

# Wavelet Methods for Inverting the Radon Transform with Noisy Data

Nam-Yong Lee and Bradley J. Lucier, *Senior Member, IEEE*

*Abstract*—Because the Radon transform is a smoothing transform, any noise in the Radon data becomes magnified when the inverse Radon transform is applied. Among the methods used to deal with this problem is the Wavelet-Vaguelette Decomposition (WVD) coupled with Wavelet Shrinkage, as introduced by David L. Donoho. We extend several results of Donoho and others here. First, we introduce a new sufficient condition on wavelets to generate a WVD. For a general homogeneous operator, which class includes the Radon transform, we show that a variant of Donoho’s method for solving inverse problems can be derived as the exact minimizer of a variational problem that uses a Besov norm as the smoothing functional. We give a new proof of the rate of convergence of wavelet shrinkage that allows us to estimate rather sharply the best shrinkage parameter needed to recover an image from noise-corrupted data. We conduct tomographic reconstruction computations that support the hypothesis that near-optimal shrinkage parameters can be derived if one can estimate only two Besov-space parameters about an image  $f$ . Both theoretical and experimental results indicate that our choice of shrinkage parameters yields uniformly better results than Kolaczky’s variant of Donoho’s method and the classical filtered back-projection method.

*Index Terms*—Wavelets, Radon transform, positron emission tomography, variational problems, wavelet shrinkage.

## I. INTRODUCTION

Inverting the Radon transform is tremendously important in many fields, especially Medical Imaging, where it is the mathematical basis of Computer Tomography, Positron Emission Tomography, etc. The difficulty in inverting the Radon transform  $A$  is that it is a smoothing transform, i.e., applying  $A$  to an image  $f$  results in data  $Af$  that has, roughly speaking, one-half derivative more smoothness than the original image. Thus,  $A^{-1}$ , the inverse Radon transform, has properties like one-half of a derivative, i.e., it is unbounded.

---

This paper will appear in the IEEE Transactions on Image Processing. This work was supported in part by the Office of Naval Research, Contract N00014-91-J-1152, and the Purdue Research Foundation. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Mila Nikolova.

N.-Y. Lee is with Department of Control and Instrumentation Engineering, Kangwon National University, Chunchon 200-701, Kangwon-Do, KOREA (e-mail: nylee@cc.kangwon.ac.kr).

B. J. Lucier is with the Department of Mathematics, Purdue University, West Lafayette, IN 47907-1395 USA (e-mail: lucier@math.purdue.edu).

One can view the Radon transform as the special case with  $\alpha = 1/2$  of the following abstract problem: Assume that a linear operator  $A$  maps the Hilbert space  $L^2(\mathbb{R}^d)$  to another Hilbert space  $\mathcal{Y}$  and satisfies

$$(1) \quad \widehat{A^*Af}(\xi) = |\xi|^{-2\alpha}\widehat{f}(\xi)$$

for some  $\alpha \geq 0$ , where  $A^*$  is the adjoint of  $A$  and  $\widehat{g}$  denotes the Fourier transform of  $g$ . We wish to recover  $f \in L^2(\mathbb{R}^d)$  from  $Af$ . In practice we are only able to observe noisy data of the form

$$(2) \quad Y = Af + Z,$$

where  $Z$  represents a perturbation error in our observation procedure. Here we assume that  $Z$  is mean-zero Gaussian noise.

Such linear inverse problems arise in various scientific fields. The noise removal problem in image processing takes the form  $Y = f + Z$ . In this case the underlying operator is the identity operator, which obviously satisfies (1) with  $\alpha = 0$ . As already mentioned, the Radon transform satisfies (1) with  $\alpha = 1/2$ . Additionally, the  $2\pi$  multiple of the one-dimensional integration operator also satisfies (1) with  $\alpha = 1$ .

Despite the simple characterization (1) of  $A$  via the Fourier transform, linear filtering methods based on (1) such as *filtered backprojection* (FBP) often exhibit degradation in recovering  $f$  from noisy data. The Fourier transform diagonalizes any convolution-type operator, and this property has been an advantage of Fourier transform methods in deconvolution problems such as (2). However, poor representation of nonsmooth functions via the Fourier transform often yields an unacceptable decision rule, which considers any low-frequency structure to be information and any high-frequency structure, no matter how strong, to be noise, in recovering  $f$  from (2).

Very recently, there has been great interest in the use of wavelet bases to represent functions, and many important advantages of wavelet bases have been discovered. One of the key features of wavelet bases is the smoothness characterization of various function spaces in terms of wavelet coefficients. This property has been used in several image processing and statistical applications, such as data compression [15], [16], [33], noise removal [5], [18], and non-parametric estimation in statistics [4], [22], [24], [25], etc.

Even though convolution-type operators, in general, are not diagonalizable with respect to wavelet bases, the operator satisfying (1) can be diagonalized with the help of

a *wavelet-vaguelette system* (WVS). In [21] Donoho proved that there exists a WVS  $\{\psi_\lambda, U_\lambda, \tilde{U}_\lambda\}$  such that  $\{\psi_\lambda\}$  is an orthogonal wavelet basis of  $L^2(\mathbb{R}^d)$ ,  $\{U_\lambda\}$  and  $\{\tilde{U}_\lambda\}$  are biorthogonal Riesz bases of  $\mathcal{Y}$ , and for any  $f \in L^2(\mathbb{R}^d)$ ,

$$(3) \quad f = \sum_{\lambda} c_{\lambda} [Af, U_{\lambda}] \psi_{\lambda}$$

for known scalars  $\{c_{\lambda}\}$ , where  $[\cdot, \cdot]$  is the inner product of  $\mathcal{Y}$ . (The definitions of a Riesz basis, biorthogonal wavelets, etc., are given in the next sections.) In [21] Donoho called (3) a *wavelet-vaguelette decomposition* (WVD) of  $f$ .

Applying the WVD (3) in solving (2), Donoho suggested the *wavelet shrinkage* method (see, e.g., [21], [22], and [30]), which shrinks the observed coefficients  $c_{\lambda}[Y, U_{\lambda}]$  towards zero by a certain amount, and proved that if the true solution  $f$  is known to lie in the Besov smoothness space  $B_{q,p}^{\beta}(\mathbb{R}^d)$  (for the definition, see [17] and Section IV of this paper), where  $\beta > (2\alpha + d)(1/p - 1/2)$ , then the solution of his method converges to  $f$  with the optimal rate if one uses the optimal shrinkage parameters (although Donoho did not give a method for finding these parameters). For details, see [21].

In this paper we consider a family of variational problems for solving (2) that are related to Donoho’s method. These variational problems take the form: Given a positive parameter  $\gamma$  and a function smoothness space  $B$ , find a function  $\tilde{f}$  that minimizes over all possible functions  $g$  in  $B$  the functional

$$(4) \quad \|Y - Ag\|_{\mathcal{Y}}^2 + 2\gamma|g|_B,$$

where  $\|\cdot\|_{\mathcal{Y}}$  is the norm defined by the inner product  $[\cdot, \cdot]$  of  $\mathcal{Y}$  and  $|\cdot|_B$  is the semi-norm of the function space  $B$ .

The function space  $B$  differs from usual smoothness subspaces in regularization techniques (see, e.g., [43]), because  $B$  could be a function space which is not *embedded* in  $L^2(\mathbb{R}^d)$ . The parameter  $\gamma$  balances the importance between the difference  $\|Y - Ag\|_{\mathcal{Y}}^2$  in  $\mathcal{Y}$  and the smoothness  $|g|_B$  of  $g$  in  $B$ .

A fast way of solving (4) is required for practical algorithms. We suggest using vaguelettes and wavelets in (4), to characterize  $\|Y - Ag\|_{\mathcal{Y}}$  and  $|g|_B$ , respectively. Using the  $L^2$ -*stability* of vaguelettes and the smoothness characterization properties of wavelets, we derive an expression involving wavelet coefficients that is equivalent to (4), and that can be minimized quickly (in a time proportional to the number of unknowns). In particular, if we choose  $B = B_{1,1}^{\beta_0}(\mathbb{R}^d)$  in (4), then the exact minimizer, denoted by  $\tilde{f}_{\gamma, \beta_0}^*$ , of the resulting equivalent sequence minimization problem has the same form as the wavelet shrinkage method proposed by Donoho [21] with possibly different shrinkage parameters.

We give a new proof of the rate of convergence of  $\tilde{f}_{\gamma, \beta_0}^*$  to  $f$  that allows us to estimate rather sharply the best wavelet shrinkage parameter for solving (2) in the presence of Gaussian noise. Our analysis reveals that only one parameter  $\beta_0$  allows finite error for all levels of resolution;

this  $\beta_0^* = d/2 - \alpha$ , which depends only on the parameter  $\alpha$  (which determines the ill-posedness of  $A$ ) and the dimension  $d$  of the domain, which will be 1 (for signals) or 2 (for images), of  $f$ . For any  $\beta_0$  the wavelet shrinkage amount for terms  $c_{\lambda}[Y, U_{\lambda}]$  at a dyadic scale  $2^k$  is given by  $\gamma 2^{k(\beta_0 - d/2 + 2\alpha)}$ ; for the optimal  $\beta_0^*$ , we should shrink terms  $c_{\lambda}[Y, U_{\lambda}]$  by  $\gamma 2^{k\alpha}$ .

It remains only to choose the parameter  $\gamma$ . We give an upper bound on the expected value of the error that depends on  $\gamma$  and two smoothness parameters of  $f$  (both of which can be estimated by a wavelet compression technique, and which are roughly invariant for images of fixed types). We suggest that one should choose  $\gamma$  by minimizing the upper bound on the error.

We have implemented this method and present the results of several experiments. In these experiments, we use a WVS that uses less-smooth wavelets than previously thought possible. To substantiate this, we rely on a sufficient condition, which is proved in Chapter 6 of [32], on regularity and *vanishing moments* (for the definition, see Section IV) of functions to have a WVS. This sufficient condition is weaker than in [21] (see Section VI), and allows more wavelets such as less smooth examples of Daubechies’s compactly supported orthogonal wavelets [10] and symmetric biorthogonal wavelets [6] to be used for solving (2). These compactly supported wavelets are generated from *scaling functions* by *refinement equations* (for the definition, see Section III); thus vaguelette coefficients can be obtained in a recursive manner as wavelet coefficients can be.

We show through tomographic reconstruction experiments that with a *translation-rotation averaging technique* (see Section X) the wavelet shrinkage method  $\tilde{f}_{\gamma, \beta_0}^*$ , where parameters  $\gamma$  and  $\beta_0$  are determined from our analysis, leads to a better reconstruction in the presence of Gaussian noise than the traditional filtered backprojection method. As compared with the shrinkage parameter suggested by Kolaczyk [30], which is motivated by the Visu-Shrink method by Donoho and Johnstone, our shrinkage parameter leads to a better reconstruction, which removes less noise, but keeps more image features.

Other researchers have introduced wavelet methods for tomographic problems. Methods based on one-dimensional wavelet decomposition of the projection data followed by linear filtered backprojection were introduced in [40], [38], [39], [45], and [47]; local reconstruction is a concern of all these authors. Bhatia, Karl, and Willsky [2] first introduce a method based on similar principles, then use a MAP model with a linear filter based on  $L^2$  regularization of each one-dimensional projection data. In a later paper [1], they again use a one-dimensional wavelet decomposition of the projection data, which is back-projected to find a different basis for the so-called “Natural Pixel” formulation of the image reconstruction problem, which yields a sparse matrix problem to solve rather than a full matrix problem, again with a MAP model of the image. Sahiner and Yagle [41] also use a one-dimensional transform of the projection data, but they threshold the wavelet coefficients before ap-

plying filtered back projection of the data. Thresholding is related to wavelet shrinkage as a means of noise removal (see, e.g., [5]), but the authors do not provide an analysis of the reduction in error due to thresholding and, to our eyes, their method results in significant artifacts. The authors add additional image constraints which do reduce these artifacts somewhat.

Other authors have used two-dimensional wavelet decompositions in the image domain, as we do. Delaney and Bresler [14] use wavelet transforms of the image as a tool to make a standard filtered backprojection algorithm more efficient numerically, and to provide a method of local reconstruction. Dobson [20] provides a two-dimensional wavelet method for electrical impedance tomography based on Tikhonov regularization with the Sobolev space  $H_0^s(I) \approx B_2^s(L^2(I))$ ; this is similar in spirit to our method (we use the space  $B_1^{\beta_0}(L_1(I))$  instead), but results in linear filtering that has worse error bounds than wavelet shrinkage. Zhu et al. [48] employ Tikhonov regularization with the space  $L^2(I)$  as the regularizing space. For a somewhat different problem, Miller, Nicolaides and Mandelis [36] use a variational penalty approach similar to ours, incorporating penalties in a norm that is roughly  $B_p^\beta(L_p)$  for  $p = 2$  and  $p = 1.2$ ; they find that  $p = 1.2$  offers somewhat better results than  $p = 2$  (we use  $p = 1$ ), but it is difficult to tell from their paper what  $\beta$  is. Finally, Zhao, Wang, and Hsieh [46] claim to apply wavelet shrinkage to a problem in fan-beam tomography, but do not give enough details to compare their method to ours.

## II. PRELIMINARIES

In this section we review some notations and definitions for the future use.

The *translation* operator  $T_h$ : For  $h \in \mathbb{R}^d$ ,  $T_h f(x) = f(x - h)$ .

The *dilation* operator  $D_a$ : For  $a > 0$ ,  $D_a f(x) = f(x/a)$ .

The *rotation* operator  $Q_\theta$ : For a function  $f$  defined on  $\mathbb{R}^2$ ,  $Q_\theta f(r, \varphi) = f(r, \varphi - \theta)$ , where  $f$  is represented in polar coordinates.

The *shrinkage* operator  $S_\mu : \mathbb{R} \rightarrow \mathbb{R}$ : For  $\mu \geq 0$ ,  $S_\mu(x) = \text{sign}(x)(|x| - \mu)^+$ .

Let  $\mathcal{H}$  be a separable Hilbert space. For real-valued functions  $S_1$  and  $S_2$  defined on  $\mathcal{H}$ , we denote  $S_1(f) \asymp S_2(f)$  if there are positive constants  $C_1$  and  $C_2$  such that for all  $f \in \mathcal{H}$ ,  $C_1 S_1(f) \leq S_2(f) \leq C_2 S_1(f)$ . A collection of functions  $\{\varphi_n\}$  in  $\mathcal{H}$  is said to be  *$L^2$ -stable* if for all  $\varphi$  in  $\mathcal{H}$ ,  $\|\varphi\|_{\mathcal{H}}^2 \asymp \sum_n |\langle \varphi, \varphi_n \rangle|^2$ . An  $L^2$ -stable basis of  $\mathcal{H}$  is also called a *Riesz basis*.

