Introduction to local tomography

Adel Faridani *† Kory A. Buglione Pallop Huabsomboon
Ovidiu D. Iancu Jeanette McGrath

1 Introduction

Computed tomography (CT) entails the reconstruction of a generalized density function $f$ from line integrals of $f$. This reconstruction is not local in the sense that reconstruction of $f$ at a point $x$ requires integrals over lines far from $x$. In a number of applications only part of an object needs to be imaged. Thus it would be desirable to only use integrals over lines intersecting this region-of-interest (ROI). This entails the loss of uniqueness, but it turns out that the null functions are nearly constant inside the ROI and that the singularities of $f$ inside the ROI can be stably recovered from such data; see, e.g., [26, §VI.4], [30].

Over the past decade a number of methods to ‘localize’ the reconstruction have been proposed. These range from methods for ‘region-of-interest tomography’ which use integrals over all lines passing through a region slightly larger than the ROI, to strictly local methods where reconstruction at a point $x$ only requires integrals over lines very close to $x$. In this introduction we will review methods of both types. The wavelet-based multiresolution local tomography of [32], discussed in section §5, uses all lines passing through a region slightly larger than the ROI, while Lambda tomography, reviewed in §3, is strictly local. Pseudolocal tomography, described in §4, can be used in both modes. Some other techniques use a limited amount of data outside the ROI [28], or extrapolate the missing data [26, §VI.4]. Extensions of local tomographic methods to more general settings have been presented in [18, 22, 25].

This article is organized as follows: In the next section we review some background material on the x-ray transform and its inversion.

Section 3 is devoted to Lambda tomography. Here not the function $f$ itself but the related function $Lf = \Lambda f + \mu \Lambda^{-1} f$ is reconstructed, where $\Lambda = \sqrt{-\Delta}$, and $\Delta$ denotes the Laplacian. This reconstruction is strictly local, and our discussion centers on what

*Dept. of Mathematics, Oregon State University, Corvallis, OR 97331.
Email: faridani@math.orst.edu; Homepage: http://ucs.orst.edu/~faridana
†This work was supported in part by NSF grant DMS-9803352.
features of $f$ can be found from reconstructions of $Lf$. Particular attention will be given to the computation of density jumps.

In section 4 we discuss the pseudolocal tomography of [21, 31] and relate this method for computing density jumps to the methods based on Lambda tomography.

Section 5 briefly reviews wavelet based multiresolution tomography [32], one of the wavelet based methods for region-of-interest tomography. The goal of this method is to reconstruct the function $f$ itself up to an almost constant error. For other wavelet based methods see, e.g., [2, 28, 41].

In section 6 we turn to the three dimensional case and examine local cone beam tomography with sources on a curve. Here additional problems arise since it is usually impractical to collect sufficiently many data to ensure stable recovery of all singularities of $f$ inside the region of interest. We discuss which singularities are stably determined and compare the two leading reconstruction algorithms.

2 The x-ray transform

We begin by introducing some notation and background material. $\mathbb{R}^n$ consists of $n$-tuples of real numbers, usually designated by single letters, $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$, etc. The inner product and absolute value are defined by $\langle x, y \rangle = \sum_1^n x_i y_i$ and $|x| = \sqrt{\langle x, x \rangle}$. The unit sphere $S^{n-1}$ consists of the points with absolute value 1. $C_c^\infty(\mathbb{R}^n)$ denotes the set of infinitely differentiable functions on $\mathbb{R}^n$ with compact support. A continuous linear functional on $C_c^\infty$ is called a distribution. If $X$ is a set, $X^c$ denotes its interior, $\overline{X}$ its closure, and $X^c$ its complement. $\chi_X$ and $\chi_n$ denote the characteristic functions (indicator functions) of $X$, and of the unit ball in $\mathbb{R}^n$, respectively. I.e., $\chi_X(x) = 1$ if $x \in X$, and $\chi_X(x) = 0$ if $x \notin X$. $|X|$ denotes the $n$-dimensional Lebesgue measure of $X \subset \mathbb{R}^n$. However, when it is clear that $X$ should be treated as a set of dimension $m < n$, $|X|$ is the $m$-dimensional area measure. Thus

$$|S^{k-1}| = 2\pi^{k/2}/\Gamma(k/2)$$

is the $(k-1)$-dimensional area of the $(k-1)$-dimensional sphere.

The convolution of two functions is given by

$$f * g(x) = \int_{\mathbb{R}^n} f(x - y) g(y) dy.$$ 

The Fourier transform is defined by

$$\hat{f}(\xi) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} f(x) e^{-i<x,\xi>} dx$$

for integrable functions $f$, and is extended to larger classes of functions or distributions by continuity or duality.

The integral transform most relevant for local tomography is the x-ray transform.
Definition 2.1 Let $\theta \in S^{n-1}$ and $\Theta^\perp$ the hyperplane through the origin orthogonal to $\theta$. We parametrize a line $l(\theta, y)$ in $\mathbb{R}^n$ by specifying its direction $\theta \in S^{n-1}$ and the point $y$ where the line intersects the hyperplane $\Theta^\perp$.

The x-ray transform of a function $f \in L_1(\mathbb{R}^n)$ is given by

$$ Pf(\theta, y) = P_\theta f(y) = \int_{\mathbb{R}} f(y + t\theta) dt, \quad y \in \Theta^\perp. \quad (1) $$

We see that $P f(\theta, x)$ is the integral of $f$ over the line $l(\theta, y)$ parallel to $\theta$ which passes through $y \in \Theta^\perp$.

The inversion formula for the x-ray transform reads as follows:

$$ f(x) = (2\pi |S^{n-2}|)^{-1} \int_{S^{n-1}} \Lambda P_\theta f(y \Theta^\perp) d\theta, \quad (2) $$

where $E_\Theta x$ denotes the orthogonal projection of $x$ onto the subspace $\Theta^\perp$, and Calderón’s operator $\Lambda$ is defined in terms of Fourier transforms by

$$ \widehat{\Lambda g}(\xi) = |\xi| \hat{g}(\xi), \quad g \in C_0^\infty(\mathbb{R}^k). $$

It is extended by duality to the class of functions $g$ for which $(1 + |x|)^{-1-k}g$ is integrable [6]. Note that

$$ \Lambda^2 = -\Delta, \quad \Delta = \text{Laplacian}. \quad (3) $$

For other inversion formulas see [26, §II.2].

