^2}2^{-2j}W^\iota(\frac{\tau-(\xi_1,0)}{2^j})\times$$
$$\times e^{-2\pi i\langle k,\frac{\tau-(\xi_1,0)}{2^j}\rangle}W^{\iota'}(\frac{\tau}{2^j})d\tau\Bigg]d\xi_1$$
$$= 2^j2h_j\int_{\mathbf{R}^2}\Bigg[\int_{\mathbf{R}}\mathrm{sinc}(2^j2h_j\xi_1)W^\iota((\tau-(\xi_1,0)))\times$$
$$\times e^{2\pi ik_1\xi_1}d\xi_1W^{\iota'}(\tau)\Bigg]e^{-2\pi i\langle k,\tau\rangle}d\tau$$
$$=: 2^j2h_j\int_{\mathbf{R}^2}\hat{g}_j(\tau)e^{-2\pi i\langle k,\tau\rangle}d\tau,$$

where

$$\hat{g}_j(\tau) \tag{6}$$
$$:= W^{\iota'}(\tau)\int_{\mathbf{R}}\mathrm{sinc}(2^j2h_j\xi_1)W^\iota((\tau-(\xi_1,0)))e^{2\pi i\langle k_1,\xi_1\rangle}d\xi_1$$

is a smooth function supported on a box independent of j. Hence, $\left|\int\hat{g}_j(\tau)e^{-2\pi ik\tau}d\tau\right|\leq c_N\|\hat{g}_j\|_\infty\langle|k|\rangle^{-N}$, and

$$\|\hat{g}_j\|_\infty \leq c\sup_\tau\int|\mathrm{sinc}(2^j2h_j\xi_1)||W^\iota(\tau-(\xi_1,0))|d\xi_1$$
$$\leq c\|\mathrm{sinc}(2^j2h_j\cdot)\|_2\leq c(2^jh)^{-1/2}.$$

Consequently, by our assumption, we have

$$\mu_c(\Lambda_j,\{\mathscr{M}_{h_j}\psi_\lambda\};\{\psi_\lambda\})\leq c_N2^jh_j(2^jh_j)^{-1/2}\sum_{k\in\mathbf{Z}^2}\langle|k|\rangle^{-N}$$
$$\leq c_N(2^jh_j)^{1/2}\to 0,\quad j\to\infty.$$

□

We would like to remark at this point that we do not need the strong condition that $h_j=o(2^{-j})$ as $j\to\infty$. In fact, carefully handling the constants in the proof of Lemma 7 will lead us to the condition

$$\mu_c(\Lambda_j,\{\mathscr{M}_{h_j}\psi_\lambda\};\{\psi_\lambda\})\leq c_N(2^jh_j)$$

with precise knowledge of the value of c_N. Since ultimately, we "only" need the cluster coherence to boundedly stay away from $1/2$, we only require the weaker condition of

$$2^jh_j\leq\frac{1}{2c_N}-\varepsilon\quad\text{for some }\varepsilon>0\text{ and for all }j\geq j_0.$$

This condition would then be also sufficient for deriving the following theorem.

We now apply Proposition 1 to Lemmata 4, 6, and 7 to obtain the desired convergence for the normalized ℓ_2 error of the reconstruction L_j derived from (1), where in this case $L=w\mathscr{L}_j$ and Φ are wavelets ψ_λ at scale j.

Theorem 1 *For $h_j=o(2^{-j})$ and L_j the solution to (1) with Φ the 2D Meyer orthonormal system,*

$$\frac{\|L_j-w\mathscr{L}_j\|_2}{\|w\mathscr{L}_j\|_2}\to 0,\qquad j\to\infty.$$

This result shows that if the size of the gap shrinks faster than 2^{-j} – i.e., the size of the gap is asymptotically smaller than 2^{-j} – or if the gap shrinks at the same rate than 2^{-j} with an exactly prescribed factor, we have asymptotically perfect inpainting.

4.2 Thresholding

We will now study thresholding as an inpainting method, which is from a computational point of view much easier to apply. Our analysis will show that we can derive the same asymptotic performance as for ℓ_1 minimization.

Our first claim concerns the set of the thresholding coefficients $\mathscr{T}_j$.

Lemma 8 *For $h_j=o(2^{-j})$ as $j\to\infty$, there exist thresholds $\{\beta_j\}_j$ such that, for all $j\geq j_0$,*

$$\{k:|k_1|\leq\rho 2^{j(1+n_1)},|k_2|\leq 2^{jn_1}\}\subseteq\mathscr{T}_j$$

for positive j_0 and n_1.

Proof By Plancherel, we can rewrite the coefficients which we have to threshold as follows:

$$|\langle (1-\mathscr{M}_{h_j}) w\mathscr{L}_j, \psi_\lambda\rangle|$$
$$= |\langle \delta_0 \star \widehat{w\mathscr{L}}_j, \hat{\psi}_\lambda\rangle - \langle \widehat{\mathscr{M}}_{h_j} \star \widehat{w\mathscr{L}}_j, \hat{\psi}_\lambda\rangle|.$$

Choose a function F such that

$$\widehat{w\mathscr{L}_j}(\xi) = \widehat{w\mathscr{L}}(\xi)F_j(\xi) = \widehat{w\mathscr{L}}(\xi)F(\xi/2^j).$$

As we are analyzing a horizontal line singularity, we only need to consider

$$\hat{\psi}_\lambda = 2^{-j} W^v(\xi/2^j) e^{-2\pi i\langle k,\xi/2^j\rangle}$$

for large wavelet coefficients. Then, the first term equals

$$\langle \delta_0 \star \widehat{w\mathscr{L}}_j, \hat{\psi}_\lambda\rangle$$
$$= 2^{-j}\int \hat{w}(\xi_1)\int (FW^v)(\xi/2^j) e^{-2\pi i\langle k/2^j,\xi\rangle} d\xi$$
$$= \int\left[\int \hat{w}(\xi_1)(FW^v)((\xi_1/2^j,\xi_2)) e^{-2\pi i\langle k_1/2^j,\xi_1\rangle} d\xi_1\right]\times$$
$$\times e^{-2\pi i\langle k_2,\xi_2\rangle} d\xi_2.$$

By using Lemma 5, we derive for the second term:

$$\langle \widehat{\mathscr{M}}_{h_j} \star \widehat{w\mathscr{L}}_j, \hat{\psi}_\lambda\rangle$$
$$= 2h_j \int \mathrm{sinc}(2h_j\tau_1)\times$$
$$\times \int \hat{w}(\xi_1)(\hat{\psi}_\lambda F_j)((\tau_1,0)+(\xi_1,\xi_2)) d\xi d\tau_1$$
$$= 2h_j 2^{-j}\int \mathrm{sinc}(2h_j\tau_1)\int \hat{w}(\xi_1)F(\xi_1/2^j,\xi_2/2^j)\times$$
$$\times W^v((\tau_1+\xi_1)/2^j,\xi_2/2^j) e^{-2\pi i\langle k/2^j,\tau_1+\xi_1,\xi_2\rangle} d\xi d\tau_1$$
$$= 2h_j\int\left[\int \hat{w}(\xi_1)\int \mathrm{sinc}((h_j/\pi)\tau_1)F(\xi_1/2^j,\xi_2)\times\right.$$
$$\left.\times W^v(((\tau_1+\xi_1)/2^j,\xi_2)) e^{-2\pi i\langle k_1,(\tau_1+\xi_1)/2^j\rangle} d\tau_1 d\xi_1\right]\times$$
$$\times e^{-2\pi i\langle k_2,\xi_2\rangle} d\xi_2.$$

Let $\hat{G}$ now be the function

$$\hat{G}(\xi_2) = \int \hat{w}(\xi_1)\left[(FW^v)((\xi_1/2^j,\xi_2))+\right.$$
$$-2h_j\int \mathrm{sinc}((h_j/\pi)\tau_1)F(\xi_1/2^j,\xi_2)\times$$
$$\left.\times W^v((\tau_1+\xi_1)/2^j,\xi_2) e^{-2\pi i(k_1/2^j)\tau_1} d\tau_1\right]\times$$
$$\times e^{-2\pi i(k_1/2^j)\xi_1} d\xi_1$$
$$= \int \hat{w}(\xi_1)\hat{H}_{\xi_2}(\xi_1) e^{-2\pi i(k_1/2^j)\xi_1} d\xi_1$$

with

$$\hat{H}_{\xi_2}(\xi_1) = (FW^v)((\xi_1/2^j,\xi_2)|)+$$
$$-2h_j\int \mathrm{sinc}((h_j/\pi)\tau_1)F(\xi_1/2^j,\xi_2)\times$$
$$\times W^v((\tau_1+\xi_1)/2^j,\xi_2) e^{-2\pi i\langle (k_1/2^j),\tau_1\rangle} d\tau_1.$$

The function $\hat{G}$ is supported on the set $[1/2,2]$, which is independent of j. By standard arguments, we can deduce that

$$|\langle (1-\mathscr{M}_{h_j}) w\mathscr{L}_j, \psi_\lambda\rangle| \le c_{N_1}\|\hat{G}\|_\infty \langle |k_2|\rangle^{-N_1}. \tag{7}$$

Let us now investigate the term $\|\hat{G}\|_\infty$ further. Using Plancherel and the support properties of w,

$$\left|\int \hat{w}(\xi_1)\hat{H}_{\xi_2}(\xi_1) e^{-2\pi i\langle k_1/2^j,\xi_1\rangle} d\xi_1\right|$$
$$= |(\hat{w}\hat{H}_{\xi_2})^\vee(-k_1/2^j)| = |(w\star H_{\xi_2})(-k_1/2^j)|$$
$$= \left|\int w(-k_1/2^j - x) H_{\xi_2}(x) dx\right| \approx c\left|\int_{-k_1/2^j-\rho}^{-k_1/2^j+\rho} H_{\xi_2}(x) dx\right|.$$

For the analysis of the function H_{ξ_2}, we use well-known properties of the Fourier transform to derive

$$H_{\xi_2}(x) = \left((FW^v)(\cdot/2^j,\xi_2)\right)^\vee(x)+$$
$$-\left(\left(2h_j \mathrm{sinc}(2h_j\cdot) e^{-2\pi i(k_1/2^j)\cdot}\right)\star\left((FW^v)(\cdot/2^j,\xi_2)\right)\right)^\vee(-x)$$
$$= \left((FW^v)(\cdot/2^j,\xi_2)\right)^\vee(x)+$$
$$-\left(2h_j\mathrm{sinc}((2h_j\cdot)) e^{-2\pi i(k_1/2^j)\cdot}\right)^\vee(-x)\times$$
$$\times\left((FW^v)(\cdot/2^j,\xi_2)\right)^\vee(-x)$$
$$= \left((FW^v)(\cdot/2^j,\xi_2)\right)^\vee(x)+$$
$$-\mathbb{1}_{[-h_j,h_j]}(x-k_1/2^j)\left((FW^v)(\cdot/2^j,\xi_2)\right)^\vee(-x).$$

Hence, since $h_j<\rho$,

$$c\left|\int_{-k_1/2^j-\rho}^{-k_1/2^j+\rho} H_{\xi_2}(x)dx\right|$$
$$= c\left|\int_{k_1/2^j-\rho}^{k_1/2^j+\rho}((FW^v)(\cdot/2^j,\xi_2))^\vee(x)+\right.$$
$$\left.-\int_{k_1/2^j-h_j}^{k_1/2^j+h_j}((FW^v)(\cdot/2^j,\xi_2)^\vee(x)dx\right| \tag{8}$$
$$= c\left|\int_{k_1-2^jh_j}^{k_1-2^j\rho}+\int_{k_1+2^j\rho}^{k_1+2^jh_j}((FW^v)(|(\cdot,\xi_2)|))^\vee(x)dx\right|. \tag{9}$$

Notice that the bounds of integration indeed make sense, since the values of k_1 which lie "in between h_j and ρ" should play an essential role. For sufficiently "nice" W, there exist some N_2 and c (possibly differing from the one before, but we do not need to distinguish constants here) such that

$$|((FW^v)(|(\cdot,\xi_2)|)^\vee(x)| \le c\langle |x|\rangle^{-N_2},$$

and hence by (9),

$$\|\hat{G}\|_\infty \le c\langle \min\{|k_1-2^j\rho|,|k_1+2^j\rho|\}\rangle^{-N_2}. \tag{10}$$

