

# Proceedings of Symposia in APPLIED MATHEMATICS

---

Volume 63

## The Radon Transform, Inverse Problems, and Tomography

American Mathematical Society  
Short Course  
January 3–4, 2005  
Atlanta, Georgia

Gestur Ólafsson  
Eric Todd Quinto  
Editors



American Mathematical Society

**AMS SHORT COURSE LECTURE NOTES**  
**Introductory Survey Lectures**  
published as a subseries of  
Proceedings of Symposia in Applied Mathematics

*This page intentionally left blank*

# Proceedings of Symposia in APPLIED MATHEMATICS

---

Volume 63

## The Radon Transform, Inverse Problems, and Tomography

American Mathematical Society  
Short Course  
January 3–4, 2005  
Atlanta, Georgia

Gestur Ólafsson  
Eric Todd Quinto  
Editors



**American Mathematical Society**  
Providence, Rhode Island

## Editorial Board

Mary Pugh    Lenya Ryzhik    Eitan Tadmor (Chair)

LECTURE NOTES PREPARED FOR THE  
AMERICAN MATHEMATICAL SOCIETY SHORT COURSE  
THE RADON TRANSFORM AND APPLICATIONS TO INVERSE PROBLEMS  
HELD IN ATLANTA, GEORGIA  
JANUARY 3–4, 2005

The AMS Short Course Series is sponsored by the Society's Program Committee for National Meetings. The series is under the direction of the Short Course Subcommittee of the Program Committee for National Meetings.

2000 *Mathematics Subject Classification*. Primary 92C55, 44A12, 65R32, 65R10, 94A20, 42C40.

---

### Library of Congress Cataloging-in-Publication Data

The radon transform, inverse problems, and tomography : American Mathematical Society short course, January 3–4, 2005, Atlanta, Georgia / Gestur Ólafsson, Eric Todd Quinto, editors.  
p. cm. — (Proceedings of symposia in applied mathematics, ISSN 0160-7634 ; v. 63 AMS short course lecture notes)

Includes bibliographical references and index.  
ISBN 0-8218-3930-6 (acid-free paper)

1. Radon transforms—Congresses. 2. Imaging systems in medicine—Congresses. 3. Inverse problems (Differential equations)—Congresses. 4. Integral transforms—Congresses. 5. Sampling (Statistics)—Congresses. 6. Wavelets (Mathematics)—Congresses. I. Ólafsson, Gestur. II. Quinto, Eric Todd, 1951— III. American Mathematical Society. IV. Proceedings of symposia in applied mathematics ; v. 63. V. Series: Proceedings of symposia in applied mathematics. AMS short course lecture notes.

QA672.R33 2006  
515'.723—dc22

2005057107

---

**Copying and reprinting.** Material in this book may be reproduced by any means for educational and scientific purposes without fee or permission with the exception of reproduction by services that collect fees for delivery of documents and provided that the customary acknowledgment of the source is given. This consent does not extend to other kinds of copying for general distribution, for advertising or promotional purposes, or for resale. Requests for permission for commercial use of material should be addressed to the Acquisitions Department, American Mathematical Society, 201 Charles Street, Providence, Rhode Island 02904-2294, USA. Requests can also be made by e-mail to [reprint-permission@ams.org](mailto:reprint-permission@ams.org).

Excluded from these provisions is material in articles for which the author holds copyright. In such cases, requests for permission to use or reprint should be addressed directly to the author(s). (Copyright ownership is indicated in the notice in the lower right-hand corner of the first page of each article.)

© 2006 by the American Mathematical Society. All rights reserved.

The American Mathematical Society retains all rights  
except those granted to the United States Government.

Copyright of individual articles may revert to the public domain 28 years  
after publication. Contact the AMS for copyright status of individual articles.

Printed in the United States of America.

⊗ The paper used in this book is acid-free and falls within the guidelines  
established to ensure permanence and durability.  
Visit the AMS home page at <http://www.ams.org/>

## Contents

Preface GESTUR ÓLAFSSON AND ERIC TODD QUINTO	vii
An Introduction to X-ray Tomography and Radon Transforms ERIC TODD QUINTO	1
Development of Algorithms in Computerized Tomography ALFRED K. LOUIS	25
Fan-Beam Tomography and Sampling Theory ADEL FARIDANI	43
Generalized Transforms of Radon Type and Their Applications PETER KUCHMENT	67
Inverse Problems in Pipeline Inspection PETER MASSOPUST	93
Robust Interferometric Imaging in Random Media LILIANA BORCEA	129
Index	157

*This page intentionally left blank*

## Preface

This volume brings together six articles on the mathematical aspects of tomography and related inverse problems. They are based on the lectures in the Short Course, *The Radon Transform and Applications to Inverse Problems*, at the American Mathematical Society meeting in Atlanta, GA, January 3-4, 2005. They covered introductory material, theoretical problems, and practical issues in 3-D tomography, impedance imaging, local tomography, wavelet methods, regularization and approximate inverse, sampling, and emission tomography. All contributions are written for a general audience, and the authors have included references for further reading.

Tomography and inverse problems are active and important fields combining pure and applied mathematics with strong interplay between applications and the diverse mathematical problems that have emerged since the first article in the field appeared almost a century ago. The applied side is best known for medical and scientific applications, in particular, medical imaging, radiotherapy, and industrial non-destructive testing. Doctors use tomography to see the internal structure of the body or to find functional information, such as metabolic processes, noninvasively. Scientists discover defects in objects, the topography of the ocean floor, and geological information using X-rays, geophysical measurements, sonar, or other data. Thus, tomography consists of a broad range of inverse problems. These are called inverse problems because information about an object is obtained from indirect data.

X-ray tomography is the most basic modality, and it can be described in the following way: a beam of X-rays is emitted with a known intensity from a source outside the material to be scanned, usually some part of the human body. A detector on the other side of the body picks up the intensity after the ray has traveled along a straight line segment,  $L$ , through the body. Some X-rays are lost due to scattering and absorption because of the attenuation effects of the material. Let  $f$  be the linear attenuation coefficient of the body. If the X-rays are monochromatic, then the attenuation coefficient is proportional to the density of the object (the proportionality depends on the energy of the photons). Choosing units so the proportionality is 1 we can view  $f$  as the density function of the object. Then, a simple derivation (see e.g., [8, (2.1)]) shows that the logarithm of the intensity ratio is proportional to the line integral of the attenuation function, so in appropriate units,

$$(1) \quad \ln \left[ \frac{I(\text{source})}{I(\text{detector})} \right] = \int_L f(x) dx =: Rf(L).$$

The right-hand side of (1) is exactly the Radon line transform of  $f$ ,  $Rf(L)$ . In short, the  $Rf(L)$  averages the function  $f$  over the line  $L$ . The problems now becomes the recovery of  $f$  from  $Rf$ .

This leads us back to the beginning of the last century, more than forty years before the first CT-scanner emerged. The Radon line transform in  $\mathbb{R}^2$ , which is the case used in many X-ray tomography scanners, was treated by Johann Radon in 1917, and he also introduced this integral transform, that now bears his name, in arbitrary dimensions [9]. The Radon transform in  $\mathbb{R}^n$  integrates over hyperplanes, and the case  $n = 3$  was first considered by H.A. Lorentz before 1906, but it was never published (see [4, p. 51]).

In the early 1960s, Allan Cormack made the first CT (computerized tomography) scanner [2], and he developed mathematics to image objects from this X-ray data. His successful algorithm is based on a singular value decomposition, and it took a long time on the computers of that time. He received the Nobel Prize in 1979 for this seminal research. Subsequently, mathematicians, scientists, and engineers developed many fast algorithms for X-ray data, and they developed applications of the Radon transform to a broad range of tomography problems.

X-ray tomography is so useful, in general, because of an efficient, easy to implement, stable inversion method, filtered back projection, that gives excellent reconstructions, at least if complete data are given (the inversion formula and general remarks on implementation are given in [8, Section 2] and a detailed description of the algorithm is given in [6, Section 4]). The concept of complete data is understood using sampling theory, and this will be discussed in [3]. Loosely, complete data are tomographic data over a fairly evenly spaced set of lines through the object in a fairly equi-spaced set of directions. Even if filtered back projection is mathematically a simple and beautiful formula, the applied problems starts here! In particular, it is a simple fact, that no finite set of lines can determine the function  $f$  uniquely [10]. On the other hand, every real life CT apparatus can only use a finite number of lines, and the data are noisy and not exact. This brings up the connection to sampling theory and the importance of error estimates [3]. Also, physical limitations such as beam hardening (e.g., [10]) and other inaccuracies in the model require considerations that are not related to the Radon transform.

Many intriguing tomography problems involve limited tomographic data, when data over some lines are missing. One of the most important is the 3-D X-ray tomography problem discussed in [6]. Reconstruction from limited tomographic data is more difficult and less successful precisely because the missing data take away important information about the object, and in any case, the algorithms for complete data cannot be used. What is missing from the reconstruction can be analyzed in several ways, including microlocal analysis [8], singular value decompositions [6], and sampling theory [3], and this analysis is important since limited data problems occur in industry, science, and medicine.

This is not the whole story; many tomographic problems are modeled by Radon transforms that integrate in non-standard weights or over sets besides lines. Such transforms are called *generalized* Radon transforms. For example, the Radon transform of SPECT (single photon emission tomography) integrates over lines in a non-standard weight that depends on the object being scanned. One transform in sonar and geological testing integrates over circles or spheres, and several of the most important examples will be given in [5].

Many other tomographic problems are not directly modeled by Radon transforms, but they are tomographic because the goal is to find the internal structure of an object or the location of objects from indirect wave-based data such as in radar or seismic testing [1] or from electromagnetic data [7].

The first contribution, [8] by Todd Quinto, gives an introduction to the mathematics of X-ray tomography including a description of the range of the Radon transform and a basic inversion formula, filtered back projection. If  $n = 2$ , which is one important case in X-ray tomography, this inversion formula involves a non-local operator, which gives extra complications in the reconstruction of  $f$ . In many applications one does not need an exact reconstruction of  $f$ , instead one just needs to see the shape of structures in the object—the singularities of the object. Boundaries and imperfections in the object are singularities, so a reconstruction of singularities can show such structure. Furthermore, standard algorithms cannot be used in limited data problems. The author describes specific limited data problems in electron microscopy and industrial non-destructive evaluation, and he gives reconstructions. The theme of the article is that the microlocal analysis of the Radon transform can be used to understand what singularities of objects (wavefront set) are visible from limited tomographic data. The basic microlocal analysis is introduced, and then the author shows how these tomographic reconstructions reflect the microlocal understanding.

Alfred Louis discusses important algorithms in X-ray tomography, including limited angle tomography, a limited data problem that is important in industry, and 3-D cone-beam X-ray tomography, the tomography of many modern CT scanners [6]. He uses the approximate inverse to put these algorithms in the same general context. He develops a singular value decomposition for the limited angle transform, and he uses the approximate inverse to develop an inversion algorithm. Many modern CT scanners are so-called cone beam scanners; an X-ray source emits X-rays in a beam shaped like a three-dimensional cone. Typically, the source moves in a circle or a spiral around the object and the scanner generates a three-dimensional data set. The problem is more difficult geometrically and analytically. Louis describes the mathematics of 3-D cone beam tomography and he gives a mathematical description of his exact algorithm as well as reconstructions from real data.

In the third contribution to this volume [3], Adel Faridani applies sampling theory to X-ray tomography. He describes how many and which line integrals should be measured in order to achieve a desired resolution in the reconstructed image. This question for two-dimensional fan-beam tomography leads to a detailed discussion of problems in sampling theory on the torus,  $\mathbb{T}^2$ . The contribution provides an excellent example of the interplay between geometry, finite subgroups of the two-dimensional torus, the Shannon sampling theory and the Poisson summation formula. The focus is on the construction of efficient sampling schemes, the identification of algorithms for accurate reconstruction from efficiently sampled data, and the qualitative understanding of artifacts. The theory is based on the classification of all finite subgroups of  $\mathbb{T}^2$  to give efficient sampling sets and secondly a non-convex set  $K$  such that the translates of  $K$  by a certain lattice are all disjoint. Numerical experiments with a simple mathematical phantom are used to show the efficiency of the construction as well as artifacts coming from undersampling. While direct reconstruction with the standard filtered backprojection algorithm is found

to be suboptimal, interpolating the data first to a denser lattice via the sampling theorem leads to good results. The last section contains several problems and ideas for those who are interested in research in this area.

Peter Kuchment's article [5] describes several of the most important examples of generalized Radon transforms in tomography. The Radon transform used in X-ray tomography is not weighted; the attenuation coefficient (or density) of the object is integrated over lines in the data set without any multiplicative weight factor in equation (1). However, some of the most interesting Radon transforms (so-called generalized Radon transforms) in tomography integrate over sets besides lines or in weights besides the canonical ones. For example, the Radon transform in single photon emission tomography (SPECT) integrates over lines but with a weight depending on the material on the line. The goal of SPECT is to detect the distribution of sources of radioactivity using this emission data. A broad range of beautiful pure and applied results about SPECT are described in this article. The Radon transform that comes up in thermoacoustic tomography, sonar, and geophysical testing involves integrals over circles and spheres, and much progress, including new inversion formulas and a better understanding of the underlying mathematics, has occurred recently. The author provides an overview of this progress as well as reconstructions for this problem. Finally, the author discusses a model of electrical impedance tomography that involves the geodesic Radon transform on the hyperbolic plane.

Peter Massopust's article [7] shows how tomographic methods are used in a specific important industrial problem, pipeline inspection. Detecting corrosion or fatigue in a pipe carrying natural gas or oil, particularly along distribution lines, can prevent natural disasters. Sensors are placed on a device that fits into the pipe and travels down the pipe. The device induces a magnetic field in the pipe wall that is measured. The author describes mathematics and physics of the problem including details of the pipeline inspection device, introduces a model for the measured data for oil pipelines, and describes the resulting inverse problem. He describes B-splines and wavelets and uses them to develop an inversion method and deals well with problems inherent in the data acquisition method, including detector sensitivity and noise. He describes limitations in the model and the data acquisition method, and he gives reconstructions.

The last article in this volume [1] is written by Liliana Borcea and describes applications of inverse problems in imaging in random media. The question is how to determine the location of sources for wave propagation or strong scatters buried in a cluttered random medium. These kinds of problems show up in reflection seismology, synthetic aperture radar imaging, interferometric radar imaging and inverse scattering of time harmonic acoustic or electromagnetic waves. The author describes the difference between homogeneous and random media and explains where classical methods break down in random media. The simplified model of a single point source buried in a finely layered medium is discussed in detail.

**Acknowledgments.** The editors would like to thank the AMS for their support and help with organizing the short course and keeping it running smoothly. In particular, the co-organizers and short course participants are indebted to Wayne Drady, AMS Conference Coordinator, for the smooth organization and help. We would also like to thank the AMS for their support in the preparation of the this volume. We thank AMS Editorial Assistant Christine Thivierge for her cooperation with the editors and her able and thorough job putting the proceedings together. Finally, we thank the short course participants for making it all worthwhile. We hope you enjoy the volume.

## Bibliography

- [1] Liliana Borcea, *Robust Interferometric Imaging in Random Media*, in these proceedings, American Mathematical Society, 2006.
- [2] Allan M. Cormack, *Representation of a function by its line integrals with some radiological applications II*, J. Appl. Physics **35** (1964), 2908–2913.
- [3] Adel Faridani, *Fan-Beam Tomography and Sampling Theory*, in these proceedings, American Mathematical Society, 2006.
- [4] Sigurdur Helgason, *The Radon Transform, Second Edition*, Birkhäuser, Boston, 1999.
- [5] Peter Kuchment, *Generalized Transforms of Radon Type and Their Applications*, in these proceedings, American Mathematical Society, 2006.
- [6] Alfred K. Louis, *Development of Algorithms in Computerized Tomography*, in these proceedings, American Mathematical Society, 2006.
- [7] Peter Massopust, *Inverse Problems in Pipeline Inspection*, in these proceedings, American Mathematical Society, 2006.
- [8] Eric Todd Quinto, *An Introduction to X-ray Tomography and Radon Transforms*, in these proceedings, American Mathematical Society, 2006.
- [9] Johann Radon, *Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten*, Ber. Verh. Sach. Akad. **69** (1917), 262–277.
- [10] Kennan T. Smith, Donald C. Solmon, and Solomon L. Wagner, *Practical and mathematical aspects of the problem of reconstructing objects from radiographs*, Bull. Amer. Math. Soc. **83** (1977), 1227–1270.

*This page intentionally left blank*

## An Introduction to X-ray Tomography and Radon Transforms

Eric Todd Quinto

**ABSTRACT.** This article provides an introduction to the mathematics behind X-ray tomography. After explaining the mathematical model, we will consider some of the fundamental theoretical ideas in the field, including the projection slice theorem, range theorem, inversion formula, and microlocal properties of the underlying Radon transform. We will use this microlocal analysis to predict which singularities of objects will be well reconstructed from limited tomographic data. We will introduce specific limited data problems: the exterior problem, region of interest tomography, and limited angle region of interest tomography, and we use some of the author's reconstructions for these problems to illustrate the microlocal predictions about singularities. The appendix includes proofs of the basic microlocal properties of the Radon transform. Our overarching goal is to show some of the ways integral geometry and microlocal analysis can help one understand limited data tomography.

### 1. Introduction

The goal of tomography is to recover the interior structure of a body using external measurements, and tomography is based on deep pure mathematics and numerical analysis as well as physics and engineering. In this article, we will introduce some of the fundamental mathematical concepts in X-ray tomography and microlocal analysis and apply them to limited data problems. In the process, we will outline how the problems come up in practice and show what the microlocal

---

2000 *Mathematics Subject Classification.* Primary: 92C55, 44A12 Secondary: 35S30, 58J40.

*Key words and phrases.* Tomography, Radon Transform, Microlocal Analysis.

I am indebted to many researchers including, but not limited to the following people. Allan Cormack was a mentor to me, and he introduced me to the field of tomography and the exterior problem (§3.1). Allan and I enjoyed solving pure mathematical problems, too. Carlos Berenstein, Larry Shepp, and Larry Zalcman helped me as I began my career. Frank Natterer and Alfred Louis have been invaluable as I learned more, showing how deep analysis and numerical work go hand in hand. I learned about and appreciated Lambda CT from Kennan Smith and Adel Faridani. Victor Guillemin and Sig Helgason gave me a love of microlocal analysis and integral geometry. Peter Kuchment provided very helpful comments on material that appeared in a related article [38], and Matthias Hahn and Gestur Ólafsson corrected some misprints. I thank these folks, the other speakers in this short course, and other friends for making it all fun. This research is based upon work supported by the National Science Foundation under grants DMS-0200788 and DMS-0456858 and Tufts University FRAC.

analysis says for these problems. This is not a survey of the various types of tomography, rather a more detailed look at X-ray tomography and the mathematics behind it. For more information, we refer the reader to the other articles in this proceedings as well as sources such as [18, 26, 73, 75, 27, 32, 52, 53, 11, 38, 59] for details and further references on tomography and related problems in integral geometry.

We will start by explaining the mathematical model of X-ray tomography (§2). Then we will prove some of the basic theorems in the field. In section 2.2 we describe the microlocal properties of the Radon transform. In Section 3 we introduce the limited data problems and explain the challenges of limited data reconstruction, illustrating them with some of the author's reconstructions. Finally, in the appendix we provide proofs of the microlocal properties of the Radon transform.

## 2. X-ray Tomography

The goal of X-ray computed tomography (CT) is to get a picture of the internal structure of an object by X-raying the object from many different directions. We consider this problem in the plane and Alfred Louis will discuss the three-dimensional case [44].

As X-rays travel on a line  $L$  from the X-ray source through the object to an X-ray detector, they are attenuated by the material on the line  $L$  (we will neglect scatter and diffraction). According to Beer's law, the X-rays at a point  $x$  are attenuated proportionally to the number there, and the proportionality constant is called the *linear attenuation coefficient*. If the X-rays are monochromatic, then the linear attenuation coefficient is proportional to the density of the object; we will assume units are chosen so that the attenuation coefficient is equal to the density<sup>1</sup>. So, let  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  be the density of the object. Mathematically, the goal of X-ray CT is to recover  $f$  from these measurements. Thus, according to Beer's Law, if  $I(x)$  is the number of X-ray photons in the beam when it arrives at  $x$ , then the intensity in a small segment of length  $\Delta x$  is decreased by the multiplicative factor  $f(x)\Delta x$ , so:

$$(2.1) \quad \Delta I \approx -(f(x)\Delta x)I(x).$$

By separating variables and integrating (2.1) from the source to the detector, we get the following integral transform:

$$\ln \left[ \frac{I(\text{source})}{I(\text{detector})} \right] = \int_L f(x) dx_L =: Rf(L).$$

This integral transform  $R$  is exactly the classical Radon transform of  $f$  on the line  $L$  [27, 52], and since  $I(\text{source})$  and  $I(\text{detector})$  are measured, the line integral  $Rf(L)$  is known.

Remarkably, Radon invented this transform in 1917 for pure mathematical reasons [69]. Apparently, Lorentz had previously developed the transform in  $\mathbb{R}^3$ , but he never published it [27, p. 51]. It wasn't until Allan Cormack reinvented it in 1963 [7, 8] that it was used in tomography. Cormack won the Nobel Prize in Medicine in 1979 because he proposed using this transform to reconstruct the density of the

---

<sup>1</sup>If the X-rays are not monochromatic, then lower energy X-rays get attenuated more than higher energy X-rays, so the average energy of X-rays increases as they go through the object. This is called *beam-hardening*, and it can create reconstruction artifacts (e.g., [75]).

body from X-ray images from different directions; he gave a mathematical formula to do the reconstruction; and he implemented his ideas by building and testing a prototype CT scanner. Godfrey Hounsfield shared the prize for his independent work deriving an algorithm and making a medical CT scanner.

*Complete tomographic data* are X-ray data over all lines. In practice, this means data are collected on a fairly evenly distributed set of lines throughout the object. The concept of complete data can be made precise using sampling theory as in [52, 53] and Faridani's article in these proceedings [12]. In this case, the commonly used reconstruction algorithm is filtered backprojection (Theorem 2.5).

*Limited data tomography* is tomography when the data set does not include all lines. For example, data for *region of interest tomography* (§3.2) are over lines that go through a region of interest in the object. Lines that do not meet that region are not in the data set.

**2.1. General Facts about the Radon transform.** In general, we will follow the notation in [52]. We will need to give coordinates on the unit sphere,  $S^1$  so to each angle  $\varphi \in [0, 2\pi]$ , we denote the unit vector in direction  $\varphi$  as  $\theta$  and the unit vector  $\pi/2$  units counterclockwise from  $\theta$  as  $\theta^\perp$ :

$$(2.2) \quad \theta = \theta(\varphi) = (\cos \varphi, \sin \varphi) \quad \theta^\perp = \theta^\perp(\varphi) = (-\sin \varphi, \cos \varphi).$$

We identify 0 and  $2\pi$  as angles, and this allows us to identify  $[0, 2\pi]$  with  $S^1$  using the angle  $\varphi$  as coordinate.

We define

$$L(\varphi, s) = \{x \in \mathbb{R}^2 \mid x \cdot \theta(\varphi) = s\}$$

to be the line perpendicular to  $\theta = \theta(\varphi)$  and  $s$  directed units from the origin. The line  $L(\varphi, s)$  represents a line along which X-rays travel.

The Radon transform (2.3) of a function  $f \in L^1(\mathbb{R}^2)$  can be naturally interpreted as a function of  $(\varphi, s)$ :

$$(2.3) \quad Rf(\varphi, s) = \int_{x \in L(\varphi, s)} f(x) dx_L = \int_{t=-\infty}^{\infty} f(s\theta + t\theta^\perp) dt,$$

and in fact,  $R$  is continuous.

**THEOREM 2.1.** *The Radon transform is a continuous map from  $L^1(\mathbb{R}^2)$  to  $L^1([0, 2\pi] \times \mathbb{R})$ , and for  $f \in L^1(\mathbb{R}^2)$ ,  $\|Rf\|_{L^1([0, 2\pi] \times \mathbb{R})} \leq 2\pi \|f\|_{L^1(\mathbb{R}^2)}$ .*

The proof will be given along with the proof of Theorem 2.2.

Note that  $R$  satisfies the following evenness condition

$$(2.4) \quad Rf(\varphi, s) = Rf(\varphi + \pi, -s)$$

since  $L(\varphi, s) = L(\varphi + \pi, -s)$  and  $dx_L$  is the arc-length measure. We define the backprojection operator, the dual Radon transform of  $g \in L^1([0, 2\pi] \times \mathbb{R})$ , as

$$(2.5) \quad R^*g(x) = \int_{\varphi=0}^{2\pi} g(\varphi, x \cdot \theta(\varphi)) d\varphi.$$

This is the integral of  $g$  over all lines through  $x$ , since  $L(\varphi, x \cdot \theta)$  is the line through  $x$  and perpendicular to  $\theta$ . Let  $g$  be a smooth function of compact support,  $g \in$

$C_c([0, 2\pi] \times \mathbb{R})$ , then the partial Fourier transform of  $g$  in the  $s$  variable is:

$$(2.6) \quad \begin{aligned} \mathcal{F}_s g(\varphi, \sigma) &= \frac{1}{\sqrt{2\pi}} \int_{s=-\infty}^{\infty} e^{-is\sigma} g(\varphi, s) ds, \\ \mathcal{F}_s^{-1} g(\varphi, s) &= \frac{1}{\sqrt{2\pi}} \int_{\sigma=-\infty}^{\infty} e^{is\sigma} g(\varphi, \sigma) d\sigma \end{aligned}$$

For  $f \in C_c(\mathbb{R}^n)$  we define the  $n$ -dimensional Fourier transform and its inverse by

$$(2.7) \quad \begin{aligned} \widehat{f}(\xi) = \mathcal{F}f(\xi) &= \frac{1}{(2\pi)^{n/2}} \int_{x \in \mathbb{R}^n} e^{-ix \cdot \xi} f(x) dx \\ \mathcal{F}^{-1} f(x) &= \frac{1}{(2\pi)^{n/2}} \int_{\xi \in \mathbb{R}^n} e^{ix \cdot \xi} f(\xi) d\xi. \end{aligned}$$

We can now define the Riesz potential,  $I_s^{-1}$ , for  $g \in C_c^\infty([0, 2\pi] \times \mathbb{R})$  as the operator with Fourier multiplier  $|\sigma|$ :

$$(2.8) \quad I_s^{-1} g = I^{-1} g = \mathcal{F}_s^{-1}(|\sigma| \mathcal{F}_s g).$$

Before we can prove the inversion formula, we need to know the fundamental relationship between the Radon and Fourier transforms.

**THEOREM 2.2** (General Projection Slice Theorem, e.g., [52, 53]). *Let  $f \in L^1(\mathbb{R}^2)$  and let  $\varphi \in [0, 2\pi]$ . Let  $h \in L^\infty(\mathbb{R})$ . Then*

$$(2.9) \quad \int_{s=-\infty}^{\infty} Rf(\varphi, s) h(s) ds = \int_{x \in \mathbb{R}^2} f(x) h(\theta \cdot x) dx.$$

*That is, integrating  $Rf(\varphi, \cdot)$  with respect to  $h(s)$  is the same as integrating  $f$  with respect to the plane wave in direction  $\theta$ ,  $h(\theta \cdot x)$ .*

A special case of the projection-slice theorem with  $\sigma \in \mathbb{R}$  and  $h(s) = e^{-is\sigma}/(2\pi)$  is especially useful.

**COROLLARY 2.3** (Fourier Slice Theorem). *Let  $f \in L^1(\mathbb{R}^2)$ . Then,*

$$(2.10) \quad \frac{1}{\sqrt{2\pi}} \mathcal{F}_s Rf(\varphi, \sigma) = \widehat{f}(\sigma \theta).$$

This corollary shows why  $R$  is injective on domain  $L^1(\mathbb{R}^2)$ : if  $Rf \equiv 0$ , then  $\widehat{f} \equiv 0$  which shows that  $f$  is zero by injectivity of the Fourier transform.

**PROOFS OF THEOREMS 2.1 AND 2.2.** First, we establish that  $R$  is defined and continuous from  $L^1(\mathbb{R}^2)$  to  $L^1([0, 2\pi] \times \mathbb{R})$  using Fubini's theorem, and the General Projection Slice Theorem will follow.

Let  $f \in L^1(\mathbb{R}^2)$  and let  $H : [0, 2\pi] \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}^2$  be defined by  $H(\varphi, s, t) = (s\theta(\varphi) + t\theta^\perp(\varphi))$ . Then  $f \circ H$  is a Lebesgue measurable function since  $f$  is measurable and  $H$  is continuous. Furthermore, for  $\varphi \in [0, 2\pi]$  fixed,  $(s, t) \mapsto H(\varphi, s, t)$  is a rotation of  $\mathbb{R}^2$ , so it preserves measure. Therefore,  $(s, t) \mapsto f(H(\varphi, s, t))$  is in  $L^1(\mathbb{R}^2)$  and

$$(2.11) \quad \int_{x \in \mathbb{R}^2} f(x) dx = \int_{s=-\infty}^{\infty} \int_{t=-\infty}^{\infty} f(H(\varphi, s, t)) dt ds$$

$$(2.12) \quad = \int_{s=-\infty}^{\infty} \int_{t=-\infty}^{\infty} f(s\theta + t\theta^\perp) dt ds$$

$$(2.13) \quad = \int_{s=-\infty}^{\infty} Rf(\varphi, s) ds.$$

The right-hand side of (2.12) consists of an inner integral in the  $\theta$  direction and an integral over the set of lines perpendicular  $\theta$ , and this is exactly (2.13). Note that (2.13) follows from (2.12) by the definition of  $R$ , (2.3).

Equation (2.11) implies that  $f \circ H \in L^1([0, 2\pi] \times \mathbb{R}^2)$  as follows.

$$\begin{aligned}
 \|f \circ H\|_{L^1([0, 2\pi] \times \mathbb{R}^2)} &= \int_{\varphi=0}^{2\pi} \left( \int_{s=-\infty}^{\infty} \int_{t=-\infty}^{\infty} |f(H(\varphi, s, t))| ds dt \right) d\varphi \\
 (2.14) \quad &= \int_{\varphi=0}^{2\pi} \|f\|_{L^1(\mathbb{R}^2)} d\varphi \\
 &= 2\pi \|f\|_{L^1(\mathbb{R}^2)}.
 \end{aligned}$$

Now that we know  $f \circ H$  is in  $L^1$ , we can use Fubini's Theorem on domain  $[0, 2\pi] \times \mathbb{R}^2$ . Using (2.3):

$$\begin{aligned}
 \|Rf\|_{L^1([0, 2\pi] \times \mathbb{R})} &= \int_{\varphi=0}^{2\pi} \int_{s=-\infty}^{\infty} |Rf(\varphi, s)| ds d\varphi \\
 &= \int_{\varphi=0}^{2\pi} \int_{s=-\infty}^{\infty} \left| \int_{t=-\infty}^{\infty} f(s\theta(\varphi) + t\theta^\perp(\varphi)) dt \right| ds d\varphi \\
 (2.15) \quad &\leq \int_{\varphi=0}^{2\pi} \int_{s=-\infty}^{\infty} \int_{t=-\infty}^{\infty} |f(H(\varphi, s, t))| dt ds d\varphi \\
 &= 2\pi \|f\|_{L^1(\mathbb{R}^2)}
 \end{aligned}$$

by (2.14). This shows  $R$  is continuous in  $L^1$ .

The projection slice theorem follows from (2.11)-(2.13) and is an exercise for the reader. First, let  $F(x) = f(x)h(x \cdot \theta)$ , then  $F$  is in  $L^1(\mathbb{R}^2)$ . Now, plug  $F$  into (2.11).  $\square$

This theorem allows us to prove the easy part of the fundamental range theorem for the Radon transform. A function  $f$  is said to be in the Schwartz space  $\mathcal{S}(\mathbb{R}^2)$  if and only if  $f$  is  $C^\infty$  and  $f$  and all its derivatives decrease faster than any power of  $1/|x|$  at  $\infty$ . A function  $g(\varphi, s)$  is said to be in the Schwartz space  $\mathcal{S}([0, 2\pi] \times \mathbb{R})$  if  $g(\varphi, s)$  can be extended to be smooth and  $2\pi$ -periodic in  $\varphi$ , and  $g$  decreases (along with all derivatives in  $s$ ) faster than any power of  $1/|s|$  uniformly in  $\varphi$ .

**THEOREM 2.4** (Range Theorem). *Let  $g(\varphi, s) \in \mathcal{S}([0, 2\pi] \times \mathbb{R})$  be even ( $g(\varphi, s) = g(\varphi + \pi, -s)$ ) (see (2.4)). Then,  $g$  is in the range of the Radon transform,  $g = Rf$  for some  $f \in \mathcal{S}(\mathbb{R}^2)$ , if and only if all of the following moment conditions hold.*

$$\begin{aligned}
 (2.16) \quad \forall k = 0, 1, 2, \dots, \quad \int_{s=-\infty}^{\infty} g(\varphi, s) s^k ds \quad &\text{is a homogeneous polynomial} \\
 &\text{of degree } k \text{ in the coordinates of } \theta.
 \end{aligned}$$

Note that the evenness condition is necessary because of (2.4). An analogous theorem is true for the Radon hyperplane transform in  $\mathbb{R}^n$  [18, 26, 27].

