

# Radon Transform Inversion using the Shearlet Representation

Flavia Colonna

*Department of Mathematics, George Mason University, Fairfax, Virginia*

Glenn Easley\*

*System Planning Corporation, Arlington, Virginia*

Kanghui Guo

*Department of Mathematics, Missouri State University, Springfield, Missouri  
65804, USA*

Demetrio Labate<sup>1</sup>

*Department of Mathematics, University of Houston, Houston, TX 77204, USA*

---

## Abstract

The inversion of the Radon transform is a classical ill-posed inverse problem where some method of regularization must be applied in order to accurately recover the objects of interest from the observable data. A well-known consequence of the traditional regularization methods is that some important features to be recovered are lost, as evident in imaging applications where the regularized reconstructions are blurred versions of the original. In this paper, we show that the affine-like system of functions known as the *shearlet system* can be applied to obtain a highly effective reconstruction algorithm which provides near-optimal rate of convergence in estimating a large class of images from noisy Radon data. This is achieved by introducing a shearlet-based decomposition of the Radon operator and applying a thresholding scheme on the noisy shearlet transform coefficients. For a given noise level  $\epsilon$ , the proposed shearlet shrinkage method can be tuned so that the estimator will attain the essentially optimal mean square error  $O(\log(\epsilon^{-1})\epsilon^{4/5})$ , as  $\epsilon \rightarrow 0$ . Several numerical demonstrations show that its performance improves upon similar competitive strategies based on wavelets and curvelets.

*Key words:* directional wavelets; inverse problems, Radon transform shearlets; wavelets

*1991 MSC:* 42C15, 42C40

## 1 Introduction

The Radon transform, introduced by Johann Radon in 1917 [25], is the underlying mathematical foundation for a number of methods employed to determine structural properties of objects by using projected information, such as computerized  $X$ -ray tomography and magnetic resonance imaging (MRI). It also provides the basic mathematical principles employed by remote sensing devices such as synthetic aperture radar (SAR) and inverse synthetic aperture radar (ISAR).

In its classical formulation (see, for example, [22]), in dimension  $n = 2$ , the Radon transform can be described as follows. For  $\theta \in S^1$  and  $t \in \mathbb{R}$ , consider the lines in  $\mathbb{R}^2$ :

$$L(\theta, t) = \{x \in \mathbb{R}^2 : x \cdot \theta = t\}.$$

These are the lines perpendicular to  $\theta$  with distance  $t$  from the origin, and represent, in  $X$ -ray tomography, the paths along which the  $X$ -rays travel. The Radon transform is used to model the attenuation of an  $X$ -ray traveling across the object  $f$  along the line  $L(\theta, t)$ , and, for  $f \in L^1(\mathbb{R}^2)$ , is defined by

$$(Rf)(\theta, t) = \int_{x \in L(\theta, t)} f(x) dx.$$

The problem of interest consists in inverting the transform to recover the function  $f$  from the Radon data  $(Rf)(\theta, t)$  and was solved, in principle, by J. Radon who obtained the inversion formula

$$f(x) = \frac{1}{4\pi^2} \int_{S^1} \int_{\mathbb{R}} \frac{\frac{d}{dt} R(\theta, t)}{x \cdot \theta - t} dt d\theta.$$

However, how to convert this inversion formula into an accurate computational algorithm is far from obvious. Indeed, this inverse problem is technically *ill-posed* since the solution is unstable with respect to small perturbations in the projection data  $R(\theta, t)$ . In practical applications, the data  $R(\theta, t)$  are known with a limited accuracy (on a discrete set only) and are typically corrupted by noise, so that some method of regularization for the inversion is needed in order to accurately recover the function  $f$  and control the amplification of noise in the reconstruction.

Starting from the rediscovery of the Radon transform in the 60's and its applications to computerized tomography [6,7], several methods have been intro-

---

\* Corresponding author

*Email addresses:* fcolonna@gmu.edu (Flavia Colonna), geasley@sysplan.com (Glenn Easley), KanghuiGuo@MissouriState.edu (Kanghui Guo), dlabate@math.uh.edu (Demetrio Labate).

<sup>1</sup> Partially supported by NSF grant DMS 0604561 and DMS (Career) 0746778.

duced to deal with the inverse problem associated with the Radon transform, including Fourier methods, backprojection and singular value decomposition [22]. A well known limitation of all these methods is that they usually yield reconstructions where high frequency features, such as edges, are smoothed away, with the result that the reconstructed images are blurred versions of the original ones. While a number of heuristic methods have been introduced to deal with the phenomenon of blurring in the Radon inversion, only in recent years, in the work of Candès and Donoho [4], a method was proposed to deal with the efficient reconstruction of images with edges and with the precise assessment of the method performance.

In fact, the approach developed in the work of Candès and Donoho relies on some recent advances in the theory of wavelets and multiscale methods which provide a new theoretical perspective on the problem of dealing with information associated with edges effectively. Following a similar theoretical framework, in this paper, we propose a novel technique for regularizing the inversion of the Radon transform by means of a multiscale and multidirectional representation known as the *shearlet representation*. Our approach provides an algorithm for inverting the Radon transform which is particularly efficient in recovering data containing edges and other distributed discontinuities. In particular, by taking advantage of the sparsity properties of the shearlet representation, we prove that the shearlet-based inversion is optimally efficient in the reconstruction of images containing edges from noisy Radon data. In addition, thanks to some specific advantages of the discrete shearlet decomposition, the shearlet-based numerical algorithm allows for a significant performance improvement over the wavelets- and curvelets-based results.

### 1.1 Historical Perspective and Motivation

To provide a more detailed perspective on the method that we propose, let us start by recalling the general framework of the Wavelet-Vaguelette decomposition (WVD), introduced by Donoho [9]. This method applies a collection of functionals called *vaguelettes* to simultaneously invert an operator and compute the wavelet coefficients of the desired function. The function is then estimated by applying a nonlinear shrinkage to the noise-contaminated wavelet coefficients and inverting the wavelet transform. The ingenuity of the Wavelet-Vaguelette decomposition is that it emphasizes the estimating capabilities of a representation best suited to approximate the underlying function. This is in contrast to constructions such as the Singular Value Decomposition (SVD) that use basis functions depending solely on the operator to regularize the inversion. Since the appearance of this work, this strategy has spurred much interest for the Radon inversion problems as well as for other inverse problems (e.g., [20,18,23]). We also recall that several other wavelet-based techniques

have been applied to the problem of inverting the Radon transform, including [2,24,26].

However, for the Radon inversion of images containing edges, while still outperforming other traditional methods, the WVD technique as well as the other wavelet-based methods fall short from being optimal in terms of its estimation capabilities. Consider the problem of recovering an image  $f$  which smooth away from regular edges, from the noisy data

$$Y = Rf + \epsilon W,$$

where  $\epsilon W$  is white Gaussian noise, and  $\epsilon$  is measuring the noise level. Then an inversion based on the WVD approach yields a Mean Squared Error (MSE) that is bounded, within a logarithmic factor, by  $O(\epsilon^{2/3})$  as  $\epsilon \rightarrow 0$ . This is better than the MSE rate of  $O(\epsilon^{1/2})$ , as  $\epsilon \rightarrow 0$ , which is achieved when the inversion techniques are SVD-based [4]. To further improve the performance, one should replace the wavelet system used in the WVD with a representation system which is more capable of dealing with edges. This is exactly the motivation for the introduction of the biorthogonal curvelet decomposition of the Radon transform in [4] whose application yields a MSE rate

$$O(\log(\epsilon^{-1})\epsilon^{4/5}) \quad \text{as } \epsilon \rightarrow 0. \quad (1)$$

This rate is essentially optimal for this class of functions.

The method that we propose also adapts the basic WVD framework. By taking advantage of the shearlet representation, we obtain a novel decomposition of the Radon transform which is optimally efficient in dealing with images containing edges and provides the same essentially optimal estimation rate given by (1). Notice that, while offering similar approximation properties, the shearlet and curvelet representations have very different mathematical constructions. In particular, unlike the curvelet representation, the shearlet approach is based on the framework of affine systems, where the representation elements are obtained by applying a countable collections of translations and dilations to a finite set of generators. As a consequence, the shearlet approach provides a simpler and more flexible mathematical setting, and a unified treatment for both the continuous formulation and the corresponding discrete implementations. Indeed, while the current curvelet description given in [3] theoretically is ideal for inverting the Radon transform from a continuous perspective, its implementation has to deal with the fact that the image to be estimated is to be described on a finite discrete set (typically, a rectangular grid). By exploiting these advantage of the shearlet setting, we can demonstrate that, in the practical numerical implementation, the shearlet-based estimation process performs significantly better than the corresponding curvelet-based estimation process.

Finally, we would like to notice that the original argument provided in [4] to prove the MSE estimation rate (1) is based on an older and somewhat cumbersome formulation of the curvelet representation. By using the simpler shearlet construction, we are able to provide a much more streamlined and straightforward set of arguments to establish the MSE estimation rate result for our approach.

## 1.2 Paper Organization

The paper is organized as follows. In Section 2, we recall the basic definitions and properties of the shearlet representation which are needed for the paper. In Section 3, we develop a decomposition of the Radon transform based on the shearlet representation. In Section 4, we analyze the performance of the shearlet-based inversion algorithm when the Radon data are corrupted by additive Gaussian noise. Finally, in Section 5 we present several numerical experiments to illustrate the performance of the shearlet-based algorithm and compare it against competing wavelet- and curvelet-based algorithms.

## 2 Shearlet Representation

Despite their spectacular success in a variety of applications from applied mathematics and signal and image processing, it is now acknowledged that traditional wavelets are not particularly efficient in dealing with multidimensional data. This limitation has stimulated a very active research during the last 10 years, which has led to the introduction of a new generation of multiscale systems with improved capability to capture the geometry of multidimensional data. These systems include the curvelets [4], the bandlets [21], the contourlets [8] and the composite wavelets [14–16] (of which the shearlets are a special realization).

The theory of composite wavelets, in particular, provides a uniquely effective approach which combines geometry and multiscale analysis by taking advantage of the theory of affine systems. In dimension  $n = 2$ , the *affine systems with composite dilations* are the collections of functions of the form

$$\mathcal{A}_{AB}(\psi) = \{\psi_{j,\ell,k}(x) = |\det A|^{j/2} \psi(B^\ell A^j x - k) : j, \ell \in \mathbb{Z}, k \in \mathbb{Z}^2\},$$

where  $\psi \in L^2(\mathbb{R}^2)$ , and  $A, B$  are  $2 \times 2$  invertible matrices with  $|\det B| = 1$ . The elements of this system are called *composite wavelets* if  $\mathcal{A}_{AB}(\psi)$  forms a

*Parseval frame* (also called *tight frame*) for  $L^2(\mathbb{R}^2)$ ; that is,

$$\sum_{j,\ell,k} |\langle f, \psi_{j,\ell,k} \rangle|^2 = \|f\|^2,$$

for all  $f \in L^2(\mathbb{R}^2)$ . In these systems, the integer powers of the dilations matrices  $A$  are associated with scale transformations, while the integer powers of the matrices  $B$  are associated to area-preserving geometric transformations, such as rotations and shear. This framework allows one to construct Parseval frames whose elements, in addition to ranging at various scales and locations, like ordinary wavelets, also range at various orientations.

In this paper, we will consider and apply a special example of affine systems with composite wavelets in  $L^2(\mathbb{R}^2)$ , called the *shearlet system*. One of the reasons for the particular significance of this system is that it is the only construction, together with the curvelets, which provide optimally sparse representations for a large class of two-dimensional functions [13]. This property will play a prominent role in the decomposition of the Radon transform.