Finally, we have to study how the function $\hat{H}$ relates to h, which will show the behavior of the coefficients as they approach the center of the mask. For this, setting

$$\hat{J}_{\xi_2}(\tau_1) = (FW^v)((\tau_1+\xi_1)/2^j,\xi_2) e^{-2\pi i\langle k_1/2^j,\tau_1\rangle},$$

we obtain

$$
\begin{aligned}
&|(FW^v)(\xi_1/2^j,\xi_2)+\\
&\quad -2h_j\int \mathrm{sinc}((h_j/\pi)\tau_1)(FW^v)((\tau_1+\xi_1)/2^j,\xi_2)\times\\
&\quad \times e^{-2\pi i\langle k_1/2^j,\tau_1\rangle}d\tau_1|\\
&=|\hat{f}_{\xi_2}(0)-2h_j\int \mathrm{sinc}((h_j/\pi)\tau_1)\hat{f}_{\xi_2}(\tau_1)d\tau_1|\\
&=|\hat{f}_{\xi_2}(0)-\int \hat{1}_{[-h_j,h_j]}(\tau_1)\hat{f}_{\xi_2}(\tau_1)d\tau_1|\\
&=|\hat{f}_{\xi_2}(0)-\int_{-h_j}^{h_j} J_{\xi_2}(x)dx|\\
&=|\int_{|x|>h_j} J_{\xi_2}(x)dx|.
\end{aligned}
$$

Hence another way to estimate $\|\hat{G}\|_\infty$ is by

$$
\begin{aligned}
&\frac{\|\hat{G}\|_\infty}{c}\le \|\hat{H}_{\xi_2}\|_\infty\\
&\le \max_{\xi_1,\xi_2}\left|\int_{|x|>h_j}((FW^v)((\cdot+\xi_1)/2^j,\xi_2))^\vee(x-k_1/2^j)dx\right|\\
&\le \max_{\xi_2}\left|\int_{|x|>2^jh_j}((FW^v)((\cdot,\xi_2))^\vee(x-k_1)dx\right|.
\end{aligned}
$$

Certainly, the minimum is attained in the center of the mask, i.e., with $k=0$. So combining this with (7) and (10),

$$
\begin{aligned}
&|\langle(1-\mathscr{M}_{h_j})w\mathscr{L}_j,\psi_\lambda\rangle|\\
&\le c\max_{\xi_2}\left|\int_{|x|>2^jh_j}((FW^v)((\cdot,\xi_2)))^\vee(x)dx\right|\times\\
&\quad\times\langle|k_2|\rangle^{-N_1}\langle\min\{|k_1-2^j\rho|,|k_1+2^j\rho|\}\rangle^{-N_2}.
\end{aligned}
$$

which is what we intend to use as a "model." Observe that this indeed is also intuitively the right estimate, since the k_2 component has to decay rapidly away from zero, thereby sensing the singularity in zero in this direction. In contrast, the k_1 component stays greater or equal to $\langle 2^j\rho\rangle^{-N_2}$ up to the point $2\rho 2^j$ and then decays rapidly in accordance with the fact that up to the point $k_1=\rho 2^j$ we are "on" the line singularity which decays smoothly with $\hat{w}$. Moreover, the first term models the behavior in the mask, which is also nicely supported by the fact that the crucial product 2^jh_j is appearing therein.

Since $2^jh_j\to 0$ as $j\to\infty$, we have

$$
\max_{\xi_2}\left|\int_{|x|>2^jh_j}((FW^v)(|(,\xi_2)|))^\vee(x)dx\right|\to C,\quad j\to\infty.
$$

We now set the thresholds β_j to be

$$
\frac{c(C-\varepsilon)}{\langle|2^{j\varepsilon}|\rangle^{N_1}\langle\min\{|(2^{j\varepsilon}-1)2^j\rho|,|(2^{j\varepsilon}+1)2^j\rho|\}\rangle^{N_2}}.
$$

This choice immediately proves the claim of the lemma. □

Lemma 9 $\delta_j=\sum_{k\in\mathscr{T}_j^c}|\langle w\mathscr{L}_j,\psi_\lambda\rangle|=o(\|w\mathscr{L}_j\|_2)$, $j\to\infty$.

Proof We observe from the proof of Lemma 6 that the desired property is automatically satisfied provided that, for all $j\ge j_0$, the set $\mathscr{T}_j$ satisfies

$$
\mathscr{T}_j\supseteq\{k:|k_1|\le\rho 2^{j(1+\nu_1)},|k_2|\le 2^{j\nu_1}\},
$$

for some $\nu_1>0$, which is implied by Lemma 8. □

We next analyze the second term in the estimate from Proposition 3.

Lemma 10 *For $h_j=o(2^{-j})$ as $j\to\infty$,*

$$
\sum_{k\in\mathscr{T}_j}|\langle\mathscr{M}_{h_j}w\mathscr{L}_j,\psi_\lambda\rangle|=o(2^{j/2}),\qquad j\to\infty.
$$

Proof We first need to derive some estimates dependent on k for the term $|\langle\mathscr{M}_{h_j}w\mathscr{L}_j,\psi_\lambda\rangle|$. By using the definitions of $\mathscr{M}_{h_j}$ and $w\mathscr{L}_j$ and a change of variables, we first obtain

$$
\begin{aligned}
&\langle\mathscr{M}_{h_j}w\mathscr{L}_j,\psi_\lambda\rangle\\
&=2h_j\int\left[\int\hat{w}(\xi_1)\int\mathrm{sinc}((2h_j)\tau_1)F(\xi_1/2^j,\xi_2)\times\right.\\
&\quad\left.\times W^v(((\tau_1+\xi_1)/2^j,\xi_2))e^{-2\pi i\langle k_1,(\tau_1+\xi_1)/2^j\rangle}d\tau_1d\xi_1\right]\times\\
&\quad\times e^{-2\pi i\langle k_2,\xi_2\rangle}d\xi_2.
\end{aligned}
$$

Let $\hat{G}$ now be the function

$$
\begin{aligned}
\hat{G}(\xi_2)&=\int\hat{w}(\xi_1)2h_j\int\mathrm{sinc}((h_j/\pi)\tau_1)F(\xi_1/2^j,\xi_2)\times\\
&\quad\times W^v((\tau_1+\xi_1)/2^j,\xi_2)e^{-2\pi i\langle k_1/2^j,\tau_1+\xi_1\rangle}d\tau_1d\xi_1\\
&=\int\hat{w}(\xi_1)\hat{H}_{\xi_2}(\xi_1)e^{-2\pi i\langle k_1/2^j,\xi_1\rangle}d\xi_1.
\end{aligned}
$$

with

$$
\begin{aligned}
\hat{H}_{\xi_2}(\xi_1)&=2h_j\int\mathrm{sinc}((h_j/\pi)\tau_1)F(\xi_1/2^j,\xi_2)\times\\
&\quad\times W^v((\tau_1+\xi_1)/2^j,\xi_2)e^{-2\pi i\langle k_1/2^j,\tau_1\rangle}d\tau_1.
\end{aligned}
$$

The function $\hat{G}$ is supported on the set $[1/2,2]$, which is independent of j. Hence, we have

$$
|\langle\mathscr{M}_{h_j}w\mathscr{L}_j,\psi_\lambda\rangle|\le c_{N_1}\|\hat{G}\|_\infty\langle|k_2|\rangle^{-N_1}. \tag{11}
$$

By Plancherel's theorem and the support properties of w,

$$
\begin{aligned}
&\left|\int\hat{w}(\xi_1)\hat{H}_{\xi_2}(\xi_1)e^{-2\pi i\langle k_1/2^j,\xi_1\rangle}d\xi_1\right|=|(\hat{w}\hat{H}_{\xi_2})^\vee(-k_1/2^j)|\\
&=|(w\star H_{\xi_2})(-k_1/2^j)|\\
&=\left|\int w(-k_1/2^j-x)H_{\xi_2}(x)dx\right|\\
&\approx c\left|\int_{-k_1/2^j-\rho}^{-k_1/2^j+\rho}H_{\xi_2}(x)dx\right|.
\end{aligned}
$$

Next, using well-known properties of the Fourier transform, we can manipulate $H_{\xi_2}(x)$:

$$
\begin{aligned}
&= \left((2h_j \mathrm{sinc}(2h_j\cdot)e^{-2\pi i k_1/2^j\cdot}) \star (FW^v(\cdot/2^j,\xi_2))\right)^\vee(-x)\\
&= \left(2h_j \mathrm{sinc}(2h_j\cdot)e^{-2\pi i k_1/2^j\cdot}\right)^\vee(-x)\times\\
&\quad\times \left((FW^v)(\cdot/2^j,\xi_2)\right)^\vee(-x)\\
&= \mathbb{1}_{[-h_j,h_j]}(-x-k_1/2^j)\left((FW^v)(\cdot/2^j,\xi_2)|)\right)^\vee(-x).
\end{aligned}
$$

Hence, since $h_j < \rho$,

$$
\begin{aligned}
&c\left|\int_{-k_1/2^j-\rho}^{-k_1/2^j+\rho} H_{\xi_2}(x)dx\right|\\
&= c\left|\int_{k_1/2^j-h_j}^{k_1/2^j+h_j} ((FW^v)(\cdot/2^j,\xi_2))^\vee(x)dx\right|\\
&= c\left|\int_{k_1-2^jh_j}^{k_1+2^jh_j} ((FW^v)((\cdot,\xi_2)))^\vee(x)dx\right|.
\end{aligned}
$$

Notice that this indeed makes sense, since due to the masking the length of the line singularity isn't allowed to play a role here. For sufficiently "nice" W, there exists some constants N_2 and c such that

$$|((FW^v)(|(\cdot,\cdot)|)^\vee(x)| \le c\langle|x|\rangle^{-N_2}.$$

Hence,

$$\|\hat{G}\|_\infty \le c\langle\min\{|k_1-2^jh_j|,|k_1+2^jh_j|\}\rangle^{-N_2}.$$

Combining this estimate with (11), we obtain

$$
\begin{aligned}
&|\langle \mathscr{M}_{h_j} w\mathscr{L}_j, \psi_\lambda\rangle|\\
&\le c\langle|k_2|\rangle^{-N_1}\langle\min\{|k_1-2^jh_j|,|k_1+2^jh_j|\}\rangle^{-N_2},
\end{aligned}
$$

which is what we intend to use.

Finally,

$$
\begin{aligned}
&\sum_{k\in\mathscr{T}_j} |\langle \mathscr{M}_{h_j} w\mathscr{L}_j, \psi_\lambda\rangle|\\
&\le c\sum_{k\in\mathscr{T}_j}\langle|k_2|\rangle^{-N_1}\langle\min\{|k_1-2^jh|,|k_1+2^jh|\}\rangle^{-N_2}\\
&\le c.
\end{aligned}
$$

□

Notice that this result holds for any $\mathscr{T}_j$, which again is intuitively clear since if it holds for the claimed on, then extending the set $\mathscr{T}_j$ does not change the estimate due to the fact that $\mathscr{M}_{h_j} w\mathscr{L}_j$ is zero "outside."

We now apply Proposition 3 to Lemmata 4, 9, and 10 to obtain the desired convergence for the normalized ℓ_2 error of the reconstruction L_j from ONE-STEP-THRESHOLDING in Figure 7. Again, in this case $x = w\mathscr{L}_j$ and Φ are wavelets ψ_λ at scale j.

Theorem 2 *For $h_j = o(2^{-j})$ and L_j the solution to (2) with Φ the 2D Meyer orthonormal system,*

$$\frac{\|L_j - w\mathscr{L}_j\|_2}{\|w\mathscr{L}_j\|_2} \to 0, \qquad j\to\infty.$$

This result shows that ONE-STEP-THRESHOLDING fills in gaps of the same size as ℓ_1 minimization (INP) in an asymptotic sense when considering the ℓ_2 error.

5 Shearlet Inpainting Positive Results

In this section, Φ is the shearlet frame as in Subsection 1.2.2. The general approach in this section is the same as in the preceding section. We show that use of the analysis coefficients of the shearlet system through either ℓ_1 minimization or thresholding will successfully inpaint a line across a missing strip. Namely, in Subsection 5.1, we investigate the inpainting results of ℓ_1 minimization by estimating the δ-relative sparsity δ_j and cluster coherence μ_c with respect to $\{\sigma_\eta : \eta = (\iota,j,k,\ell), \iota\in\{h,v\}; |\ell|\le 2^{j/2}; k\in\mathbf{Z}^2\}$ and a properly chosen index set Λ_j. In Subsection 5.2, we similarly give the estimation of δ_j and μ_c for inpainting using thresholding. Some of the proofs in this section are very similar in spirit to the corresponding ones in Section 4 but decidedly more technical due to the structural difference between wavelets and shearlets. The auxiliary functions (6) and (12) in the proofs of Lemma 7 and Theorem 3 demonstrate this relationship quite well.

5.1 ℓ_1 Minimization

For our analysis we choose the set of significant shearlet coefficients to be

$$\Lambda_j = \{(\iota;j,k,\ell) : |k_1| \le \rho n_j 2^{j/2}, |k_2|\le n_j, \ell = 0; \iota = v\},$$

where we revive the notion $n_j = 2^{\varepsilon j}$ from the previous subsection.