**PROOF SKETCH.** Necessity is the easy part. We use (2.9) with  $g(s) = s^k$ . Then,

$$(2.17) \quad \int_{s=-\infty}^{\infty} Rf(\varphi, s) s^k ds = \int_{x \in \mathbb{R}^2} f(x) (x \cdot \theta)^k dx$$

and expand (2.17) in the coordinates of  $\theta$ . This shows (2.17) is a polynomial in  $\theta$  that is homogeneous of degree  $k$  when the unit vector  $\theta$  is viewed as a vector in  $\mathbb{R}^2$ . The difficult part of the proof is to show that if  $g$  satisfies the moment conditions, then  $g = Rf$  for some  $f$  that is smooth and rapidly decreasing. One uses the Fourier Slice Theorem 2.3 to get a function  $f$  that has Radon transform  $g$ . The subtle part of the proof is to show  $f$  is smooth at the origin, and this is where the moment conditions are used. The interested reader is referred to [18, 26, 27] for details.  $\square$

Now we have the background to state the filtered back projection inversion formula, which will be proved at the end of the section. Recall that  $R^*$  is defined by (2.5).

**THEOREM 2.5.** [70, 73, 52] *Let  $f \in C_c^\infty(\mathbb{R}^2)$ . Then  $f = \frac{1}{4\pi} R^*(I_s^{-1} Rf)(x)$ .*

Note that this theorem is true on a larger domain than  $C_c^\infty(\mathbb{R}^2)$ , but even for  $f \in L^1(\mathbb{R}^2)$ ,  $I^{-1}Rf$  could be a distribution rather than a function.

The theorem is applied in practice by truncating and smoothing the multiplier  $|\sigma|$  in  $I^{-1}$  and writing this truncated multiplier as a convolution operator in  $s$  [70, 52, 53]. The resulting approximate inversion algorithm becomes  $f \approx \frac{1}{4\pi} R^*(\Phi *_s Rf)$  where  $\Phi$  is the inverse Fourier transform of the truncated multiplier and  $*_s$  denotes convolution in the  $s$  variable,

$$g(\varphi, \cdot) *_s h(\varphi, \cdot) = \int_{s=-\infty}^{\infty} g(\varphi, s - \tau) h(\varphi, \tau) d\tau.$$

Here is some historical background. Old X-ray CT scanners took data using this parameterization,  $Rf(\varphi, s)$ , so-called parallel beam data. A single X-ray emitter and detector (or parallel emitter/detector sets) were oriented perpendicular to  $\theta$  and then were translated through the object (fixing  $\varphi$  and changing  $s$ ) then rotated to a new angle and then translated. The simplest kernel  $\Phi$  was given by cutting off  $|\sigma|$  and taking the inverse Fourier transform of

$$(2.18) \quad (\mathcal{F}_s \Phi)(\sigma) = \begin{cases} |\sigma| & |\sigma| \leq \Omega \\ 0 & |\sigma| > \Omega \end{cases}.$$

However,  $\Phi$  oscillates too much (an exercise shows that  $\Phi(s) = \frac{\sqrt{2}}{s^2 \sqrt{\pi}} (\Omega s \sin(\Omega s) + \cos(\Omega s) - 1)$ ). Various approximations to (2.18) and other kernels were used. See [73] for a discussion of the inversion methods and kernels of that time. Data collection for these scanners was time consuming because of the translation and rotation steps, but the inversion method was a simple application of Theorem 2.5. Once  $|\sigma|$  is approximated by a compactly supported function and  $\Phi(s)$  is the inverse Fourier transform, then the formula can be written  $f \approx \frac{1}{4\pi} R^*(\Phi *_s Rf)(x)$  and this is easy to implement since  $\Phi *_s Rf$  can be done as the data are collected and the backprojection step is just averaging as  $\theta$  is incremented around the circle.

Modern two-dimensional scanners take fan beam data, in which there is one point-source that emits X-rays in a fan at a bank of detectors on the other side of the body. The advantage is that the emitter/detectors need only to be rotated (not translated) to get data around the body. However, other adaptations of Theorem 2.5 are used since the data would have to be rebinned (a change of variable done) to change to the parallel beam parameterization of lines  $L = L(\varphi, s)$  so the convolution

$\Phi *_s Rf(\varphi, \cdot)$  could easily be done. Modern references such as [52, 53] provide excellent descriptions of the new inversion methods, and a detailed description of the method is given in Alfred Louis' article in these proceedings [44].

**PROOF OF THEOREM 2.5.** The proof uses some of the key elementary formulas for the Radon transform. One writes the two-dimensional Fourier inversion formula in polar coordinates and uses Fourier Slice Theorem 2.3 to get:

$$\begin{aligned} f(x) &= \frac{1}{2(2\pi)} \int_{\varphi=0}^{2\pi} \int_{\sigma \in \mathbb{R}} e^{ix \cdot (\sigma\theta)} \widehat{f}(\sigma\theta) |\sigma| d\sigma d\varphi \\ &= \frac{1}{4\pi} \int_{\varphi=0}^{2\pi} \int_{\sigma \in \mathbb{R}} \frac{e^{i\sigma(\theta \cdot x)}}{\sqrt{2\pi}} |\sigma| (\mathcal{F}_s Rf)(\varphi, \sigma) d\sigma d\varphi \\ &= \frac{1}{4\pi} \int_{\varphi=0}^{2\pi} I^{-1} Rf(\varphi, \theta \cdot x) d\varphi = \frac{1}{4\pi} R^* I^{-1} Rf(x). \end{aligned}$$

The factor of  $1/2$  before the first integral occurs because the integral has  $\sigma \in \mathbb{R}$  rather than  $\sigma \in [0, \infty)$ .  $\square$

**2.2. Wavefront Sets and Singularity Detection.** We will now use microlocal analysis to learn about how the Radon transform  $R$  detects singularities. To do this, we need a concept of singularity, the Sobolev wavefront set. The distribution  $f$  is in  $H^\alpha(\mathbb{R}^n)$  if and only if its Fourier transform,  $\widehat{f} = \mathcal{F}f$  is in  $L^2(\mathbb{R}^n, (1+|\xi|^2)^\alpha)$ . This relates global smoothness of  $f$  to integrability of its Fourier transform. A local version of this at a point  $x_0 \in \mathbb{R}^n$  is obtained by multiplying  $f$  by a smooth cut-off function  $\psi \in C_c^\infty(\mathbb{R}^n)$  (with  $\psi(x_0) \neq 0$ ) and seeing if the Fourier transform  $(\widehat{\psi f})$  is in this weighted  $L^2$  space. However, this localized Fourier transform  $(\widehat{\psi f})$  gives even more specific information—microlocal information—namely, the *directions* near which  $(\widehat{\psi f})$  is in  $L^2(\mathbb{R}^2, (1+|\xi|^2)^\alpha)$ . The precise definition is:

**DEFINITION 2.6** ([61], p. 259). A distribution  $f$  is in the Sobolev space  $H^\alpha$  locally near  $x_0 \in \mathbb{R}^n$  if and only if there is a cut-off function  $\psi \in C_c^\infty(\mathbb{R}^n)$  with  $\psi(x_0) \neq 0$  such that the Fourier transform  $(\widehat{\psi f})(\xi) \in L^2(\mathbb{R}^n, (1+|\xi|^2)^\alpha)$ . Let  $\xi_0 \in \mathbb{R}^n \setminus 0$ . The distribution  $f$  is in  $H^\alpha$  microlocally near  $(x_0, \xi_0)$  if and only if there is a cut-off function  $\psi \in \mathcal{D}(\mathbb{R}^n)$  with  $\psi(x_0) \neq 0$  and a function  $u(\xi)$  homogeneous of degree zero and smooth on  $\mathbb{R}^n \setminus 0$  and with  $u(\xi_0) \neq 0$  such that the product  $u(\xi)(\widehat{\psi f})(\xi) \in L^2(\mathbb{R}^n, (1+|\xi|^2)^\alpha)$ . The  $H^\alpha$  wavefront set of  $f$ ,  $\text{WF}^\alpha(f)$ , is the complement of the set of  $(x_0, \xi_0)$  near which  $f$  is microlocally in  $H^\alpha$ .

Note that  $\text{WF}^\alpha(f)$  is conic (if  $(x, \xi) \in \text{WF}^\alpha(f)$  then so is  $(x, a\xi)$  for any  $a > 0$ ) and closed. Also, note that the cut-off function,  $\psi$ , makes the calculation of wavefront sets intrinsically local: one needs only values of  $f(x)$  near  $x_0$  to find the wavefront set of  $f$  above  $x_0$ .

The Sobolev wavefront set and microlocal Sobolev smoothness are usually defined on  $T^*(\mathbb{R}^n) \setminus 0$ , the cotangent space of  $\mathbb{R}^n$  with its zero section removed, because such a definition can be extended invariantly to manifolds. To this end, let  $x_0 \in \mathbb{R}^n$ . If  $\vec{r} = (r_1, \dots, r_n) \in \mathbb{R}^n$ , then we let  $\vec{r}\mathbf{d}\mathbf{x} = r_1\mathbf{d}\mathbf{x}_1 + \dots + r_n\mathbf{d}\mathbf{x}_n$  be the cotangent vector corresponding to  $\vec{r}$  in  $T_{x_0}^*\mathbb{R}^n$ . A basic example will give a feeling for the definition.

**EXAMPLE 2.7.** Consider a function  $f$  in the plane that is smooth except for a jump singularity along a smooth curve  $C$ . Let  $x \in C$  and let  $\theta$  be normal to  $C$  at  $x$ . Because of this, we say that the covector  $\theta\mathbf{dx}$  is conormal to  $C$  at  $x$ . Then, clearly  $f$  is not smooth at  $x$ ;  $f$  is not even in  $H^1$  locally near  $x$ . In fact,  $(x, \theta\mathbf{dx}) \in \text{WF}^1(f)$  and it can be shown that  $\text{WF}^1(f)$  is the set of all conormals to  $C$ . So, the wavefront set gives a precise concept of singularity, not only points at which  $f$  is not smooth, but also directions in which  $f$  is not smooth.

The reader is encouraged to illustrate this principle by calculating  $\text{WF}^1(f)$  for the special case,  $f(x, y) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$ , when the curve  $C$  is the  $y$ -axis. This calculation is easier if one uses a cut-off function that is a product of cut-off functions in  $x$  and in  $y$ .

We will be dealing with Sobolev spaces of functions on  $Y = [0, 2\pi] \times \mathbb{R}$ . To do this, we extend functions  $g(\varphi, s)$  on  $Y$  periodically in  $\varphi$  and take localizing functions  $\psi$  with support in  $\varphi$  less than a period so  $\psi g$  can be viewed as a function on  $\mathbb{R}^2$  and the two-dimensional Fourier transform can be calculated using these coordinates. We let  $d\varphi$  and  $d\mathbf{s}$  be the standard basis of  $T_{(\varphi, s)}^*([0, 2\pi] \times \mathbb{R})$ , where the basis covector  $d\varphi$  is the dual covector to  $\partial/\partial\varphi$  and  $d\mathbf{s}$  is the dual covector to  $\partial/\partial s$ . The wavefront set is extended to distributions on  $[0, 2\pi] \times \mathbb{R}$  using these local coordinates, and it is a subset of  $T^*([0, 2\pi] \times \mathbb{R})$ .

If the reader is not familiar with cotangent spaces, one can just envision  $(x; \vec{r}\mathbf{dx})$  as the vector  $(x; \vec{r})$  where  $x$  represents a point in the plane and  $\vec{r}$  a tangent vector at  $x$ . In a similar way,  $(\varphi, s; a d\varphi + b d\mathbf{s})$  can be viewed as the vector  $(\varphi, s; a, b)$ .

The fundamental theorem that gives the relation between Sobolev wavefront of a function and its Radon transform is the following.

**THEOREM 2.8** (Theorem 3.1 [66]). *Let  $f$  be a distribution of compact support,  $f \in \mathcal{E}'(\mathbb{R}^2)$ . Let  $x_0 \in L(\varphi_0, s_0)$ ,  $\theta_0 = \theta(\varphi_0)$ ,  $\eta_0 = \mathbf{ds} - (x_0 \cdot \theta_0^\perp) d\varphi$  and  $a \neq 0$ . The Sobolev wavefront correspondence is*

$$(2.19) \quad (x_0; a\theta_0\mathbf{dx}) \in \text{WF}^\alpha(f) \quad \text{if and only if} \quad (\varphi_0, s_0; a\eta_0) \in \text{WF}^{\alpha+1/2}(Rf).$$

Given  $(\varphi_0, s_0; a\eta_0)$ ,  $(x_0; a\theta_0\mathbf{dx})$  is uniquely determined by (2.19). Sobolev singularities of  $Rf$  above  $(\varphi_0, s_0)$  give no stable information about Sobolev singularities of  $f$  above points not on  $L(\varphi_0, s_0)$  or at points on this line in directions not conormal to the line. These other singularities are smoothed by data  $Rf$  near  $(\varphi_0, s_0)$ . Singularities above points not on the line do not affect singularities of  $Rf$  on  $(\varphi_0, s_0)$ .

The proof of this theorem is in the appendix along with proofs that  $R$  and  $R^*$  are elliptic Fourier integral operators and  $R^*R$  is an elliptic pseudodifferential operator.

Theorem 2.8 allows one to understand what  $R$  does to singularities in a precise and rigorous way and it provides an application of microlocal analysis to tomography. It gives an exact correspondence between singularities of  $f$  and those of  $Rf$ . Moreover, it states that the singularities of  $Rf$  that are detected from the data are of Sobolev order  $1/2$  smoother than the corresponding singularities of  $f$ . A simple illustration will give an intuitive feeling for the theorem.

**EXAMPLE 2.9.** Let  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  be equal to one on the unit disk and zero outside. Then,  $Rf(\varphi, s) = 2\sqrt{1-s^2}$  for  $|s| \leq 1$  and  $Rf(\theta, s) = 0$  for  $|s| > 1$ . The only lines

where  $Rf$  is not smooth are those with  $|s| = 1$  and these lines are tangent to the unit circle, the curve on which  $f$  is discontinuous. Since  $\text{WF}^1(f)$  is the set of covectors conormal to the unit circle (see Example 2.7), singularities of  $Rf$  precisely locate the corresponding singularities of  $f$ .

**REMARK 2.10.** Here is how to use Theorem 2.8 to determine  $\text{WF}^\alpha(f)$ . Let  $x_0 \in \mathbb{R}^2$ , and  $\varphi_0 \in [0, 2\pi]$  and  $a \neq 0$ . To see if  $(x_0, a\theta_0 \mathbf{d}\mathbf{x}) \in \text{WF}^\alpha(f)$  we need to know if the covector  $(\varphi_0, x_0 \cdot \theta_0; a(\mathbf{d}s - (x_0 \cdot \theta_0^\perp) \mathbf{d}\varphi)) \in \text{WF}^{\alpha+1/2}(Rf)$ . To determine this, we need only data  $Rf$  near  $(\varphi_0, s_0)$  where  $\theta_0 = \theta(\varphi_0)$ ,  $s_0 = x_0 \cdot \theta_0$  because the calculation of wavefront sets is local.

It follows from Theorem 2.8 that if  $Rf$  is in  $H^{\alpha+1/2}$  near  $(\varphi_0, s_0)$ , then  $f$  is  $H^\alpha$  in directions  $\pm\theta_0$  at all points on the line  $L(\varphi_0, s_0)$ , and if  $Rf$  is not in  $H^{\alpha+1/2}$  near  $(\varphi_0, s_0)$ , then at some point  $x \in L(\varphi_0, s_0)$ ,  $(x, \theta_0)$  or  $(x, -\theta_0)$  is in  $\text{WF}^\alpha(f)$ .

We will apply this theorem to limited data sets, that is  $Rf(\varphi, s)$  for  $(\varphi, s)$  in an open proper subset of  $\mathcal{A} \subset [0, 2\pi] \times \mathbb{R}$ . With limited data in  $\mathcal{A}$ , the only points at which we can find wavefront sets are points  $x \in L(\varphi, s)$  for  $(\varphi, s) \in \mathcal{A}$ , and if  $x \in L(\varphi, s)$  the only wavefront directions we see at  $x$  are the directions perpendicular to the line, directions  $\pm\theta$ . Other wavefront directions at points on  $L(\varphi, s)$  are not visible from this data, and wavefront at points off of  $L(\varphi, s)$  are not visible from this data since they do not affect this data.

**DEFINITION 2.11.** We will say a singularity of  $f$   $(x, \pm\theta \mathbf{d}\mathbf{x})$  is *visible* from a limited data set if the line  $L(\varphi, x \cdot \theta)$  is in the data set. Other singularities will be called *invisible*.

Of course, visible singularities at a point  $x$  are ones that, according to (2.19), affect the smoothness of  $Rf$  near  $(\varphi_0, x \cdot \theta_0)$ . The associated singularities of  $Rf$  are  $1/2$  order smoother in Sobolev scales than the corresponding singularities of  $f$ , so they should be stably detectable from Radon data for lines near  $L(\varphi, x \cdot \theta)$ . Invisible singularities are not really invisible but are harder to reconstruct because they are smoothed to  $C^\infty$  by data near  $(\varphi_0, x \cdot \theta_0)$ . A similar definition was given by Palamodov [58] for sonar. We now illustrate the idea by an example.

**EXAMPLE 2.12.** Assume we are given limited tomographic data of a function  $f$  over an open set  $\mathcal{A} \subset Y$ . Assume  $f$  has a jump singularity along a smooth curve  $C$ , and  $x \in C$ . Let  $\theta$  be perpendicular to the curve  $C$  at  $x$ . Then, the line  $L(\varphi, x \cdot \theta)$  is tangent to  $C$  at  $x$  and, as noted in Example 2.7,  $(x, \theta \mathbf{d}\mathbf{x}) \in \text{WF}^1(f)$ . If this line  $L(\varphi, x \cdot \theta)$  is in the data set, i.e.,  $(\varphi, x \cdot \theta) \in \mathcal{A}$ , then according to Theorem 2.8, this singularity at  $x$  will be stably detectable from the data, but if the line is not in the data set, then the singularity will not be stably detectable but will be smoothed by the data  $Rf$  on  $\mathcal{A}$ .

In other words, if the line tangent to  $C$  at  $x$  is in the data set, then the jump singularity at  $x$  along  $C$  will be visible and if not, the singularity will be invisible.

One can use this paradigm to understand which singularities of an object will be visible from limited data, and we will illustrate this for three limited data problems in Section 3. Palamodov stated a closely related idea in [57]. The “tangent casting” effects of [74] are related to Example 2.12.

**REMARK 2.13.** It is important to note that this paradigm explains only part of the issue. A good algorithm should be able to reconstruct singularities more clearly

if they are visible from the data set. But other issues, such as noisy data or a bad algorithm, could have a larger effect on the reconstruction than the paradigm. In any case, the paradigm does not predict how an algorithm will reconstruct invisible singularities. Some algorithms, like my ERA [67] (see Figure 1) smear these singularities and others, like limited data Lambda CT can just make them disappear (e.g., [37]). Finally, other issues such as data sampling can have a dramatic effect on reconstruction as will be discussed in Faridani's article [12] in this collection.

Now, we will give some history of the microlocal perspective on Radon transforms. Guillemin first developed the microlocal analysis of the Radon transform. In broad generality, he proved that  $R$  is an elliptic Fourier integral operator, and he proved that  $R^*R$  is an elliptic pseudodifferential operator under a specific assumption, the Bolker Assumption (see Remark A.3) [22]. Because Guillemin showed  $R$  is a Fourier integral operator associated to a specific canonical relation (A.6) [22, 24], (2.19) follows. Sobolev continuity is a basic property for Fourier integral operators; any FIO, such as  $R$ , of order  $-1/2$  will map functions in  $H^\alpha$  of fixed compact support continuously to functions in  $H_{\text{loc}}^{\alpha+1/2}$ . If the operator is elliptic, then the original function must be  $1/2$  order less smooth than its image. Theorem 2.8 is a refinement of this observation. Not only does  $R$  smooth of order  $1/2$  but it maps functions that are in  $H^\alpha$  microlocally near a given covector to functions that are in  $H^{\alpha+1/2}$  near the covector given by the correspondence (2.19).

The author [62] described the symbol of  $R^*R$  for all generalized Radon transforms satisfying the Bolker Assumption (see Remark A.3) in terms of the measures involved into the transform. He also proved more concrete results for the hyperplane transform. Beylkin [3] proved related results for Radon transforms satisfying the Bolker Assumption integrating over surfaces in  $\mathbb{R}^n$ . One type of generalized Radon transform will be discussed in Peter Kuchment's article in these proceedings [36].

The author's article [66] that included Theorem 2.8 was written to show the connection between microlocal analysis and singularity detection in tomography. The microlocal analysis was known to Fourier analysts, and tomographers understood the heuristic ideas about singularity detection, but the explicit and precise connection in (2.19) was not generally known in tomography. Subsequent many authors have used microlocal analysis to understand problems in tomography including radar [54, 55], and X-ray tomography (e.g., [37, 33]).

### 3. Limited Data Tomography

*Limited tomographic data* are tomographic data given on some proper open subset  $\mathcal{A} \subset [0, 2\pi] \times \mathbb{R}$ . Theorem 2.8 and Remark 2.10 provide a paradigm to decide which singularities of  $f$  are stably visible from limited tomographic data, and we will examine what this predicts for three common types of limited data: exterior CT, (§3.1), the interior problem or region of interest CT (§3.2), and limited angle region of interest CT (§3.3).

Density functions  $f$  in tomography can often be modelled by piecewise continuous functions that are continuous on open sets with well-behaved boundaries. So, singularities of  $f$  occur at the boundaries, and the singularities are in  $H^{1/2-\epsilon}$  for  $\epsilon > 0$ . By Theorem 2.8, the corresponding singularities of the Radon data  $Rf$  will be in  $H^{1-\epsilon}$ . A limitation of this analysis is that any discrete data  $Rf$  can

be considered an approximation of a smooth function. However, singularities of  $Rf$  should have large norm in  $H^1$  when given by discrete data and so should be visible. Moreover, the paradigm of the preceding section is observed in all typical CT reconstructions from limited data including those in Figures 1, 2, and 3 below.

We now give a little historical perspective. Ramm and Zaslavsky [71] have analyzed how the Radon transform itself behaves on functions that are smooth except at smooth boundary surfaces. They give precise asymptotics of  $Rf$  at lines tangent to boundaries depending on the curvature of the boundaries, and they have proposed a singularity detection method using this information on the raw data [71]. This method has been tested on simulated data [34]. A more general method using wavefront sets has been proposed [66] based on correspondence (2.19) and Theorem 2.8. Candès and Donoho, have developed ridgelets [5], an exciting way to make this correspondence and wavefront sets in general numerically tractable [6].

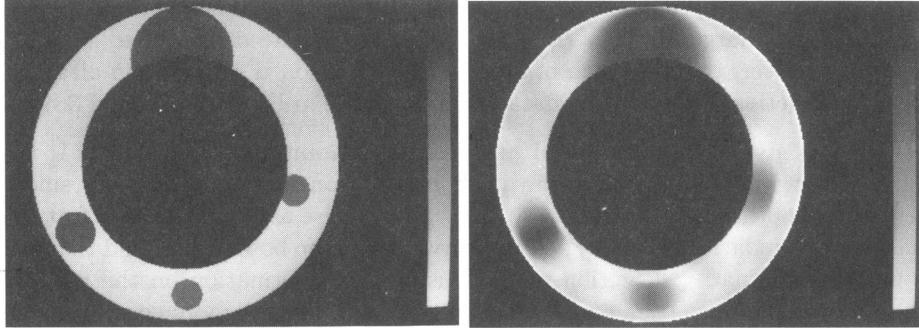
It should be pointed out that one can also understand stability of limited data tomography using singular value decompositions [9, 49, 43, 46, 47, 48], and they reflect the principle predicted by the microlocal analysis: singular functions associated with large singular values (which are easy to reconstruct) oscillate in directions in which wavefront is easily detectable, and vice versa. This is discussed more completely in [68] and examples are given there.

**3.1. The Exterior Problem.** Let  $M > 1$  and assume  $\text{supp } f \subset \{x \in \mathbb{R}^2 \mid |x| \leq M\}$ . In the exterior problem, one has data  $Rf(\varphi, s)$  for all  $\varphi \in [0, 2\pi]$  but only for  $|s| > 1$ . By the support theorem for the Radon transform [18, 26], one can reconstruct  $f(x)$  for  $|x| > 1$ . This problem comes up in studies of the solar corona [1] in which data are total intensities of the corona of the sun along lines exterior to the core of the sun that go from the solar corona to the observer on earth. Exterior data also occur in industrial tomography of very large objects such as rocket shells, for which X-ray data through their centers is too highly attenuated to be usable [67].

Here is some history. An example of Finch's can be adapted to show inversion of the exterior transform is discontinuous in any range of Sobolev norms. This is reflected in the fact that some singularities are invisible; exterior data smooths them. Lissianoi [40] has extended Finch's Sobolev discontinuity result to show that in the exterior problem even recovery of the function in a smaller ring than where the data are given and does not improve stability. However, logarithmic stability has been proved by Isakov and Sun [31].

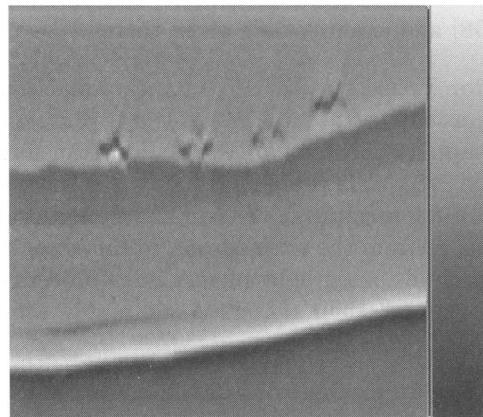
Lewitt and Bates [2], Louis, and Natterer [51] have developed good reconstruction algorithms that use exterior data. Lewitt and Bates' algorithm completes the exterior data by projecting it on the range of the complete Radon transform as Louis did for limited angle data [42]. The projection step is unstable because the singular functions are not orthogonal on the annulus but on a disk. Natterer's algorithm is an effective regularization method.

The author has developed an exterior reconstruction algorithm which employs a singular value decomposition [60] for the Radon transform on domain  $L^2(\{x \in \mathbb{R}^2 \mid |x| \geq 1\})$  and *a priori* information about the shape of the object to be reconstructed. Reconstructions for ‘medical’ phantoms are given in Figure 1 [63], and reconstructions from industrial phantoms are in [65, 64] and real industrial data in [67]. Industrial reconstruction from the author’s limited angle exterior



**Figure 1.** Phantom (left) and reconstruction (right) from simulated data [63] using the author's exterior reconstruction algorithm. Note how the boundaries tangent to lines in the data set (lines not intersecting the inner disk) are sharper than boundaries tangent to lines not in the data set.

Lambda tomography algorithm [68] are given in Figure 2. The article [37] has Lambda reconstructions of simulated exterior data.



**Figure 2.** Polar coordinate display of the author's limited angle exterior Lambda reconstruction from a  $3\pi/4$  angular range limited angle exterior data set of a Perceptics rocket motor mockup [68]. The horizontal axis corresponds to  $\varphi \in [0, \pi/2]$ , the vertical corresponds to  $r \in [0.9453, 1.0]$  (with  $r = 1$  at the bottom and magnified by a factor of 27). The bottom of the picture shows some area outside the object. The part of the reconstruction in  $[\pi/2, 3\pi/4]$  is of the same quality but less interesting. Data were taken over 1350 sources in the range  $[0, 3\pi/4]$  and 280 detectors.

The reconstructions in Figures 1 and 2 illustrate the paradigm in Remark 2.10 perfectly. The lines in the data set are those that do not meet the center disk, and the boundaries tangent to those lines (with wavefront perpendicular to those lines) are better defined than other parts of the boundaries. For example, in Figure 1, for each little circle, the inside and outside boundaries are better defined than the sides. The reconstruction in Figure 2 is displayed in polar coordinates, and if it is mapped into rectangular coordinates, almost all of the singularity curves are tangent to lines in the data set since the vertical axis is the radial direction so these curves map to curves of approximately constant radius.

**3.2. Region of Interest Tomography.** Let  $M > 1$  and assume  $\text{supp } f \subset \{x \in \mathbb{R}^2 \mid |x| \leq M\}$ . *Region of interest data* (or *interior tomographic data*) are data  $Rf(\varphi, s)$  for all  $\varphi$  but only for  $|s| < 1$ . Data are missing over lines outside the unit disk, even though  $\text{supp } f$  can meet the annulus  $\{x \in \mathbb{R}^2 \mid 1 \leq |x| \leq M\}$ . This is the opposite of exterior data, in which data are given *outside* the unit disk.

The goal of region of interest CT is to reconstruct information about  $f(x)$  in the region of interest, the unit disk. This problem comes up whenever scientists want information only about some region of interest in an object, not the whole object, or in problems, such as high-resolution tomography of very small parts of objects, for which it is difficult or impossible to get complete high-resolution CT data [14].

Simple examples (derived using Range Theorem 2.4) show the interior transform, the Radon transform with this limited data, is not injective.

However, according to Theorem 2.8, all singularities of  $f$  in  $|x| < 1$  are visible. To see this, choose a point  $x$  inside the unit disk and choose an angle  $\varphi \in [0, 2\pi]$ . Then the line through  $x$  and normal to  $\theta = \theta(\varphi)$  is in the data set for interior tomography. Therefore, by Theorem 2.8, any singularity of  $f$  at  $(x; \theta \mathbf{d}\mathbf{x})$  is stably detected by interior data. This explains why singularity detection methods work so well for interior data.

Lambda tomography [73, 15, 13, 78] is a well developed singularity detection algorithm that uses interior data. The key is that by (2.8),  $I^{-2} = I^{-1} \circ I^{-1} = -\frac{d^2}{ds^2}$  which is a local operator. In Lambda tomography, one replaces  $I^{-1}$  in the filtered backprojection inversion formula,  $f = \frac{1}{4\pi} R^* I^{-1} Rf$  by  $I^{-2}$  to get

$$(3.1) \quad \sqrt{-\Delta} f = \frac{1}{4\pi} R^* I^{-2} Rf = \frac{-1}{4\pi} R^* \left( \frac{d^2}{ds^2} Rf \right).$$

One shows the first equality in (3.1) using arguments similar to those in the proof of Theorem 2.5. Equation (3.1) is a local reconstruction formula because one needs only data  $Rf$  near  $(\varphi, x \cdot \theta)$  to calculate  $\frac{d^2}{ds^2} Rf(\varphi, x \cdot \theta)$  and then to calculate  $\frac{-1}{4\pi} R^* \left( \frac{d^2}{ds^2} Rf(x) \right)$ . Moreover,  $\sqrt{-\Delta}$  is an elliptic pseudodifferential operator of order 1 and so  $\text{WF}^\alpha(f) = \text{WF}^{\alpha-1}(\sqrt{-\Delta}f)$ . This reconstruction formula takes singularities of  $f$  and makes them more pronounced; any  $\text{WF}^\alpha$  singularity of  $f$  becomes a more pronounced singularity, in  $\text{WF}^{\alpha-1}$ , because  $\Lambda$  is an elliptic pseudodifferential operator of order 1. At the same time, there is a cupping effect at the boundaries [15]. Because of this, the developers of Lambda CT chose to add a multiple of  $R^* Rf = f * (2/|x|)$  (see the proof of Theorem A.1) to the reconstruction. The effect is to provide a smoothed version of the density  $f$ ; moreover, with a good multiple, the cupping effects are decreased [15, 13].

Here is some background and perspective. The authors of [2] have developed a reconstruction method that projects the interior data onto the range of the Radon transform with complete data and then inverts this completed data. Because of the non-uniqueness of the problem, this projection step is not unique. Maaß has developed a singular value decomposition for this problem, and he showed that singular functions associated to small singular values are fairly constant inside the region of interest, the unit ball [48]. These singular functions corresponding to small singular values are difficult to reconstruct, but they do not add much detail inside the region since they are relatively constant there. See also [46]. This

reflects the fact that all Sobolev singularities inside the region of interest are stably reconstructed.

Limited angle and exterior versions of Lambda CT have been developed, and they are promising on simulations [37] and tests on industrial and electron microscope data (see [68] and section 3.3).

Other algorithms for region of interest CT have been developed. Ramm and Katsevich developed pseudolocal tomography [35]. In industrial collaboration, Louis has used wavelets to help detect boundaries of bone and metal in X-ray CT. Madych [50] used wavelet analysis to show the strong relationship between Lambda CT, pseudolocal CT, and regular filtered backprojection. Authors of the papers [10, 72] have also used wavelet techniques for local tomography. They use Radon data to calculate wavelet coefficients of the density to be reconstructed. Although they need some data slightly outside the region of interest, their methods are fairly local.

Candès and Donoho [5] have recently defined ridgelets, wavelets that are not radially symmetric and are more sensitive to singularities in specified directions. They have developed a local tomographic reconstruction algorithm using ridgelets that provides high-quality reconstructions in practice. Their ideas reflect the singularity detection predictions of Theorem 2.8. In an exciting development, they have also used wavelets to detect wavefront sets of functions and develop a ridgelet theory of Fourier integral operators [6].

For the X-ray transform over a curve in  $\mathbb{R}^3$ , Louis and Maaß [45] have developed a very promising generalization of Lambda CT (see [33] for related ideas). Greenleaf and Uhlmann [20, 21] completely analyzed the microlocal properties of  $R^*R$  for admissible transforms on geodesics, including this case. The microlocal analysis of this three-dimensional X-ray transform has been investigated by Finch, Lan, and Uhlmann [16, 39], and Katsevich [33]. The operator  $R^*I^{-2}R$  adds singularities to  $f$  because  $R$  is not well enough behaved; for the classical line transform in the plane,  $R^*I^{-2}R$  is an elliptic pseudodifferential operator and therefore preserves singularities. The theorem corresponding to Theorem 2.8 for the X-ray transform on lines through a curve is given in [66], and it explains which singularities are stably reconstructed by this transform. The prediction is observed in the reconstructions in [45]. The general analysis in [23] provides the basic microlocal results needed to prove this theorem, and the specific microlocal properties of this operator on manifolds is given [20]. The microlocal properties of this line transform were also developed in [4] as a way to prove support and uniqueness theorems for the X-ray transform.