The shearlet system is an affine systems with composite dilations where  $A = A_0$  is an anisotropic dilation matrix and  $B = B_0$  is the shear matrix, which are defined by

$$A_0 = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}, \quad B_0 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

For any  $\xi = (\xi_1, \xi_2) \in \widehat{\mathbb{R}}^2$ ,  $\xi_1 \neq 0$ , let

$$\hat{\psi}^{(0)}(\xi) = \hat{\psi}^{(0)}(\xi_1, \xi_2) = \hat{\psi}_1(\xi_1) \hat{\psi}_2\left(\frac{\xi_2}{\xi_1}\right), \quad (2)$$

where  $\hat{\psi}_1, \hat{\psi}_2 \in C^\infty(\widehat{\mathbb{R}})$ ,  $\text{supp } \hat{\psi}_1 \subset [-\frac{1}{2}, -\frac{1}{16}] \cup [\frac{1}{16}, \frac{1}{2}]$  and  $\text{supp } \hat{\psi}_2 \subset [-1, 1]$ . Hence,  $\hat{\psi}^{(0)}$  is a compactly-supported  $C^\infty$  function with support contained in  $[-\frac{1}{2}, \frac{1}{2}]^2$ . In addition, we assume that

$$\sum_{j \geq 0} |\hat{\psi}_1(2^{-2j}\omega)|^2 = 1 \quad \text{for } |\omega| \geq \frac{1}{8}, \quad (3)$$

and, for each  $j \geq 0$ ,

$$\sum_{\ell=-2^j}^{2^j-1} |\hat{\psi}_2(2^j\omega - \ell)|^2 = 1 \quad \text{for } |\omega| \leq 1. \quad (4)$$

From the conditions on the support of  $\hat{\psi}_1$  and  $\hat{\psi}_2$  one can easily deduce that the functions  $\psi_{j,\ell,k}$  have frequency support contained in the set

$$\{(\xi_1, \xi_2) : \xi_1 \in [-2^{2j-1}, -2^{2j-4}] \cup [2^{2j-4}, 2^{2j-1}], |\frac{\xi_2}{\xi_1} + \ell 2^{-j}| \leq 2^{-j}\}.$$

Thus, each element  $\hat{\psi}_{j,\ell,k}$  is supported on a pair of trapezoids of approximate size  $2^{2j} \times 2^j$ , oriented along lines of slope  $\ell 2^{-j}$  (see Figure 1(a)).

From equations (3) and (4) it follows that the functions  $\{\hat{\psi}^{(0)}(\xi A_0^{-j} B_0^{-\ell})\}$  form a tiling of the set

$$\mathcal{D}_0 = \{(\xi_1, \xi_2) \in \widehat{\mathbb{R}}^2 : |\xi_1| \geq \frac{1}{8}, \left| \frac{\xi_2}{\xi_1} \right| \leq 1\}.$$

Indeed, for  $(\xi_1, \xi_2) \in \mathcal{D}_0$

$$\sum_{j \geq 0} \sum_{\ell = -2^j}^{2^j-1} |\hat{\psi}^{(0)}(\xi A_0^{-j} B_0^{-\ell})|^2 = \sum_{j \geq 0} \sum_{\ell = -2^j}^{2^j-1} |\hat{\psi}_1(2^{-2j} \xi_1)|^2 |\hat{\psi}_2(2^j \frac{\xi_2}{\xi_1} - \ell)|^2 = 1. \quad (5)$$

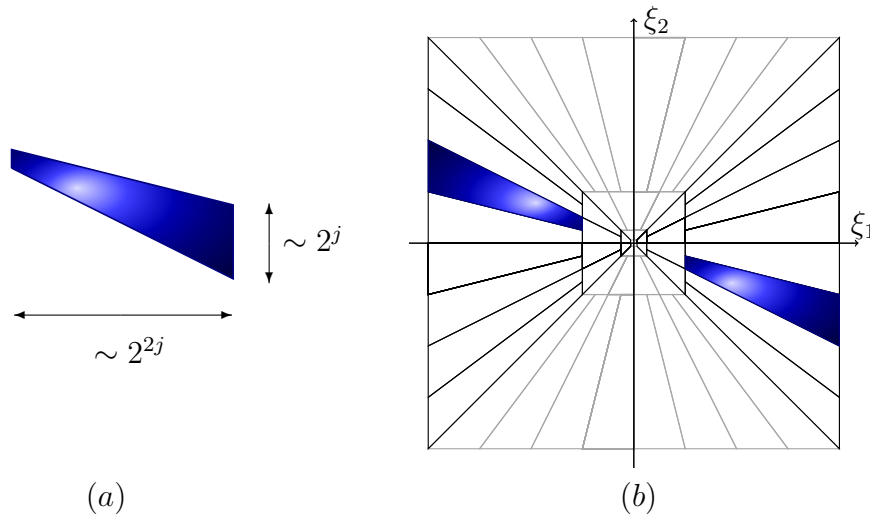


Fig. 1. (a) The frequency support of a shearlet  $\psi_{j,\ell,k}$  satisfies parabolic scaling. The figure shows only the support for  $\xi_1 > 0$ ; the other half of the support, for  $\xi_1 < 0$ , is symmetrical. (b) The tiling of the spatial-frequency plane  $\widehat{\mathbb{R}}^2$  induced by the shearlets. The tiling of  $\mathcal{D}_0$  is illustrated in black lines, the tiling of  $\mathcal{D}_1$  is shown in gray lines.

Letting  $L^2(\mathcal{D}_0)^\vee = \{f \in L^2(\mathbb{R}^2) : \text{supp } \hat{f} \subset \mathcal{D}_0\}$ , property (5) and the fact that  $\hat{\psi}^{(0)}$  is supported inside  $[-\frac{1}{2}, \frac{1}{2}]^2$  imply that the collection  $\{\psi_{j,\ell,k}^{(0)} : j \geq 0, -2^j \leq \ell \leq 2^j - 1, k \in \mathbb{Z}^2\}$  defined by

$$\psi_{j,\ell,k}^{(0)}(x) = 2^{\frac{3j}{2}} \psi^{(0)}(B_0^\ell A_0^j x - k) \quad (6)$$

is a Parseval frame for  $L^2(\mathcal{D}_0)^\vee$ . Similarly, we can construct a Parseval frame for  $L^2(\mathcal{D}_1)^\vee$ , where  $\mathcal{D}_1$  is the vertical cone  $\mathcal{D}_1 = \{(\xi_1, \xi_2) \in \widehat{\mathbb{R}}^2 : |\xi_2| \geq \frac{1}{8}, \left| \frac{\xi_1}{\xi_2} \right| \leq 1\}$ . Specifically, let

$$A_1 = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}, \quad B_1 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix},$$

and  $\psi^{(1)}$  be given by

$$\hat{\psi}^{(1)}(\xi) = \hat{\psi}^{(1)}(\xi_1, \xi_2) = \hat{\psi}_1(\xi_2) \hat{\psi}_2\left(\frac{\xi_1}{\xi_2}\right).$$

Then the collection  $\{\psi_{j,\ell,k}^{(1)} : j \geq 0, -2^j \leq \ell \leq 2^j - 1, k \in \mathbb{Z}^2\}$  defined by

$$\psi_{j,\ell,k}^{(1)}(x) = 2^{\frac{3j}{2}} \psi^{(1)}(B_1^\ell A_1^j x - k) \quad (7)$$

is a Parseval frame for  $L^2(\mathcal{D}_1)^\vee$ .

Finally, let  $\hat{\varphi} \in C_0^\infty(\mathbb{R}^2)$  be chosen to satisfy

$$\begin{aligned} 1 &= |\hat{\varphi}(\xi)|^2 + \sum_{j \geq 0} \sum_{\ell = -2^j}^{2^j - 1} |\hat{\psi}^{(0)}(\xi A_0^{-j} B_0^{-\ell})|^2 \chi_{\mathcal{D}_0}(\xi) \\ &\quad + \sum_{j \geq 0} \sum_{\ell = -2^j}^{2^j - 1} |\hat{\psi}^{(1)}(\xi A_1^{-j} B_1^{-\ell})|^2 \chi_{\mathcal{D}_1}(\xi), \end{aligned}$$

where  $\chi_{\mathcal{D}}$  is the indicator function of the set  $\mathcal{D}$ . This implies that  $\text{supp } \hat{\varphi} \subset [-\frac{1}{8}, \frac{1}{8}]^2$ ,  $|\hat{\varphi}(\xi)| = 1$  for  $\xi \in [-\frac{1}{16}, \frac{1}{16}]^2$ , and the collection  $\{\varphi_k : k \in \mathbb{Z}^2\}$  defined by  $\varphi_k(x) = \varphi(x - k)$  is a Parseval frame for  $L^2([-\frac{1}{16}, \frac{1}{16}]^2)^\vee$ .

An illustration of the frequency tiling provided by the the shearlet system is shown in Figure 1(b).

Thus, we have the following result:

**Theorem 2.1** *Let  $\varphi$  and  $\psi_{j,\ell,k}^{(d)}$  be defined as above. Also, for  $d = 0, 1$ , let  $\tilde{\psi}_{j,\ell,k}^{(d)}(\xi) = \hat{\psi}_{j,\ell,k}^{(d)}(\xi) \chi_{\mathcal{D}_d}(\xi)$ . The collection of shearlets*

$$\begin{aligned} &\{\varphi_k : k \in \mathbb{Z}^2\} \cup \{\psi_{j,\ell,k}^{(d)}(x) : j \geq 0, -2^j + 1 \leq \ell \leq 2^j - 2, k \in \mathbb{Z}^2, d = 0, 1\} \\ &\cup \{\tilde{\psi}_{j,\ell,k}^{(d)}(x) : j \geq 0, \ell = -2^j, 2^j - 1, k \in \mathbb{Z}^2, d = 0, 1\}, \end{aligned}$$

is a Parseval frame for  $L^2(\mathbb{R}^2)$ .

Notice that the ‘‘corner’’ elements  $\tilde{\psi}_{j,\ell,k}^{(d)}(x)$ ,  $\ell = -2^j, 2^j - 1$ , are simply obtained by truncation on the cones  $\chi_{\mathcal{D}_d}$  in the frequency domain and that the corner elements in the horizontal cone  $\mathcal{D}_0$  match nicely with those in the vertical cone  $\mathcal{D}_1$ . We refer to [13,16] for additional detail about the construction of the shearlet system.

In the following, for brevity of notation, it will be convenient to introduce the index set  $\mathcal{M} = N \cup M$ , where  $N = \mathbb{Z}^2$ ,  $M = \{\mu = (j, \ell, k, d) : j \geq 0, -2^j \leq$

$\ell \leq 2^j - 1, k \in \mathbb{Z}^2, d = 0, 1\}$ . Hence, we will denote the shearlet system as the collection  $\{s_\mu : \mu \in \mathcal{M} = N \cup M\}$ , where  $s_\mu = \psi_\mu = \psi_{j,\ell,k}^{(d)}$  if  $\mu \in M$  and  $s_\mu = \varphi_\mu$  if  $\mu \in N$ . For  $\psi_\mu, \mu \in M$ , it is understood that the corner elements are modified as in Theorem 2.1.

The *shearlet transform* is the mapping on  $L^2(\mathbb{R}^2)$  defined by:

$$\mathcal{SH} : f \rightarrow \mathcal{SH}f(\mu) = \langle f, s_\mu \rangle, \mu \in \mathcal{M}.$$

### 3 Inversion of Radon Transform via Shearlet Representation

We will obtain a formula for the decomposition of the Radon transform based on the shearlet representation. In order to do that, some construction is needed.

#### 3.1 Companion Representation

We start by introducing a companion representation of the shearlet system which is obtained under the action of the fractional Laplacian.

**Definition 3.1** For a rational number  $\alpha$  and  $f \in C^\infty(\mathbb{R}^2)$ , define  $(-\Delta)^\alpha f$  by the relation  $((-\Delta)^\alpha f)^\wedge(\xi) = |\xi|^{2\alpha} \hat{f}(\xi)$ . For  $\mu \in \mathcal{M}$ , let  $s_\mu^+ = 2^{-j}(-\Delta)^{1/4} s_\mu$ . In particular, we use the notation  $\psi_\mu^+ = 2^{-j}(-\Delta)^{1/4} \psi_\mu$ , for  $\mu \in M$ , and  $\phi_\mu^+ = (-\Delta)^{1/4} \phi_\mu$ , for  $\mu \in N$ .