Now we can show the relative sparsity of the shearlet coefficients with the choice of Λ_j.

Lemma 11 *For $\varepsilon < 1/4$,*

$$\delta_j = o(2^{j/2}), \qquad j\to\infty.$$

Proof By the definition, we have

$$
\begin{aligned}
\delta_j \;=\; &\sum_{|k_1|\ge\rho n_j 2^{j/2}, |k_2|\le n_j, \ell=0} |\langle w\mathscr{L}_j, \sigma^v_{j,\ell,k}\rangle| +\\
&+ \sum_{|k_2|\ge n_j,\ell=0} |\langle w\mathscr{L}_j, \sigma^v_{j,\ell,k}\rangle| + \sum_{k\in\mathbf{Z}^2,\ell\ne 0} |\langle w\mathscr{L}_j, \sigma^v_{j,\ell,k}\rangle| +\\
&+ \sum_{k\in\mathbf{Z}^2,\ell} |\langle w\mathscr{L}_j, \sigma^h_{j,\ell,k}\rangle|\\
=:\; &T_1+T_2+T_3+T_4.
\end{aligned}
$$

To estimate T_1, we first estimate $\langle w\mathscr{L}, \sigma_\eta\rangle$ for the case $\ell = 0$ and $\iota = v$. By Lemma 18 in Section 7,

$$
\begin{aligned}
&\langle w\mathscr{L}, \sigma^v_{j,k,0}\rangle\\
&\le c_N a_j^{-1/4}\langle|k_2|\rangle^{-1}\langle[k_2^2 + a_j^{-2}\min_{\pm}(a_j^{1/2}k_1 \pm \rho)^2]^{1/2}\rangle^{2-N}
\end{aligned}
$$

$$\le c_N a_j^{-1/4}\langle |k_2|\rangle^{-1}\langle [k_2^2+\min_{\pm}(a_j^{-1/2}k_1 \pm a_j^{-1}\rho)^2]^{1/2}\rangle^{2-N}$$
$$\le c_N a_j^{-1/4}\langle |k_2|\rangle^{-1}\langle a_j^{-1}\min_{\pm}|a_j^{1/2}k_1 \pm \rho|\rangle^{2-N},$$

Therefore, we have

$$T_1 \le c_N a_j^{-1/4} a_j^{-\varepsilon} \sum_{|k_1|\ge \rho a_j^{-1/2-\varepsilon}} \langle a_j^{-1}\min_{\pm}|a_j^{1/2}k_1 \pm \rho)|\rangle^{2-N}$$
$$\le c_N a_j^{-1/4} a_j^{-\varepsilon} \sum_{|k_1|\ge \rho a_j^{-1/2-\varepsilon}} \langle \min_{\pm}|a_j^{-1/2}k_1 \pm a_j^{-1}\rho)|\rangle^{2-N}.$$

Note that $a_j^{-\varepsilon} = n_j = 2^{j\varepsilon}$. Since

$$\int_{|x|>\rho a_j^{-1/2-\varepsilon}} \langle |a_j^{-1/2}x - a_j^{-1}\rho)|\rangle^{2-N} dx$$
$$= a_j^{1/2}\int_{|y|>\rho a_j^{-1-\varepsilon}} \langle |y - a_j^{-1}\rho|\rangle^{2-N} dy$$
$$\le a_j^{1/2}\int_{|y|>\rho a_j^{-1}} \langle |y|\rangle^{2-N} dy$$
$$\le c_N a_j^{1/2+N-3},$$

we obtain

$$T_1 \le c_N a_j^{1/4-\varepsilon+N-3}.$$

For T_2, we have

$$\frac{T_2}{c_N a_j^{-1/4}}$$
$$\le \sum_{k_1\in\mathbf{Z},|k_2|\ge a_j^{-\varepsilon}} \langle [k_2^2+\min_{\pm}(a_j^{-1/2}k_1 \pm a_j^{-1}\rho)^2]^{1/2}\rangle^{2-N}$$
$$\le \sum_{|k_1|\le \rho a_j^{-1/2-\varepsilon},|k_2|\ge a_j^{-\varepsilon}} \langle [k_2^2+\min_{\pm}(a_j^{-1/2}k_1 \pm a_j^{-1}\rho)^2]^{1/2}\rangle^{2-N}$$
$$+ \sum_{|k_1|> \rho a_j^{-1/2-\varepsilon},|k_2|\ge a_j^{-\varepsilon}} \langle [k_2^2+\min_{\pm}(a_j^{-1/2}k_1 \pm a_j^{-1}\rho)^2]^{1/2}\rangle^{2-N}$$
$$=: T_{2,1}+T_{2,2}.$$

For $T_{2,1}$, we have

$$T_{2,1} \le c\int_{|x_1|<\rho a_j^{-1/2-\varepsilon}}\int_{|x_2|>a_j^{-\varepsilon}} \langle |x_2|\rangle^{2-N} dx_2 dx_1$$
$$\le c a_j^{-1/2+(N-4)\varepsilon}.$$

For $T_{2,2}$, we have

$$T_{2,2} \le c a_j^{1/2}\int_{x_1>\rho a_j^{-1-\varepsilon}}\int_{x_2>a_j^{-\varepsilon}} \langle |(x_1,x_2)|\rangle^{2-N} dx_2 dx_1$$
$$\le c a_j^{(N-3)(1+2\varepsilon)}.$$

Therefore,

$$T_2 \le c_N a_j^{-3/4+(N-1)\varepsilon}.$$

For T_3, we convert the result in Lemma 19 in Section 7 to the discrete case.

Lemma 12 *Let $t_1 = a_j(k_1-\ell k_2)$ and $t_2 = a_j^{1/2}k_2$ with $a_j = 2^{-j}$.*

(i) *For $t_1\neq 0$ and $t_2\neq 0$, we have*

$$|\langle w\mathscr{L}_j, \sigma^h_{j,\ell,k}\rangle|$$
$$\le c_N e^{-c_\rho a_j^{-1/2}} a_j^{-1/4}|a_j(k_1-\ell k_2)|^{-N}|a_j^{1/2}k_2|^{-N} a_j^{N/2},$$

and

$$|\langle w\mathscr{L}_j, \sigma^v_{j,\ell,k}\rangle|$$
$$\le c_N e^{-c_\rho a_j^{-1}} a_j^{-1/4}|a_j(k_1-\ell k_2)|^{-N}|a_j^{1/2}k_2|^{-N} a_j^{N}.$$

(ii) *When exactly one of t_1 or t_2 is 0 and $\iota\in\{h,v\}$, we have*

$$|\langle w\mathscr{L}, \sigma^\iota_{j,\ell,k}\rangle|$$
$$\le c_L\Big[\max\{a_j|k_1-\ell k_2|, a_j^{1/2}|k_2|\}\Big]^{-L} a_j^{-1/4} e^{-c_\rho a_j^{-1/2}\ell}.$$

(iii) *For $t_1=t_2=0$ and $\iota\in\{h,v\}$, we have*

$$|\langle w\mathscr{L}, \sigma^\iota_{j,\ell,k}\rangle|$$
$$\le c a_j^{-1/4} e^{-c_\rho a_j^{-1/2}}.$$

For $t_1 := a_j(k_1-\ell k_2)\neq 0$ and $t_2 := a_j^{1/2}k_2\neq 0$, we have

$$a_j^{3/2}\sum_{k\in\mathbf{Z}^2,t_1\neq 0,t_2\neq 0} |a_j(k_1-\ell k_2)|^{-N}|a_j^{1/2}k_2|^{-N}$$
$$\le a_j^{3/2}\int_{\{x\,:\,x_1\neq \ell x_2, x_2\neq 0\}} |a_j(x_1-\ell x_2)|^{-N}|a_j^{1/2}x_2|^{-N} dx_1 dx_2$$
$$< c\cdot\int_{|x_1|\ge 1,|x_2|\ge 1} |x_1|^{-N}|x_2|^{-N} dx_1 dx_2$$
$$< \infty.$$

Hence

$$\sum_{k\in\mathbf{Z}^2,t_1\neq 0,t_2\neq 0} |a_j(k_1-\ell k_2)|^{-N}|a_j^{1/2}k_2|^{-N} < c a_j^{-3/2}.$$

Similarly, for $t_1=0$ or $t_2=0$, we have

$$\sum_{k\in\mathbf{Z}^2,t_1=0 \text{ or } t_2=0} \Big[\max\{a_j|k_1-\ell k_2|, a_j^{1/2}|k_2|\}\Big]^{-N} < c a_j^{-3/2}.$$

The estimate for (iii) follows by direct computation. Therefore, by the above estimates (i), (ii), and (iii), and that

$$T_3 = \sum_{\ell=1}^{a_j^{-1/2}}\sum_{k\in\mathbf{Z}^2,(t_1,t_2)\neq 0} |\langle w\mathscr{L}_j, \sigma^v_{j,\ell,k}\rangle| + \sum_{\ell=1}^{a_j^{-1/2}} |\langle w\mathscr{L}_j, \sigma^v_{j,\ell,0}\rangle|,$$

we obtain

$$T_3 \le \sum_{\ell=1}^{a_j^{-1/2}} c_N a_j^{-1/4} e^{-c_\rho a_j^{-1/2}}(a_j^{-3/2}+1) \le c_N a_j^N \quad \forall N\ge 0.$$

Similarly, for T_4,

$$T_4 \le \sum_{\ell=1}^{a_j^{-1/2}} c_N a_j^{-1/4} e^{-c_\rho a^{-1/2}}(a_j^{-3/2}+1) \le c_N a_j^N \quad \forall N\ge 0.$$

Combining the estimates for $T_1,\dots,T_4$, we are done.

Next we estimate the cluster coherence

$$\mu_c(\Lambda_j, \{\mathscr{M}_{h_j}\sigma_\eta\}; \{\sigma_\eta\})$$

and show that it converges to zero as $j \to \infty$ when h_j is related j by $h_j = o(2^{-j/2})$ as $j \to \infty$. We wish to remark that the size of the gaps which can be filled with asymptotically high precision is dramatically larger than the corresponding size for wavelet inpainting.

Theorem 3 *For $h_j = o(2^{-j/2})$*

$$\mu_c(\Lambda_j, \{\mathscr{M}_{h_j}\sigma_\eta\}; \{\sigma_\eta\}) \to 0, \quad j \to \infty$$

with $\eta = (\iota, j, \ell, k)$ and $\iota \in \{h, v\}$.

Proof We have

$$\begin{aligned}
&\mu_c(\Lambda_j, \{\mathscr{M}_{h_j}\sigma_\eta\}; \{\sigma_\eta\}) = \max_{\eta_2} \sum_{\eta_1 \in \Lambda_j} |\langle \mathscr{M}_{h_j}\sigma_{\eta_1}, \sigma_{\eta_2}\rangle| \\
&\leq \max_{\eta_2, \iota = v} \sum_{\eta_1 \in \Lambda_j} |\langle \mathscr{M}_{h_j}\sigma_{\eta_1}, \sigma_{\eta_2}\rangle| + \\
&\quad + \max_{\eta_2, \iota = h} \sum_{\eta_1 \in \Lambda_j} |\langle \mathscr{M}_{h_j}\sigma_{\eta_1}, \sigma_{\eta_2}\rangle| \\
&=: T_1 + T_2.
\end{aligned}$$

We bound T_1:

$$\begin{aligned}
T_1 &\leq \sum_{(\iota;j,\ell,k)\in\Lambda_j} |\langle \mathscr{M}_{h_j}\sigma^v_{j,\ell,k}, \sigma^v_{j,0,0}\rangle| \\
&\leq \sum_{(\iota;j,\ell,k)\in\Lambda_j} \int_{\mathbf{R}} 2h_j \mathrm{sinc}(2h_j\xi_1) \Bigg[\int_{\mathbf{R}^2} 2^{-3j/2} |W(\tau_2/2^j)|^2 \times \\
&\quad \times V(\ell + 2^{j/2}\frac{\tau_1 - \xi_1}{\tau_2}) V(2^{j/2}\frac{\tau_1}{\tau_2}) e^{-2\pi i \langle t, \tau - (\xi_1, 0)\rangle} d\tau \Bigg] d\xi_1 \\
&\leq 2(2^{j/2}h_j) \sum_{(\iota;j,\ell,k)\in\Lambda_j} \int_{\mathbf{R}} \mathrm{sinc}(2^{j/2}2h_j\xi_1) \Bigg[\int_{\mathbf{R}^2} |W(\tau_2)|^2 \times \\
&\quad \times V(\ell + \frac{\tau_1 - \xi_1}{\tau_2}) V(\frac{\tau_1}{\tau_2}) e^{it_1 2^{j/2}\xi_1} e^{-2\pi i \langle t, A^v_{1/a_j}\tau\rangle} d\tau \Bigg] d\xi_1 \\
&\leq 2(2^{j/2}h_j) \sum_{(\iota;j,\ell,k)\in\Lambda_j} \int_{\mathbf{R}} \hat{g}_j(\tau) e^{-2\pi i \langle t, A^v_{1/a_j}\tau\rangle} d\tau,
\end{aligned}$$

where $t = A^v_{1/a_j} S^v_\ell k$ with $a_j = 2^{-j}$ and

$$\begin{aligned}
\hat{g}_j(\tau) := \int_{\mathbf{R}} &\mathrm{sinc}(2^{j/2}2h_j\xi_1) V(\ell + \frac{\tau_1 - \xi_1}{\tau_2}) e^{2\pi i t_1 2^{j/2}\xi_1} d\xi_1 \times \\
&\times |W(\tau_2)|^2 V(\frac{\tau_1}{\tau_2}). \qquad (12)
\end{aligned}$$