### 3.3. Limited angle region of interest tomography and electron microscopy.

Disregarding scatter and assuming an ideal detector, one can interpret electron microscope data as the X-ray transform of the three-dimensional structure of the object being scanned [17, 28, 53]. In collaboration with Ulf Skoglund at the Karolinska Institute and Ozan Öktem at Sidec Technologies, the author is developing a limited data singularity detection algorithm [68] for electron microscopy data. The starting point is the author's adaptation of Lambda tomography to limited data (see [68] for exterior and limited angle exterior data and see [37] for a related algorithm for general measures).

Here is some background on other tomographic methods in electron microscopy. It should be pointed out that there is a rich theory that considers electron microscopy as a discrete problem. In one model, the goal is to find the location of atoms and the model is a pixel function that is either one or zero. Authors such as Lohmann (Ph.D. Thesis, Uni. Münster), Shepp and others have developed good methods to detect such objects using data from a few angles. In another type of experiment, so called single-view microscopy, the object consists of multiple copies of the same rigid (typically crystalline) molecule, and the goal is to reconstruct the shape of the molecule from one view of the sample. In this case, the data are images of the same molecule in different *but unknown* directions. Researchers (e.g., [19]) have developed methods to reconstruct structures from projections in random directions.

However, our model is more intrinsically integral geometric. The Karolinska Institute has an electron microscope which images the specimen in a stack of two-dimensional slices. One can rotate the object in a limited angular range in the microscope, typically  $120^\circ$ , so these data are so called *limited angle tomographic data*. This method takes tomographic data of individual molecules from different, known angles, and the result is a reconstruction of the specific molecules. As opposed to the “discrete” methods above, this method can image individual molecules among many, varied objects and find how each is shaped.

Furthermore, the sample is so long relative to the thin focused electron beam that only a small region of interest is being imaged. Therefore the data are limited angle region of interest data.

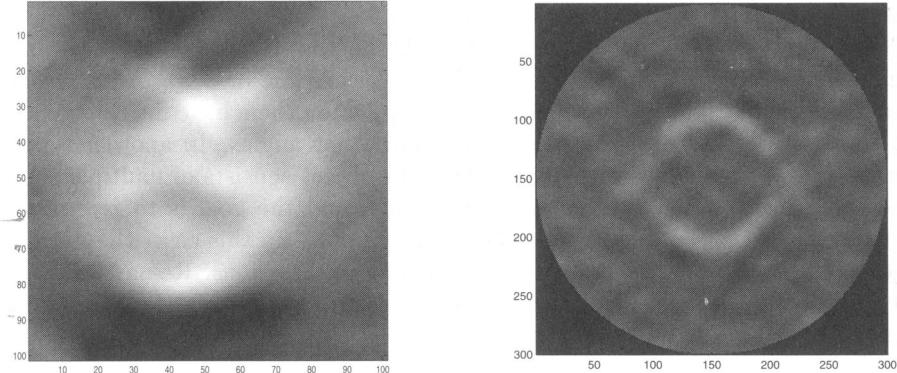
Ulf Skoglund and coresearchers at Karolinska Institute have developed a constrained minimum relative entropy method of reconstruction [56]. This method works by starting with a prior, which is a guess for the object to be reconstructed, and my algorithm could be used as such a prior, or as an independent reconstruction method.

Let  $V$  be the region of interest, a small open set in a cross-section of the slide, and let  $U$  be the set of perpendicular directions the microscope is rotated in angles. So, the data set consists of lines for  $\{(\varphi, x \cdot \theta) \mid \varphi \in U, x \in V\}$ . If we apply Theorem 2.8 and Remark 2.10, the only wave front directions of  $f$  at points  $x_0 \in V$  we will be able to see are those with directions  $\varphi \in U$ .

The visible singularities (those tangent to lines in the data set [66], Remark 2.10 and Example 2.12) should be precisely reconstructed by the algorithm. Two of my reconstructions are given in Figure 3. The left picture is a reconstruction of a virus from relatively clean data, and the right one uses very noisy simulated data of the cross-section of a spherical shell. The left reconstruction used higher electron flux than is typical (making for less noise and easier reconstruction), and the mottling in the right reconstruction is typical of reconstructions of real data with lower electron flux. Both reconstructions illustrate the principle that object boundaries tangent to lines in the data set (lines with perpendicular angle between  $30^\circ$  and  $150^\circ$ ) are well imaged, and others are not.

## Appendix A. The microlocal properties of $R$ and $R^*R$

In this section, we will show that  $R^*R$  is an elliptic pseudodifferential operator and then that  $R$  and  $R^*$  are elliptic Fourier integral operators. In the process, we



**Figure 3.** Left: reconstruction from my local tomography algorithm of virus from clean data. Right: reconstruction using the same algorithm of a cross-section of a sphere from very noisy simulated data, both from the Karolinska Institute and Sidec. In both pictures the visible singularities (those with normal angles between  $30^\circ$  and  $120^\circ$ ) are better reconstructed.

will calculate their canonical relations. Finally, we will use these results to prove Theorem 2.8.

**THEOREM A.1.**  *$R^*R$  is an elliptic pseudodifferential operator of order  $-1$  and*

$$(A.1) \quad R^*Rf(x) = \int_{\xi \in \mathbb{R}^2} e^{ix \cdot \xi} \frac{2}{|\xi|} \widehat{f}(\xi) d\xi.$$

**PROOF.** Here is a fun exercise for the reader. Show

$$(A.2) \quad R^*Rf(x) = \left( \frac{2}{|x|} \right) * f.$$

One can begin the proof with the fact  $Rf(\varphi, x_0 \cdot \theta) = \int_{t=-\infty}^{\infty} f(x_0 + t\theta^\perp) dt$  and then use polar coordinates.

Next, we use the fact that the Fourier transform of  $1/|x|$  is  $1/|\xi|$  [27, (42) p. 161]. Now, we take the Fourier transform and then inverse Fourier transform of (A.2) to get

$$(A.3) \quad R^*Rf(x) = 2\pi \mathcal{F}^{-1} \left( \frac{2}{|\xi|} \widehat{f} \right) = \int_{\xi \in \mathbb{R}^2} e^{ix \cdot \xi} \frac{2}{|\xi|} \widehat{f}(\xi) d\xi$$

which proves (A.1). The factor of  $2\pi$  in front of the middle term in (A.3) comes about because, with our normalizations,  $\mathcal{F}(f * g) = 2\pi(\mathcal{F}f)(\mathcal{F}g)$ . Note that the integral is in  $L^1$  for  $f \in C_c^\infty(\mathbb{R}^2)$  and it can be defined more generally, as with pseudodifferential operators, using integration by parts [61]. Because the symbol of  $R^*R$ ,  $\frac{2}{|\xi|}$ , is homogeneous and nowhere zero,  $R^*R$  is a classical elliptic pseudodifferential operator [61]. Because the symbol is homogeneous of degree  $-1$ ,  $R^*R$  is a pseudodifferential operator of order  $-1$ .  $\square$

Our theorems about the Radon transform itself will be easier to describe if we recall our identification on  $[0, 2\pi]$ . We have identified  $0$  and  $2\pi$  so  $\varphi \in [0, 2\pi]$  becomes a smooth coordinate on  $S^1$ , and we let

$$(A.4) \quad y = (\varphi, s) \in Y := [0, 2\pi] \times \mathbb{R}.$$

We denote the dual covector to  $\partial/\partial\varphi$  by  $\mathbf{d}\varphi$ , and we denote covectors in  $T^*\mathbb{R}^2$  by  $(r_1, r_2)\mathbf{dx} = r_1\mathbf{dx}_1 + r_2\mathbf{dx}_2$ .  $Rf$  is a function of  $y = (\varphi, s) \in Y$ ,  $Rf(y) = Rf(\varphi, s)$ .

We need some notation about subsets of vector bundles. Let  $X$  and  $Y$  be manifolds and let  $A \subset T^*X \times T^*Y$ , then we define

$$\begin{aligned} A' &= \{(x, y; \xi, -\eta) \mid (x, y; \xi, \eta) \in A\}, \\ A^t &= \{(y, x; \eta, \xi) \mid (x, y; \xi, \eta) \in A\}. \end{aligned}$$

If, in addition,  $B \subset T^*Y$  then

$$(A.5) \quad A \circ B = \{(x, \xi) \in T^*X \mid \exists (y, \eta) \in B \text{ such that } (x, y; \xi, \eta) \in A\}$$

**THEOREM A.2.**  *$R$  is an elliptic Fourier integral operator (FIO) of order  $-1/2$  and with canonical relation*

$$(A.6) \quad C = \{(x, \varphi, s; a(\theta\mathbf{dx} + \mathbf{ds} - (x \cdot \theta^\perp)\mathbf{d}\varphi)) \mid a \neq 0, x \cdot \theta = s\}.$$

The projection  $p_Y : C \rightarrow T^*Y$  is an injective immersion. Therefore,  $C$  is a local canonical graph [77, Chapter VIII, §6].

$R^*$  is an elliptic Fourier integral operator of order  $-1/2$  and with local canonical graph  $C^t$ .

The canonical relation of a FIO is used to tell the microlocal properties of a FIO including what it does to singularities, and we will use it in our calculation in our proof of Theorem 2.8 at the end of the appendix.

**PROOF.** First, we will use the Fourier-Slice Theorem 2.3 to show  $R$  is an elliptic FIO. Define the ‘‘polar projection’’  $J : C^\infty(\mathbb{R}^2) \rightarrow C^\infty([0, 2\pi] \times \mathbb{R})$  by  $Jf(\varphi, s) = f(s\theta)$ . We take the one-dimensional inverse Fourier transform in (2.10):

$$\begin{aligned} (A.7) \quad Rf(\varphi, s) &= \mathcal{F}_s^{-1} \mathcal{F}_s Rf(\varphi, s) \\ &= \sqrt{2\pi} (\mathcal{F}_s^{-1} \circ (J \circ \mathcal{F}_{x \rightarrow \xi}))(f)(\varphi, s) \\ &= \frac{1}{2\pi} \int_{\sigma=-\infty}^{\infty} \int_{x \in \mathbb{R}^2} e^{i(s-\theta \cdot x)\sigma} f(x) dx d\sigma. \end{aligned}$$

Of course, this integral does not converge absolutely, although the operation  $\mathcal{F}_{\sigma \rightarrow p}^{-1} \circ J \circ \mathcal{F}_{x \rightarrow \xi}$  is defined for  $f \in C_c^\infty(\mathbb{R}^2)$ . To make the integrals converge, one does integrations by parts as with Fourier integral operators in general to show (A.7) can be defined on distributions [30, 76]. Equation (A.7) shows that  $R$  is an elliptic Fourier integral operator with phase function  $\phi(x, \varphi, s, \sigma) = (s - \theta \cdot x)\sigma$  and amplitude  $1/2\pi$ .

To calculate the canonical relation for  $R$ , we follow the general methods in [30, p. 165] or [77, (6.1) p. 462]. We first calculate the differentials of  $\phi$ ,

$$(A.8) \quad d_x \phi = -\sigma \theta \mathbf{dx}, \quad d_y \phi = \sigma (\mathbf{ds} - x \cdot \theta^\perp \mathbf{d}\varphi), \quad d_\sigma \phi = (s - x \cdot \theta) \mathbf{d}\sigma.$$

Next we define an auxiliary manifold  $\Sigma_\phi$ , the set of points at which  $d_\sigma \phi = 0$ :

$$(A.9) \quad \Sigma_\phi = \{(x, \varphi, s, \sigma) \in \mathbb{R}^2 \times Y \times (\mathbb{R} \setminus \{0\}) \mid s - x \cdot \theta = 0\}.$$

The set

$$(A.10) \quad Z = \{(x, \varphi, s) \in \mathbb{R}^2 \times Y \mid s - x \cdot \theta = 0\}$$

is called the *incidence relation* of  $R$  because it is the set of all  $(x, \varphi, s)$  with  $x \in L(\varphi, s)$ , and it is the projection of  $\Sigma_\phi$  on the first coordinates.

Note that the conditions for  $\phi$  to be a nondegenerate phase function [77, (2.2)-(2.4), p. 315] hold because  $d_x \phi$  and  $d_y \phi$  are not zero for  $\sigma \neq 0$ . Therefore  $R$  is a

Fourier integral operator.  $R$  has order  $-1/2$  because its symbol  $1/2\pi$  is homogeneous of degree zero,  $2 = \dim \mathbb{R}^2 = \dim Y$ , and  $\sigma$  is one dimensional (see [77, p. 462 under (6.3)]). Since the symbol,  $1/2\pi$ , is homogeneous and nowhere zero,  $R$  is elliptic.

The canonical relation,  $C$  associated to  $R$  is defined by the map

$$\Sigma_\phi \ni (x, \varphi, s, \sigma) \mapsto (x, \varphi, s; d_x \phi, -d_y \phi) \in C.$$

Therefore,

$$(A.11) \quad C = \{(x, \varphi, s; -\sigma(\theta \mathbf{dx} + \mathbf{ds} - (x \cdot \theta^\perp) \mathbf{d}\varphi)) \mid s - x \cdot \theta = 0, \sigma \neq 0\}.$$

The equation (A.6) is gotten from (A.11) by letting  $a = -\sigma$ .

To show the projection  $p_Y : C \rightarrow T^*Y$  is an injective immersion, we need only observe that the second coordinates in (A.6),  $(\varphi, s, a(\mathbf{ds} - x \cdot \theta^\perp \mathbf{d}\varphi))$ , determine the factor  $a$  (from the  $\mathbf{ds}$  coordinate) and since  $a \neq 0$ , these coordinates smoothly determine  $x$  by  $x = (x \cdot \theta^\perp)\theta^\perp + s\theta$ . Since this projection is an injective immersion, the projection to  $T^*\mathbb{R}^2$  is also an immersion. This can also be seen by a direct calculation showing that  $p_X : C \rightarrow T^*\mathbb{R}^2$  is a two-to-one immersion (see Remark A.4). Since these projections are immersions,  $C$  is a local canonical graph and the properties of FIO associated to local canonical graphs are easier to prove (see the discussion in [77, Chapter VIII, §6]).

The proof for  $R^*$  is similar and will be outlined. We let  $g \in C_c^\infty(Y)$  and calculate  $\mathcal{F}_s^{-1} \mathcal{F}_s g$  evaluated at  $(\varphi, x \cdot \theta)$  and finally integrate with respect to  $\varphi$ :

$$(A.12) \quad \begin{aligned} R^* g(x) &= \int_{\varphi \in [0, 2\pi]} g(\varphi, x \cdot \theta) d\varphi \\ &= \int_{\varphi \in [0, 2\pi]} \int_{\sigma=-\infty}^{\infty} \int_{s=-\infty}^{\infty} \frac{e^{i(x \cdot \theta)\sigma}}{\sqrt{2\pi}} \frac{e^{-is\sigma}}{\sqrt{2\pi}} g(\varphi, s) ds d\sigma d\varphi \\ &= \int_{\varphi \in [0, 2\pi]} \int_{\sigma=-\infty}^{\infty} \int_{s=-\infty}^{\infty} e^{i((x \cdot \theta)-s)\sigma} \frac{1}{2\pi} g(\varphi, s) ds d\sigma d\varphi. \end{aligned}$$

Note that the integrals can be made to converge using integration by parts arguments as with (A.7). Thus, the phase function of  $R^*$  is  $\tilde{\phi}(\varphi, s, x, \sigma) = (x \cdot \theta - s)\sigma$  and the arguments showing  $R$  is an elliptic FIO of order  $-1/2$  associated to  $C$  show  $R^*$  is an elliptic FIO of order  $-1/2$  associated to  $C^t$ .  $\square$

**REMARK A.3.** It should be pointed out that since  $R$  is a FIO associated to  $C$ , it is immediate that its adjoint is a FIO associated to  $C^t$  [30, Theorem 4.2.1], however we gave our direct proof for  $R^*$  because it is so elementary. Note also that the Schwartz Kernel of  $R$  (and of  $R^*$ ) as a distribution on  $\mathbb{R}^2 \times Y$  is the set  $Z$  in (A.10). Such distributions (including all Radon transforms [22]) are called conormal distributions, and their properties as Fourier integral distributions are especially simple. Note that the Lagrangian manifold of a Fourier integral operator is just the “prime,” (A.5), of its canonical relation, and sometimes one associates FIO to their Lagrangian manifolds rather than their canonical relations. In fact, the Lagrangian manifold of  $R$  is  $C' = N^*Z \setminus \{0\}$  (e.g., [62, pp. 335-337]).

We showed that the projection from  $C$  to  $T^*Y$  is an injective immersion, and this assumption is called the *Bolker Assumption* [24, pp. 364-365], [62, equation (9)]. If a generalized Radon transform satisfies this assumption, then one can compose  $R^*$  and  $R$  and show  $R^*R$  is an elliptic pseudodifferential operator. See

[29, Theorem 4.2.2], and discussion at the bottom of p. 180 for how to compose FIO and [24] [62] for this specific result.

PROOF OF THEOREM 2.8. The relationship (2.19) follows from the following general fact about what FIO do to wavefront sets. Let  $X$  and  $Y$  be manifolds and let  $S$  be a Fourier integral operator of order  $m$  associated to canonical relation  $C \subset (T^*X \setminus \{0\}) \times (T^*Y \setminus \{0\})$ . Let  $f \in \mathcal{E}'(X)$ , and  $s \in \mathbb{R}$ . Then, there is a natural relation between singularities of  $f$  and those of  $Sf$ :

$$(A.13) \quad \text{WF}^{\alpha-m}(Sf) \subset (C^t) \circ \text{WF}^\alpha(f).$$

Relation (A.13) for the  $C^\infty$  wavefront set is known (e.g., [77, Theorem 5.4, p. 461]). Sobolev continuity of  $S$  from  $H_c^\alpha(X)$  to  $H_{\text{loc}}^{\alpha-m}(Y)$  is also known [77, Theorem 6.1, p. 466], and elementary proofs exist for global Sobolev continuity for  $R$  and  $R^{-1}$  (e.g., [41, 25, 52]). To prove (A.13) for Sobolev wavefront, one uses pseudodifferential operators of order zero to microlocalize near a specific cotangent direction in  $T^*X$ ,  $L_X$  and in directions in  $T^*Y$  that correspond by  $C^t$ ,  $L_Y$ . Then, the microlocalized operator  $L_Y S L_X$  is of order  $m$  and smoothing away from these directions.

We apply this theorem to both  $R$  and  $R^*$ . Let  $\alpha \in \mathbb{R}$ . First, since  $R$  is a FIO associated to the local canonical graph  $C$ , (A.6), by (A.13)

$$\text{WF}^{\alpha+1/2}(Rf) \subset (C^t) \circ \text{WF}^\alpha(f).$$

Since  $R^* R$  is an elliptic pseudodifferential operator of order  $-1$ ,

$$\text{WF}^\alpha(f) = \text{WF}^{\alpha+1}(R^* R f)$$

[77, Proposition 6.10, p. 70]. Applying (A.13) to  $R^*$  and then to  $R$  we see

$$\text{WF}^\alpha(f) = \text{WF}^{\alpha+1}(R^* R f) \subset C \circ \text{WF}^{\alpha+1/2}(Rf) \subset C \circ C^t \text{WF}^\alpha(f) = \text{WF}^\alpha(f).$$

Here we have used that  $C \circ C^t = \text{Id}$  because of the Bolker assumption:  $p_Y$  is an injective immersion. This proves

$$(A.14) \quad \text{WF}^\alpha(f) = C \circ \text{WF}^{\alpha+1/2}(Rf).$$

If you trace back (A.14) using the expression for  $C$ , you get (2.19). Then, the covector  $(\theta_0, s_0; a\eta_0)$  in (2.19) determines  $(x_0, a\theta_0 \mathbf{d}\mathbf{x})$  by the Bolker assumption. Note that (A.14) is an equality, so other wavefront directions (those not conormal to  $L(\varphi_0, s_0)$  or not over points on  $L(\varphi_0, s_0)$ ) will not be visible.  $\square$

REMARK A.4. Note that two points in  $T^*([0, 2\pi] \times \mathbb{R})$  correspond to  $(x_0, a\theta_0)$  in (2.19). We will show they are caused by the ambiguity in parametrization of lines, (2.4). They are the  $T^*([0, 2\pi] \times \mathbb{R})$  coordinates of the two points in  $C$  that include  $(x_0, a\theta_0 \mathbf{d}\mathbf{x})$ :

$$\begin{aligned} & (x_0, \varphi_0, s_0; a(\theta_0 \mathbf{d}\mathbf{x} + \mathbf{d}s - (x_0 \cdot \theta_0^\perp) \mathbf{d}\varphi)) \text{ and} \\ & (x_0, \varphi_0 + \pi, -s_0; (-a)(\theta(\varphi_0 + \pi) \mathbf{d}\mathbf{x} + \mathbf{d}s - (x_0 \cdot \theta^\perp(\varphi_0 + \pi)) \mathbf{d}\varphi)) \\ & = (x_0, \varphi_0 + \pi, -s_0; (-a)(-\theta_0 \mathbf{d}\mathbf{x} + \mathbf{d}s + (x_0 \cdot \theta_0^\perp) \mathbf{d}\varphi)). \end{aligned}$$

These points in  $T^*Y$  are  $(\varphi_0, s_0; a\mathbf{d}s - a(x_0 \cdot \theta_0^\perp) \mathbf{d}\varphi)$  and  $(\varphi_0 + \pi, -s_0; -a\mathbf{d}s - a(x_0 \cdot \theta_0^\perp) \mathbf{d}\varphi)$  and they are mapped into each other under the map in (2.4),  $(\varphi, s) \mapsto (\varphi + \pi, -s)$ , since  $\mathbf{d}\varphi$  stays the same under translation  $\varphi \mapsto \varphi + \pi$  and  $\mathbf{d}s$  changes to  $-\mathbf{d}s$  under the reflection  $p \mapsto -p$ . Since the Radon transform is invariant under

(2.4),  $Rf$  either has wavefront at both points or at neither point, so there is no ambiguity and (2.19) is valid in both directions.

## References

- [1] M. D. Altschuler, *Reconstruction of the global-scale three-dimensional solar corona*, Topics in Applied Physics, vol. 32, pp. 105–145, Springer-Verlag, New York/Berlin, 1979.
- [2] R.H.T. Bates and R.M. Lewitt, *Image reconstruction from projections: I: General theoretical considerations, II: Projection completion methods (theory), III: Projection completion methods (computational examples)*, Optik **50** (1978), I: 19–33, II: 189–204, III: 269–278.
- [3] Gregory Beylkin, *The inversion problem and applications of the generalized Radon transform*, Comm. Pure Appl. Math. **37** (1984), 579–599.
- [4] Jan Boman and Eric Todd Quinto, *Support theorems for real analytic Radon transforms on line complexes in  $\mathbb{R}^3$* , Trans. Amer. Math. Soc. **335** (1993), 877–890.
- [5] E.J. Candès and D. L. Donoho, *Curvelets and Reconstruction of Images from Noisy Radon Data*, Wavelet Applications in Signal and Image Processing VIII (M. A. Unser A. Aldroubi, A. F. Laine, ed.), vol. Proc. SPIE 4119, 2000.
- [6] Emmanuel Candès and Laurent Demanet, *Curvelets and Fourier Integral Operators*, C. R. Math. Acad. Sci. Paris. Serie I **336** (2003), no. 5, 395–398.
- [7] Allan M. Cormack, *Representation of a function by its line integrals with some radiological application*, J. Appl. Physics **34** (1963), 2722–2727.
- [8] ———, *Representation of a function by its line integrals with some radiological applications II*, J. Appl. Physics **35** (1964), 2908–2913.
- [9] M.E. Davison, *The ill-conditioned nature of the limited angle tomography problem*, SIAM J. Appl. Math. **43** (1983), 428–448.
- [10] J. DeStefano and T. Olsen, *Wavelet localization of the Radon transform in even dimensions*, IEEE Trans. Signal Proc. **42** (1994), 2055–2067.
- [11] Charles L. Epstein, *Introduction to the Mathematics of Medical Imaging*, Prentice Hall, Upper Saddle River, NJ, USA, 2003.
- [12] Adel Faridani, *Tomography and Sampling Theory*, The Radon Transform and Applications to Inverse Problems (Providence, RI, USA), AMS Proceedings of Symposia in Applied Mathematics, American Mathematical Society, 2006.
- [13] Adel Faridani, David Finch, E. L. Ritman, and Kennan T. Smith, *Local tomography, II*, SIAM J. Appl. Math. **57** (1997), 1095–1127.
- [14] Adel Faridani and E. L. Ritman, *High-resolution computed tomography from efficient sampling*, Inverse Problems **16** (2000), 635–650.
- [15] Adel Faridani, E. L. Ritman, and Kennan T. Smith, *Local tomography*, SIAM J. Appl. Math. **52** (1992), 459–484.
- [16] David Victor Finch, Ih-Ren Lan, and Gunther Uhlmann, *Microlocal Analysis of the Restricted X-ray Transform with Sources on a Curve*, Inside Out, Inverse Problems and Applications (Gunther Uhlmann, ed.), MSRI Publications, vol. 47, Cambridge University Press, 2003, pp. 193–218.
- [17] Joachim Frank, *Three-dimensional electron microscopy of macromolecular assemblies*, Academic Press, San Diego, 1996.
- [18] Israel M. Gelfand, M. I. Graev, and N. Ya. Vilenkin, *Generalized Functions*, vol. 5, Academic Press, New York, 1966.
- [19] A. B. Goncharov, *Integral geometry and three-dimensional reconstruction of randomly oriented identical particles from their electron microphotos*, Acta Applicandae Mathematicae **11** (1988), 199–211.
- [20] Allan Greenleaf and Gunther Uhlmann, *Non-local inversion formulas for the X-ray transform*, Duke Math. J. **58** (1989), 205–240.
- [21] ———, *Microlocal techniques in integral geometry*, Contemporary Math. **113** (1990), 121–136.
- [22] Victor Guillemin, *Some remarks on integral geometry*, Tech. report, MIT, 1975.
- [23] ———, *On some results of Gelfand in integral geometry*, Proceedings Symposia Pure Math. **43** (1985), 149–155.
- [24] Victor Guillemin and Shlomo Sternberg, *Geometric Asymptotics*, American Mathematical Society, Providence, RI, 1977.

- [25] Marjorie G. Hahn and Eric Todd Quinto, *Distances between measures from 1-dimensional projections as implied by continuity of the inverse Radon transform*, Zeitschrift Wahrscheinlichkeit **70** (1985), 361–380.
- [26] Sigurdur Helgason, *The Radon transform on Euclidean spaces, compact two-point homogeneous spaces and Grassmann manifolds*, Acta Math. **113** (1965), 153–180.
- [27] ———, *The Radon Transform, Second Edition*, Birkhäuser, Boston, 1999.
- [28] W. Hoppe and R. Hegerl, *Three-dimensional structure determination by electron microscopy*, Computer Processing of Electron Microscope Images, Springer Verlag, New York, 1980.
- [29] Lars Hörmander, *Fourier Integral Operators, I*, Acta Mathematica **127** (1971), 79–183.
- [30] ———, *The Analysis of Linear Partial Differential Operators*, vol. I, Springer Verlag, New York, 1983.
- [31] V. Isakov, *Inverse Problems for Partial Differential Equations*, Applied Mathematical Sciences, vol. 127, Springer-Verlag, New York, 1998.
- [32] Avinash C. Kak and Malcolm Slaney, *Principles of Computerized Tomographic Imaging*, Classics in Applied Mathematics, vol. 33, SIAM, Philadelphia, PA, 2001.
- [33] Alexander I. Katsevich, *Cone beam local tomography*, SIAM J. Appl. Math. **59** (1999), 2224–2246.
- [34] Alexander I. Katsevich and Alexander G. Ramm, *A method for finding discontinuities from the tomographic data*, Lectures in Appl. Math. **30** (1994), 105–114.
- [35] ———, *Pseudolocal tomography*, SIAM J. Appl. Math. **56** (1996), 167–191.
- [36] Peter Kuchment, *Generalized Transforms of Radon Type and their Applications*, The Radon Transform and Applications to Inverse Problems (Providence, RI, USA), AMS Proceedings of Symposia in Applied Mathematics, American Mathematical Society, 2006.
- [37] Peter Kuchment, Kirk Lancaster, and Ludmilla Mogilevskaya, *On local tomography*, Inverse Problems **11** (1995), 571–589.
- [38] Peter Kuchment and Eric Todd Quinto, *Some problems of integral geometry arising in tomography*, The Universality of the Radon Transform, by Leon Ehrenpreis, Oxford University Press, London, 2003.
- [39] I.-R. Lan, *On an operator associated to a restricted x-ray transform*, Ph.D. thesis, Oregon State University, 2000.
- [40] Serguei Lissianoi, *On stability estimates in the exterior problem for the Radon transform*, Lecture Notes in Appl. Math., vol. 30, pp. 143–147, Amer. Math. Soc., Providence, RI, USA, 1994.
- [41] Alfred K. Louis, *Analytische Methoden in der Computer Tomographie*, Universität Münster, 1981, Habilitationsschrift.
- [42] ———, *Ghosts in tomography, the null space of the Radon transform*, Mathematical Methods in the Applied Sciences **3** (1981), 1–10.
- [43] ———, *Incomplete data problems in X-ray computerized tomography I. Singular value decomposition of the limited angle transform*, Numerische Mathematik **48** (1986), 251–262.
- [44] ———, *Development of Algorithms in CT*, The Radon Transform and Applications to Inverse Problems (Providence, RI, USA) (Gestur Olafsson and Eric Todd Quinto, eds.), AMS Proceedings of Symposia in Applied Mathematics, American Mathematical Society, 2006.
- [45] Alfred K. Louis and Peter Maas, *Contour reconstruction in 3-D X-Ray CT*, IEEE Trans. Medical Imaging **12** (1993), no. 4, 764–769.
- [46] Alfred K. Louis and Andreas Rieder, *Incomplete data problems in X-ray computerized tomography II. Truncated projections and region-of-interest tomography*, Numerische Mathematik **56** (1986), 371–383.
- [47] Peter Maas, *3D Röntgentomographie: Ein Auswahlkriterium für Abtastkurven*, Zeitschrift für angewandte Mathematik und Mechanik **68** (1988), T 498–T499.
- [48] ———, *The interior Radon transform*, SIAM J. Appl. Math. **52** (1992), 710–724.
- [49] Wolodymir R. Madych and Stewart A. Nelson, *Polynomial based algorithms for computed tomography*, SIAM J. Appl. Math. **43** (1983), 157–185.
- [50] W.R. Madych, *Tomography, approximate reconstruction, and continuous wavelet transforms*, Appl. Comput. Harmon. Anal. **7** (1999), no. 1, 54–100.
- [51] Frank Natterer, *Efficient implementation of ‘optimal’ algorithms in computerized tomography*, Math. Methods in the Appl. Sciences **2** (1980), 545–555.
- [52] ———, *The mathematics of computerized tomography*, Classics in Mathematics, Society for Industrial and Applied Mathematics, New York, 2001.

- [53] Frank Natterer and Frank Wübbeling, *Mathematical Methods in image Reconstruction*, Monographs on Mathematical Modeling and Computation, Society for Industrial and Applied Mathematics, New York, 2001.
- [54] Clifford J. Nolan and Margaret Cheney, *Synthetic Aperture inversion*, Inverse Problems **18** (2002), no. 1, 221–235.
- [55] ———, *Microlocal Analysis of Synthetic Aperture Radar Imaging*, Journal of Fourier Analysis and Applications **10** (2004), 133–148.
- [56] Ozan Öktem, *Mathematical problems related to the research at CMB and SIDEC Technologies AB*, Tech. report, SIDEC Technologies, AB, Stockholm, 2001.
- [57] Victor Palamodov, *Nonlinear artifacts in tomography*, Soviet Physics Doklady **31** (1986), 888–890.
- [58] ———, *Reconstruction from Limited Data of Arc Means*, J. Fourier Analysis and Applications **6** (2000), 26–42.
- [59] ———, *Reconstructive Integral Geometry*, Monographs in Mathematics, Birkhauser, Boston, MA, USA, 2004.
- [60] R. M. Perry, *On reconstructing a function on the exterior of a disc from its Radon transform*, J. Math. Anal. Appl. **59** (1977), 324–341.
- [61] Bent Petersen, *Introduction to the Fourier Transform and Pseudo-Differential Operators*, Pittman, Boston, 1983.
- [62] Eric Todd Quinto, *The dependence of the generalized Radon transform on defining measures*, Trans. Amer. Math. Soc. **257** (1980), 331–346.
- [63] ———, *Tomographic reconstructions from incomplete data—numerical inversion of the exterior Radon transform*, Inverse Problems **4** (1988), 867–876.
- [64] ———, *Computed tomography and rockets*, Mathematical Methods in Tomography, Lecture Notes in Mathematics, vol. 1497, Springer Verlag, 1990, pp. 261–268.
- [65] ———, *Limited data tomography in non-destructive evaluation*, IMA Volumes in Mathematics and its Applications, vol. 23, pp. 347–354, Springer-Verlag, 1990.
- [66] ———, *Singularities of the X-ray transform and limited data tomography in  $\mathbb{R}^2$  and  $\mathbb{R}^3$* , SIAM J. Math. Anal. **24** (1993), 1215–1225.
- [67] ———, *Exterior and Limited Angle Tomography in Non-destructive Evaluation*, Inverse Problems **14** (1998), 339–353.
- [68] ———, *Local Algorithms in Exterior Tomography*, Journal of Computational and Applied Mathematics (2005), to appear.
- [69] Johann Radon, *Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten*, Ber. Verh. Sach. Akad. **69** (1917), 262–277.
- [70] G.N. Ramachandran and A.V. Lakshminarayanan, *Three dimensional reconstruction from radiographs and electron micrographs: applications of convolutions instead of Fourier Transforms*, Proc. Natl. Acad. Sci. USA **68** (1971), 262–277.
- [71] Alexander G. Ramm and Alexander I. Zaslavsky, *Singularities of the Radon transform*, Bull. Amer. Math. Soc. **25** (1993), 109–115.
- [72] F. Rashid-Farrokhi, K.J.R. Liu, Carlos A. Berenstein, and David Walnut, *Wavelet-based multiresolution local tomography*, IEEE Trans. Image Processing **6** (1997), 1412–1430.
- [73] Lawrence A. Shepp and J. B. Kruskal, *Computerized Tomography: The new medical X-ray technology*, Amer. Mathematical Monthly **85** (1978), 420–439.
- [74] Lawrence A. Shepp and S. Srivastava, *Computed tomography of PKM and AKM exit cones*, A. T & T. Technical J. **65** (1986), 78–88.
- [75] Kennan T. Smith, Donald C. Solomon, and Solomon L. Wagner, *Practical and mathematical aspects of the problem of reconstructing objects from radiographs*, Bull. Amer. Math. Soc. **83** (1977), 1227–1270.
- [76] O. J. Tretiak, *Attenuated and exponential Radon transforms*, Proceedings Symposia Appl. Math. **27** (1982), 25–34.
- [77] F. Treves, *Introduction to Pseudodifferential and Fourier integral operators*, Volumes 1 & 2, Plenum Press, New York, 1980.