In two dimensions we parametrize $\theta \in S^1$ by its polar angle $\varphi$ and define a vector $\theta^\perp$ orthogonal to $\theta$ such that

$$ \theta = (\cos \varphi, \sin \varphi), \quad \theta^\perp = (-\sin \varphi, \cos \varphi). \quad (4) $$

Then the points in the subspace $\Theta^\perp$ are given by $\Theta^\perp = \{s \theta^\perp, \quad s \in \mathbb{R}\}$. When working in two dimensions, we will often use the simplified notation $P f(\theta, s)$ or $P_\theta f(s)$ instead of $P f(\theta, s \theta^\perp)$. Occasionally we will also replace $\theta$ by the polar angle $\varphi$ according to (4) and write $P f(\varphi, s)$.

For $g$ a function of one variable we have $\Lambda g = \mathcal{H} \partial g$, where $\partial g$ denotes the derivative of $g$ and $\mathcal{H}$ denotes the Hilbert transform

$$ \mathcal{H} g(s) = \frac{1}{\pi} \int_{\mathbb{R}} \frac{g(t)}{s - t} dt, \quad (5) $$

where the integral is understood as a principal value.

In dimension $n = 2$, i.e., when $f$ is a function of 2 variables, $P_\theta f$ is a function of one variable and the inversion formula (2) becomes

$$ f(x) = \frac{1}{4\pi^2} \int_0^{2\pi} \int_{\mathbb{R}} \frac{\partial P_\theta f(s)}{s < x, \theta^\perp > -s} ds d\varphi. \quad (6) $$
From equation (6) we see that computation of \( f(x) \) requires integrals over lines far from \( x \), because the Hilbert transform kernel has unbounded support. Note that \( P_\theta f(< x, \theta^\perp >) \) is the integral over the line with direction \( \theta \) which passes through \( x \). Hence the inversion formula is not “local”. A local inversion formula would utilize only integrals over lines passing close to \( x \), i.e., values \( P_\theta f(s) \) with \( s \) close to \( < x, \theta^\perp > \). For dimension \( n > 2 \) the inversion formula (2) is also not local.

The operator \( \Lambda \) is not continuous in an \( L_2 \) setting. Hence, in order to use the inversion formula in practice we have to stabilize it. This involves a well-known trade-off between stability and accuracy of the reconstruction. Here we give up the goal of recovering the function \( f \) itself, and aim instead at reconstructing an approximation \( e * f \), where \( e \) is an approximate delta function whose Fourier transform \( \hat{e}(\xi) \) decays sufficiently fast for large \( |\xi| \). The price to pay for the stabilization is limited resolution, so \( e \) must be chosen carefully, depending on the amount and accuracy of the available measurements.

In order to allow for local reconstruction formulas we reconstruct \( \Lambda^m f \) instead of \( f \), with \( m > -1 \) an integer. This yields the approximate inversion formula

\[
e * \Lambda^m f(x) = \int_{S^{n-1}} (k * P_\theta f)(E_{\Theta^\perp} x) d\theta, \quad m \geq -1,
\]

with the convolution kernel

\[
k(y) = (2\pi |S^{n-2}|)^{-1} \Lambda^{m+1} P_\theta e(y), \quad y \in \Theta^\perp.
\]

If \( e \) is a radial function, then \( P_\theta e \) and the convolution kernel \( k \) are independent of \( \theta \).

Of greatest interest are the case \( m = 0 \), which gives the formulas for reconstructing the function \( f \) itself, and the cases \( m = \pm 1 \) which give local reconstruction formulas.

Since the parameters \( \theta \) and \( y \in \Theta^\perp \) of a line passing through a point \( x \) must satisfy the equation \( E_{\Theta^\perp} x = y \), reconstruction according to (7) will be local if the kernel \( k \) is supported in a small neighborhood of the origin. However, for \( m \) even and \( \int_{\mathbb{R}^n} e(x) dx \neq 0 \), \( k \) is not analytic, so \( k \) cannot have compact support. This again reflects the fact that ordinary tomography is global, not local. On the other hand, it follows from (8) and (3) that \( k \) does have compact support if \( m \geq -1 \) is odd and \( e \) has compact support. This explains the interest in the cases \( m = \pm 1 \). Computing \( \Lambda^{-1} f(x) \) consists of taking the average of all integrals over lines passing through \( x \). This was done in early imaging techniques preceding CT. However, since \( \Lambda^{-1} \), the inverse of \( \Lambda \), is given by convolution with the Riesz kernel \( R_1 \),

\[
\Lambda^{-1} f = R_1 * f, \quad R_1(x) = (\pi |S^{n-2}|)^{-1} |x|^{1-n},
\]

the result is a very blurry image of \( f \) which by itself is of limited usefulness; see the bottom left image in Fig. 1. Current Lambda tomography avoids this disadvantage by computing a linear combination of \( \Lambda f \) and \( \Lambda^{-1} f \).
3 Lambda tomography

Lambda tomography was introduced independently in [41] and [38], further developed in work including [6, 7, 8, 20, 31, 42], and generalized in [18, 22]. It does not attempt to reconstruct the function \( f \) itself but instead produces the related function \( Lf = \Lambda f + \mu \Lambda^{-1} f \). This has the advantage that the reconstruction is strictly local in the sense that computation of \( Lf(x) \) requires only integrals over lines passing arbitrarily close to \( x \). Lambda tomography has found applications in medical imaging [39], nondestructive testing [34, 42], and microtomography [7, 8, 33, 35]. Reconstructions from efficiently sampled data are analyzed in [8].