Let  $X_1$  and  $X_2$  be normed vector spaces. The space  $X_1$  is said to be *embedded* in the space  $X_2$ , denoted by  $X_1 \hookrightarrow X_2$ , if for each  $f$  in  $X_1$ ,  $f$  is in  $X_2$  and there is a constant  $C$  such that for all  $f \in X_1$ ,  $\|f\|_{X_2} \leq C\|f\|_{X_1}$ .

We denote by  $S(\mathbb{R}^d)$  the space of rapidly decreasing  $C^\infty$  functions on  $\mathbb{R}^d$  and by  $S'(\mathbb{R}^d)$  its topological dual, the space of tempered distributions. The Fourier transform  $\widehat{f}$

of a function  $f \in S(\mathbb{R}^d)$  is defined by

$$\widehat{f}(\xi) = \int_{\mathbb{R}^d} e^{-2\pi i \xi \cdot x} f(x) dx,$$

while the inverse Fourier transform gives  $f$  back from  $\widehat{f}$  by

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

One extends the Fourier transform and its inverse from  $S(\mathbb{R}^d)$  to  $S'(\mathbb{R}^d)$  by duality.

## III. WAVELETS

In this section we briefly review basic wavelet theory. Throughout this paper we consider only compactly supported wavelets such as Daubechies' orthogonal wavelets [10], symmetric biorthogonal wavelets by Cohen, Daubechies, and Feauveau [6] and Herley and Vetterli [27], and modified wavelets [7] and [29], which are designed to deal with functions defined on a bounded domain.

We begin with orthogonal wavelets on  $\mathbb{R}$ . Let  $\psi$  be a bounded and compactly supported function on  $\mathbb{R}$ . We define  $\psi_{k,j}(x) = 2^{k/2} \psi(2^k x - j)$  for all integers  $k$  and  $j$ . The collection of functions  $\{\psi_{k,j}\}_{k \in \mathbb{Z}, j \in \mathbb{Z}}$  is called an orthonormal wavelet basis for  $L^2(\mathbb{R})$  if for any  $f \in L^2(\mathbb{R})$ ,

$$(5) \quad f = \sum_{k \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} \langle f, \psi_{k,j} \rangle \psi_{k,j}$$

and

$$\|f\|_{L^2}^2 = \sum_{k \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} |\langle f, \psi_{k,j} \rangle|^2.$$

Associated with  $\psi$  is a scaling function  $\phi$ , from which one generates the functions  $\phi_{k,j}(x) = 2^{k/2} \phi(2^k x - j)$ . The set  $\{\phi_{k,j}\}_{j \in \mathbb{Z}}$  is orthonormal for each  $k$ . With these functions  $\phi_{k,j}$ , we have

$$(6) \quad f = \sum_{k \geq k_0} \sum_{j \in \mathbb{Z}} \langle f, \psi_{k,j} \rangle \psi_{k,j} + \sum_{l \in \mathbb{Z}} \langle f, \phi_{k_0,l} \rangle \phi_{k_0,l}$$

and

$$\|f\|_{L^2}^2 = \sum_{k \geq k_0} \sum_{j \in \mathbb{Z}} |\langle f, \psi_{k,j} \rangle|^2 + \sum_{l \in \mathbb{Z}} |\langle f, \phi_{k_0,l} \rangle|^2$$

for each integer  $k_0$ .

We assume that for a given scaling function  $\phi$ , there exist finite sequence  $(h_n)$  such that  $\phi = \sum_n h_n \phi_{1,n}$  and  $\psi = \sum_n g_n \phi_{1,n}$ , where  $g_n = (-1)^{n+1} h_{-n+1}$ . We call the above equations *refinement equations*. These equations are closely related to the so-called *fast wavelet transform* (FWT) and inverse FWT. For details, see, e.g., [10].

For orthogonal wavelets on  $L^2(\mathbb{R}^d)$  with  $d > 1$ , we use the tensor product of  $\phi$  and  $\psi$ . We make  $2^d$  functions defined on  $\mathbb{R}^d$  by  $\varphi_1(x_1) \times \cdots \times \varphi_d(x_d)$ , where either  $\varphi_i = \phi$  or  $\varphi_i = \psi$ . Among them, we denote  $\phi(x_1) \times \cdots \times \phi(x_d)$  by  $\Phi$ , and the remaining  $2^d - 1$  functions by  $\psi^{(i)}$  with  $i = 1, 2, \dots, 2^d - 1$ . (It is convenient to fix that  $\varphi(x_j)$  is  $\psi(x_j)$  if the  $j$ th binary digit of  $i$  from the left is one, and  $\varphi(x_j)$  is  $\phi(x_j)$  if the  $j$ th binary digit of  $i$  is zero.) We define

$\Phi_{k,j} = 2^{kd/2}\Phi(2^k \cdot -j)$  for  $k \in \mathbb{Z}$  and  $j \in \mathbb{Z}^d$ , and similarly for  $\psi_{k,j}^{(i)}$ . Then any  $f \in L^2(\mathbb{R}^d)$  can be written as

$$f = \sum_{k \geq k_0} \sum_{j \in \mathbb{Z}^d} \sum_i \langle f, \psi_{k,j}^{(i)} \rangle \psi_{k,j}^{(i)} + \sum_{l \in \mathbb{Z}^d} \langle f, \Phi_{k_0,l} \rangle \Phi_{k_0,l},$$

and, moreover,

$$\|f\|_{L^2}^2 = \sum_{k \geq k_0} \sum_{j \in \mathbb{Z}^d} \sum_i |\langle f, \psi_{k,j}^{(i)} \rangle|^2 + \sum_{l \in \mathbb{Z}^d} |\langle f, \Phi_{k_0,l} \rangle|^2,$$

where the index  $i$  in above two equations ranges over  $1, 2, \dots, 2^d - 1$ .

When one is concerned with a bounded domain, for example, the unit cube  $\Omega$  in  $\mathbb{R}^d$ , then one does not consider all shifts  $j \in \mathbb{Z}^d$ , but only those shifts for which  $\psi_{k,j}^{(i)}$  intersects  $\Omega$  nontrivially. Moreover, one must adapt wavelets that overlap the boundary of  $\Omega$  to preserve  $L^2$ -stability on the domain. For details, see, e.g., [7] and [29]. To ignore all complication of this sort, we shall use indices without precisely specifying the domains of the indices of the sums whenever this abbreviation does not cause any confusion.

Before we close this section, we state the two-dimensional FWT and its inverse FWT associated with  $\{\psi_{k,j}^{(i)}\}$  for future use.

Fast wavelet transform:

$$\begin{aligned} \langle f, \Phi_{k,j} \rangle &= \sum_{n_1, n_2} h_{n_1-2j_1} h_{n_2-2j_2} \langle f, \Phi_{k+1,n} \rangle, \\ \langle f, \psi_{k,j}^{(1)} \rangle &= \sum_{n_1, n_2} h_{n_1-2j_1} g_{n_2-2j_2} \langle f, \Phi_{k+1,n} \rangle, \\ \langle f, \psi_{k,j}^{(2)} \rangle &= \sum_{n_1, n_2} g_{n_1-2j_1} h_{n_2-2j_2} \langle f, \Phi_{k+1,n} \rangle, \quad \text{and} \\ \langle f, \psi_{k,j}^{(3)} \rangle &= \sum_{n_1, n_2} g_{n_1-2j_1} g_{n_2-2j_2} \langle f, \Phi_{k+1,n} \rangle. \end{aligned}$$

Inverse fast wavelet transform:

$$\begin{aligned} \langle f, \Phi_{k+1,j} \rangle &= \sum_{n_1} h_{j_1-2n_1} \sum_{n_2} h_{j_2-2n_2} \langle f, \Phi_{k,n} \rangle \\ &+ \sum_{n_1} h_{j_1-2n_1} \sum_{n_2} g_{j_2-2n_2} \langle f, \psi_{k,n}^{(1)} \rangle \\ &+ \sum_{n_1} g_{j_1-2n_1} \sum_{n_2} h_{j_2-2n_2} \langle f, \psi_{k,n}^{(2)} \rangle \\ &+ \sum_{n_1} g_{j_1-2n_1} \sum_{n_2} g_{j_2-2n_2} \langle f, \psi_{k,n}^{(3)} \rangle. \end{aligned}$$

#### IV. BESOV SPACES

One of the major advantages of wavelets is smoothness characterization, which means that we can determine the membership of a function in many different function spaces by examining its wavelet coefficients. For details, see, e.g., [17], [19], [26], [29], [31], and [34].

We are interested in Besov spaces  $B_q^\beta(L^p(\mathbb{R}^d))$  for  $1 < p \leq \infty$  and  $B_q^\beta(H^p(\mathbb{R}^d))$  for  $0 < p \leq 1$ , where  $H^p(\mathbb{R}^d)$  is the real Hardy space (for the definition, see, e.g., [44]). To simplify our presentation, throughout this paper we shall

use the following convention:

$$B_{q,p}^\beta(\mathbb{R}^d) = \begin{cases} B_q^\beta(L^p(\mathbb{R}^d)), & 1 < p \leq \infty, \\ B_q^\beta(H^p(\mathbb{R}^d)), & 0 < p \leq 1. \end{cases}$$

With this convention, we get more familiar spaces for certain values of parameters. When  $p = q = 2$ ,  $B_{2,2}^\beta(\mathbb{R}^d)$  is the Sobolev space  $W^\beta(\mathbb{R}^d)$ , and when  $0 < \beta < 1$  and  $1 \leq p \leq \infty$ ,  $B_{\infty,p}^\beta(\mathbb{R}^d)$  is the Lipschitz space  $\text{Lip}(\beta, L^p(\mathbb{R}^d))$ .

When  $0 < p < 1$  or  $0 < q < 1$ , then  $B_{q,p}^\beta(\mathbb{R}^d)$  are no longer normed spaces. However, they are always quasi-normed spaces, and with a certain abuse of terminology, we shall continue to call these quasi-norms norms.

We will not give the precise definition of Besov spaces here. What is important to us is that one can determine whether  $f$  is in  $B_{q,p}^\beta(\mathbb{R}^d)$  simply by examining its wavelet coefficients. Let  $\phi$  and  $\psi$  be bounded, compactly supported, and univariate functions as described in Section III. Suppose  $\phi$  has  $R$  continuous derivatives and  $\psi$  has *vanishing moments* of order  $M$ , i.e.,  $\int_{\mathbb{R}} x^m \psi(x) dx = 0$ , for  $m = 0, 1, \dots, M$ . We assume that  $\psi^{(i)}$  and  $\Phi$  are functions described in Section III by tensor products of  $\phi$  and  $\psi$ . Then as long as  $\beta < \min(R, M)$ , for any  $k_0$ ,

$$(7) \quad |f|_{B_{q,p}^\beta} \asymp \left( \sum_{k \geq k_0} 2^{ksq} \left[ \sum_{j,i} |\langle f, \psi_{k,j}^{(i)} \rangle|^p \right]^{\frac{q}{p}} \right)^{\frac{1}{q}}$$

and

$$(8) \quad \|f\|_{B_{q,p}^\beta} \asymp |f|_{B_{q,p}^\beta} + 2^{k_0 d(\frac{1}{2} - \frac{1}{p})} \left( \sum_l |\langle f, \Phi_{k_0,l} \rangle|^p \right)^{\frac{1}{p}},$$

where  $s = \beta + d(1/2 - 1/p)$ .

The proof of (7) and (8) can be found in [31]. In this paper we shall use the right hand side of (7) as the *definition* of  $|f|_{B_{q,p}^\beta}$  and the right hand side of (8) as the definition of  $\|f\|_{B_{q,p}^\beta}$ , where we assume that  $k_0$  is a fixed, nonnegative small integer throughout this paper.

There are various embedding relations among the Besov spaces  $B_{q,p}^\beta(\mathbb{R}^d)$ . For example, for  $q' > q$ ,  $B_{q,p}^\beta(\mathbb{R}^d) \hookrightarrow B_{q',p}^\beta(\mathbb{R}^d)$ ; if  $\beta' < \beta$  and  $(\beta - \beta')/d = 1/p - 1/p'$ , then  $B_{q,p}^\beta(\mathbb{R}^d) \hookrightarrow B_{q,p'}^{\beta'}(\mathbb{R}^d)$ ; and for any  $q_1$  and  $q_2$ ,  $B_{q_1,p}^\beta(\mathbb{R}^d) \hookrightarrow B_{q_2,p}^{\beta'}$  as long as  $\beta' < \beta$ . These relations show that the parameter  $q$  in  $B_{q,p}^\beta(\mathbb{R}^d)$  is not as important as  $p$  or  $\beta$ . Thus, we consider only Besov spaces of the form  $B_{p,p}^\beta(\mathbb{R}^d)$  from now on.

Based on a standard argument from interpolation theory (see, e.g., [5]), one can show that if  $L^\infty(\mathbb{R}^d) \cap B_{p,p}^\beta(\mathbb{R}^d) \hookrightarrow B_{p',p'}^{\beta'}(\mathbb{R}^d)$ , where  $\beta' < \beta$  and  $\beta' p' = \beta p$ .

In analogy with the special case of Sobolev spaces  $W^\beta(\mathbb{R}^d)$ , the Besov space  $B_{p,p}^\beta(\mathbb{R}^d)$  with  $\beta < 0$  is understood as the dual space of  $B_{p',p'}^{\beta'}(\mathbb{R}^d)$ , where  $\beta' = -\beta$  and  $1/p + 1/p' = 1$ . (Here we assume that  $1 \leq p \leq \infty$ .) For details, see, e.g., [34] and [32].

## V. LINEAR HOMOGENEOUS EQUATIONS

We now discuss three examples of linear homogeneous equations—the identity transform, the Radon transform, and  $2\pi$  times the one-dimensional integral.

The identity operator on  $L^2(\mathbb{R}^d)$  satisfies (1) with  $\alpha = 0$ . In this case, we have  $Y = f + Z$  in (2). Recovering  $f$  can be viewed as a noise removal problem.