- [78] E.I. Vainberg, I. A. Kazak, and V. P. Kurozaev, *Reconstruction of the internal three-dimensional structure of objects based on real-time integral projections*, Soviet Journal of Nondestructive Testing **17** (1981), 415–423.

DEPARTMENT OF MATHEMATICS, TUFTS UNIVERSITY, MEDFORD, MA 02155, USA  
*E-mail address:* Todd.Quinto@tufts.edu    *Web Site:* <http://www.tufts.edu/~equinto>

*This page intentionally left blank*

## Development of Algorithms in Computerized Tomography

Alfred K. Louis

*This paper is dedicated to Hermann and Dr. Charlotte Deutsch.*

**ABSTRACT.** In this paper we present a general approach to derive inversion algorithms for tomographic applications, the so-called approximate inverse. Three different techniques for calculating a reconstruction kernel are shown and applied to invert the Radon transform, to compute approximations in the limited angle problem and to solve the 3D cone beam reconstruction problem. Reconstructions from real data in this case are presented.

### 1. Introduction

In this paper we present some principles in designing inversion algorithms in tomography. We concentrate on linear problems arising in connection with the Radon and the x-ray transform. In the original 2D x-ray CT problem the Radon transform served as mathematical model. Here one integrates over lines and the problem is to recover a function from its line integrals. The same holds in the 3D x-ray case, but in 3D the Radon transform integrates over planes, in general over  $N - 1$  - dimensional hyper planes in  $\mathbb{R}^N$ . Hence here the so-called x-ray transform is the mathematical model. Further differences are in the parametrization of the lines. The 3D - Radon transform merely appears as tool to derive inversion formula. In the early days of MRI ( magnetic resonance imaging ), at those days called NMR, nuclear magnetic resonance, it served as a mathematical model, see for example Marr-Chen-Lauterbur [MCL81], but then, due to the limitations of computer power in those days one changed the measuring procedure and scanned the Fourier transform of the searched-for function in two dimensions. Nowadays the Radon transform reappeared, now in three and even four dimensions as mathematical model in EPRI ( electron parametric resonance imaging ) where spectral - spatial information is the goal, see e.g. Kuppusamy et al [KCSWZ95].

The paper is organized as following. We start with a general principle for reconstruction information from measured data, the so-called approximate inverse, see Louis [Lou96], Louis-Maass [LM90]. The well-known inversion of the Radon

---

1991 *Mathematics Subject Classification.* Primary 65R10, 65D25; Secondary 45L10, 44A15.

The author was supported in part by a Grant of the Hermann und Dr. Charlotte Deutsch Stiftung and by the Deutsche Forschungsgemeinschaft under grant LO 310/8-1.

transform is considered a model case for inversion. The singular value decomposition is then used to compute reconstruction kernels for the limited angle problem. Finally we consider a 3D x-ray problem and present reconstructions from real data.

## 2. Approximate Inverse as a Tool for Deriving Inversion Algorithms

The integral operators appearing in medical imaging are typically compact operators between suitable Hilbert spaces. The inverse operator of those compact operators with infinite dimensional range are not continuous, which means that the unavoidable data errors are amplified in the solution. Hence one has to be very careful in designing inversion algorithms. They have to balance the demand for highest possible accuracy and the necessary damping of the influence of the unavoidable data errors. From the theoretical point of view, exact inversion formulae are nice, but they do not take care of data errors. The way out of this dilemma is the use of approximate inversion formulas whose principles are explained in the following.

For approximating the solution of

$$Af = g$$

we apply the method of approximate inverse, see Louis [Lou96]. The basic idea works as follows: choose a so-called mollifier  $e_\gamma(x, y)$  which, for a fixed reconstruction point  $x$  is a function of the variable  $y$  and which approximates the delta distribution for the point  $x$ . The parameter  $\gamma$  acts as regularization parameter. Simply think in the case of one spatial variable  $x$  of

$$e_\gamma(x, y) = \frac{1}{2\gamma} \chi_{[x-\gamma, x+\gamma]}(y)$$

where  $\chi_\Omega$  denotes the characteristic function of  $\Omega$ . Then the mollifier fulfills

$$(2.1) \quad \int e_\gamma(x, y) dy = 1$$

for all  $x$  and the function

$$f_\gamma(x) = \int f(y) e_\gamma(x, y) dy$$

converges for  $\gamma \rightarrow 0$  to  $f$ . The larger the parameter  $\gamma$  the larger is the interval where the averaging takes place and hence the stronger is the smoothing. Now solve for fixed reconstruction point  $x$  the auxiliary problem

$$(2.2) \quad A^* \psi_\gamma(x, \cdot) = e_\gamma(x, \cdot)$$

where  $e_\gamma(x, \cdot)$  is the chosen approximation to the delta distribution for the point  $x$ , and put

$$\begin{aligned} f_\gamma(x) &= \langle f, e_\gamma(x, \cdot) \rangle \\ &= \langle f, A^* \psi_\gamma(x, \cdot) \rangle = \langle Af, \psi_\gamma(x, \cdot) \rangle = \langle g, \psi_\gamma(x, \cdot) \rangle \\ &=: S_\gamma g(x). \end{aligned}$$

The operator  $S_\gamma$  is called the approximate inverse and  $\psi_\gamma$  is the reconstruction kernel. To be precise it is the approximate inverse for approximating the solution  $f$  of  $Af = g$ . If we choose instead of  $e_\gamma$  fulfilling (2.1) a wavelet, then  $f_\gamma$  can be interpreted as a wavelet transform of  $f$ . Wavelet transforms are known to approximate in a certain sense derivatives of the transformed function  $f$ , see [LMR97].

Hence this is a possibility to find jumps in  $f$  as used in contour reconstructions, see [LM93].

The advantage of this method is that  $\psi_\gamma$  can be pre-computed independently of the data. Furthermore, invariances and symmetries of the operator  $A^*$  can directly be transformed into corresponding properties of  $S_\gamma$  as the following consideration shows, see Louis [Lou96]. Let  $T_1$  and  $T_2$  be two operators intertwining with  $A^*$

$$A^*T_2 = T_1A^*.$$

If we choose a standard mollifier  $E$  and solve  $A^*\Psi = E$  then the solution of Eq. (2.2) for the special mollifier  $e_\gamma = T_1E$  is given as

$$\psi_\gamma = T_2\Psi.$$

As an example we mention, that if  $A^*$  is translation invariant; i.e.,  $T_1f(x) = T_2f(x) = f(x - a)$ , then also the reconstruction kernel is translation invariant.

Sometimes it is easier to check these conditions for  $A$  itself. Using  $AT_1^* = T_2^*A$  we get the above relations by using the adjoint operators.

This method is presented in [Lou99] as general regularization scheme to solve inverse problems. Generalizations are also given. The application to vector fields is derived by Schuster [Sch00].

If the auxiliary problem is not solvable then its minimum norm solution leads to the minimum norm solution of the original problem.

There are several possibilities for solving (2.2). Besides the straight forward approach we can easily solve this equation if the inverse operator is of the form

$$A^{-1} = A^*B.$$

The solution of (2.2) is then given as

$$(2.3) \quad \psi_\gamma = BAe_\gamma$$

as the following equations show

$$A^*\psi_\gamma = e_\gamma = A^{-1}Ae_\gamma = A^*BAe_\gamma.$$

The next situation we consider is that a singular system of the operator is known. Assume the operator being compact as mapping from a Hilbert space  $X$  into a Hilbert space  $Y$ . Then the operator has a complete singular system consisting of normalized functions  $v_n \in X$ ,  $u_n \in Y$  and nonnegative numbers  $\sigma_n$  such that

$$Av_n = \sigma_n u_n, A^*u_n = \sigma_n v_n.$$

Information on singular value decomposition in connection with the Radon transform can be found in [Lou89], [Nat86]. The solution of (2.2) can then be calculated as

$$(2.4) \quad \psi_\gamma(x, y) = \sum_{n=0}^{\infty} \sigma_n^{-1} \langle e_\gamma(x, \cdot), v_n \rangle u_n(y).$$

Finally we want to mention the case, where the inverse is again given as

$$A^{-1} = A^*B$$

and we want to save the structure of this formula using the operator  $A^*$ . Here we start by first smoothing the data with a smoothing operator  $\tilde{M}_\gamma$  and then we invert, see [Lou99] Theorems 6 and 7. The approximate inverse has then the form

$$S_\gamma = A^{-1} \tilde{M}_\gamma.$$

We pre-compute a reconstruction kernel in the following way. Let  $w_\gamma(y, \cdot)$  be an approximation of the delta distribution on the data for the point  $y$ . Then we put

$$(2.5) \quad \phi_\gamma(y, \cdot) = B w_\gamma(y, \cdot)$$

and compute

$$f_\gamma = A^* B_\gamma g$$

with

$$B_\gamma g(y) = \langle g, \phi_\gamma(y, \cdot) \rangle.$$

### 3. Inversion of the Radon Transform

We apply the above approach to derive inversion algorithms for the Radon transform. The Radon transform in  $\mathbb{R}^N$  is defined as

$$\mathbf{R}f(\theta, s) = \int_{\mathbb{R}^N} f(x) \delta(s - x^\top \theta) dx$$

for unit vectors  $\theta \in S^{N-1}$  and  $s \in \mathbb{R}$ . Its inverse is

$$(3.1) \quad \mathbf{R}^{-1} = c_N \mathbf{R}^* I^{1-N}$$

where  $\mathbf{R}^*$  is the adjoint operator from  $L_2$  to  $L_2$ , also called the backprojection, defined as

$$\mathbf{R}g(x) = \int_{S^{N-1}} g(\theta, x^\top \theta) d\theta,$$

$I^\alpha$  is the Riesz potential defined via the Fourier transform as

$$\widehat{(I^\alpha g)}(\xi) = |\xi|^{-\alpha} \widehat{g}(\xi),$$

acting on the second variable of  $\mathbf{R}f$  and the constant

$$c_N = \frac{1}{2} (2\pi)^{1-N}.$$

see e.g. [Nat86]. We start with a mollifier  $e_\gamma(x, \cdot)$  for the reconstruction point  $x$  and get

$$\begin{aligned} \mathbf{R}^* \psi_\gamma(x, \cdot) &= e_\gamma(x, \cdot) \\ &= c_N \mathbf{R}^* I^{1-N} \mathbf{R} e_\gamma(x, \cdot) \end{aligned}$$

leading to

$$\psi_\gamma(x; \theta, s) = c_N I^{1-N} \mathbf{R} e_\gamma(x; \theta, s).$$

The Radon transform for fixed  $\theta$  is translational invariant; i.e., if we denote by  $\mathbf{R}_\theta f(s) = \mathbf{R}f(\theta, s)$ , then

$$\mathbf{R}_\theta T_1^a f = T_2^{a^\top \theta} \mathbf{R}_\theta f$$

with the shift operators  $T_1^a f(x) = f(x - a)$  and  $T_2^t g(s) = g(s - t)$ . If we chose a mollifier  $\bar{e}_\gamma$  supported in the unit ball centred around 0 that is shifted to  $x$  as

$$e_\gamma(x, y) = 2^{-N} \bar{e}_\gamma\left(\frac{x - y}{2}\right)$$

then also  $e_\gamma$  is supported in the unit ball and the reconstruction kernel fulfills

$$\psi_\gamma(x; \theta, s) = \frac{1}{2} \bar{\psi}_\gamma(\theta, \frac{s - x^\top \theta}{2})$$

as follows from the general theory in [Lou96] and as was used for the 2D case in [LS96].

Furthermore, the Radon transform is invariant under rotations; i.e.,

$$\mathbf{R}T_1^U = T_2^U \mathbf{R}$$

for the rotation  $T_1^U f(x) = f(Ux)$  with unitary  $U$  and  $T_2^U g(\theta, s) = g(U\theta, s)$ . If the mollifier is invariant under rotation; i.e.,

$$\bar{e}_\gamma(x) = \bar{e}_\gamma(\|x\|)$$

then the reconstruction kernel is independent of  $\theta$  leading to the following observation.

**THEOREM 3.1.** *Let the mollifier  $e_\gamma(x, y)$  be of the form*

$$e_\gamma(x, y) = 2^{-N} \bar{e}_\gamma(\|x - y\|/2)$$

*then the reconstruction kernel is a function only of the variable  $s$  and the algorithms is of filtered backprojection type*

$$(3.2) \quad f_\gamma(x) = \int_{S^{n-1}} \int_{\mathbb{R}} \psi_\gamma(x^\top \theta - s) \mathbf{R}f(\theta, s) ds d\theta.$$

We described here the approach mentioned in (2.3). First references to this technique can be found in the work of Grünbaum [DG81] and Solmon, [HS88].

#### 4. The Filtered Backprojection for the Radon Transform in 2 and 3 Dimensions

In the following we describe the derivation of the filtered backprojection, see Theorem 3.1, for two and three dimensions. As seen in Formula (3.1) the inverse operator of the Radon transform in  $\mathbb{R}^N$  has the representation

$$\mathbf{R}^{-1} = \mathbf{R}^* B$$

with

$$B = c_N I^{1-N}.$$

Hence we can apply Formula (2.2) for deriving reconstruction kernels. They then can be represented as

$$(4.1) \quad \psi_\gamma = c_N I^{1-N} \mathbf{R} e_\gamma.$$

As mollifier we choose a translational and rotational invariant function

$$\bar{e}_\gamma(x, y) = e_\gamma(\|x - y\|)$$

whose Radon transform then is a function of the variable  $s$  only. Taking the Fourier transform of Equation (4.1) we get

$$\begin{aligned} \hat{\psi}_\gamma(\sigma) &= c_N (\widehat{I^{1-N}(\mathbf{R} e_\gamma)})(\sigma) \\ &= \frac{1}{2} (2\pi)^{(1-N)/2} |\sigma|^{N-1} \hat{e}_\gamma(\sigma), \end{aligned}$$

where in the last step we have used the projection theorem

$$\hat{f}(\sigma\theta) = (2\pi)^{(1-N)/2} \widehat{\mathbf{R}_\theta f}(\sigma).$$

So, we can proceed in the following two ways. Either we prescribe the mollifier  $e_\gamma$ , where the Fourier transform is then computed to

$$\hat{e}_\gamma(\sigma) = \sigma^{1-N/2} \int_0^\infty e_\gamma(s) s^{N/2} J_{N/2-1}(s\sigma) ds$$

where  $J_\nu$  denotes the Bessel function of order  $\nu$ . On the other hand we prescribe

$$\hat{e}_\gamma(\sigma) = (2\pi)^{-N/2} F_\gamma(\sigma)$$

with a suitably chosen filter  $F_\gamma$  leading to

$$\hat{\psi}_\gamma(\sigma) = \frac{1}{2} (2\pi)^{1/2-N} |\sigma|^{N-1} F_\gamma(\sigma).$$

If  $F_\gamma$  is the ideal low-pass; i.e.,  $F_\gamma(\sigma) = 1$  for  $|\sigma| \leq \gamma$  and 0 otherwise, then the mollifier is easily computed as

$$e_\gamma(x, y) = (2\pi)^{-N/2} \gamma^N \frac{J_{N/2}(\gamma \|x - y\|)}{(\gamma \|x - y\|)^{N/2}}.$$

In the two-dimensional case the calculation of  $\psi$  leads to the so called RAM-LAK filter, which has the disadvantage to produce ringing artefacts due to the discontinuity in the Fourier domain.

More popular for 2D is the filter

$$F_\gamma(\sigma) = \begin{cases} \text{sinc} \frac{\sigma \pi}{2\gamma} & , \quad |\sigma| \leq \gamma, \\ 0 & , \quad |\sigma| > \gamma. \end{cases}$$

From this we compute the kernel  $\psi_\gamma$  by inverse Fourier transform to get for  $\gamma = \pi/h$  where  $h$  is the stepsize on the detector; i.e.,  $h = 1/q$  if we use  $2q + 1$  points on the interval  $[-1, 1]$  and  $s = s_\ell = \ell h$ ,  $\ell = -q, \dots, q$

$$\psi_\gamma(s_\ell) = \frac{\gamma^2}{\pi^4} \frac{1}{1 - 4\ell^2},$$

known as Shepp - Logan kernel.

The algorithm of filtered backprojection is a stable discretization of the above described method using the composite trapezoidal rule for computing the discrete convolution. Instead of calculating the convolution for all points  $\theta^\top x$  the convolution is evaluated for equidistant points  $\ell h$  and then a linear interpolation is applied. Nearest neighbour interpolation is not sufficiently accurate, higher order interpolation is not bringing any improvement because the interpolated functions are not smooth enough. Then the composite trapezoidal rule is used for approximating the backprojection. Here one integrates a periodic function, hence, as shown with the Euler- Maclaurin summation formula, this formula is highly accurate. The filtered backprojection then consists of two steps. Let the data  $\mathbf{R}f(\theta, s)$  be given for the directions  $\theta_j = (\cos \varphi_j, \sin \varphi_j)$ ,  $\varphi_j = \pi(j-1)/p$ ,  $j = 1, \dots, p$  and the values  $s_k = kh$ ,  $h = 1/q$  and  $k = -q, \dots, q$ .

*Step 1:* For  $j=1,\dots,p$ , evaluate the discrete convolutions

$$(4.2) \quad v_{j,\ell} = h \sum_{k=-q}^q \psi_\gamma(s_\ell - s_k) \mathbf{R}f(\theta_j, s_k), \quad \ell = -q, \dots, q.$$

*Step 2:* For each reconstruction point  $x$  compute the discrete backprojection

$$(4.3) \quad \tilde{f}(x) = \frac{2\pi}{p} \sum_{j=1}^p ((1-\eta)v_{j,\ell} + \eta v_{j,\ell+1})$$

where, for each  $x$  and  $j$ ,  $\ell$  and  $\eta$  are determined by

$$s = \theta_j^\top x, \ell \leq s/h < \ell + 1, \eta = s/h - \ell$$

see e.g. [Nat86].

In the three - dimensional case we can use the fact, that the operator  $I^{-2}$  is local,

$$I^{-2}g(\theta, s) = \frac{\partial^2}{\partial s^2}g(\theta, s).$$

If we want to keep this local structure in the discretization we choose

$$F_\gamma(\sigma) = 2(1 - \cos(h\sigma))/(h\sigma)^2$$

leading to

$$(4.4) \quad \psi_\gamma(s) = (\delta_\gamma - 2\delta_0 + \delta_{-\gamma})(s).$$

Hence, the application of this reconstruction kernel is nothing but the central difference quotient for approximating the second derivative. The corresponding mollifier then is

$$e_\gamma(y) = \begin{cases} (2\pi)^{-1}h^{-2}|y|^{-1} & , \text{ for } |y| < h, \\ 0 & , \text{ otherwise,} \end{cases}$$

see [Lou83]. The algorithm has the same structure as mentioned above for the 2D case.

In order to get reconstruction formulas for the fan beam geometry coordinate transforms can be used, the structure of the algorithms does not change.

## 5. The Reconstruction Kernel for the Limited Angle Transform

As an example for the use of the singular value decomposition for computing the reconstruction kernel we consider the so-called limited angle problem. For practical applications see the paper of Quinto, [Qui06] in this volume. Singular value decompositions are known for several of these integral transforms appearing in tomography. For the classical 2D Radon transform the early result of Cormack [Cor64] can be interpreted as singular value decomposition. For the Radon transform in arbitrary dimensions it is given in [Lou84], for the x-ray transform in arbitrary dimensions and the parallel geometry it was derived by Maass, [Maa87]. Singular value decompositions for the fan-beam geometry are simple modifications. In the 2D case it has been used by Louis-Rieder [LR89] for deriving algorithms for region-of-interest tomography. Quinto [Qui88] used a singular value decomposition for developing algorithms for the exterior problem.

The singular value decomposition can also be used to answer the question of resolution and the practically invisible objects. An approach which gives more insight is based on wavefront sets, see Quinto [Qui93]. In the following we want to use the svd for computing a reconstruction kernel, rather than using them directly on the data to compute approximations in the missing range as done in Louis [Lou80].

We first describe the limited angle problem and formulate its singular value decomposition, originally presented in Louis [Lou86], now using the notion introduced by Slepian [Sle78].

We consider the parallel geometry of x-ray CT where the mathematical model is the *Radon transform*. Let  $f$  be an  $L_2$ -function of compact support in the unit disk  $\Omega \subset \mathbb{R}^2$ , after a possible rescaling. Denote by  $\theta \in S^1$  the unit vector

$$\theta = \theta(\varphi) = (\cos \varphi, \sin \varphi)^\top$$

and by  $\theta^\perp = \theta(\varphi + \pi/2) = (-\sin \varphi, \cos \varphi)^\top$  the vector orthogonal to  $\theta$ .

The Radon transform is defined as

$$\begin{aligned} \mathbf{R}f(\theta, s) &= \int_{\mathbb{R}^2} f(x) \delta(s - x^\top \theta) dx \\ &= \int_{\mathbb{R}} f(s\theta + t\theta^\perp) dt. \end{aligned}$$

We assume  $\mathbf{R}f(\theta, \cdot)$  to be given for all  $\theta$  with

$$\theta \in S_\phi := \{\theta(\varphi) : |\varphi| \leq \phi\} \cup \{\theta(\varphi) : |\varphi - \pi| \leq \phi\} \subset S^1$$

with  $0 < \phi < \frac{\pi}{2}$ . Note that in [Lou86] this was the missing range, hence the results change accordingly. Due to the symmetry of the Radon transform it suffices to know the data on one of the subsets of  $S_\phi$ .

With the truncated cylinder

$$Z_\phi = S_\phi \times [-1, 1]$$

and the weight

$$w(s) = (1 - s^2)^{1/2}$$

we consider

$$\mathbf{R}_\phi : L_2(\Omega) \rightarrow L_2(Z_\phi, w^{-1})$$

where the scalar product in the latter space is defined as

$$\langle f, g \rangle_{w^{-1}} = \int_{-\Phi}^{\Phi} \int_{-1}^1 w^{-1}(s) f(\varphi, s) g(\varphi, s) ds d\varphi.$$

In a first step we present the singular value decomposition of the limited angle transform  $\mathbf{R}_\phi$ . Note that  $\mathbf{R}_{\pi/2} = \mathbf{R}$ .

We make use of the following notion, introduced by Slepian [Sle78].

Let the  $N \times N$  matrix  $\rho(N, \phi/\pi)$  be given as

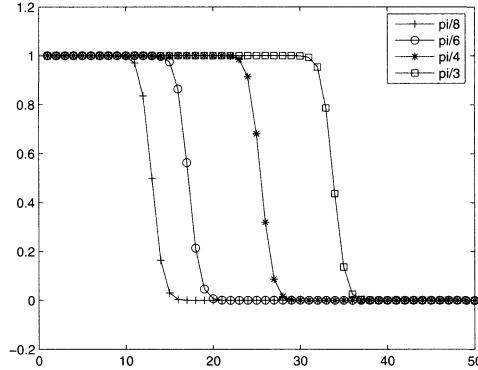
$$\rho(N, \phi/\pi)_{mn} = \frac{\sin 2\phi(m-n)}{\pi(m-n)}, \quad m, n = 0, \dots, N-1.$$

with the diagonal elements  $2\phi/\pi$ .

Let  $\lambda_k(N, \phi/\pi)$ ,  $k = 0, \dots, N-1$  be its eigenvalues and the

$$v^{(k)}(N, \phi/\pi) \in \mathbb{R}^N, \quad k = 0, \dots, N-1$$

its with respect to the Euclidean norm normalized eigenvectors. They are sections of the discrete prolate spheroidal sequences and are related to the *discrete spheroidal wave functions*  $u_k(N, \phi/\pi, \varphi/\pi)$  via

FIGURE 1. Eigenvalues of the Slepian matrix for different values of  $\Phi$ .

$$(5.1) \quad u_k(N, W; f) = \varepsilon_k \sum_{n=0}^{N-1} v_n^{(k)}(N, W) e^{-i\pi(N-1-2n)f}, \quad k = 0, \dots, N-1,$$

where

$$\varepsilon_k = \begin{cases} 1 & , \quad k \text{ even}, \\ i & , \quad k \text{ odd}, \end{cases}$$

see [Sle78, Formula (26)].

Finally we denote by  $U_m$  the Chebychev polynomials of the second kind and by

$$Q_{ml}(r) = \left(2(m+1)\right)^{1/2} r^l P_{(m-l)/2}^{(0,l)}(2r^2 - 1)$$

the normalized Zernike polynomials with  $P_k^{(\alpha,\beta)}$  the Jacobi polynomials.

**THEOREM 5.1.** *Let  $\mathbf{R}_\phi : L_2(\Omega) \rightarrow L_2(Z_\phi, w^{-1})$ . Then  $(f_{ml}^\phi, g_{ml}^\phi; \sigma_{ml}^\phi)$ ,  $m \geq 0$ ,  $0 \leq l \leq m$  with*

$$\begin{aligned} f_{ml}^\phi(r\omega(\vartheta)) &= (2\pi)^{-1/2} \sum_{\mu=0}^m v_\mu^{(l)}(m+1, \phi/\pi) Q_{m,|2\mu-m|}(r) e^{i(2\mu-m)\vartheta}, \\ g_{ml}^\phi(\theta(\varphi), s) &= \frac{1}{\pi} w(s) U_m(s) \varepsilon_l^{-1} u_l(m+1, \phi/\pi, \varphi/\pi) \lambda_l(m+1, \phi/\pi)^{-1/2} \\ \sigma_{ml}^\phi &= 2 \left( \frac{\pi}{m+1} \lambda_l(m+1, \phi/\pi) \right)^{1/2} \end{aligned}$$

form a complete singular system for the limited angle problem.

PROOF. The proof follows [Lou86] and is sketched here. Using the singular value decomposition for the full range case, see e.g. [Lou84], [Nat86]

$$\begin{aligned} f_{ml}(r\omega(\vartheta)) &= (2\pi)^{-1/2} Q_{m,|2l-m|}(r) e^{i(2l-m)\vartheta}, \\ g_{ml}(\theta(\varphi), s) &= \frac{1}{\pi} w(s) U_m(s) e^{i(2l-m)\varphi}, \\ \sigma_{ml} = \sigma_m &= 2 \left( \frac{\pi}{m+1} \right)^{1/2} \end{aligned}$$

it remains to orthogonalize on the truncated cylinder.

The  $f_{ml}^\phi$  are a linear combination of the  $f_{ml}$ ,

$$f_{ml}^\phi = \sum_{\mu=0}^m v_\mu(\ell)(m+1, \Phi/\pi) f_{m\mu}$$

and hence they are complete on  $L_2(\Omega)$  due to the completeness and orthogonality of the functions  $f_{ml}$  and the orthonormality of the vectors  $v^{(\ell)}(m+1, \Phi/pi)$ ,  $\ell = 1, \dots, m$ . Then we expand  $\mathbf{R}^* \mathbf{R} f_{ml}^\phi$  in terms of these functions. The Fourier coefficients are

$$\begin{aligned} c_{nk} &= \langle \mathbf{R}_\phi^* \mathbf{R}_\phi f_{ml}^\phi, f_{nk}^\phi \rangle_{L_2(\Omega)} \\ &= \langle \mathbf{R}_\phi f_{ml}^\phi, \mathbf{R}_\phi f_{nk}^\phi \rangle_{L_2(Z_\phi, w^{-1})}. \end{aligned}$$

Using  $\mathbf{R} f_{ml} = \sigma_m g_{ml}$  we get

$$\begin{aligned} c_{nk} &= \frac{4}{\pi} \frac{1}{\sqrt{(m+1)(n+1)}} \underbrace{\int_{-1}^1 w(s) U_m(s) U_n(s) ds}_{=(\pi/2)\delta_{mn}} \times \\ &\quad \times \sum_{\mu=0}^m \sum_{\nu=0}^n v_\mu^{(l)}(m+1, \phi/\pi) v_\nu^{(k)}(n+1, \phi/\pi) \underbrace{\int_{S_\phi} e^{i(2\mu-m+n-2\nu)\varphi} d\varphi}_{=2\pi\rho(m+1, \phi/\pi)_{\mu\nu}} \\ &= \frac{4\pi}{m+1} \delta_{mn} v^{(l)}(m+1, \phi/\pi)^\top \rho(m+1, \phi/\pi) v^{(k)}(m+1, \phi/\pi) \\ &= \frac{4\pi}{m+1} \lambda_l(m+1, \phi/\pi) \delta_{mn} \delta_{kl}, \end{aligned}$$

which means that

$$\mathbf{R}_\phi^* \mathbf{R}_\phi f_{ml}^\phi = \frac{4\pi}{m+1} \lambda_l(m+1, \phi/\pi) f_{ml}^\phi = (\sigma_{ml}^\phi)^2 f_{ml}^\phi.$$

Using  $R_\phi f_{ml} = \sigma_{ml}^\phi g_{ml}^\phi$  and relation (5.1) complete the proof. □

The eigenvalues  $\lambda_l(m+1, \phi/\pi)$  show the well known behaviour that they are close to 1, if  $l < (m+1)\phi/\pi$  and the rest is close to zero, if they are ordered from large to small. Their exponential decay, see [Sle78], means that the problem is severely ill – posed . As a consequence some of the components of the solution, namely those belonging to small singular values, are practically invisible. For a more geometrical interpretation see Qunito's approach with the wave front sets,

[Qui06]. This also affects the computation of the reconstruction kernel, see the discussion after Theorem 5.2.

The aim is now to represent an approximation  $f_\gamma$  of the equation  $\mathbf{R}_\phi f = g$  in the form

$$f_\gamma(x) = \langle g, \tilde{\psi}_\gamma^\phi \rangle_{L_2(Z, w^{-1})}.$$

To this end we solve, according to (2.4)

$$\mathbf{R}_\phi^* \tilde{\psi}_\gamma^\phi = e_\gamma$$

for a prescribed mollifier resulting in  $f_\gamma = \langle f, e_\gamma \rangle$ .

The limited angle transform intertwines with two shift operators, namely

$$T_1^x f(y) = f(x - y)$$

and

$$T_2^t g(\omega, s) = g(\omega, s - t)$$

as

$$\mathbf{R}_\Phi T_1^x = T_2^{x^\top \omega} \mathbf{R}_\Phi.$$

This means that, if we choose a shift invariant mollifier  $e_\gamma(x, y) = T_1^x E_\gamma(y)$  with a standard mollifier  $E_\gamma$ , then the reconstruction kernel  $\tilde{\psi}_\gamma(x; \omega, s)$  is of the form

$$\tilde{\psi}_\gamma(x; \omega, s) = \bar{\psi}_\gamma(\omega, s - x^\top \omega),$$

see [Lou96].

This means that it suffices to compute the reconstruction kernel for one point, preferably  $x = 0$ , and than shift the kernel.

Using the singular value decomposition of the limited angle transform we can represent the minimum norm solution of

$$\mathbf{R}_\phi^* \bar{\psi}_\gamma = E_\gamma$$

as

$$(5.2) \quad \bar{\psi}_\gamma(\omega, s) = \sum_{m=0}^{\infty} \sum_{l=0}^m \frac{1}{\sigma_{ml}^\phi} \langle E_\gamma, f_{ml}^\phi \rangle_{L_2(\Omega)} g_{ml}^\phi(\omega, s).$$

Using polar coordinates the scalar products are computed as

$$\begin{aligned} \langle E_\gamma, f_{ml}^\phi \rangle_{L_2(\Omega)} &= (2\pi)^{-1/2} \sum_{\mu=0}^m v_\mu^{(l)}(m+1, \phi/\pi) \times \\ &\times \int_0^1 r Q_{m,|2\mu-m|}(r) \int_0^{2\pi} E_\gamma(r \omega(\vartheta)) e^{i(2\mu-m)\vartheta} d\vartheta dr. \end{aligned}$$

If the mollifier is radially symmetric, i.e.

$$E_\gamma(y) = \bar{E}_\gamma(|y|),$$

then we get

$$\begin{aligned} \langle E_\gamma, f_{ml}^\phi \rangle_{L_2(\Omega)} &= (2\pi)^{-1/2} \sum_{\mu=0}^m v_\mu^{(l)}(m+1, \phi/\pi) \times \\ &\quad \times \int_0^1 r Q_{m,|2\mu-m|}(r) \bar{E}_\gamma(r) dr \cdot 2\pi \delta_{\mu, m/2} \\ &= (2\pi)^{1/2} v_{m/2}^{(l)}(m+1, \phi/\pi) \int_0^1 r Q_{m,0}(r) \bar{E}_\gamma(r) dr \end{aligned}$$

for  $m$  even, and 0 otherwise.

Using the symmetry relation of the discrete spheroidal sequence

$$v_n^{(l)}(N, W) = (-1)^l v_{N-1-n}^{(l)}(N, W),$$

see [Sle78, Formula (23)], we observe

$$v_{m/2}^{(l)}(m+1, \phi/\pi) = 0 \quad \text{for } l \text{ odd}.$$

This means that both  $m$  and  $l$  have to be even.