Since $\hat{g}_j(\tau)$ is smooth and compactly supported on a box Ξ of volume independent of j,

$$|\int \hat{g}_j(\tau) e^{it\tau} d\tau| \leq c_N \|\hat{g}_j\|_\infty \langle |t| \rangle^{-N}.$$

Note that

$$\|\hat{g}_j\|_\infty \leq c(2^{j/2}h_j)^{-1/2};$$

□

therefore,

$$T_1 \leq c(2^{j/2}h_j)^{1/2} \sum_{k \in \mathbf{Z}^2} \langle |k| \rangle^{-N} \to 0, j \to \infty.$$

We now bound T_2:

$$\begin{aligned}
T_2 &\leq \sum_{(\iota;j,\ell,k)\in\Lambda_j} |\langle \hat{\mathscr{M}}_{h_j} \hat{\sigma}^v_{j,\ell,k}, \hat{\sigma}^h_{j,\ell,0}\rangle| \\
&\leq \sum_{(\iota;j,\ell,k)\in\Lambda_j} \int_{\mathbf{R}} 2h_j \mathrm{sinc}(2h_j\xi_1) \Bigg[\int_{\mathbf{R}^2} \hat{\sigma}^v_{a_j,s,0}(\tau - (\xi_1, 0)) \times \\
&\quad \times \hat{\sigma}^h_{a_j,s',0}(\tau) e^{-2\pi i \langle t, \tau - (\xi_1, 0)\rangle} d\tau \Bigg] d\xi_1 \\
&\leq \sum_{(\iota;j,\ell,k)\in\Lambda_j} \int_{\mathbf{R}^2} \Bigg[\int_{\mathbf{R}} 2h_j \mathrm{sinc}(2h_j\xi_1) \hat{\sigma}^v_{a_j,s,0}(\tau - (\xi_1, 0)) \times \\
&\quad \times \hat{\sigma}^h_{a_j,s',0}(\tau) d\xi_1 \Bigg] e^{-2\pi i \langle t, \tau - (\xi_1, 0)\rangle} d\tau \\
&=: \sum_{(\iota;j,\ell,k)\in\Lambda_j} \int_{\mathbf{R}^2} \hat{g}_j(\tau) e^{-2\pi i \langle t, \tau\rangle} d\tau,
\end{aligned}$$

where

$$\begin{aligned}
\hat{g}_j(\tau) := \int_{\mathbf{R}} &2h_j \mathrm{sinc}(2h_j\xi_1) \hat{\sigma}^v_{a_j,s,0}(\tau - (\xi_1, 0)) \times \\
&\times \hat{\sigma}^h_{a_j,s',0}(\tau) e^{2\pi i t_1 \xi_1} d\xi_1.
\end{aligned}$$

Using integration by parts, we obtain

$$\begin{aligned}
&|\int_{\mathbf{R}^2} \hat{g}_j(\tau) e^{-2\pi i \langle t, \tau\rangle} d\tau| \\
&\leq c_{L,M} \langle |t_1| \rangle^{-L} \langle |t_2| \rangle^{-M} \|D^{L,M}\hat{g}_j\|_\infty \mathrm{supp}\,(\hat{g}_j) \\
&\leq c_{L,M} \langle |t_1| \rangle^{-L} \langle |t_2| \rangle^{-M} \|D^{L,M}\hat{g}_j\|_\infty a_j^{-2},
\end{aligned}$$

where

$$\begin{aligned}
&|D^{L,M}\hat{g}_j| \\
&\leq 2h_j \int_{\mathbf{R}} |\mathrm{sinc}(2h_j\xi_1)| |D^{L,M}(\hat{\sigma}^v_{a_j,s,0}(\tau - (\xi_1, 0)) \times \\
&\quad \times \hat{\sigma}^h_{a_j,s',0}(\tau))| d\xi_1 \\
&\leq 2h_j \|\mathrm{sinc}(2h_j)\|_2 \|D^{L,M}(\hat{\sigma}^v_{a_j,s,0}(\tau - (,0)) \hat{\sigma}^h_{a_j,s',0}(\tau))\|_2 \\
&\leq c_{L,M} 2h_j^{1/2} \|D^{L,M}(\hat{\sigma}^v_{a_j,s,0}(\tau - (,0)) \hat{\sigma}^h_{a_j,s',0}(\tau))\|_\infty a_j^{-1/2}.
\end{aligned}$$

Since

$$\begin{aligned}
&\frac{\partial^N}{\partial \tau_1^N}(\hat{\sigma}^v_{a,s,0} \hat{\sigma}^h_{a,s',0}) = O(a_j^{3/4} a_j^{N/2}) \text{ and} \\
&\frac{\partial^N}{\partial \tau_2^N}(\hat{\sigma}^v_{a,s,0} \hat{\sigma}^h_{a,s',0}) = O(a_j^{3/4} a_j^{N/2}).
\end{aligned}$$

Consequently, as $j \to \infty$,

$$\begin{aligned}
T_2 &\leq h_j \sum_{(\iota;j,\ell,k)\in\Lambda_j} c_N \langle |t_1| \rangle^{-N} \langle |t_2| \rangle^{-N} a_j^{-2} h^{-1/2} a_j^{-1/2} a_j^{3/4} a_j^N \\
&\leq a_j^{N-3/4} h \to 0.
\end{aligned}$$

□

Notice that – in contrast to the wavelet result – here we require the stronger condition $(2^{j/2}h_j) \to 0$ as $j \to \infty$ to handle the additional angular component. However, we can slightly relax this condition, since again we only need boundedness away from $1/2$.

We now apply Proposition 1 to Lemmata 4, 11, and 3 to obtain the desired convergence for the normalized ℓ_2 error of the reconstruction L_j from (1). In this case $L = w\mathscr{L}_j$ and Φ are shearlets $\sigma^{\iota}_{j,\ell,k}$ at scale j.

Theorem 4 *For $h_j = o(2^{-j/2})$ and L_j the solution to (1) with Φ the shearlet system defined using the Meyer wavelet*

$$\frac{\|L_j - w\mathscr{L}_j\|_2}{\|w\mathscr{L}_j\|_2} \to 0, \qquad j \to \infty.$$

This result shows that we have asymptotically perfect inpainting as long as the size of the gap shrinks faster than $2^{-j/2}$. The similar result for wavelet inpainting, Theorem 1, only guarantees such successful inpainting when the gap is asymptotically smaller than 2^{-j}.

5.2 Thresholding

Our first claim concerns the set of the thresholding coefficients $\mathscr{T}_j := \{\eta = (\iota; j, \ell, k) : |\langle w\mathscr{L}_j, \sigma_\eta\rangle| \geq \beta_j\}$ for some $\beta_j > 0$.

Lemma 13 *For $h_j = o(2^{-j/2})$ as $j \to \infty$, there exist thresholds $\{\beta_j\}_j$ such that, for all $j \geq j_0$,*

$$\{(\iota; j, \ell, k) : |k_1| \leq \rho 2^{j(1+\nu_1)}, |k_2| \leq 2^{j\nu_1}, \ell = 0; \iota = v\} \subseteq \mathscr{T}_j$$

for some j_0, ν_1, and $\nu_2 < 1/4$.

Proof We first observe that

$$\begin{aligned}
&|\langle (1 - \mathscr{M}_{h_j}) w\mathscr{L}_j, \sigma^v_{j,\ell,k}\rangle| \\
&= |\langle \delta_0 \star \widehat{w\mathscr{L}}_j, \hat{\sigma}^v_{j,\ell,k}\rangle - \langle \hat{\mathscr{M}}_{h_j} \star \widehat{w\mathscr{L}}_j, \hat{\sigma}^v_{j,\ell,k}\rangle|.
\end{aligned}$$

The first term equals

$$\begin{aligned}
&\langle \delta_0 \star \widehat{w\mathscr{L}}_j, \hat{\sigma}^v_{j,\ell,k}\rangle \\
&= 2^{j/4} \int \left[\int \hat{w}(\xi_1) F(\xi_1/2^j, \xi_2) W(\xi_2) \times \right. \\
&\qquad \left. \times V(\ell + 2^{-j/2}\xi_1/\xi_2) e^{-i\langle b_1, \xi_1\rangle} d\xi_1 \right] e^{-2\pi i \langle 2^j b_2, \xi_2\rangle} d\xi_2; \quad (13)
\end{aligned}$$

whereas, by using Lemma 5, we derive for the second term

$$\begin{aligned}
&\langle \mathscr{M}_{h_j} w\mathscr{L}_j, \sigma^v_{j,\ell,k}\rangle \\
&= 2h_j \int \mathrm{sinc}(2h_j\tau_1) \int \hat{w}(\xi_1) F_j(\xi_1, \xi_2) \times \\
&\qquad \times \hat{\sigma}^v_{j,\ell,k}(\xi_1 + \tau_1, \xi_2) d\xi d\tau_1 \\
&= 2^{j/4} \int \left[\int \hat{w}(\xi_1) 2h_j \int \mathrm{sinc}(2h_j\tau_1) F(\xi_1/2^j, \xi_2) \times \right. \\
&\qquad \left. \times W(\xi_2) V(\ell + 2^{-j/2} \frac{\tau_1 + \xi_1}{\xi_2}) e^{-2\pi i \langle b_1, \tau_1 + \xi_1\rangle} d\tau_1 d\xi_1 \right] \times \\
&\qquad \times e^{-2\pi i \langle 2^j b_2, \xi_2\rangle} d\xi_2 \\
&=: 2^{j/4} \int \hat{G}(\xi_2) e^{-2\pi i \langle 2^j b_2, \xi_2\rangle} d\xi_2.
\end{aligned}$$

By standard arguments, we can deduce that

$$|\langle (1 - \mathscr{M}_{h_j}) w\mathscr{L}_j, \sigma^v_{j,\ell,k}\rangle| \leq c_{N_1} 2^{j/4} \|\hat{G}\|_\infty \langle |2^j b_2|\rangle^{-N_1}.$$

By $b_2 = k_2/2^j$ due to $b = (A^v_{2^{-j}} S^v_{-\ell})^T k$, we have

$$|\langle (1 - \mathscr{M}_{h_j}) w\mathscr{L}_j, \sigma^v_{j,\ell,k}\rangle| \leq c_{N_1} 2^{j/4} \|\hat{G}\|_\infty \langle |k_2|\rangle^{-N_1} \quad (14)$$

Let us now investigate the term $\|\hat{G}\|_\infty$ further. We define

$$\begin{aligned}
\hat{H}_{\xi_2}(\xi_1) &= F(\xi_1/2^j, \xi_2) W(\xi_2) V(\ell + 2^{-j/2}\xi_1/\xi_2) + \\
&-2h_j \int \mathrm{sinc}(2h_j\tau_1) F(\xi_2/2^j, \xi_2) W(\xi_2) \times \\
&\times V(\ell + 2^{-j/2} \frac{\xi_1 + \tau_1}{\xi_2}) e^{-2\pi i \langle b_1, \tau_1\rangle} d\tau_1
\end{aligned}$$

and hence need to analyze

$$\|\hat{G}\|_\infty = \left| \int \hat{w}(\xi_1) \hat{H}_{\xi_2}(\xi_1) e^{-2\pi i \langle b_1, \xi_1\rangle} d\xi_1 \right|. \quad (15)$$

By Plancherel's theorem and the support properties of w,

$$\begin{aligned}
\left| \int \hat{w}(\xi_1) \hat{H}_{\xi_2}(\xi_1) e^{-2\pi i \langle b_1, \xi_1\rangle} d\xi_1 \right| &= |(\hat{w}\hat{H}_{\xi_2})^\vee(-b_1)| \\
&\approx c \left| \int_{-b_1-\rho}^{-b_1+\rho} H_{\xi_2}(x) dx \right|.
\end{aligned}$$