The Radon transform  $\mathcal{R} : L^2(\mathbb{R}^2) \rightarrow L^2([0, \pi], L^2(\mathbb{R}))$  is defined by

$$\mathcal{R}f(\theta, u) = \int_{L_{\theta, u}} f(x, y) ds(x, y),$$

where  $ds(x, y)$  is Euclidean measure on the line  $L_{\theta, u} = \{(x, y) \mid x \cos \theta + y \sin \theta = u\}$ . The inner product  $[\cdot, \cdot]$  in  $L^2([0, \pi], L^2(\mathbb{R}))$  is defined by

$$[F, G] = \int_0^\pi \int_{\mathbb{R}} F(\theta, u) \overline{G(\theta, u)} du d\theta.$$

The *Fourier slice theorem* states that

$$(9) \quad \mathcal{R}f(\theta, \cdot)^\wedge(w) = \widehat{f}(w \cos \theta, w \sin \theta),$$

where  $\mathcal{R}f(\theta, \cdot)^\wedge(w)$  is the one-dimensional Fourier transform of  $\mathcal{R}f(\theta, u)$  as a function of  $u$ , and  $\widehat{f}$  is the two-dimensional Fourier transform of  $f$ . The proof of (9) can be found in [37].

Using (9), one can show  $\widehat{\mathcal{R}^* \mathcal{R}f}(\xi) = |\xi|^{-1} \widehat{f}(\xi)$ ; thus, the Radon transform satisfies (1) with  $\alpha = 1/2$ . Therefore, the domain of  $\mathcal{R}$ , denoted by  $\mathcal{D}(\mathcal{R})$ , is  $L^2(\mathbb{R}^2) \cap S^{-1/2}(\mathbb{R}^2)$ , where

$$S^\beta(\mathbb{R}^d) = \left\{ f \in S'(\mathbb{R}^d) \mid \int_{\mathbb{R}^d} |\xi|^{2\beta} |\widehat{f}(\xi)|^2 d\xi < \infty \right\}.$$

Our third example,  $\mathcal{I} : L^2(\mathbb{R}) \rightarrow L^2(\mathbb{R})$ , is defined by

$$\mathcal{I}f(x) = 2\pi \int_{-\infty}^x f(t) dt.$$

Since  $2\pi \widehat{f}(\xi) = 2\pi i \xi \widehat{\mathcal{I}f}(\xi)$ , we have  $\widehat{\mathcal{I}f}(\xi) = -i \xi^{-1} \widehat{f}(\xi)$  and  $\mathcal{D}(\mathcal{I}) = L^2(\mathbb{R}) \cap S^{-1}(\mathbb{R})$ . Therefore,  $\widehat{\mathcal{I}^* \mathcal{I}f}(\xi) = |\xi|^{-2} \widehat{f}(\xi)$ . Hence the operator  $\mathcal{I}$  on  $L^2(\mathbb{R})$  satisfies (1) with  $\alpha = 1$ .

Since  $\widehat{A^* A f}(\xi) = |\xi|^{-2\alpha} \widehat{f}(\xi)$ , one's first attempt to solve (2) might be

$$\widetilde{f}(x) = \int_{\mathbb{R}^d} e^{2\pi i \xi \cdot x} |\xi|^{2\alpha} \widehat{A^* Y}(\xi) d\xi.$$

For  $\alpha > 0$ , however, this method generates an unacceptable solution for (2) because the high frequency components of  $A^* Y$ , whose size is determined by  $\widehat{A^* Y}(\xi)$  for  $\xi$  large, are multiplied by  $|\xi|^{2\alpha}$ , which is unbounded as  $\xi \rightarrow \infty$ . To avoid this phenomenon, one can consider a filtered solution  $\widetilde{f}_w$  such that

$$(10) \quad \widetilde{f}_w(x) = \int_{\mathbb{R}^d} e^{2\pi i \xi \cdot x} w(\xi) |\xi|^{2\alpha} \widehat{A^* Y}(\xi) d\xi,$$

where the *weight filter*  $w(\xi)$  satisfies  $0 \leq w(\xi) \leq 1$  and  $w(\xi) \rightarrow 0$  as  $|\xi| \rightarrow \infty$ . For example, the famous *filtered backprojection* (FBP) method for inverting the Radon transform in medical image processing takes the form (10) with  $\alpha = 1/2$ .

Finding a proper weight filter  $w(\xi)$  is very important for the performance of  $\widetilde{f}_w$  (10). We consider a family of weight filters  $\{w_M\}$  such that  $w_M(\xi) = 1$  if  $|\xi| \leq M$  and  $w_M(\xi) = 0$  otherwise.

A calculation (see, e.g., [32]) shows that with the optimal value of

$$M \approx \left( \frac{|f|_{W^\beta}}{\sigma} \right)^{1/(\beta+d/2+\alpha)}$$

we have

$$E \|f - \widetilde{f}_{w_M}\|_{L^2}^2 \leq C |f|_{W^\beta}^{(2\alpha+d)/(\beta+d/2+\alpha)} \sigma^{2r},$$

with rate exponent

$$(11) \quad r = \frac{\beta}{\beta + d/2 + \alpha},$$

when  $f$  is in the Sobolev space  $W^\beta(\mathbb{R}^d) = B_{2,2}^\beta(\mathbb{R}^d)$ .

It is known (see Theorem 4 of [21]) that the rate of convergence of any method for recovering  $f \in W^\beta(\mathbb{R}^d)$  is at most the  $r$  of (11). Thus no filtered backprojection method can provide an asymptotically better solution than  $\widetilde{f}_{w_M}$  with the optimal  $M$ .

The filtered backprojection method  $\widetilde{f}_{w_M}$  considers any low-frequency structure to be signal, and any high-frequency structure to be noise, no matter how large  $|\widehat{f}(\xi)|$  might be. This is not acceptable when we wish to recover functions that can be more meaningfully characterized by their discontinuities (e.g., an image with strong edges and small extent), since such information lies in the high-frequency domain. Even with a more general weight filter  $w$ , we cannot avoid this kind of degradation in recovering nonsmooth functions, since  $w(\xi) \rightarrow 0$  as  $|\xi| \rightarrow \infty$ .

## VI. WAVELET-VAGUELETTE DECOMPOSITIONS

The terminology vaguelette was used by Meyer in [35] to describe a collection of functions which are wavelet-like. In [21] Donoho constructed a *wavelet-vaguelette system* (WVS) for solving homogeneous equations described in Section V and gave a sufficient condition on wavelets to generate a WVS of  $A$ .

A collection of functions  $\{\psi_{k,j}^{(i)}, U_{k,j}^{(i)}, \widetilde{U}_{k,j}^{(i)}\}$  is called a WVS of  $A : L^2(\mathbb{R}^d) \rightarrow \mathcal{Y}$ , where  $\widehat{A^* A f}(\xi) = |\xi|^{-2\alpha} \widehat{f}(\xi)$ , if, first,  $\{\psi_{k,j}^{(i)}\}$  forms an orthonormal wavelet basis of  $L^2(\mathbb{R}^d)$ ; second,  $\{U_{k,j}^{(i)}\}$  and  $\{\widetilde{U}_{k,j}^{(i)}\}$  are biorthogonal Riesz bases of  $\mathcal{Y}$ , i.e., for any  $Y \in \mathcal{Y}$ ,

$$\sum_{k,j,i} |[Y, U_{k,j}^{(i)}]|^2 \asymp \sum_{k,j,i} |[Y, \widetilde{U}_{k,j}^{(i)}]|^2 \asymp \|Y\|_{\mathcal{Y}}^2$$

and  $[U_{k,j}^{(i)}, \widetilde{U}_{k',j'}^{(i')}] = \delta_{(k,j,i),(k',j',i')}$ ; finally,

$$(12) \quad A^* U_{k,j}^{(i)} = 2^{-k\alpha} \psi_{k,j}^{(i)} \quad \text{and} \quad A \psi_{k,j}^{(i)} = 2^{-k\alpha} \widetilde{U}_{k,j}^{(i)}$$

for all  $k, j, i$ .

Once we have a WVS of  $A$ , we can determine wavelet coefficients  $\langle f, \psi_{k,j}^{(i)} \rangle$  of  $f$  from vaguelette coefficients  $[Af, U_{k,j}^{(i)}]$  of  $Af$ . Notice that by (12),

$$(13) \quad \langle f, \psi_{k,j}^{(i)} \rangle = 2^{k\alpha} \langle f, A^* U_{k,j}^{(i)} \rangle = 2^{k\alpha} [Af, U_{k,j}^{(i)}].$$

Plugging (13) into (5), we have the following reproducing formula via the WVS  $\{\psi_{k,j}^{(i)}, U_{k,j}^{(i)}, \tilde{U}_{k,j}^{(i)}\}$  of  $A$ :

$$(14) \quad f = \sum_{k,j,i} 2^{k\alpha} [Af, U_{k,j}^{(i)}] \psi_{k,j}^{(i)}.$$

We call (14) a wavelet-vaguelette decomposition (WVD) of  $f$ .

We now consider how to construct a WVS of  $A$ . We first note that for any operator  $A$  satisfying  $A^* A f(\xi) = |\xi|^{-2\alpha} \widehat{f}(\xi)$ ,  $A^* A$  is translation-invariant, i.e.,

$$(15) \quad A^* A T_h = T_h A^* A,$$

and homogeneous of order  $2\alpha$ , i.e.,

$$(16) \quad A^* A D_a = a^{2\alpha} D_a A^* A.$$

On the other hand, all elements of  $\{\psi_{k,j}^{(i)}\}$  are generated by dilation and translation of basic  $2^d - 1$  functions  $\{\psi^{(i)}\}$  as  $\psi_{k,j}^{(i)} = 2^{kd/2} D_{2^{-k}} T_j \psi^{(i)}$ . We define  $\pi^{(i)}$  by  $A^* A \pi^{(i)} = \psi^{(i)}$  or equivalently  $\widehat{\pi^{(i)}}(\xi) = |\xi|^{2\alpha} \widehat{\psi^{(i)}}(\xi)$ . Let  $\pi_{k,j}^{(i)} = 2^{kd/2} D_{2^{-k}} T_j \pi^{(i)}$ . Then, using (15) and (16), it is not difficult to show that  $A^* A \pi_{k,j}^{(i)} = 2^{-2k\alpha} \psi_{k,j}^{(i)}$ . Hence

$$(17) \quad A^* (2^{k\alpha} A \pi_{k,j}^{(i)}) = 2^{-k\alpha} \psi_{k,j}^{(i)}.$$

Furthermore,

$$\begin{aligned} [2^{k\alpha} A \pi_{k,j}^{(i)}, \tilde{U}_{k',j'}^{(i')}] &= 2^{(k+k')\alpha} \langle A \pi_{k,j}^{(i)}, A \psi_{k',j'}^{(i')} \rangle \\ &= 2^{-(k-k')\alpha} \langle \psi_{k,j}^{(i)}, \psi_{k',j'}^{(i')} \rangle \quad (\text{by (17)}) \\ &= \delta_{(k,j,i),(k',j',i')}. \end{aligned}$$

This implies that if  $\{2^{k\alpha} A \pi_{k,j}^{(i)}\}$  and  $\{\tilde{U}_{k,j}^{(i)}\}$  form  $L^2$ -stable bases of  $\mathcal{Y}$  simultaneously, then by setting

$$(18) \quad U_{k,j}^{(i)} = 2^{k\alpha} A \pi_{k,j}^{(i)},$$

we have a WVS  $\{\psi_{k,j}^{(i)}, U_{k,j}^{(i)}, \tilde{U}_{k,j}^{(i)}\}$  of  $A$ . It is obvious that this construction is the only way of having a WVS of  $A$  for given orthogonal wavelets  $\{\psi_{k,j}^{(i)}\}$ . From now on we use the right-hand side of (18) as the definition of  $U_{k,j}^{(i)}$ .

In [21] Donoho gave a sufficient condition for the existence of a WVS. He proved that if  $\phi$  has  $R$  continuous derivatives, and vanishing moments of order  $M$ , then  $\{\psi_{k,j}^{(i)}, U_{k,j}^{(i)}, \tilde{U}_{k,j}^{(i)}\}$  forms a wavelet-vaguelette system of  $A$  as long as

$$(19) \quad \alpha + d + 1 < \min(R, M).$$

Condition (2) requires high regularity and large vanishing moments for wavelets to be usable for a WVS. For instance, one needs  $R \geq 4$  for the Radon transform in  $\mathbb{R}^2$ .

Notice that none of the first ten Daubechies's compactly supported orthogonal wavelets have this amount of regularity. Many symmetric biorthogonal wavelets (the WVS can be defined for biorthogonal wavelets with a slight modification; see [21] and [32]) with relatively small support do not satisfy condition (2), either. See, e.g., [6], [12], and [13].

There are several disadvantages in using smoother wavelets to approximate a function that is much less smooth than the given wavelets. First, as long as (7) and (8) hold, there is no gain in approximation order with smoother wavelets. Second, since smoother wavelets have wider support, they are not as good as less smooth wavelets, which can have smaller support, in approximating functions with discontinuities.

As an alternative to (19), we can prove (see Theorem 6.3.1 of [32]) that if

$$(20) \quad \begin{aligned} |\widehat{\psi}(w)| &\leq C |w|^\alpha (1 + |w|^2)^{-(b+a)/2} \quad \text{and} \\ |\widehat{\phi}(w)| &\leq C (1 + |w|^2)^{-b/2}, \quad a > \alpha, b > \alpha + 1, \end{aligned}$$

then  $\{\psi_{k,j}^{(i)}, U_{k,j}^{(i)}, \tilde{U}_{k,j}^{(i)}\}$  forms a WVS for  $A$ .

The conditions in (20) are weaker than (19) and allow more wavelets to be used to construct a WVS for a given operator  $A$  than condition (19) does. More importantly, the sufficient conditions (20) do not depend on the dimension  $d$ .

## VII. VAGUELETTE COEFFICIENTS AND VARIATIONAL PROBLEMS

We begin with a recursive algorithm to compute vaguelette coefficients. Here we shall only consider the case  $d = 2$ . Higher-dimensional algorithms follow easily from the 2-dimensional one. Throughout this section we assume that  $A$  is onto, i.e., the range of  $A$  is  $\mathcal{Y}$ .