It is easy to verify that the functions  $\{s_\mu^+ : \mu \in \mathcal{M}\}$  are smooth and compactly supported in the frequency domain. In addition, since each element of the fine-scale shearlet system  $\psi_\mu, \mu \in M$ , is supported, in the frequency domain, on the compact region  $C_j = [-2^{2j-1}, 2^{2j-1}]^2 \setminus [-2^{2j-4}, 2^{2j-4}]^2$ , it follows that, in this region,  $2^{-j}|\xi|^{1/2} \approx 1$  and, thus,  $\|\psi_\mu^+\| \approx \|\psi_\mu\|$ . This is the key observation which is used in the following result and is similar to the one given in [4] for the case of curvelets.

**Theorem 3.2** The system  $\{\psi_\mu^+\}_{\mu \in M}$  is a frame for  $L^2(\mathbb{R}^2 \setminus [-\frac{1}{8}, \frac{1}{8}]^2)^\vee$ . That is, there are constants  $0 < A \leq B < \infty$  such that

$$A \|f\|_2 \leq \left\| \sum_{\mu} \langle f, \psi_\mu^+ \rangle \right\|_{\ell^2} \leq B \|f\|_2,$$

for all functions  $f$  such that  $\text{supp } \hat{f} \subset \mathbb{R}^2 \setminus [-\frac{1}{8}, \frac{1}{8}]^2$ .

Notice that the larger system  $\{s_\mu^+ : \mu \in M\}$  is not a frame for  $L^2(\mathbb{R}^2)$  since the lower frame bound condition is not satisfied. To enlarge  $\{\psi_\mu^+\}_{\mu \in M}$  and

make it into a frame for the whole space  $L^2(\mathbb{R}^2)$ , one can include a “coarse scale” system of the form  $\{\varphi(x - k) : k \in \mathbb{Z}^n\}$  as done for the original shearlet system.

**Proof.** To show that the system  $\{\psi_\mu^+\}_{\mu \in M}$  is a frame, we introduce the following notation. For  $j \geq 0$ , let  $M_j = \{\mu = (j, \ell, k, d) : -2^j \leq \ell \leq 2^j - 1, k \in \mathbb{Z}^2, d = 0, 1\}$ . Also, denote  $\mathcal{C} = \mathbb{R}^2 \setminus [-\frac{1}{8}, \frac{1}{8}]^2$ .

Using the properties of the shearlet function  $\psi$ , for  $h \in L^2(\mathcal{C})^\vee$ , we have:

$$\begin{aligned}
\sum_{\mu \in M_j} |\langle h, \psi_\mu \rangle|^2 &= \sum_{\ell, k, d} \left| \int_{\mathcal{C}} \hat{h}(\xi) \overline{\hat{\psi}_{j\ell k}^{(d)}(\xi)} d\xi \right|^2 \\
&= \sum_{\ell, k, d} \left| \int_{\mathcal{C}} \hat{h}(\xi) \overline{\hat{\psi}^{(d)}(\xi A_d^{-j} B_d^{-\ell})} e^{-2\pi i \xi A_d^{-j} B_d^{-\ell} k} d\xi \right|^2 \\
&= \int_{\mathcal{C}} |\hat{h}(\xi)|^2 \sum_{\ell, d} |\hat{\psi}^{(d)}(\xi A_d^{-j} B_d^{-\ell})|^2 d\xi \\
&= \int_{\mathcal{C}} |\hat{h}(\xi)|^2 \hat{V}_j(\xi) d\xi, \tag{8}
\end{aligned}$$

where

$$\hat{V}_j(\xi) = |\hat{\psi}_1(2^{-2j}\xi_1)|^2 \chi_{\mathcal{D}_0}(\xi) + |\hat{\psi}_1(2^{-2j}\xi_2)|^2 \chi_{\mathcal{D}_1}(\xi).$$

Notice that the window function  $\hat{V}_j$  is supported on the set

$$C_j = [-2^{2j-1}, 2^{2j-1}]^2 \setminus [-2^{2j-4}, 2^{2j-4}]^2,$$

and that, by the assumptions on  $\psi_1$ ,

$$\sum_{j \geq 0} \hat{V}_j(\xi) = 1 \quad \text{for } \xi \in \mathbb{R}^2 \setminus [-\frac{1}{8}, \frac{1}{8}].$$

Let  $\hat{w}_j$  be a smooth window function supported in  $C_{j-1} \cup C_j \cup C_{j+1}$  which is equal 1 on the set  $C_j$ . It follows that  $\hat{w}_j \hat{V}_j = \hat{V}_j$  and that  $w_j * g = g$  for each function  $g$  such that  $\hat{g}$  is supported on the set  $C_j$ . Hence, using the fact that  $\text{supp } \hat{\psi}_\mu \subset C_j$ , when  $\mu \in M_j$ , we have:

$$\begin{aligned}
\sum_{\mu \in M_j} |\langle f, \psi_\mu^+ \rangle|^2 &= \sum_{\mu \in M_j} |\langle f, 2^{-j} (-\Delta)^{1/4} \psi_\mu \rangle|^2 \\
&= \sum_{\mu \in M_j} |\langle w_j * f, 2^{-j} (-\Delta)^{1/4} \psi_\mu \rangle|^2 \\
&= \sum_{\mu \in M_j} |\langle 2^{-j} (-\Delta)^{1/4} (w_j * f), \psi_\mu \rangle|^2 \\
&= \sum_{\mu \in M_j} |\langle h_j, \psi_\mu \rangle|^2,
\end{aligned}$$

where  $h_j = 2^{-j}(-\Delta)^{1/4}(w_j * f)$ . Notice that the smoothness assumption on  $w_j$  guarantees that  $h_j$  is well defined.

Using (8), we have:

$$\begin{aligned} \sum_{\mu \in M_j} |\langle h_j, \psi_\mu \rangle|^2 &= \int_{\mathcal{C}} |\hat{h}_j(\xi)|^2 \hat{V}_j(\xi) d\xi \\ &= 2^{-2j} \int_{C_j} |\xi| |\hat{f}(\xi)|^2 \hat{V}_j(\xi) d\xi. \end{aligned}$$

Notice that, for  $\xi \in C_j$ , the function  $2^{-2j}|\xi|$  is bounded above and below by positive constants, independently of  $j$ . Hence, from the expression above, we have that

$$\sum_{\mu \in M_j} |\langle h_j, \psi_\mu \rangle|^2 \simeq \int_{C_j} |\hat{f}(\xi)|^2 \hat{V}_j(\xi) d\xi$$

Adding up over all  $j \geq 0$ , it follows that

$$\begin{aligned} \sum_{\mu \in M} |\langle f, \psi_\mu^+ \rangle|^2 &= \sum_{j \geq 0} \sum_{\mu \in M_j} |\langle f, \psi_\mu^+ \rangle|^2 \\ &\simeq \sum_{j \geq 0} \int_{C_j} |\hat{f}(\xi)|^2 \hat{V}_j(\xi) d\xi \\ &= \int_{\mathcal{C}} |\hat{f}(\xi)|^2 d\xi \end{aligned}$$

for all  $f \in L^2(\mathcal{C})^\vee$ .  $\square$

### 3.2 Shearlet Decomposition of the Radon Transform

We start by recalling some important properties of the Radon transform, which will be useful in the construction of our decomposition based on the shearlet representation. We refer to [17] for additional detail about these properties and their derivation.

In dimension  $n = 2$ , the *Radon operator*  $R$  associates to each suitable function  $f$  and each pair  $(\theta, t) \in [0, 2\pi) \times \mathbb{R}$ , the value

$$Rf(\theta, t) = \int f(x, y) \delta(x \cos \theta + y \sin \theta - t) dx dy,$$

where  $\delta$  is the Dirac distribution at the origin. It is easy to verify that  $Rf$ , the Radon transform of  $f$ , satisfies the *antipodal symmetry*  $Rf(\theta + \pi, -t) = Rf(\theta, t)$ . Let  $\mathcal{D}_R$  denote the space of all functions in  $L^2([0, 2\pi) \times \mathbb{R})$  satisfying the antipodal symmetry. For a rational number  $\alpha$ , define the operator of

fractional differentiation (of a single variable) as

$$D^\alpha f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\omega|^\alpha \widehat{f}(\omega) e^{i\omega t} d\omega,$$

and let

$$\mathcal{R} = (I \otimes D^{\frac{1}{2}}) \circ R,$$

where  $I$  denotes the identity operator. The operator  $\mathcal{R}$  is well-defined on functions on  $L^2([0, 2\pi) \times \mathbb{R})$  and it satisfies the *Radon isometry*

$$[\mathcal{R}f, \mathcal{R}g] = \langle f, g \rangle,$$

where  $[\cdot, \cdot]$  is the inner product in  $L^2([0, 2\pi) \times \mathbb{R})$ . Also, one can show that  $\mathcal{R}$  maps functions which are smooth and of rapid decay on  $\mathbb{R}^2$  into functions which are smooth and of rapid decay on  $[0, 2\pi) \times \mathbb{R}$ .

Using the companion shearlet representations  $\{s_\mu^+ : \mu \in \mathcal{M}\}$  from Section 3.1, we define the system  $\{U_\mu : \mu \in \mathcal{M}\}$  by the formula

$$U_\mu = \mathcal{R}s_\mu^+, \quad \mu \in \mathcal{M}. \quad (9)$$

Using the Radon isometry and Theorem 3.2, one can show that  $\{U_\mu : \mu \in \mathcal{M}\}$  is a frame sequence (that is, a frame for its span), even if not a frame for the whole Hilbert space  $\mathcal{D}_R$  since  $\{U_\mu : \mu \in \mathcal{M}\}$  is not a frame for  $L^2(\mathbb{R}^2)$ :

We are now ready to introduce a decomposition formula for the Radon operator based on the shearlet representation. As mentioned above, our construction adapts the general principles of the Wavelet-Vaguelette Decomposition [9] and is similar to the Curvelet Biorthogonal Decomposition given in [4].

**Theorem 3.3** *Let  $\{s_\mu : \mu \in M\}$  be the Parseval frame of shearlets defined above and  $\{U_\mu : \mu \in M\}$  be the systems defined by (9). For all  $f \in L^2(\mathbb{R}^2)$  the following reproducing formula holds:*

$$f = \sum_{\mu} [Rf, U_\mu] 2^j s_\mu.$$

**Proof.** This proof is very similar to the one in [4] and is based on the intertwining relation:

$$R \circ (-\Delta)^\alpha = (I \otimes D^{2\alpha}) \circ R.$$

Direct computations show that:

$$\begin{aligned}
\langle f, s_\mu \rangle &= [\mathcal{R}f, \mathcal{R} s_\mu] \\
&= [(I \otimes D^{\frac{1}{2}}) \circ Rf, (I \otimes D^{\frac{1}{2}}) \circ R s_\mu] \\
&= [Rf, (I \otimes D^{\frac{1}{2}}) \circ (I \otimes D^{\frac{1}{2}}) \circ R s_\mu] \\
&= [Rf, (I \otimes D^{\frac{1}{2}}) \circ R \circ (-\Delta)^{\frac{1}{4}} s_\mu] \\
&= [Rf, \mathcal{R}(2^j s_\mu^+)] \\
&= 2^j [Rf, U_\mu]. \quad \square
\end{aligned}$$

## 4 Inversion of Noisy Radon Data

In our model, we assume that Radon transform data are corrupted by white Gaussian noise, that is, we have the observations:

$$Y = Rf + \epsilon W, \quad (10)$$

where  $f$  is the function to be recovered,  $W$  is a Wiener sheet and  $\epsilon$  is measuring the noise level. This means that each measurement  $[Y, U_\mu]$  of the observed data is normally distributed with mean  $[Rf, U_\mu]$  and variance  $\epsilon^2 \|U_\mu\|_{L^2((0,2\pi) \times \mathbb{R})}^2$ .

While the white noise model does not precisely describes the types of noise typically found in practical applications, the asymptotic theory derived from this assumption has been found to lead to very acceptable results in practice. In addition, this framework allows one to derive a theoretical assessment of the performance of the method which would be extremely complicated to handle otherwise.