We now need to compute H. Using well-known properties of the Fourier transform, we manipulate $H_{\xi_2}(x)$ to obtain

$$\begin{aligned}
&= \Big(F(\cdot/2^j, \xi_2) W(\xi_2) V(\ell + 2^{-j/2}(\cdot/\xi_2)) \Big)^\vee(x) \\
&\quad - \Big((2h_j \mathrm{sinc}(2h_j\cdot) e^{-2\pi i b_1 \cdot}) \star (F(\cdot/2^j, \xi_2) W(\xi_2) \times \\
&\quad \times V(\pi\ell + 2^{-j/2}(\cdot/\xi_2)) \Big)^\vee(-x) \\
&= \Big(F(\cdot/2^j, \xi_2) W(\xi_2) V(\ell + 2^{-j/2}(\cdot/\xi_2)) \Big)^\vee(x) \\
&\quad - \Big(2h_j \mathrm{sinc}(2h_j\cdot) e^{-2\pi i b_1 \cdot} \Big)^\vee(-x) \times \\
&\quad \times \Big(F(\cdot/2^j, \xi_2) W(\xi_2) V(\ell + 2^{-j/2} \cdot/\xi_2) \Big)^\vee(-x) \\
&= \Big(F(\cdot/2^j, \xi_2) W(\xi_2) V(\ell + 2^{-j/2}(\cdot/\xi_2)) \Big)^\vee(x) \\
&\quad - \mathbb{1}_{[-h_j, h_j]}(x - b_1) \times \\
&\quad \times \Big(F(\cdot/2^j, \xi_2) W(\xi_2) V(\ell + 2^{-j/2} \cdot/\xi_2) \Big)^\vee(-x).
\end{aligned}$$

Hence, since $h_j < \rho$,

$$\begin{aligned}
&\Big|\int_{-b_1-\rho}^{-b_1+\rho} H_{\xi_2}(x)dx\Big| \\
&= \Big|\int_{b_1-\rho}^{b_1+\rho} \Big(F(\cdot/2^j,\xi_2)W(\xi_2)V(\ell+2^{-j/2}(\cdot/\xi_2))\Big)^\vee(x) \\
&\quad -\int_{b_1-h_j}^{b_1+h_j}\Big(F(\cdot/2^j,\xi_2)W(\xi_2)V(\ell+2^{-j/2}(\cdot/\xi_2))\Big)^\vee(x)dx\Big| \\
&= \Big|\int_{2^{j/2}(b_1-h_j)}^{2^{j/2}(b_1-\rho)} + \int_{2^{j/2}(b_1+h_j)}^{2^{j/2}(b_1+\rho)} \Big(F(\cdot/2^j,\xi_2)W(\xi_2)\times \\
&\quad \times V(\ell+2^{-j/2}(\cdot/\xi_2))\Big)^\vee(x)dx\Big|.
\end{aligned}$$

Notice that this indeed makes sense, since the values k_1 "in between h_j and ρ" should play an essential role. As already observed in the proof of (14), we have $b_1 \approx k_1/2^{j/2}$ for j large and small $|\ell k_2|$ (since $b_1 = 2^{-j/2}k_1 + 2^{-j}\ell k_2$), and hence

$$\begin{aligned}
&c\Big|\int_{-b_1-\rho}^{-b_1+\rho} H(x)dx\Big| \\
&\approx c\Big|\int_{k_1-2^{j/2}\rho}^{k_1-2^{j/2}h_j} + \int_{k_1+2^{j/2}h_j}^{k_1+2^{j/2}\rho}\Big(F(\cdot/2^j,\xi_2)W(\xi_2)\times \\
&\quad \times V(\ell+2^{-j/2}(\cdot/\xi_2))\Big)^\vee(x)dx\Big|.
\end{aligned}$$

Notice that this fact also implies that the function

$$\Big(F(\cdot/2^j,\xi_2)W(\xi_2)V(\ell+2^{-j/2}(\cdot/\xi_2))\Big)^\vee$$

is independent of j. For sufficiently "nice" W, there exist some N_2 and c such that

$$|\Big(F(\cdot/2^j,\xi_2)W(\xi_2)V(\ell+2^{-j/2}(\cdot/\xi_2))\Big)^\vee(x)| \le c\langle|x|\rangle^{-N_2},$$

and hence by (15) and the previous computation,

$$\|\hat{G}\|_\infty \le c\langle\min\{|k_1-2^{j/2}\rho|,|k_1+2^{j/2}\rho|\}\rangle^{-N_2}. \tag{16}$$

Finally, we study how the term $\hat{H}$ relates to h_j. For this, we set

$$\hat{J}_{\xi_2}(\tau_1) = F(\xi_1/2^j,\xi_2)W(\xi_2)V(\ell+2^{-j/2}\frac{\xi_1+\tau_1}{\xi_2})e^{-i\langle b_1,\tau_1\rangle}$$

Now,

$$\begin{aligned}
|\hat{H}_{\xi_2}(\xi_1)| &= |F(\xi_1/2^j,\xi_2)W(\xi_2)V(\ell+2^{-j/2}\xi_1/\xi_2)+ \\
&\quad -2h_j\int \mathrm{sinc}(2h_j\tau_1)F(\xi_1/2^j,\xi_2)W(\xi_2)\times \\
&\quad \times V(\ell+2^{-j/2}\frac{\xi_1+\tau_1}{\xi_2})e^{-2\pi i\langle b_1,\tau_1\rangle}d\tau_1| \\
&= |\hat{J}_{\xi_2}(0) - 2h_j\int \mathrm{sinc}(2h_j\tau_1)\hat{J}_{\xi_2}(\tau_1)d\tau_1| \\
&= |\hat{J}_{\xi_2}(0) - \int \hat{1}_{[-h_j,h_j]}(\tau_1)\hat{J}_{\xi_2}(\tau_1)d\tau_1| \\
&= |\hat{J}_{\xi_2}(0) - \int_{-h_j}^{h_j} J_{\xi_2}(x)dx| \\
&= |\int_{|x|>h_j} J_{\xi_2}(x)dx|.
\end{aligned}$$

Hence another way to estimate (15) is by

$$\begin{aligned}
\|\hat{G}\|_\infty &\le c\|\hat{H}\|_\infty \\
&\le c\max_{\xi_1,\xi_2}\Big|\int_{|x|>h_j}(F(\xi_1/2^j,\xi_2)W(\xi_2)\times \\
&\quad \times V(\ell+2^{-j/2}\frac{\cdot+\xi_1}{\xi_2}))^\vee(x-b_1)dx\Big| \\
&\le c\max_{\xi_1,\xi_2}\Big|\int_{|x|>2^{j/2}h_j}(F(\xi_1/2^j,\xi_2)W(\xi_2)\times \\
&\quad \times V(\ell+\frac{\cdot+2^{-j/2}\xi_1}{\xi_2}))^\vee(x-2^{j/2}b_1)dx\Big|.
\end{aligned}$$

Certainly, the minimum is attained in the center of the mask, i.e., with $b=0$. So by combining this with (14) and (16),

$$\begin{aligned}
&|\langle(1-\mathscr{M}_{h_j})w\mathscr{L}_j,\sigma^v_{j,\ell,k}\rangle| \\
&\le c2^{j/4}\Big|\int_{|x|>2^{j/2}h_j}\max_{\xi_1,\xi_2}\Big|\int_{|x|>2^{j/2}h_j}(F(\xi_1/2^j,\xi_2)\times \\
&\quad \times W(\xi_2)V(\ell+\frac{\cdot+2^{-j/2}\xi_1}{\xi_2}))^\vee(x-2^{j/2}b_1)dx\Big|\times \\
&\quad \times\langle\min\{|k_1-2^{j/2}\rho|,|k_1+2^{j/2}\rho|\}\rangle^{-N_2}\langle|k_2|\rangle^{-N_1},
\end{aligned}$$

which is what we intend to use as a "model." Observe that this indeed is the right intuitive estimate, since the k_2 component has to decay rapidly away from zero thereby sensing the singularity in zero in this direction. In contrast, the k_1 component stays greater or equal to $\langle 2^j\rho\rangle^{-N_2}$ up to the point $2\rho 2^j$ and then decays rapidly in accordance with the fact that until the point $k_1 = \rho 2^j$ we are "on" the line singularity which decays smoothly up with $\hat{w}$. Also, the required angle sensitivity is represented. Finally, the first term models the behavior in the mask, which is also nicely supported by the fact that the crucial product $2^j h_j$ is appearing therein. Set

$$J(\cdot) = F(\cdot/2^j,\xi_2)W(\xi_2)V(\ell+2^{-j/2}(\cdot/\xi_2)).$$

Since $2^{j/2}h_j \to 0$ as $j\to\infty$, letting $j\to\infty$ we have

$$\Big|\int_{|x|>2^{j/2}h_j}\check{J}(x)dx\Big| \le C.$$

We now use

$$\begin{aligned}
\beta &= c2^{j/4}(C-\varepsilon)\langle|2^{j\varepsilon}|\rangle^{-N_1}\times \\
&\quad \times\langle\min\{|(2^{j\varepsilon}-1)2^{j/2}\rho|,|(2^{j\varepsilon}+1)2^{j/2}\rho|\}\rangle^{-N_2}
\end{aligned}$$

as a threshold. It follows immediately that, for all $j \ge j_0$,

$$\{(\iota;j,\ell,k):|k_1|\le\rho 2^{j(1+\nu_1)},|k_2|\le 2^{j\nu_1},\ell=0;\iota=v\}\subseteq\mathscr{T}_j$$

for some j_0 and ν_1.

□

Lemma 14 $\sum_{\eta\in\mathscr{T}_j^c} |\langle w\mathscr{L}_j, \sigma_\eta\rangle| = o(2^{j/2}),\ j\to\infty.$

Proof We observe from the proof of Lemma 11, that the desired property is automatically satisfied provided that, for all $j \geq j_0$, the set $\mathscr{T}_j$ contains

$$\{(\iota; j,\ell,k) : |k_1| \leq \rho 2^{j(1/2+\nu_1)}, |k_2| \leq 2^{j\nu_1}, \ell = 0, \iota = v\},$$

for some $\nu_1 > 0$, which is the content of Lemma 8. □

We next analyze the second term in the estimate from Proposition 3.

Lemma 15 *For $h_j = o(2^{-j/2})$ as $j\to\infty$,*

$$\sum_{\eta\in\mathscr{T}_j} |\langle \mathscr{M}_{h_j} w\mathscr{L}_j, \sigma_\eta\rangle| = o(2^{j/2}), \qquad j\to\infty.$$

Proof First, we need to derive some estimates dependent on (k,ℓ) for the term $|\langle \mathscr{M}_{h_j} w\mathscr{L}_j, \sigma^\iota_{j,\ell,k}\rangle|$. By using the definitions of $\mathscr{M}_{h_j}$ and $w\mathscr{L}_j$ and a change of variables, we obtain

$$\begin{aligned}
&\langle \mathscr{M}_{h_j} w\mathscr{L}_j, \sigma^v_{j,\ell,k}\rangle \\
&= 2^{j/4}\int\Bigg[\int \hat{w}(\xi_1)2h_j \int \operatorname{sinc}(2h_j\tau_1)F(\xi_1/2^j,\xi_2)\times \\
&\quad\times W(\xi_2)V(\ell+2^{-j/2}\frac{\tau_1+\xi_1}{\xi_2})\times \\
&\quad\times e^{-2\pi i b_1(\tau_1+\xi_1)}d\tau_1 d\xi_1\Bigg] e^{-2\pi i\langle 2^j b_2,\xi_2\rangle}d\xi_2.
\end{aligned}$$

Let $\hat{G}$ now be the function

$$\begin{aligned}
\hat{G}(\xi_2) &= \int \hat{w}(\xi_1)2h_j\int \operatorname{sinc}(2h_j\tau_1)F(\xi_1/2^j,\xi_2)\times \\
&\quad\times W(\xi_2)V(\ell+2^{-j/2}\frac{\tau_1+\xi_1}{\xi_2})e^{-2\pi i\langle b_1,\tau_1+\xi_1\rangle}d\tau_1 d\xi_1.
\end{aligned}$$

This function is supported on the set $[1/2,2]$, which is independent of j. By standard arguments, we can deduce that

$$|\langle \mathscr{M}_{h_j} w\mathscr{L}_j, \sigma^v_{j,\ell,k}\rangle| \leq c_{N_1} 2^{j/4}\|\hat{G}\|_\infty \langle |k_2|\rangle^{-N_1}. \tag{17}$$

Let us now investigate the term $\|\hat{G}\|_\infty$ further. We define

$$\begin{aligned}
\hat{H}_{\xi_2}(\xi_1) &= 2h_j\int \operatorname{sinc}(2h_j\tau_1)F(\xi_1/2^j,\xi_2)W(\xi_2)\times \\
&\quad\times V(\ell+2^{-j/2}\frac{\tau_1+\xi_1}{\xi_2})e^{-2\pi i\langle b_1,\tau_1+\xi_1\rangle}d\tau_1,
\end{aligned}$$

and hence need to analyze

$$\|\hat{G}\|_\infty = \left|\int \hat{w}(\xi_1)\hat{H}_{\xi_2}(\xi_1)e^{-2\pi i\langle b_1,\xi_1\rangle}d\xi_1\right|. \tag{18}$$