We define  $\rho \in S'(\mathbb{R}^2)$  by  $\widehat{\rho}(\xi) = |\xi|^{2\alpha} \widehat{\Phi}(\xi)$ . Let  $\rho_{k,j} = 2^k \rho(2^k \cdot - j)$  and  $V_{k,j} = 2^{k\alpha} A \rho_{k,j}$ . Then by following the same argument used to show  $A^* A \pi_{k,j}^{(i)} = 2^{-2k\alpha} \psi_{k,j}^{(i)}$  in (17), one can show that  $A^* A \rho_{k,j} = 2^{-2k\alpha} \Phi_{k,j}$ . Thus

$$\begin{aligned} 2^{k\alpha} [Af, V_{k,j}] &= 2^{2k\alpha} [Af, A \rho_{k,j}] \\ &= 2^{2k\alpha} \langle f, A^* A \rho_{k,j} \rangle = \langle f, \Phi_{k,j} \rangle. \end{aligned}$$

With a similar argument, we can show that  $2^{k\alpha} [Af, U_{k,j}^{(i)}] = \langle f, \psi_{k,j}^{(i)} \rangle$  for  $i = 1, 2, 3$ . Therefore, by rewriting (6), we have

$$f = \sum_{k \geq k_0} \sum_{j,i} 2^{k\alpha} [Af, U_{k,j}^{(i)}] \psi_{k,j}^{(i)} + \sum_l 2^{k_0\alpha} [Af, V_{k_0,l}] \Phi_{k_0,l}.$$

Furthermore, since  $A$  is onto, for each  $Y \in \mathcal{Y}$  we have

$$\begin{aligned} 2^{(k-1)\alpha}[Y, V_{k-1,j}] &= \sum_{n_1, n_2} h_{n_1-2j_1} h_{n_2-2j_2} 2^{k\alpha}[Y, V_{k,n}], \\ 2^{(k-1)\alpha}[Y, U_{k-1,j}^{(1)}] &= \sum_{n_1, n_2} h_{n_1-2j_1} g_{n_2-2j_2} 2^{k\alpha}[Y, V_{k,n}], \\ 2^{(k-1)\alpha}[Y, U_{k-1,j}^{(2)}] &= \sum_{n_1, n_2} g_{n_1-2j_1} h_{n_2-2j_2} 2^{k\alpha}[Y, V_{k,n}], \\ 2^{(k-1)\alpha}[Y, U_{k-1,j}^{(3)}] &= \sum_{n_1, n_2} g_{n_1-2j_1} g_{n_2-2j_2} 2^{k\alpha}[Y, V_{k,n}], \end{aligned}$$

where we have used the FWT at the end of Section III. Applying these four equations to  $\{2^{k\alpha}[Y, V_{k,j}]\}$ , we can get  $\{2^{k\alpha}[Y, U_{k,j}^{(i)}]\}_{\{k_0 \leq k < m, j, i=1,2,3\}}$  and  $\{2^{k_0\alpha}[Y, V_{k_0,n}]\}_n$  from  $\{2^{m\alpha}[Y, V_{m,l}]\}_l$ .

We now consider a family of variational problems that naturally give rise to a parametrized solution class  $\tilde{f}_{\gamma, \beta_0}$ : Given a positive parameter  $\gamma$  and Besov space  $B_{1,1}^{\beta_0}(\mathbb{R}^d)$ , find a function  $\tilde{f}_{\gamma, \beta_0}$  that minimizes over all possible functions  $g$  in  $B_{1,1}^{\beta_0}(\mathbb{R}^d)$  the functional

$$(21) \quad \|Y - Ag\|_{\mathcal{Y}}^2 + 2\gamma|g|_{B_{1,1}^{\beta_0}}.$$

Using the  $L^2$ -stability of  $\{U_{k,j}^{(i)}\}$  and the smoothness characterization of  $\{\psi_{k,j}^{(i)}\}$ , we have

$$\begin{aligned} \|Y - Ag\|_{\mathcal{Y}}^2 &\asymp \sum_{k,j,i} |[Y - Ag, U_{k,j}^{(i)}]|^2 \\ &= \sum_{k,j,i} |[Y, U_{k,j}^{(i)}] - 2^{k\alpha}\langle g, A^* A \pi_{k,j}^{(i)} \rangle|^2 \text{ (by (18))} \\ &= \sum_{k,j,i} |[Y, U_{k,j}^{(i)}] - 2^{-k\alpha}\langle g, \psi_{k,j}^{(i)} \rangle|^2 \text{ (by (17))} \end{aligned}$$

and

$$|g|_{B_{1,1}^{\beta_0}} \asymp \sum_{k \geq k_0} \sum_{j,i} 2^{k(\beta_0-d/2)} |\langle g, \psi_{k,j}^{(i)} \rangle|.$$

Combining these sequence sums, we have following equivalent sequence sums to the functional (21):

$$(22) \quad \sum_{k,j,i} 2^{-2k\alpha} |2^{k\alpha}[Y, U_{k,j}^{(i)}] - \langle g, \psi_{k,j}^{(i)} \rangle|^2 + 2\gamma \sum_{k \geq k_0} \sum_{j,i} 2^{k(\beta_0-d/2)} |\langle g, \psi_{k,j}^{(i)} \rangle|,$$

which can be minimized by minimizing each term separately. Notice that  $a|b-x|^2 + 2c|x|$ , where  $a > 0$ ,  $c > 0$ , is minimized when  $x = \text{sign}(b)(|b| - c/a)^+$ , i.e., when  $x = S_{c/a}(b)$ . We apply this to the case where  $a = 2^{-2k\alpha}$ ,  $b = 2^{k\alpha}[Y, U_{k,j}^{(i)}]$ , and  $c = \gamma 2^{k(\beta_0-d/2)}$  for  $k \geq k_0$ ; then (22) is minimized by choosing the function  $\tilde{f}_{\gamma, \beta_0}^*$  such that

$$(23) \quad \tilde{f}_{\gamma, \beta_0}^* = \sum_{k,j,i} S_{\mu_k}(2^{k\alpha}[Y, U_{k,j}^{(i)}]) \psi_{k,j}^{(i)},$$

where

$$(24) \quad \mu_k = \begin{cases} \gamma 2^{k(\beta_0-d/2+2\alpha)}, & \text{for } k \geq k_0, \\ 0, & \text{for } k < k_0. \end{cases}$$

Since  $\mu_k = 0$  if  $k < k_0$ , one can rewrite (23) as

$$(25) \quad \begin{aligned} \tilde{f}_{\gamma, \beta_0}^* &= \sum_{k \geq k_0} \sum_{j,i} S_{\mu_k}(2^{k\alpha}[Y, U_{k,j}^{(i)}]) \psi_{k,j}^{(i)} \\ &\quad + \sum_l 2^{k_0\alpha}[Y, V_{k_0,l}] \Phi_{k_0,l}. \end{aligned}$$

With equivalence between function norm and wavelet sequence sums, we now suggest  $\tilde{f}_{\gamma, \beta_0}^*$  as a solution method for solving (2).

## VIII. DISCRETIZATION AND NOISE MODEL

We recall problem (2) of this paper:  $Y = Af + Z$ . In practice we have only finitely many data points  $\{Y_i\}_{i=0,1,\dots,N-1}$  for  $Y$ . With this data, we wish to approximate the true solution  $f$  by

$$(26) \quad \begin{aligned} \tilde{f}_{\gamma, \beta_0, m}^* &= \sum_{k \geq k_0} \sum_{j,i}^{m-1} S_{\mu_k}(2^{k\alpha}[Y, U_{k,j}^{(i)}]) \psi_{k,j}^{(i)} \\ &\quad + \sum_l 2^{k_0\alpha}[Y, V_{k_0,l}] \Phi_{k_0,l}. \end{aligned}$$

In this approach, we have three different types of errors. First, in approximating  $Y$  from  $\{Y_i\}_{i=0,1,\dots,N-1}$  and computing  $[Y, U_{k,j}^{(i)}]$  and  $[Y, V_{k_0,l}]$  from the approximation to  $Y$ , we cannot avoid a certain error. However, we shall ignore this type of error in this paper by assuming that we have sufficient observations  $\{Y_i\}_{i=0,1,\dots,N-1}$  to approximate the underlying  $Y$  with a negligible error. Second, an error is introduced in having finite  $m$  in  $\tilde{f}_{\gamma, \beta_0, m}^*$  (26). Notice that the smallest mean-square error we can have with  $\tilde{f}_{\gamma, \beta_0, m}^*$  (26) is equal to the sequence sum  $\sum_{k \geq m} \sum_{j,i} |\langle f, \psi_{k,j}^{(i)} \rangle|^2$ . The integer  $m$  in (26) is closely related to  $N$ , the number of observations. Increasing  $m$  for a fixed  $N$  does not necessarily give a better solution since it may generate another error in computing  $[Y, U_{k,j}^{(i)}]$  and  $[Y, V_{k_0,l}]$ . We assume that we can take a positive integer  $m$  in (26) such that  $2^{md} = N$ . Again, we ignore this type of error in this paper. Third, the solution method  $\tilde{f}_{\gamma, \beta_0, m}^*$  (26) itself makes an error in approximating

$$f_m = \sum_{k \geq k_0} \sum_{j,i}^{m-1} \langle f, \psi_{k,j}^{(i)} \rangle \psi_{k,j}^{(i)} + \sum_l \langle f, \Phi_{k_0,l} \rangle \Phi_{k_0,l}.$$

Notice that  $\|f_m - \tilde{f}_{\gamma, \beta_0, m}^*\|_{L^2}$  is

$$\begin{aligned} &\sum_{k=k_0}^{m-1} \sum_{j,i} |\langle f, \psi_{k,j}^{(i)} \rangle - S_{\mu_k}(2^{k\alpha}[Y, U_{k,j}^{(i)}])|^2 \\ &\quad + \sum_l |\langle f, \Phi_{k_0,l} \rangle - 2^{k_0\alpha}[Y, V_{k_0,l}]|^2. \end{aligned}$$

In this paper we are only interested in this type of error.

We assume a white noise model for the observation error  $Z$  in (2). Let

$$(27) \quad Z = \sigma W$$

for a constant  $\sigma > 0$ , where  $W$  is the white noise process defined on the underlying space of functions in  $\mathcal{Y}$ . This assumption naturally impose a noise model for the discrete data  $\{Y_i\}$  such that  $Y_i = (Af)_i + Z_i$ , where  $Z_i$  are independent and identically distributed as  $N(0, \sigma_0^2)$  for some  $\sigma_0 > 0$ . Since the number of observations is  $N = 2^{md}$ , it is reasonable to assume that  $\sigma_0^2 = 2^{md}\sigma^2$ .

We now examine the effect of this noise model (27) on  $[Y, U_{k,j}^{(i)}]$  and  $[Y, V_{k_0,l}]$  in  $\tilde{f}_{\gamma, \beta_0, m}^*$ . Notice that for the white noise process  $W$ ,  $[W, U_{k,j}^{(i)}]$  are mean zero Gaussian random variables for all  $k, j, i$  and  $\text{Var}[W, U_{k,j}^{(i)}] = \|U_{k,j}^{(i)}\|_{\mathcal{Y}}^2$ . Using the definition of  $U_{k,j}^{(i)}$  (see (18)), one can compute  $\|U_{k,j}^{(i)}\|_{\mathcal{Y}}^2 = \int_{\mathbb{R}^d} |\xi|^{2\alpha} |\widehat{\psi^{(i)}}(\xi)|^2 d\xi$ . Therefore,  $[Z, U_{k,j}^{(i)}]$  are mean zero Gaussian random variables with

$$\text{Var}[Z, U_{k,j}^{(i)}] = \sigma^2 \int_{\mathbb{R}^d} |\xi|^{2\alpha} |\widehat{\psi^{(i)}}(\xi)|^2 d\xi.$$

Similarly, we have mean zero Gaussian random variables  $[Z, V_{k_0,l}]$  such that

$$\text{Var}[Z, V_{k_0,l}] = \sigma^2 \int_{\mathbb{R}^d} |\xi|^{2\alpha} |\widehat{\Phi}(\xi)|^2 d\xi.$$

Notice that

$$\begin{aligned} 2^{k\alpha} [Y, U_{k,j}^{(i)}] &= 2^{k\alpha} \langle f, A^* U_{k,j}^{(i)} \rangle + 2^{k\alpha} [Z, U_{k,j}^{(i)}] \\ &= \langle f, \psi_{k,j}^{(i)} \rangle + 2^{k\alpha} [Z, U_{k,j}^{(i)}]. \end{aligned}$$

With a similar argument, we have

$$2^{k_0\alpha} [Y, V_{k_0,l}] = \langle f, \Phi_{k_0,l} \rangle + 2^{k_0\alpha} [Z, V_{k_0,l}].$$

Thus, the solution method  $\tilde{f}_{\gamma, \beta_0, m}^*$  (26) can be rewritten as

$$\begin{aligned} \sum_{k_0 \leq k < m} \sum_{j,i} S_{\mu_k} (\langle f, \psi_{k,j}^{(i)} \rangle + 2^{k\alpha} [Z, U_{k,j}^{(i)}]) \psi_{k,j}^{(i)} \\ + \sum_l (\langle f, \Phi_{k_0,l} \rangle + 2^{k_0\alpha} [Z, V_{k_0,l}]) \Phi_{k_0,l}. \end{aligned}$$

Before we close this section, for future use we define

$$c_i = \int_{\mathbb{R}^d} |\xi|^{2\alpha} |\widehat{\psi^{(i)}}(\xi)|^2 d\xi$$

for  $i = 1, \dots, 2^d - 1$ , and

$$c_0 = \int_{\mathbb{R}^d} |\xi|^{2\alpha} |\widehat{\Phi}(\xi)|^2 d\xi.$$

## IX. ERROR ESTIMATES

Here we calculate an upper bound of  $E\|f_m - \tilde{f}_{\gamma, \beta_0, m}^*\|_{L^2}^2$ . While doing that, we determine  $\beta_0$ , which gives the main effect on the wavelet shrinkage procedure, and then we choose  $\gamma$ , which corresponds to a more subtle, but still critical, part of the algorithm.