From (5.2) then follows

$$\bar{\psi}_\gamma(\omega, s) = \sum_{\substack{m=0 \\ m \text{ even}}}^{\infty} \sum_{\substack{l=0 \\ l \text{ even}}}^m \frac{1}{\sigma_{ml}^\phi} \langle E_\gamma, f_{ml}^\phi \rangle_{L_2(\Omega)} g_{ml}^\phi(\omega, s).$$

Denoting by

$$\alpha_{m,\gamma} := \int_0^1 r Q_{m,0}(r) \bar{E}_\gamma(r) dr$$

and using that  $g_{ml}^\phi$  contains the factor  $w(s)$ , the scalar product in  $L_2(Z_\phi, w^{-1})$  we can state the result avoiding weighted scalar products in the following way.

**THEOREM 5.2.** *The reconstruction kernel  $\psi_\gamma^\phi$  for the limited angle problem has the form*

$$\begin{aligned} \tilde{\psi}_\gamma^\phi(\omega, s) &= \frac{1}{\pi\sqrt{2}} \sum_{m=0}^{\infty} (2m+1)^{1/2} U_{2m}(s) \alpha_{2m,\gamma} \times \\ &\quad \times \sum_{l=0}^m u_{2l}(2m+1, \phi/\pi, \varphi/\pi) \lambda_{2l}(2m+1, \phi/\pi)^{-1} \end{aligned}$$

leading to the representation of the solution of  $\mathbf{R}_\phi f = g$   
as

$$f_\gamma(x) = \int_{-\phi}^\phi \int_{-1}^1 \tilde{\psi}_\gamma^\phi(\omega(\varphi), s - x^\top \omega) g(\omega(\varphi), s) ds d\varphi.$$

The numerical problem in the evaluation of  $\psi_\gamma^\phi$  is based on the decay of the eigenvalues  $\lambda_l(m+1, \phi/\pi)$ . The series has to be truncated early enough not to produce numerical instabilities. But this restricts the resolution in the reconstruction too much, which we avoid by computing the reconstruction kernel as a correction

term for the full range kernel.

Using

$$\psi_\gamma(s) = \frac{1}{\pi\sqrt{2}} \sum_{m=0}^{\infty} (2m+1)^{1/2} U_{2m}(s) \alpha_{2m,\gamma},$$

we get the following result:

**THEOREM 5.3.** *The reconstruction kernel for the limited angle problem has the following, numerically more attractive, form:*

$$\tilde{\psi}_\gamma^\phi(\omega, s) = \psi_\gamma(s) + \tilde{\psi}_\gamma^{\phi,C}(\omega, s)$$

with

$$\begin{aligned} \tilde{\psi}_\gamma^{\phi,C}(\omega, s) &= \frac{1}{\pi\sqrt{2}} \sum_{m=0}^{\infty} (2m+1)^{1/2} U_{2m}(s) \alpha_{2m,\gamma} \times \\ &\quad \left( \sum_{l=0}^m u_{2l}(2m+1, \phi/\pi; \varphi/\pi) \lambda_{2l}(2m+1, \phi/\pi)^{-1} - 1 \right). \end{aligned}$$

**REMARK 5.4.** The following symmetry conditions hold

$$\psi_\gamma^\phi(\omega(-\varphi), s) = \psi_\gamma^\phi(\omega(\varphi), s), \psi_\gamma^{\phi,C}(\omega(-\varphi), s) = \psi_\gamma^{\phi,C}(\omega(\varphi), s),$$

$$\psi_\gamma^\phi(\omega, -s) = \psi_\gamma^\phi(\omega, s), \psi_\gamma^{\phi,C}(\omega, -s) = \psi_\gamma^{\phi,C}(\omega, s).$$

For practical computations we cut-off the representation of the correction term in the above theorem.

## 6. Inversion Formula for the 3D Cone Beam Transform

In the following we consider the X-ray reconstruction problem in three dimensions when the data is measured by firing an X-ray tube emitting rays to a 2D detector. The movement of the combination source - detector determines the different scanning geometries. In many real - world applications the source is moved on a circle around the object. From a mathematical point of view this has the disadvantage that the data are incomplete, the condition of Tuy-Kirillov is not fulfilled. This condition says, that essentially the data are complete for the three - dimensional Radon transform. All planes through a point  $x$  have to cut the scanning curve  $\Gamma$ . We base our considerations on the assumptions that this condition is fulfilled, the reconstruction from real data nevertheless is then from the above described circular scanning geometry, because other data is not available to us so far.

A first theoretical presentation of the reconstruction kernel was given by Finch [Fin87], invariances were then used in the group of the author to speed-up the computation time considerably, so that real data could be handled, see [Lou03]. See also the often used algorithm from Feldkamp et al. [FDK84] and the contribution of Defrise and Clack [DC94]. The approach of Katsevich [Kat02] differs from our approach that he avoids the Crofton symbol by restricting the backprojection to a range dependent on the reconstruction point  $x$ .

The presentation follows Louis [Lou04].

The mathematical model here is the so-called X-ray transform, where we denote with  $a \in \Gamma$  the source position, where  $\Gamma \subset \mathbb{R}^3$  is a curve,  $\theta \in S^2$  is the direction of the ray:

$$\mathbf{D}f(a, \theta) = \int_0^\infty f(a + t\theta) dt$$

The adjoint operator of  $D$  as mapping from  $L_2(\mathbb{R}^3) \rightarrow L_2(\Gamma \times S^2)$  is given as

$$\mathbf{D}^*g(x) = \int_\Gamma |x - a|^{-2} g\left(a, \frac{x - a}{|x - a|}\right) da$$

Most attempts to find inversion formulae are based on a relation between X-ray transform and the 3D Radon transform, the so-called *Formula of Grangeat*, first published in Grangeat's PhD thesis [Gr87], see also [Gr91] :

$$\frac{\partial}{\partial s} \mathbf{R}f(\omega, a^\top \omega) = - \int_{S^2} \mathbf{D}f(a, \theta) \delta'(\theta^\top \omega) d\theta.$$

PROOF. We copy the proof from [NW01] ). It consists of the following two steps.

i)  $\int_{\mathbb{R}} \mathbf{R}f(\omega, s) \psi(s) ds = \int_{\mathbb{R}^3} f(x) \psi(x^\top \omega) dx,$

ii)  $\int_{S^2} \mathbf{D}f(a, \theta) h(\theta) d\theta = \int_{\mathbb{R}^3} f(x) h\left(\frac{x-a}{|x-a|}\right) |x-a|^{-2} dx.$

Putting  $\psi(s) = \delta'(s - a^\top \omega)$  and use  $h(\theta) = \delta'(\theta^\top \omega)$  and the fact that  $\delta'$  is homogeneous of degree  $-2$  in  $\mathbb{R}^3$  completes the proof.  $\square$

We note the following rules for  $\delta'$ :

i)

$$\int_{S^2} \psi(a^\top \omega) \delta'(\theta^\top \omega) d\omega = -a^\top \theta \int_{S^2 \cap \theta^\perp} \psi'(a^\top \omega) d\omega,$$

ii)

$$\int_{S^2} \psi(\omega) \delta'(\theta^\top \omega) d\omega = - \int_{S^2 \cap \theta^\perp} \frac{\partial}{\partial \theta} \psi(\omega) d\omega.$$

Starting point is now the inversion formula for the 3D Radon transform

$$(6.1) \quad f(x) = -\frac{1}{8\pi^2} \int_{S^2} \frac{\partial^2}{\partial s^2} \mathbf{R}f(\omega, x^\top \omega) d\omega$$

rewritten as

$$f(x) = \frac{1}{8\pi^2} \int_{S^2} \int_{\mathbb{R}} \frac{\partial}{\partial s} \mathbf{R}f(\omega, s) \delta'(s - x^\top \omega) ds d\omega.$$

We assume in the following that the Tuy-Kirillov condition is fulfilled. Then we can change the variables as:  $s = a^\top \omega$ ,  $n$  is the Crofton symbol; i.e., the number of source points  $a \in \Gamma$  such that  $a^\top \omega = x^\top \omega$ ,  $m = 1/n$  and get

$$\begin{aligned} f(x) &= \frac{1}{8\pi^2} \int_{S^2} \int_{\Gamma} (\mathbf{R}f)'(\omega, a^\top \omega) \delta'((a-x)^\top \omega) |a'^\top \omega| m(\omega, a^\top \omega) dad\omega \\ &= -\frac{1}{8\pi^2} \int_{S^2} \int_{\Gamma} \int_{S^2} \mathbf{D}f(a, \theta) \delta'(\theta^\top \omega) d\theta \delta'((a-x)^\top \omega) |a'^\top \omega| m(\omega, a^\top \omega) dad\omega \\ &= -\frac{1}{8\pi^2} \int_{\Gamma} |x-a|^{-2} \int_{S^2} \int_{S^2} \mathbf{D}f(a, \theta) \delta'(\theta^\top \omega) d\theta \delta'\left(\frac{(x-a)^\top}{|x-a|} \omega\right) \\ &\quad \times |a'^\top \omega| m(\omega, a^\top \omega) dad\omega \end{aligned}$$

where we again used that  $\delta'$  is homogeneous of degree  $-2$ . We now introduce the following operators

$$(6.2) \quad T_1 g(\omega) = \int_{S^2} g(\theta) \delta'(\theta^\top \omega) d\theta$$

and we use  $T_1$  acting on the second variable as

$$T_{1,a} g(\omega) = T_1 g(a, \omega).$$

We also use the multiplication operator

$$(6.3) \quad M_{\Gamma,a} h(\omega) = |a'^\top \omega| m(\omega, a^\top \omega) h(\omega).$$

and state the following result.

**THEOREM 6.1.** *Let the condition of Tuy-Kirillov be fulfilled. Then the inversion formula for the cone beam transform is given as*

$$(6.4) \quad f = -\frac{1}{8\pi^2} \mathbf{D}^* T_1 M_{\Gamma,a} T_1 \mathbf{D} f$$

with the adjoint operator  $\mathbf{D}^*$  of the cone beam transform and  $T_1$  and  $M_{\Gamma,a}$  as defined above.

Note that the operators  $\mathbf{D}^*$  and  $M$  depend on the scanning curve  $\Gamma$ .

This form allows for computing reconstruction kernels. To this end we have to solve the equation

$$\mathbf{D}^* \psi_\gamma = e_\gamma$$

in order to write the solution of  $\mathbf{D}f = g$  as

$$f(x) = \langle g, \psi_\gamma(x, \cdot) \rangle.$$

In the case of exact inversion formula  $e_\gamma$  is the delta distribution, in the case of the approximate inversion formula it is an approximation of this distribution, see the method of approximate inverse. Using that  $\mathbf{D}^{-1} = -\frac{1}{8\pi^2} \mathbf{D}^* T_1 M_{\Gamma,a} T_1$  we get

$$\mathbf{D}^* \psi = \delta = -\frac{1}{8\pi^2} \mathbf{D}^* T_1 M_{\Gamma,a} T_1 \mathbf{D} \delta$$

and hence

$$(6.5) \quad \psi = -\frac{1}{8\pi^2} T_1 M_{\Gamma,a} T_1 \mathbf{D} \delta.$$

We can explicitly give the form of the operators  $T_1$  and  $T_2 = M T_1$ . The index at  $\nabla$  indicates the variable with respect to which the differentiation is performed.

$$\begin{aligned} T_1 g(a, \omega) &= \int_{S^2} g(a, \theta) \delta'(\theta^\top \omega) d\theta \\ &= -\omega^\top \int_{S^2 \cap \omega^\perp} \nabla_2 g(a, \theta) d\theta \end{aligned}$$

and

$$\begin{aligned}
 T_1 M_{\Gamma,a} h(a, \alpha) &= \int_{S^2} \delta'(\omega^\top \alpha) |a'^\top \omega| m(\omega, a^\top \omega) h(a, \omega) d\omega \\
 &= -a'^\top \alpha \int_{S^2 \cap \alpha^\perp} \text{sign}(a'^\top \omega) m(\omega, a^\top \omega) h(a, \omega) d\omega \\
 &\quad -\alpha^\top \int_{S^2 \cap \alpha^\perp} |a'^\top \alpha| \nabla_1 m(\omega, a^\top \omega) h(a, \omega) d\omega \\
 &\quad -a^\top \alpha \int_{S^2 \cap \alpha^\perp} |a'^\top \omega| \nabla_2 m(a, a^\top \omega) h(a, \omega) d\omega \\
 &\quad - \int_{S^2 \cap \alpha^\perp} |a'^\top \omega| m(\omega, a^\top \omega) \frac{\partial}{\partial \alpha} h(a, \omega) d\omega.
 \end{aligned}$$

Note that the function  $m$  is piecewise constant, the derivatives are then Delta-distributions at the discontinuities with factor equal to the height of the jump; i.e., 1/2.

Depending on the scanning curve  $\Gamma$  invariances have to be used. For the circular scanning geometry this leads to similar results as mentioned in [Lou03]. In the following we present a reconstruction from data provided by the Fraunhofer Institut for Nondestructive Testing (Izfp) in Saarbrücken. The detector size was  $(204.8\text{mm})^2$  with  $512^2$  pixels and 400 source positions on a circle around the object. The second data set was provided by the Deutsches Krebsforschungszentrum (DKFZ) Heidelberg, with the same number of data, namely 10.4 million. The mollifier used is

$$e_\gamma(y) = (2\pi)^{-3/2} \gamma^{-3} \exp\left(-\frac{1}{2} \left|\frac{y}{\gamma}\right|^2\right).$$

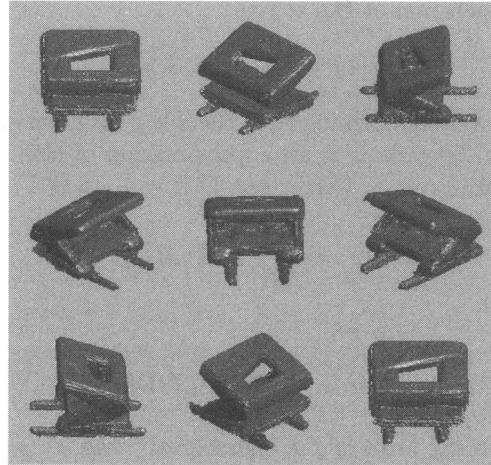


FIGURE 2. Reconstruction of a perforator.

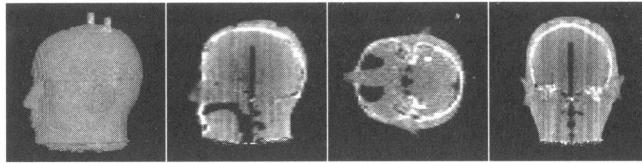


FIGURE 3. A human skull filled with plastic, and with skin, ears, nose etc. modelled with plastic. The bars which fix this skull are visible.

## References

- [Cor64] A.M. Cormack, *Representation of a function by its line integral, with some radiological applications II*. J. Appl.Phys. **35** (1964) 195–207.
- [DC94] M. Defrise, R. Clack, *A cone-beam reconstruction algorithm using shift-invariant filtering and cone-beam backprojection*. IEEE TMI **13** (1994), 186–195.
- [FDK84] L.A. Feldkamp, L.C. Davis, J.W. Kress, *Practical cone beam algorithm*. J. Opt. Soc. Am. A **6** (1984), 612–619.
- [Fin87] D. Finch, *Approximate reconstruction formulae for the cone beam transform, I*. Preprint (1987).
- [DG81] M.E. Davison, F.A. Grünbaum, *Tomographic reconstruction with arbitrary directions*. IEEE TNS **26** (1981) 77–120.
- [Gr87] P. Grangeat, *Analyse d'un Système d'Imagerie 3D par Reconstruction à partir de Radiographies X en Géométrie Conique*, Dissertation, Ecole Nationale Supérieure des Télécommunications (1987)
- [Gr91] P. Grangeat, *Mathematical framework of cone beam 3-D reconstruction via the first derivative of the Radon transform* in Herman, G. T., Louis, A.K. and Natterer, F., editors, *Mathematical Methods in Tomography*. Springer, Berlin (1991) 66–97.
- [HS88] I. Hazou, D.C. Solmon, *Inversion of the exponential X-ray transform. I:Analysis*. Math. Methods Appl. Sci. **10** (1988), 561–574.
- [Kat02] A. Katsevich, *Analysis of an exact inversion algorithm for spiral-cone beam CT*. Phys. Med. Biol. **47** (2002), 2583–2597.
- [KCSWZ95] P. Kuppusamy, M. Chzhan, A. Samouilov, P. Wang, J.L. Zweier, *Mapping the spin-density and lineshape distribution of free radicals using 4D spectral-spatial EPR imaging*. J. Magnetic Resonance, Series B **197** (1995), 116–125.
- [Lou80] A.K. Louis, *Picture reconstruction from projections in restricted range*. Math. Methods Appl. Sci. **2** (1980) 209–220.
- [Lou83] A.K. Louis, *Approximate Inversion of the 3 D Radon Transform*. Math. Methods Appl. Sci. **5** (1983) 176–185.
- [Lou84] A.K. Louis, *Orthogonal function series expansion and the null space of the Radon transform*. SIAM J. Math. Anal. **15** (1984), 621–633.
- [Lou84a] A.K. Louis, *Nonuniqueness in inverse Radon problems: the frequency distributions of the ghosts*. Math. Z. **185** (1984) 429–440.
- [Lou86] A.K. Louis, *Incomplete data problems in X-ray computerized tomography, I: Singular value decomposition of the limited angle transform*. Numer. Math. **48** (1986), 251–262.
- [Lou89] A.K. Louis, *Inverse und schlecht gestellte Probleme* Teubner, Stuttgart (1989).
- [Lou96] A.K. Louis, *The approximate inverse for linear and some nonlinear problems*. Inverse Problems **12** (1996), 175–190.
- [Lou99] A.K. Louis, *A unified approach to regularization methods for linear ill-posed problems*. Inverse Problems **15** (1999), 489–498.
- [Lou03] A.K. Louis, *Filter design in three-dimensional cone beam tomography: circular scanning geometry*. Inverse Problems **19** (2003), S31–S40.
- [Lou04] A.K. Louis. *Inversion formula for 3D cone beam transform*. to appear.
- [LM90] A.K. Louis, P. Maass, *A mollifier method for linear operator equations of the first kind*. Inverse Problems **6** (1990), 427–440 427-440.

- [LM93] A.K. Louis, P. Maass, *Contour reconstruction in 3D X-ray CT*, IEEE Trans. Med. Imaging **TMI12** (1993), 764–769
- [LMR97] A.K. Louis, P. Maass, A. Rieder *Wavelets : Theory and Applications*. Wiley, Chichester (1997).
- [LR89] A.K. Louis, A. Rieder, *Incomplete data problems in X-ray computerized tomography, II: Truncated projections and region-of-interest tomography*. Numer. math. **56** (1989), 371–383.
- [LS96] A.K. Louis, T. Schuster, *A novel filter design technique in 2D computerized tomography*. Inverse Problems **12** (1996) 685–696.
- [Maa87] P. Maass, *The X-ray transform: singular value decomposition and resolution*. Inverse Problems **3** (1987), 729–741.
- [MCL81] R.B. Marr, C.N. Chen, P.C. Lauterbur, *On Two Approaches to 3D Reconstruction in NMR Zeugmatography* in Herman, G. T. and Natterer, F., editors, *Mathematical Aspects of Computerized Tomography*. Springer, Berlin (1981) 225.
- [Nat86] F. Natterer *The mathematics of computerized tomography* Teubner-Wiley, Stuttgart (1986).
- [NW01] F. Natterer, F. Wübbeling, *Mathematical Methods in Image Reconstruction*, SIAM, Philadelphia (2001).
- [Qui88] E.T. Quinto, *Tomographic reconstruction from incomplete data – numerical inversion of the exterior Radon transform*. Inverse Problems **4** (1988), 867–876.
- [Qui93] E.T. Quinto, *Singularities of the X-ray transform and limited data tomography in  $\mathbb{R}^2$  and  $\mathbb{R}^3$* . SIAM J. Math. Anal. **24** (1993), 1215–1225.
- [Qui06] E.T. Quinto, *An introduction to X-ray tomography and Radon transforms* in The Radon Transform and Applications to Inverse Problems, G. Olafsson and E.T. Quinto, eds. AMS Proceedings of Symposia in Applied Mathematics, Providence, RI, 2006, American Mathematical Society
- [Rie01] A. Rieder, *Principles of reconstruction filter design in 2D-computerized tomography*. Contemporary Mathematics, **278**, (2001), 207–226.
- [Sch00] T. Schuster, *The 3D-Doppler transform: elementary properties and computation of reconstruction kernels*. Inverse Problems **16** (2000), 701–723
- [Sle78] D. Slepian, *Prolate spheroidal wave functions, Fourier analysis and uncertainty - V: The discrete case*. Bell System Technical Journal, **57** (1978), 1371–1430.

DEPARTMENT OF MATHEMATICS, SAARLAND UNIVERSITY, 66041 SAARBRÜCKEN GERMANY  
*E-mail address:* louis@num.uni-sb.de

## Fan-Beam Tomography and Sampling Theory

Adel Faridani

**ABSTRACT.** Computed tomography entails the reconstruction of a function from measurements of its line integrals. In this article we explore the question: How many and which line integrals should be measured in order to achieve a desired resolution in the reconstructed image? Answering this question may help to reduce the amount of measurements and thereby the radiation dose, or to obtain a better image from the data one already has. Our exploration leads us to a mathematically and practically fruitful interaction of Shannon sampling theory and tomography. For example, sampling theory helps to identify efficient data acquisition schemes, provides a qualitative understanding of certain artifacts in tomographic images, and facilitates the error analysis of some reconstruction algorithms. On the other hand, applications in tomography have stimulated new research in sampling theory, e.g., on nonuniform sampling theorems and estimates for the aliasing error. The focus of this article will be the application of sampling theory to the so-called fan-beam geometry. Its dual aim is an exposition of the main principles involved as well as the development of some new insights.

### 1. Introduction

Computed tomography (CT) entails the reconstruction of a function  $f$  from measurements of line integrals of  $f$ . Naturally, one would like to reconstruct a high-resolution image with a minimal amount of measured data. The fundamental question underlying the line of research reported here was posed in a seminal 1978 paper by A. Cormack, one of the pioneers of tomography:

‘In practice one can make only a finite number of measurements with beams of finite width, and the question which arises is how many observations should be made, and how should they be related to each other in order to reconstruct the object’ [8].

While Cormack himself proceeded with geometric arguments to discover an efficient data acquisition scheme, we will use two-dimensional sampling theory to address this question, an approach first introduced by Lindgren and Rattey [40, 55].

---

1991 *Mathematics Subject Classification*. Primary 94A20, 65R10; Secondary 42A15, 42B99, 65R32, 65T40.

*Key words and phrases.* tomography, fan-beam, sampling, resolution, sampling theory, periodic sampling, Radon transform, filtered backprojection.

This work was supported by NSF Grant DMS-0206752.

Sampling theory originated from the classical sampling theorem, which permits the reconstruction of a bandlimited function from its values on a regular grid or lattice. The classical sampling theorem has been extend in many directions, giving rise to a lively field of contemporary research.

It turns out that sampling theory is not only useful in identifying efficient sampling schemes for tomographic data, but among other things provides a qualitative understanding of certain artifacts and facilitates the numerical analysis of reconstruction algorithms. On the other hand, applications in tomography have stimulated research in sampling theory, for example on estimates for the aliasing error and on non-uniform sampling.

This article is organized as follows. In the next section we lay the foundation by describing the two-dimensional Radon transform which furnishes the mathematical model for tomography, as well as the classical sampling theorem. We introduce the two popular data acquisition geometries and characterize the sampling lattices for the so-called fan-beam geometry, which will serve as the focus of this article. Section 3 is devoted to some applications of the classical sampling theorem, namely identification of efficient sampling schemes and the qualitative understanding of artifacts resulting from undersampling. In § 4 we briefly describe a recent development: the use of non-equidistant periodic sampling in achieving higher resolution in fan-beam tomography. In the final section we summarize our conclusions and present a brief overview over the themes and topics of the interaction between tomography and sampling theory with references for further study.

## 2. Foundations

**2.1. The two-dimensional Radon transform.** The 2D Radon transform maps a density function  $f$  into its line integrals. Throughout this paper we will assume that  $f \in C_0^\infty(\Omega)$ , i.e.,  $f$  is infinitely differentiable and vanishes outside the unit disk  $\Omega$  of  $\mathbb{R}^2$ . The smoothness assumption simplifies the mathematical proofs, and although the density functions occurring in practice are not necessarily smooth, the theoretical results seem to describe the phenomena observed in practice well.

Let  $\mathbb{Z}, \mathbb{R}, \mathbb{C}$  denote the integers, real and complex numbers, respectively. Let  $\theta = (\cos \varphi, \sin \varphi)$  be the unit vector in  $\mathbb{R}^2$  with polar angle  $\varphi$ , and  $\theta^\perp = (-\sin \varphi, \cos \varphi)$ . For  $f \in C_0^\infty(\Omega)$  define its Radon transform  $Rf$  by

$$(2.1) \quad \begin{aligned} Rf(\varphi, s) &= \int_{-\infty}^{\infty} f(s \cos \varphi - t \sin \varphi, s \sin \varphi + t \cos \varphi) dt \\ &= \int_{\mathbb{R}} f(s\theta + t\theta^\perp) dt, \end{aligned}$$

i.e.,  $Rf(\varphi, s)$  is the integral of  $f$  over the line in direction  $\theta^\perp$  with signed distance  $s$  from the origin. Sometimes  $Rf$  is considered as a function of  $s$  for fixed  $\varphi$ . In this case we write  $R_\varphi f(s)$  for  $Rf(\varphi, s)$ .

Observe that the parameter choices  $(\varphi, s)$  and  $(\varphi + \pi, -s)$  lead to one and the same line. We therefore have the symmetry relation

$$(2.2) \quad Rf(\varphi, s) = Rf(\varphi + \pi, -s).$$

The goal of x-ray tomography is to reconstruct an approximation to  $f(x)$  from sampled values of  $Rf$ . An explanation of how the Radon transform arises as the

mathematical model for x-ray tomography is given in Quinto's article in this volume [54].

The Fourier transform of a function  $g \in C_0^\infty(\mathbb{R}^n)$  is defined by

$$\hat{g}(\xi) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} g(x) e^{-i\langle x, \xi \rangle} dx$$

and is extended to larger classes of functions or distributions by continuity or duality. Here  $\langle x, \xi \rangle$  denotes the usual inner product in  $\mathbb{R}^n$ .

In particular, the Fourier transform of  $R_\varphi f$  is given by

$$(R_\varphi f)^\wedge(\sigma) = (2\pi)^{-1/2} \int_{\mathbb{R}} R_\varphi f(s) e^{-is\sigma} ds.$$

The following relation between the Fourier transforms of  $R_\varphi f$  and  $f$  is very useful and straightforward to verify:

$$(2.3) \quad (R_\varphi f)^\wedge(\sigma) = (2\pi)^{1/2} \hat{f}(\sigma\theta).$$

Equation (2.3) is called the projection-slice theorem.

Discretizing the Radon transform in the variables  $(\varphi, s)$  of (2.1) leads to the so-called parallel-beam sampling geometry. It derives its name from the fact that keeping  $\varphi$  fixed and varying  $s$  leads to a collection of parallel lines.

Many medical scanners employ an x-ray source which circles around the object. This leads to the so-called fan-beam sampling geometry, where for each of a number of source positions distributed around a circle of radius  $r > 1$  the integrals over rays emanating from that source position are measured. To describe this type of data collection a parametrization of lines as in (2.1) is less convenient. Instead, we introduce the divergent beam x-ray transform

$$(2.4) \quad D_z f(\omega) = \int_0^\infty f(z + t\omega) dt, \quad z \in \mathbb{R}^2, \quad \omega \in S^1,$$

which gives the integral of  $f$  over the ray with direction  $\omega$  emanating from the source point  $z$ . Let  $\beta$  denote the polar angle of  $z$ , that is  $z = r(\cos \beta, \sin \beta)$ . We parameterize the direction  $\omega$  of a ray emanating from  $z$  by  $\omega = -(\cos(\alpha + \beta), \sin(\alpha + \beta))$ , where  $\alpha$  is the angle between the ray from  $z$  in direction  $\omega$  and the central ray connecting  $z$  and the origin. The angle  $\alpha$  is taken to be positive when the ray in direction  $\omega$  lies to the left of the central ray when viewed from the point  $z$ . With this parametrization we have

$$D_z f(\omega) = Df(\beta, \alpha), \quad 0 \leq \beta < 2\pi, \quad -\pi/2 \leq \alpha \leq \pi/2.$$

We extend  $Df(\beta, \alpha)$  as a  $2\pi$ -periodic function in  $\alpha$  and obtain the correspondence

$$Df(\beta, \alpha) = \begin{cases} Rf(\alpha + \beta - \pi/2, r \sin \alpha), & |\alpha| < \pi/2 \\ 0 & |\alpha| \geq \pi/2 \end{cases}, \quad \alpha \in [-\pi, \pi).$$

The symmetry relation (2.2) now becomes

$$(2.5) \quad Df(\beta, \alpha) = Df(\beta + 2\alpha + \pi, -\alpha).$$

**REMARK 2.1.** Since  $f$  is supported in the unit disk, its ray integrals can only be non-zero for rays intersecting the unit circle. These are the rays with  $|\alpha| < \arcsin(1/r)$ . Hence we could consider  $Df$  as a  $2a$ -periodic function in  $\alpha$  for any  $a$  with  $\arcsin(1/r) \leq a \leq \pi$ . The choice  $a = \pi$  is made here for reasons of simplicity of exposition and of consistency with the notation in [49]. Choosing a smaller value of  $a$  may have some advantages for computer implementations.

For readers interested in a more detailed introduction to tomography we recommend the other articles in this volume, the monographs [45, 49] or the introductory survey [20] and the references given there.

**2.2. The Classical Sampling Theorem.** The origin of sampling theory is the classical sampling theorem. In its simplest version it reads as follows.

**THEOREM 2.2.** *Let  $g \in L_2(\mathbb{R})$  such that its Fourier transform  $\hat{g}(\xi)$  vanishes for  $|\xi| \geq b$ . If  $0 < h \leq \pi/b$  then*

$$g(x) = \frac{hb}{\pi} \sum_{l \in \mathbb{Z}} g(hl) \operatorname{sinc}(b(x - hl)),$$

where  $\operatorname{sinc}(t) = \frac{\sin t}{t}$ . The series converges in  $L_2$  as well as uniformly.

**REMARK 2.3.** The term  $\operatorname{sinc}$  stands for ‘sinus cardinalis’, and the series is called ‘cardinal series’. The following observations will be useful for generalizing the theorem; cf. Theorem 2.7 below.

- (1) The theorem permits recovery of  $g$  from its values on a subgroup  $\mathbf{L} = h\mathbb{Z}$  of  $\mathbb{R}$ .  $\mathbf{L}$  is also called a lattice.
- (2) The set  $\mathbf{L}^\perp = \{2\pi l/h, l \in \mathbb{Z}\}$  satisfies  $e^{iy\eta} = 1$  for all  $y \in \mathbf{L}$ ,  $\eta \in \mathbf{L}^\perp$  and is called the dual or reciprocal lattice of  $\mathbf{L}$ . According to the hypothesis of Theorem 2.2 the support of  $\hat{g}$  is contained in the closure of the set  $K = [-b, b]$ . The density condition  $h \leq \pi/b$  is therefore equivalent to the condition that the translates  $K + \eta$ ,  $\eta \in \mathbf{L}^\perp$  are mutually disjoint.
- (3) Note that the function  $s(x) = \operatorname{sinc}(bx)$  is up to a multiplicative constant equal to the inverse Fourier transform  $\tilde{\chi}_K$  of the indicator function  $\chi_K$  of  $K = [-b, b]$ . Recall that  $\chi_K(\xi) = 1$  for  $\xi \in K$  and  $\chi_K(\xi) = 0$  otherwise. We have

$$(2.6) \quad \tilde{\chi}_K(x) = (2\pi)^{-1/2} \int_{-b}^b e^{ix\xi} d\xi = \sqrt{\frac{2}{\pi}} \frac{\sin bx}{x} = \sqrt{\frac{2}{\pi}} b \operatorname{sinc}(bx).$$

The sampling theorem is closely related to the *Poisson summation formula*, which is a fundamental tool for all results which will be discussed in this paper. Its one-dimensional version reads as follows.

**THEOREM 2.4** (Poisson summation formula for  $\mathbb{R}$ ). *Let  $h > 0$ ,  $z, \sigma \in \mathbb{R}$ , and  $g \in C(\mathbb{R})$  such that  $|g(x)| \leq C(1 + |x|)^{-1-\epsilon}$ , and  $|\hat{g}(\xi)| \leq C(1 + |\xi|)^{-1-\epsilon}$  for some  $C > 0$ ,  $\epsilon > 0$ . Then*

$$(2.7) \quad (2\pi)^{-1/2} h \sum_{l \in \mathbb{Z}} g(z + hl) e^{-i\sigma(z + hl)} = \sum_{l \in \mathbb{Z}} \hat{g}(\sigma + 2\pi l/h) e^{iz2\pi l/h}.$$

For a proof see, e.g., [24, Theorem (8.36)]. The result holds also under less restrictive hypotheses and in a very general setting; see [26, p. 217].

The sampling theorem can be formally derived from (2.7) as follows. Let  $z = 0$  in (2.7). If  $\hat{g}(\xi)$  vanishes for  $|\xi| \geq b$  and  $h \leq \pi/b$ , then the right-hand side of (2.7) simplifies to  $\hat{g}(\sigma)$  for  $|\sigma| \leq b$ . Now take the inverse Fourier transform

$$g(x) = (2\pi)^{-1/2} \int_{-b}^b \hat{g}(\sigma) e^{ix\sigma} d\sigma,$$

replace in the integral  $\hat{g}(\sigma)$  by the left-hand side of (2.7) and use (2.6).