By Plancherel's theorem and the support properties of w,

$$\begin{aligned}
\left|\int \hat{w}(\xi_1)\hat{H}_{\xi_2}(\xi_1)e^{-2\pi i\langle b_1,\xi_1\rangle}d\xi_1\right| &= |(\hat{w}\hat{H}_{\xi_2})^\vee(-b_1)| \\
&\approx c\left|\int_{-b_1-\rho}^{-b_1+\rho} H_{\xi_2}(x)dx\right|.
\end{aligned}$$

Next,

$$\begin{aligned}
H_{\xi_2}(x) &= \Big((2h_j\operatorname{sinc}(2h_j\cdot)e^{-2\pi i b_1\cdot})\star(F(\cdot/2^j,\xi_2)W(\xi_2)\times \\
&\quad\times V(\ell+2^{-j/2}(\cdot/\xi_2)))\Big)^\vee(-x) \\
&= \left(2h_j\operatorname{sinc}(2h_j\cdot)e^{-2\pi i b_1\cdot}\right)^\vee(-x)\times \\
&\quad\times\left(F(\cdot/2^j,\xi_2)W(\xi_2)V(\ell+2^{-j/2}(\cdot/\xi_2))\right)^\vee(-x) \\
&= \mathbb{1}_{[-h_j,h_j]}(-x-b_1)\times \\
&\quad\times\left(F(\cdot/2^j,\xi_2)W(\xi_2)V(\ell+2^{-j/2}(\cdot/\xi_2))\right)^\vee(-x).
\end{aligned}$$

Hence, since $h_j < \rho$,

$$\begin{aligned}
&\left|\int_{b_{-1}-\rho}^{-b_1+\rho} H_{\xi_2}(x)dx\right| \\
&= \left|\int_{b_1-h_j}^{b_1+h_j}\Big(F(\cdot/2^j,\xi_2)W(\xi_2)V(\ell+2^{-j/2}(\cdot/\xi_2))\Big)^\vee(-x)dx\right| \\
&= \left|\int_{2^{j/2}(b_1-h_j)}^{2^{j/2}(b_1+h_j)}\Big(F(\cdot/2^{j/2},\xi_2)W(\xi_2)V(\ell+(\cdot/\xi_2))\Big)^\vee(-x)dx\right|.
\end{aligned}$$

Notice that this indeed makes sense, since due to the masking, the length of the line singularity is not allowed to play a role here. Since $(k,\ell)\in\mathscr{T}_j$, we have

$$\begin{aligned}
&\left|\int_{-b_1-\rho}^{-b_1+\rho} H(x)dx\right| \\
&= \left|\int_{k_1-2^{j/2}h_j}^{k_1+2^{j/2}h_j}(F(\cdot/2^{j/2},\xi_2)W(\xi_2)V(\ell+(\cdot/\xi_2)))^\vee(-x)dx\right|.
\end{aligned}$$

For sufficiently "nice" W, there exists some N_2 and c (possibly differing from the one before, but we do not need to distinguish those) such that

$$|(F(\cdot/2^{j/2},\xi_2)W(\xi_2)V(\ell+(\cdot/\xi_2)))^\vee(-x)| \leq c\langle|x|\rangle^{-N_2},$$

and hence by (18) and the previous computation,

$$\|\hat{G}\|_\infty \leq c\langle\min\{|k_1-2^{j/2}h_j|,|k_1+2^{j/2}h_j|\}\rangle^{-N_2}.$$

Combining this estimate with (17), we obtain

$$\begin{aligned}
&|\langle \mathscr{M}_{h_j} w\mathscr{L}_j, \sigma^v_{j,\ell,k}\rangle| \\
&\leq c2^{j/4}\langle|k_2|\rangle^{-N_1}\langle\min\{|k_1-2^{j/2}h_j|,|k_1+2^{j/2}h_j|\}\rangle^{-N_2},
\end{aligned}$$

which is what we intend to use. Hence,

$$\begin{aligned}
&\frac{1}{c}\sum_{\eta\in\mathscr{T}_j}|\langle \mathscr{M}_{h_j} w\mathscr{L}_j, \sigma_\eta\rangle| \\
&\leq 2^{j/4}\sum_{\eta\in\mathscr{T}_j}\langle|k_2|\rangle^{-N_1}\langle\min\{|k_1-2^{j/2}h_j|,|k_1+2^{j/2}h_j|\}\rangle^{-N_2} \\
&\leq 2^{j(1/4+\nu_2)}.
\end{aligned}$$

Since $\nu_2 < 1/4$, the lemma is proven.

□

We now apply Proposition 3 to Lemmata 4, 14, and 15 to obtain the desired convergence for the normalized ℓ_2 error of the reconstruction L_j from ONE-STEP-THRESHOLDING in Figure 7. In this case $x = w\mathscr{L}_j$ and Φ are shearlets $\sigma^{\iota}_{j,\ell,k}$ at scale j.

Theorem 5 *For $h_j = o(2^{-j/2})$ and L_j the solution to (2) with Φ the shearlet system defined using the Meyer wavelet*

$$\frac{\|L_j - w\mathscr{L}_j\|_2}{\|w\mathscr{L}_j\|_2} \to 0, \qquad j \to \infty.$$

This result shows that if the size of the gap shrinks faster than $2^{-j/2}$, the gap can be asymptotically perfect inpainted.

6 A Comparison of Shearlet vs. Wavelets

From the results of previous sections, we see that the size of the gaps which can be filled by shearlets ($h_j = o(2^{-j/2})$) with asymptotically high precision is larger than the corresponding size for wavelets ($h_j = o(2^{-j})$); however, certainly we still need to prove that we cannot do better than the presented rates for wavelet in order to show that shearlets perform better than wavelets. In fact, we show that the rates presented for wavelets are indeed the "critical scales" for the thresholding case.

Theorem 6 *Let ψ_λ be the Meyer orthonormal wavelets. Let $\mathscr{T}$ be a index set such that*

$$\mathscr{T} \supseteq \{(\iota, j, 0, (k_1, 0)) : |k_1| \le 2^j h_j - K_0\}$$

for some $K_0 > 0$ and $h_j > 0$. Then, we have

$$\sum_{\lambda \in \mathscr{T}} |\langle \mathscr{M}_{h_j} w\mathscr{L}_j, \psi_\lambda \rangle| = O(2^j h_j).$$

Proof At level j, the signal $w\mathscr{L}$ is filtered with the three corresponding frequency strips: $\check{F}_j = \sum_{\iota \in \{h,v,d\}} \check{F}^{\iota}_j(x,y)$ with $\check{F}^{\iota}_j(x,y) := 2^{2j}\check{W}^{\iota}(2^j x, 2^j y)$, $\iota = \{h,v,d\}$. Note that the Fourier transform of $\check{F}^{\iota}_j$ is $F^{\iota}_j = W^{\iota}(\xi_1/2^j, \xi_2/2^j)$. We can consider each of the filtered signals; i.e., consider $w\mathscr{L}^{\iota}_j := w\mathscr{L} \star F^{\iota}_j$ with $\iota = v,h,d$. Since the signal is a horizontal line segment, we only need to consider $w\mathscr{L}^h_j$. For simplicity, we denote $w\mathscr{L}_j := w\mathscr{L}^h_j$, $F_j := F^h_j$, and $\psi_\lambda = \psi^h_{j,k} =: \psi_{j,k}$. Note that $\check{F}_j(x,y) = 2^{2j}\phi(2^j x)\check{W}(2^j y)$. We want to estimate the coefficients $|\langle \mathscr{M}_{h_j} w\mathscr{L}_j, \psi_\lambda \rangle|$. By definition, we have

$$\begin{aligned}
&\langle \mathscr{M}_{h_j} w\mathscr{L}_j, \psi_\lambda \rangle \\
&= \int_{|x|<h_j} \int_{y \in \mathbf{R}} w\mathscr{L}_j(x,y)\psi_\lambda(x,y)dydx \\
&= \int_{|x|<h_j} \int_{y \in \mathbf{R}} (w\mathscr{L} \star \check{F}_j)(x,y)\psi_\lambda(x,y)dydx \\
&= \int_{|x|<h_j} \int_{y \in \mathbf{R}} \int_{z \in \mathbf{R}^2} w\mathscr{L}(z_1,z_2)\check{F}_j((x,y)-(z_1,z_2))dz \times \\
&\quad \times \psi_\lambda(x,y)dydx.
\end{aligned}$$

Now, by the definition of $w\mathscr{L}$, we have

$$\begin{aligned}
&\langle \mathscr{M}_{h_j} w\mathscr{L}_j, \psi_\lambda \rangle \\
&= \int_{|x|<h_j} \int_{y \in \mathbf{R}} \int_{-\rho}^{\rho} w(z)\check{F}_j(x-z,y)dz\psi_\lambda(x,y)dydx \\
&\approx c \int_{|x|<h_j} \int_{y \in \mathbf{R}} \int_{-\rho}^{\rho} \check{F}_j(x-z,y)dz\psi_\lambda(x,y)dydx \\
&= c \int_{|x|<h_j} \int_{y \in \mathbf{R}} \int_{-\rho}^{\rho} 2^{2j}\phi(2^j(x-z))\check{W}(2^j y)dz \times \\
&\quad \times 2^j \phi(2^j x - k_1)\check{W}(2^j y - k_2)dydx \\
&= c2^j \int_{y \in \mathbf{R}} \check{W}(2^j y)\check{W}(2^j y - k_2)dy 2^{2j} \times \\
&\quad \times \int_{|x|<h_j} \int_{-\rho}^{\rho} \phi(2^j(x-z))dz\phi(2^j x - k_1)dx \\
&= c2^{2j} \int_{|x|<h_j} \int_{-\rho}^{\rho} \phi(2^j(x-z))dz\phi(2^j x - k_1)dx \\
&= c2^j \int_{|x|<h_j} \int_{-2^j\rho+2^j x}^{2^j\rho+2^j x} \phi(z)dz\phi(2^j x - k_1)dx \\
&= c \int_{-k_1-2^j h_j}^{-k_1+2^j h_j} \int_{-2^j\rho+x+k_1}^{2^j\rho+x+k_1} \phi(z)dz\phi(x)dx.
\end{aligned}$$

For each $x \in [-k_1 - 2^j h_j, -k_1 + 2^j h_j]$, we have $x + k_1 \in [-2^j h_j, 2^j h_j]$. Consequently, we have

$$[-2^j\rho + x + k_1, 2^j\rho + x + k_1] \supseteq [-2^j(\rho - h_j), 2^j(\rho - h_j)]$$

for all $x \in [-k_1 - 2^j h_j, -k_1 + 2^j h_j]$. Note that $\rho > h_j$. Hence, when j is large enough, we have $\int_{-2^j\rho+x+k_1}^{2^j\rho+x+k_1} \phi(z)dz \approx c \neq 0$ due to $\int \phi(x)dx \neq 0$. Therefore, we have

$$\langle \mathscr{M}_{h_j} w\mathscr{L}_j, \psi_\lambda \rangle \approx c \int_{-k_1-2^j h_j}^{-k_1+2^j h_j} \phi(x)dx.$$

As $\int \phi(x)dx \neq 0$, there exists $K_0 > 0$ such that

$$\int_{|x|<K} \phi(x)dx \ge c_0$$

for some $c_0 > 0$ as long as $K > K_0$. Hence, when j is large enough so that $2^j h_j > K_0$ and $k_1 \in [-(2^j h_j - K_0), 2^j h - K_0]$, we have about $2^j h_j - K_0$ many coefficients that are larger than c_0. Consequently, when j is large enough, we have

$$\sum_{k \in \mathscr{T}} |\langle \mathscr{M}_{h_j} w\mathscr{L}_j, \psi_\lambda \rangle| = O(2^j h_j)$$

as long as the index set $\mathscr{T} \supseteq \{(\iota, j, 0, (k_1, 0)) : |k_1| \le 2^j h_j - K_0\}$.