In [21] Donoho assumed the same white noise model as in this paper, and suggested

$$(28) \quad \begin{aligned} \tilde{f}_{(a_{k,j,i})} &= \sum_{k \geq k_0} \sum_{j,i} S_{a_{k,j,i}} (2^{k\alpha} [Y, U_{k,j}^{(i)}]) \psi_{k,j}^{(i)} \\ &+ \sum_l 2^{k_0\alpha} [Y, V_{k_0,l}] \phi_{k_0,l} \end{aligned}$$

as a solution method for solving (2). This is exactly the same as  $\tilde{f}_{\gamma, \beta_0}^*$  of (25) except for possibly different shrinkage parameters. For the performance of  $\tilde{f}_{(a_{k,j,i})}$ , he proved that

$$\inf_{(a_{k,j,i})} \sup_{f \in \mathcal{F}(C)} E\|f - \tilde{f}_{(a_{k,j,i})}\|_{L^2}^2 \asymp \sigma^{2r_M}$$

as  $\sigma \rightarrow 0$ , with rate exponent

$$(29) \quad r_M = \frac{\beta}{\beta + d/2 + \alpha}$$

for the true solution  $f$ , which is known to lie in a ball  $\mathcal{F}(C) = \{f \mid \|f\|_{B_{q,p}^\beta} \leq C\}$  of the Besov space  $B_{q,p}^\beta(\mathbb{R}^d)$  with  $\beta > (2\alpha + d)(1/p - 1/2)$ . He also showed that this wavelet shrinkage method attains the optimal rate of convergence:

$$\begin{aligned} \inf_{(a_{k,j,i})} \sup_{f \in \mathcal{F}(C)} E\|f - \tilde{f}_{(a_{k,j,i})}\|_{L^2}^2 \\ \leq C \inf_{\tilde{f}} \sup_{f \in \mathcal{F}(C)} E\|f - \tilde{f}\|_{L^2}^2, \end{aligned}$$

where  $\tilde{f}$  ranges over all possible methods.

This asymptotical result implies that the wavelet shrinkage method (28) with the optimal shrinkage parameters  $(a_{k,j,i})$  is the best *estimator* in *minimax* sense. (For definitions of these statistical terms, see, e.g., [3].) However, Donoho in [21] did not give a method for finding those parameters.

In [30] Kolaczyk used the shrinkage parameters

$$(30) \quad a_{k,j,i} = \sqrt{2 \log(2^{2k})} 2^{k/2} \sigma c_i^{1/2},$$

in (28) for tomographic reconstruction. For the definition of  $c_i$ , see Section VIII. This choice of shrinkage amount is motivated by the VisuShrink method of Donoho and Johnstone [22]. For details, see, e.g., [22] and [30]. However, Kolaczyk [30] did not give a proof that the shrinkage parameter (30) gives the optimal rate of convergence  $r_M$  (29).

We propose to find an explicit shrinkage scheme that provides the optimal rate of convergence  $r_M$  (29). To do so, we assume that the true solution  $f$  is in the Besov spaces  $B_{p,p}^\beta(\mathbb{R}^d)$  with

$$p = \frac{2\alpha + d}{\beta + d/2 + \alpha}$$

for some positive  $\beta$ . (It will turn out that this one-parameter family of spaces (depending on  $\beta$ ) have minimal smoothness for the inversion of  $A$ .) We also assume that  $f$  is bounded and compactly supported in the unit cube of  $\mathbb{R}^d$ . Thus we need only  $j$  and  $l$  for which  $\langle f, \psi_{k,j}^{(i)} \rangle \neq 0$  and  $\langle f, \Phi_{k_0,l} \rangle \neq 0$  in (26). Moreover, the number of such  $j$

$\leq C2^{kd}$  for fixed  $k$  and  $i$ , and that of such  $l \leq C2^{k_0d}$  for a constant  $C$ . Having these results in mind, we start to estimate an upper bound of  $E\|f_m - \tilde{f}_{\gamma, \beta_0, m}^*\|_{L^2}^2$ .

Notice that  $E\|f_m - \tilde{f}_{\gamma, \beta_0, m}^*\|_{L^2}^2$  is

$$\begin{aligned} & \sum_{k=k_0}^{m-1} \sum_{\langle f, \psi_{k,j}^{(i)} \rangle \neq 0} E|\langle f, \psi_{k,j}^{(i)} \rangle - S_{\mu_k}(\langle f, \psi_{k,j}^{(i)} \rangle) + 2^{k\alpha}[Z, U_{k,j}^{(i)}]|^2 \\ & \quad + \sum_{\langle f, \Phi_{k_0,l} \rangle \neq 0} 2^{2k_0\alpha} \text{Var}[Z, V_{k_0,l}], \end{aligned}$$

where we have used the expectation operator  $E$  to deal with the statistical noise model. We note (see Section VIII) that  $\text{Var}[Z, V_{k_0,l}] = c_0\sigma^2$  for all  $l$ . Thus  $E\|f_m - \tilde{f}_{\gamma, \beta_0, m}^*\|_{L^2}^2$  is bounded by

$$\begin{aligned} & \sum_{k=k_0}^{m-1} \sum_{\langle f, \psi_{k,j}^{(i)} \rangle \neq 0} E|\langle f, \psi_{k,j}^{(i)} \rangle - S_{\mu_k}(\langle f, \psi_{k,j}^{(i)} \rangle) + 2^{k\alpha}[Z, U_{k,j}^{(i)}]|^2 \\ & \quad + Cc_02^{2k_0(2\alpha+d)}\sigma^2, \end{aligned}$$

where we have used the fact that the number of  $l$  for which  $\langle f, \Phi_{k_0,l} \rangle \neq 0$  is less than  $C2^{k_0d}$  for a constant  $C$ .

Let  $\Lambda_k = \{(j, i) \mid |\langle f, \psi_{k,j}^{(i)} \rangle| > \mu_k\}$  and  $\tilde{\Lambda}_k = \{(j, i) \mid |\langle f, \psi_{k,j}^{(i)} \rangle| \leq \mu_k\}$ .

We now use the following inequality (see [5] and [23]): If  $X \sim N(0, \tau^2)$ , then

$$E|t - S_\mu(t + X)|^2 \leq \begin{cases} \mu^2 + \tau^2, & |t| > \mu, \\ t^2 + E|S_\mu(X)|^2, & |t| \leq \mu, \end{cases}$$

where

$$E|S_\mu(X)|^2 = 2\tau^2 \int_{y > \frac{\mu}{\tau}} (y - \frac{\mu}{\tau})^2 P(y) dy,$$

where  $P(y) = \frac{1}{\sqrt{2\pi}}e^{-y^2/2}$ . We apply above inequality to the case where  $t = \langle f, \psi_{k,j}^{(i)} \rangle$ ,  $X = 2^{k\alpha}[Z, U_{k,j}^{(i)}]$ ,  $\tau^2 = 2^{2k\alpha}\sigma^2$ , and  $\mu = \mu_k$  (see (24)). Then we have

$$(31) \quad E\|f_m - \tilde{f}_{\gamma, \beta_0, m}^*\|_{L^2}^2 \leq C \left( S_0(f) + S_1(f) + S_2(f) + S_3(f) + S_4(f) \right),$$

where

$$\begin{aligned} S_0(f) &= c_02^{k_0(2\alpha+d)}\sigma^2, \\ S_1(f) &= \sum_{k=k_0}^{m-1} \sum_{(j,i) \in \Lambda_k} \mu_k^2, \\ S_2(f) &= \sum_{k=k_0}^{m-1} \sum_{(j,i) \in \Lambda_k} 2^{2k\alpha}\sigma^2, \\ S_3(f) &= \sum_{k=k_0}^{m-1} \sum_{(j,i) \in \tilde{\Lambda}_k} |\langle f, \psi_{k,j}^{(i)} \rangle|^2, \quad \text{and} \end{aligned}$$

$$S_4(f) = \sum_{k=k_0}^{m-1} \sum_{(j,i) \in \tilde{\Lambda}_k} 2^{2k\alpha+1}\sigma^2 \int_{t > t_k} (t - t_k)^2 P(t) dt,$$

where

$$t_k = 2^{k(\beta_0-d/2+\alpha)} \frac{\gamma}{\sigma}.$$

For  $S_1(f)$ , notice that

$$\begin{aligned} S_1(f) &= \sum_{k=k_0}^{m-1} \sum_{(j,i) \in \Lambda_k} \mu_k^{2-p} \cdot \mu_k^p \\ &= \sum_{k=k_0}^{m-1} 2^{-ksp} \mu_k^{2-p} 2^{ksp} \sum_{(j,i) \in \Lambda_k} \mu_k^p, \end{aligned}$$

where  $s = \beta + d(1/2 - 1/p)$ . Let

$$(32) \quad a = \frac{\gamma}{\sigma}.$$

Since  $\mu_k = a\sigma 2^{k(\beta_0-d/2+2\alpha)}$  (see (24) and (32)), one has

$$(33) \quad S_1(f) \leq a^{2-p}\sigma^{2-p}T(\beta_0) \sum_{k=k_0}^{m-1} 2^{ksp} \sum_{(j,i) \in \Lambda_k} \mu_k^p,$$

where

$$(34) \quad T(\beta_0) = \max_{k_0 \leq k < m} \left\{ 2^{k(\beta_0(2-p)+4\alpha-2\alpha p-\beta p)} \right\}.$$

Thus we have

$$(35) \quad S_1(f) \leq |f|_{B_{p,p}^\beta}^p a^{2-p}\sigma^{2-p}T(\beta_0),$$

because in (33),

$$\begin{aligned} \sum_{k=k_0}^{m-1} 2^{ksp} \sum_{(j,i) \in \Lambda_k} \mu_k^p &\leq \sum_{k=k_0}^{m-1} 2^{ksp} \sum_{j,i} |\langle f, \psi_{k,j}^{(i)} \rangle|^p \\ &\leq |f|_{B_{p,p}^\beta}^p. \end{aligned}$$

For  $S_2(f)$ , since  $\sigma^2 = a^{-2}2^{-2k(\beta_0-d/2+2\alpha)}\mu_k^2$  (see (24) and (32)), one has

$$S_2(f) \leq \sum_{k_0=k}^{m-1} \sum_{(j,i) \in \Lambda_k} 2^{2k\alpha} a^{-2} 2^{-k(2\beta_0-d+4\alpha)} \mu_k^2.$$

We use the same argument as for  $S_1(f)$ . Then we have

$$S_2(f) \leq \sum_{k_0=k}^{m-1} a^{-2} 2^{-k(sp+2\beta_0-d+2\alpha)} \mu_k^{2-p} 2^{ksp} \sum_{(j,i) \in \Lambda_k} \mu_k^p.$$

Moreover, since  $a^{-2}\mu_k^{2-p} = a^{-p}\sigma^{2-p}2^{k(\beta_0-d/2+2\alpha)(2-p)}$ , one has

$$S_2(f) \leq a^{-p}\sigma^{2-p}T^*(\beta_0) \sum_{k_0=k}^{m-1} 2^{ksp} \sum_{(j,i) \in \Lambda_k} \mu_k^p,$$

where

$$(36) \quad T^*(\beta_0) = \max_{k_0 \leq k < m} \left\{ 2^{-k(\beta_0 p + \beta p - 2\alpha - d + 2\alpha p)} \right\}.$$

With a similar argument used for  $S_1(f)$ , we have

$$\sum_{k=k_0}^{m-1} 2^{ksp} \sum_{(j,i) \in \Lambda_k} \mu_k^p \leq |f|_{B_{p,p}^\beta}^p.$$

Thus

$$(37) \quad S_2(f) \leq |f|_{B_{p,p}^\beta}^p a^{-p} \sigma^{2-p} T^*(\beta_0).$$

For  $S_3(f)$ , we can bound them as follows:

$$\begin{aligned} S_3(f) &= \sum_{k_0=k}^{m-1} \sum_{(j,i) \in \tilde{\Lambda}_k} |\langle f, \psi_{k,j}^{(i)} \rangle|^2 \\ &\leq \sum_{k_0=k}^{m-1} \sum_{(j,i) \in \tilde{\Lambda}_k} |\langle f, \psi_{k,j}^{(i)} \rangle|^p \cdot \mu_k^{2-p} \\ &= \sum_{k_0=k}^{m-1} 2^{-ksp} \mu_k^{2-p} 2^{ksp} \sum_{(j,i) \in \tilde{\Lambda}_k} |\langle f, \psi_{k,j}^{(i)} \rangle|^p. \end{aligned}$$

We now use the same argument used for  $S_1(f)$ . Then we have

$$(38) \quad S_3(f) \leq |f|_{B_{p,p}^\beta}^p a^{2-p} \sigma^{2-p} T(\beta_0).$$

Combining (35), (37), and (38), we can bound  $S_1(f) + S_2(f) + S_3(f)$  by

$$(49) \quad |f|_{B_{p,p}^\beta}^p \sigma^{2-p} \left( 2a^{2-p} T(\beta_0) + a^{-p} T^*(\beta_0) \right).$$

We now determine  $\beta_0$ , which minimizes  $S_1(f) + S_2(f) + S_3(f)$  for a fixed  $a$ . To do so, we examine the exponents of 2 in  $T(\beta_0)$  and  $T^*(\beta_0)$ . Obviously, to reduce the contributions from  $S_1(f) + S_2(f) + S_3(f)$ , it is desirable if our choice of  $\beta_0$  satisfies

$$\begin{aligned} \beta_0(2-p) + 4\alpha - 2\alpha p - \beta p &\leq 0 \quad \text{and} \\ \beta_0 p + \beta p - 2\alpha - d + 2\alpha p &\geq 0, \end{aligned}$$

so that  $T(\beta_0)$  (34) and  $T^*(\beta_0)$  (36) are bounded independently of  $m$ . Simple computations reveal that it is only possible when

$$\beta_0 = \beta_0^* = d/2 - \alpha.$$

With this  $\beta_0^*$ , we have  $T(\beta_0^*) = T^*(\beta_0^*) = 1$ . Therefore, from (39), we have

$$(40) \quad S_1(f) + S_2(f) + S_3(f) \leq |f|_{B_{p,p}^\beta}^p \sigma^{2-p} (2a^{2-p} + a^{-p}).$$

For  $S_4(f)$ , with  $\beta_0^* = d/2 - \alpha$ ,  $t_k \leq a$ , we note that as a function of  $x$ ,  $\int_{t>x} (t-x)^2 P(t) dt$  is strictly decreasing for  $x > 0$ . Thus by bounding  $|\tilde{\Lambda}_k|$  by  $(2^d - 1)2^{kd}$ , we have

$$S_4(f) \leq C_1 \sum_{k=k_0}^{m-1} 2^{k(2\alpha+d)} \sigma^2 \int_{t>a} (t-a)^2 P(t) dt,$$

where  $C_1 = 2(2^d - 1)$ . Finally, combining (31), (40), and above inequality, we can bound  $E\|f - \tilde{f}_{\gamma, \beta_0^*, m}^*\|_{L^2}^2$  by

$$(41) \quad |f|_{B_{p,p}^\beta}^p \sigma^{2-p} (2a^{2-p} + a^{-p}) + C_2 \cdot 2^{m(2\alpha+d)} \sigma^2 \int_{t>a} (t-a)^2 P(t) dt,$$

where  $C_2 = 2(2^d - 1)/(2^{2(d/2+\alpha)} - 1)$ . Here we ignored the contribution from  $S_0(f)$ , because it is small and independent of  $m$ .