In order to obtain an upper bound on the risk of the estimator, it is necessary to specify the type of functions we are dealing with. Following [4], let  $A$  be a positive constant and  $STAR^2(A)$  be a class of indicator functions of sets  $B$  with  $C^2$  boundaries  $\partial B$  satisfying the following conditions. In polar coordinates, let  $\rho(\theta) : [0, 2\pi) \rightarrow [0, 1]^2$  be a radius function and define  $B$  by  $x \in B$  if and only if  $|x| \leq \rho(\theta)$ . In particular, the boundary  $\partial B$  is given by the curve in  $\mathbb{R}^2$ :

$$\beta(\theta) = \begin{pmatrix} \rho(\theta) \cos(\theta) \\ \rho(\theta) \sin(\theta) \end{pmatrix}. \quad (11)$$

The class of boundaries of interest to us are defined by

$$\sup |\rho''(\theta)| \leq A, \quad \rho \leq \rho_0 < 1. \quad (12)$$

We say that a set  $B \in STAR^2(A)$  if  $B \subset [0, 1]^2$  and  $B$  is a translate of a set whose boundary obeys (11) and (12). In addition, we set  $C_0^2([0, 1]^2)$  to be the collection of twice differentiable functions supported inside  $[0, 1]^2$ . Finally, we

define the set  $\mathcal{E}^2(A)$  of *functions which are  $C^2$  away from a  $C^2$  edge* as the collection of functions of the form

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

where  $f_0, f_1 \in C_0^2([0, 1]^2)$ ,  $B \in STAR^2(A)$  and  $\|f\|_{C^2} = \sum_{|\alpha| \leq 2} \|D^\alpha f\|_\infty \leq 1$ .

Projecting the data (10) onto the frame  $\{U_\mu : \mu \in M\}$ , and rescaling, we obtain

$$\begin{aligned} y_\mu &:= 2^j [Y, U_\mu] \\ &= 2^j [Rf, U_\mu] + \epsilon 2^j [W, U_\mu] \end{aligned} \quad (13)$$

$$= \langle f, \psi_\mu \rangle + \epsilon 2^j n_\mu, \quad (14)$$

where  $n_\mu$  is a (non-i.i.d) Gaussian noise with zero mean and variance  $\sigma_\mu = \|\psi_\mu^+\|_2$ . In order to estimate  $f$ , we need to estimate the shearlet coefficients  $\langle f, \psi_\mu \rangle$ ,  $\mu \in M$ , from the data  $y_\mu$ . To achieve that, we will devise a thresholding rule to be applied to  $\{y_\mu : \mu \in M\}$ , which exploits the sparsity properties of the shearlet representation.

#### 4.1 Modified Shearlet System

For the application of the shearlet representation to the estimation problem, it is useful to introduce the following simple variant of the shearlet system given in Section 2, which is obtained by rescaling the coarse scale system and changing the range of scales for which the directional fine-scale system is defined. Namely, for a fixed  $j_0 \in N$ , let  $\varphi \in C_0^\infty(\mathbb{R}^2)$  be such that

$$\begin{aligned} 1 &= |\hat{\varphi}(2^{-j_0} \xi)|^2 + \sum_{j \geq j_0} \sum_{\ell = -2^j}^{2^j - 1} |\hat{\psi}^{(0)}(\xi A_0^{-j} B_0^{-\ell})|^2 \chi_{\mathcal{D}_0}(\xi) \\ &\quad + \sum_{j \geq j_0} \sum_{\ell = -2^j}^{2^j - 1} |\hat{\psi}^{(1)}(\xi A_1^{-j} B_1^{-\ell})|^2 \chi_{\mathcal{D}_1}(\xi). \end{aligned}$$

Then we obtain the Parseval frame of shearlets:

$$\begin{aligned} &\{2^{j_0} \varphi(2^{j_0} x - k) : k \in \mathbb{Z}^2\} \cup \{\tilde{\psi}_{j,\ell,k}^{(d)}(x) : j \geq j_0, \ell = -2^j, 2^j - 1, k \in \mathbb{Z}^2, d = 0, 1\} \\ &\cup \{\psi_{j,\ell,k}^{(d)}(x) : j \geq j_0, -2^j + 1 \leq \ell \leq 2^j - 2, k \in \mathbb{Z}^2, d = 0, 1\}. \end{aligned}$$

Similarly to the original shearlet system given in Theorem 2.1, the modified shearlet system is made of coarse and fine scale systems, with the coarse scale

system now associated with the coarse scale  $j_0$ . Similarly to what we did above, we introduce the index set  $\mathcal{M}^0 = N \cup M^0$ , where  $N = \mathbb{Z}^2$ ,  $M^0 = \{\mu = (j, \ell, k, d) : j \geq j_0, -2^j \leq \ell \leq 2^j - 1, k \in \mathbb{Z}^2, d = 0, 1\}$ , and denote the new shearlet system using the compact notation  $\{s_\mu : \mu \in \mathcal{M}^0\}$ , where  $s_\mu = \psi_\mu = \psi_{j,\ell,k}^{(d)}$  if  $\mu \in M^0$  and  $s_\mu = 2^{j_0} \varphi(2^{j_0} x - \mu)$  if  $\mu \in N$ . For  $\psi_\mu$ ,  $\mu \in M^0$ , it is understood that the corner elements are modified as in Theorem 2.1.

In our construction, the selection of the scale  $j_0$  will depend on the noise level  $\epsilon$ . Namely, we set  $j_0 = \frac{2}{15} \log_2(\epsilon^{-1})$ . We also introduce the scale index  $j_1 = \frac{2}{5} \log_2(\epsilon^{-1})$  (so that  $2^{j_0} = \epsilon^{-2/15}$  and  $2^{j_1} = \epsilon^{-2/5}$ ).

Hence, depending on the noise level  $\epsilon$ , we can now define the *set of significant coefficients* in the shearlet representation of a function  $f \in \mathcal{E}^2(A)$ . These are the elements  $\langle f, s_\mu \rangle$  for which  $\mu$  belongs to the *set of significant indices*  $\mathcal{N}(\epsilon) \subset \mathcal{M}^0$  which is given by the union  $\mathcal{N}(\epsilon) = M_1(\epsilon) \cup N_0(\epsilon)$ , where

$$N_0(\epsilon) = \{\mu = k \in \mathbb{Z}^2 : |k| \leq 2^{2j_0+1}\};$$

$$M_1(\epsilon) = \{\mu = (j, \ell, k, d) : j_0 < j \leq j_1, |k| \leq 2^{2j_0+1}, d = 0, 1\}.$$

We have the following result which is proved in the Appendix.

**Theorem 4.1** *Let  $\epsilon$  denote the noise level, and  $\mathcal{N}(\epsilon)$  be the set of significant indices associated with the shearlet representation of  $f$  given by*

$$f = \sum_{\mu \in \mathcal{M}^0} \langle f, s_\mu \rangle s_\mu.$$

*The following properties hold:*

(1) *The neglected shearlet coefficients  $\{\langle f, s_\mu \rangle : \mu \notin \mathcal{N}(\epsilon)\}$  satisfy:*

$$\sup_{f \in \mathcal{E}^2(A)} \sum_{\mu \notin \mathcal{N}(\epsilon)} |\langle f, s_\mu \rangle|^2 \leq C' \epsilon^{4/5}.$$

(2) *The risk proxy satisfies:*

$$\sup_{f \in \mathcal{E}^2(A)} \sum_{\mu \in \mathcal{N}(\epsilon)} \min(|\langle f, s_\mu \rangle|^2, 2^{2j} \epsilon^2) \leq C'' \epsilon^{4/5}.$$

(3) *The cardinality of  $\mathcal{N}(\epsilon)$  obeys:*

$$\#\mathcal{N}(\epsilon) \leq C''' \epsilon^{-2}.$$

$C', C'', C'''$  are positive constants independent of  $f$ .

## 4.2 Estimation Rate

To estimate  $f$  from the noisy observations (14), we will apply the soft thresholding function  $T_s(y, t) = \text{sgn}(y)(|y| - t)_+$ . The analysis of the estimation error follows the general framework of the wavelet shrinkage developed in [10]. Letting  $\#\mathcal{N}(\epsilon)$  be the number of significant coefficients of the shearlet representation of  $f$ , we estimate function  $f$  by

$$\tilde{f} = \sum_{\mu \in \mathcal{M}^0} \tilde{c}_\mu s_\mu, \quad (15)$$

where the coefficients are obtained by the rule

$$\tilde{c}_\mu = \begin{cases} T_s(y_\mu, \epsilon \sqrt{2 \log(\#\mathcal{N}(\epsilon))} 2^j \sigma_\mu), & \mu \in \mathcal{N}(\epsilon) \\ 0 & \text{otherwise.} \end{cases} \quad (16)$$

and  $\sigma_\mu = \|s_\mu^+\|_2$ . Notice that the terms  $\sigma_\mu$ ,  $\mu \in \mathcal{M}^0$ , are uniformly bounded.

The main theorem can now be established.

**Theorem 4.2** *Let  $f \in \mathcal{E}^2(A)$  be the solution of the problem  $Y = Rf + \epsilon W$  and  $\tilde{f}$  be the approximation to  $f$  given by equations (15)-(16). Then there is a constant  $C > 0$  such that*

$$\sup_{\mathcal{E}^2(A)} E \|\tilde{f} - f\|_2^2 \leq C \log(\epsilon^{-1}) \epsilon^{4/5}, \quad \text{as } \epsilon \rightarrow 0.$$

*In the expression above,  $E$  is the expectation operator.*

**Proof.** For  $\mu \in \mathcal{M}^0$ , let  $c_\mu = \langle f, s_\mu \rangle$  and  $\tilde{c}_\mu$  be given by (16). By the Parseval frame property of the shearlet system  $\{s_\mu : \mu \in \mathcal{M}^0\}$ , it follows that

$$\left\| \sum_{\mu \in \mathcal{M}^0} c_\mu s_\mu \right\|_2^2 \leq \sum_{\mu \in \mathcal{M}^0} |c_\mu|^2,$$

and that

$$E \|\tilde{f} - f\|_2^2 \leq E \left( \sum_{\mu \in \mathcal{M}^0} |\tilde{c}_\mu - c_\mu|^2 \right). \quad (17)$$

On the other hand, by the oracle inequality [10] we have

$$E \left( \sum_{\mu \in \mathcal{N}(\epsilon)} |\tilde{c}_\mu - c_\mu|^2 \right) \leq L(\epsilon) \left( \epsilon^2 \sum_{\mu \in \mathcal{N}(\epsilon)} \left( \frac{2^{2j} \sigma_\mu^2}{\#\mathcal{N}(\epsilon)} + \min(c_\mu^2, \epsilon^2 2^{2j} \sigma_\mu^2) \right) \right) \quad (18)$$

where  $L(\epsilon) = (1 + 2\log(\#\mathcal{N}(\epsilon)))$ . Now observe that, by Theorem 4.1, there exists positive constants  $C', C'', C'''$  such that

$$\begin{aligned} \sum_{\mu \in \mathcal{N}(\epsilon)} \min(c_\mu^2, \epsilon^2 2^{2j} \sigma_\mu^2) &\leq C' \epsilon^{4/5}, \\ \sum_{\mu \in \mathcal{N}(\epsilon)^c} c_\mu^2 &\leq C'' \epsilon^{4/5}, \\ \log(\#\mathcal{N}(\epsilon)) &\leq C''' \log(\epsilon^{-1}). \end{aligned}$$

By the assumption on  $\mathcal{N}(\epsilon)$ , we have that there is a constant  $C_1 > 0$  such that

$$\epsilon^2 \sum_{\mu \in \mathcal{N}(\epsilon)} 2^{2j} \frac{\sigma_\mu^2}{\#\mathcal{N}(\epsilon)} \leq C_1 \epsilon^2 2^{2j_1} = C_1 \epsilon^{6/5}.$$

Thus, using these observations and equations (17) and (18), we obtain that there is a constant  $C > 0$  such that

$$E\|\tilde{f} - f\|_2^2 \leq E \left( \sum_{\mu \in \mathcal{N}(\epsilon)} |\tilde{c}_\mu - c_\mu|^2 \right) + \sum_{\mu \in \mathcal{N}(\epsilon)^c} c_\mu^2 \leq C \log(\epsilon^{-1}) \epsilon^{4/5}. \quad \square$$

As also observed in [4], Theorem 4.2 remains valid if the soft thresholding operator  $T_s(y, t)$  is replaced by the hard thresholding operator  $T_h(y, t) = y \chi_{|y| \geq t}$ . In fact, also in the case of hard thresholding, one can obtain estimates similar to (18).