For the other orientations $w\mathscr{L}^v_j$ and $w\mathscr{L}^d_j$, the coefficients are negligible following calculations similar to above. □

In the proof of Proposition 4, we have

$$\begin{aligned}
&\|x^\star - x^0\|_2 \\
&= \|\Phi 1_{\mathscr{T}^c} \Phi^* P_K x^0 + \Phi 1_{\mathscr{T}} \Phi^* P_M x^0\|_2 \\
&=: \|T_1 + T_2\|_2 \ge \|T_2\|_2 - \|T_1\|_2.
\end{aligned}$$

In the wavelet threshold case, the first term corresponds to $T_1 = \sum_{k\in\mathscr{T}^c} |\langle w\mathscr{L}_j, \psi_\lambda\rangle|$, while the second term corresponds to $T_2 = \sum_{k\in\mathscr{T}} \langle \mathscr{M}_{h_j} \cdot w\mathscr{L}_j, \psi_\lambda\rangle$ for some index set $\mathscr{T}$. As shown in the wavelet threshold, to guarantee that the first term $\|T_1\|_2$ is small, the index set $\mathscr{T}$ is chosen such that $\mathscr{T} \supseteq \{(k_1,k_2) : |k_1| \le \rho 2^{j(1+\nu_1)}, |k_1| \le 2^{j\nu_2}\}$. But then the second term $\|T_2\|_2$ will be of order $O(2^j h_j)$ as shown above. If h_j decays slower than order of $O(2^{-j/2})$, then we have $\|L_j - w\mathscr{L}_j\| = O(2^{j/2})$. Thus, we have the following theorem:

Theorem 7 *For $h_j = \omega(2^{-j/2})$ and L_j the solution to (2) where Φ is the 2D Meyer orthonormal system,*

$$\frac{\|L_j - w\mathscr{L}_j\|_2}{\|w\mathscr{L}_j\|_2} \nrightarrow 0, \quad j \to \infty.$$

That is, the wavelet threshold method does not fill the gap. Heuristically, one can think about the situation when the gap size h_j is fixed as 1. Consider the wavelets $2^j\phi(2^j x - k_1)\check{W}(2^j y)$. Then as $j \to \infty$, the number of such wavelets that fall in the gap is about $O(2^j)$. The norm $\langle \mathscr{M}_{h_j} w\mathscr{L}, \psi_\lambda\rangle$ for any such wavelets in the gap is about the same. Consequently, the total energy concentrated in the gap will be about $O(2^j)$.

When $2^j h_j \to 0$ and since $|\phi(x)| \le c_N \langle|x|\rangle^{-N}$ for any N, we have

$$\begin{aligned}&|\langle \mathscr{M}_{h_j} w\mathscr{L}_j, \psi_\lambda\rangle| \\ &\le c 2^j h_j \langle \min\{|k_1 - 2^j h_j|, |k_1 + 2^j h_j|\}\rangle^{-N}.\end{aligned}$$

For the Meyer mother wavelets $W^v = \check{W}(x)\phi(y)$ and $W^d = \check{W}(x)\psi(y)$, the above inequality still holds. In this case, the threshold method fills the gap.

Contrasting Theorem 5 and Theorem 7, we see that when the gap size h_j decays like $2^{j/2}$, the using the ONE-STEP-THRESHOLDING algorithm produces a good approximation of the original image if shearlets are used but does not if wavelets are used.

7 Appendix: Decay of Shearlet Coefficients Related to Line Singularity

We introduce the idea of a *continuous shearlet system* in order to prove various auxiliary results. For $\iota \in \{h,w\}$, $a > 0$, $s \in \mathbf{R}$, and $t \in \mathbf{R}^2$, define

$$\sigma^\iota_{a,s,t} := a^{-3/4} \sigma^\iota (S^\iota_s A^\iota_{a^{-1}}(\cdot - t)).$$

The cone-adaptive shearlet system is then obtained by sampling $\sigma^\iota_{a,s,t}$ on the discrete set of points

$$\begin{aligned}\{\iota = h,w\} \times \{a = 2^j : j \in \mathbf{N}\} \times \\ \times \{s = \ell : \ell \in \mathbf{Z}, |\ell| \le \lceil 2^{j/2} \rceil\} \times \{t \in A^\iota_{2^j} S^\iota_{-\ell} Z^2\}.\end{aligned}$$

To prove that the choice of Λ_j offers relative sparsity for the shearlet frame, we need some auxiliary results. The following lemma gives the decay estimate of the shearlet elements.

Note that if we define $\langle |t|_{a,s;\iota}\rangle := \langle |S^\iota_s A^\iota_{a^{-1}} t|\rangle$, then

$$|\sigma^\iota_{a,s,t}(x)| \le c_N a^{-3/4} \langle |x-t|_{a,s;\iota}\rangle^{-N}.$$

The following lemma is needed later for estimating the decay coefficients of the shearlet aligned with the singularity.

Lemma 16 *Let the line segment with respect to $(a,s,t;v)$ to be $Seg(a,s,t;v) := \{S^v_s A^v_{a^{-1}}(x - t_1, -t_2) : |x| \le \rho\}$. Then*

1. *Given the line*

   $$Line(a,s,t;v) := \{S^v_s A^v_{a^{-1}}(x - t_1, -t_2) : x \in \mathbf{R}\},$$

   the closest point P_L to the origin on this line satisfies

   $$d_1^2 := \|P_L\|_2^2 = \frac{a^{-2}}{1+s^2} t_2^2.$$

2. *Set $x_0 = \frac{a^{-1/2}s}{1+s^2} t_2 + t_1$. If P_S is the closest point on the segment $Seg(a,s,t;v)$ to the origin, then*

   $$\begin{aligned} d_2^2 &:= \|P_S - P_L\|_2^2 \\ &= \begin{cases} \min_\pm a^{-1}(1+s^2)(\pm\rho - x_0)^2 & x_0 \in [-\rho,\rho] \\ 0 & x_0 \notin [-\rho,\rho] \end{cases}. \end{aligned}$$

Proof Let $L(x) := S^v_s A^v_{a^{-1}}(x - t_1, -t_2)$. Then

$$\begin{aligned}&\|L(x)\|_2^2 = \|(a^{-1/2}(x-t_1), a^{-1/2}s(x-t_1) - a^{-1}t_2)\|_2^2 \\ &= a^{-1}(x-t_1)^2 + a^{-1}s^2(x-t_1)^2 + a^{-2}t_2^2 - 2a^{-3/2}s(x-t_1)t_2 \\ &= a^{-1}(1+s^2)(x-t_1)^2 + a^{-2}t_2^2 - 2a^{-3/2}s(x-t_1)t_2.\end{aligned}$$

Solving $\frac{d}{dx}\|L(x)\|_2 = 2(x-t_1)a^{-1}(1+s^2) - 2a^{-3/2}st_2 = 0$, we have $x_0 = \frac{a^{-1/2}s}{1+s^2}t_2 + t_1$. It follows that

$$P_L = L(x_0) = L(\frac{a^{-1/2}s}{1+s^2}t_2 + t_1) = \frac{a^{-2}}{1+s^2}t_2^2 =: d_1^2.$$

Note that $P_L \in Seg(a,s,t;v)$ if and only if $x \in [-\rho,\rho]$, in which case $d_2 = 0$. Otherwise,

$$\begin{aligned} d_2^2 &= \min_\pm \|L(\pm\rho) - P_L\|_2^2 \\ &= \min_\pm \|L(\pm\rho) - P_L\|_2^2 \\ &= \min_\pm \|(a^{-1/2}(\pm\rho - x_0), -a^{-1/2}s(\pm\rho - x_0))\|_2^2 \\ &= \min_\pm a^{-1}(1+s^2)(\pm\rho - x_0)^2, \end{aligned}$$

which completes the proof. □

We need another auxiliary lemma. Note that

$$\langle w\mathscr{L}, \sigma^\iota_{a,s,t}\rangle = \langle w\mathscr{L}_j, \sigma^\iota_{a,s,t}\rangle.$$

Lemma 17 *Define $R_N(x_0,y_0) := \int_{y_0}^\infty \langle |(x_0,\alpha)|\rangle^{-N} d\alpha$ (which may be thought of as a ray integral). Then for $y_0 \ge 0$,*

$$R_N(x_0,y_0) \le \pi \langle |x_0|\rangle^{-1} \langle |(x_0,y_0)|\rangle^{2-N}.$$

Proof Choose $\beta \in (0,1)$. Then

$$\int_0^\infty |f(\alpha)|d\alpha \leq (\sup_{t\in(0,\infty)} |f(\alpha)|^\beta) \int_0^\infty |f(\alpha)|^{1-\beta} d\alpha.$$

If we set $(1-\beta)N = 2$ and $f(t) = \langle |(x_0, y_0+\alpha)|\rangle^{-N}$, then we obtain

$$R_N(x_0,y_0) \leq (\sup_{v\in R(x_0,y_0)} \langle |v|\rangle^{2-N}) \int_0^\infty \langle |(x_0,y_0+\alpha|\rangle^{-2} d\alpha.$$

Since

$$\begin{aligned}\int_{-\infty}^\infty \langle |(x_0,y)|\rangle^{-M} dy &= \langle |x_0|\rangle^{-M} \int_{-\infty}^\infty \left\langle \frac{y}{\langle |x_0|\rangle} \right\rangle^{-M} dy \\ &= \langle |x_0|\rangle^{-M+1} \int_{-\infty}^\infty \langle \alpha\rangle^{-M} d\alpha,\end{aligned}$$

fixing $M = 2$ and recalling the classic identity $\pi = \int_{-\infty}^\infty (1+\alpha^2)^{-1} d\alpha$ yield the bound

$$\int_0^\infty \langle |(x_0,y_0+\alpha)|\rangle^{-2} d\alpha \leq \pi \langle |x_0|\rangle^{-1}.$$

Furthermore, since $y_0 \geq 0$,

$$\sup_{v\in R(x_0,y_0)} \langle |v|\rangle^{2-N} = \langle |(x_0,y_0)|\rangle^{2-N}.$$

This completes the proof. □

Now we can estimate the decay of the shearlet coefficients aligned with the line singularity $w\mathscr{L}$ as follows.

Lemma 18 *Retaining the notation as above, we have*

$$\langle w\mathscr{L}, \sigma^v_{a,s,t}\rangle \leq c_N \frac{a^{-1/4}}{\sqrt{1+s^2}} R_N(d_1, a^{-1/2}\sqrt{1+s^2} d_2)$$

$$\leq c_N \frac{a^{-1/4}}{\sqrt{1+s^2}} \langle |d_1|\rangle^{-1} \langle |(d_1, a^{-1/2}\sqrt{1+s^2} d_2|\rangle^{2-N}.$$

Proof We have

$$\begin{aligned}|\langle w\mathscr{L}, \sigma^v_{a,s,t}\rangle| &= |\int_{-\rho}^{\rho} w_1(x)\sigma^v_{a,s,t}(x,0)dx| \\ &\leq \int_{-\rho}^{\rho} |\sigma^v_{a,s,t}(x,0)dx \\ &\leq c_N a^{-3/4} \int_{Seg(a,s,t;v)} \langle |w|\rangle^{-N} dw, \qquad (19)\end{aligned}$$

where we use an affine transformation of variables to turn the anisotropic norm $|(x,0)|_{a,s,t;v}$ into the Euclidean norm $|w|$. Application of the same transformation to $[-\rho,\rho]\times\{0\}$ yields $Seg(a,s,t;v)$. The integral in (19) is along a curve traversing $Seg(a,s,t;v)$ at speed $v_1 = a^{-1/2}\sqrt{1+s^2}$. If we let $Ray(a,s,t;v)$ denote the ray starting from P_S and initially traversing $Seg(a,s,t;v)$, then

$$\begin{aligned}&a^{-3/4} \int_{Seg(a,s,t;v)} \langle |w|\rangle^{-N} dw \leq a^{-3/4} \int_{Ray(a,s,t;v)} \langle |w|\rangle^{-N} dw \\ &\leq a^{-3/4} v^{-1} \int_{v_1 Ray(a,s,t;v)} \langle |w|\rangle^{-N} dw \\ &\leq \frac{a^{-1/4}}{\sqrt{1+s^2}} \int_{v_1 d_2}^\infty \langle |(d_1,t)|\rangle^{-N} dw \\ &\leq \frac{a^{-1/4}}{\sqrt{1+s^2}} R_N(d_1, v_1 d_2).\end{aligned}$$

□

Next, we estimate the decay of the shearlet coefficients associated with those shearlets not aligned with the line singularity.