Formula (41) is our main bound of  $E\|f_m - \tilde{f}_{\gamma, \beta_0^*, m}^*\|_{L^2}^2$ . We emphasize that given only two parameters characterizing the smoothness of  $f$  ( $\beta$  and  $|f|_{B_{p,p}^\beta}$ ), the known parameter  $\alpha$  that determines the ill-posedness of the reconstruction procedure, and an estimate of the standard deviation  $\sigma$  of the noise in the observation procedure, one can numerically compute the minimum point  $a^*$  of the equation (41) as a function of  $a$  and use  $\gamma^* = a^* \sigma$  to determine  $\tilde{f}_{\gamma^*, \beta_0^*, m}^*$  that minimizes our upper bound on the error.

To have an asymptotic result for (41), we use following inequality; for any  $x \geq 0$ ,

$$(42) \quad \int_{t>x} (t-x)^2 P(t) dt \leq \frac{\sqrt{2\pi}}{2} P(x).$$

This is rather rough estimation since one can replace the right hand side of (42) by  $2x^{-3}P(x)$  for  $x \geq 1$  (see, e.g., [5]). Using (42), we can bound the second term in (41) by a constant multiple of

$$2^{m(2\alpha+d)} \sigma^2 e^{-a^2/2}.$$

We now compare two dominant terms  $|f|_{B_{p,p}^\beta}^p \sigma^{2-p}$  and  $2^{m(2\alpha+d)} \sigma^2 e^{-a^2/2}$  to get a simple approximation to the critical  $a$ . If we assume that  $\sigma_0 \leq |f|_{B_{p,p}^\beta}$ , where  $\sigma_0^2 = 2^{md} \sigma^2$ , and  $2^{md}$  is large enough that

$$\frac{|f|_{B_{p,p}^\beta}}{\sigma_0} \leq 2^{m(\beta+d+\alpha)},$$

then by setting

$$|f|_{B_{p,p}^\beta}^p \sigma^{2-p} = 2^{m(2\alpha+d)} \sigma^2 e^{-a^2/2},$$

we can get the critical point  $a$  such that

$$a = \sqrt{D_1 \log 2^{md} - D_2 \log \frac{|f|_{B_{p,p}^\beta}}{\sigma_0}},$$

where  $D_1 = \frac{4\alpha}{d} + \frac{2\beta}{\beta+d/2+\alpha}$  and  $D_2 = 2 \frac{2\alpha+d}{\beta+d/2+\alpha}$ . With this  $a$ , we have

$$\begin{aligned} E\|f_m - \tilde{f}_{\gamma, \beta_0^*, m}^*\|_{L^2}^2 &\leq C |f|_{B_{p,p}^\beta}^p \left( \frac{\sigma_0^2}{2^{md}} \right)^{\beta/(\beta+d/2+\alpha)} \\ &\quad \times \left( \left[ \left( \frac{4\alpha}{d} + \frac{2\beta}{\beta+d/2+\alpha} \ln 2^{md} \right)^R + 1 \right] \right), \end{aligned}$$

where

$$R = \begin{cases} \beta/(\beta+d/2+\alpha), & \text{if } a \geq 1, \\ (d/2+\alpha)/(\beta+d/2+\alpha), & \text{if } a < 1. \end{cases}$$

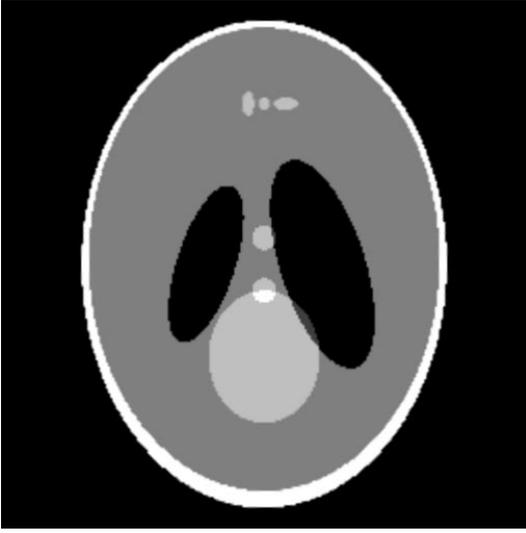


FIG. 1.  $512 \times 512$  phantom image

## X. COMPUTATIONS: TOMOGRAPHIC RECONSTRUCTION

We conducted tomographic reconstruction ( $\alpha = 1/2$ ) experiments using  $\tilde{f}_{\gamma^*, \beta_0^*, m}^*$  in (26) with  $\beta_0^* = 1/2$  and  $\gamma^* = a^* \sigma$ , where  $a^*$  is chosen as the (numerical) minimum point of (41). Our main conclusion is that this shrinkage parameter leads to smaller errors and better reconstruction than using the parameter suggested by Kolaczyk [30]. The wavelet shrinkage method often exhibits certain artifacts, which do not appear in reconstructed images from the filtered backprojection method. Those artifacts in the wavelet shrinkage method are largely due to the fact that the wavelet shrinkage method is neither translationally nor rotationally invariant. However, a “translation-rotation averaging technique” associated with the wavelet shrinkage method reduces those artifacts dramatically, and outperforms not only the standard wavelet shrinkage method but also the traditional filtered backprojection method in the mean square error sense.

Our main computations are applied to the phantom image  $f_9$  of  $512 \times 512$  pixels in Figure 1. With the assumption that the true intensity field  $f$  of the digital image  $f_9$  is bounded, we scaled  $f_9$  to have 255 as the brightest pixel and 0 as the darkest one. We computed Radon projection data,  $\mathcal{R}f$ , at 512 uniformly spaced angles, and at 512 uniformly spaced points for each angle. Thus, the number of data values,  $N = 262,144$ , and  $m = 9$  in  $\tilde{f}_{\gamma^*, \beta_0^*, m}^*$ .

We added i.i.d. Gaussian noise with standard deviations  $\sigma_0^{(10)} = 22.6113$ ,  $\sigma_0^{(15)} = 12.7152$ ,  $\sigma_0^{(20)} = 7.1503$ ,  $\sigma_0^{(25)} = 4.0209$ , and  $\sigma_0^{(30)} = 2.2611$  to  $\mathcal{R}f$ . Here the superscripts in standard deviations denote *signal-to-noise ratio* (SNR) defined by

$$\text{SNR} = 10 \log_{10} \left( \frac{\sum_{i=0}^{511} \sum_{j=0}^{511} |\mathcal{R}f(\theta_i, u_j)|^2}{262144 \times \sigma_0^2} \right).$$

As we mentioned before, we can modify the definition of a WVS to incorporate biorthogonal wavelets. We use biorthogonal wavelets  ${}_{3\phi}$ ,  ${}_{3\psi}$ ,  ${}_{3,\phi}$ , and  ${}_{3,\psi}$  illustrated

on page 276 of [10]. These wavelets and dual wavelets have the required regularity and zero-moment properties to generate a WVS of the Radon transform in two dimension. For details, see Theorem 6.3.1 of [32]. We modified the biorthogonal wavelets at the boundary in a way equivalent to assuming that the phantom image is periodic, and rescaled them to make  $c_i = 1$  for  $i = 1, 2, 3$ . We also choose  $k_0 = 5$  in (26).

We assume that  $f \in B_{p,p}^\beta(\mathbb{R}^2)$ , where  $\beta > 0$  and  $p = \frac{3}{\beta+3/2}$ . Notice that when  $|\langle f, \psi_{k,j}^{(i)} \rangle| < 2^{k/2}\gamma$ ,

$$\begin{aligned} |\langle f, \psi_{k,j}^{(i)} \rangle|^2 &= |\langle f, \psi_{k,j}^{(i)} \rangle|^{2-p} \cdot |\langle f, \psi_{k,j}^{(i)} \rangle|^p \\ &\leq \gamma^{2-p} \sum_{k \geq 5} 2^{k(2-p)/2} \sum_{j,i} |\langle f, \psi_{k,j}^{(i)} \rangle|^p. \end{aligned}$$

Since  $(\beta + 2(1/2 - 1/p))p = (2-p)/2$ , if we define

$$E(\gamma) = \left( \sum_{k \geq 5} \sum_{|\langle f, \psi_{k,j}^{(i)} \rangle| < 2^{k/2}\gamma} |\langle f, \psi_{k,j}^{(i)} \rangle|^2 \right)^{1/2},$$

then we have

$$(43) \quad E(\gamma)^2 \leq \gamma^{2-p} |f|_{B_{p,p}^\beta}^p.$$

On the other hand, if we denote the number of  $(j, i)$  for which  $|\langle f, \psi_{k,j}^{(i)} \rangle| \geq 2^{k/2}\gamma$  by  $N_k(\gamma)$ , then we have

$$(44) \quad \begin{aligned} \sum_{k \geq 5} N_k(\gamma) (2^{k/2}\gamma)^p 2^{k(\beta+2(1/2-1/p))p} \\ \leq \sum_{k \geq 5} 2^{k(\beta+2(1/2-1/p))p} \sum_{j,i} |\langle f, \psi_{k,j}^{(i)} \rangle|^p = |f|_{B_{p,p}^\beta}^p, \end{aligned}$$

since  $2^{k(\beta+2(1/2-1/p)+1/2)p} = 2^k$ . We define

$$N(\gamma) = \sum_{k \geq 5} N_k(\gamma) 2^k.$$

Then from (44), we have

$$(45) \quad \gamma^p \leq N(\gamma)^{-1} |f|_{B_{p,p}^\beta}^p.$$

Combining (43) and (45), we have

$$(46) \quad E(\gamma) \leq N(\gamma)^{-\beta/3} |f|_{B_{p,p}^\beta}.$$

It is remarked in [8] that (46) is invertible, i.e., if we observe

$$E(\gamma) \leq CN(\gamma)^{-\beta/3}$$

for some  $\beta$  and  $C$ , then one can conclude that  $f \in B_{p,p}^\beta(\mathbb{R}^2)$ ,  $p = \frac{3}{\beta+3/2}$ , and one can define an equivalent semi-norm on  $B_{p,p}^\beta(\mathbb{R}^2)$  such that in this semi-norm  $|f|_{B_{p,p}^\beta} = C$ . (This statement is not correct, but is close enough to the truth to be used in practice; see [8] for the precise statement.) We use (46) in estimating the smoothness order  $\beta$  of  $f$ .

We computed  $E(\gamma)$  and  $N(\gamma)$  for several  $\gamma$ , and estimated  $\beta$  and  $|f|_{B_{p,p}^\beta}$  from the log-log graph of  $E(\gamma)$  versus  $N(\gamma)$ . (Cf. the smoothness estimates in [5].) With this approach, we estimated  $\beta \approx 0.7745$  and  $|f|_{B_{p,p}^\beta} \approx 166.4654$  with  $p = 1.3189$ .

We compute the minimum point of (41) as a function of  $a$ , with  $p = 1.3189$ ,  $\alpha = 1/2$ , and  $|f|_{B_{p,p}^\beta}^p$  ( $\approx 850.5585$ ), by the bisection method for each  $\sigma = 2^{-9}\sigma_0$ . We denote the resulting algorithm  $\tilde{f}_{\gamma^*,\beta_0^*,m}^*$  by  $f_C$ , i.e.,

$$\sum_{k \geq 5} \sum_{j,i} S_{a\sigma 2^{k/2}} (2^{k/2} [Y, U_{k,j}^{(i)}]) \psi_{k,j}^{(i)} + \sum_l 2^{2.5} [Y, V_{k_0,l}] \Phi_{5,l},$$

where  $a$  is chosen as the numerical minimum point of (41) for given  $\sigma = 2^{-9}\sigma_0$ , and we denote  $E \|f_9 - f_C\|_{L^2}^2$  by  $E_C$ .

We also consider a rotational averaging version  $f_C^{(R)}$  of  $f_C$ , where superscript  $R$  refers to the rotation number. For instance,

$$f_C^{(2)} = \frac{1}{2} \left[ Q_{\frac{\pi}{4}} \left( \sum_{k \geq 5} \sum_{j,i} S_{a\sigma 2^{k/2}} (2^{k/2} [Y(\theta - \frac{\pi}{4}, \cdot), U_{k,j}^{(i)}]) \psi_{k,j}^{(i)} + \sum_l 2^{2.5} [Y(\theta - \frac{\pi}{4}, \cdot), V_{k_0,l}] \Phi_{4,l} \right) + f_C \right],$$

where  $Q_{\frac{\pi}{4}}$  denotes rotation by  $\frac{\pi}{4}$  and  $a$  is, again, chosen as the numerical minimum point of (41) for each  $\sigma$  with same estimated values of  $\alpha$ ,  $p$ , and  $|f|_{B_{p,p}^\beta}^p$  used in  $f_C$ . The

mean square error  $E \|f_9 - f_C^{(R)}\|_{L^2}^2$  is denoted by  $E_C^{(R)}$ . Notice that  $f_C^{(1)} = f_C$  and  $E_C^{(1)} = E_C$ , because the shrinkage method associated with the wavelets constructed using tensor products is invariant with respect to rotation by  $\pi/2$ .

We denote the translation averaging version of  $f_C^{(R)}$  by  $f_T^{(R)}$ , where  $f_T^{(R)}$  uses the same rotational averaging techniques as  $f_C^{(R)}$ , but the translation-invariant wavelet shrinkage method introduced by Coifman and Donoho [9] in the shrinkage step. The mean square error  $\|f_9 - f_T^{(R)}\|_{L^2}^2$  is denoted by  $E_T^{(R)}$ .

In [30] Kolaczyk used the shrinkage parameter (30) (with  $c_i = 1$  for  $i = 1, 2, 3$ ) in (28) for tomographic reconstruction. This choice of shrinkage parameter is motivated by the VisuShrink method of Donoho and Johnstone [22]. For comparison purposes, we define

$$f_V = \sum_{k=7}^8 \sum_{j,i} S_{a_{k,j,i}} (2^{k/2} [Y, U_{k,j}^{(i)}]) \psi_{k,j}^{(i)} + \sum_l 2^{3.5} [Y, V_{7,l}] \Phi_{7,l}.$$

Notice that  $f_V$  only shrinks the two highest levels of wavelet coefficients. (In [30] Meyer's wavelets are used for experiments.) The mean square error  $\|f_9 - f_V\|_{L^2}^2$  is denoted by  $E_V$ .

