Tomography, Approximate Reconstruction, and Continuous Wavelet Transforms

W. R. Madych

Department of Mathematics, University of Connecticut, Storrs, Connecticut 06269-3009

Communicated by Charles K. Chui

Received October 20, 1997; revised April 10, 1998

It has been recognized for some time now that certain high-frequency information concerning planar densities \( f \) in a neighborhood of a point can be recovered from data which consist of averages of \( f \) over lines that are relatively close to that point. The wavelet transform of \( f \) is a classical tool for analyzing local frequency content. In this article we introduce continuous wavelet transforms which are particularly well suited to producing high-resolution local reconstructions from local data of the type described above. We also show how such transforms can be realized numerically via simple modifications of well-established convolution backprojection-type algorithms.

As part of our development we review the concepts of “local tomography” and “pseudolocal tomography” introduced by several authors and indicate that, in effect, these notions basically involve the computation of a wavelet transform. The results in this paper are based on the observation that Radon’s classical inversion formula is a summability formula with an integrable convolution-type summability kernel.

1. INTRODUCTION

1.1. Background

The Radon transform of a sufficiently well-behaved scalar density or function \( f(x) \) defined on the plane, \( x \in \mathbb{R}^2 \), is the function \( Rf(\theta, t) \) which is defined on the cylinder \( 0 \leq \theta < 2\pi, -\infty < t < \infty \), and which is related to \( f \) via the formula

\[
Rf(\theta, t) = \int_{-\infty}^{\infty} f(tu + sv) ds,
\]

(1)

This research was partially supported by AFOSR Grant F49620-93-1-0565.
where $u_\theta = (\cos \theta, \sin \theta)$ and $v_\theta = (-\sin \theta, \cos \theta)$. By taking $t = \langle x, u_\theta \rangle$, where $\langle x, u_\theta \rangle$ denotes the usual scalar product of $x$ and $u_\theta$, one can see that $Rf(\theta, \langle x, u_\theta \rangle)$ is the integral of $f$ along the line through $x$ in the direction $v_\theta$.

Formula (1) is used to model the data acquisition schemes of various physical experiments and technical devices; for example, see [7, 16, 22, 32, 34, 41]. In many of these scenarios samples of $Rf$, namely discrete values such as $\{Rf(\theta_j, t_k) : j = 1, \ldots, m \text{ and } k = 1, \ldots, n\}$, are collected and used to recover or approximate $f$ or some feature of $f$.

Under appropriate conditions on $f$ its values, $f(x)$, can be recovered from values of its Radon transform via Radon’s celebrated classical formula

$$f(x) = \frac{-1}{\pi} \int_0^\infty \frac{dF_s(q)}{q},$$

where $F_s(q)$ is the average of all the integrals of $f$ over lines which are distance $q$ from $x$. For example, see [41, p. 245] and also [7, Appendix A; 12, 32, 37].

Discrete variants of (2) are useful in practical recovery schemes alluded to above. However, this formula contains features which make it difficult to apply directly in numerical and otherwise practical situations. Among these are the fact that (i) the integral in the formula is not proper and (ii) the inversion is not local, namely, the evaluation of $f(x)$ requires the integrals of $f$ over all lines, not only those through $x$ or close to $x$.

Over the years many publications have dealt with (i); for example, see [16, 32] for detailed expositions and extensive bibliographies. The celebrated algorithm of Shepp and Logan [40] played a significant role in this development. Relatively recently several publications have addressed (ii); see [9–11, 19, 42]. The basic idea in these works seems to be the use of integrals of $f$ over lines close to $x$ not to recover the value $f(x)$ but to obtain certain high-frequency content of $f$ near $x$.

The concept of the wavelet transform is a classical tool for analyzing the frequency content of functions [4, 13, 14, 23, 31, 35, 43–45, 50], which has currently become quite fashionable in the study of signals and images [1, 5, 6, 16, 18, 21, 31, 46, 48] including tomography [2, 3, 8, 17, 36, 38, 47, 49]. As is evident from a perusal of the literature, there are several related notions involving this term and frequency analysis. In this article we use the following definition: For a scalar-valued function $f$ defined on the plane $\mathbb{R}^2$, its continuous wavelet transform is a one parameter family $W_{\Psi}f(a, x)$, $a > 0$, of functions defined by the convolution

$$W_{\Psi}f(a, x) = \Psi_a * f(x),$$

where $\Psi_a(x) = a^{-2}\Psi(x/a)$ and $\Psi$ is an integrable function on $\mathbb{R}^2$ with $\int_{\mathbb{R}^2} \Psi(x)dx = 0$; more explicitly,

$$\Psi_a * f(x) = \frac{1}{a^2} \int_{\mathbb{R}^2} \Psi(y/a)f(x - y)dy.$$
The function $\Psi$ is sometimes referred to as the analyzing wavelet.\(^3\) If the function $\Psi$ in formula (3) does not have total integral zero then $\Psi \ast f(x)$ may still be well defined but is not considered to be a wavelet transform. Indeed, if $\Phi$ is an integrable function on $\mathbb{R}^2$ with total integral one then it is very well known that for sufficiently well-behaved $f$,

$$f(x) = \lim_{a \to 0} \Phi \ast f(x),$$

so that for fixed $a$ the function $\Phi \ast f(x)$ is a low-frequency approximation of $f(x)$ which improves with decreasing $a$. A limit such as (4) is sometimes referred to as a summability method with kernel $\Phi$; see [18, 44, 50].

Since any nice radial function enjoys a convenient representation as a uniform sum of ridge functions, in the case when $\Psi$ is radial the transformation (3) can be computed directly from the values of the Radon transform of $f$ via a relatively simple formula involving convolution and backprojection.\(^4\) The resulting formulas, unlike (2), usually involve proper integrals and, in view of (4), can provide a theoretical foundation for many numerical reconstruction algorithms. On the other hand, like (2), these formulas are generally not local. Indeed, low-frequency approximations such as those suggested by (4) cannot be computed in terms of local Radon transform data. To surmount this difficulty Smith and Keinert [42] suggested the use of combinations of certain high-pass and low-pass frequency filters. Other articles featuring high-pass frequency filters, including wavelets, followed [2, 8–11, 19, 36, 38], confirming the soundness of this idea.

Beginning with the work of Olsen and DeStefano [36] most of the articles proposing wavelets as high-pass frequency filters utilize various variants of the so-called discrete wavelet expansions introduced and popularized by Daubechies, Mallat, Meyer, and their collaborators [6, 31]. Berenstein and Walnut [2] were the first to employ a variant of continuous wavelet transforms in this regard.

One objective of this article is to introduce continuous wavelet transforms which are particularly well suited to producing high-resolution local reconstructions from local Radon transform data. This is accomplished by first showing that Radon’s inversion formula (2) and its cousins are really summability methods (4) and then using the corresponding kernels to produce analyzing wavelets with very desirable properties for numerically efficient local high-pass or, more accurately, mid-pass-frequency filtering of bivariate functions in terms of local Radon transform data.

A more detailed outline of the objectives and contents of this article can be found in the next subsection.

---

\(^3\) Note that the transform depends on the analyzing wavelet $\Psi$. A popular choice is the Laplacian of the Gaussian or the so-called “Mexican hat”

$$\Psi(x) = (|x|^2 - 2)e^{-|x|^2}$$

but, of course, many other choices are possible. This definition and some of its consequences are discussed in more detail in Subsection 2.3. The particular variant of the continuous wavelet transform used here employs a different power of the parameter $a$ than that which is currently in vogue.

\(^4\) See [29]. This is also reviewed in more detail in Subsection 2.1.
1.2. Overview

Because of the improper integral in (2), Radon’s inversion formula should be expressed as

\[ f(x) = -\frac{1}{\pi} \lim_{\varepsilon \to 0} \int_{\varepsilon}^{\infty} \frac{dF_{\varepsilon}(t)}{t} \]  

(5)

or, after integration by parts,

\[ f(x) = \frac{1}{\pi} \lim_{\varepsilon \to 0} \left\{ \frac{F_{\varepsilon}(\varepsilon)}{\varepsilon} - \int_{\varepsilon}^{\infty} \frac{F_{\varepsilon}(t)}{t^2} dt \right\}, \]  

(6)

where

\[ F_{\varepsilon}(t) = \frac{1}{2\pi} \int_{0}^{2\pi} f_{\varepsilon}(x, u_{\theta}) + t d\theta; \]  

(7)

see [37, formula III']. Radon showed the validity of this formula under the assumption that \( f \) is continuous and satisfies two other technical conditions. His development used what currently might be referred to as a Laplace transform-related fractional integral method.

We show that

\[ \int_{\varepsilon}^{\infty} \frac{dF_{\varepsilon}(t)}{t} = \Phi_{\varepsilon} * f(x), \]  

(8)

where \( \Phi_{\varepsilon} \) is a positive radial kernel with total integral one. Identity (8) means that Radon’s inversion formula is, in effect, a summability type of procedure of convolution type with kernel \( \Phi \). Thus it provides not only an alternate proof of Radon’s result but also a significant improvement of it.

Formulas like (8) can be used as the basis for obtaining localized reconstructions from Radon transform data. Namely, if \( \Phi \) is the same as in (8) then \( \Psi(x) = 2\Phi(2x) - \Phi(x) \) has mean value 0 and the corresponding wavelet transform can be computed via

\[ W_{\Psi}f(a, x) = -\frac{1}{\pi} \int_{a/2}^{\infty} \frac{dF_{\varepsilon}(t)}{t}. \]  

(9)

In other words, the value of the wavelet transform \( W_{\Psi}f(a, x) \) can be computed using only the integrals of \( f \) over lines which intersect the disc of radius \( a \) about \( x \). Furthermore, for any positive \( \varepsilon \),

\footnote{We use slightly different variants of the inversion formula in most of our development.}
\[
\sum_{j=-\infty}^{\infty} W_{2^j e}(x) = f(x),
\]

and, since for sufficiently large \( j \) the terms \( W_e(2^j e, x) \) contain mainly low-frequency information which can often be ignored, the sum in formula (10) can be truncated to obtain a local-type approximate inversion formula.

The inversion method suggested by (9) and (10) is potentially of some practical significance. For instance, we show how, in one particular realization, (9) can be approximately evaluated using parallel beam X-ray data and a slightly modified version of the celebrated Logan–Shepp algorithm [40].

The classical convolution–backprojection algorithm introduced by Logan and Shepp in [40] is usually viewed as a filter or regularization method. We show that it can also be viewed as a direct and natural discretization of a slightly modified variant of Radon’s original inversion formula. This is done as part of the development of the numerical variants of formulas like (9) and (10).

In our development we include the following:

- a brief review of some connections between the Radon transform, ridge functions, convolution–backprojection methods, and inversion (Subsection 2.1);
- a precise statement and proof of (8) and related formulas, including specific convergence results (Subsection 2.2);
- a brief review of the definition and the motivation of wavelet transforms, the convolution–backprojection method for computing such transforms in terms of Radon transform data, and several procedures for inverting wavelet transforms (Subsection 2.3);
- a recollection of the notions of “local tomography” and “pseudolocal tomography” as found in [9–11, 42] and [19], respectively, and an indication of a connections between these notions and wavelet transforms (Subsection 2.4); an indication that some of the theoretical results found in [19] can be significantly strengthened (Subsection 3.4);
- a derivation of the localized-type approximate reconstruction formula from Radon transform data like the one implied by the pair of Eqs. (9), (10) (Subsection 2.5);
- discrete analogues of formulas like (9) and (10) which lead to various reconstruction algorithms (Subsection 2.6);
- numerical examples illustrating potential applications of these algorithms (Section 4).

Most of the development can be found in Section 2, where to maintain readability certain technical details and other pertinent, but not essential, remarks are kept to a minimum. The main development is contained in Subsections 2.1, 2.2, 2.3, 2.5, and 2.6. Subsection 2.4 contains material on “local tomography” and “pseudolocal tomography” and is not critical to this development; it is included to bring attention to the natural connection between these notions and wavelet transforms.