Intelligent use of Lambda tomography requires knowledge of what kind of useful information about \( f \) is retained in \( Lf \). Let us consider an example. The upper left of Fig. 1 shows an ordinary, global reconstruction of the density function \( f \) of a calibration object used by the Siemens company. The data come from an old generation Siemens hospital scanner. Units are such that the radius of the global reconstruction circle is one. The figure displays the reconstruction inside the rectangle \([-0.5, 0.5] \times [0, 0.5]\). The scanning geometry is a fan-beam geometry with source radius \( R = 2.868 \), \( p = 720 \) source positions, and \( 2q = 512 \) rays per source; cf. [26, p. 75]. The upper right of Fig. 1 shows a reconstruction of \( \Lambda f \). Reconstructions of \( \Lambda^{-1} f \) and \( Lf = \Lambda f + 46 \Lambda^{-1} f \) are shown in the lower left and lower right, respectively. The similarity between the images of \( f \) and \( \Lambda f \) is at first glance surprising. We expect that a good local reconstruction method should detect the singularities of \( f \), since these are stably determined by the data. Indeed, since \( \Lambda \) is an invertible elliptic pseudo-differential operator, \( f \) and \( \Lambda f \) have precisely the same singular set. However, we see that \( \Lambda f \) is cupped where \( f \) is constant, and that the singularities are amplified in \( \Lambda f \). The image of \( \Lambda^{-1} f \) by itself seems less useful, but it provides a counterpoint for the cup in \( \Lambda f \). Thus, the image of \( Lf \) shows less cupping and looks even more similar to \( f \) than the image of \( \Lambda f \). For example, the image of \( Lf \) indicates that the density just inside the boundary of the object is larger than the density outside the object, while this can not be clearly seen from the image of \( \Lambda f \). To achieve this effect, a good selection of \( \mu \) is necessary. Here \( \mu = 46 \) was chosen by trial and error. A prescription for selecting \( \mu \) can be found in [7].

A more detailed understanding of images of \( \Lambda f \) or \( Lf \) is obtained from studying quantitative relations between \( \Lambda f \), \( \Lambda^{-1} f \) and \( f \) [6, 7]. Some of the results for \( \Lambda f \) are stated in Theorem 3.1 and discussed in Remark 3.2 below. For corresponding results on \( \Lambda^{-1} \) see [6].

**Theorem 3.1** ([6]) Let \( X \) and \( Y \) be measurable subsets of \( \mathbb{R}^n \), \( n \geq 2 \), and let \((1 + |x|)^{-1-n} f \) be integrable.

(a) If \( f_r(x) = f(x/r) \), then \( \Lambda f_r(x) = r^{-1} \Lambda f(x/r) \).

(b) \( \Lambda \chi_X(x) > 0 \) on \( X^c \), and \( < 0 \) on \( X^c \); \( \Lambda \chi_X = -\Lambda \chi_X \).

(c) \( \Lambda \chi_X \) is subharmonic (Laplacian \( \geq 0 \)) on \( X^c \), and superharmonic on \( X^c \). This implies that \( \Lambda \chi_X \) cannot have a local maximum in \( X^c \), nor a local minimum in \( X^c \).
(d) If $x$ is outside the support of $f$, then
\[ \Lambda f(x) = \frac{1 - n}{\pi |S^{n-2}|} \int_{\mathbb{R}^n} |x - y|^{-1-n} f(y) \, dy. \]

(e) Near $\partial X$, $|\Lambda \chi_X(y)| \sim \frac{1}{d(x, \partial X)}$, where $d(x, \partial X)$ denotes the distance of $x$ to $\partial X$.

**Remark 3.2** The results for $\Lambda \chi_X$ are of practical interest, since in many applications the function $f$ can be modeled as a linear combination of characteristic functions.

- As a consequence of (a), small features are amplified in images of $\Lambda f$. This is beneficial for the detection of small, low contrast details. For example, in Fig. 1 the small holes in the rectangular pieces are more clearly visible in the image of $\Lambda f$ than in the image of $f$.

- Part (b) indicates that the jumps of $\Lambda f$ at discontinuities of $f$ have the same direction as those of $f$.

- Part (c) explains why there are no oscillations which could be mistaken for actual details in images of $\Lambda f$.

- Part (d) shows that if $f$ has compact support, then $\Lambda f$ cannot. This means that there are global effects in images of $\Lambda f$ in the sense that the value of $\Lambda f(x_0)$ depends on the values of $f$ everywhere. However, Part d) implies that $\Lambda f(x)$ will decay at least as $O(|x|^{-1-n})$ for $|x| \to \infty$. More refined estimates are derived in [7] and used to develop a procedure to reduce the global effects.

- Part (e) shows that a finite jump in $f$ causes an infinite jump in $\Lambda f$. In a neighborhood of $\partial X$, $\Lambda f$ is not a function but a principal value distribution [6].

While $\Lambda f$ retains the signs of jumps in density, it does not give direct information about the size of these jumps. However, such information about density differences may be extracted in certain cases. In this and the following sections we will describe several methods. We assume that $f$ is a linear combination of a smooth function and of characteristic functions of sets:

\[ f = f_0 + \sum c_i \chi_{X_i}, \quad f_0 \in C_0^\infty, \quad |\partial X_i| = 0, \quad X_i = \overline{X_i^\circ}, \quad X_i^\circ \cap X_j^\circ = \emptyset \text{ if } i \neq j. \]  

(10)

We are interested in estimating $c_j - c_i$ when $X_j$, $X_i$ have a common nontrivial boundary $\Gamma$,

\[ \Gamma = \partial X_i \cap \partial X_j \cap W \neq \emptyset, \quad W = (X_i \cup X_j)^\circ. \]  

(11)
We first discuss the method developed in [7]. It is based on Theorem 3.3 below. The theorem expresses the fact that for \( x \) sufficiently close to \( \Gamma \),

\[
c_j - c_i = \frac{\Lambda f(x)}{\Lambda \chi_{x_j}(x)} + O(d), \quad \text{and}
\]

\[
|c_j - c_i| = \frac{|\nabla \Lambda f(x)|}{|\nabla \Lambda \chi_{x_j}(x)|} + O(d^2),
\]

where \( d \) is the distance from \( x \) to \( \Gamma \).