We also consider the filtered backprojection method using the Hamming weight filter

$$w_H(n) = \begin{cases} 0.5 + 0.5 \cos(n\pi/N_s), & n \in [-N_s, N_s], \\ 0, & \text{otherwise,} \end{cases}$$

where  $n$  is the frequency number. For details, see, e.g., [42]. We define  $f_F$  by

$$(47) \quad f_F = \int_{\mathbb{R}^2} e^{2\pi i \xi \cdot x} w_H(\xi) |\xi| \widehat{\mathcal{R}^* Y}(\xi) d\xi.$$

We use  $E_F$  to denote the mean square error of  $f_F$  with  $N_s = 512$ . Obviously, the algorithm  $f_F$  depends on the filter size  $N_s$ . With a similar argument used in Section V, we expect that the filter size  $N_s$  should be decreased for a fixed image as the amount of noise increases. We denote by  $E_F^\#$  the mean square error of  $f_F$  with an optimal filter size. In our experiments, we computed numerically an optimal filter size of  $f_F$  for each noise level.

In order to compute vaguelette coefficients in  $f_C^{(R)}$ ,  $f_T^{(R)}$ , and  $f_V$ , we first compute  $2^{4.5} [Y, V_{9,l}]$  by

$$2^{-9} \int_{\mathbb{R}^2} |\xi| \widehat{\mathcal{R}^* Y}(\xi) e^{2\pi i \xi \cdot l 2^{-9}} \widehat{\Phi}(2^{-9}\xi) d\xi,$$

and use the recursive algorithm described in Section VII to get  $\{2^{k/2} [Y, U_{k,j}^{(i)}]\}_{\{5 \leq k < 9, j\}}$  and  $\{2^{2.5} [Y, V_{5,l}]\}_l$ .

For  $\mathcal{R}^* Y$ , we note that

$$\begin{aligned} \widehat{\mathcal{R}^* \mathcal{R} f}(w \cos \theta, w \sin \theta) &= |w|^{-1} \widehat{f}(w \cos \theta, w \sin \theta) \\ &= |w|^{-1} \mathcal{R} f(\theta, \cdot)^\wedge(w), \end{aligned}$$

where we have used the Fourier Slice Theorem (9). Thus we can compute  $\mathcal{R}^* Y$  by

$$\widehat{\mathcal{R}^* Y}(w \cos \theta, w \sin \theta) = |w|^{-1} Y(\theta, \cdot)^\wedge(w).$$

In numerical computations involving the Fourier transform, we used the *fast Fourier transform* (see, e.g., [28]). To reduce the aliasing effect in the fast Fourier transform, we *zero-padded* the projection data of  $512 \times 512$  to  $512 \times 1024$  (see, e.g., [42]).

TABLE 1  
Errors for Different Methods and Noise Levels

SNR	10	15	20	25	30
$E$	51874	18176	6747	2482	876
$E_V$	2721	991	408	219	156
$E_F$	7974	2942	1038	396	189
$E_F^\#$	555	365	249	180	139
$E_C^{(1)}$	1168	638	366	231	161
$E_C^{(2)}$	551	343	232	173	141
$E_C^{(4)}$	501	315	218	167	139
$E_T^{(1)}$	666	387	247	175	137
$E_T^{(2)}$	456	291	206	161	136
$E_T^{(4)}$	452	288	205	160	136

Table 1 contains the results of our tests. We calculated mean square errors  $E_V$ ,  $E_F$ ,  $E_C^{(R)}$ , and  $E_T^{(R)}$  with  $R = 1, 2$ , and 4 for five different noise levels with SNR = 10, 15, 20, 25, and 30. The first row in Table 1 shows the mean square errors of the direct backprojection method (with neither filtering nor wavelet shrinkage) for each noise level. We also report that  $\|f_9\|_{L^2}^2 = 9769$ .

The results in Table 1 shows that  $f_T^{(R)}$  outperforms  $f_F$ ,  $f_F^\#$ ,  $f_V$ , and  $f_C^{(R)}$ . The performance of  $f_C^{(R)}$  and  $f_T^{(R)}$  is improved as the rotation number  $R$  increases. This improvement comes at the expense of greater computations. Since the translation-invariant wavelet transform requires

the same computational complexity as the fast Fourier transform,  $f_T^{(R)}$  needs, roughly speaking,  $R + 1$  times the computations as  $f_F$  does. On the other hand, the traditional wavelet transform can be performed in a time proportional to the number of terms to be processed. Thus, as long as the rotation number  $R$  is small, (for  $512 \times 512$  phantom images,  $R \leq 8$ ), the number of computations in  $f_C^{(R)}$  is less than twice of that in  $f_F$ .

Figures 2, 3, 4, 5, and 6 show  $512 \times 512$  reconstructions based on  $f_C^{(R)}$ ,  $f_T^{(R)}$ ,  $f_F$ ,  $f_F^\#$ , and  $f_V$ , respectively, for noise level with SNR 20. The algorithm  $f_C^{(1)}$  has certain artifacts in the reconstructed image Figure 2-a, where we can see “square blocks” near edges. Such artifacts are largely due to the fact that the wavelet shrinkage method is not rotationally invariant. Thus by doing rotational averaging, we can reduce those artifacts dramatically. See Figure 2-b and 2-c.

As compared with the reconstructed images in Figure 3-a, 3-b, and 3-c, those in Figure 2-a, 2-b, and 2-c still suffer from certain artifacts. Most of these artifacts disappear in Figure 3 after applying the translation-invariant wavelet shrinkage algorithm introduced by Coifman and Donoho [9]. Notice that  $f_T^{(R)}$  is invariant with respect to all translations (in pixel units). However,  $f_C^{(R)}$  and  $f_T^{(R)}$  with  $R \geq 2$  are invariant only with respect to rotations of integer multiple of  $\pi/(2R)$ .

TABLE 2  
Projected and Optimal Shrinkage Parameters

SNR	10	15	20	25	30
$a^*$	1.4695	1.3589	1.2438	1.1251	1.0060
$a_C^{(1)}$	2.5000	2.2000	1.9000	1.7000	1.4000
$a_C^{(2)}$	1.9000	1.5000	1.3000	1.0000	0.8000
$a_C^{(4)}$	1.7000	1.4000	1.2000	1.0000	0.8000
$a_T^{(1)}$	2.1000	1.9000	1.7000	1.5000	1.4000
$a_T^{(2)}$	1.6000	1.4000	1.2000	1.0000	0.8000
$a_T^{(4)}$	1.6000	1.4000	1.2000	1.0000	0.8000

In Table 2, the first row contains the shrinkage parameters  $a^*$  used to compute  $f_C^{(R)}$  and  $f_T^{(R)}$ . As we mentioned before, the shrinkage parameter  $a^*$  is obtained by computing the minimum of (41) numerically as a function of  $a$ . For comparison purpose, in Table 2 we also give a near-optimal shrinkage parameter for each noise level and rotation number. For example,  $a_C^{(R)}$  is the best constant for  $f_C^{(R)}$  in  $\{a|a = n/10, n = 0, 1, \dots, 40\}$  in the sense that it gives the smallest mean square error for each given noise level. We define  $a_T^{(R)}$  in a similar manner. The shrinkage parameters suggested by this paper (the first row in Table 2) are closer to the real optimal shrinkage parameters as the rotation number increases.

Figures 4 and 5 show  $512 \times 512$  reconstructions based on the filtered backprojection method  $f_F$  for noise level SNR 20. The filter size  $N_s = 512$  and  $N_s = 192$  are used in Figure 4 and 5, respectively. As one can see from the

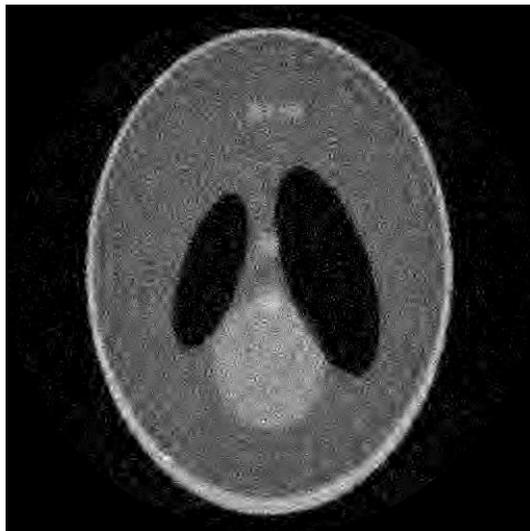


FIG. 2-A.  $f_C^{(1)}$ .

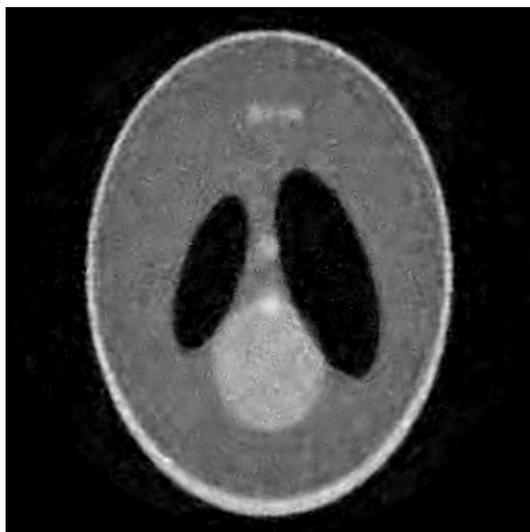


FIG. 2-B.  $f_C^{(2)}$ .

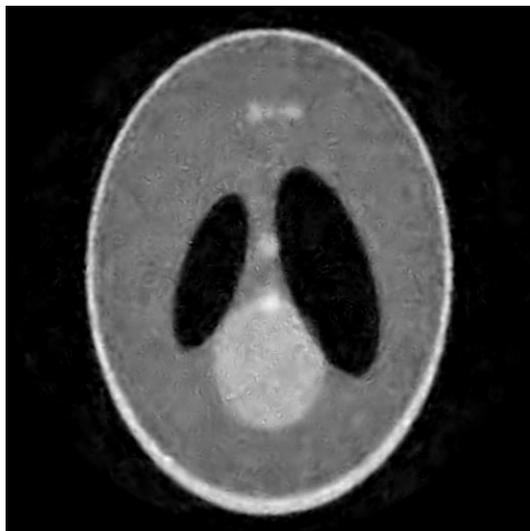


FIG. 2-C.  $f_C^{(4)}$ .

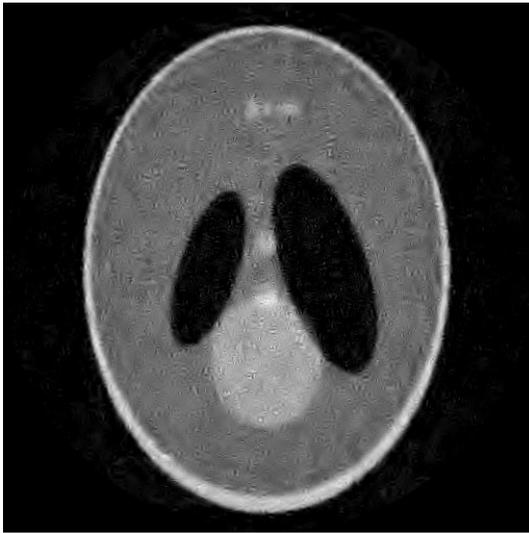


FIG. 3-A.  $f_T^{(1)}$ .

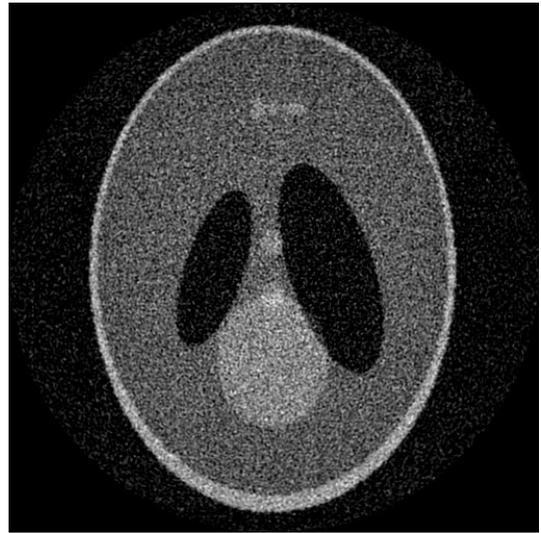


FIG. 4.  $f_F$  with  $N_s = 512$ .



FIG. 3-B.  $f_T^{(2)}$ .



FIG. 5.  $f_F$  with  $N_s = 192$ .

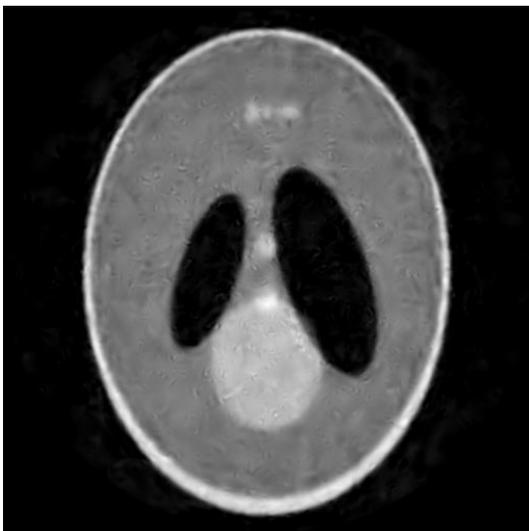


FIG. 3-C.  $f_T^{(4)}$ .



FIG. 6.  $f_V$ .

fourth row in Table 1 and Figure 4, the universal filter size, which works for all noise levels,  $N_s = 512$  does not remove the noise effectively. By reducing the filter size  $N_s$  to 192 we have a better reconstructed image in Figure 5 than that in Figure 4. In this case, however, we can not avoid certain degradation in reconstruction by losing information from the frequency number  $n$  with  $|n| > 192$ .