Subsection 3.n is devoted to the details and remarks omitted in Subsection 2.n.
1.3. Notation

We use standard terminology, notation, and conventions. Here we simply remind the reader of the following:

- The convolution $f * g$ of two scalar-valued functions $f$ and $g$ on $\mathbb{R}^2$ (or $\mathbb{R}$) is defined by

$$f * g(x) = \int f(y) g(x - y) dy,$$

where the integral is taken over all of $\mathbb{R}^2$ (or $\mathbb{R}$) whenever the integral is well defined and distributionally otherwise.

- The Fourier transform $\hat{f}$ of a function $f$ on $\mathbb{R}^2$ (or $\mathbb{R}$) is defined by

$$\hat{f}(\xi) = \int e^{-i\langle \xi, x \rangle} f(x) dx,$$

where the integral is taken over all of $\mathbb{R}^2$ (or $\mathbb{R}$) whenever the integral is well defined and distributionally otherwise. $\langle \xi, x \rangle$ denotes the usual scalar product of $\xi$ and $x$.

- Whether the convolution or Fourier transform is to be interpreted in the bivariate or univariate sense should be clear from the context.

- Generic constants, whose meaning should be clear from the context, are denoted by $c$.

In what follows $f$ always denotes an integrable scalar-valued function on the plane $\mathbb{R}^2$, that is, $\int_{\mathbb{R}^2} |f(x)| dx$ is finite. Other restrictions on $f$ will be given as needed.

2. DEVELOPMENT

2.1. Ridge Functions and the Radon Transform

Recall that the Radon transform $Rf(\theta, t)$, $0 \leq \theta < 2\pi$, $-\infty < t < \infty$, of an integrable function $f$ on $\mathbb{R}^2$ may be defined by

$$Rf(\theta, t) = \int_{-\infty}^{\infty} f(tu_\theta + sv_\theta) ds,$$

where $u_\theta = (\cos \theta, \sin \theta)$ and $v_\theta = (-\sin \theta, \cos \theta)$. Since it is often convenient to view $Rf(\theta, t)$ as a family of functions of $t$ parametrized by $\theta$ we use the abbreviated notation $f_\theta(t) = Rf(\theta, t)$. 
A scalar-valued function $F$ on the plane $\mathbb{R}^2$ is said to be a ridge function if it can be expressed as $F(x) = \phi(\langle x, u_\theta \rangle)$ for some unit vector $u_\theta$ and some univariate function $\phi$. The transformation relating the univariate function $\phi(t)$ to the bivariate ridge function $\phi(\langle x, u_\theta \rangle)$ is often referred to as backprojection.

If $\Phi$ is such a ridge function then convolving it with another function $f$ on $\mathbb{R}^2$ and expressing the integral in the $\{u_\theta, v_\theta\}$ coordinate system results in

$$
\int_{\mathbb{R}^2} \Phi(x - y) f(y) dy = \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} \phi(\langle x, u_\theta \rangle - t) f(tu_\theta + sv_\theta) ds \right) dt
$$

or, in more compact and suggestive notation,

$$
\Phi \ast f(x) = \phi \ast f(\langle x, u_\theta \rangle).
$$

In other words, the bivariate convolution of $\Phi$ and $f$ evaluated at $x$ is equal to the univariate convolution of $\phi$ and $f_{\theta}$ evaluated at $\langle x, u_\theta \rangle$. We are assuming, of course, that all the functions are sufficiently well behaved so that the integrals make sense.

Despite its elementary nature formula (11) is very useful. For instance, it should be quite easy to see that if a convolution kernel $\Phi$ can be expressed as a sum of ridge functions then $\Phi \ast f$ can be readily computed in terms of $Rf$. The definitions and formulas below are simply more precise versions of this observation.

A locally integrable function $\Phi$ on $\mathbb{R}^2$ is a uniform sum of ridge functions if there is an even locally integrable univariate function $\phi$ such that

$$
\Phi(x) = \frac{1}{2\pi} \int_0^{2\pi} \phi(\langle x, u_\theta \rangle) d\theta
$$

for all $x$ in $\mathbb{R}^2$. If (12) holds in the distributional sense we simply say the $\Phi$ is a uniform sum of ridge functions in the distributional sense.

It should be clear that if both $\Phi$ and $f$ are suitably well behaved and $\Phi$ is a uniform sum of ridge functions then

$$
\Phi \ast f(x) = \frac{1}{2\pi} \int_0^{2\pi} \phi \ast f(\langle x, u_\theta \rangle) d\theta.
$$

Note that if $f$ has compact support and is sufficiently regular then (13) is valid even if $\Phi$ is a uniform sum of ridge functions only in the distributional sense. This will be the case in some of the applications below.

Formula (13) is very convenient for dealing with Radon transform data; it allows us to...
compute the convolution of $\Phi$ and $f$ in terms of the Radon transform of $f$. For example, if $\Phi$ can be taken to be a sufficiently narrow approximation of the identity then (13) may be viewed as an approximate inversion formula for the Radon transform. Therefore it is of some interest to understand the nature of such functions $\Phi$ which satisfy (12). Here we briefly review some of the material in [29].

Observe that, as a consequence of the definition, if $\Phi$ is a uniform sum of ridge functions then it must be a radial function; this means that there is a univariate function $\Phi_0$ such that $\Phi(x) = \Phi_0(|x|)$. The converse is also true; namely, any reasonable radial function can be uniquely expressed as a uniform sum of ridge functions.

Recall that $\Phi_a(x) = a^{-2}\Phi(x/a)$ so that

$$\Phi_a * f(x) = \frac{1}{2\pi a} \int_0^{2\pi} \phi_a * f_\theta(x, u_\theta)d\theta, \quad (14)$$

where $\phi_a(t) = a^{-1}\phi(t/a)$. Thus (14) and (4) together with the fact that any reasonable radial function can be expressed as a uniform sum of ridge functions imply a multitude of summability-type inversion formulas.

If $\Phi$ and $\phi$ satisfy (12) there are various alternate formulas expressing $\Phi$ in terms of $\phi$ and vice versa; see [29]. A particularly convenient one relates their Fourier transforms,

$$2\hat{\phi}(|\xi|) = |\xi|\hat{\Phi}(\xi), \quad (15)$$

whenever $\hat{\Phi}$ is sufficiently regular.

For example, if $\Phi$ is the summability kernel

$$\Phi(x) = \frac{1}{2\pi(1 + |x|^2)^{3/2}}$$

whose Fourier transform is

$$\hat{\Phi}(\xi) = e^{-|\xi|},$$

then relation (15) can be used to obtain the corresponding ridge function representation, namely, the function $\phi$ in (12), which is

$$\phi(t) = \frac{1 - t^2}{2\pi(1 + t^2)^{3/2}}.$$
2.2. Radon's Inversion Formula and Variations

Consider the natural approximation \( g_{\varepsilon f} \) of Radon's inversion formula (6) parametrized by \( \varepsilon \) and defined by

\[
g_{\varepsilon f}(x) = \frac{1}{\pi} \left\{ \frac{F_{\varepsilon}(t)}{\varepsilon} - \int_{\varepsilon}^{\infty} \frac{F_{\varepsilon}(t)}{t^2} \, dt \right\},
\]

where

\[
F_{\varepsilon}(t) = \frac{1}{2\pi} \int_{0}^{2\pi} f_\varepsilon(x, u_\theta + t) d\theta
\]
is the average of all the integrals of \( f \) over lines which are distance \( t \) from \( x \). The fact that the transformation \( f \to g_{\varepsilon f} \) is translation invariant implies that \( g_{\varepsilon f} \) is the convolution of \( f \) with some distribution. Appropriate calculations show that

\[
g_{\varepsilon f}(x) = \Phi_{\varepsilon} * f(x),
\]

where

\[
\Phi_{\varepsilon}(x) = \varepsilon^{-2} \Phi(x/\varepsilon),
\]

\[
\Phi(x) = \frac{1}{\pi |x|^2 \sqrt{|x|^2 - 1}} \chi(|x|),
\]

and

\[
\chi(t) = \begin{cases} 
0 & \text{if } |t| \leq 1 \\
1 & \text{if } |t| > 1.
\end{cases}
\]

Note that if \( \Phi \) is the kernel in formula (17) then \( \Phi \) enjoys the following properties:

(a) \( \Phi \) is a non-negative radial function on \( \mathbb{R}^2 \).
(b) \( \int_{\mathbb{R}^2} \Phi(x) dx = 1 \).
(c) \( \int_{\mathbb{R}^2} |x|^\alpha \Phi(x) dx \) is finite for \( 0 < \alpha < 1 \).
(d) \( \Phi(x) \leq c|x|^{-3} \) for \( |x| > 2 \).

Thus \( g_{\varepsilon f} \) is simply the convolution of \( f \) with a radial summability kernel, or approximation of the identity, of “thickness” \( \varepsilon \). Indeed, in view of (16), formula (17) routinely implies (6) whenever \( f \) is continuous. For other standard consequences of identity (17) see the Theorem below.
Use of the representation\(^6\)

\[
f(x) = -\frac{1}{8\pi^2} \int_{-\infty}^{\infty} \int_{0}^{2\pi} \frac{f_\delta(x, u_\delta) + t - 2f_\delta(x, u_\delta)}{t^2} \, dt\,d\theta \tag{18}
\]

together with the approximation suggested by (16) results in a similar conclusion with a different summability kernel. To wit, consider the approximation \(G_\epsilon f\) of (18) defined by

\[
G_\epsilon f(x) = -\frac{1}{8\pi^2} \int_{|t| > \epsilon} \int_{0}^{2\pi} \frac{f_\delta(x, u_\delta) + t - 2f_\delta(x, u_\delta)}{t^2} \, dt\,d\theta; \tag{19}
\]

then the analogous manipulations which gave (17) from (16) result in

\[
G_\epsilon f(x) = K_\epsilon * f(x), \tag{20}
\]

where

\[
K_\epsilon(x) = \epsilon^{-2} K(x/\epsilon),
\]

\[
K(x) = \frac{1}{\pi} \left\{ \frac{1}{|x|} - \sqrt{|x|^2 - 1} \chi(|x|) \right\}
\]
and \(\chi\) is the indicator function of \(\{t : |t| > 1\}\) as above. The kernel \(K\) may be reexpressed as

\[
K(x) = \frac{1}{\pi} \left\{ \frac{1}{|x|} \left(1 - \chi(|x|)\right) + \frac{1}{|x|^2 \sqrt{|x|^2 - 1}} \chi(|x|) \right\}
\]

to see that it is an integrable radial function which is decreasing as a function of \(|x|\) and \(K(x) = O(|x|^{-3})\) as \(|x|\) tends to \(\infty\).

Note that \(K\) enjoys properties (a)-(d) enjoyed by the kernel \(\Phi\) in formula (17) above. Thus \(G_\epsilon f\), like \(g_\epsilon f\), is the convolution of \(f\) with a radial summability kernel of thickness \(\epsilon\). Because \(K\) is dominated by an integrable radially decreasing function, namely itself, it is significantly better than \(\Psi\), which does not enjoy this property. Relationship (20), together with the properties of \(K\) routinely implies various results concerning the convergence of \(G_\epsilon f(x)\) to \(f(x)\) as \(\epsilon\) goes to 0. For example, recalling that we always assume that \(f\) is integrable on \(\mathbb{R}^2\), we have the following:

**Theorem.** (i) \(\lim_{\epsilon \to 0} G_\epsilon f(x) = f(x)\) almost everywhere.

(ii) \(\lim_{\epsilon \to 0} \|f - G_\epsilon f\|_1 = 0\).

(iii) If \(f\) is in \(L^p\) for some \(p, 1 < p < \infty\), then \(\lim_{\epsilon \to 0} \|f - G_\epsilon f\|_p = 0\).

\(^6\) This inversion formula is derived in [29, p. 196]. For a connection with (6) see Subsection 3.2.
(iv) If \( f \) is in \( L^\infty \) and continuous at \( x \) then \( \lim_{\epsilon \to 0} G_{\epsilon}f(x) = f(x) \). Furthermore, if \( f \) is uniformly continuous then this holds uniformly in \( x \).