We say that a set \( Y \) has curvature \( \leq 1/r \) along a subset \( Y_0 \) of \( \partial Y \) if for each point \( \tilde{y} \in Y_0 \) there are open balls \( B \subset Y \) and \( B' \subset Y^c \) of radius \( r \) with \( \tilde{y} \in \overline{B} \cap \overline{B'} \). The distance of a point \( x \) to a set \( Y \) is denoted by \( d(x, Y) \).

**Theorem 3.3** ([7]) *Let \( f \) be as in (10). Fix \( i, j \), let \( W = (X_i \cup X_j)^\circ \) and assume that \( \Gamma = \partial X_i \cap \partial X_j \cap W \neq \emptyset \).

Let \( X_j \) have curvature \( \leq 1/r, \ r > 0 \), along a closed subset \( \Gamma_0 \) of \( \Gamma \). Let \( x \in W \setminus \Gamma \) be such that \( d(x, \partial X_j) = d(x, \Gamma_0) = d \). Then

\[
\left| \frac{\Lambda f(x)}{\Lambda \chi_{x_j}(x)} - (c_j - c_i) \right| \leq F_1(d/r) \left( \max |\Lambda f_0| + C_1 \frac{\left( \max_{k \neq j} |c_k| \right)}{d(x, \partial W)} \right) d \tag{12}
\]

\[
\left| \frac{|\nabla \Lambda f(x)|}{|\nabla \Lambda \chi_{x_j}(x)|} - |c_j - c_i| \right| \leq F_2(d/r) \left( \max |\nabla \Lambda f_0| + C_2 \frac{\left( \max_{k \neq j} |c_k| \right)}{d(x, \partial W)^2} \right) d^2 \tag{13}
\]

The constants \( C_1 \) and \( C_2 \) and the functions \( F_1, F_2 \) can be given explicitly. E.g., for \( n = 2 \), \( C_1 = 2 \), and \( C_2 = 3 \). Furthermore,

\[
\lim_{t \to 0^+} F_1(t) = \lim_{t \to 0^+} F_2(t) = \pi.
\]

The error terms on the right-hand sides of (12) and (13) indicate that in general the estimate (13) should be more accurate than (12) when \( d \) is small. The terms involving
$d(x, \partial W)$ come from the influence of other boundaries than $\Gamma$ and reflect the global effects mentioned above.

Numerical implementation of (12) or (13) requires computation of reconstructions of $\Lambda f$ and $\Lambda \chi_{X_j}$ inside a region of interest $R$. In the following let $\hat{\Lambda} f$ and $\hat{\Lambda} \chi_{X_j}$ denote these reconstructions, rather than the functions $\Lambda f$ and $\Lambda \chi_{X_j}$ themselves. It is also assumed that $f$ has the form (10) with sets $X_i$ such that $X_i \subset \hat{R}$ or $X_i \cap R = \emptyset$. This entails no loss of generality since any set $X_i$ violating this condition can be replaced by the two sets $X_i \cap R$ and $X_i \cap \hat{R}$. $\hat{\Lambda} \chi_{X_j}$ is computed using simulated x-ray data, after $\partial X_j$ has been found from $\hat{\Lambda} f$. In principle, either (12) or (13) can be used, but as mentioned above the method based on (13) is likely to be more accurate. This gives only $|c_j - c_i|$, but since the sign of $c_j - c_i$ is preserved in $\Lambda f$, this is all that is needed.

The method consists of the following steps:

1. Compute $\hat{\Lambda} f$ from local data inside a region of interest $R$.

2. Determine $X_j$ by finding $\partial X_j$ from $\hat{\Lambda} f$.

3. Compute $\hat{\Lambda} \chi_{X_j}$ inside the region of interest from simulated x-ray data, using the same sampling geometry as for the original data.

4. If $x \in \partial X_j$, take the ratio $|\nabla \hat{\Lambda} f(x)|/|\nabla \hat{\Lambda} \chi_{X_j}(x)|$ as an estimate for the magnitude of the density jump. It is advisable to use suitable averages of the gradients over points near the boundary of $X_j$ instead of the gradient at a single point $x$. This reduces effects due to measurement noise.

A detailed discussion of the implementation of this method and numerical tests using real-world data have been reported in [7]; see also [5, 33].

The method described above can be simplified by making a priori assumptions about the unknown boundary $\partial X_j$. This can be used to simplify the edge detection in step 2 and to avoid the reconstruction from simulated data in step 3. For example, $X_j$ could be assumed to be a halfspace $H$. If the filtered backprojection algorithm, i.e., a discretization of the approximate inversion formula (7) is used, then the reconstruction $\hat{\Lambda} f$ will, apart from discretization errors, be equal to $\epsilon \ast \Lambda f$. Hence $\Lambda f$ and $\Lambda \chi_{X_j}$ in (12) and (13) can be replaced by $\epsilon \ast \Lambda f$ and $\epsilon \ast \Lambda \chi_H$, respectively. We can compute $\epsilon \ast \Lambda \chi_H$ analytically in the following way: For $x \notin \partial H$ one has ([6, Theorem 4.5])

$$\Lambda \chi_H(x) = (\pi \hat{d}(x))^{-1},$$

where $\hat{d}(x)$ is the signed distance of $x$ from $\partial H$, i.e., $\hat{d}(x) = d(x, \partial H)$ for $x \in H$, and $\hat{d}(x) = -d(x, \partial H)$ for $x \notin H$. Computing $\epsilon \ast \Lambda \chi_H$ involves the Radon transform of $\epsilon$. It is given by

$$R_{\theta} \epsilon(s) = \int_{\Theta} \epsilon(s \theta + y) dy, \quad \theta \in S^{n-1}, \quad s \in \mathbb{R}.$$
We assume that $e$ is radial, so that $R_\theta e$ does not depend on $\theta$. Therefore the subscript $\theta$ will be suppressed and $Re(s)$ viewed as a function of the one variable $s$. It now follows that
\begin{equation}
  e \ast \Lambda \chi_H(x) = \mathcal{H}Re(\hat{d}(x)),
\end{equation}
where $\mathcal{H}$ denotes the Hilbert transform as defined in (5). Recalling that for functions $g$ of one variable $\Lambda g(t) = \frac{d}{dt}\mathcal{H}g(t)$ gives
\begin{equation}
  |\nabla(e \ast \Lambda \chi_H(x))| = |\Lambda Re(\hat{d}(x))|.
\end{equation}