Notice that in (47), by choosing a radially symmetric  $w_H(\xi)$ , which is commonly used in the practice of the filtered backprojection, we have

$$w_H(\xi) |\xi| \widehat{\mathcal{R}^* Y}(\xi) = w_H(w) Y(\theta, \cdot)^\wedge(w),$$

where  $\xi = (w \cos \theta, w \sin \theta)$ . Therefore, unlike for  $f_C^{(R)}$  and  $f_T^{(R)}$ , the translation-rotation averaging techniques do not give an improved result in  $f_F$ , since the Fourier basis itself is translation-invariant and the same weight filter is applied to the projection data  $Y(\theta, u)$  for all  $\theta$ .

TABLE 3  
Optimal Filter Widths

SNR	10	15	20	25	30
$\sigma \times 10^3$	44.1626	24.8343	13.9654	7.8533	4.4162
$N_s$	112	144	192	240	320

In Table 3 we give an optimal filter size of  $f_F$  for each noise level. We obtained those numbers experimentally.

Figure 5 shows a  $512 \times 512$  reconstruction based on  $f_V$  for noise level SNR 20. The algorithm  $f_V$  has fewer artifacts, but it is outperformed overall by most of other algorithms. See Table 1. Since  $f_V$  employs the wavelet shrinkage method for the wavelet coefficients only in the two highest levels, it can be viewed a combined method of the filtered backprojection and wavelet shrinkage. By shrinking the wavelet coefficients in more levels and applying translation and rotation averaging techniques, one can improve the result of  $f_V$ . However, since  $f_V$  uses a larger shrinkage parameter, even with translation and rotation averaging, one can not avoid certain over-smoothing in reconstruction.

We applied  $f_F$  with  $N_s = 128$  and  $f_T^{(2)}$  to a real *positron emission tomography* (PET) data set, taken from a physical phantom, which contains the Radon projection data at 256 uniformly spaced angles, and at 256 uniformly spaced points for each angle. For these reconstruction, we do not have an original image. We believe that  $f_T^{(2)}$  removes as much noise as  $f_F$  does, while preserving more of the fine structure of the image.

Based on these experiments, we believe that minimizing our bound on the error (41) leads to near-optimal shrinkage parameters for tomographic reconstruction with wavelet shrinkage. Moreover, our technique for estimating the smoothness of images leads to accurate estimates of the true smoothness of images. We also can predict the performance of the wavelet tomographic reconstruction algorithm using only two smoothness parameters  $\beta$  and  $|f|_{B_{p,p}^\beta}$ . We also believe that translation-rotation averaging techniques

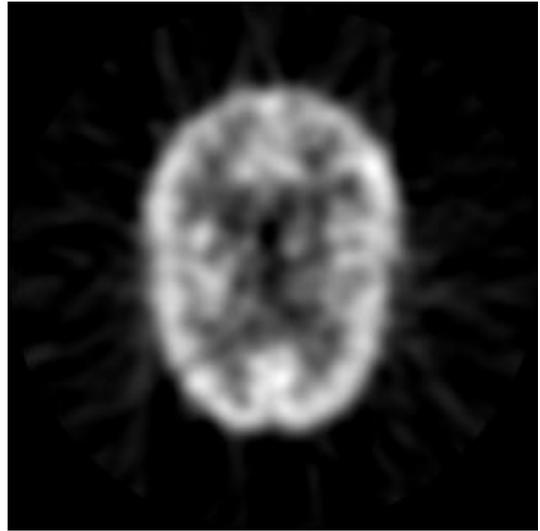


FIG. 7. PET image by  $f_F$  with  $N_s = 128$ .

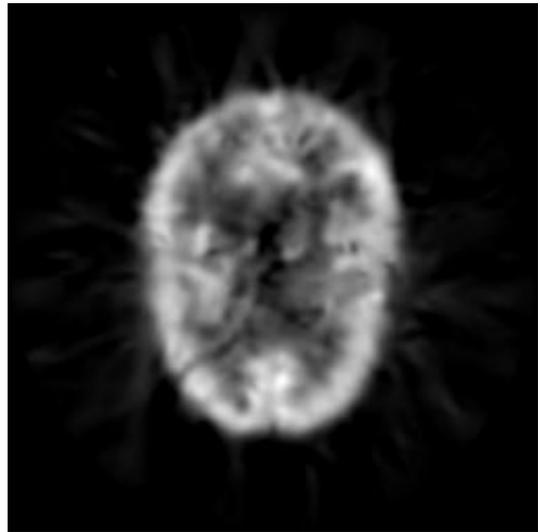


FIG. 8. PET image with  $f_T^{(2)}$ .

remove most of the artifacts of wavelet shrinkage methods, while removing a great amount of noise.

## REFERENCES

- [1] M. Bhatia, W. C. Karl, and A. S. Willsky, *Tomographic reconstruction and estimation based on multiscale natural-pixel bases*, IEEE Trans. Image Processing, 6 (1997), pp. 463–478.
- [2] ———, *A wavelet-based method for multiscale tomographic reconstruction*, IEEE Trans. on Medical Imaging, 15 (1996), pp. 92–101.
- [3] J. O. Berger, *Statistical decision theory and Bayesian analysis*, Springer-Verlag, 1985.
- [4] T. Cai, *Minimax wavelet estimations via block thresholding*, Technical Report #96-41, Purdue University.
- [5] A. Chambolle, R. A. DeVore, N.-Y. Lee, and B. J. Lucier, *Non-linear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage*, IEEE Trans. Image Processing, 7 (1998), pp. 319–335.
- [6] A. Cohen, I. Daubechies, and J.-C. Feauveau, *Biorthogonal bases of compactly supported wavelets*, Comm. Pure Appl. Math., 45 (1992), pp. 485–560.

- [7] A. Cohen, I. Daubechies, and P. Vial, *Wavelets on the interval and fast wavelet transforms*, Appl. Comput. Harmonic Analysis, 1 (1993), pp. 54–81.
- [8] A. Cohen, R. A. DeVore, and R. Hochmuth, *Restricted nonlinear approximation*, in preparation.
- [9] R. R. Coifman and D. L. Donoho, *Translation-invariant denoising*, in Wavelets and Statistics, A. Antoniadis and G. Oppenheim, eds., Springer-Verlag, New York, 1995, pp. 125–150.
- [10] I. Daubechies, *Orthonormal basis of compactly supported wavelets*, Comp. Pure Appl. Math., 46 (1988), pp. 909–996.
- [11] ———, *Ten Lectures on Wavelets*, CBMS-NSF Regional Conference series in Applied Mathematics 91, SIAM, Philadelphia, 1992.
- [12] I. Daubechies and J. C. Lagarias, *Two-scale difference equations I. Existence and global regularity of solutions*, SIAM J. Math. Anal., 22 (1991), pp. 1388–1410.
- [13] ———, *Two-scale difference equations II. Local regularity, infinite products of matrices and fractals*, SIAM J. Math. Anal., 23 (1992), pp. 1031–1079.
- [14] A. H. Delaney and Y. Bresler, *Multiresolution tomographic reconstruction using wavelets*, IEEE Trans. Image Processing, 4 (1995), pp. 799–813.
- [15] R. A. DeVore, B. Jawerth, and B. J. Lucier, *Image Compression through wavelet transform coding*, IEEE Trans. Information Theory, 38, 2 (1992), pp. 719–746; Special issue on Wavelet Transforms and Multiresolution Analysis.
- [16] ———, *Surface compression*, Computer Aided Geom. Design, 9 (1992), pp. 219–239.
- [17] R. A. DeVore, B. Jawerth, and V. Popov, *Compression of wavelet decompositions*, Amer. J. Math., 114 (1992), pp. 737–785.
- [18] R. A. DeVore and B. J. Lucier, *Fast wavelet techniques for near-optimal image processing*, in IEEE Military Communications Conference Record, San Diego, October 11-14, 1992, IEEE Press, Piscataway, NJ, 1992, pp. 1129–1135.
- [19] R. A. DeVore and V. Popov, *Interpolation of Besov spaces*, Trans. Amer. Math. Soc., 305 (1988), pp. 397–414.
- [20] D. C. Dobson, *Convergence of a reconstruction method for the inverse conductivity problem*, SIAM J. Appl. Math., 52 (1992), pp. 442–458.
- [21] D. L. Donoho, *Nonlinear solution of linear inverse problem by wavelet-vaguelette decomposition*, Appl. Comput. Harmon. Anal., 2 (1995), pp. 101–126.
- [22] D. L. Donoho and I. M. Johnstone, *Adapting to unknown smoothness via wavelet shrinkage*, J. Amer. Statist. Assoc., 90 (1995), pp. 1200–1224.
- [23] ———, *Neo-classical minimax problems, thresholding and adaptive function estimations*, Bernoulli, 2 (1996), pp. 39–62.
- [24] ———, *Ideal spatial adaptation by wavelet shrinkage*, Biometrika, 81 (1994), pp. 425–455.
- [25] D. L. Donoho, I. M. Johnstone, G. Kerkycharian, and D. Picard, *Wavelet shrinkage: Asymptopia ?*, J. Roy. Statist. Soc. Ser. B, 57 (1995), pp. 301–369.
- [26] M. Frazier and B. Jawerth, *A discrete transform and decompositions of distribution spaces*, J. of Functional Anal., 93 (1990), pp. 34–170.
- [27] C. Herley and M. Vetterli, *Biorthogonal bases of symmetric compactly supported wavelets*, in Wavelets, Fractals, and Fourier Transforms, Oxford Univ. Press, New York, 1993, pp. 991–1008.
- [28] A. C. Kak and M. Slaney, *Principles of computerized tomographic imaging*, IEEE Press, 1988.
- [29] D.-G. Kim, *Wavelet decompositions and function spaces on the unit cube*, Ph.D. Thesis, Purdue University, 1994.
- [30] E. D. Kolaczyk, *A wavelet shrinkage approach to tomographic image reconstruction*, J. Amer. Statist. Assoc., 91 (1996), pp. 1079–1090.
- [31] G. C. Kyriazis, *Wavelet decompositions and spaces of functions*, Ph.D. Thesis, University of South Carolina, 1992.
- [32] N.-Y. Lee, *Wavelet-vaguelette decompositions and homogeneous equations*, Ph.D. Thesis, Purdue University, 1997.
- [33] B. J. Lucier, M. Kallergi, W. Qian, R. A. DeVore, R. A. Clark, E. B. Saff, and L. P. Clarke, *Wavelet compression and segmentation of mammographic images*, J. of Digital Imaging, 7 (1994), pp. 27–38.
- [34] Y. Meyer, *Ondelettes et opérateurs I: Ondelettes*, Hermann, Paris, 1990; English transl. by D. H. Salinger, *Wavelets and operators*, Cambridge Univ. Press, Cambridge, 1992.
- [35] Y. Meyer, *Ondelettes et opérateurs II: Opérateurs de Calderón-Zygmund*, Hermann, Paris, 1990.
- [36] E. L. Miller, L. Nicolaides, and A. Mandelis, *Nonlinear inverse scattering methods for thermal-wave slice tomography: a wavelet domain approach*, J. Optical Soc. Amer. A, 15 (1998), pp. 1545–1556.
- [37] F. Natterer, *The mathematics of computerized tomography*, John Wiley and Sons, New York, 1986.
- [38] T. Olson, *Optimal time-frequency projections for localized tomography*, Annals of Biomedical Engin., 23 (1995), pp. 622–636.
- [39] T. Olson and J. DeStefano, *Wavelet localization of the Radon transform*, IEEE Trans. Signal Processing, 42 (1994), pp. 2055–2067.
- [40] F. RashidFarrokhi, K. J. R. Liu, C. A. Berenstein, and D. Walnut, *IEEE Trans. Image Processing*, 6 (1997), pp. 1412–1430.
- [41] B. Sahiner and A. E. Yagle, *Reconstructions from projections under time-frequency constraints*, IEEE Trans. Medical Imaging, 14 (1995), pp. 193–204.
- [42] J. L. C. Sanz, E. B. Hinkle, and A. K. Jain, *Radon and projection transform-based computer vision*, Springer-Verlag, 1987.
- [43] A. N. Tikhonov, *Solution of incorrectly formulated problems and regularization method*, Soviet Math. Dok., 4 (1963), pp. 1035.
- [44] H. Triebel, *Theory of function spaces II*, Monographs in Mathematics, vol. 78, Birkhauser Verlag, Basel, 1983.
- [45] S. Y. Zhao, *Wavelet filtering for filtered backprojection in computed tomography*, Appl. and Comp. Harmonic Anal., 6 (1999), pp. 346–373.
- [46] S. Y. Zhao, G. Wang, and J. Hsieh, *Wavelet filtering algorithm for fan-beam CT*, Electronics Letters, 34 (1998), pp. 2395–2396.
- [47] S. Y. Zhao, G. Welland, and G. Wang, *Wavelet sampling and localization schemes for the Radon transform in two dimensions*, SIAM J. Appl. Math., 57 (1997), pp. 1749–1762.
- [48] W. W. Zhu, Y. Wang, Y. N. Deng, Y. Q. Yao, and R. L. Barbour, *A wavelet-based multiresolution regularized least squares reconstruction approach for optical tomography*, IEEE Trans. Medical Imaging, 16 (1997), pp. 210–217.

**Nam-Yong Lee** was born in Seoul, Korea, in 1967. He received the B.S. degree in mathematics from Seoul National University in 1990, and the Ph.D. degree in mathematics from Purdue University in 1997.

Dr. Lee worked as a Postdoctoral Research Associate in the Department of Mathematics, University of South Carolina from September 1997 to May 1998. From September 1998 to August 1999, he was a Postdoctoral Research Associate in the Global Analysis Research Center, Korea. He is currently a Visiting Professor in the Department of Control and Instrumentation Engineering, Kangwon National University, Korea. His research interests include applied mathematics, image processing, tomography, and multimedia data security.

**Bradley J. Lucier** (M'91, SM'94) received a B.Sc. (Hon.) degree in mathematics from the University of Windsor, Ontario, in 1976 and S.M. and Ph.D. degrees in applied mathematics from the University of Chicago in 1978 and 1981, respectively. He has been at Purdue University since 1981, holding the position of Professor of mathematics and computer sciences since 1991. He has been a long-term visitor at the Mathematics Research Center (University of Wisconsin, Madison); the Institute for Mathematics and its Applications (University of Minnesota, Minneapolis); and CEREMADE (Université de Paris-Dauphine, Paris). Dr. Lucier's research is motivated by problems of computation, recently in image processing, where he applies wavelet techniques to problems in compression, noise removal, and tomography, and in partial differential equations.