(v) If \( f \) is \( L^\infty \) and Hölder continuous\(^7\) at \( x \) of order \( \alpha \) for some positive \( \alpha \) then

\[
|f(x) - G_{\epsilon}f(x)| \leq c \begin{cases} 
\epsilon^\alpha & \text{if } 0 < \alpha < 1, \\
\epsilon \{1 + |\log \epsilon|\} & \text{if } \alpha = 1, \\
\epsilon & \text{if } \alpha > 1,
\end{cases}
\]

where \( c \) is independent of \( \epsilon \). Furthermore, if \( f \) is uniformly Hölder continuous on \( \mathbb{R}^2 \) then this estimate is valid for all \( x \) with a constant \( c \) independent of \( \epsilon \) and \( x \).

(vi) Suppose \( \Gamma \) is an analytic arc, \( x \in \Gamma \), and \( B \) is an open neighborhood of \( x \) such that the complement of \( \Gamma \cap B \) in \( B \) has two components \( B_1 \) and \( B_2 \). Furthermore, suppose that the restrictions of \( f \) to \( B_1 \) and \( B_2 \) have extensions to the closures of \( B_1 \) and \( B_2 \), respectively, which are uniformly Hölder continuous of order \( \alpha \) for some positive \( \alpha \). Let

\[
f(x-) = \lim_{y \in B_1, y \to x} f(y) \quad \text{and} \quad f(x+) = \lim_{y \in B_2, y \to x} f(y),
\]

then (21) holds with the left-hand side of the inequality replaced with

\[
\left| \frac{1}{2} \{f(x-) + f(x+)\} - G_{\epsilon}f(x) \right|.
\]

(vii) If \( f \) is in \( L^p \) for some \( p \), \( 2 < p \leq \infty \), then \( G_{\epsilon}f \) is Hölder continuous of order \( 1 - \frac{2}{p} \). If \( f \) is Hölder continuous of order \( \alpha \) then \( G_{\epsilon}f \) is Hölder continuous of order \( \alpha + 1 \).

Items (ii)–(vii) remain valid if \( G_{\epsilon}f \) is replaced with \( g_{\epsilon}f \). See Subsection 3.2 for more details.

Note that item (iv) concerns the behavior of \( G_{\epsilon}f(x) \) when \( x \) is on an analytic "edge" of \( f \) and is basically a consequence of the \( L^1 \) smoothness, positivity, and symmetry properties of the kernel \( K \). Similar results are valid for somewhat more general edges and various types of "corners."

2.3. Wavelet Transforms and Tomography

Recall that the wavelet transform of \( f \) with the "wavelet" \( \Psi \) is defined by

\[
W_\Psi f(a, x) = \Psi_{a} \ast f(x),
\]

\(^7\)The function \( f \) is said to be Hölder continuous of order \( \alpha \) at \( x \) if the \( k \)th order difference on \( f \) in \( y \) at \( x \) is dominated by a constant times \(|y|^k\), where \( k \) is the least integer greater than \( \alpha \). Thus, if \( \alpha \) is in the range \( 0 < \alpha < 1 \), this means that

\[
|f(x + y) - f(x)| \leq C|y|^\alpha,
\]

where \( C \) is a constant independent of \( y \); if \( \alpha \) is in the range \( 1 \leq \alpha < 2 \), this means that

\[
|f(x + y) - 2f(x) + f(x - y)| \leq C|y|^\alpha,
\]

etc. The function \( f \) is said to be uniformly Hölder continuous of order \( \alpha \) if it is Hölder continuous of order \( \alpha \) at every point \( x \) and the constant \( C \) is also independent of \( x \). Note that if \( f \) is differentiable then it is Hölder continuous of order one, if \( f \) is twice differentiable then it is Hölder continuous of order two, etc.
where \( a > 0 \), \( \Psi_a(x) = a^{-2} \Psi(x/a) \), and \( \Psi \) is an integrable function or measure on \( \mathbb{R}^2 \) with mean value 0 or, in other words, \( \hat{\Psi}(0) = 0 \). To obtain a hint of the motivation behind this transform note that by virtue of Plancherel’s formula

\[
W_\Psi f(a, x) = (2\pi)^{-2} \int_{\mathbb{R}^2} e^{i(a\xi \cdot x)} \hat{\Psi}(a\xi) \hat{f}(\xi) d\xi.
\]

So if the support of \( \hat{\Psi} \) is the annulus \( 0 < b_0 \leq |\xi| \leq b_1 < \infty \), which is “roughly” the case for some \( b_0 \) and \( b_1 \) whenever \( \Psi \) is a smooth and integrable radial function with mean value zero, then \( W_\Psi f(a, x) \) may be viewed as the frequency content of \( f \) in the band \( b_0/a \leq |\xi| \leq b_1/a \) at \( x \).

In theoretical applications, see, for example, [4, 23, 35], the specific \( \Psi \) is not really important, only certain of its properties play a major role. However, in numerical work it is nice to have explicit expressions for both \( \Psi \) and its Fourier transform. In addition to the Mexican hat wavelet mentioned earlier, typical examples are the following:

(i) Derivatives of the Poisson kernel

\[
P_a(x) = \frac{a}{2\pi(a^2 + |x|^2)^{3/2}}
\]

whose Fourier transform is

\[
\hat{P}_a(\xi) = e^{-|\xi|}.\]

For instance,

\[
\Psi_a(x) = -a \frac{\partial P_a}{\partial a}(x) = \frac{2a^3 - a|x|^2}{2\pi(a^2 + |x|^2)^{5/2}},
\]

whose Fourier transform is

\[
\hat{\Psi}_a(\xi) = |a\xi| e^{-|\xi|}.
\]

(ii) Appropriate differences of known summability kernels or measures. For instance,

\[
\Psi_a(x) = \delta(x) - P_a(x)
\]

or

\[
\Psi_a(x) = 2P_a(2x) - P_a(x),
\]

where \( P_a(x) \) is the Poisson kernel and \( \delta(x) = a^{-2}\delta(x/a) \) is the bivariate normalized Dirac delta “function.” In this case
\[ \hat{\Psi}_a(\xi) = 1 - e^{i|\xi|} \]

or

\[ \hat{\Psi}_a(\xi) = e^{-\frac{|\xi|^2}{2}} - e^{-i|\xi|}, \]

respectively.

Depending on the nature of \( \Psi \), there are various formulas for recovering \( f \) from its wavelet transform. One of the simplest seems to be the following: Suppose \( \Psi \) is a sufficiently well-behaved radial function such that

\[ c = \frac{1}{2\pi} \int_{\mathbb{R}^2} \Psi(\xi) \frac{d\xi}{|\xi|^2} \]

is finite and not zero; then using Plancherel’s identity it is not difficult to see that

\[ f(x) = \frac{1}{c} \int_{0}^{\infty} \mathcal{W}_\psi f(a,x) \frac{da}{a}. \quad (22) \]

Another formula which is very useful arises when \( \Phi \) is of the form

\[ \Psi(x) = 2\Phi(2x) - \Phi(x), \quad (23) \]

where \( \Phi \) is an integrable function whose total integral is one. In this case for any integers \( M < N \) and any positive \( \epsilon \) we may write

\[ \sum_{j=M+1}^{N} \Phi_{2^j} * f(x) = \Phi_{2^M} * f(x) - \Phi_{2^N} * f(x). \]

Since

\[ \lim_{M \to -\infty} \Phi_{2^j} * f(x) = f(x) \quad \text{and} \quad \lim_{N \to \infty} \Phi_{2^j} * f(x) = 0 \]

it follows that

\[ f(x) = \sum_{j=-\infty}^{\infty} \Phi_{2^j} * f(x). \quad (24) \]
Using the $W$ notation this may be expressed as the inversion formula

$$f(x) = \sum_{j=\infty}^{\infty} W_{\Psi} f(2^j e, x),$$

(25)

which is valid for any positive $\epsilon$ whenever $\Psi$ is of the form (23).

Recall that if $\Psi$ is a sufficiently well-behaved radial function then it can be represented as a uniform sum of ridge functions, that is,

$$\Psi(x) = \frac{1}{2\pi} \int_{0}^{2\pi} \psi(\langle x, u \rangle) d\theta$$

(26)

for some even univariate function $\psi(t)$. As noted in Subsection 2.1 this leads to

$$\Psi_a * f(x) = \frac{1}{2\pi a} \int_{0}^{2\pi} \psi_a * f(\langle x, u \rangle) d\theta,$$

(27)

where $\psi_a(t) = a^{-1} \psi(t/a)$. In short, the wavelet transform of $f$ with wavelet $\Psi$ can be computed terms of the Radon transform of $f$.

The function $\psi$ in representation (26) can be computed in terms of $\Psi$ by any one of the formulas alluded to in Subsection 2.1. In particular, if $\hat{\Psi}$ is sufficiently regular then the relationship (15) implies that the corresponding $\psi$ is integrable and has mean value zero, and $\psi_a * f(\theta)$ is a (univariate) wavelet transform of $f(\theta)$. Thus in this case (27) implies that the (bivariate) wavelet transform of $f$ is a sum of appropriately backprojected (univariate) wavelet transforms of the $f(\theta)$’s; in other words,

$$W_{\Psi} f(a, x) = \frac{1}{2\pi a} \int_{0}^{2\pi} W_{\Psi} f(\theta, \langle x, u \rangle) d\theta.$$ 

For example, the bivariate wavelet

$$\Psi_a(x) = \frac{2a^3 - a|x|^2}{2\pi(a^2 + |x|^2)^{3/2}}$$

mentioned earlier is related via (26) to the univariate wavelet

$$\psi_a(t) = \frac{a^5 - 3a^3 t^2}{\pi(a^2 + t^2)^{3/2}}$$
with
\[ \hat{\psi}_a(\tau) = \frac{a^2 \tau e^{-|\tau|}}{2}. \]

Other examples can be found in Subsections 2.5 and 3.5.

It is quite transparent and has been recognized for some time that wavelet inversion formulas when combined with representation (26) can result in inversion formulas for the Radon transform. However generally, if the analyzing wavelet is not carefully chosen with a view to some particular outcome, the resulting formula is simply a complicated expression for the usual type summability inversion method.

Before continuing with this development in Subsection 2.5 we review some allied notions in Subsection 2.4.

2.4. Local or Pseudolocal Tomography and Wavelet Transforms

2.4.1. Local Tomography: Low- and High-Bandpass Filters

Note that in order to determine \( f(x) \) via (2) knowledge of \( f_\theta(t) \) for all \( t \) and \( \theta \) is required. On the other hand, applying the formal adjoint \( R^\# \) of \( R \) to \( f_\theta(t) \) results in
\[
R^\# Rf(x) = \int_0^{2\pi} f_\theta(x, u_\theta) d\theta. \tag{29}
\]

Observe that the above formula for \( R^\# Rf(x) \) in terms of \( f_\theta(t) \) is very local in the sense that only integrals of \( f \) over lines through \( x \) are used to compute \( R^\# Rf(x) \). The relationship between \( R^\# Rf \) and \( f \) may be more transparent from
\[
R^\# Rf(x) = 4\pi J \ast f(x),
\]

where
\[
J(x) = \frac{1}{2\pi |x|}.
\]

Since the Fourier transform of \( J \ast f \) is
\[
\hat{J \ast f}(\xi) = |\xi|^{-1} \hat{f}(\xi),
\]

\( R^\# Rf \) may be regarded as a low-bandpass-filtered version of \( f \).

If \( f \) is sufficiently well behaved and the Laplacian \( \Delta \) is applied to both sides of (29) then interchanging the order of integration and differentiation on the right-hand side results in
\[
\Delta R^\theta f(x) = \int_0^{2\pi} f''_\theta(x, u_\theta) d\theta,
\]
where \(f''_\theta(t)\) is the second derivative of \(f_\theta(t)\) with respect to \(t\). Note that formula (30) for \(\Delta R^\theta Rf(x)\) in terms of \(f_\theta(t)\) is only slightly less local than (29) in the sense that only integrals of \(f\) over all lines which pass through an arbitrarily small neighborhood of \(x\) are needed to compute \(\Delta R^\theta Rf(x)\).