Another important application of the Poisson summation formula consists in error estimates for numerical integration with the trapezoidal rule. Assume for example that  $\hat{g}(\sigma) = (2\pi)^{-1/2} \int_{\mathbb{R}} g(x) e^{-i\sigma x} dx$  is to be approximated by

$$\hat{g}(\sigma) \simeq (2\pi)^{-1/2} h \sum_{l \in \mathbb{Z}} g(hl) e^{-i\sigma hl}.$$

Letting  $z = 0$  in (2.7) we obtain the following expression for the integration error:

$$(2\pi)^{-1/2} h \sum_{l \in \mathbb{Z}} g(hl) e^{-i\sigma hl} - \hat{g}(\sigma) = \sum_{l \neq 0} \hat{g}(\sigma + 2\pi l/h).$$

The classical sampling theorem is fundamental for signal processing, has been generalized in many directions, and given rise to the field of sampling theory. Readers seeking more information on sampling theory will find ample material in survey articles including [1, 7, 28, 34], monographs such as [29, 41, 60], and collections of research articles, e.g., [3, 4, 30, 42].

**2.3. Sampling lattices for the divergent beam transform.** We have seen that  $Df(\beta, \alpha)$  is a function with domain  $[0, 2\pi) \times [-\pi, \pi)$ . For the subsequent analysis it is more convenient to transform this domain to  $[0, 1)^2$  by means of the change of variables

$$(2.8) \quad \begin{aligned} g(s, t) &= Df(\beta, \alpha) \\ (s, t) &\in [0, 1)^2, \quad (\beta, \alpha) \in [0, 2\pi) \times [-\pi, \pi) \\ s &= \frac{\beta}{2\pi}, \quad t = \frac{\alpha}{2\pi} + \frac{1}{2}, \end{aligned}$$

that is, we will henceforth consider the function

$$(2.9) \quad g(s, t) = Df(2\pi s, 2\pi t - \pi), \quad (s, t) \in [0, 1)^2.$$

This will allow us to directly use the theory and algorithms developed in [18, §4].

The subsequent analysis of sampling and resolution will make use of Fourier analysis. This requires both the domain of  $g$  as well as the sampling sets to have a group structure. Equipped with addition modulo 1 the interval  $[0, 1)$  becomes a group, called the circle group, which we denote by  $\mathbb{T}$ . Then the domain of  $g$  may be identified with the group  $\mathbb{T}^2$ , called the torus group. The Fourier transform of  $g$  is given by

$$\hat{g}(k, m) = \int_0^1 \int_0^1 g(s, t) e^{-2\pi i(k s + m t)} ds dt, \quad (k, m) \in \mathbb{Z}^2.$$

Using the notation  $z = (s, t)$ ,  $\zeta = (k, m)$ ,  $\langle z, \zeta \rangle = sk + tm$ , this can be written as

$$\hat{g}(\zeta) = \int_{\mathbb{T}^2} g(z) e^{-2\pi i \langle z, \zeta \rangle} dz, \quad \zeta \in \mathbb{Z}^2.$$

The inverse Fourier transform in this setting is given by

$$\begin{aligned} \tilde{G}(s, t) &= \sum_{k \in \mathbb{Z}} \sum_{m \in \mathbb{Z}} G(k, m) e^{2\pi i(k s + m t)} \\ &= \int_{\mathbb{Z}^2} G(\zeta) e^{2\pi i \langle z, \zeta \rangle} d\zeta, \end{aligned}$$

with  $z = (s, t) \in \mathbb{T}^2$ ,  $\zeta = (k, m) \in \mathbb{Z}^2$ , and  $d\zeta$  denoting the counting measure on  $\mathbb{Z}^2$ .

The task of fan-beam tomography is to reconstruct  $f$  from finitely many measurements of  $g$ . We require the set of all points  $(s, t)$  where  $g$  is measured to be a (possibly shifted) finite subgroup of  $\mathbb{T}^2$ . We call a finite subgroup  $\mathbf{L}$  of  $\mathbb{T}^2$  a lattice or sometimes a sampling lattice. Every lattice  $\mathbf{L}$  in  $\mathbb{T}^2$  has a corresponding ‘reciprocal lattice’  $\mathbf{L}^\perp$  in the Fourier domain  $\mathbb{Z}^2$ .  $\mathbf{L}^\perp$  is the set of all  $\eta \in \mathbb{Z}^2$  such that  $e^{2\pi i \langle y, \eta \rangle} = 1$  for all  $y \in \mathbf{L}$ .

In order to systematically investigate sampling on  $\mathbb{T}^2$  we need to characterize and parametrize all lattices. Fortunately such a characterization of the finite subgroups of  $\mathbb{T}^2$  is available in the literature. Let  $|\mathbf{L}|$  denote the number of elements of a lattice  $\mathbf{L}$  and let  $[x]$  denote the fractional part of a real number  $x$ , i.e.,  $[x] \in [0, 1)$  and  $x - [x] \in \mathbb{Z}$ .

**PROPOSITION 2.5.** If  $\mathbf{L}$  is a finite subgroup of  $\mathbb{T}^2$ , then there exists a unique non-singular lower triangular  $2 \times 2$  matrix  $W$  such that

- (i) As a set,  $\mathbf{L} = [W\mathbb{Z}^2]$ , and  $\mathbf{L}^\perp = W^{-T}\mathbb{Z}^2$ .
- (ii)  $|\mathbf{L}| = |\det W|^{-1}$
- (iii) The matrix  $W^{-T}$  has Hermite normal form, i.e.  $W^{-T}$  is an upper triangular matrix

$$(2.10) \quad W^{-T} = \begin{pmatrix} P & -N \\ 0 & Q \end{pmatrix} \quad \text{with } N, P, Q \in \mathbb{Z}, \quad P, Q > 0, \quad 0 \leq N \leq P - 1.$$

$W$  and  $W^{-T}$  are called generator matrices of  $\mathbf{L}$  and  $\mathbf{L}^\perp$ , respectively.

*Proof:* This is the two-dimensional case of Proposition 4.2 in [18], which in turn is based on [50, pp. 125-126, 131-132] and [51, Theorems II.2, II.3].  $\square$

From Proposition 2.5 we conclude the following. Any sampling lattice  $\mathbf{L} = \mathbf{L}(N, P, Q)$  is characterized by three integers  $N, P, Q$  such that  $P, Q > 0$ ,  $0 \leq N \leq P - 1$ , the generator matrix  $W$  is given by

$$W = \begin{pmatrix} 1/P & 0 \\ N/(PQ) & 1/Q \end{pmatrix},$$

and

$$(2.11) \quad \begin{aligned} \mathbf{L} &= [W\mathbb{Z}^2] \\ &= \{(s_j, t_{jl}) : s_j = j/P, t_{jl} = [(l + Nj/P)/Q], \\ &\quad j = 0, \dots, P - 1, l = 0, \dots, Q - 1\}. \end{aligned}$$

Furthermore, the reciprocal lattice is given by

$$(2.12) \quad \mathbf{L}^\perp(N, P, Q) = \{(Pk_1 - Nk_2, Qk_2), k_1, k_2 \in \mathbb{Z}\}.$$

Going back for a moment to the  $(\beta, \alpha)$  coordinates we see that sampling lattices for fan-beam tomography have the following structure:

- There are  $P$  source positions  $z_j = r(\cos \beta_j, \sin \beta_j)$  corresponding to the equidistant angles

$$\beta_j = 2\pi j/P, \quad j = 0, \dots, P - 1.$$

- For each of these source positions  $Q$  integrals over rays corresponding to an equidistant set of angles  $\alpha_{jl}$  with spacing  $2\pi/Q$  are measured.

- This collection of equiangular rays is shifted by an amount which varies with the angle  $\beta_j$ , so that the  $\alpha$ -values corresponding to the angle  $\beta_j$  are given by

$$\alpha_{jl} = -\pi + 2\pi[(l + Nj/P)/Q], \quad 0 \leq l \leq Q - 1$$

where the parameter  $N \in \mathbb{Z}$ ,  $0 \leq N \leq P - 1$  determines the shifting pattern.

Hence every sampling lattice  $\mathbf{L} = \mathbf{L}(N, P, Q)$  is characterized by its number of source positions  $P$ , the number  $Q$  of equiangularly spaced rays in each fan, and the shift parameter  $N$ . The total number of samples equals  $PQ$ .

The most frequently used lattice is the *standard lattice*

$$\mathbf{L}_S = \mathbf{L}(0, P, Q) = \{(s_j, t_l) : s_j = j/P, t_l = l/Q, j = 0, \dots, P - 1, l = 0, \dots, Q - 1\}$$

which is obtained by letting  $N = 0$ .

We wish to apply Shannon sampling theory in order to find the best sampling lattices. In order to do this we need to state the Poisson summation formula and the classical sampling theorem for functions defined on  $\mathbb{T}^2$ .

**THEOREM 2.6** (Poisson summation formula for  $\mathbb{T}^2$ ). *Let  $z \in \mathbb{T}^2$ ,  $\zeta \in \mathbb{Z}^2$ ,  $\mathbf{L} = \mathbf{L}(N, P, Q)$  be a sampling lattice and  $g \in C^\infty(\mathbb{T}^2)$ . Then*

$$(2.13) \quad \frac{1}{PQ} \sum_{y \in \mathbf{L}} g(z + y) e^{-2\pi i \langle z + y, \zeta \rangle} = \sum_{\eta \in \mathbf{L}^\perp} \hat{g}(\zeta + \eta) e^{2\pi i \langle z, \eta \rangle}.$$

The classical sampling theorem for this setting reads as follows.

**THEOREM 2.7.** *Let  $g \in C^\infty(\mathbb{T}^2)$ ,  $\mathbf{L} = \mathbf{L}(N, P, Q)$  a sampling lattice and  $K$  be a finite subset of  $\mathbb{Z}^2$  such that its translates  $K + \eta$ ,  $\eta \in \mathbf{L}^\perp$  are disjoint. Let  $\chi_K$  denote the indicator function of  $K$ , i.e.,  $\chi_K(\zeta) = 1$  if  $\zeta \in K$  and  $\chi_K(\zeta) = 0$  otherwise. For  $z \in \mathbb{T}^2$  define*

$$(2.14) \quad Sg(z) = \frac{1}{PQ} \sum_{y \in \mathbf{L}} \tilde{\chi}_K(z - y) g(y).$$

Then

$$(2.15) \quad |g(z) - Sg(z)| \leq 2 \int_{\mathbb{Z}^2 \setminus K} |\hat{g}(\zeta)| d\zeta.$$

For a proof see, e.g., [17] or [45, pp. 62–64]. In these references the domain of the function  $g$  is different, but the proof is readily transferred to the present setting. The key idea is to start out with the observation

$$|g(z) - Sg(z)| \leq \int_{\mathbb{Z}^2} |\hat{g}(\zeta) - \widehat{Sg}(\zeta)| d\zeta,$$

split the integral into integrals over  $K$  and  $\mathbb{Z}^2 \setminus K$ , observe that  $\widehat{Sg}$  vanishes outside  $K$ , and estimate the integral over  $K$  using the Poisson summation formula and the disjointness of the translates of  $K$  under elements of  $\mathbf{L}^\perp$ .

Observe that if  $\hat{g}$  vanishes outside of  $K$  then  $g = Sg$ , i.e.,  $g$  can be recovered exactly from its samples on the lattice  $\mathbf{L}$ .

The right-hand side of (2.15) provides an estimate for the so-called aliasing error. If  $g$  as in (2.9) is the divergent beam transform of a function with compact support, then the Fourier transform  $\hat{g}$  cannot have compact support and an aliasing error will always be present.

### 3. Applications of the Classical sampling theorem.

In this section we illustrate some of the applications of the classical sampling theorem, namely the identification of efficient sampling schemes and the qualitative understanding of artifacts. Our presentation will focus on the fan-beam geometry. An analogous discussion for the parallel-beam geometry can be found in [21].

**3.1. Identification of efficient sampling schemes.** Theorem 2.7 indicates what needs to be done in order to find efficient sampling schemes for tomography, i.e., when  $g(s, t) = Df(2\pi s, 2\pi t - \pi)$ . First we need to find a suitable set  $K$  so that the right-hand side of (2.15) is sufficiently small. Then we need to identify lattices  $\mathbf{L}(N, P, Q)$  as sparse as possible but such that the translated sets  $K + \eta$ ,  $\eta \in \mathbf{L}^\perp$  are disjoint.

The set  $K$  will of course depend on the function  $f$ . The crucial parameter turns out to be a cut-off frequency  $b$  such that  $|\hat{f}(\xi)|$  is negligible for  $|\xi| > b$ , in the sense that the integral  $\int_{|\xi|>b} |\hat{f}(\xi)| d\xi$  is sufficiently small. The parameter  $b$  may be viewed as an ‘essential bandwidth’ of  $f$ . In [46] it is shown that then  $|\hat{g}(k, m)|$  will be small for  $|k - m| \geq rb$ . On the other hand, it follows from [52, Theorem 2] that due to  $f$  being supported in the unit disk,  $|\hat{g}(k, m)|$  is small and decays exponentially with increasing  $k$  in the region  $r|k| \geq \vartheta^{-1}|k - m|$ , where the ‘safety parameter’  $\vartheta$  satisfies  $0 < \vartheta < 1$  and can usually be chosen very close to 1. Combining these estimates and leaving some safety margin near the origin (cf. [25]) leads us to the set

(3.1)

$$K(\vartheta, b) = \{(k, m) \in \mathbb{Z}^2 : |k - m| < rb, r|k| < \vartheta^{-1} \max(|k - m|, (1 - \vartheta)rb)\};$$

see Figure 1. In this connection it is important to note that the only assumption made about  $f$  when determining  $K$  is the essential bandwidth  $b$ . If  $f$  has additional properties the essential support of  $\hat{g}$  may be smaller than  $K(\vartheta, b)$ . For example, if  $f$  is rotationally symmetric, then  $\hat{g}(k, m) = 0$  for  $|k| > 0$ .

The crucial feature both for finding efficient sampling schemes and for understanding some of the artifacts caused by undersampling is the particular, non-convex shape of  $K$ . It stems from the fact that the function  $f$  has compact support.

The next step in applying Theorem 2.7 is to find conditions for the lattice parameters  $N, P, Q$  such that the translated sets  $K(\vartheta, b) + \eta$ ,  $\eta \in \mathbf{L}^\perp(N, P, Q)$  are disjoint. This is a requirement for the reciprocal lattice to be sparse which means that the sampling lattice itself must be sufficiently dense.

We begin with the standard lattice, i.e.,  $N = 0$ . According to (2.12) the reciprocal lattice is given by

$$\mathbf{L}_S^\perp = \mathbf{L}^\perp(0, P, Q) = \{(Pk_1, Qk_2), k_1, k_2 \in \mathbb{Z}\}.$$

In particular, the points  $\eta_1 = (0, Q)$ ,  $\eta_2 = (P, 0)$ , and  $\eta_3 = (P, Q)$  are elements of  $\mathbf{L}_S^\perp$ . Because of the symmetries of  $K(\vartheta, b)$  and  $\mathbf{L}_S$  the sets  $K(\vartheta, b) + \eta$ ,  $\eta \in \mathbf{L}_S^\perp$  will be disjoint if and only if the sets  $K(\vartheta, b) + \eta_i$ ,  $i = 1, 2, 3$ , do not intersect  $K(\vartheta, b)$ . The sets  $K(\vartheta, b) + (0, Q)$  and  $K(\vartheta, b) + (P, 0)$  will be disjoint from  $K(\vartheta, b)$  if and only if  $Q \geq 2rb$  and  $P \geq \max(2rb/(1 + \vartheta r), (2 - \vartheta)b/\vartheta)$ , respectively. Usually  $\vartheta$  is chosen sufficiently close to 1 for the latter condition to simplify to  $P \geq 2rb/(1 + \vartheta r)$ . In this case it turns out that for the minimal choice

$$(3.2) \quad P = \frac{2rb}{1 + \vartheta r}, \quad Q = 2rb$$

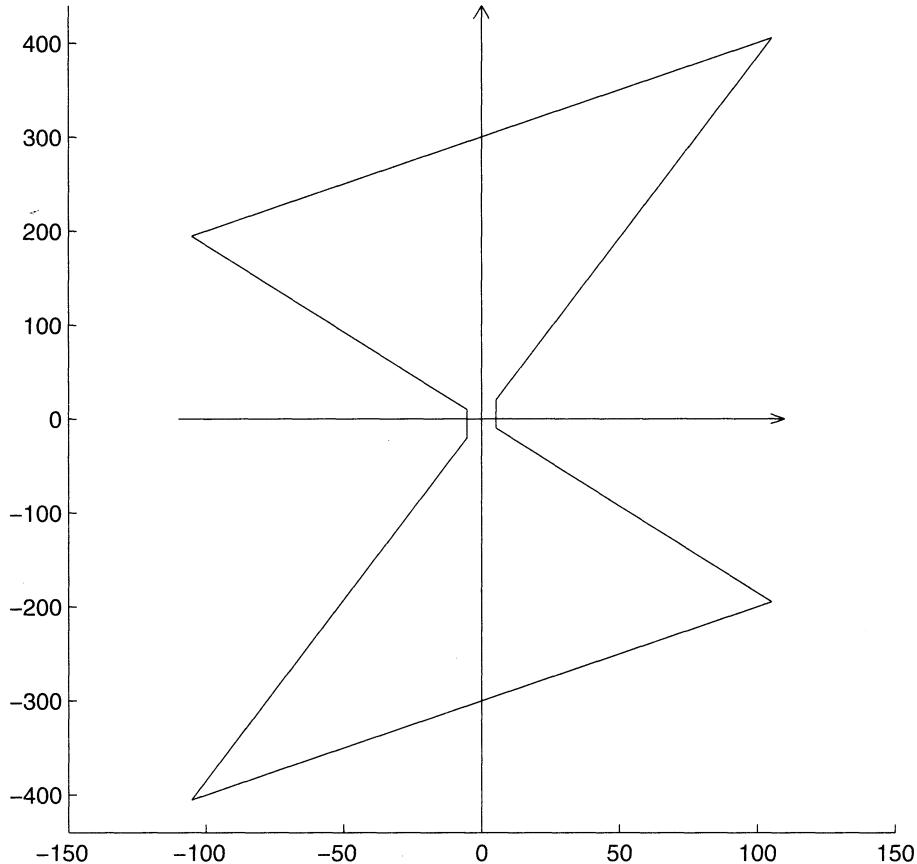


FIGURE 1. The set  $K(\vartheta, b)$  of (3.1) for  $\vartheta = 0.95$ ,  $b = 100$ , and  $r = 3$ . The coordinates of the four outer corners are  $\pm(b/\vartheta, (1+\vartheta r)b/\vartheta)$  and  $\pm(b/\vartheta, (1-\vartheta r)b/\vartheta)$ . The coordinates of the four corners near the origin equal  $(1-\vartheta)$  times the coordinates of the outer corners. In the case  $\vartheta = 1$  these four inner corners coincide at the origin, resulting in the set  $K$  given in [46] and [49, p. 75].

the set  $K(\vartheta, b) + (P, Q)$  is also disjoint from  $K(\vartheta, b)$ , so that all sets  $K(\vartheta, b) + \eta$ ,  $\eta \in \mathbf{L}_S^\perp$  are disjoint; see Figure 2. From this figure we can also see that keeping  $P$  fixed and slightly increasing  $Q$  would move the set  $K(\vartheta, b) + (P, Q)$  higher and may lead to its intersection with  $K(\vartheta, b)$ . Thus the sampling conditions would be violated in spite of having sampled more data. This phenomenon comes from the non-convexity of  $K(\vartheta, b)$  and will be discussed further in §3.3 below. Avoiding this intersection of sets requires further restrictions on the choice of  $Q$  as long as  $P < 2b/\vartheta$ . We obtain the following conditions.

$$(3.3) \quad \text{If } \max\left(\frac{2rb}{1+\vartheta r}, (2-\vartheta)b/\vartheta\right) \leq P < 2b/\vartheta, \quad \text{then choose either } 2rb \leq Q \leq (1+\vartheta r)P \quad \text{or} \quad Q \geq 2rb + P.$$

The conditions simplify considerably if  $P \geq 2b/\vartheta$ , since then the sets  $K(\vartheta, b)$  and  $K(\vartheta, b) + (P, Q)$  cannot intersect. Hence the sets  $K + \eta$ ,  $\eta \in \mathbf{L}_S^\perp$  will be disjoint if

$$(3.4) \quad P \geq 2b/\vartheta, \quad Q \geq 2rb.$$

Together the conditions (3.3) and (3.4) form a necessary and sufficient set of sampling conditions for the standard lattice. For comparison of these conditions with earlier results in the literature it is sometimes helpful to rewrite them in terms of the angular increments  $\Delta\beta = 2\pi/P$  and  $\Delta\alpha = 2\pi/Q$ .

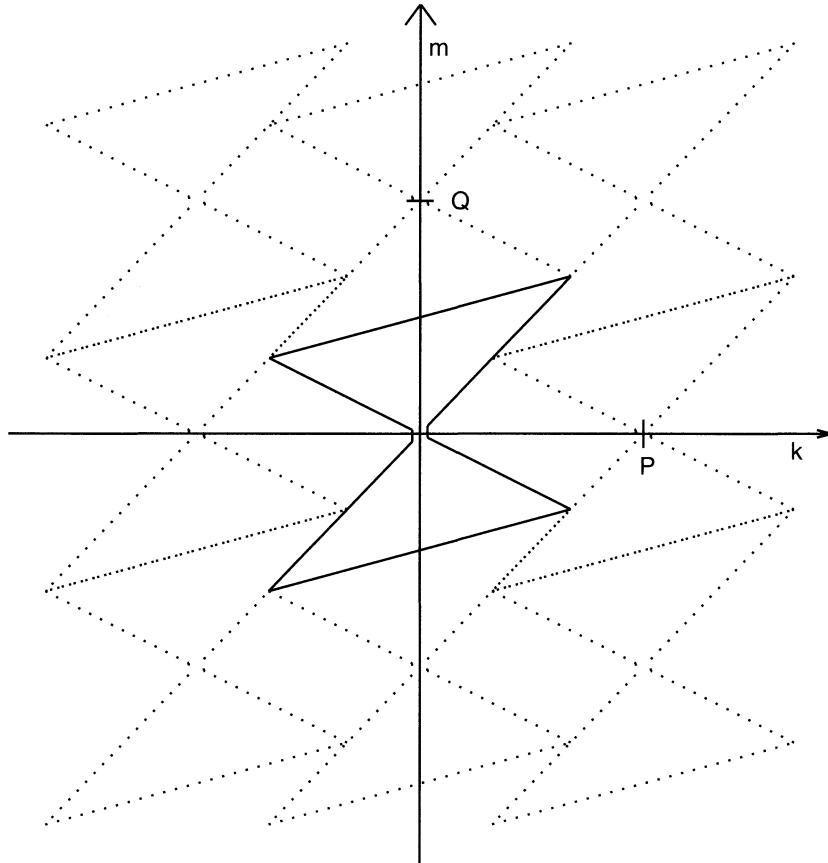


FIGURE 2. The translated sets  $K(\vartheta, b) + \eta$  for the standard lattice in case of  $P = \frac{2rb}{1+\vartheta r}$ ,  $Q = 2rb$ ,  $b = 100$ ,  $r = 3$ ,  $\vartheta = 0.95$ .

The translated sets in Figure 2 do not appear to be packed as densely as possible. Another arrangement corresponding to a different lattice may result in a denser packing, giving a denser reciprocal lattice and therefore a sparser sampling lattice. However, it is apparent from letting  $k_1 = 1$ ,  $k_2 = 0$  in (2.12) that the point  $\eta = (P, 0)$  always belongs to  $\mathbf{L}^\perp(N, P, Q)$ . Hence for every sampling lattice  $P$  needs to be chosen no smaller than  $\max(2rb/(1 + \vartheta r), (2 - \vartheta)b/\vartheta)$ , since otherwise the sets  $K(\vartheta, b)$  and  $K(\vartheta, b) + (P, 0)$  would intersect. Hence the standard lattice is optimal in the sense that it allows for parameter choices with a

minimal number of source positions. Since the total number of measurements is  $PQ$ , other lattices can only be more efficient overall by allowing values of  $Q$  less than  $2rb$  or by exploiting the symmetry (2.5). On the other hand the choice  $k_1 = N/\gcd(P, N)$ ,  $k_2 = P/\gcd(P, N)$  in (2.12) reveals that the reciprocal lattice always contains the point  $\eta = (0, QP/\gcd(P, N))$ . This yields the necessary condition  $Q \geq 2rb\gcd(P, N)/P$ . If  $\gcd(P, N) < P$  this would allow for values of  $Q$  smaller than  $2rb$ . However, the above condition is only necessary but in general not sufficient to avoid overlap of the translated sets  $K + \eta$ ,  $\eta \in \mathbf{L}^\perp$ .

Natterer [46] showed that in the case  $\vartheta = 1$  a lattice  $\mathbf{L}$  of optimal sparsity is given if  $b$  and  $rb$  are integers and the generator matrix of the reciprocal lattice has the form

$$(3.5) \quad W^{-T} = \begin{pmatrix} b & 0 \\ (1-r)b & 2rb \end{pmatrix}, \quad b, rb \in \mathbb{Z}.$$

(see also [47] and [49, p. 77]). According to Proposition 2.5 there exists a unique matrix of the form (2.10) that generates the same reciprocal lattice. This matrix can be found using an algorithm similar to the one described in [6]. This yields the lattice parameters

$$(3.6) \quad Q = \gcd((r-1)b, 2rb), \quad P = 2rb^2/Q, \quad N = -\text{mod}(nb, -P),$$

where the integer  $n$  is found by the relation  $m2rb - n(r-1)b = Q$ , with  $m \in \mathbb{Z}$ , and  $\text{mod}(nb, -P)$  denotes the unique integer in  $[1-P, 0]$  which differs from  $nb$  by a multiple of  $P$ . For example, let  $r = 3$  and  $b = 100$ . Then  $Q = \gcd(200, 600) = 200$ ,  $P = 300$ ,  $m = 0$ ,  $n = -1$ ,  $N = 100$ . In comparison, the sampling conditions for the standard lattice require at least  $P = 2rb/(1+r) = 150$  and  $Q = 2rb = 600$ . So the total number  $PQ$  of lattice points for the efficient lattice is in this case  $2/3$  of the number of points required for the standard lattice. In general this ratio equals  $(1+r)/(2r)$ , which approaches 1 for  $r \rightarrow 1$  and  $1/2$  in the limit  $r \rightarrow \infty$ . Our numerical experiments indicate that the set  $K$  may be slightly too small in the case  $\vartheta = 1$ . In case of  $\vartheta < 1$  the values of  $P$  and  $N$  have to be slightly increased for the efficient lattice, leaving the ratio  $P/N$  unchanged.

Figure 3 shows the case of (nearly) optimally sparse sampling with the efficient lattice, for  $b = 100$ ,  $r = 3$ ,  $\vartheta = 0.95$ ,  $P = 330$ ,  $N = 110$ ,  $Q = 200$ . It obtains the same theoretical resolution (as determined by the bandwidth  $b$ ) as the standard lattice with approximately 73% of the amount of data required for the standard lattice.

Practical drawbacks of the efficient lattice include the presence of the dynamic detector shift since  $N \neq 0$  which may be inconvenient to realize in practice, and that for some values of  $r$  and  $b$  the value for  $P$  in (3.6) can become very large and the value for  $Q$  very small. We will disregard these difficulties for the moment and explore if indeed accurate reconstructions can be obtained from efficiently sampled data.

**3.2. Reconstruction.** There are at least two ways in which to approach the reconstruction of images from the sampled data. First, one could use the sampled data directly as input for a reconstruction algorithm, for example the filtered back-projection algorithm. Second, one could first interpolate the sampled data to a denser lattice using the sampling theorem, and then reconstruct from these interpolated data. In the parallel-beam case it was shown in [38, 17, 22] that filtered backprojection can be used directly, even with efficiently sampled data, although

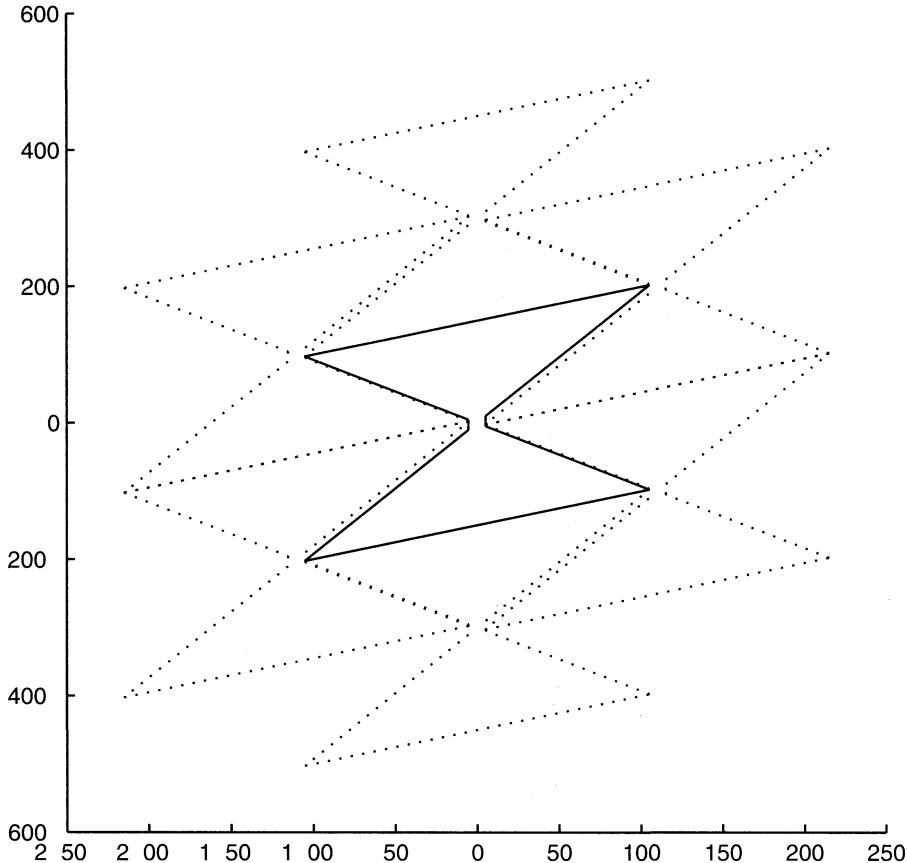


FIGURE 3. Some of the translated sets  $K(\vartheta, b) + \eta$ ,  $\eta \in \mathbf{L}^\perp(N, P, Q)$  for the efficient lattice in case of  $r = 3$ ,  $b = 100$ ,  $\vartheta = 0.95$ ,  $N = 110$ ,  $P = 330$ ,  $Q = 200$ .

experiments in [16] indicate that the second approach may be slightly better. We now investigate this situation for the fan-beam geometry and compare using the filtered backprojection algorithm directly with the method of first bandlimiting and interpolating the data onto a denser lattice and only then reconstructing with the filtered backprojection algorithm. We use the version of the fan-beam filtered backprojection algorithm as described in [45] or [35], with a convolution kernel similar to [35, §3.4.1], but replace the straight ramp filter  $h(t)$  used in [35, p. 82] with the Shepp-Logan kernel as given in [45, p. 111]. The algorithm used for bandlimiting and interpolating the data is described in [18, §4].