Replacing $\Lambda f$ and $\Lambda \chi_X$ in (12) and (13) by $e \ast \Lambda f$ and $e \ast \Lambda \chi_H$, and using (14) and (15) gives the approximate formulas
\begin{equation}
  c_j - c_i \simeq \frac{e \ast \Lambda f(x)}{\mathcal{H}Re(\hat{d}(x))},
\end{equation}
\begin{equation}
  |c_j - c_i| \simeq \frac{|\nabla(e \ast \Lambda f(x))|}{|\Lambda Re(\hat{d}(x))|}.
\end{equation}

These two formulas are the basis of two of the algorithms proposed in [20, 31] for dimension $n = 2$, cf. formulas (2.17) and (2.21) in [20]. The derivation in [20, 31] is different and employs an asymptotic expansion for $\Lambda f$, where $f$ is smooth except for jumps across smooth boundaries. An algorithm based on (17) given in [31] uses the fact that $|\nabla(e \ast \Lambda f(x))|$ will be maximal for $x \in \Gamma$, and that $\hat{d}(x) = 0$ for $x \in \Gamma$. Hence one can find the points $x \in \Gamma$ by looking for the local maxima of $|\nabla \Lambda f|$ and then estimate the jump by
\begin{equation}
  |c_j - c_i| \simeq \frac{|\nabla \Lambda f(x)|}{|\Lambda Re(0)|}.
\end{equation}

In our numerical experiments this algorithm tended to be somewhat less accurate than the more elaborate method of [7].

4 Pseudolocal tomography

Another method to compute jumps of a function from essentially local data is pseudolocal tomography. It was introduced in [21] and further developed in [31]. Here we follow the presentation given in [3] which allows us to understand the numerical implementation of this method in the framework of (16) and (17).

The starting point for pseudolocal tomography is the two-dimensional inversion formula (6) which we repeat here:
\begin{align}
f(x) &= \frac{1}{4\pi} \int_{S^1} \mathcal{H}\partial \theta f(<x, \theta^\perp>)d\theta \\
&= \frac{1}{4\pi^2} \int_0^{2\pi} \int_{\mathbb{R}} \frac{4\pi P_\theta f(s)}{<x, \theta^\perp > -s} ds \, d\varphi.
\end{align}
Now truncate the Hilbert transform integral and define
\[
    f_d(x) = \frac{1}{4\pi^2} \int_0^{2\pi} \int_{\left< x, \theta^+ > + d}^{\left< x, \theta^+ > - d} \frac{d}{ds} P_{\theta} f(s) \ ds \ d\varphi.
\]  

It was shown in [21] that \( f - f_d \) is continuous, hence \( f_d \) has the same jumps as \( f \). Recalling that \( P_{\theta} f(\left< x, \theta^+ > \right) \) is the integral over the line in direction \( \theta \) which passes through \( x \), we see that computation of \( f_d(x) \) requires only integrals over lines with distance at most \( d \) from \( x \) ("pseudo-local" reconstruction.)

In practice one has to use an approximate inversion formula and computes
\[
    f_{d,r}(x) = e_r * f_d(x) = \int_0^{2\pi} \int_{\mathbb{R}} \tilde{k}_{d,r}(\left< x, \theta^\perp > - s) P_{\theta} f(s) \ ds \ d\varphi,
\]
\[
    \tilde{k}_{d,r}(t) = \frac{1}{4\pi^2} \int_{t-d}^{t+d} \frac{d}{ds} P_{\theta} e_r(s) \ ds,
\]

where \( e_r \) is a radial function satisfying
\[
    e_r(x) = r^{-2} e_1(x/r), \quad e_1(x) = 0 \text{ for } |x| > 1, \quad \int_{\mathbb{R}^2} e_1 \ dx = 1.
\]

Note that \( \tilde{k}_{d,r}(t) = 0 \) for \( |t| > d + r \), i.e., computation of \( f_{d,r}(x) \) requires integrals over lines with distance at most \( d + r \) from \( x \). Furthermore, \( \lim_{d \to \infty} \tilde{k}_{d,r}(t) = (4\pi)^{-1} \mathcal{H} P_{\theta} e_r(t) \). Hence (7) gives that \( \lim_{d \to \infty} f_{d,r}(x) = e_r * f(x) \). Indeed, the convolution kernel \( \tilde{k}_{d,r} \) can be obtained from the kernel \( k \) in (8) by letting \( m = 0 \) and truncating the Hilbert transform integral. The relation \( f_{d,r} = e_r * f_d \) was shown in [21].

It turns out that for small \( d \) (i.e., local data), \( f_d \) is significantly different from zero only in a narrow region near a boundary (cf. [21, Fig. 3]), and that the convolution with the point spread function \( e_r \) alters these values so much that the jumps cannot just be simply read off the reconstructed image \( f_{d,r} \). We need an algorithm to obtain information about the jumps of \( f \). The methods developed by Katsevich and Ramm [21, 31] can be understood in the framework developed for Lambda tomography. According to (16) and (17) we have for \( x \) close to \( \Gamma \)
\[
    c_j - c_i \simeq \frac{E * \Lambda f(x)}{\mathcal{H}E(d(x))} \quad \text{(20)}
\]
\[
    |c_j - c_i| \simeq \frac{|\nabla E * \Lambda f(x)|}{|\mathcal{A}E(d(x))|} \quad \text{(21)}
\]

The task now is to find \( E_{d,r} \) such that \( E_{d,r} * \Lambda f = f_{d,r} = e_r * f_d \).
Proposition 4.1 ([31, 3]) Define $E_{d,r}$ by

$$P_0 E_{d,r} = (P_0 e_r) * M_d$$

with

$$M_d(s) = -\frac{1}{\pi} \ln(|s/d|) \chi_{[-d,d]}(s).$$

Then

$$f_{d,r}(x) = E_{d,r} * \Lambda f(x).$$

With this result (20) and (21) give

$$c_j - c_i \simeq \frac{f_{d,r}(x)}{\mathcal{H}RE_{d,r}(d(x))} \quad (22)$$

$$|c_j - c_i| \simeq \frac{\left|\nabla f_{d,r}(x)\right|}{\left|\Lambda RE_{d,r}(d(x))\right|} \quad (23)$$

and we can apply the same algorithms for recovering the jumps as in Lambda tomography. Some remarks are in order.