Finally, for completeness, we recall the following theorem from [4] showing that the rate of convergence of our estimator is near optimal; no estimator can achieve an essentially better rate uniformly over  $\mathcal{E}^2(A)$ .

**Theorem 4.3** ([4]) *Let  $f \in \mathcal{E}^2(A)$  and consider the minimax mean square error*

$$\mathcal{M}(\epsilon, \mathcal{E}^2(A)) = \inf_{\tilde{f}} \sup_{f \in \mathcal{E}^2(A)} E\|\tilde{f} - f\|_2^2.$$

*This satisfies*

$$\mathcal{M}(\epsilon, \mathcal{E}^2(A)) \geq C \epsilon^{4/5} (\log(\epsilon^{-1}))^{-2/5}, \quad \epsilon \rightarrow 0,$$

*for some  $C \in \mathbb{R}^+$ .*

## 5 Numerical Experiments

In this section, we demonstrate the derived estimation rates and compare our proposed shearlet estimation method against a curvelet and wavelet-based ver-

sion. The numerical implementations of curvelets and shearlets are described in [3] and [12], respectively.

In the first experiment, we used the star image shown in the upper left-hand side of Figure 2. White Gaussian noise with zero mean was added to the Radon transform of the image with various levels of standard deviations and the inversion processes was regularized by the proposed method. For comparison, a version using curvelets and wavelets was also included.

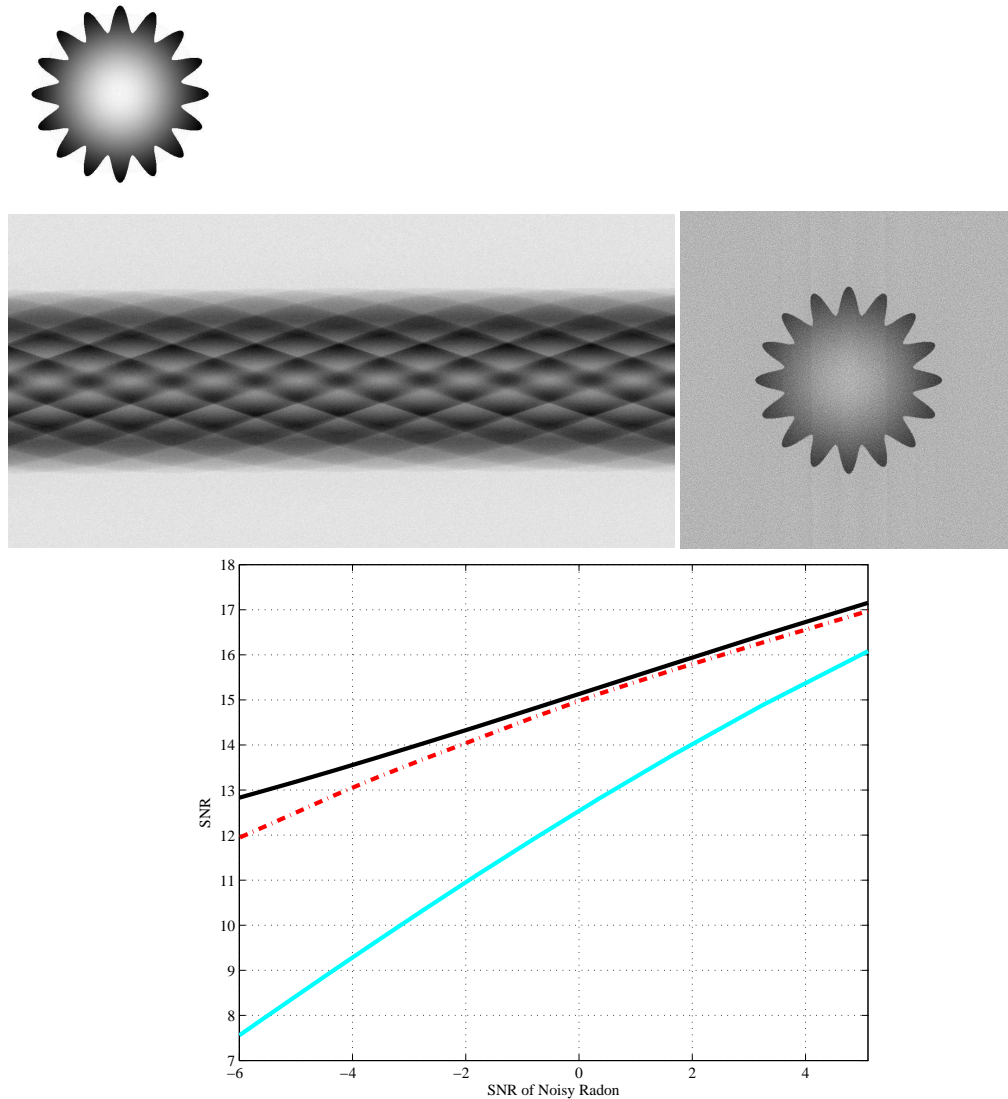


Fig. 2. SNR performances of shearlet-based estimate (solid line), curvelet-based estimate (dash-dot line), and wavelet-based estimate (solid-gray line) as a function of SNR of the unfiltered reconstruction.

For another set of experiments, we tested both a soft thresholding and hard thresholding version for each of the methods using a modified Shepp-Logan Phantom image and an MRI brain scan image.

For the soft thresholding  $T_s(y, t)$ , we estimated the standard deviation  $\tilde{\sigma}_\mu$  of  $[Y, U_\mu]$  and set  $t = 2^j \tilde{\sigma}_\mu \sqrt{2 \log(2^j)}$ . For the hard thresholding, Monte Carlo simulations were used to estimate the standard deviation of the noise through each subband and a scalar multiple of the estimated standard deviations was chosen as the threshold value. In particular, the threshold for the the finest scales was chosen to be four times the estimated standard deviation of the noise and the threshold for the remaining scales except for the coarsest scale was chosen to be three times the estimated standard deviations. To assess the performance, we used the measure

$$SNR(f, f_{\text{est}}) = 10 \log_{10} \left[ \frac{\text{var}(f)}{\text{mean}(f - f_{\text{est}})} \right],$$

where  $f, f_{\text{est}}$ , and  $\text{var}(f)$  are the original image, the estimated image, and the variance of the image, respectively. Both the  $SNR$  of the noisy projections (treated as an image) and the unfiltered inversion are given in Tables I and II. Some results are also displayed in Figures 3, 4, and 5.

TABLE I - Estimates based on soft-thresholding  
(Performance in SNR)

Radon dom.	Unfiltered	Shearlet	Curvelet	Wavelet
Star image				
41.38 dB	15.27 dB	20.46 dB	19.87 dB	19.70 dB
35.35 dB	10.68 dB	19.36 dB	18.63 dB	18.08 dB
31.83 dB	7.48 dB	18.62 dB	17.86 dB	17.14 dB
Phantom				
39.63 dB	14.08 dB	18.82 dB	18.60 dB	17.81 dB
33.62 dB	9.23 dB	17.62 dB	17.45 dB	16.25 dB
30.09 dB	5.97 dB	16.80 dB	16.68 dB	15.36 dB
Brain scan				
41.34 dB	15.01 dB	22.08 dB	21.36 dB	20.89 dB
35.32 dB	9.39 dB	20.79 dB	20.09 dB	19.43 dB
31.80 dB	5.95 dB	19.88 dB	19.25 dB	18.57 dB

As subtle artifacts remained after reconstruction from the shearlet-based estimates, we applied the following additional filtering scheme which is based on Total Variation minimization and is described in [11]. This post-filtering applies only to the hard thresholding estimate and can be roughly described as follows. Denote by  $M^C$  the set of indices of  $M$  in the shearlet domain that correspond to the coefficients that would be set to zero in the above reconstruction (namely, the indices of the thresholded coefficients). Define the

projection operator  $P_S$  onto the reconstruction from these coefficients as

$$P_S(u) = \sum_{j,\ell,k \in M^C} \langle u, \psi_{j,\ell,k} \rangle \psi_{j,\ell,k}.$$

The proposed method is then to essentially solve

$$\frac{\partial u}{\partial t} = \nabla \cdot \left( \frac{\phi'(\|\nabla P_S(u)\|)}{\|\nabla P_S(u)\|} \nabla P_S(u) \right) - \lambda_{x,y}(u - u_0)$$

with the boundary condition  $\frac{\partial u}{\partial n} = 0$  on  $\partial\Omega$  and the initial condition  $u(x, y, 0) = u_0(x, y)$  for  $x, y \in \Omega$ , where  $\Omega$  is the image domain. The quantity  $\lambda_{x,y}$  is a spatially varying penalty term based on a measure of local variances that is updated after a number of iterations or progressions of artificial time step. For this post-processing step, we set  $\lambda_{x,y} = 0$  and  $u_0(x, y)$  to be the initial shearlet-based estimate.

Notice that, as shown in Table II, this post-filtering scheme produces a slight improvement in performance. However, the numerical tests show that, even without this filtering, the shearlet based method would still outperform the corresponding wavelet- and curvelet-based tests.

TABLE II - Estimates based on hard-thresholding  
(Performance in SNR)

Radon dom.	Unfiltered	Shear+TV	Shearlet	Curvelet	Wavelet
Star image					
41.38 dB	15.27 dB	21.69 dB	21.60 dB	20.73 dB	20.72 dB
35.35 dB	10.68 dB	20.75 dB	20.67 dB	19.62 dB	19.03 dB
31.83 dB	7.48 dB	20.07 dB	20.01 dB	18.98 dB	17.88 dB
Phantom					
39.63 dB	14.08 dB	20.27 dB	20.21 dB	19.41 dB	18.59 dB
33.62 dB	9.23 dB	19.07 dB	18.99 dB	18.27 dB	16.92 dB
30.09 dB	5.97 dB	18.29 dB	18.23 dB	17.78 dB	15.89 dB
Brain scan					
41.34 dB	15.01 dB	23.26 dB	22.98 dB	22.22 dB	21.66 dB
35.32 dB	9.39 dB	21.86 dB	21.55 dB	20.85 dB	20.14 dB
31.80 dB	5.95 dB	20.82 dB	20.49 dB	19.92 dB	19.08 dB

For the final demonstrations of the methods, we tested the performance using an ISAR dataset collected by System Planning Corporation’s Mark V radar of a SAAB 9000 car. The collected data (Radon projections) and an image of the unfiltered reconstruction are shown in Figure 6. The hard thresholding