Using \(\Lambda\) to denote the transformation defined by
\[
\Lambda f(x) = -\Delta J \ast f(x)
\]
we see that it may also be expressed as
\[
\Lambda f(x) = -\frac{1}{4\pi} \Delta R^\theta Rf(x)
\]
and computed via (30). Note that mapping \(f \rightarrow J \ast f\) is the inverse of \(\Lambda\) since \(J \ast \Lambda f = f\); in other words,
\[
\Lambda^{-1} f(x) = J \ast f(x).
\]
Since the Fourier transform of \(\Lambda f\) is
\[
\widehat{\Lambda f}(\xi) = |\xi| \hat{f}(\xi),
\]
\(\Lambda f\) may be regarded as a high-bandpass-filtered version of \(f\). Thus one may obtain both low- and high-bandpass versions of \(f\), \(\Lambda^{-1} f = (1/4\pi) R^\theta Rf\) and \(\Lambda f = (-1/4\pi) \Delta R^\theta Rf\), from local Radon transform data via formulas (29) and (30). Both versions can be used to obtain information on \(f\) from its Radon transform data. For example, in [9–11, 42] the authors suggest the reconstruction of linear combinations of \(\Lambda f\) and \(\Lambda^{-1} f\) rather than the density function \(f\) itself; they refer to this as “local tomography.”

2.4.2. Pseudolocal Tomography

For sufficiently well-behaved \(f\) the inversion formula (2) can be reexpressed as
\[
f(x) = \frac{1}{4\pi^2} \int_0^{2\pi} \int_{-\infty}^{\infty} \frac{f'_\theta(x, u_\theta) - t}{t} dt d\theta,
\]
where \(f'_\theta(t)\) is the derivative of \(f_\theta(t)\) with respect to \(t\). In view of this expression the authors of [19] consider the decomposition
\[
f(x) = h_\varepsilon f(x) + \{f(x) - h_\varepsilon f(x)\},
\]
where \(h_\varepsilon\) is the characteristic function of the interval \([-\varepsilon, \varepsilon]\).
where

$$h_\epsilon f(x) = \frac{1}{4\pi} \int_{-\epsilon}^{\epsilon} \int_0^{2\pi} \frac{f_\epsilon'(x, u_\theta) - t}{t} \, dt \, d\theta,$$  \hspace{1cm} (33)$$

and suggest the reconstruction of the quantity $h_\epsilon f$ rather than the density function $f$ itself to obtain high-frequency information concerning $f$. They refer to this as “pseudolocal tomography.”

Note that (32) may be expressed as

$$f(x) = h_\epsilon f(x) + g_\epsilon f(x),$$

where

$$h_\epsilon f(x) = f(x) - g_\epsilon f(x) \hspace{1cm} (34)$$

and $g_\epsilon f$ is defined by (16). The last relation implies that the Fourier transform of $h_\epsilon f$ may be expressed as

$$\hat{h}_\epsilon f(\xi) = \{1 - \hat{\Phi}(\epsilon \xi)\} \hat{f}(\xi),$$

where, as indicated in Subsection 2.2, $\hat{\Phi}(\epsilon \xi)$ is a low-frequency bandpass filter with $\hat{\Phi}(0) = 1$. Thus $h_\epsilon f$ may be regarded as the result of processing $f$ through a high-frequency bandpass filter.

A decomposition analogous to (32) can be made using representation (18). Namely,

$$f(x) = H_\epsilon f(x) + G_\epsilon f(x),$$

where

$$H_\epsilon f(x) = f(x) - G_\epsilon f(x)$$

and $G_\epsilon f$ is defined by (19). In this case, for sufficiently well-behaved $f$, the formula for $H_\epsilon f(x)$ may be expressed as

$$H_\epsilon f(x) = -\frac{1}{8\pi^2} \int_{-\epsilon}^{\epsilon} \int_0^{2\pi} \frac{f_\epsilon'(x, u_\theta) + t - 2f_\epsilon(x, u_\theta) + f_\epsilon(x, u_\theta) - t}{t^2} \, dt \, d\theta. \hspace{1cm} (35)$$

It should be clear that $H_\epsilon f$ may also be regarded as the result of processing $f$ through a high-frequency bandpass filter.

It is important to note that, by virtue of (33) and (35), both $h_\epsilon f(x)$ and $H_\epsilon f(x)$ are, in
principle, computable\(^8\) in terms of integrals of \(f\) over lines which intersect the disk of radius \(\epsilon\) centered at \(x\). For this reason they may be regarded as being computable from local Radon transform data of \(f\).

Thus “local tomography” whose objective is to compute \(\Lambda f\) and “pseudolocal tomography” whose objective is to compute \(h_\epsilon f\) appear to take different routes to roughly the same goal. However, the fact that both \(\Lambda f\) and \(h_\epsilon f\) are high-frequency band-filtered versions of \(f\) which can be computed in terms of local Radon transform data is not the only similarity between these functions. Observe that the integrand in (33) may be replaced with

\[
\frac{f'(\langle x, u_\theta \rangle - t) - f'_\theta(\langle x, u_\theta \rangle)}{t},
\]

which in turn is well approximated by \(-f''_\theta(\langle x, u_\theta \rangle)\) if \(\epsilon\) is sufficiently small. Thus for sufficiently small \(\epsilon\) we may write

\[
h_\epsilon f(x) = -\frac{2\epsilon}{4\pi} \int_0^{2\pi} f''_\theta(\langle x, u_\theta \rangle) d\theta = \frac{2\epsilon}{\pi} \Lambda f(x)
\]

or, more precisely, for sufficiently well-behaved \(f\),

\[
h_\epsilon f(x) = \frac{2\epsilon}{\pi} \Lambda f(x) + o(\epsilon) \quad \text{as} \quad \epsilon \to 0.
\]

Further connections are discussed in Subsection 3.4.

2.4.3. A Connection with the Wavelet Transform

According to the authors of [9, 10, 42], in practical applications of local tomography they do not compute \(\Lambda f(x)\), but rather attempt to reconstruct \(\Lambda(\Phi_{\epsilon} * f)(x)\) for some approximation of the identity \(\Phi_{\epsilon}(x) = \epsilon^{-2}\Phi(x/\epsilon), \epsilon > 0\), where \(\Phi\) is an integrable function with total integral one. Now, for sufficiently well-behaved \(\Phi\), \(\Lambda\Phi\) is an integrable function with mean value zero, \(\Lambda(\Phi_{\epsilon}) = \epsilon^{-1}(\Lambda\Phi)_{\epsilon}\), and we may write

\[
\Lambda(\Phi_{\epsilon} * f)(x) = \epsilon^{-1}(\Lambda\Phi)_{\epsilon} * f(x).
\]

This may be reexpressed as

\[
\Lambda(\Phi_{\epsilon} * f)(x) = \epsilon^{-1}W_\psi f(\epsilon, x),
\]

\(^8\) These formulas are valid if \(f\) is sufficiently smooth. Otherwise one should use “regularized” analogues; see Subsection 2.4.3.
where

\[ \Psi(x) = \Lambda \Phi(x). \] (39)

In other words, \( \epsilon \Lambda(\Phi \ast f)(x) \) with \( a = \epsilon \) is a wavelet transform of \( f \).

If \( \Phi \) is a sufficiently well-behaved radial function then, in view of (26), the wavelet transform suggested by (38) can be conveniently computed in terms of the Radon transform of \( f \). Indeed, the corresponding univariate wavelet \( \psi \) can, in certain cases, be easily determined from \( \Phi \). For example, by virtue of (15),

\[ 2\hat{\psi}(|\xi|) = |\xi|^2 \hat{\Phi}(\xi) \] (40)

whenever relations (26) and (39) are valid and \( \Phi \) is sufficiently well behaved. In particular, if

\[ \Phi(x) = \frac{1}{2\pi(1 + |x|)^{3/2}}, \]

then \( \Phi(x) \) is the so-called Poisson kernel for the upper half space \( \{(x, \epsilon) : x \in \mathbb{R} \text{ and } 0 < \epsilon < \infty\} \) mentioned earlier; (39) gives

\[ \Psi(x) = \frac{2 - |x|^2}{2\pi(1 + |x|)^{3/2}}, \]

and the relationship between \( \Psi \) and \( \psi \) results in the formula

\[ \psi(t) = \frac{1 - 3t^2}{\pi(1 + t^2)^{3/2}}. \]

The Gaussian kernel

\[ \Phi(x) = \frac{e^{-|x|^2/2}}{2\pi} \]

by virtue of (40) gives rise to

\[ \psi(t) = \frac{(1 - t^2)e^{-t^2/2}}{2\sqrt{2\pi}}. \]

In this case

\[ (\Lambda \Phi)_x f(x) = \frac{1}{2\pi \epsilon} \int_{0}^{2\pi} W_{\Phi} f_\phi(\epsilon, \langle x, u_\phi \rangle) d\theta, \]

where \( \psi \) is the univariate Mexican hat wavelet.
The expression $h_\varepsilon f(x)$ from the “pseudolocal tomography” found in [19] is a wavelet transform of $f$ with $a = \varepsilon$ and

$$\Psi_\varepsilon(x) = \delta(x) - k_\varepsilon(x), \quad (41)$$

where $k_\varepsilon(x) = \Phi_\varepsilon(x)$ is the kernel described by the formulas immediately succeeding (17) and $\delta(x)$ is the bivariate normalized Dirac delta “function.” Thus this analyzing wavelet $\Psi_\varepsilon$ is simply a special case of the general form mentioned in example (ii) in the previous subsection.

On the other hand, the authors of [19] also suggest the computation of $h_\varepsilon(\Phi_{\varepsilon_2} * f)(x)$ for positive $\varepsilon_1$ and $\varepsilon_2$ and appropriately selected $\Phi$ in lieu of the computation of $h_\varepsilon f(x)$. Since, by virtue of (34) and (41),

$$h_\varepsilon(\Phi_{\varepsilon_2} * f)(x) = (\Phi_{\varepsilon_2} - k_{\varepsilon_1} * \Phi_{\varepsilon_2}) * f(x),$$

it is clear that for suitably well-behaved $\Phi$ this expression is some sort of wavelet transform. Indeed, if we write $\varepsilon_1 = c\varepsilon_2 = c\varepsilon$ then

$$h_\varepsilon(\Phi_{\varepsilon_2} * f)(x) = \varepsilon^{-1}W_\varepsilon f(\varepsilon, x),$$

where $\Psi(x) = \Phi(x) - k_\varepsilon * \Phi(x)$. If $\Phi$ is radial then this wavelet transform may, in theory, be computed in terms of the Radon transform of $f$ in a manner similar to that suggested above. Furthermore, if $\varepsilon$ is sufficiently small then one may use the approximation suggested by (36)

$$h_\varepsilon(\Phi_{\varepsilon_2} * f)(x) \approx \frac{2c\varepsilon}{\pi} \Lambda(\Phi_{\varepsilon} * f)(x)$$

together with the formulas for computing $\Lambda(\Phi_{\varepsilon} * f)(x)$.

Of course, similar remarks are also valid for $H_\varepsilon f(x)$.

### 2.5. Wavelet Transforms and Local Approximate Reconstruction

As indicated in the Introduction and supported by the development in Subsections 2.1–2.3, reconstruction methods for Radon transform data which are based on the ridge function–convolution–backprojection paradigm are generally not local. Indeed, if $\int_{\mathbb{R}^2} \Phi(x)dx \neq 0$ then if $\Phi$ enjoys representation (12) the corresponding univariate function $\phi$ cannot be compactly supported. Thus low-pass frequency filters cannot give rise to local reconstruction algorithms.

The material in Subsection 2.4 suggests the use of high-pass frequency filters. Indeed, in retrospect it is quite transparent that there are kernels $\Phi$ with $\int_{\mathbb{R}^2} \Phi(x)dx = 0$ which enjoy representation (12) with the corresponding univariate function $\phi$ having compact
support. This suggests that the wavelet transform may be useful in producing high-pass frequency filters for local approximate reconstruction procedures.