1. Note that because $E_{d,r}$ is radial, $\mathcal{H}RE_{d,r}(0) = 0$, so $f_{d,r}(x) \simeq 0$ for $x \in \Gamma$. This makes it difficult to use the relation (22) in practice, since finding $\hat{d}(x)$ is not easy, cf. the algorithm given [21] and further discussed in [3]. However, since $|\nabla f_{d,r}|$ is maximal for $x \in \Gamma$ one can find the points $x \in \Gamma$ by looking for the local maxima of $|\nabla f_{d,r}|$ and then estimate the jump by

$$|c_j - c_i| \simeq \frac{|\nabla f_{d,r}(x)|}{|\Lambda RE_{d,r}(0)|}, \quad x \in \Gamma.$$ This approach has essentially been used in [31] for pseudolocal tomography and in [20] for Lambda tomography.

2. The property that $f_d$ has the same jumps as $f$ is not used in the algorithm.

3. $E_{d,r}(x) = 0$ for $|x| > d + r$. Hence our derivation of the algorithm is only justified for $d + r$ sufficiently small. In practice the method seems to work also for much larger values of $d + r$.

5 Wavelet-based multiresolution local tomography

Wavelet-based multiresolution local tomography is a method for region of interest tomography developed in [32]. The goal here is to reconstruct the function $f$ itself within the
region of interest up to an almost constant error. The method illustrates the possible uses of wavelets to 'localize' the x-ray transform, or, more precisely, to separate the features which are well determined by local data from those who are not. The following discussion assumes some background on wavelets which can be found in [43] or other texts on this subject.

Consider a (two-dimensional) multiresolution analysis of nested subspaces $V_j, j \in \mathbb{Z}$ of $L_2(\mathbb{R}^2)$. We use the notation

$$f_{j,k}(x) = 2^j f(2^j x - k), \quad j \in \mathbb{Z}, \quad k \in \mathbb{Z}^2, \quad x \in \mathbb{R}^2.$$ 

Let $\Phi$ be the scaling function and $\Psi^\mu, \mu = 1, 2, 3$ the associated wavelets. Since the $\Phi_{j+1,k}, k \in \mathbb{Z}^2$ are a Riesz basis of the subspace $V_{j+1}$, a function $f \in V_{j+1}$ can be written as

$$f(x) = \sum_{k \in \mathbb{Z}^2} \tilde{A}_{j+1,k} \Phi_{j+1,k}(x).$$

The so-called approximation coefficients $\tilde{A}_{j,k}$ are given by

$$\tilde{A}_{j,k} = \langle f, \tilde{\Phi}_{j,k} \rangle$$

where $\langle, \rangle$ denotes the inner product in $L_2$ and $\tilde{\Phi}$ is the biorthogonal scaling function. Alternatively we can use the relation $V_{j+1} = V_j + W_j$ and obtain the expansion

$$f(x) = \sum_{k \in \mathbb{Z}^2} \tilde{A}_{j,k} \Phi_{j,k}(x) + \sum_{\mu=1}^3 \sum_{k \in \mathbb{Z}^2} \tilde{D}_{j,k}^{\mu} \Psi_{j,k}^\mu(x).$$

We can interpret the first sum as an approximation to $f$ in $V_j \subset V_{j+1}$, i.e., at a lower resolution. The second sum supplies the missing detail information. Therefore the coefficients

$$\tilde{D}_{j,k}^{\mu} = \langle f, \tilde{\Psi}_{j,k}^\mu \rangle$$

are called detail coefficients. The Fast Wavelet Transform and its inverse allow efficient computation of the $\tilde{A}_{j,k}$ and $\tilde{D}_{j,k}^{\mu}, k \in \mathbb{Z}^2$ from the $\tilde{A}_{j+1,k}, k \in \mathbb{Z}^2$, and vice versa.

We now observe that the approximation and detail coefficients can be computed directly from the x-ray data. Let $f^\vee(x) = f(-x)$. Then

$$\tilde{A}_{j,k} = \langle f, \tilde{\Phi}_{j,k} \rangle = (f \star \overline{\Phi_{j,0}^\vee})(2^{-j}k)$$ \hspace{1cm} (24)

Similarly,

$$\tilde{D}_{j,k}^{\mu} = \langle f, \tilde{\Psi}_{j,k}^\mu \rangle = (f \star \overline{\Psi_{j,0}^\mu})(2^{-j}k)$$ \hspace{1cm} (25)

Hence we can use the approximate inversion formula (7) with $e(x) = \overline{\Phi_{j,0}^\vee}(x)$ and reconstruction on the grid $x = 2^{-j}k, k \in \mathbb{Z}^2$, to obtain the approximation coefficients directly from the x-ray data. For the detail coefficients we let $e = \overline{\Psi_{j,0}^\vee}$. Alternatively one
could first compute the approximation coefficients \( \tilde{A}_{j+1,k} \) by letting \( \epsilon(x) = \phi_{j+1,0}(x) \) and choosing the finer grid \( x = 2^{-j-1}k, \ k \in \mathbb{Z}^2 \), and then use the Fast Wavelet Transform to obtain the approximation and detail coefficients at level \( j \). Since the additional computational burden of applying the Fast Wavelet Transform is negligible compared to the effort required for the reconstruction from the x-ray data, this alternative method seems preferable, since only one point-spread function and corresponding convolution kernel need to be used. However, if not all coefficients on level \( j \) are needed, the first method will be more efficient.

The next question is how this approach allows to ‘localize’ the x-ray transform, i.e., to separate features which are determined by local data from those which are not. It was observed in \([28]\) that the detail coefficients for sufficiently large \( j \) should be well determined by local data, if the wavelets \( \Psi^u \) have vanishing moments. Let us see why.