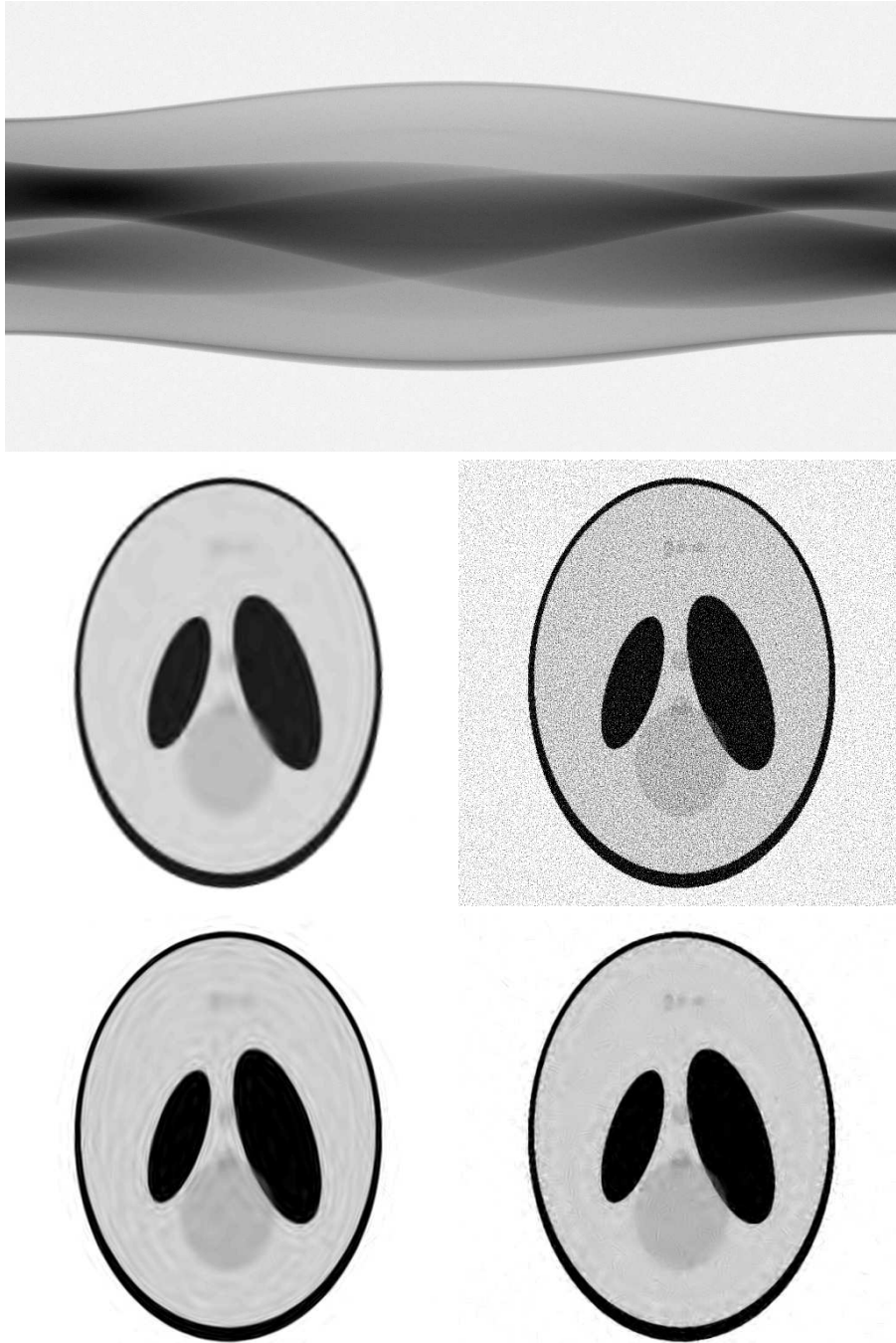


Fig. 3. From the top, clockwise: noisy Radon projections (SNR=30.09 dB); unfiltered reconstruction (SNR=5.97 dB); shearlet and TV-based estimate (SNR=18.29 dB); curvelet-based estimate (SNR=17.78 dB); wavelet-based estimate (SNR=15.89 dB).

values were determined using the noise level found by means of a median estimator. Since there is no reference image to compare against, no measure of performance can be given. Yet, it is clear the shearlet-based estimate provides the best looking estimate. Note that some of the noise present in ISAR imagery can be attributed to speckle but when displayed on a logarithmic scale as

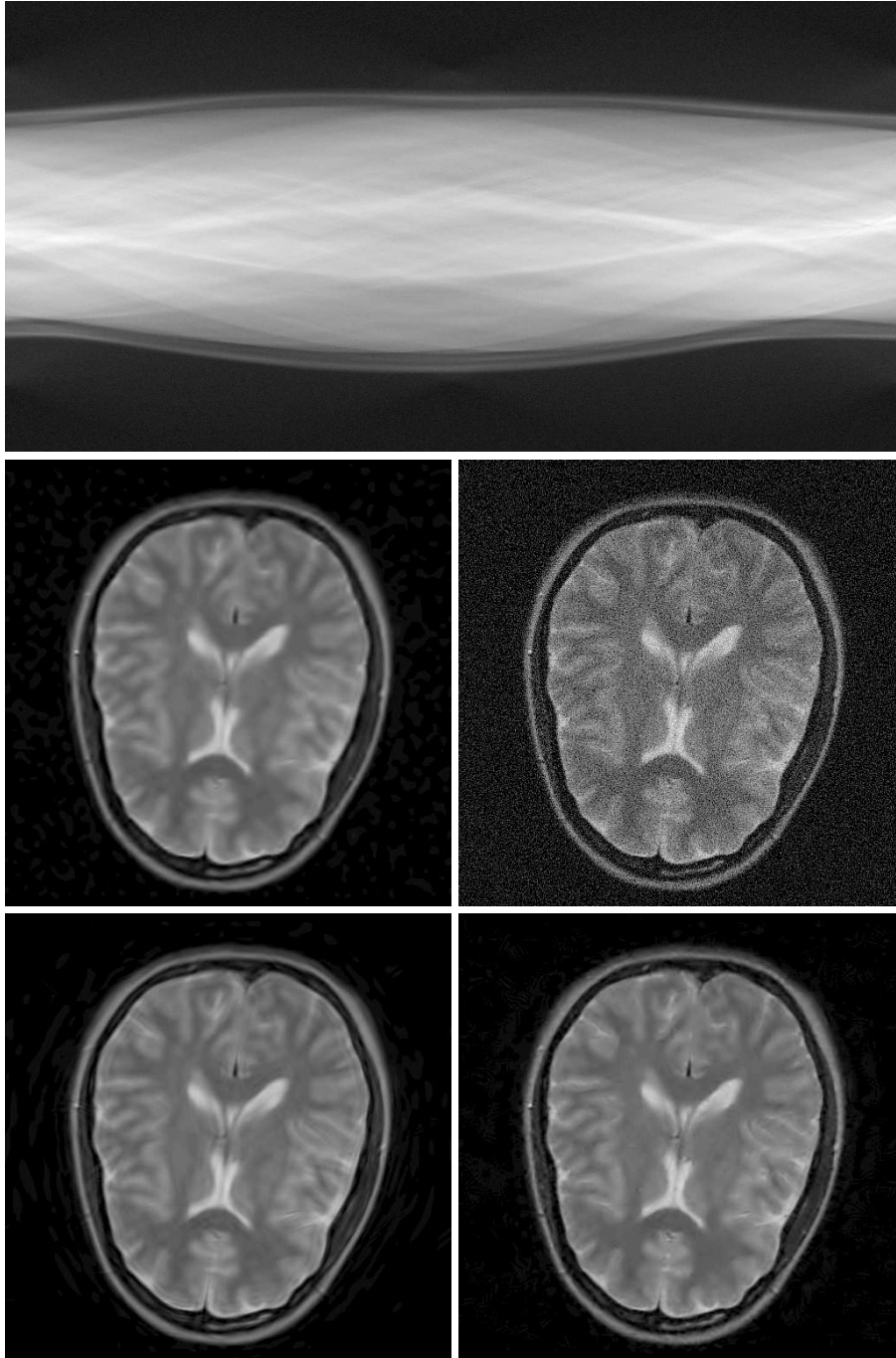


Fig. 4. From the top, clockwise: noisy Radon projections (SNR=35.32 dB); unfiltered reconstruction (SNR=9.39 dB); shearlet and TV-based estimate (SNR=21.86 dB); curvelet-based estimate (SNR=20.85 dB); wavelet-based estimate (SNR=20.14 dB).

shown in Figure 6, the noise can be approximated as additive Gaussian [1].

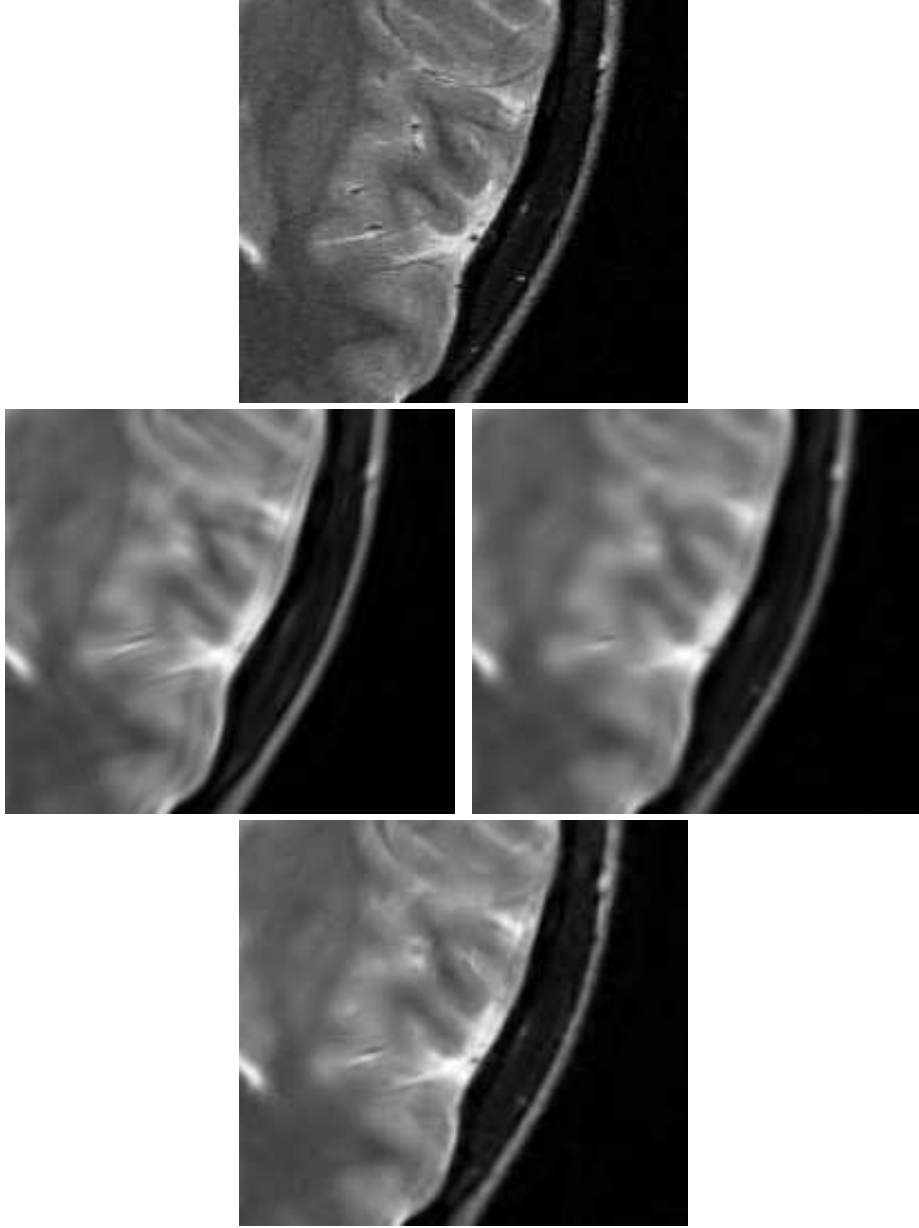


Fig. 5. Close ups from the top, clockwise: original, wavelet-based estimate (SNR=20.14 dB); shearlet and TV-based estimate (SNR=21.86 dB); curvelet-based estimate (SNR=20.85 dB).

## 6 Appendix. Proofs from Section 4

We are considering the modified shearlet system  $\{s_\mu : \mu \in \mathcal{M}^0\}$  introduced in Section 4.1. Recall that  $\mathcal{M}^0 = N \cup M^0$ , where  $N = \mathbb{Z}^2$ ,  $M^0 = \{\mu = (j, \ell, k, d) : j \geq j_0, -2^j \leq \ell \leq 2^j - 1, k \in \mathbb{Z}^2, d = 0, 1\}$ , and the system is made of the coarse scale system  $\{s_\mu = 2^{j_0} \varphi(2^{j_0} x - \mu) : \mu \in N\}$  and the fine scale system  $\{s_\mu = \psi_\mu = \psi_{j,\ell,k}^{(d)} : \mu \in M^0\}$ .