The fact that continuous wavelet transforms are useful in producing local high-frequency reconstructions for Radon transform data was first recognized by Berenstein and Walnut in [2]. Here we give an alternate procedure, which can be easily discretized, based on the development in Subsections 2.2 and 2.3.

Suppose

$$C(x) = K_{1/2}(x) - K_1(x),$$

(42)

where $$K_{e}$$, $$e = \frac{1}{2}$$ or 1, is the convolution kernel in (20). This can also be described via

$$C(x) = \frac{1}{\pi^2} \left\{ \frac{1 - \sqrt{2|x|^2 - 1}}{|x|^2} \chi(|2x|) + \frac{2|x|^2 - 1}{|x|^2} \chi(|x|) \right\}.$$  

In view of (19) and (20) it follows that

$$\Psi_a * f(x) = -\frac{1}{8 \pi^2} \int \int \int_{a \leq |t| < 1} \frac{f_d(x, u_\phi + t) - 2f_d(x, u_\phi) + f_d(x, u_\phi - t)}{t^2} \, dt \, d\theta.$$

(43)

From (43) it is clear that the wavelet transform

$$W_{\Psi} f(a, x) = \Psi_a * f(x)$$

does not require knowledge of the full Radon transform of $$f$$. It can be computed in terms of only those integrals of $$f$$ over lines whose intersection with the disk $$\{x : |x| < a\}$$ is nonempty.

Viewed in terms of the material in Subsections 2.1 and 2.3, the function $$\psi$$ in the ridge function representation of $$\Psi$$ is the distribution

$$\psi(t) = \frac{1}{\pi} \left\{ \delta(t) - \frac{\chi(2t) - \chi(t)}{2t^2} \right\},$$

(44)

where $$\delta(t)$$ is the univariate normalized Dirac delta “function” and $$\chi(t)$$ is the indicator function of the set $$\{t : |t| > 1\}$$. This is an immediate consequence of (43).

Furthermore, by virtue of (42) formula (24) is valid, namely,

$$f(x) = \sum_{j=-\infty}^{\infty} \Psi_{2^j e} * f(x).$$

(45)

In view of (43), formula (45) is an inversion formula for the Radon transform. Truncating (45) appropriately leads to approximate reconstruction formulas.

9 For examples see Subsection 3.5.
For example, if \( f \) is supported in the unit disk \( \{ x : |x| \leq 1 \} \) then the integrand in formula (43) reduces to \(-2f_d(x, u_d)/r^2\) whenever \(|x| \leq 1\) and \(|t| > 2\). Hence if \(|x| \leq 1\) it follows that

\[
\Psi_a \ast f(x) = \frac{1}{2\pi a} \int_0^{2\pi} f_d(x, u)d\theta
\]

whenever \( a \) is greater than 4. In this case formula (45) reduces to

\[
f(x) = \frac{1}{2\pi 2^{-N_0}\epsilon} \int_0^{2\pi} f_d(x, u_d)d\theta + \sum_{j=-\infty}^{N_0} \Psi_{2^j}\ast f(x), \tag{46}
\]

where \( N_0 \) is the least integer so that \( 2^{N_0} > 4/\epsilon \). In numerical applications considerations of resolution impose natural truncation points for the last sum. See the next subsection.

**Remark 1.** It is important to note that the wavelet \( \Psi \) defined by (42) can be replaced with

\[
K_{\epsilon_1}(x) - K_{\epsilon_2}(x)
\]

for any \( \epsilon_1 \) and \( \epsilon_2 \) which satisfy \( 0 < \epsilon_1 < \epsilon_2 \) with consequences analogous to those indicated above for the special case \( \epsilon_1 = \frac{1}{2} \) and \( \epsilon_2 = 1 \). This flexibility may be significant in numerical experiments and practical applications.

**Remark 2.** Explicit error estimates are available for truncations of (44). That is, properties of the kernel \( K \) give rise to estimates of

\[
|f(x) - \sum_{j=M}^{N} \Psi_{2^j}\ast f(x)|
\]

in terms of \( f, M, \) and \( N \). Many of these estimates are routine consequences of the general theory found in [4, 23, 35, 43]. Others also utilize the positivity and symmetry of the kernel \( K \) and can be related to some of the error bounds listed in the Theorem in Subsection 2.2.
and we wish to approximate

$$f(x) = \frac{-1}{8\pi} \int_{-\infty}^{\infty} f_{\phi}(\langle x, u_\phi \rangle + t) - 2f_{\phi}(\langle x, u_\phi \rangle) + f_{\phi}(\langle x, u_\phi \rangle - t) \frac{1}{t^2} \, dt \, d\theta$$

or

$$W_\phi f(a, x) = \frac{1}{2\pi a} \int_{0}^{2\pi} \psi(u) * f_{\phi}(\langle x, u_\phi \rangle) \, d\theta,$$

where

$$\psi(t) = \frac{1}{\pi} \left( \delta(t) - \frac{\chi(2t) - \chi(t)}{2t^2} \right).$$

Here $\delta(t)$ is the univariate normalized Dirac delta “function” and $\chi(t)$ is the indicator function of the set $\{ t : |t| > 1 \}$.

View the inner integral in either formula as the (univariate) convolution of an appropriate distribution $w$ with $f_{\phi}$ and approximate the outer integration via the periodic trapezoid rule. This results in the approximation

$$c = \frac{2\pi}{M} \sum_{m=0}^{M-1} w * f_{\phi_m}(\langle x, u_\phi \rangle),$$

where $\theta_m = m\pi/M$, $m = 0, 1, \ldots, M - 1$, and $c$ is the appropriate constant.

There are several ways to approximate the convolutions $w * f_{\phi_m}$. We use the paradigm suggested in [25]. Namely, first replace $w$ with the distribution

$$\tilde{w}(t) = \sum_{n=-\infty}^{\infty} w_n \delta(t - \frac{n}{N}),$$

where $\delta(t)$ is the univariate unit Dirac delta “function” at the origin,

$$w_n = \int_{-\infty}^{\infty} w(t) \phi_n(t) \, dt.$$
and \( \phi_{\mu}(t) \) is sufficiently smooth, is centered and concentrated around \( n/N \), and has total integral one.\(^{10}\) Next replace \( f_{\theta_n} \) with

\[
\tilde{f}_m(t) = \sum_{n=\pm N} f_{\theta_n} \left( \frac{n}{N} \right) \lambda \left( t - \frac{n}{N} \right),
\]

(52)

where \( \lambda \) is a continuous function which satisfies

\[
\lambda \left( \frac{n}{N} \right) = \begin{cases} 
1 & \text{if } n = 0 \\
0 & \text{otherwise}
\end{cases}
\]

so that \( \tilde{f}_m(t) \) is a continuous function which interpolates\(^{11}\) \( f_{\theta_n}(t) \) at \( t = n/N, n = 0, \pm 1, \pm 2, \ldots \). Observe that

\[
\tilde{w} \ast \tilde{f}_m(t) = \sum_{n=\pm N} \left\{ \sum_{l=-N}^{N} w_{n-l} f_{\theta_n} \left( \frac{l}{N} \right) \right\} \lambda \left( t - \frac{n}{N} \right).
\]

(53)

Since

\[
\tilde{w} \ast \tilde{f}_m \left( \frac{n}{N} \right) = \sum_{l=-N}^{N} w_{n-l} f_{\theta_n} \left( \frac{l}{N} \right)
\]

it follows that if \( \lambda \) has compact support then the sum in \( n \) in (53) has a finite number of nonzero terms.

Since \( f_{\theta_n} \left( \frac{n}{N} \right) = R f \left( \frac{m}{N}, \frac{n}{N} \right) \), the approximations suggested by (50) and (53) give rise to approximations of (48) and (49) which can be computed in terms of the discrete Radon transform data (47) of \( f \).

For example, if

\[
\phi_n(t) = N \phi(N t - n),
\]

(54)

where

\[
\phi(t) = \begin{cases} 
1 & \text{if } |t| \leq 1/2 \\
0 & \text{otherwise}
\end{cases}
\]

(55)

\(^{10}\) \( w_n \) is the distribution \( w \) evaluated at \( \phi_{\mu} \), which, instead of the integral, may be more properly expressed as \( w_{\phi} = (w, \phi_{\mu}) \).

\(^{11}\) \( \lambda \) should also be chosen so that \( \tilde{f}_m(t) \) approximates \( f_{\theta_n}(t) \) well.
then the distribution \( w \) in (48) gives rise to

\[
w_n = N \int_{-\infty}^{\infty} \frac{\phi(Nt - n) - 2\phi(n) + \phi(Nt + n)}{t^2} dt,
\]

which simplifies to

\[
w_n = \frac{8N^2}{4n^2 - 1}.
\]  

(56)

The resulting approximation \( Af \) of \( f \) is given by

\[
Af(x) = \frac{N^2}{4\pi M} \sum_{m=0}^{M-1} \sum_{n=-\infty}^{\infty} \left( \sum_{l=-N}^{N} \frac{f_{n+m}(l/N)}{1 - 4(n - l)^2} \right) \lambda \left( \langle x, u_n \rangle - \frac{n}{N} \right).
\]  

(57)

In the case

\[
\lambda(t) = \begin{cases} 
1 - N|t| & \text{if } |t| \leq \frac{1}{N} \\
0 & \text{otherwise}
\end{cases}
\]  

(58)

the algorithm suggested by (57) is the celebrated algorithm introduced by Logan and Shepp in [40].

To approximate the wavelet transform (49) observe that the distribution \( w = \psi_n \) in (49) and the \( \phi \) in (54) give rise to

\[
w_n = N \int_{|t| \leq a} \frac{\phi(Nt - n) - 2\phi(n) + \phi(Nt + n)}{t^2} dt.
\]

To simplify evaluation of this expression we choose a convenient relationship between the parameters \( a \) and \( N \), namely,

\[
a = \frac{2^\ell}{N}
\]  

(59)

for some integer \( \ell, \ell = 0, 1, 2, \ldots \). If (59) holds then

\[
w_n = \begin{cases} 
- \frac{4N^2}{2^{2\ell}} & \text{if } n = 0 \\
\frac{8N^2}{4n^2 - 1} & \text{if } 2^{\ell-1} < |n| < 2^\ell \\
\frac{4N^2}{2^{2\ell}(2^{2\ell+1} - 1)} & \text{if } |n| = 2^{\ell-1} \\
\frac{2^{\ell+1}(2^{\ell+1} - 1)}{4N^2} & \text{if } |n| = 2^\ell \\
0 & \text{otherwise.}
\end{cases}
\]  

(60)
Denote the train of $\delta$ functions with weights given by (60) as $\tilde{w}_a$ to indicate its dependence on $a$. Here $a$ takes on the possible values $2^\ell/N$, $\ell = 0, 1, 2, \ldots$. Then

$$ AW_{\Psi} f(a, x) = \frac{-1}{4\pi M} \sum_{m=0}^{M-1} \tilde{w}_a * \tilde{f}_m((x, u_m)) $$

is a natural approximation to a wavelet transform of $f$ which is computable in terms of its Radon transform data (47). Furthermore, if $\lambda(t)$ is chosen to have small support such as (58) then $AW_{\Psi} f(a, x)$ depends only on integrals of $f$ over lines which intersect the disc of radius $a$ centered at $x$; more precisely, $AW_{\Psi} f(a, x)$ depends only on $Rf(m \pi/M, n/N)$ for $|n/N - (x, u_m \pi/M)| \leq a$ and $m = 0, 1, \ldots, M - 1$.

Finally, if we denote the train of $\delta$ functions with weights given by (56) as $\tilde{w}$, it is not difficult to see that

$$ \tilde{w}(t) = \sum_{\ell=0}^{\infty} \tilde{w}_{2^\ell N}(t). $$

Relationship (62) is essentially a discrete analogue of (45). Using reasoning similar to that which led to (46) results in the approximation

$$ Af(x) = \sum_{j=0}^{N_0} AW_{\Psi} f(2^j N, x) + \frac{N}{\pi M 2^{N_0}} \sum_{n=0}^{M-1} \tilde{f}_m((x, u_m)), $$

where $N_0$ is the smallest integer such that $2^{N_0} > 2N$. What is remarkable about this simple decomposition is the fact that often the first few terms of the first sum are sufficient to obtain a reasonable reconstruction. This is illustrated by Example 1 in Section 4.