Lemma 19 *Let $t = (t_1,t_2)$. We consider the following three cases:*

(i) *$t_1 \neq 0$ and $t_2 \neq 0$. Then we have*

$$|\langle w\mathscr{L}, \sigma^v_{a,s,t}\rangle| \leq c_{L,M} |t_1|^{-L} |t_2|^{-M} a^{-1/4} e^{-c_\rho a^{-1/2} s} a^M,$$

when $1 \leq |s| \leq a^{-1/2}$ and

$$|\langle w\mathscr{L}, \sigma^h_{a,s,t}\rangle| \leq c_{L,M} |t_1|^{-L} |t_2|^{-M} a^{-1/4} e^{-c_\rho a^{-1}} a^{M/2}.$$

(ii) *If exactly one of t_1 or t_2 is 0, then we have*

$$|\langle w\mathscr{L}, \sigma^\iota_{a,s,t}\rangle| \leq c_L |t_1^2 + t_2^2|^{-L/2} a^{-1/4} e^{-c_\rho a^{-1/2} s}, \iota = h, v.$$

(iii) *$t_1 = t_2 = 0$. Then we have*

$$|\langle w\mathscr{L}, \sigma^\iota_{a,s,t}\rangle| \leq c a^{-1/4} e^{-c_\rho a^{-1/2}}, \iota = h, v.$$

Proof First, it is easy to show that

$$|\frac{\partial^L}{\partial \xi_1^L} \frac{\partial^M}{\partial \xi_2^M} |\hat{\sigma}^v_{a,s,0}| \leq c_{L,M} a^{3/4} a^{L/2} a^M,$$

By definition of the line singularity $w\mathscr{L}$, we have

$$\begin{aligned}\langle w\mathscr{L}, \sigma^v_{a,s,t}\rangle &= \int\int \hat{w}(\xi_1)\hat{\sigma}^h_{a,s,t}(\xi_1,\xi_2) d\xi_1 d\xi_2 \\ &= \int e^{-it_2\xi_2} \left[\int \hat{w}(\xi_1)\hat{\sigma}^v_{a,s,0}(\xi_1,\xi_2) e^{-t_1\xi_1} d\xi_1\right] d\xi_2.\end{aligned}$$

For $t_1 \neq 0$ and $t_2 \neq 0$, when we repeatedly apply integration by parts, we have

$$|\langle w\mathscr{L}, \sigma^v_{a,s,t}\rangle| \leq C |t_2|^{-M} |t_1|^{-L} \|h_{L,M}\|_{L^1(\mathbf{R})},$$

where

$$h_{L,M}(\xi_2) = \int D^{L,M}(\hat{w}(\xi_1)\hat{\sigma}^v_{a,s,0}(\xi_1,\xi_2)) d\xi_1,$$

and for some nice function f,

$$D^{L,M} f(\eta_1,\eta_2) = \left(\frac{\partial}{\partial \eta_1}\right)^L \left(\frac{\partial}{\partial \eta_2}\right)^M f(\eta_1,\eta_2).$$

The next step is to estimate the term $|h_{L,M}(\xi_2)|$.

Let $\Xi_{a,s}(\xi_2)$ be the support of the function

$$\xi_2 \mapsto D^{L,M}(\hat{w}(\xi_1)\hat{\sigma}^v_{a,s,0}(\xi_1,\xi_2)).$$

Note that for fixed a,s, the function $\xi_1 \mapsto \hat{w}(\xi_1)\hat{\sigma}^v_{a,s,0}(\xi_1,\xi_2)$ is supported inside $[c_\rho a^{-1/2}|s|,\infty)$ for a constant c_ρ depends only on ρ. $h_{L,M}$ can then be written as

$$h_{L,M}(\xi_2) = \int_{\Xi_{a,s}(\xi_2)} D^{L,M}(\hat{w}(\xi_1)\hat{\sigma}^v_{a,s,0}(\xi_1,\xi_2)) d\xi_1.$$

We then rewrite the integrand as

$$\begin{aligned}&D^{L,M}(\hat{w}(\xi_1)\hat{\sigma}^v_{a,s,0}(\xi_1,\xi_2)) \\ &= \sum_{\ell=0}^L \binom{L}{\ell} \hat{w}^{(\ell)}(\xi_1) D^{L-\ell,M}(\hat{\sigma}^v_{a,s,0}(\xi_1,\xi_2))\end{aligned}$$

Thus $|h_{L,M}(\xi_1)|$ is bounded by

$$
\begin{aligned}
&h_{L,M}(\xi_2)\\
&\leq \sum_{\ell=0}^{L}\binom{L}{\ell}\left|\int_{\Xi(a,s)(\xi_2)}\hat{w}^{(\ell)}(\xi_1)D^{L-\ell,M}(\hat{\sigma}^{h}_{a,s,0}(\xi_1,\xi_2))d\xi_1\right|\\
&\leq \sum_{\ell=0}^{L}\binom{L}{\ell}\|\hat{w}^{(\ell)}()\|_{L^1[c_\rho a^{-1/2}|s|,\infty)}N^{L-\ell,M}(a,s)\\
&\leq c_{L,M}e^{-c_\rho a^{-1/2}s}\sum_{\ell=0}^{L}\binom{L}{\ell}N^{L-\ell,M}(a,s)\\
&\leq c_{L,M}e^{-c_\rho a^{-1/2}s}a^{3/4}a^{M}
\end{aligned}
$$

where

$$N^{L-\ell,M}(a,s)=\|D^{L-\ell,M}\hat{\sigma}^{v}_{a,s,0}(\xi_1,\xi_2)\|_{L^\infty(\Xi_{a,s}(\xi_2))}$$

Consequently, we have

$$
\begin{aligned}
\|h_{L,M}\|_{L^1(\mathbf{R})} &\leq c_{L,M}a^{-1}e^{-c_\rho a^{-1/2}s}a^{3/4}a^{M}\\
&\leq c_{L,M}a^{-1/4}e^{-c_\rho a^{-1/2}s}a^{M}.
\end{aligned}
$$

Therefore,

$$|\langle w\mathscr{L},\sigma^{v}_{a,s,t}\rangle|\leq c_{L,M}|t_1|^{-L}|t_2|^{-M}a^{-1/4}e^{-c_\rho a^{-1/2}s}a^{M}.$$

Using the same approach, it is not difficult to show that for $|s|\leq a^{-1/2}$,

$$|\langle w\mathscr{L},\sigma^{h}_{a,s,t}\rangle|\leq c_{L,M}|t_1|^{-L}|t_2|^{-M}a^{-1/4}e^{-c_\rho a^{-1}}a^{M/2}.$$

The proofs for other cases are similar with simple modifications of the above procedure.

□

References

1. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans. Signal Process. **54**, 4311–4322 (2006)
2. Ballester, C., Bertalmio, M., Caselles, V., Sapiro, G., Verdera, J.: Filling-in by joint interpolation of vector fields and gray levels. IEEE Trans. Image Process. **10**(8), 1200–1211 (2001)
3. Bertalmio, M., Bertozzi, A., Sapiro, G.: Navier-stokes, fluid dynamics, and image and video inpainting. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001 (CVPR 2001), pp. I–355–I362. IEEE (2001)
4. Bertalmío, M., Sapiro, G., Caselles, V., Ballester, C.: Image inpainting. In: Proceedings of SIGGRAPH 2000, New Orleans, pp. 417–424 (2000)
5. Cai, J.F., Cha, R.H., Shen, Z.: Simultaneous cartoon and texture inpainting. Inverse Probl. Imag. **4**(3), 379 – 395 (2010)
6. Cai, J.F., Dong, B., Osher, S., Shen, Z.: Image restoration: Total variation, wavelet frames, and beyond. J. Amer. Math. Soc. (2012). To appear
7. Candès, E.J., Donoho, D.L.: New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities. Comm. Pure Appl. Math. **57**(2), 219–266 (2004)
8. Candès, E.J., Donoho, D.L.: Continuous curvelet transform. I. Resolution of the wavefront set. Appl. Comput. Harmon. Anal. **19**(2), 162–197 (2005)
9. Chan, T.F., Kang, S.H.: Error analysis for image inpainting. J. Math. Imaging Vision **26**(1-2), 85–103 (2006)
10. Chan, T.F., Kang, S.H., Shen, J.: Euler's elastica and curvature based inpainting. SIAM J. Appl. Math. **63**(2), 564–592 (2002)
11. Chan, T.F., Shen, J.: Mathematical models for local nontexture inpaintings. SIAM J. Appl. Math. **62**(3), 1019–1043 (electronic) (2001/02)
12. Daubechies, I.: Ten lectures on wavelets, *CBMS-NSF Regional Conference Series in Applied Mathematics*, vol. 61. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (1992)
13. Dong, B., Ji, H., Li, J., Shen, Z., Xu, Y.: Wavelet frame based blind image inpainting. Appl. Comput. Harmon. Anal. **32**(2), 268–279 (2012)
14. Donoho, D.L., Elad, M.: Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization. Proc. Natl. Acad. Sci. USA **100**(5), 2197–2202 (electronic) (2003)
15. Donoho, D.L., Huo, X.: Uncertainty principles and ideal atomic decomposition. IEEE Trans. Inform. Theory **47**(7), 2845–2862 (2001)
16. Donoho, D.L., Kutyniok, G.: Microlocal analysis of the geometric separation problem. Comm. Pure Appl. Math. (2012). Accepted
17. Elad, M., Bruckstein, A.M.: A generalized uncertainty principle and sparse representation in pairs of bases. IEEE Trans. Inform. Theory **48**(9), 2558–2567 (2002)
18. Elad, M., Starck, J.L., Querre, P., Donoho, D.L.: Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA). Appl. Comput. Harmon. Anal. **19**(3), 340–358 (2005)
19. Engan, K., Aase, S., Hakon Husoy, J.: Method of optimal directions for frame design. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, 1999. (ICASSP '99), vol. 5, pp. 2443–2446 (1999)
20. Grohs, P.: Continuous shearlet frames and resolution of the wavefront set. Monatsh. Math. **164**(4), 393–426 (2011). DOI 10.1007/s00605-010-0264-2. URL http://dx.doi.org/10.1007/s00605-010-0264-2
21. Grohs, P., Kutyniok, G.: Parabolic molecules. preprint. (2012)
22. Guo, K., Labate, D.: Analysis and detection of surface discontinuities using the 3D continuous shearlet transform. Appl. Comput. Harmon. Anal. **30**(2), 231–242 (2011)
23. Hennenfent, G., Fenelon, L., Herrmann, F.J.: Nonequispaced curvelet transform for seismic data reconstruction: A sparsity-promoting approach. Geophysics **75**(6), WB203–WB210 (2010)
24. Herrmann, F.J., Hennenfent, G.: Non-parametric seismic data recovery with curvelet frames. Geophys. J. Int. **173**, 233–248 (2008)
25. Hörmander, L.: The analysis of linear partial differential operators. I. Classics in Mathematics. Springer-Verlag, Berlin (2003). Distribution theory and Fourier analysis, Reprint of the second (1990) edition [Springer, Berlin; MR1065993 (91m:35001a)]
26. King, E.J., Kutyniok, G., Zhuang, X.: Analysis of data separation and recovery problems using clustered sparsity. SPIE Proceedings: Wavelets and Sparsity XIV **8138** (2011)
27. Kowalski, M., Torrésani, B.: Sparsity and persistence: mixed norms provide simple signal models with dependent coefficients. Signal, Image and Video Processing **3**, 251–264 (2009)
28. Kutyniok, G.: Geometric separation by single pass alternating thresholding (2010). Preprint
29. Kutyniok, G., Labate, D.: Resolution of the wavefront set using continuous shearlets. Trans. Amer. Math. Soc. **361**(5), 2719–2754 (2009)
30. Kutyniok, G., Labate, D. (eds.): Shearlets: Multiscale Analysis for Multivariate Data. Applied and Numerical Harmonic Analysis. Birkhäuser (2012)
31. Kutyniok, G., Lemvig, J., Lim, W.: Compactly supported shearlets. In: Approximation Theory XIII (San Antonio, TX, 2010). Springer (2010)

32. Kutyniok, G., Lemvig, J., Lim, W.: Shearlets and optimally sparse approximations. In: Shearlets: Multiscale Analysis for Multivariate Data. Springer (2011). To appear
33. Kutyniok, G., Lemvig, J., Lim, W.Q.: Compactly supported shearlet frames and optimally sparse approximations of functions in $l^2(r^3)$ with piecewise c^α singularities. SIAM J. Appl. Math. (to appear)
34. Kutyniok, G., Lim, W.: Compactly supported shearlets are optimally sparse. J. Approx. Theory **163**, 1564–1589 (2011)
35. Meyer, Y.: Principe d'incertitude, bases hilbertiennes et algèbres d'opérateurs. Astérisque (145-146), 4, 209–223 (1987). Séminaire Bourbaki, Vol. 1985/86
36. Meyer, Y.: Oscillating patterns in image processing and nonlinear evolution equations, *University Lecture Series*, vol. 22. American Mathematical Society, Providence, RI (2001). The fifteenth Dean Jacqueline B. Lewis memorial lectures
37. Nam, S., Davies, M., Elad, M., Gribonval, R.: Cosparse analysis modeling - uniqueness and algorithms. In: International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2011). IEEE (2011)
38. Nam, S., Davies, M.E., Elad, M., Gribonval, R.: The Cosparse Analysis Model and Algorithms. Appl. Comput. Harmon. Anal. (2012). In Press
39. Olshausen, B.A., Field, D.J.: Sparse coding with an overcomplete basis set: A strategy employed by V1? Vision Res. **37**(23), 3311–3325 (1997)