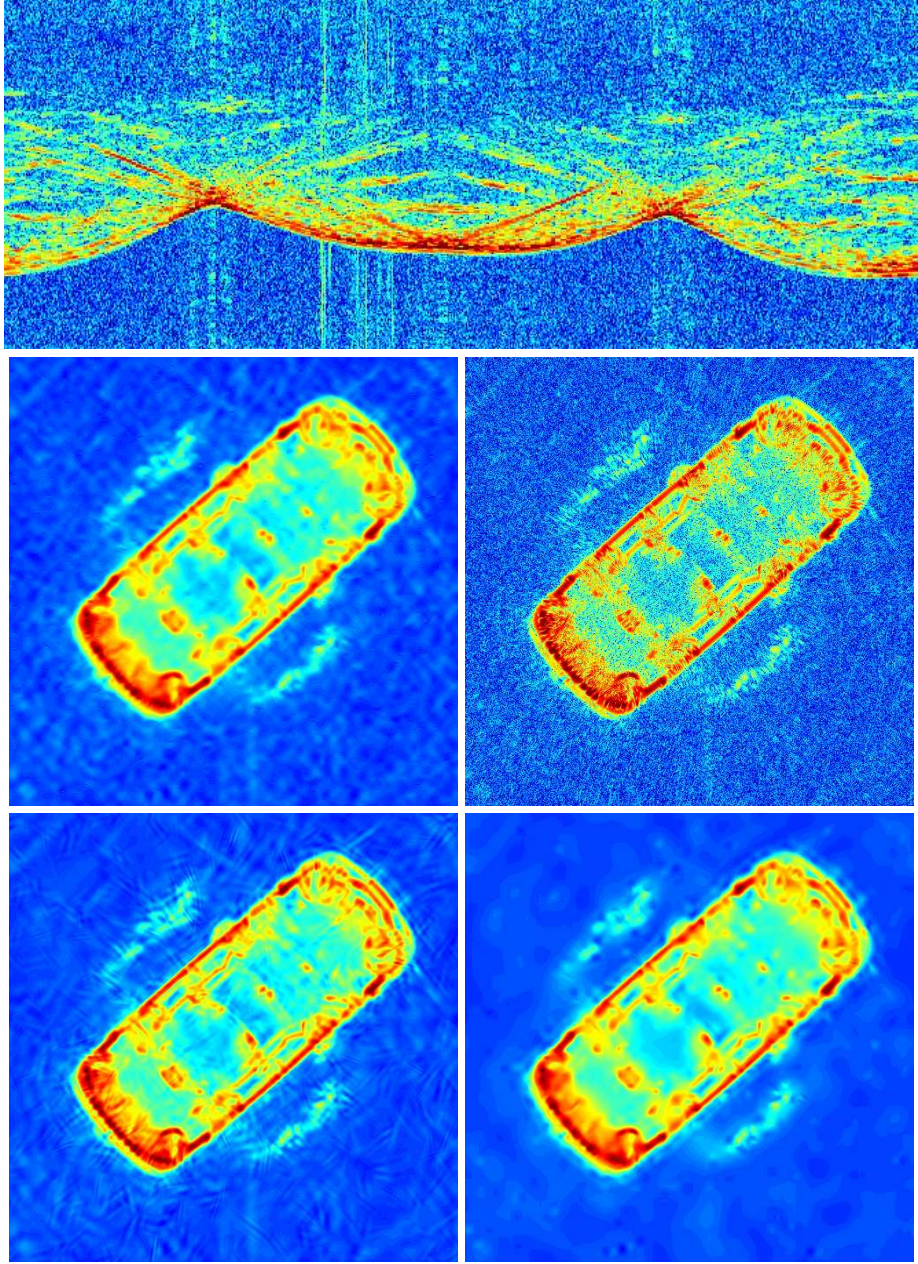


Fig. 6. From the top, clockwise: noisy Inverse Synthetic Aperture Radar (Radon) projections; unfiltered reconstruction; shearlet-based estimate; curvelet-based estimate; wavelet-based estimate.

In order to prove Theorem 4.1, we need the following lemma which provides an estimate for the size of the shearlet coefficients at a fixed scale  $j$  (where  $j \geq j_0$ ). For such  $j$  fixed, let  $M_j = \{(j, \ell, k, d) : -2^j \leq \ell \leq 2^j, k \in \mathbb{Z}^2\}$ . We have:

**Lemma 6.1** *Let  $f \in \mathcal{E}^2(A)$  and  $j \geq j_0$ . Then there is a positive constant  $C$*

such that

$$\sum_{\mu \in M_j} |\langle f, \psi_\mu \rangle|^2 \leq C 2^{-2j}$$

**Proof.**

It is useful to introduce a smooth localization of the function  $f$  near dyadic squares. Let  $\mathcal{Q}_j$  be the collection of dyadic squares of the form  $Q = [\frac{\nu_1}{2^j}, \frac{\nu_1+1}{2^j}] \times [\frac{\nu_2}{2^j}, \frac{\nu_2+1}{2^j}]$ , with  $\nu_1, \nu_2 \in \mathbb{Z}$ . For a nonnegative  $C^\infty$  function  $w$  with support in  $[-1, 1]^2$ , we define a smooth partition of unity

$$\sum_{Q \in \mathcal{Q}_j} w_Q^2(x) = 1, \quad x \in \mathbb{R}^2,$$

where, for each dyadic square  $Q \in \mathcal{Q}_j$ ,  $w_Q(x) = w(2^j x_1 - \nu_1, 2^j x_2 - \nu_2)$ .

Given  $f \in \mathcal{E}^2(A)$ , the coefficients  $\{\langle f, \psi_\mu \rangle\}$  will exhibit a very different behavior depending on whether the edge curve of  $f$  intersects the support of  $w_Q$  or not. We split  $\mathcal{Q}_j$  into the disjoint sets  $\mathcal{Q}_j^0$  and  $\mathcal{Q}_j^1$  that indicate whether the collection of dyadic squares  $Q$  intersects an edge curve or not. Since each dyadic square  $Q$  has sidelength  $2 \cdot 2^{-j}$  and  $f$  has compact support in  $[0, 1]^2$ , there are  $O(2^j)$  dyadic cubes in  $Q \in \mathcal{Q}_j^0$  intersecting the edge curve and  $O(2^{2j})$  dyadic cubes in  $Q \in \mathcal{Q}_j^1$  not intersecting the edge curve.

For each such cube  $Q \in \mathcal{Q}_j^0$ , it is shown in [13] that

$$\sum_{k \in \mathbb{Z}^2} |\langle f_Q, \psi_\mu \rangle|^2 \leq C 2^{-3j} (1 + |\ell|)^{-5}.$$

More precisely, in the proof of Theorem 1.3 in [13], it is shown that

$$\sum_{k \in R_K} |\langle f_Q, \psi_\mu \rangle|^2 \leq C L_K^{-2} 2^{-3j} (1 + |\ell|)^{-5},$$

where  $K \in \mathbb{Z}^2$ ,  $\cup_{K \in \mathbb{Z}^2} R_K = \mathbb{Z}^2$  and  $\sum_{K \in \mathbb{Z}^2} L_K^{-2} < \infty$ . Hence

$$\sum_{\mu \in M_j} |\langle f_Q, \psi_\mu \rangle|^2 = \sum_{|\ell| \leq 2^j} \sum_{k \in \mathbb{Z}^2} |\langle f_Q, \psi_\mu \rangle|^2 \leq C 2^{-3j}.$$

Adding up over all cubes in  $\mathcal{Q}_j^0$ , we have that

$$\sum_{\mu \in M_j} \sum_{Q \in \mathcal{Q}_j^0} |\langle f_Q, \psi_\mu \rangle|^2 \leq C 2^{-2j}. \quad (19)$$

For  $Q \in \mathcal{Q}_j^1$ , it is shown in [13, Thm.1.4] that

$$\sum_{\mu \in M_j} |\langle f_Q, \psi_\mu \rangle|^2 \leq C 2^{-4j}.$$

Hence adding up over all  $Q \in \mathcal{Q}_j^1$ , it follows that

$$\sum_{\mu \in M_j} \sum_{Q \in \mathcal{Q}_j^1} |\langle f_Q, \psi_\mu \rangle|^2 \leq C 2^{-2j}. \quad (20)$$

The proof is completed by combining the estimates (19) and (20).  $\square$

We now proceed with the proof of Theorem 4.1.

**Proof.** (Theorem 4.1.1(1)) We need to establish

$$\sup_{f \in \mathcal{E}^2(A)} \sum_{\mu \notin \mathcal{N}(\epsilon)} |\langle f, s_\mu \rangle|^2 \leq C' \epsilon^{4/5}.$$

(I) We start by examining the situation at fine scales, for  $j \geq j_1(\epsilon) = \frac{2}{5} \log_2(\epsilon^{-1})$ , so that  $2^{-j} \leq \epsilon^{\frac{2}{5}}$ . Notice that, for these values of the index  $j$ ,  $s_\mu = \psi_\mu$ .

By Lemma 6.1, for each  $f \in \mathcal{E}^2(A)$  and each  $j \geq j_0$ , we have that

$$\sum_{\mu \in M_j} |\langle f, \psi_\mu \rangle|^2 \leq C 2^{-2j}.$$

Hence

$$\sum_{j > j_1} \sum_{\mu \in M_j} |\langle f, \psi_\mu \rangle|^2 \leq C \sum_{j > j_1} 2^{-2j} \leq C \epsilon^{\frac{4}{5}}.$$

(II) Let  $Q_0 = [0, 1]^2$  and  $\text{supp } f \subset Q_0$ . We will show that the terms  $\langle f, s_\mu \rangle$  decay very rapidly for locations  $k$  away from  $Q_0$ .

We will start by examining the decay of a fine-scale term  $\langle f, s_\mu \rangle$ , for  $\mu = (j, \ell, k, d) \in M^0$  and  $d = 0$ . Let  $E_{j\ell} = B^\ell A^j = \begin{pmatrix} 2^{2j} & \ell 2^j \\ 0 & 2^j \end{pmatrix}$ . By the assumptions on  $\psi$ , it follows that, for each  $m \in \mathbb{N}$ , there is a constant  $C_m > 0$  such that

$$|\psi(x)| \leq C_m (1 + \|x\|)^{-m}. \quad (21)$$

It follows that

$$|\psi_{j,\ell,k}^{(0)}(x)| \leq C_m 2^{3j/2} (1 + \|E_{j\ell}x - k\|)^{-m}.$$

We will use two simple facts. The first one is  $\|E_{j\ell}x\| \leq \|E_{j\ell}\|_{op} \|x\| = 2^{2j} \|x\| \leq \sqrt{2} 2^{2j}$  for  $x \in Q_0$  and the second one is that for  $a > 0$ ,  $0 \leq b \leq c \leq a$ , we have  $a - b \geq a - c$ .