We will use simulated data from a very simple object for our investigation, namely the function

$$(3.7) \quad f(y) = (1 - 100|y - y_0|^2)_+^3, \quad y_0 = (0.4, 0.7)$$

where the  $+$  symbol indicates that  $f(y) = 0$  whenever  $(1 - 100|y - y_0|^2) < 0$ . The function  $f(y)$  is supported in the region  $|y - y_0| \leq 0.1$ . Since  $f$  is quite smooth, it is essentially bandlimited and therefore provides a good test for our theory. Choosing

a cut-off frequency  $b = 100$  will be sufficient for a good reconstruction. Since  $f$  is known analytically, we can use not only the visual impression of the images but also the discrete relative  $l_2$ -error as a measure for the accuracy of the reconstruction. If  $I(x_n, y_m)$  is the reconstructed image, the relative  $l_2$ -error is given by

$$E = \left( \frac{\sum_{n,m} |f(x_n, y_m) - I(x_n, y_m)|^2}{\sum_{n,m} |f(x_n, y_m)|^2} \right)^{\frac{1}{2}}.$$

The upper left image in Figure 4 shows the direct reconstruction with the filtered backprojection algorithm from data sampled on the standard lattice with  $b = 100$ ,  $r = 3$ ,  $P = \text{ceil}(2rb/(1 + \vartheta r)) = 156$ ,  $N = 0$ , and  $Q = 2rb = 600$ . The discrete convolution occurring in the algorithm is computed on a very dense grid with stepsize  $H = \pi/(8Q) = \pi/4800$  in order to suppress errors stemming from the interpolation step of the algorithm; cf. [21, §6], [22]. The reconstruction is computed on a  $256 \times 256$  grid. The maximum of  $f(x)$  is 1, but since we want to study small artifacts the display window is such that values below  $-0.01$  are rendered black and values above  $0.01$  are rendered white. The relative  $l_2$ -error is about 5.4%. The lower left image shows the reconstruction after interpolating the data first onto the denser lattice with  $P = 274$ ,  $Q = 892$ ,  $N = 0$  and then using the filtered backprojection algorithm. The interpolation was done using the sampling theorem and the algorithm described in [18, §4]. It computes the Fourier transform of the data inside  $K$ , sets it to zero outside  $K$ , and then computes the data on the denser lattice using an Inverse Fast Fourier Transform. The relative  $l_2$ -error is now about 2.4%, less than half of the error of the direct reconstruction. The difference between the two methods of reconstruction is even more pronounced in case of the efficient lattice. The upper right image in Figure 4 shows the direct reconstruction from efficiently sampled data with  $P = 330$ ,  $Q = 200$ ,  $N = 110$ , and the other parameters as above. There are strong artifacts and the relative  $l_2$ -error is now about 52%. The lower right image shows the reconstruction with prior interpolation onto the standard lattice with parameters  $P = 274$ ,  $Q = 892$ ,  $N = 0$ , and the discrete convolution computed on a grid with a stepsize of  $H = \pi/4906$ . The relative  $l_2$ -error is about 2.4%, almost exactly the same as for the reconstruction in the lower left image. This experiment indicates that accurate reconstruction from efficiently sampled data is possible if the data are first bandlimited to  $K$  and then interpolated onto a denser grid. On the other hand, the filtered backprojection algorithm in its usual form does not appear to achieve the theoretically possible resolution when reconstructing directly from the data without prior interpolation to a denser grid. A heuristic explanation for this behavior could be as follows: The change of variables from parallel-beam to fan-beam yields for the convolution step of the algorithm a cut-off frequency  $b' = |x - z(\beta)|b$  which depends on the reconstruction point  $x$  and the source location  $z(\beta) = r(\cos(\beta), \sin(\beta))$ . In order to obtain a fast algorithm  $b'$  is replaced by a constant  $b_c$ ; cf. [45, p. 113]. It appears that in order to obtain the desired resolution  $b_c$  should not be smaller than the maximum of  $|x - z(\beta)|b$ , i.e.,  $(r + 1)b$ . On the other hand, the values of  $P$  and  $Q$  used for the reconstruction should be large enough for the numerical integrations occurring in the algorithm to be accurate for such a value of  $b_c$ . The reconstructions from efficiently sampled data presented in [46] where performed with a modified filtered backprojection algorithm (see [46, Eq. (4.7)]) which apparently avoided replacing  $b'$  by a constant, but may be slow as a consequence. Since the bandlimiting

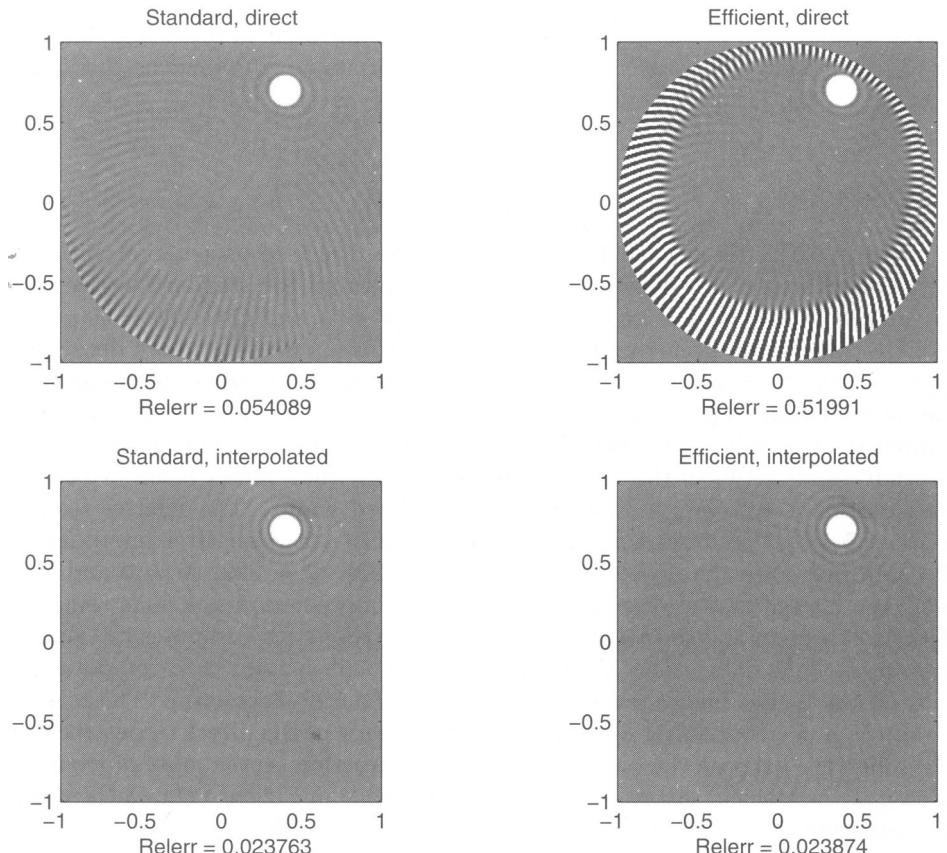


FIGURE 4. Reconstructions of the function  $f(x)$  of (3.7). All reconstructions are computed with  $b = 100$ ,  $r = 3$ , and displayed on a  $256 \times 256$  grid. The function takes values between 0 and 1, but in order to highlight small artifacts the display is such that values smaller than  $-0.01$  are rendered black and values larger than  $0.01$  are rendered white. Upper row: Direct reconstructions with filtered backprojection algorithm. Upper left: Data sampled on the standard lattice with  $P = 156$ ,  $N = 0$ ,  $Q = 600$ . Upper right: Efficient lattice with  $P = 330$ ,  $N = 110$ ,  $Q = 200$ . Lower row: Reconstructions after bandlimiting the data to  $K(\vartheta, b)$ ,  $\vartheta = 0.95$ , and then interpolating onto a denser standard lattice with  $P = 274$ ,  $N = 0$ ,  $Q = 892$  prior to reconstruction with filtered backprojection. Lower left: Original data sampled on standard lattice as in upper left. Lower right: Original data sampled on efficient lattice as in upper right.

and interpolation by the sampling theorem is much faster than reconstruction with the filtered backprojection algorithm, adding this step does not lead to a significant slowdown of the reconstruction.

**3.3. Qualitative understanding of artifacts from undersampling.** Now we investigate the effects of undersampling, that is, of violating the sampling conditions requiring the translates  $K + \eta$ ,  $\eta \in \mathbf{L}^\perp$  to be disjoint. Consider the following numerical experiments. Figure 5 again shows various reconstructions of the func-

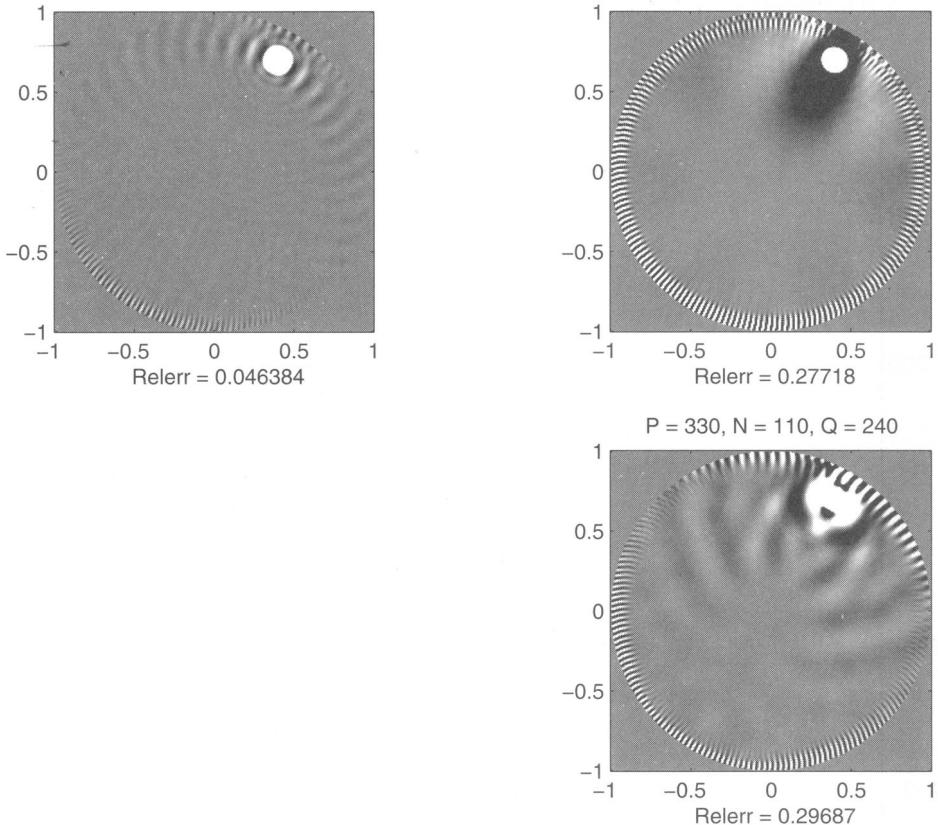


FIGURE 5. Artifacts resulting from undersampling. The function to be reconstructed is  $f(x)$  from (3.7). All reconstructions are computed with  $b = 100$ ,  $r = 3$ , and displayed on a  $256 \times 256$  grid such that values less than  $-0.01$  are rendered black and values greater than  $0.01$  are rendered white. Top row: Effects of  $P$  being too small. Top left: Reconstruction from standard lattice with  $P = 140$  instead of 156. Top right: Reconstruction from efficient lattice with  $P$  reduced to 300 from 330. Bottom right: Reconstruction from efficient lattice with  $Q$  increased from 200 to 240.

tion  $f(x)$  given in (3.7), from fan beam data with source radius  $r = 3$  and cut-off frequency  $b = 100$ .

The upper left picture shows a reconstruction with the standard lattice with  $Q = 600$  and  $P$  reduced to 140 from the value of 156 required by the sampling conditions (3.3) for  $r = 3$ ,  $b = 100$ , and  $\vartheta = 0.95$ . This reduction of  $P$  by about

10% results in a relative  $l_2$ -error of about 4.6%. The upper right picture shows the corresponding case for the efficient lattice, where  $P$  has been reduced by about 10% from 330 to 300. The resulting artifacts are much stronger, resulting in a relative error of about 28%. The picture in the lower right shows an at first glance surprising result. Here we used the efficient lattice and *increased* the parameter  $Q$  from 200 to 240. In spite of having sampled more data we obtain strong artifacts and a relative error of almost 30%.

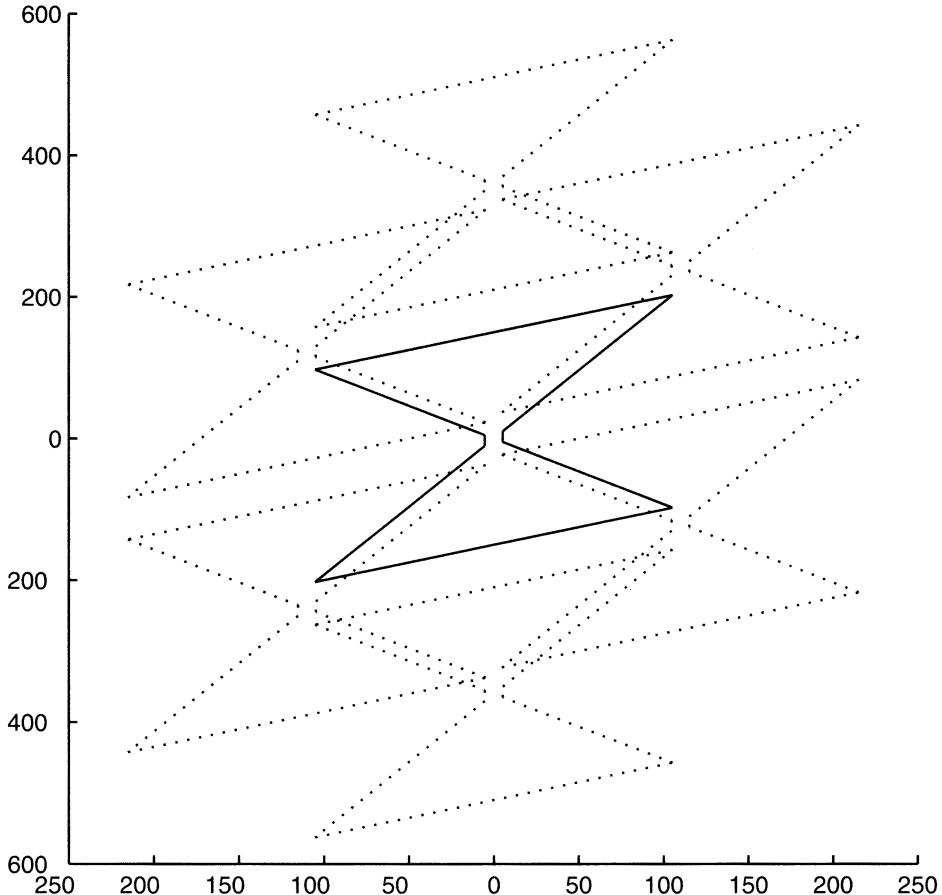


FIGURE 6. The translated sets  $K(\vartheta, b) + \eta$ ,  $\eta \in \mathbf{L}^\perp$  for the efficient lattice in case of  $r = 3$ ,  $b = 100$ ,  $\vartheta = 0.95$ ,  $P = 330$ ,  $N = 110$ , and  $Q = 240$ . The increased value of  $Q$  compared to Figure 3 leads to overlap of the translated sets and thus aliasing, in spite of sampling more data than with the correct choice  $Q = 200$ .

The classical sampling theorem can help us understand these experiments at least qualitatively. In Figure 6 we see the translates  $K(\vartheta, b) + \eta$ ,  $\eta \in \mathbf{L}^\perp$  for the case corresponding to the lower right reconstruction in Figure 5. We see that increasing  $Q$  has led to an overlap of the translates of  $K$  and thus undersampling, in spite of having sampled more data. This explains the presence of the artifacts. Increasing

$P$  while leaving both  $Q$  and  $N$  fixed can also lead to overlap of the translated sets. On the other hand, sampling more data by leaving  $Q$  fixed and increasing  $P$  and  $N$  such that  $P/N$  remains constant does not lead to undersampling. In this case the translated sets  $K + \eta$  in Figure 3 move further apart from each other in the horizontal direction and remain disjoint.

As our numerical experiment indicated, the efficient lattice shows much greater sensitivity with regard to undersampling in  $P$  than the standard lattice. The pattern of overlap of the sets  $K + \eta$  provides a qualitative explanation. Figure 7 shows the translated sets  $K + \eta$  in case of the standard lattice with  $P = 140$ ,  $N = 0$ , and  $Q = 600$ , which partially overlap for this choice of parameters. The effect of such overlap can be investigated with the help of the Poisson summation formula (2.13). Assume we wish to compute an approximation for the Fourier transform  $\hat{g}(\zeta)$  from

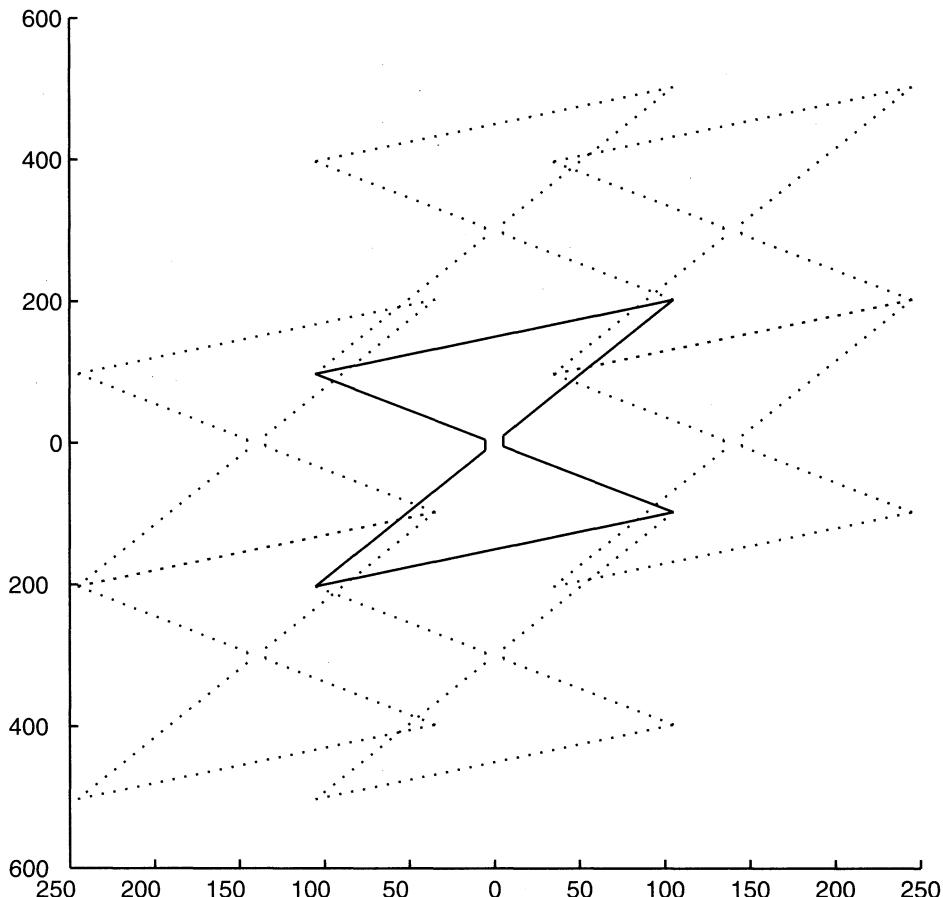


FIGURE 7. The translated sets  $K(\vartheta, b) + \eta$ ,  $\eta \in \mathbf{L}^\perp$  for the standard lattice with parameters  $r = 3$ ,  $b = 100$ ,  $\vartheta = 0.95$ ,  $P = 140$ ,  $N = 0$ , and  $Q = 600$ . The decreased value of  $P$  compared to Figure 2 leads to overlap.

the values of  $g$  on a sampling lattice  $\mathbf{L}$  by means of taking the discrete Fourier transform with respect to  $\mathbf{L}$ . The result would be  $\hat{g}(\zeta) \simeq (1/PQ) \sum_{y \in \mathbf{L}} g(y) e^{-2\pi i \langle y, \zeta \rangle}$ , i.e., just the left-hand side of the Poisson summation formula (2.13) for  $z = 0$ . According to the Poisson summation formula we have

$$\hat{g}(\zeta) \simeq (1/PQ) \sum_{y \in \mathbf{L}} g(y) e^{-2\pi i \langle y, \zeta \rangle} = \sum_{\eta \in \mathbf{L}^\perp} \hat{g}(\zeta + \eta) = \hat{g}(\zeta) + \sum_{0 \neq \eta \in \mathbf{L}^\perp} \hat{g}(\zeta + \eta).$$

If the translated points  $\zeta + \eta$  lie outside  $K = K(\vartheta, b)$  for  $\eta \neq 0$  we may assume that their contribution is small, so that we obtain a good approximation for  $\hat{g}(\zeta)$ . For  $\zeta \in K$  this will be the case if the translated sets  $K + \eta$  are disjoint. If the sets  $K + \eta$  are not disjoint, there will be non-negligible error terms  $\hat{g}(\zeta + \eta)$ , with  $\zeta + \eta$  lying in the part of  $K$  which is overlapped by other translates. In the case of Figure 7 we see that these regions lie well away from the origin. Since  $f$  is non-negative (as are almost all functions encountered in tomography) the data function  $g$  will be non-negative as well. Hence  $|\hat{g}|$  will assume its maximum at the origin and decrease away from the origin. So we can assume that the error resulting from the terms  $\hat{g}(\zeta + \eta)$  with  $\zeta + \eta$  not close to the origin will be at most moderately large. This gives a qualitative explanation that the standard lattice is not overly sensitive to undersampling with regard to  $P$ . By a similar argument we may expect that violating the sampling conditions (3.3) by choosing  $Q$  in the excluded range  $(1 + \vartheta r)P < Q < 2rb + P$  will in most cases not cause significant artifacts.

The situation is different with the efficient lattice. Figure 8 shows the translated sets  $K + \eta$  corresponding to the reconstruction in the lower right of Figure 5. Comparing this to Figure 3 we see that decreasing  $P$  from 330 to 300 leads to overlap of  $K$  by some of the translated sets, in particular near the origin where  $\hat{g}$  is largest. So some of the error terms  $\hat{g}(\zeta + \eta)$  will be large, making the efficient lattice considerably more sensitive with regard to undersampling in  $P$  than the standard lattice.

A similar discussion has been given for the parallel-beam case in [21], followed by a detailed quantitative error analysis explaining also the location of the artifacts resulting from undersampling. Comparing the two sampling geometries we find that the standard lattice for the fan-beam geometry is more sensitive to undersampling with regard to  $P$  than the parallel-beam standard lattice is with regard to undersampling in the variable  $\varphi$ .

**3.4. Error analysis of reconstruction algorithms.** The most popular tomographic reconstruction algorithm is the so-called filtered backprojection algorithm. It is based on the approximate inversion formula

$$(3.8) \quad e * f(x) = \int_0^{2\pi} \int_{\mathbb{R}} k(\langle x, \theta \rangle - s) Rf(\varphi, s) ds d\varphi$$

where  $e$  is an approximate  $\delta$ -function and the kernel  $k$  can be computed from  $e$ ; see, e.g., [45, p. 102]. The relation (3.8) can be verified by writing  $e * f$  as  $e * f(x) = \int \hat{e}(\xi) \hat{f}(\xi) e^{i\langle x, \xi \rangle} d\xi$ , expressing the integral in polar coordinates, and using the relation (2.3); see, e.g., [20].

Discretizing the integrals in equation (3.8) by using the trapezoidal rule yields the filtered backprojection algorithm for the parallel-beam geometry. The appropriate Poisson summation formula furnishes an error estimate. If the sampling

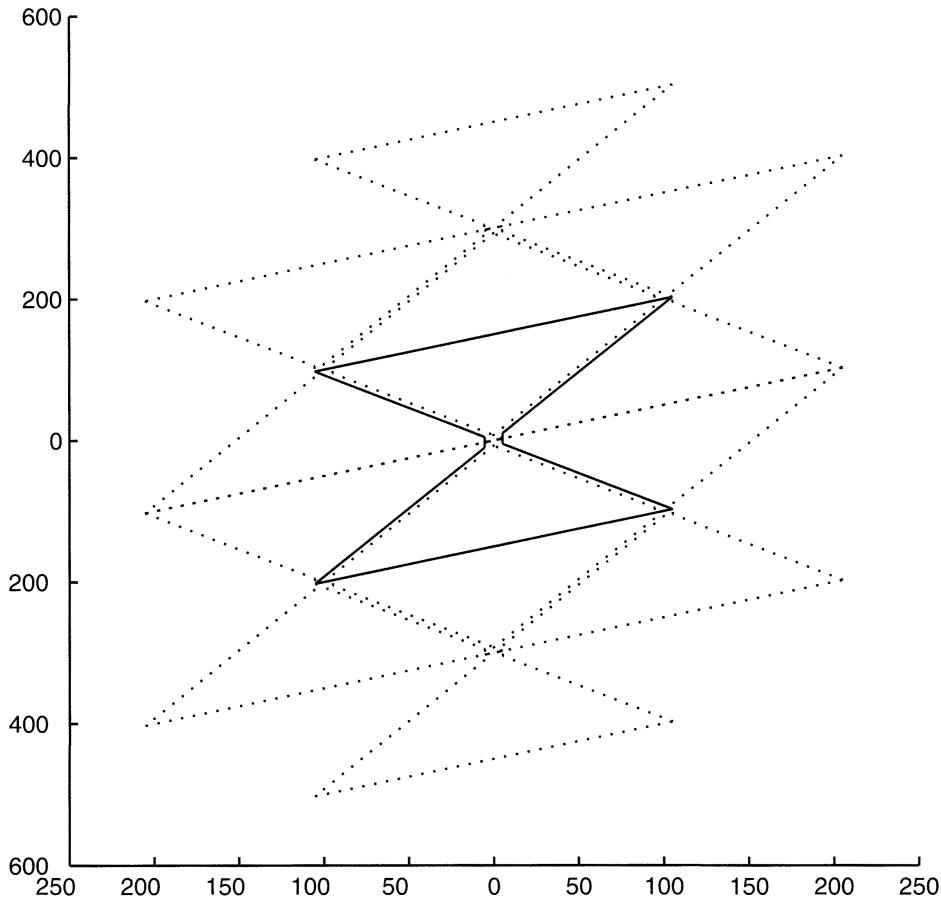


FIGURE 8. The translated sets  $K(\vartheta, b) + \eta$ ,  $\eta \in \mathbf{L}^\perp$  for the efficient lattice with parameters  $r = 3$ ,  $b = 100$ ,  $\vartheta = 0.95$ ,  $P = 300$ ,  $N = 100$ , and  $Q = 200$ . The decreased value of  $P$  compared to Figure 3 leads to overlap.

conditions are satisfied, this error will be small [38]. Nevertheless, at first reconstructions with the efficient parallel-beam lattice at full resolution showed large errors. The reason turned out to be an interpolation step in the algorithm. This interpolation is harmless for the standard lattice, but has to be carried out very accurately for the efficient lattice in order to obtain good reconstructions. For details see [38, 17, 22, 21]. Convergence rates for the filtered backprojection algorithm were obtained in [56] with an approach unrelated to sampling theory.

The filtered backprojection algorithm for the fan-beam geometry is obtained by using the change of variables  $\varphi = \alpha + \beta - \pi/2$ ,  $s = r \sin \alpha$  in (3.8) and making additional approximations as mentioned above; see, e.g., [45, §V.1.2]. A complete error analysis analogous to the parallel-beam case appears to be still outstanding, but current research may result in considerable progress [33]. Kruse [38] provided

an error analysis for Marr's algorithm [43], which was found to perform superior to direct reconstruction with the fan-beam filtered backprojection algorithm.

#### 4. Exploiting the symmetry: An application of periodic sampling to increase resolution in fan-beam tomography

Periodic sampling sets are unions of shifted copies of a lattice. Such sets are called periodic since they are invariant under shifts with elements of the lattice, since lattices are subgroups. Sampling theorems for periodic sampling sets are more complicated than the classical sampling theorem, but can still be proved using the Poisson summation formula; see, e.g., [16, 18]. In this section we briefly describe a very interesting recent development involving periodic sampling. So far we have not yet made use of the symmetry property (2.5). In the parallel-beam case the analogous relation (2.2) is used by choosing lattices which also possess the symmetry, so that only half of the lattice points need to be measured. For special lattices in the fan-beam geometry e.g., the efficient lattice (3.5) introduced in [46] this is also the case, but not in general. Izen, Rohler, and Sastry [32] discovered a way to exploit the symmetry relation (2.5) to increase resolution. If  $Q = 2q$  is even, the (shifted) standard sampling lattice in the  $(\beta, \alpha)$  coordinates is given by

$$(4.1) \quad \begin{aligned} \mathbf{L}_S = & \{(\beta_j, \alpha_l) : \beta_j = 2\pi j/P, \alpha_l = \pi(l + \delta)/q, \\ & j = 0, \dots, P - 1, l = -q, \dots, q - 1, \delta \geq 0\}, Q = 2q. \end{aligned}$$

Note the constant shift  $\delta\pi/q$  in the  $\alpha$  variable. Many scanners use  $\delta = 1/4$  to reduce data redundancy. If  $Df$  is sampled on the standard lattice (4.1), then using the symmetry (2.5) gives us additional data on a 'reflected lattice'

$$\mathbf{L}_R = \{(\beta_j + 2\alpha_l + \pi, -\alpha_l) : (\beta_j, \alpha_l) \in \mathbf{L}_S\}.$$

Since the union  $\mathbf{L}_S \cup \mathbf{L}_R$  is in general not a lattice (or shifted lattice) itself, the classical sampling theorem cannot be applied. In addition, since  $\mathbf{L}_R$  differs from  $\mathbf{L}_S$  by more than a constant shift,  $\mathbf{L}_S \cup \mathbf{L}_R$  is a union of two different lattices, so a periodic sampling theorem which handles sampling sets which are unions of shifted copies of the same lattice does at first glance also not apply. However, Izen et al. [32] discovered that  $\mathbf{L}_S \cup \mathbf{L}_R$  is a union of  $2q/\gcd(P, q)$  shifted copies of the smaller lattice

$$\mathbf{L}_P = \{(2\pi j/P, \pi l/\gcd(P, q)), j = 0, \dots, P - 1, |l| \leq \gcd(P, q)\}.$$

While Izen, Rohler, and Sastry used a different approach to reconstruction, their discovery makes it possible to apply the periodic sampling theorem and interpolation algorithms of [18]. This is somewhat more complicated but also more general than the approach of [32]. A first demonstration has recently been given by Mitchell [44]. While normally doubling the resolution would require four times the data, using the reflected data allows us to achieve twice the resolution achievable with  $\mathbf{L}_S$  alone by only having to double the number  $P$  of source positions. According to [32] this also overcomes the following problem, related to the beams of finite width mentioned by Cormack in the quote at the beginning of this article. It can be shown that the natural band-limiting of the data caused during the measurement process by the averaging over the finite detector width  $d$  (i.e., by using beams of finite width instead of measuring line integrals) corresponds to a maximum frequency of approximately  $2\pi/d$ ; cf. [35, §5.1.2]. On the other hand, the highest sampling rate in the  $\alpha$  variable occurs when the detectors are adjacent, so it is equal to  $d$ . Hence

according to Theorem 2.2 the highest resolvable frequency would be  $b = \pi/d$ , only half as large as the band-width of the measured data. Using the reflected data as described above now allows to achieve the full resolution.

## 5. Conclusions

This article has given an introduction to the use of sampling theory in tomography using the example of the two-dimensional fan-beam geometry. The approach taken here involved a change of variables so that existing theory and software [18] for sampling on the torus group  $\mathbb{T}^2$  could be used. A parameterization of all sampling lattices was given. The commonly used standard lattice and an efficient lattice similar to the one of [46] were given special attention, including a complete set of sampling conditions for the standard lattice.

Numerical experiments were presented, indicating that unlike the parallel-beam case, a direct application of the fan-beam filtered backprojection algorithm does give suboptimal, and in case of the efficient lattice unacceptable, results. On the other hand, using the sampling theorem to first bandlimit the data to the set  $K(\vartheta, b)$  and then interpolating it to a dense lattice prior to reconstruction yields accurate reconstructions, at least for a smooth test object.

A qualitative explanation of artifacts resulting from undersampling was given with results similar to the parallel-beam case, the most significant difference being the somewhat higher sensitivity of the fan-beam standard lattice with regard to undersampling in the number of source positions. Finally, a new way to exploit the symmetry of the data to increase resolution was reviewed.

Two practically important issues left out of the preceding discussion are functions with discontinuities and the effect of noise. Discontinuities cause a slower decay of the Fourier transform and thus the assumption of an essential bandwidth  $b$  is less well satisfied. A common way to deal with the resulting artifacts is to filter the data prior to reconstruction. The trade-off is that this gives up some of the higher resolution gained by efficient sampling. Noise in the data is similarly more critical when reconstructing at high resolution with sparsely sampled data. One way to deal with this problem is to denoise the reconstructed image with a denoising method that preserves edges, such as the method of Rudin-Osher-Fatemi [58] based on minimizing the total variation. Recently Hass [27] has demonstrated that such denoising may remove the increased effects of noise and still retain somewhat higher resolution.

We conclude by giving a list of examples of areas where sampling theory and tomography have interacted and stimulated each other. An area of increasing current interest is sampling in three-dimensional tomography. The list below is by no means complete but is meant to provide interested readers with an opportunity for further study.

**5.1. Examples for the interaction of sampling theory and tomography.** A brief overview of how sampling theory and tomography have interacted and stimulated each other may contain at least the themes and topics listed below.

### 5.1.1. Applications of the Classical Sampling Theorem.

- *Identification of efficient sampling lattices in 2D and 3D.* The goal is to obtain a desired resolution with a minimum number of measurements. See, e.g., [11, 12, 13, 14, 19, 31, 40, 45, 46, 47, 48, 49, 55].

- *Qualitative understanding of artifacts.* The classical sampling theorem permits a qualitative understanding of artifacts caused by aliasing from undersampling; see, e.g., [17, 21, 31].
- *Error analysis of reconstruction algorithms.* The Poisson summation formula connects sampling theory to the error analysis of the filtered back-projection algorithm; see, e.g., [17, 21, 22, 33, 38, 39, 45, 49]. For results on algebraic algorithms see [10, 36].

5.1.2. *Research in Sampling Theory stimulated by CT.* Different data acquisition geometries in tomography stimulate interest in:

- *Sampling theorems where the sampling set is not a lattice.*
  - *Multidimensional periodic sampling sets.* Here the sampling set is a union of several cosets of a lattice. See [16, 18].
  - *Non-periodic sampling on unions of shifted lattices* [2, 59].
  - *Non-uniform sampling.* Here the sampling set has no structure which could be exploited with the Poisson summation formula. Examples include polar or spiral sampling in Fourier space; see, e.g., [1, 5, 23, 53].
- *An estimate for the aliasing error sharper than (2.15)* [19].
- *A unified mathematical framework.* Many applications of sampling can be unified in the simple and elegant framework of Fourier analysis on locally compact abelian groups; see, e.g., [2, 15, 18, 37].

5.1.3. *Applications of periodic sampling in CT.*

- *Additional efficient 2D sampling schemes* [9, 16].
- ‘Preferred pitch’ in 3D helical CT [25, 57].
- *Higher resolution in 2D fan-beam CT* [32].

5.1.4. *Applications of non-periodic sampling.*

- *Higher resolution in 2D fan-beam CT* [25].