Remark 1. Similar formulas that involve the more general $\Psi$’s mentioned in Remark 1 at the end of the previous subsection are valid. The significance of this lies in the fact that one may want to experiment with the various choices of parameters $\epsilon_1$ and $\epsilon_2$ to determine selections most suitable for specific applications.

Remark 2. The choice of $\phi$ suggested by (55) was motivated by the connection with the classical algorithm of Shepp and Logan [40]. However, in view of the distribution $w$ in (48) and (49), this choice seems somewhat less than ideal.

A smooth, yet extremelly simple, selection is

$$ \phi(t) = \begin{cases} 1 - 3t^2 + 2|t|^3 & \text{if } |t| \leq 1 \\ 0 & \text{otherwise}, \end{cases} $$

which is twice continuously differentiable, is supported in $\{t : |t| \leq 1\}$, and satisfies

$$ \phi(0) = 1, \quad \sum_{n=-\infty}^{\infty} \phi(t - n) = 1, \quad \text{and} \quad \int_{-\infty}^{\infty} \phi(t) dt = 1. $$
In this case the distribution of $\tilde{w}$ corresponding to the $w$ in (48) has the weights

$$w_n = c \begin{cases} 
1 & \text{if } n = 0 \\
1 - 2 \log 2 & \text{if } |n| = 1 \\
1 - (|n| + n^2) \log \left( \frac{|n| + 1}{|n|} \right) - (|n| - n^2) \log \left( \frac{|n|}{|n| - 1} \right) & \text{if } |n| \geq 2,
\end{cases}$$

(65)

where $c$ is an appropriately chosen constant. In other words, (65) is the analogue of (56) if (64) is used instead of (55).

The weights for the distributions $\tilde{w}_a$, $a = 2^\ell/N$, $\ell = 0, 1, 2, \ldots$, which may be used to approximate the wavelet transform (49) are the following: when $\ell = 0$,

$$w_n = c \begin{cases} 
3 & \text{if } n = 0 \\
-3/2 & \text{if } |n| = 1 \\
0 & \text{if } |n| \geq 2;
\end{cases}$$

when $\ell = 1$,

$$w_n = c \begin{cases} 
2 & \text{if } n = 0 \\
-24 \log 2 - 16 & \text{if } |n| = 1 \\
24 \log 2 - 17 & \text{if } |n| = 2 \\
0 & \text{if } |n| \geq 3;
\end{cases}$$

and when $\ell \geq 2$,

$$w_n = 2c \begin{cases} 
1/2^\ell & \text{if } n = 0 \\
6 \left( 1 - (|n| + n^2) \log \left( \frac{|n| + 1}{|n|} \right) - (|n| - n^2) \log \left( \frac{|n|}{|n| - 1} \right) \right) & \text{if } 2^{\ell-1} < |n| < 2^\ell \\
-1/|n| + 3 + 6|n| - 6(|n| + n^2) \log \left( \frac{|n| + 1}{|n|} \right) & \text{if } |n| = 2^{\ell-1} \\
1/|n| + 3 - 6|n| - 6(|n| - n^2) \log \left( \frac{|n|}{|n| - 1} \right) & \text{if } |n| = 2^\ell \\
0 & \text{otherwise},
\end{cases}$$

(66)

where $c$ is an appropriately chosen constant and is the same in all three formulas. In other words, (66) and the two preceding formulas\footnote{Since all three formulas are special cases of one general formula we will refer to them collectively as (66) in Section 4.} are the analogue of (60) if (64) is used instead of (55). The wavelet transforms in our numerical examples are computed with these weights.

3. DETAILS

3.1. Details of Section 2.1

$\Phi$ is the uniform sum of ridge functions in the distributional sense if there is an even univariate distribution $\phi$ such that
for all test functions \( f \). Here \( B_\phi \) is the backprojection of \( \phi \) defined by

\[
\langle B_\phi \phi, f \rangle = \langle \phi, f \rangle.
\]

Note that if \( f \) is a locally integrable function then \( B_\phi \) reduces to the expected expression, namely \( B_\phi f(x) = \phi(x, u_\theta) \).

### 3.2 Details of Section 2.2

To see that \( F_t(x) \) is the average of the integrals of \( f \) over lines which are a distance \(|t|\) from \( x \) simply write

\[
2\pi F_t(x) = \int_0^{2\pi} \int_{-\infty}^{\infty} f((x, u_\theta) + tu_\theta + sv_\theta) ds d\theta
\]

because the integration in \( s \) is translation invariant. Since \( x = \langle x, u_\theta \rangle u_\theta + \langle x, v_\theta \rangle v_\theta \), we may write

\[
F_t(x) = \frac{1}{2\pi} \int_0^{2\pi} \left\{ \int_{-\infty}^{\infty} f(x + tu_\theta + sv_\theta) ds \right\} d\theta,
\]

which is the desired result since the expression in braces is the integral of \( f \) over the line which is \( t \) units away from \( x \) in the direction \( u_\theta \).

The inversion formula (18) is not equivalent to (2). However, it is a natural consequence of certain summability formulas for the inversion of the Radon transform. The details can be found in [29, p. 196; 30].

The formulas \( g_e f(x) = \Phi_e * f(x) \) and \( G_e f(x) = K_e * f(x) \) follow from the identities

\[
\int_0^{2\pi} f_d(\langle x, u_\theta \rangle - t)d\theta = 2 \int_{|y|>|t|} \frac{f(x - y)}{\sqrt{|y|^2 - t^2}} dy,
\]

\[
\int_0^{2\pi} \int_{|y|>|t|} \frac{f_d(\langle x, u_\theta \rangle)}{t} dtd\theta = 4 \int_{|y|>\varepsilon} \frac{1}{|y|} f(x - y) dy.
\]
and
\[ \int_0^{2\pi} \int_{|y|<\epsilon} \frac{f_\delta(x, u_\delta) - t}{t^2} \, dt \, d\theta = 4 \int_{|y|<\epsilon} \frac{\sqrt{|y/e|^2 - 1}}{|y|^2} f(x - y) \, dy. \] (69)

Let \( I_1(t), I_2(\epsilon), \) and \( I_3(\epsilon) \) denote the right-hand sides of formulas (67), (68), and (69), respectively. Note that \( I_1(t) \) and \( I_3(\epsilon) \) can also be expressed as
\[ I_1(t) = \int_0^{2\pi} f_\delta(x, u_\delta + t) \, d\theta \]
and
\[ I_3(\epsilon) = \int_0^{2\pi} \int_{|y|<\epsilon} \frac{f_\delta(x, u_\delta) + t}{t^2} \, dt \, d\theta \]
or
\[ I_3(\epsilon) = 2 \int_0^{2\pi} \int_\epsilon^\infty f_\delta(x, u_\delta + t) \, dt \, d\theta. \]

Thus
\[ g_\epsilon f(x) = \frac{1}{2\pi^2} \left\{ \frac{I_1(\epsilon)}{\epsilon} - \frac{I_3(\epsilon)}{2} \right\} \]
\[ = \int_{|y|>\epsilon} \frac{1}{\epsilon^2 \pi^2} \left\{ \frac{1}{\sqrt{|y/e|^2 - 1}} - \frac{\sqrt{|y/e|^2 - 1}}{|y/e|^2} \right\} f(x - y) dy = \Phi_\epsilon * f(x) \]
and
\[ G_\epsilon f(x) = \frac{-1}{8\pi^2} \left\{ 2I_2(\epsilon) - 2I_3(\epsilon) \right\} \]
\[ = \int_{\mathbb{R}^2} \frac{1}{\epsilon^2 \pi^2} \left\{ \frac{1}{|y/e|} - \frac{\sqrt{|y/e|^2 - 1}}{|y/e|^2} \chi(\epsilon) \right\} f(x - y) dy = K_\epsilon * f(x), \]
which are the desired results.
Formula (68) is simply a consequence of

\[ \int_{0}^{2\pi} f_\theta(x, u_\theta) \, d\theta = \int_{|y|} \frac{2}{|y|} f(x - y) \, dy \]

and

\[ \int \frac{dt}{t^2} = \frac{2}{\epsilon}. \]

To see formula (69) use identity (67) to write

\[ \int_{0}^{2\pi} \int_{|y|>\epsilon} \frac{f_\theta(x, u_\theta - t)}{t^2} \, d\theta \, dt = 2 \int_{|y|>\epsilon} \int_{|y|>\epsilon} \frac{f(x - y)}{t^2 \sqrt{|y|^2 - t^2}} \, dy \, dt \]

\[ = 2 \int_{|y|>\epsilon} \left\{ \int_{|y|>\epsilon} \frac{dt}{t^2 \sqrt{|y|^2 - t^2}} \right\} f(x - y) \, dy \]

\[ = 2 \int_{|y|>\epsilon} 2 \frac{|y/\epsilon|^2 - 1}{|y|^2} f(x - y) \, dy, \]

which is the desired result.

Finally, to see (67) first recall the second displayed identity in this subsection, namely

\[ \int_{0}^{2\pi} f_\theta(x, u_\theta - t) \, d\theta = \int_{0}^{2\pi} f(x - tu_\theta + sv_\theta) \, dsd\theta. \]

Now

\[ -tu_{\theta} + sv_{\theta} = -\sqrt{t^2 + s^2} (\alpha \cos \theta + \beta \sin \theta, \alpha \sin \theta - \beta \cos \theta) \]

where

\[ \alpha = \frac{t}{\sqrt{t^2 + s^2}} = \cos \phi \quad \text{and} \quad \beta = \frac{s}{\sqrt{t^2 + s^2}} = \sin \phi \]

so that

\[ -tu_{\theta} + sv_{\theta} = -\sqrt{t^2 + s^2} (\cos(\theta - \phi), \sin(\theta - \phi)) = -\sqrt{t^2 + s^2} u_{\theta - \phi}. \]
Thus we may write

\[
\int_0^{2\pi} \int_{-\infty}^{\infty} f(x - iu_\theta + s\nu_\theta)dsd\theta = \int_{-\infty}^{\infty} \int_0^{2\pi} f(x - \sqrt{r^2 + s^2}u_\theta)d\theta ds
\]

\[
= \int_{-\infty}^{\infty} \left\{ \int_0^{2\pi} f(x - \sqrt{r^2 + s^2}u_\theta)d\theta \right\} ds,
\]

\[
= 2\int_0^{2\pi} \left\{ \int_{-\infty}^{\infty} f(x - ru_\theta \frac{rdr}{\sqrt{r^2 - t^2}})d\theta \right\} dt,
\]

\[
= 2\int_{|t|>|y|} \frac{f(x - y)}{\sqrt{|y|^2 - t^2}}dy.
\]

where the second equality follows from translation invariance in the \( \theta \) variable, the third is a consequence of the fact that the integrand is even in the \( s \) variable, the fourth is a result of the change of variable \( r = \sqrt{t^2 + s^2} \), and the last follows from the polar change of variable \( y = ru_\theta \). This implies (67) and completes the proof of the formulas \( g_\epsilon f(x) = \Phi_\epsilon * f(x) \) and \( G_\epsilon f(x) = K_\epsilon * f(x) \).

As mentioned earlier the various assertions of the Theorem are routine consequences of the representation \( G_\epsilon f = K_\epsilon * f \) and the properties of \( K \). More specifically:

- Statements (ii), (iii), and (iv) are well-known consequences of the fact that \( K \) is integrable and \( \int_{\mathbb{R}^2} K(x)dx = 1 \).
- Statement (v) follows from the additional fact that \( \int_{\mathbb{R}^2} |x|^\alpha |K(x)|dx \) is finite if \( 0 < \alpha < 1 \) and \( |x||K(x)| \leq C|x|^{-2} \) for \( |x| > 2 \).
- Assertion (vi) follows from the additional fact that \( K \) is a positive radial function.
- Statement (vii) follows from the general theory of Hölder classes of functions and the fact that

\[
\int_{\mathbb{R}^2} |K(x + y) - 2K(x) + K(x - y)|^qdx \leq C|y|^{2-q}
\]

if \( 1 \leq q < 2 \), where \( C \) is a constant independent of \( y \). See, for example [4, 23, 35, 44].