It follows that, for  $|k| \geq 2^{2j+1}$ , we have

$$\begin{aligned}
|\langle f, \psi_{j,\ell,k}^{(0)} \rangle| &\leq \|f\|_\infty \int_Q |\psi_{j,\ell,k}^{(0)}(x)| dx \\
&\leq C_m 2^{\frac{3}{2}j} \int_{Q_0} \left(1 + \|E_{j\ell}x - k\|\right)^{-m} dx \\
&\leq C_m 2^{\frac{3}{2}j} \int_{Q_0} \left(1 + |k| - \|E_{j\ell}x\|\right)^{-m} dx \\
&\leq C_m 2^{\frac{3}{2}j} \int_{Q_0} (|k| - 2^{2j}\|x\|)^{-m} dx \\
&\leq C_m 2^{\frac{3}{2}j} (|k| - \sqrt{2} 2^{2j})^{-m}.
\end{aligned} \tag{22}$$

Thus,

$$\begin{aligned}
\sum_{\ell=-2^j}^{2^j-1} \sum_{|k| \geq 2^{2j+1}} |\langle f, \psi_{j,\ell,k}^{(0)} \rangle|^2 &\leq C_m \sum_{\ell=-2^j}^{2^j-1} 2^{3j} \sum_{|k| \geq 2^{2j+1}} (|k| - \sqrt{2} 2^{2j})^{-2m} \\
&\leq C_m 2^{4j} \sum_{|k| \geq 2^{2j+1}} (|k| - \sqrt{2} 2^{2j})^{-2m} \\
&\leq C_m 2^{4j} 2^{-2j(2m-2)} \\
&= C_m 2^{8j} 2^{-4jm}
\end{aligned}$$

Now we can add up all contributions for  $j \geq j_0$ . Since we can choose  $m$  arbitrarily, we have:

$$\sum_{j \geq j_0} \sum_{\ell=-2^j}^{2^j-1} \sum_{|k| \geq 2^{2j+1}} |\langle f, \psi_{j,\ell,k}^{(0)} \rangle|^2 \leq C_m \sum_{j \geq j_0} 2^{8j} 2^{-4jm} \leq C 2^{-6j_0} = C \epsilon^{\frac{4}{5}}.$$

The analysis in the case where  $\mu \in M^0$  and  $d = 1$  is essentially the same as the one given above. For the coarse case terms, notice first that  $\phi$  satisfies the same decay behavior as (21) for  $\psi$ . Hence, letting  $\phi_{j_0,k}(x) = 2^{j_0}\phi(2^{j_0}x - k)$  and proceeding as in (22), we have that

$$\begin{aligned}
|\langle f, \phi_{j_0,k} \rangle| &\leq \|f\|_\infty \int_Q |\phi_{j_0,k}(x)| dx \\
&\leq C_m 2^{j_0} (|k| - \sqrt{2} 2^{j_0})^{-m}.
\end{aligned} \tag{23}$$

Now we can proceed as above, by summing over  $|k| \geq 2^{2j+1}$  and using the fact

that  $m$  can be chosen arbitrarily, to conclude that

$$\sum_{|k| \geq 2^{2j+1}} |\langle f, \phi_{j_0, k} \rangle|^2 \leq C \epsilon^{\frac{4}{5}}.$$

Combining the estimates from parts (I) and (II) and of the proof, we finally have that

$$\sum_{\mu \notin \mathcal{N}(\epsilon)} |\langle f, s_\mu \rangle|^2 \leq C \epsilon^{4/5}. \quad \square$$

**Proof.** (Theorem 4.1.1(2))

For  $\mu \in \mathcal{M}^0$ , we use the notation  $c_\mu = \langle f, s_\mu \rangle$  and we define the set

$$R(j, \epsilon) = \{\mu \in M_j : |c_\mu| > \epsilon\},$$

to denote the set of “large” shearlet coefficients, at a fixed scale  $j$ .

By Corollary 1.5 in [13] (which is valid both for coarse and fine scale shearlets), there is a constant  $C > 0$  such that, as  $\epsilon \rightarrow 0$ ,

$$\#R(j, \epsilon) \leq C \epsilon^{-2/3}.$$

It follows by rescaling that

$$\#R(j, 2^j \epsilon) \leq C 2^{-2j/3} \epsilon^{-2/3}.$$

Since  $\widehat{\psi} \in C_0^\infty(\mathbb{R}^2)$ , for  $\mu = (j, \ell, k, d) \in M^0$  and  $d = 0$ , we have that

$$\begin{aligned} |c_\mu| &= |\langle f, \psi_\mu \rangle| = \left| \int_{\mathbb{R}^2} f(x) 2^{3j/2} \psi(B^\ell A^j x - k) dx \right| \\ &\leq 2^{-3j/2} \|f\|_\infty \int_{\mathbb{R}^2} |\psi(z)| dz = C' 2^{-\frac{3}{2}j}. \end{aligned}$$

Thus,  $R(j, 2^j \epsilon) = \emptyset$  when  $2^j > \epsilon^{-2/5}$  (that is,  $j > j_1(\epsilon) = \frac{2}{5} \log_2(\epsilon^{-1})$ ). Similarly,  $R(j, \epsilon) = \emptyset$  when  $2^j > \epsilon^{-2/3}$  (that is,  $j > j_2(\epsilon) = \frac{2}{3} \log_2(\epsilon^{-1})$ ). For  $\mu \in M^0$  and  $d = 1$ , we get exactly the same estimates.

For the risk proxy, notice that

$$\sum_{\{\mu \in \mathcal{N}(\epsilon)\}} \min(c_\mu^2, 2^{2j} \epsilon^2) = S_1(\epsilon) + S_2(\epsilon),$$

where

$$S_1(\epsilon) = \sum_{\{\mu \in \mathcal{N}(\epsilon) : |c_\mu| \geq 2^j \epsilon\}} \min(c_\mu^2, 2^{2j} \epsilon^2)$$

$$S_2(\epsilon) = \sum_{\{\mu \in \mathcal{N}(\epsilon): |c_\mu| < 2^j \epsilon\}} \min(c_\mu^2, 2^{2j} \epsilon^2)$$

Hence, using the observations above, we have:

$$\begin{aligned} S_1(\epsilon) &= \sum_{\{\mu \in \mathcal{N}(\epsilon): |c_\mu| \geq 2^j \epsilon\}} 2^{2j} \epsilon^2 \\ &\leq \sum_{j \leq j_1} \sum_{\{\mu \in M_j: |c_\mu| \geq 2^j \epsilon\}} 2^{2j} \epsilon^2 \\ &\leq C \sum_{j \leq j_1} (2^{-2j/3} \epsilon^{-2/3}) 2^{2j} \epsilon^2 \\ &= C \sum_{j \leq j_1} 2^{\frac{4}{3}j} \epsilon^{\frac{4}{3}} \\ &\leq C \epsilon^{\frac{4}{5}} \end{aligned}$$

For  $S_2$ , we have

$$\begin{aligned} S_2(\epsilon) &= \sum_{\{\mu \in \mathcal{N}(\epsilon): |c_\mu| < 2^j \epsilon\}} |c_\mu|^2 \\ &= \sum_{j_0 \leq j \leq j_1} \sum_{n=0}^{\infty} \sum_{\{2^{j-n-1} \epsilon \leq |c_\mu| < 2^{j-n} \epsilon\}} |c_\mu|^2 \\ &\leq C \sum_{j_0 \leq j \leq j_1} \sum_{n=0}^{\infty} 2^{-\frac{2}{3}(j-n-1)} \epsilon^{-\frac{2}{3}} 2^{2(j-n)} \epsilon^2 \\ &= C \sum_{j_0 \leq j \leq j_1} \sum_{n=0}^{\infty} 2^{-\frac{4}{3}n} 2^{\frac{2}{3}} 2^{\frac{4}{3}j} \epsilon^{\frac{4}{3}} \\ &\leq C \sum_{j_0 \leq j \leq j_1} 2^{\frac{4}{3}j} \epsilon^{\frac{4}{3}} \\ &\leq C \epsilon^{\frac{4}{5}} \quad \square \end{aligned}$$

**Proof.** (Theorem 4.1.1(3)) For each fixed scale  $j_0 \leq j \leq j_1$ , the number of indices  $\mu$  in  $\mathcal{N}(\epsilon) \cap M_j$  is of the order  $O(2^{5j})$ . In fact,  $\mathcal{N}(\epsilon) \cap M_j \subset \{(j, \ell, k, d) : |k| \leq 2^{2j+1}, |\ell| \leq 2^j\}$  and this set contains  $O(2^{4j})$  terms for the  $k$  variable and  $O(j)$  terms for the  $\ell$  variable. Hence, adding up the contributions corresponding to the various scales, we obtain:

$$\#\mathcal{N}(\epsilon) \leq C \sum_{j \leq j_1} 2^{5j} \leq C \epsilon^{-2}. \quad \square$$

## References

- [1] H. H. Arsenault, G. April, Properties of speckle integrated with a finite aperture and logarithmically transformed, *J. Opt. Soc. Am.* **66** (1976), 1160–1163.
- [2] M. Bhatia, W. C. Karl, and A. S. Willsky, A wavelet-based method for multiscale tomographic reconstruction, *IEEE Trans. Med. Imag.*, **15** (1996), 92–101.
- [3] E. J. Candès, L. Demanet, D. L. Donoho, and L. Ying, Fast discrete curvelet transforms, *Multiscale Model. Simul.*, **5** (3) (2006), 861–899.
- [4] E. J. Candès, and D. L. Donoho, Recovering edges in ill-posed inverse problems: optimality of curvelet frames, *Annals Stat.*, **30**(3) (2002), 784–842.
- [5] O. Christensen, *An introduction to frames and Riesz bases*, Birkhäuser, Boston, 2003.
- [6] A. M. Cormack, Representation of a function by its line integrals, with some radiological applications, *J. Appl. Phys.* **34** (1963), 2722–2727.
- [7] A. M. Cormack, Representation of a function by its line integrals, with some radiological applications II, *J. Appl. Phys.* **35** (1964), 195–207.
- [8] Do, M. N., and M. Vetterli, The contourlet transform: an efficient directional multiresolution image representation, *IEEE Trans. Image Process.*, **14**(12) (2005), 2091–2106.
- [9] D. L. Donoho, Nonlinear solution of linear inverse problem by wavelet-vaguelettedecomposition, *Appl. Comput. Harmon. Anal.*, **2** (1995), pp. 101–126.
- [10] D. L. Donoho, I. M. Johnstone, Ideal spatial adaptation via wavelet shrinkage, *Biometrika*, **81** (1994), 425–455.
- [11] G. R. Easley, D. Labate, F. Colonna, Shearlet-Based Total Variation Diffusion for Denoising, *IEEE Trans. Image Proc.* **18** (2) (2009), 260–268.
- [12] G. R. Easley, D. Labate, and W. Lim, Sparse Directional Image Representations using the Discrete Shearlet Transform, *Appl. Comput. Harmon. Anal.*, **25** (1) (2008), 25–46.
- [13] K. Guo, D. Labate, Optimally Sparse Multidimensional Representation using Shearlets, *SIAM J. Math. Anal.*, **39** (2007), 298–318.
- [14] K. Guo, W. Lim, D. Labate, G. Weiss, E. Wilson, Wavelets with composite dilations, *ERA Amer. Math. Soc.*, **10** (2004), 78–87.
- [15] K. Guo, W. Lim, D. Labate, G. Weiss, E. Wilson, The theory of wavelets with composite dilations, in: *Harmonic Analysis and Applications*, C. Heil (ed.), pp. 231–249, Birkhäuser, Boston, 2006.

- [16] K. Guo, W. Lim, D. Labate, G. Weiss, E. Wilson, Wavelets with composite dilations and their MRA properties, *Appl. Computat. Harmon. Anal.*, **20** (2006) 231–249.
- [17] S. Helgason, *The Radon Transform*, Birkhäuser, Boston, 1980.
- [18] E. D. Kolaczyk, A wavelet shrinkage approach to tomographic image reconstruction, *J. Amer. Statist. Assoc.*, **91** (1996), 1079–1090.
- [19] G. Kutyniok and D. Labate, Resolution of the Wavefront Set using Continuous Shearlets, *Trans. Am. Math. Soc.*, **361** (2009), 2719–2754.
- [20] N.Y. Lee, and B. J. Lucier, Wavelet Methods for Inverting the Radon Transform with Noisy Data, *IEEE Trans. Image Process.* **10** (1) (2001), 79–94.
- [21] Le Pennec, E., and S. Mallat, Sparse geometric image representations with bandelets, *IEEE Trans. Image Process.* **14** (2005), 423–438.
- [22] F. Natterer, and F. Wübbeling, *Mathematical Methods in Image Reconstruction*, SIAM Monographs on Mathematical Modeling and Computation, Philadelphia, 2001.
- [23] R. Neelamani, R. Nowak, and R. Baraniuk, Model-based inverse halftoning with Wavelet-Vaguelette Deconvolution, in *Proc. IEEE Int. Conf. Image Processing ICIP 2000*, (Vancouver, Canada), pp. 973–976, Sept. 2000.
- [24] E. L. Miller, L. Nicolaides, and A. Mandelis, Nonlinear inverse scattering methods for thermal-wave slice tomography: A wavelet domain approach, *J. Opt. Soc. Amer. A*, **15** (1998), 1545–1556.
- [25] J. Radon, Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten, *Berichte Sächsische Akademie der Wissenschaften, Leipzig, Math.-Phys. Kl.*, **69**, 262–267, reprinted in [17], 177–192, 1917.
- [26] S. Y. Zhao, Wavelet filtering for filtered backprojection in computed tomography, *Appl. Comput. Harmon. Anal.*, **6** (2009), 346–373.