**Definition 5.1** A function \( f \) of \( n \) variables has vanishing moments of order up to \( N \), if

\[
\int_{\mathbb{R}^n} x^\alpha f(x) \, dx = 0
\]

for all multiindices \( \alpha = (\alpha_1, \ldots, \alpha_n) \) with \( |\alpha| = \sum \alpha_i \leq N \). Recall that the \( \alpha_i \) are non-negative integers and that \( x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_n^{\alpha_n} \).

The nonlocality in the approximate inversion formula comes from the convolution kernel \( k \) in (8) in case of \( m = 0 \). In two dimensions this is caused by the presence of the Hilbert transform in the formula \( k = (4\pi)^{-1} \Lambda P_\theta \epsilon = (4\pi)^{-1} \mathcal{H} \partial P_\theta \epsilon \). The key observation now is that the Hilbert transform of a function with vanishing moments decays fast.

**Lemma 5.2** ([32, p. 1418]) Let \( f(t) \in L_2(\mathbb{R}) \) vanish for \( |t| > A \) and have vanishing moments of order up to \( N \). Then, for \( |s| > A \),

\[
|\mathcal{H}f(s)| \leq \frac{1}{\pi|s-A|^{N+2}} \int_{-A}^{A} |f(t)t^{N+1}| dt
\]

It is well known how to construct wavelets with vanishing moments, and it turns out that the functions \( \partial P_\theta \tilde{\Psi}^u, v \) inherit the vanishing moments from the \( \tilde{\Psi}^u \). Therefore the convolution kernels \( k = (4\pi)^{-1} \mathcal{H} \partial P_\theta \tilde{\Psi}^u, v \) will decay rapidly outside the support of \( P_\theta \tilde{\Psi}^u, v \).

So we see that the detail coefficients for large \( j \), when \( \tilde{\Psi}^u, v \) has small support, are well determined by local data. This is intuitively plausible since these coefficients contain high-frequency information, and we know already from Lambda tomography that high-frequency information is well-determined. So the nonlocality shows its greatest impact in the approximation coefficients. Since the scaling function satisfies \( \int \Phi(x) \, dx = 1 \), its zero order moment does not vanish. One could still choose \( \Phi \) so that the moments of order 1 through \( N \) vanish. It is shown in [32, p. 1419] that in such a case the resulting convolution kernel \( k \) satisfies

\[
|k(s)| = O(s^{-2}) + O(s^{-N-3}).
\]
It seems that this does not achieve much, since we cannot remove the leading $O(s^{-2})$ term. Nevertheless, the authors of [32] found that some scaling functions having vanishing moments lead to convolution kernels with sufficiently rapid decay for practical purposes. In their reconstructions the authors of [32] also extrapolated the missing data by constant values, thus reducing cupping artifacts. While it is suggested in [32] to first compute the approximation and detail coefficients at level $j$ and then use an inverse fast wavelet transform to obtain the approximation coefficients at level $j+1$, numerical tests in [36] indicated that the simpler approach of directly computing the approximation coefficients at level $j+1$ yields equivalent results. We observe that this can be done without using wavelet theory, namely just by specifying the particular point spread function $e = \Phi_{j+1,0}$ in (7).

6 Cone-beam local tomography with sources on a curve

A problem of great practical interest which still poses many open problems is three-dimensional cone-beam reconstruction with sources on a curve. See, e.g., [40] for an inversion formula, [10] for a general stability result, [30] for conditions to detect singularities, and [4, 9, 11, 24, 27] for reconstruction algorithms and other developments.

To describe data collection with an x-ray source moving on a curve, the parameterization of lines by $\theta \in S^{n-1}$ and $y \in \Theta^+$ is less convenient. It is more suitable to introduce the divergent beam x-ray transform

$$Df(a, \theta) = D_a f(\theta) = \int_0^\infty f(a + t\theta)dt, \quad \theta \in S^{n-1},$$

which gives the integral of $f$ over the ray with direction $\theta$ emanating from the source point $a$. If $f$ is supported in the unit ball, and the source points $a$ lie on a sphere $A$ with center in the origin and radius $R > 1$, then the approximate inversion formula for the divergent beam x-ray transform reads [38]

$$e \ast \Lambda^m f(x) = R^{-1} \int_A \int_{S^{n-1}} D_a f(\theta) \mid < a, \theta > \mid k(E_{\Theta^+}(x - a)) d\theta da,$$

with $m \geq -1$ and $k$ as in (8). This formula is very useful in two dimensions, but not so in three dimensions. It needs integrals over all lines, but in three dimensions the lines form a four parameter family, so (27) requires far more data than should be needed to determine a function of three variables. In practical 3D tomography an x-ray source moves on a curve, so only integrals over lines intersecting the curve are measured. The conditions on the source curve $\Gamma$ for stable inversion are restrictive, so that in most practical situations one has an incomplete data problem.

Microlocal analysis has proved to be a useful tool in determining which singularities of $f$ are stably determined by the available data. Based on the exposition in [30] we
now state the relevant microlocal concepts and apply them to this situation. The reader interested in a deeper treatment may wish to first read [30] and [15], and then proceed to articles such as [1, 12, 13, 14, 29].

The following concept of a wavefront set uses the fact that the Fourier transform of a $C^\infty_0$ function decays rapidly. A local version of this fact can be obtained by first multiplying $f$ with a $C^\infty_0$ cut-off function $\Phi$ with small support, and seeing if the Fourier transform of the product $\Phi f$ decays rapidly. The wavefront set gives even more specific, so-called microlocal information, inasmuch as it identifies the directions in which the Fourier transform of $\Phi f$ does not decrease rapidly.

**Definition 6.1** Let $f$ be a distribution and let $x_0, \xi_0 \in \mathbb{R}^n$, $\xi_0 \neq 0$. Then $(x_0, \xi_0)$ is in the wavefront set of $f$ if and only if for each cut-off function $\Phi$ in $C^\infty_0$ with $\Phi(x_0) \neq 0$, the Fourier transform of $\Phi f$ does not decrease rapidly in any conic neighborhood of the ray $\{t\xi_0, t > 0\}$.