- Finally, (i) follows from the facts that \( \int_{\mathbb{R}^2} K(x)dx = 1 \) and that \( |K(x)| \) is dominated by a integrable radially decreasing function; see [44].

As alluded to earlier all the items except (i) of this Theorem remain true if \( G_\epsilon f \) is replaced with \( g_\epsilon f \); the arguments are essentially identical. Since the kernel \( \Phi(x) \) in (17)
is not dominated by an integrable radially decreasing function, the argument used to prove (i) is not valid for $g \ast f$. Indeed, without further restrictions on $f$, item (i) fails to hold in this case.

### 3.3 Details of Section 2.3

To see (22) write

$$
\int_0^\infty W_\varphi f(a, x) \frac{da}{a} = (2\pi)^{-2} \int \int \int e^{i\xi \cdot \varphi} \hat{\varphi}(a, \xi) \hat{f}(\xi) d\xi \frac{da}{a}
$$

$$
= c(2\pi)^{-2} \int e^{i\xi \cdot \varphi} \hat{f}(\xi) d\xi = cf(x),
$$

where

$$
c = \int_0^\infty \hat{\varphi}(a, \xi) \frac{da}{a} = \frac{1}{2\pi} \int \int \hat{\varphi}(\xi) \frac{d\xi}{|\xi|^2}.
$$

The fact that

$$
\lim_{M \to -\infty} \Phi_{2^m} \ast f(x) = f(x) \quad \text{and} \quad \lim_{N \to -\infty} \Phi_{2^m} \ast f(x) = 0
$$

is elementary. See, for example [28].

### 3.4 Details of Section 2.4

#### 3.4.1

The formal adjoint of $R$ is the transformation $R^g$, which maps suitably well-behaved scalar-valued functions $g(\theta, t)$ on $[0, 2\pi) \times \mathbb{R}$ to functions $R^g(x)$ on $\mathbb{R}^2$ and satisfies the relation

$$
\int_{\mathbb{R}^2} f(x) R^g(x) dx = \int_{0}^{2\pi} \int_{-\infty}^{\infty} Rf(\theta, t) g(\theta, t) dt d\theta.
$$

Routine calculations, namely

$$
\int_{0}^{2\pi} \int_{-\infty}^{\infty} Rf(\theta, t) g(\theta, t) dt d\theta = \int_{0}^{2\pi} \int_{-\infty}^{\infty} f(tu_\theta + s\nu_\theta) g(\theta, t) ds dt d\theta
$$

$$
= \int_{0}^{2\pi} \int_{\mathbb{R}^2} f(x) g(\theta, \langle x, u_\theta \rangle) dxd\theta,
$$
where \( t = \langle x, u \rangle \) and \( s = \langle x, v \rangle \), show that

\[
R^\theta g(x) = \int_0^{2\pi} g(\theta, \langle x, u \rangle) d\theta.
\]

To see that \( R^\theta f = 4\pi J \ast f \) write

\[
R^\theta R f(x) = \int_0^{2\pi} \int_{-\infty}^{\infty} f(\langle x, u \rangle + s \nu_\theta) ds d\theta.
\]

Because the inner integral is translation invariant the variable \( s \) may be replace by \( \langle x, \nu_\theta \rangle + s \), which, in view of the fact that \( x = \langle x, u \rangle u_\theta + \langle x, \nu_\theta \rangle \nu_\theta \), results in

\[
R^\theta R f(x) = 2 \int_0^{2\pi} \int_{-\infty}^{\infty} f(x + s \nu_\theta) ds d\theta
\]

\[
= 2 \int_0^{2\pi} \int_{0}^{\infty} \frac{f(x - su_\theta)}{s} s ds d\theta = \int_{\mathbb{R}^2} \frac{f(x - y)}{|y|} dy
\]

when the polar change of variables \( y = su_\theta \) is used.

The fact that the Fourier transform of \( |x|^{-1} \) is a constant multiple of itself follows from the elementary fact that the bivariate Fourier transform of a radial distribution homogeneous of degree minus one is a radial distribution homogeneous of degree minus one. The value of the constant follows from the Fourier inversion formula

\[
f(x) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} \hat{f}(\xi) e^{i\langle x, \xi \rangle} d\xi.
\]

Recall that the Laplacian \( \Delta \) is the differential operator defined by

\[
\Delta f(x) = \frac{\partial^2 f}{\partial x_1^2}(x) + \frac{\partial^2 f}{\partial x_2^2}(x)
\]

whose Fourier transform is

\[
\hat{\Delta f}(\xi) = -|\xi|^2 \hat{f}(\xi).
\]

In view of this all the formulas involving \( \Delta \) and \( \Lambda^{-1} \) are routine consequences of previous identities.
3.4.2

There are many ways to arrive at (31). Perhaps the most direct and simplest method is to use (i) the fact that the (univariate) Fourier transform of $f_u$ is equal to the (bivariate) Fourier transform of $f$ restricted to the line $\tau u_\theta, -\infty < \tau < \infty$, or symbolically $\hat{f}_u(\tau) = \hat{f}(\tau u_\theta)$; (ii) the Fourier inversion formula; and (iii) a polar change of variables. See for example [32, 41]. It is also a natural consequence of various summability formulas for the inversion of the Radon transform; see [29, p. 196]. Of course, it can also be derived from (2) by using the fact that

$$f_y(t) = f_{y+a}(-t)$$

and an interchange of order of differentiation and integration.

Reasonable smoothness conditions on $f$ needed to make formula (31) valid are not clear. Certainly $f \in C^{1+a}$ for some positive $\alpha$ is sufficient\(^{13}\) but it is far too restrictive for typical applications. For this reason we take (34) to be the definition of $h_\varepsilon f(x)$ and view (33) as an expression valid for sufficiently smooth $f$.

In view of (36), if for a given point $x$ the ratio $h_\varepsilon f(x)/\varepsilon$ fails to be bounded as $\varepsilon$ goes to 0 it should be clear that for the same $x$ the quantity $\Lambda f(x)$ will not be finite. The behavior of $h_\varepsilon f(x)$ for small $\varepsilon$ is a consequence of the behavior of $f$ at $x$ and the properties of the kernel $\Phi$ in representation (17). We will not detail this behavior here but only mention that, in view of the fact that $h_\varepsilon f(x) = f(x) - g_\varepsilon f(x)$, some of this behavior is described by the statements in the Theorem found in Subsection 2.2. In particular, if $f$ fails to be sufficiently smooth at $x$ the ratio $h_\varepsilon f(x)/\varepsilon$ fails to be bounded as $\varepsilon$ goes to 0 and $\Lambda f(x)$ fails to be finite. For such reasons the authors of “local tomography” and “pseudo-local tomography” suggest that $\Lambda f$ and $h_\varepsilon f$, respectively, and their regularized variants could be useful in studying the singularities or edges of $f$.

Note that representation (17) not only implies the theoretical properties of $h_\varepsilon f$ and other theoretic convergence results recorded in [19] but in many instances implies stronger variants of those results. For example, to obtain the conclusion of [19, Theorem 1, (3.4)] the hypothesis that $f$ is $C^2$ in a neighborhood of $x$ can be weakened to $f$ being Hölder continuous of order $\alpha$ at $x$ for some $\alpha > 1$; furthermore, in the case that $f$ is Hölder continuous of order $\alpha$ at $x$ for some $\alpha$ satisfying $0 < \alpha \leq 1$, representation (17) implies corresponding error bounds; see item (v) in the Theorem found in Subsection 2.2. In view of relationships such as (36), representation (17) can also be used to obtain theoretical results concerning the behavior of $\Lambda f$; certain detailed results concerning this behavior can be found in [11].

3.4.3

The way (37) is computed in terms of the Radon transform data of $f$ is dictated by formula (29) and

$$(\Phi_\varepsilon * f)_d(t) = (\Phi_\varepsilon)_d * f_d(t).$$

---

\(^{13}\) $C^{1+a}$ is the class of those differentiable functions whose derivatives are Hölder continuous of order $\alpha$. 
This results in
\[ \Lambda(\Phi \ast f)(x) = -\frac{1}{4\pi} \int_0^{2\pi} (\Phi_\lambda)_{\theta}^x \ast f_\theta(x, u_\theta) \, d\theta. \]

For more details see [10, 20, 42]. Thus
\[ \psi(t) = -\frac{\tau^2}{2} (\Phi_\lambda)^{\prime\prime}(t), \]
which is another variant of (40).

### 3.5 Details of Section 2.5

To see that there are radial functions \( \Psi \) whose ridge function representatives \( \psi \) are compactly supported if we allow the total integral of \( \Psi \) to be zero, let \( \phi(t) \) be any (univariate) continuous function with compact support. If \( \phi \) is odd, that is, \( \phi(-t) = -\phi(t) \), consider the function
\[ \psi(t) = \frac{\phi(t + s) - \phi(t - s)}{2s}, \quad (70) \]
where \( s \) is any fixed positive number or, if \( \phi \) is differentiable, the limiting case
\[ \psi(t) = \phi'(t), \quad (71) \]
where \( \phi' \) is the derivative of \( \phi \). If \( \phi \) is even, that is, \( \phi(-t) = \phi(t) \), consider
\[ \psi(t) = \alpha \phi(\alpha t) - \beta \phi(\beta t), \quad (72) \]
where \( \alpha \) and \( \beta \) are any fixed pair of positive numbers such that \( \alpha \neq \beta \). Note that in each case the function \( \psi \) is even and has compact support which can be determined in terms of the support of \( \phi \).

Suppose that \( \psi \) is any one of the functions defined by Eqs. (70)–(72) and the support of \( \psi \) is contained in the interval \([-r, r]\). Then the following is true:

- \( \psi \) is a univariate wavelet with compact support. Furthermore, \( \hat{\psi}(\tau) = O(\tau^2) \) as \( \tau \) goes to 0.
- The function
\[
\Psi(x) = \frac{1}{2\pi} \int_0^{2\pi} \psi(x, u_\theta) \, d\theta
\]
is well defined and is integrable,

\[
\int_{\mathbb{R}^2} \Psi(x)dx = 0,
\]

and

\[
\int_{\mathbb{R}^2} \frac{|\Psi(\xi)|}{|\xi|} d\xi < \infty.
\]

- The value of the corresponding wavelet transform \( W^{\Psi}_f(a, x) \) can be computed in terms of \( f_{\theta}(x, u_\theta) \), \( 0 \leq \theta < \pi \), \(-ar \leq t \leq ar\). In other words, \( W^{\Psi}_f(a, x) \) can be evaluated in terms of the integrals of \( f \) over lines which are no greater distance than \( ar \) from \( x \).

Of course, there are many other ways, including a combinations of the above methods, to obtain compactly supported univariate wavelets \( \phi \) which give rise to bivariate wavelet transforms with the above properties. Explicit examples of functions \( f \) which do the job in (70), (71), or (72) are the even functions

\[
\phi(t) = \begin{cases} (1 - t^2)^p & \text{if } |t| < 1 \\ 0 & \text{if } |t| \geq 1, \end{cases}
\]

where \( p \) is any number greater than 0,

\[
\phi(t) = \begin{cases} \exp((t^2 - 1)^{-1}) & \text{if } |t| < 1 \\ 0 & \text{if } |t| \geq 1, \end{cases}
\]

and their odd cousins \( t\phi(t) \). The reader should have no difficulty devising many other explicit examples.

In view of the application under consideration, explicit formulas for the corresponding bivariate wavelets \( \Psi \) are not needed for computational purposes. In any event, \( \Psi \) can be numerically evaluated from (26) or its substitutes; see [29].

To see formula (44) for the wavelet \( \psi \) observe that for any sufficiently well-behaved univariate function \( g \),

\[
\psi_a \ast g(s) = \frac{-a}{4\pi} \int_{a|t| < a} \frac{g(s + t) - 2g(s) + g(s - t)}{t^2} dt
\]

\[
= \frac{g(s)}{\pi} - \frac{a}{2\pi} \int_{a|t| < a} \frac{g(s - t)}{a^2(t/a)} dt = \frac{g(s)}{\pi} - \tilde{\psi}_a \ast g(s).
\]
where

\[ \tilde{\psi}(t) = \frac{\chi(2t) - \chi(t)}{2\pi t^2}, \]

\[ \psi_{\alpha}(t) = a^{-1}\tilde{\psi}(t/a) , \]

and \( \chi(2t) - \chi(t) \) is the indicator function of \( \{ t : 1/2 < |t| \leq 1 \} \).

### 3.6 Details of Section 2.6

There are several points of view concerning convolution–backprojection-type reconstruction algorithms in computed tomography [7, 9, 16, 20, 29, 32, 33, 39–42]. We will not outline them all here and only mention that the connections between the various inversion methods and their discrete implementations are based on well-established, but somewhat heuristic, principles. See, for example, the arguments on pp. 102–108, particularly Theorem 1.1, in Natterer’s excellent treatise [32]. The relatively recent papers [9, 22, 33, 34] nicely illustrate the state of the art.

Since [25] may not be easily accessible, we now briefly outline the rationale for the paradigm used to approximate

\[ w \ast g(s) = \int_{-\infty}^{\infty} w(s-t)g(t)dt, \]

where \( w \) is a known univariate integrable function or measure, whose derivative is very large or even nonexistent in the classical sense, and \( g \) is a univariate function supported in the interval \(-1 \leq t \leq 1\) whose samples \( g(n/N), n = 0, \pm 1, \pm 2, \ldots, \pm N \), are known.

First assume that \( g \) is reasonably well behaved so that approximation

\[ \tilde{g}(t) = \sum_{n=-N}^{N} g(n/N)\lambda(t - n/N) \]

enjoys the correct order of approximation. For example, if \( g \) is Hölder continuous of order \( \alpha \) then

\[ |g(t) - \tilde{g}(t)| \leq \frac{C}{N^\alpha}. \]  \hspace{1cm} (73)

The primary rationale of this method is that the approximant is easy to compute and the approximation of \( w \ast g \) is of the same order as \( |g(t) - \tilde{g}(t)| \), e.g., (73). Now, consider the method outlined in Subsection 2.6; that is, replace \( w \) with

\[ \tilde{w}(t) = \sum_{n=-N}^{N} w_n\delta\left(t - \frac{n}{N}\right), \]
where \( \delta(t) \) is the univariate unit Dirac delta “function” at the origin,

\[
w_n = \int_{-\infty}^{\infty} w(t) \phi_n(t) \, dt,
\]

and \( \phi_n(t) \) is sufficiently smooth, is centered and concentrated around \( n/N \), and has total integral one. It is not difficult to see that (i) \( \tilde{w} * \tilde{g} \) is of the same form as \( \tilde{g} \), that is, a linear combination of translates of \( \lambda \); (ii) if the \( \phi_n \)'s are chosen appropriately then \( \tilde{w} * \tilde{g} \) is easy
to compute; and (iii) \( |w \ast g(t) - \tilde{w} \ast \tilde{g}(t)| \) is of the same order as \( |g(t) - \tilde{g}(t)| \), e.g., if (73) is the case then

\[
|w \ast g(t) - \tilde{w} \ast \tilde{g}(t)| \leq \frac{LC}{N^n},
\]

where \( L \) is the \( L^1 \) norm or total variation of \( w \).

FIG. 3. The wavelet transform \( AW_{\varphi f(2/N, x)} \) via (47) using weights (53).

FIG. 4. The wavelet transform \( AW_{\varphi f(4/N, x)} \) via (47) using weights (52).
4. NUMERICAL EXAMPLES

We present the results of two numerical experiments based on the algorithms suggested by the formulas in Subsection 2.6. As is customary, we used the celebrated phantom \( \Phi \) which was originally introduced and documented in [40]. That is, \( \Phi(x) \) is the linear combination of indicator functions of ellipses as described in [40]. MATLAB software was used to implement the algorithms and display the results.

4.1. Example 1

The point of this example is to compare a classical reconstruction with that suggested by the discrete analogues of the wavelet transform.

We used the data described by (47) with \( f(x) = \Phi(x) \), \( M = 90 \), and \( N = 100 \). The results are summarized in Figs. 1–5. Each “reconstruction” was evaluated on a 201 \( \times \) 201 grid determined by \( x = ((j_1 - 100)/100, (j_2 - 100)/100) \) and is represented by a

![FIG. 5. The reconstruction \( A_f(x) \) via (49) using weights (53) and \( N_0 = 7 \).](image)

![FIG. 6. Full data, conventional reconstruction, and local wavelet transform.](image)
matrix $I = (I_{j_2,j_1})$, where $j_1$ and $j_2$ take on the values 0, 1, ..., 200. Each figure contains three plots.

- The first plot is of the “cross section” $I_{101,j_1}$, $j_1 = 0, 1, ..., 200$.
- The second plot is of $I_{101,j_1}$, $j_1 = 0, 1, ..., 200$, where $I_\tau$ is the range truncated variant of $I$ defined by

$$I_{\tau,j_1} = \begin{cases} 
\mu_4 & \text{if } I_{j_1,j_1} \geq \mu_4 \\
I_{j_1,j_1} & \text{if } \mu_0 \leq I_{j_1,j_1} \leq \mu_4 \\
\mu_0 & \text{if } I_{j_1,j_1} \leq \mu_0.
\end{cases}$$

The values of $\mu_0$ and $\mu_4$ were chosen so that the variations inside the “skull” would be visible and were determined experimentally by viewing the plot of $I_{101,j_1}$ with different scalings on the vertical axis.

- The third plot is a gray level representation (image) of $I_\tau$ using 256 gray levels uniformly distributed between $\mu_0$ and $\mu_4$.

Figure 1 gives the results of the computation of $Af(x)$, the approximation of $f$ given by formula (57). In other words, this is the result of applying the celebrated Shepp–Logan algorithm [40]. It is presented for comparison purposes.
We computed the approximations \( AW_{f_2}(2^j/N, x) \), \( j = 0, 1, \ldots, 7 \), to the wavelet transform using the algorithm suggested by (61) and the weights given by (66).

Figures 2, 3, and 4 show \( AW_{f_2}(a, x) \) with \( a = 1/N, 2/N \), and \( 4/N \), respectively. Figure 5 shows the reconstruction

\[
A_f(x) = \sum_{j=0}^{7} AW_{f_2}(2^j/N, x) + \frac{N}{\pi M^2} \sum_{m=0}^{M-1} \int u((x, u_m)).
\]

The algorithms for computing \( A_f(x) \) were not normalized so that the plots in Figs. 1 and 5 should only be compared relatively, not absolutely.

### 4.2. Example 2

The point of this example is to numerically illustrate the local nature of the wavelet transform \( W_{\Psi}(a, x) \) with \( \Psi \) as described by (42), when computed from Radon transform data.

We used the data

\[
Rf\left(\frac{m\pi}{M}, \frac{2n}{N}\right), \quad m = 0, 1, \ldots, M - 1, \quad n = 0, \pm 1, \pm 2, \ldots, \pm N,
\]

(74)

with \( f(x) = \Phi(x + x_0) \), \( M = 90 \), and \( N = 200 \). Here \( x_0 = (0, -0.605) \) is the center of the central small ellipse in the lower portion of the original phantom \( \Phi \). We computed “reconstructions” from various truncations of the data matrix, \( D_{m,n} = Rf\left(\frac{m\pi}{M}, \frac{2n}{N}\right) \), in the second variable; more specifically \( n = 0, \pm 1, \pm 2, \ldots, \pm n \), with \( r = 1, \frac{1}{3}, \frac{1}{5}, \frac{1}{7}, \) and \( \frac{1}{9} \). The results are summarized in Figs. 6–15.

Each “reconstruction” was evaluated on a 401 \( \times \) 401 grid determined by \( x = ((j_1 - 200)r/100, (j_2 - 200)r/100) \) and is represented by a matrix \( I = (I_{j_1,j_2}) \), where \( j_1 \) and \( j_2 \) take on the values 0, 1, \ldots, 400. Each of Figs. 6–10 contains three plots.

FIG. 9. One-quarter of full data, conventional reconstruction, and local wavelet transform. Scaled to be proportional to Fig. 8.

FIG. 10. One-eighth of full data, conventional reconstruction, and local wavelet transform. Scaled to be proportional to Figs. 8 and 9.
• The first plot is a gray level representation (image) of the data using 256 gray levels uniformly distributed between 0 and the maximum of the data.

• The second plot is a gray level representation (image) of the truncated matrix \( It \), with values between \( \mu_0 \) and \( \mu_1 \), which represents the result of simply applying the Shepp–Logan algorithm (formula (57)) to the data; 256 gray levels uniformly distributed between \( \mu_0 \) and \( \mu_1 \) were used. The range-truncated matrix \( It \) was obtained by essentially the same procedure as described in Example 1; namely, the values of \( \mu_0 \) and \( \mu_1 \) were chosen so that the variations inside the “skull” would be visible and were determined experimentally by viewing the plot of \( I_{201,j} \) with different scalings on the vertical axis. See Figs. 11–15. In the case \( r = \frac{1}{2} \) it was impossible to find appropriate values of \( \mu_0 \) and \( \mu_1 \); the case \( r = \frac{1}{2} \) was at best borderline.

FIG. 11. Cross sections corresponding to the reconstruction in Fig. 6.

FIG. 12. Cross sections corresponding to the reconstruction in Fig. 7.
• The third plot is a gray level representation (image) of the truncated matrix \( I_r \), with values between \( \mu_0 \) and \( \mu_1 \), which represents \( AW_{\Psi} f(1/N, x) \) as computed by the algorithm suggested by (61) and the weights given by (66); 256 gray levels uniformly distributed between \( \mu_0 \) and \( \mu_1 \) were used. The range-truncated matrix \( I_r \) was obtained as described in the previous item. Note that even in the cases \( r = \frac{1}{4} \) and \( \frac{1}{8} \) it was possible to find appropriate values of \( \mu_0 \) and \( \mu_1 \).

Figures 6–8 are scaled so that they occupy the same area. Figures 9 and 10 are scaled in correct proportion.

Figures 11–15 contain the central cross sections of the reconstructions found in Figs.
6–10, respectively, and document how the range truncated matrices $I_t$ were obtained. Each figure consists of four plots arranged as

1 3
2 4

and which contain, respectively, the following:

1. a plot of the cross section $I_{201,j_1}; j_1 = 0, 1, \ldots, 400$, where the $401 \times 401$ matrix $I$ represents the result of simply applying the Shepp–Logan algorithm (formula (57)) to the corresponding data in Fig. $n - 5$;
2. a plot of $I_{201,j_1}; j_1 = 0, 1, \ldots, 400$, where $I_t$ is the range truncated variant of the matrix $I$ defined in item 1; as mentioned earlier, this truncation was obtained in the same manner as indicated in Example 1;
3. A plot of the cross section $I_{201,j_1}; j_1 = 0, 1, \ldots, 400$, where the $401 \times 401$ matrix $I$ represents $AW_qf(1/N, x)$ as computed from the data in Fig. $n - 5$ by the algorithm suggested by (61) and the weights given by (66);
4. A plot of $I_{201,j_1}; j_1 = 0, 1, \ldots, 400$, where $I_t$ is the range truncated variant of the matrix $I$ defined in item 3; again, this truncation was obtained in the same manner as indicated in Example 1.

ACKNOWLEDGMENTS

This article was initially motivated by my attempt to understand the article “Pseudolocal Tomography” by Katsevich and Ramm [19], the appearance of which was brought to my attention by Dr. Arje Nachman. I am grateful to the reviewers for criticism which led to a clearer presentation and, most importantly, for alerting me to the article [2].
REFERENCES