Loosely speaking, we say that a singularity of $f$ can be stably detected from available x-ray data, if there exists a corresponding singularity of comparable strength in the data. The strength of a singularity can be quantified microlocally using Sobolev space concepts:

**Definition 6.2** A distribution $f$ is in the Sobolev space $H^s$ microlocally near $(x_0, \xi_0)$ if and only if there is a cut-off function $\Phi \in C^\infty_0(\mathbb{R}^n)$ with $\Phi(x_0) \neq 0$ and function $u(\xi)$ homogeneous of degree zero and smooth on $\mathbb{R}^n \setminus \{0\}$ and with $u(\xi_0) \neq 0$ such that $u(\xi)(\overline{\Phi f})(\xi) \in L^2(\mathbb{R}^n, (1 + |\xi|^2)^s)$.

First, one localizes near $x_0$ by multiplying $f$ by $\Phi$, then one microlocalizes near $\xi_0$ by forming $u\overline{\Phi f}$ and sees how rapidly $\overline{\Phi f}$ decays at infinity.

For 3D tomography with sources on a curve we have the following result:

**Theorem 6.3** (cf. [30, Theorem 4.1], and [1, 12]) Let $\Gamma$ be a smooth curve in $\mathbb{R}^3$ and $f$ a distribution whose support is compact and disjoint from $\Gamma$. Then any wavefront set of $f$ at $(x_0, \xi_0)$ is stably detected from divergent beam x-ray data $Df$ with sources on $\Gamma$ if and only if

the plane $P$ through $x_0$ and orthogonal to $\xi_0$, intersects $\Gamma$ transversally.

If data are taken over an open set of rays with sources on $\Gamma$, then a ray in $P$ from $\Gamma$ to $x_0$ must be in the data set for stable detection to apply. In these cases $f$ is in $H^s$ microlocally near $(x_0, \xi_0)$ if and only if the corresponding singularity of $Df$ is in $H^{s+1/2}$.

We see that the corresponding singularities of $Df$ are weaker by $1/2$ Sobolev order, but this is still strong enough to allow stable detection in practice.

It is now interesting to ask if the available numerical algorithms can actually reconstruct all the stable singularities. The results for a general class of restricted x-ray transforms obtained in [12, 13, 14] show that microlocal analysis is also a powerful tool
to answer such a question. For an introduction to these results see [15]. Explicit calculations analysing an algorithm for contour reconstruction proposed by Louis and Maass in [24] and some closely related methods have recently been given in [19, 23].

The algorithm of [24] aims to reconstruct the function

$$f_R = -\Delta D^* D f,$$

with

$$D^* g(x) = \int_{\Gamma} \left\| x - a \right\|^{-1} g \left( a, \frac{x - a}{\left\| x - a \right\|} \right) \, da.$$  

An advantage of the formula (28) is that reconstruction of $f_R$ is local. In [24] it is shown that $f_R$ approximates $\Lambda f$ in certain cases.

The results in [12, 19, 23] show that the wavefront set of $f_R$ consists of two parts. The first part contains those wavefronts $(x, \xi)$ of $f$ for which the plane through $x$ and normal to $\xi$ intersects \( \Gamma \). The second part may introduce new singularities, namely on the line from a source point $a \in \Gamma$ to $x$, the location of the original singularity in $f$. This will happen if the plane through $x$ and normal to $\xi$ contains $a$ and the tangent vector to \( \Gamma \) at $a$ is orthogonal to $\xi$, i.e., the plane touches \( \Gamma \) but does not intersect \( \Gamma \) transversally. In addition, the acceleration vector of the curve at $a$ should not be orthogonal to $\xi$.

The Sobolev strength of these additional singularities is the same as the reconstructed part of the original wavefront set [13, 14, 19], and they appear as artifacts in numerical simulations [17, 19].

Another, and apparently the historically first method for 3D local tomography is an adaptation of the algorithm by Feldkamp, Davis and Kress [9] (FDK algorithm) which was developed by P.J. Thomas at the Mayo Clinic. While the details of this local FDK algorithm have not been published, it has been used in various papers, e.g., [39, 6]. A recent implementation of a local FDK algorithm has been reported in [16]. Figures 2 and 3 provide a comparison of the two algorithms, using the implementations in [17] and [16], respectively. The experiments use a mathematical phantom consisting of a superposition of four balls with the following parameters:

<table>
<thead>
<tr>
<th>Center</th>
<th>Radius</th>
<th>Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,0,0)</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>(0,0,0.125)</td>
<td>0.1</td>
<td>-1</td>
</tr>
<tr>
<td>(-0.3, 0, -0.125)</td>
<td>0.02</td>
<td>-1</td>
</tr>
<tr>
<td>(0.3, 0, 0.2)</td>
<td>0.01</td>
<td>-1</td>
</tr>
</tbody>
</table>

The source is assumed to move on a circle in the $x$-$y$ plane with radius $R = 3$ and center in the origin. We used 400 equidistant source positions and a $240 \times 240$ detector array. The local FDK algorithm [16] used a planar detector array, while our code for the Louis-Maass algorithm assumed a spherical array; see [17]. The images consist of $131 \times 131$ pixels.
Figure 3 shows reconstructions in the vertical plane $y = 0$, with the Louis-Mass method in the upper left, and the local FDK algorithm in the upper right. The additional singularities predicted for the Louis-Mass algorithm by the references given above are clearly visible in the upper left image as lines tangential to each ball and intersecting the source curve. The reconstruction with the local FDK algorithm does not show these lines, but does appear to have greater distortions with regard to the small objects. This is confirmed in Figure 4 which shows reconstructions of the horizontal plane $z = 0.2$. 
References


Figure 1: Top left: Global reconstruction of density $f(x)$ of calibration object. Top right: Reconstruction of $\Lambda f$. Bottom left: Reconstruction of $\Lambda^{-1}f$. Bottom right: Reconstruction of $L f = \Lambda f + \mu \Lambda^{-1}f$, $\mu = 46$. 
Figure 2: For explanation see text.
Figure 3: For explanation see text.