## References

- [1] A. Aldroubi and K. Gröchenig, *Nonuniform sampling and reconstruction in shift-invariant spaces*, SIAM Review **43** (2001), 585–620.
- [2] H. Behmard and A. Faridani, *Sampling of bandlimited functions on unions of shifted lattices*, J. Fourier Anal. Appl. **8** (2002), 43–58.
- [3] J. J. Benedetto and P. J. S. G. Ferreira (eds.), *Modern Sampling Theory: Mathematics and Applications*, Birkhäuser, Boston, 2001.
- [4] J. J. Benedetto and A. I. Zayed (eds.), *Sampling, Wavelets, and Tomography*, Birkhäuser, Boston, 2004.
- [5] M. Bourgeois, F. T. A. W. Wajer, D. van Ormondt, and D. Graveron-Demilly, *Reconstruction of MRI images from non-uniform sampling and its application to intrascan motion correction in functional MRI*, in [3], pp. 343–363.
- [6] G. H. Bradley, *Algorithms for Hermite and Smith normal matrices and linear Diophantine equations*, Math. Comp. **25** (1971), 897–907.
- [7] P. L. Butzer, W. Splettstößer and R. L. Stens, *The sampling theorem and linear prediction in signal analysis*, Jber. d. Dt. Math.-Verein. **90** (1988), 1–70.
- [8] A. M. Cormack, *Sampling the Radon transform with beams of finite width*, Phys. Med. Biol. **23** (1978), 1141–1148.
- [9] L. Desbat, *Efficient sampling on coarse grids in tomography*, Inverse Problems **9** (1993), 251–269.
- [10] L. Desbat, *Algebraic reconstructions from efficient sampling schemes in tomography*, Computerized Tomography (M.M. Lavrent'ev, ed.), VSP, Utrecht, 1995, pp. 124–137.

- [11] L. Desbat, *Echantillonnage parallèle efficace en tomographie 3D*, C.R. Acad. Sci. Paris, Série I, **324** (1997), 1193–1199.
- [12] L. Desbat, *Interpolation of lacking data in tomography*, Proceedings of SampTA 2001, Intern. Conference on Sampling Theory and Applications, Orlando, FL, 2001, pp. 123–128.
- [13] L. Desbat, S. Roux, P. Grangeat, and A. Koenig, *Sampling conditions of 3D parallel and fan-beam x-ray CT with application to helical tomography*, Phys. Med. Biol. **49** (2004), 2377–2390.
- [14] L. Desbat and C. Mennessier, *Efficient sampling of the rotation invariant Radon transform*, in [3], pp. 365–377.
- [15] M. M. Dodson and M. G. Beaty, *Abstract harmonic analysis and the sampling theorem*, in [30], pp. 233–265.
- [16] A. Faridani, *An application of a multidimensional sampling theorem to computed tomography*, Contemporary Mathematics, **113** (1990), 65–80.
- [17] A. Faridani, *Reconstructing from efficiently sampled data in parallel-beam computed tomography*, Inverse Problems and Imaging (G.F. Roach, ed.), Pitman Research Notes in Mathematics Series, vol. 245, Longman, 1991, pp. 68–102.
- [18] A. Faridani, *A generalized sampling theorem for locally compact abelian groups*, Math. Comp. **63** (1994), 307–327.
- [19] A. Faridani, *Sampling in parallel-beam tomography*, Inverse Problems, Tomography, and Image Processing (A. G. Ramm, ed.), Plenum, New York, 1998, pp. 33–53.
- [20] A. Faridani, *Introduction to the Mathematics of Computed Tomography*, Inside Out: Inverse Problems and Applications (G. Uhlmann, ed.), MSRI Publications, vol. 47, Cambridge University Press, 2003, pp. 1–46.
- [21] A. Faridani, *Sampling theory and parallel-beam tomography*, in [4], pp. 225–254.
- [22] A. Faridani and E. L. Ritman, *High-resolution computed tomography from efficient sampling*, Inverse Problems **16** (2000), 635–650.
- [23] K. Fourmont, *Non-equispaced fast Fourier transforms with applications to tomography*, J. Fourier Anal. Appl. **9** (2003), 431–450.
- [24] G. B. Folland, *Real Analysis*, Wiley, New York, 1984.
- [25] L. A. Gratton, *Nonuniform Sampling Problems in Computed Tomography*, Ph.D. thesis, Dept. of Mathematics, Oregon State University, 2005.
- [26] K. Gröchenig, *Aspects of Gabor analysis on locally compact abelian groups*, Gabor Analysis and Algorithms (H. G. Feichtinger and T. Strohmer, eds.), Birkhäuser, Boston, 1998, pp. 211–231.
- [27] R. A. Hass, *De-noising tomography images through bounded variation*, M.S. paper, Dept. of Mathematics, Oregon State University, 2005.
- [28] J. R. Higgins, *Five short stories about the cardinal series*, Bull. Amer. Math. Soc. **12** (1985), 45–89.
- [29] J. R. Higgins, *Sampling Theory in Fourier and Signal Analysis: Foundations*, Clarendon Press, Oxford, 1996.
- [30] J. R. Higgins and R. L. Stens (eds), *Sampling Theory in Fourier and Signal Analysis: Advanced Topics*, Clarendon Press, Oxford, 1999.
- [31] T.-C. Hsung and D. P. K. Lun, *New sampling scheme for region-of-interest tomography*, IEEE Trans. Signal Processing **48** (2000), 1154–1163.
- [32] S. H. Izen, D. P. Rohler, and K.L.A Sastry, *Exploiting symmetry in fan-beam CT: Overcoming third generation undersampling*, SIAM J. Appl. Math. **65** (2005), 1027–1052.
- [33] S. H. Izen, *An analysis of the fan beam CT reconstruction kernel*, talk presented at the Special Session on Radon Transform and Inverse Problems, AMS National Meeting, Atlanta, January 2005.
- [34] A. J. Jerri, *The Shannon sampling theorem - its various extensions and applications: a tutorial review*, Proc. IEEE **65** (1977), 1565–1596.
- [35] A. C. Kak and M. Slaney, *Principles of computerized tomographic imaging*, IEEE Press, New York, 1988.
- [36] W. Klaverkamp, *Tomographische Bildrekonstruktion mit direkten algebraischen Verfahren*, Ph.D. thesis, Fachbereich Mathematik, Westfälische Wilhelms-Universität Münster, Germany, 1991.
- [37] I. Kluvanek, *Sampling theorem in abstract harmonic analysis*, Mat. Casopis Slovens. Akad. Vied., **15** (1965), 43–48.

- [38] H. Kruse, *Resolution of reconstruction methods in computerized tomography*, SIAM J. Sci. Stat. Comput. **10** (1989), 447–474.
- [39] R. M. Lewitt, R. H. T. Bates, and T. M. Peters, *Image reconstruction from projections: II: Modified back-projection methods*, Optik **50** (1978), 85–109.
- [40] A. G. Lindgren and P. A. Rattey, *The inverse discrete Radon transform with applications to tomographic imaging using projection data*, Advances in Electronics and Electron Physics **56** (1981), 359–410.
- [41] R. J. Marks II, *Introduction to Shannon Sampling and Interpolation Theory*, Springer, New York, 1991.
- [42] R. J. Marks II (ed.), *Advanced Topics in Shannon Sampling and Interpolation Theory*, Springer, New York, 1993.
- [43] R. B. Marr, *On the reconstruction of a function on a circular domain from a sampling of its line integrals*, J. Math. Anal. Appl. **45** (1974), 357–374.
- [44] N. Mitchell, *Multi-dimensional sampling in fan beam tomography*, M.S. paper, Dept. of Mathematics, Oregon State University, 2005.
- [45] F. Natterer, *The Mathematics of Computerized Tomography*, Wiley, 1986.
- [46] F. Natterer, *Sampling in fan-beam tomography*, SIAM J. Appl. Math. **53** (1993), 358–380.
- [47] F. Natterer, *Sampling and resolution in CT*, Computerized Tomography (M.M. Lavrent'ev, ed.), VSP, Utrecht, 1995, pp. 343–354.
- [48] F. Natterer, *Resolution and reconstruction for a helical CT scanner*, Technical Report 20/96-N, Fachbereich Mathematik, Universität Münster, Germany, 1996.
- [49] F. Natterer and F. Wübbeling, *Mathematical Methods in Image Reconstruction*, SIAM, Philadelphia, 2001.
- [50] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, Philadelphia, 1992.
- [51] M. Newman, *Integral Matrices*, Academic Press, 1972.
- [52] V. P. Palamodov, *Localization of harmonic decomposition of the Radon transform*, Inverse Problems **11** (1995), 1025–1030.
- [53] D. Potts and G. Steidl, *Fourier reconstruction of functions from their nonstandard sampled Radon transform*, J. Fourier Anal. Appl. **8** (2002), 513–533.
- [54] E. T. Quinto, *An introduction to x-ray tomography and Radon transforms*, this volume.
- [55] P. A. Rattey and A. G. Lindgren, *Sampling the 2-D Radon transform*, IEEE Trans. Acoust. Speech Signal Processing **29** (1981), 994–1002.
- [56] A. Rieder and A. Faridani, *The semi-discrete filtered backprojection algorithm is optimal for tomographic inversion*, SIAM J. Num. Anal. **41** (2003), 869–892.
- [57] P. J. La Riviere and X. Pan, *Pitch dependence of longitudinal sampling and aliasing effects in multi-slice helical computed tomography (CT)*, Phys. Med. Biol. **47** (2002), 2797–2810.
- [58] L.I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D **60** 1-4 (1992), 259–268.
- [59] D. Walnut *Nonperiodic sampling of bandlimited functions on unions of rectangular lattices*, J. Fourier Anal. Appl. **2** (1996), 435–452.
- [60] A. I. Zayed, *Advances in Shannon's Sampling Theory*, CRC Press, Boca Raton, Florida, 1993.

DEPARTMENT OF MATHEMATICS, OREGON STATE UNIVERSITY, CORVALLIS, OREGON, 97331  
*E-mail address:* `faridani@math.oregonstate.edu`

## Generalized Transforms of Radon Type and Their Applications

Peter Kuchment

**ABSTRACT.** These notes represent an extended version of the contents of a lecture delivered at the AMS Short Course “Radon Transform and Applications to Inverse Problems” in Atlanta in January 2005. They contain a brief description of properties of some generalized Radon transforms arising in inverse problems. Here by generalized Radon transforms we mean transforms that involve integrations over curved surfaces and/or weighted integrations. Such transformations arise in many areas, e.g. in Single Photon Emission Tomography (SPECT), Electrical Impedance Tomography (EIT) thermoacoustic Tomography (TAT), and other areas.

### 1. Introduction

The notes by E. T. Quinto in this volume have already introduced the reader to the properties of the Radon transform and its role in inverse problems, in particular in computerized tomography. In this text we show that in some applications one has to work with weighted (*attenuated*) transforms of Radon type, where the lines (planes) of integration are equipped with certain weights that need to be incorporated into the transform. On the other hand, there are also important applied problems, where the data provides the values of the integrals of an unknown quantity over a family of curved manifolds (e.g., spheres) rather than lines or planes. These manifolds of integrations might be equipped with some weights as well. Such transforms have been studied in rather general situation (e.g., [20, 22, 23, 34, 35, 40, 54, 55, 57, 58, 61, 63, 62, 123, 124, 127, 128, 129, 132, 133, 134, 139] and references therein), but a richer theory can be developed for more specific examples. As it often happens, transforms arising in applications, have a special structure that allows for a deep and beautiful analytic theory.

Although this does not exhaust all situations that fall under our topic, we will restrict ourselves to the following three (probably the most prominent) areas: Thermoacoustic Tomography (TAT), where integrations over spheres are involved, Single Photon Emission Computed Tomography (SPECT), where weighted transforms arise, and Electrical Impedance Tomography (EIT), where hyperbolic Radon transforms appear naturally.

In these notes we are unable to provide a comprehensive bibliography (which would take at least as much space as the whole notes). Apologies are extended to the authors whose work should have been, but was not mentioned explicitly.

## 2. Thermoacoustic tomography and the circular Radon transform

Tomographic methods of medical imaging, as well as of industrial non-destructive evaluation and geological prospecting are based on the following general procedure: one sends towards a non-transparent body some kind of a signal (acoustic or electromagnetic wave, X-ray, visual light photons, etc.) and measures the wave after it passes through the body. Then one tries to use the measured information to recover the internal structure of the object of study. The common feature of most traditional methods of tomography is that the same kinds of physical signals are sent and measured. Each of the methods has its own drawbacks. For instance, sometimes when imaging biological tissues, microwaves and optical imaging might provide good contrasts between different types of tissues, but are inferior in terms of resolution in comparison with ultrasound or X-rays. This, in particular, is responsible for the common low resolution of optical or electrical impedance tomography. On the other hand, ultrasound, while giving good resolution, often does not do a good job in terms of contrast. One of the recent trends is to combine different types of waves in a single imaging process. The best developed example is probably the thermoacoustic tomography (TAT or TCT) and its sibling photoacoustic tomography (PAT) (e.g., [79],[152]-[155]). In TAT, a short microwave pulse is sent through a biological object. At each internal location  $x$  certain energy  $H(x)$  is absorbed. It is known, that cancerous cells often absorb several times more microwave (or radio frequency) energy than the normal ones, which means that significant contrast is expected between the values of  $H(x)$  at tumorous and healthy locations. The absorbed energy causes a thermoelastic expansion, which in turn creates a pressure wave. This wave can be detected by ultrasound transducers placed at the edges of the object. Now the former weakness of ultrasound (low contrast) becomes an advantage. Indeed, in many cases (e.g., for mammography) one can assume the sound speed to be approximately constant. Hence, the sound waves detected by a transducer at any moment  $t$  of time are coming from points at a constant distance (depending on time  $t$  of travel and the sound speed) from its location. The strength of the signal coming from a location  $x$  reflects the energy absorption  $H(x)$ . Thus, one effectively measures the integrals of  $H(x)$  over all spheres centered at the transducers' locations. In other words, in order to reconstruct  $H$  (and thus find cancerous locations) one needs to invert a generalized Radon transform that provides the integrals of  $H$  over spheres centered at all available transducers' locations [79], [152]-[155]. This method amazingly combines advantages of two types of radiation, while avoiding their deficiencies.

This motivates the study of the following “circular” Radon transform<sup>1</sup>. Let  $f(x)$  be a continuous function on  $\mathbb{R}^n$ ,  $n \geq 2$ . We define its circular Radon transform as

$$Rf(p, r) = \int_{|y-p|=r} f(y) d\sigma(y),$$

---

<sup>1</sup>Numerous other reasons to study this transform are known, e.g. Radar and Sonar imaging, approximation theory, PDEs, potential theory, complex analysis, etc. [2, 96]. Although in dimensions higher than two one should probably use the word “spherical” rather than “circular,” we will use for simplicity the latter. This should not create any confusion.

where  $d\sigma(y)$  is the surface area on the sphere  $|y - p| = r$  centered at  $p \in \mathbb{R}^n$ .

The mapping from  $f$  to  $Rf$  is overdetermined, since the dimension of pairs  $(p, r)$  is  $n + 1$ , while the function  $f$  depends on  $n$  variables only. This suggests to restrict the centers  $p$  to a set (hypersurface)  $S \subset \mathbb{R}^n$ , while not imposing any restrictions on the radii<sup>2</sup>. This restricted transform will be denoted by  $R_S$ :

$$R_S f(p, r) = Rf(p, r)|_{p \in S}.$$

Among central problems that naturally arise are:

- **Uniqueness of reconstruction:** is the information collected for a given set  $S$  of centers sufficient for the unique determination of the function  $f$ ? In other words, is the operator  $R_S$  injective (on a specific function space)?
- **Inversion formulas** and algorithms for  $R_S$ .
- **Stability** of the reconstruction.
- Description of the **range** of the transform: what conditions must ideal data satisfy?

All these questions have been resolved for the classical Radon transform [69, 105, 107]. However, they are more complex and not too well understood for the circular Radon transform.

We will now provide a brief survey of the known results and approaches to the problems listed above.

**2.1. Injectivity.** Here one is interested in finding which sets  $S$  guarantee uniqueness of reconstruction of a function  $f$  from its transform  $R_S$ . We introduce the following

**DEFINITION 1.** *The circular transform  $R$  is said to be injective on a set  $S$  ( $S$  is a set of injectivity) if for any compactly supported continuous function  $f$  on  $\mathbb{R}^n$  the condition  $Rf(p, r) = 0$  for all  $r \geq 0$  and all  $p \in S$  implies  $f \equiv 0$ . In other words,  $S$  is a set of injectivity, if the mapping  $R_S$  is injective on  $C_c(\mathbb{R}^n)$ .*

The condition of compactness of support on  $f$  is essential in what follows. The situation is significantly different and much harder to study without compactness of support (or at least a sufficiently fast decay) [1, 2]. Fortunately, tomographic problems usually yield compactly supported functions.

One now arrives to the

**Problem:** Describe all sets of injectivity for the circular Radon transform  $R$  on  $C_c(\mathbb{R}^n)$ . In other words, we are looking for a description of those sets of positions of transducers that enable one to recover uniquely the energy deposition function.

This problem has been around in different guises for quite a while (e.g., [2, 40, 93, 94] and references therein). One of its most useful reformulations is the following: finding all possible nodal sets of oscillating infinite membranes. Namely, consider the initial value problem for the wave equation in  $\mathbb{R}^n$ :

$$(1) \quad u_{tt} - \Delta u = 0, \quad x \in \mathbb{R}^n, t \in \mathbb{R}$$

$$(2) \quad u|_{t=0} = 0, \quad u_t|_{t=0} = f.$$

Then

$$u(x, t) = \frac{1}{(n-2)!} \frac{\partial^{n-2}}{\partial t^{n-2}} \int_0^t r(t^2 - r^2)^{(n-3)/2} (Rf)(x, r) dr, \quad t \geq 0.$$

---

<sup>2</sup>The most popular in TAT geometries of these surfaces (curves)  $S$  of centers (transducers) are spheres, planes, and cylinders [152]-[154].

Hence, it is not hard to show [2] that the original problem is equivalent to the problem of recovering  $u_t(x, 0)$  from the value of  $u(x, t)$  on subsets of  $S \times (-\infty, \infty)$ .

**LEMMA 2.** [2] *A set  $S$  is a non-injectivity set for  $C_c(\mathbb{R}^n)$  if and only if there exists a non-zero compactly supported continuous function  $f$  such that the solution  $u(x, t)$  of the problem (1)-(2) vanishes for any  $x \in S$  and any  $t$ .*

In other words, injectivity sets are those for which the motion of the membrane over  $S$  gives complete information about the motion of the whole membrane. An analogous relation holds also for solutions of the heat equation [2].

So, what could the injectivity sets be? As it turns out, they are more common than the non-injectivity ones. So, one should better try to describe the “bad” (i.e., non-injectivity) sets, i.e. sets of transducers’ positions from which one cannot recover the energy absorption function.

A simple example of a non-injectivity surface is any hyperplane  $S$ . Indeed, if  $f$  is odd with respect to this plane, then clearly  $R_S f = 0$ , so one cannot recover  $f$  from the data. It is known that in this case the odd functions are the only ones “eliminated” by  $R_S$  [38, 76]. In particular, any line on the plane is a non-injectivity set. There are other options as well. Let us consider for any  $N \in \mathbb{N}$  the Coxeter system  $\Sigma_N$  of  $N$  lines  $L_0, \dots, L_{N-1}$  in the plane passing through the origin and forming equal angles  $\pi/N$ :  $L_k = \{te^{i\pi k/n} | -\infty < t < \infty\}$ . There exist non-zero compactly supported functions that are simultaneously odd with respect to all lines  $\Sigma_N$  (look at the Fourier series expansion with respect to the polar angle). So,  $\Sigma_N$  is a non-injectivity set as well. Any rigid motion  $\omega$  preserves non-injectivity property, so  $\omega\Sigma_N$  is also a non-injectivity set. It is not that obvious, but still not hard to prove that adding a finite set  $F$  does not change this property. The following remarkable theorem was conjectured by V. Lin and A. Pincus [93, 94] and proven by M. Agranovsky and E. Quinto [2]:

**THEOREM 3.** [2] *A set  $S \subset \mathbb{R}^2$  is an injectivity set for the circular Radon transform on  $C_c(\mathbb{R}^2)$ , if and only if it is not contained in any set of the form  $\omega(\Sigma_N) \cup F$ , where  $\omega$  is a rigid motion in the plane and  $F$  is a finite set.*

The beautiful proof of this theorem in [2] is based on microlocal analysis and geometric properties of zeros of harmonic polynomials<sup>3</sup>. There are, however, some comments about non-injectivity sets that can be made without heavy techniques being involved.

The first important observation concerning non-injectivity sets is that they must be algebraic (i.e., sets of zeros of non-zero polynomials). In fact, if  $R_S f = 0$  and  $f$  decays faster than any power, it is not hard to see that the following polynomials vanish on  $S$ :  $Q_k(x) = \int_{\mathbb{R}^n} \|x - \xi\|^{2k} f(\xi) d\xi$ . One might wonder whether they could all vanish identically and thus imply nothing about  $S$ . It is easy to prove that this cannot happen if the function decays exponentially.

Now applying Laplace operator one readily concludes that the lowest degree polynomial among  $Q_k$  is in fact harmonic:

**LEMMA 4.** *Let  $f \in C_c(\mathbb{R}^n)$ , and  $P = Q_{k_0}[f]$  be the minimal degree nontrivial polynomial among  $Q_k$ , then  $P$  is harmonic and vanishes on  $S$ .*

Thus, if  $R$  is not injective on  $S$ , then  $S$  is the zero set of a harmonic polynomial.

---

<sup>3</sup>Albeit some simpler approaches have started to appear, e.g. in [47, 5], there is still no alternative proof of this theorem available, except some partial solutions in [5].

**COROLLARY 5.** *Any set  $S \subset \mathbb{R}^n$  of uniqueness for the harmonic polynomials is a set of injectivity for the transform  $R$ . E.g., if  $U \subset \mathbb{R}^n$  is any bounded domain, then  $S = \partial U$  is an injectivity set of  $R$ .*

This corollary is already good enough for many practical applications. Indeed, in one of the common practical set-ups one places transducers around a sphere  $S$ . The corollary guarantees uniqueness of reconstruction. In fact, algebraicity implies that the data over any open piece of the sphere has as much information as the data collected over the whole sphere, and thus also guarantees uniqueness (albeit at the price of significantly reduced stability of reconstruction [96, 155]).

The conjecture that describes non-injectivity sets in higher dimensions (still for compactly supported functions) is:

**CONJECTURE 6.** [2] *A set  $S \subset \mathbb{R}^n$  is an injectivity set for the circular Radon transform on  $C_c(\mathbb{R}^n)$ , if and only if it is not contained in any set of the form  $\omega(\Sigma) \cup F$ , where  $\omega$  is a rigid motion of  $\mathbb{R}^n$ ,  $\Sigma$  is the zero set of a non-zero **homogeneous** harmonic polynomial, and  $F$  is an algebraic subset in  $\mathbb{R}^n$  of co-dimension at least 2.*

For  $n = 2$ , this boils down to Theorem 3.

The uniqueness problem remains unresolved in dimensions higher than 2, and even in dimension 2 it is not resolved for functions that are not compactly supported (albeit possibly very fast decaying). For instance, there is a belief that the statement of Theorem 3 should hold true for functions that decay fast, say super-exponentially. So far no-one has succeeded in proving this. In [1], a limited scope question was posed: does the claim of Corollary 5 concerning the boundaries of bounded domains hold true for functions from  $L^p(\mathbb{R}^n)$ ,  $p < \infty$ ? The answer, given by the following theorem, shows that the situation is non-trivial:

**THEOREM 7.** [1] *The boundary  $S$  of any bounded domain in  $\mathbb{R}^n$  is uniqueness set for  $f \in L^p(\mathbb{R}^n)$  if  $p \leq \frac{2n}{n-1}$ . This is not true when  $p > \frac{2n}{n-1}$ , where spheres provide counterexamples.*

A new approach based on the wave equation interpretation that we have mentioned above and which promises possible new advances in this problem, is introduced in [47] (see also its further development in [5]; an early indication of this technique can be found in [1]). It uses essentially only the finite speed of propagation and domain of dependence for the wave equation. It boils down to the following idea: one has a free infinite oscillating membrane, but a (non-injectivity) set  $S$  stays fixed (nodal). Thus, one can also adopt a point of view that the membrane is just fixed along  $S$ . In this interpretation, the waves have to bypass  $S$ , while on the other hand the membrane is free and the waves are free to go without any obstacle. This sometimes gives a contradiction between two times of arrival, which in turn eliminates some sets  $S$  as possible non-injectivity sets.

**2.2. Inversion.** When  $S$  is a uniqueness set (e.g., a sphere) one is interested in reconstruction formulas for recovery of  $f$  from its transform  $R_S f$ . There are very few examples when such a formula is known, e.g. when  $S$  is a plane. Although in this case, as we know, there is no uniqueness, functions that are even with respect to  $S$ , or functions that are supported on one side of  $S$  can be reconstructed (e.g., [7, 39, 54, 56, 107, 120]). For the most interesting for TAT case of  $S$  being a sphere, inversion algorithms using special functions expansions are known (see

[7, 39, 43, 54, 107, 108, 112, 113, 120] and references therein concerning all these inversion formulas). However, analytic formulas (e.g., of backprojection type similar to the ones for the standard Radon transform) had not been known until recently. In [47], such explicit formulas were derived for odd dimensions under the condition that the unknown function is supported inside of the sphere  $S$  (which is not a restriction for TAT). The 3D version of one of the results of [47] is presented below:

**THEOREM 8.** *Let  $f$  be a smooth function supported in the unit ball centered at the origin in  $\mathbb{R}^3$ . Then for any  $x$  in this ball, the following reconstruction formulas hold true:*

$$(3) \quad \begin{aligned} f(x) &= -\frac{1}{8\pi^2} \int_{|p|=1} \frac{1}{|x-p|} \frac{\partial^2 R_S}{\partial r^2}(p, |x-p|) dp \\ &= -\frac{1}{8\pi^2} \Delta_x \int_{|p|=1} \frac{1}{|x-p|} R_S(p, |x-p|) dp. \end{aligned}$$

Here the set of centers  $S$  is the unit sphere centered at the origin.

Notice that if the function is not supported inside the unit ball, the formula would give incorrect values even inside the ball.

Such formulas for 2D and higher even dimensions are still not known. However, it is easy to write approximate formulas (parametrices) either by using ideas of microlocal analysis in the spirit of [20, 85, 109] or just by mimicking the Radon case. Microlocal analysis of such formulas usually shows that they recover the singularities of the function correctly, albeit they do not reconstruct the values of  $f$  precisely. Simple iterative correction procedures significantly improve the quality of reconstruction and seem to provide reconstructions adequate for practical purposes (e.g., [152]-[155]).

One should also mention an important analytic tool, unfortunately not that well known in the applied community, the so called  $\kappa$ -operator developed in I. Gelfand's school (one can find its description for instance in [53, 54]). It provides a unified approach to inverting various Radon type transforms.

**2.3. Stability of reconstruction.** The microlocal analysis (i.e., in terms of wave front sets) approach, similar to the one used for singularity detection in the Radon transform (see E. T. Quinto's lecture in this volume), provides the general answer of what can and what cannot be stably reconstructed. Notice that uniqueness results by themselves do not guarantee stability. For instance, as we have mentioned before, a small portion of the sphere covered by transducers guarantees uniqueness of reconstruction. However, most of the sharp details will disappear, since their reconstruction is unstable. Namely, the following rule describes in general the situation. If at each point of the object to be reconstructed and for each line passing through this point there is a transducer located on this line, then reconstruction of the object can be made as stable as from the regular Radon transform data. However, if there is a line through a point that does not pass through any transducer's location, then any possible boundary between tissues at this point that is normal to the line, will be blurred in the reconstructed image. One can find further details for the case of the standard Radon (or X-ray) transform in [131], for attenuated transforms in [77, 78, 85], and for circular transforms in [96, 155].

**2.4. Range description.** Knowing the range of the transform  $R_S$  could be very useful. For instance, the range theorems for the Radon transform have been used to correct errors in measured data, to complete incomplete data, and for other purposes. The paper [121] contains a series of range conditions for the circular transform in the case of  $S$  being a sphere. As for the standard Radon transform, these conditions are not hard to derive. Indeed, let  $S$  be the unit sphere in  $\mathbb{R}^n$  centered at the origin and we know the function  $g(p, r) = \int_{|x-p|=r} f(x)dx$  for any  $p \in S$ . Then for any natural  $k$  we immediately conclude that the momentum

$$(4) \quad Q_k(p) = \int_0^\infty r^{2k} g(p, r) dr = \int_{\mathbb{R}^n} (|x|^2 - 2x \cdot p + |p|^2)^k f(x) dx,$$

viewed on the unit sphere, is the restriction of a (non-homogeneous) polynomial of degree at most  $k$  with respect to  $p$ . This gives us a series of necessary conditions. It has recently been discovered [6] that this set of conditions is incomplete; the missing conditions were obtained in the two-dimensional case. They can be expressed in terms of vanishing of some momenta with respect to Bessel functions [6].

**2.5. Implementation.** We finish this section with examples of reconstructions from synthesized, as well as real data. These results and figures are borrowed from [155].

Fig. 1 shows a mathematical phantom that was used for reconstructions shown in the next picture.

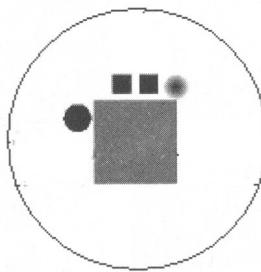


FIGURE 1. A mathematical phantom

Fig. 2 shows in different columns reconstructions of the phantom using different amounts of data. Namely, the detectors were placed correspondingly along an arc of approximately 90 degrees in the first quadrant, an arc containing two first quadrants, and finally a 360 degrees arc. The blurred parts of the boundaries are due to the limited view, which agrees with the microlocal analysis of the problem (see the discussion in the stability sub-section). Namely, a part of the boundary is blurred when its normals do not contain any detector locations. One can see how the existence of blurred parts depends on the detector arc. Different rows represent different reconstruction methods (see details in [155]).

In Fig. 3 one can see the photograph of a physical phantom (a piece of meat immersed into fat) and its reconstructions from the experimentally measured TAT data (measurements were performed in Prof. L. Wang's Optical Imaging Lab at

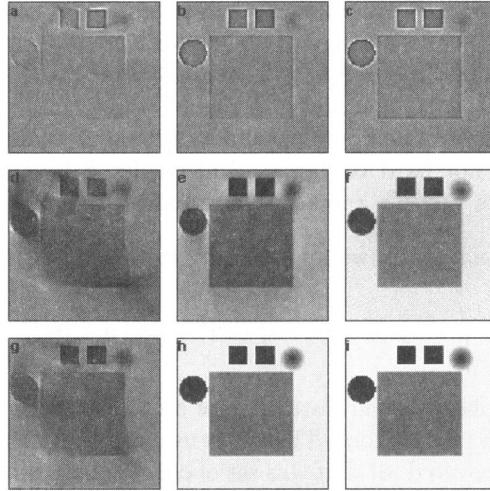


FIGURE 2. TAT reconstructions from the phantom data

Texas A& M University). We show TAT reconstructions that used limited data (left to right: detection arcs of 92 degrees, 202 degrees, and 360 degrees). The blurred parts of the boundaries again behave according to the theory.

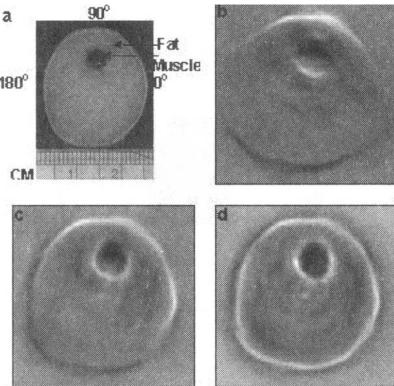


FIGURE 3. TAT reconstructions from experimental data

### 3. Emission tomography and attenuated Radon transform

Emission tomography deals with imaging of self-radiating bodies (as opposite to transmission imaging methods, where the source of radiation is outside of the object to be imaged). Let us describe briefly the main principle of the so called Single Photon Emission Computed Tomography (SPECT), a popular method of medical diagnostics (one can find more details in [29, 71, 105, 107])<sup>4</sup>. In SPECT, a patient is given a medication labelled by a radionuclide. The resulting emission is observed outside the body by collimated detectors that allow in only narrow beams of radiation (see Fig. 4). The goal is by measuring the intensity of the

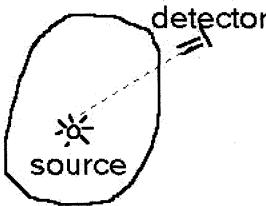


FIGURE 4. Single Photon Emission Computed Tomography

outgoing radiation to reconstruct the interior intensity distribution  $f(x)$  of the radiation sources<sup>5</sup>. Let  $\mu(x)$  be the linear attenuation coefficient (or just attenuation) of the body at the location  $x$ . Due to the absorption, a beam passing through the body, suffers losses. Assuming that effects of scatter are small, one can solve a simple transport equation to conclude (e.g., [105]) that the total detected intensity along a beam (straight line)  $L$  reaching the detector is

$$(5) \quad T_\mu f(L) = \int_L f(x) e^{-\int_{L_x} \mu(y) dy} dx.$$

Here  $L_x$  is the segment of the line  $L$  between the emission point  $x$  and the detector and  $dy$  denotes the standard linear measure on  $L$ . The operator  $T_\mu$  is said to be the **attenuated Radon transform** with attenuation  $\mu(x)$  of the function  $f(x)$ . One can make the formula above a little more specific by parametrizing the lines. Namely, let  $\omega$  be the unit vector normal to the direction of  $L$ ,  $\omega^\perp$  be its  $90^\circ$  degrees counterclockwise rotation, and  $s$  be the signed distance from the origin to  $L$ . Then the line  $L$  consists of points  $s\omega + t\omega^\perp$ ,  $t \in \mathbb{R}$ . Now

$$(6) \quad T_\mu f(\omega, s) = \int_{-\infty}^{\infty} f(s\omega + t\omega^\perp) e^{-\int_t^\infty \mu(s\omega + \tau\omega^\perp) d\tau} dt.$$

<sup>4</sup>We will consider here the 2D version, i.e. only the beams that belong to a specific plane will be taken into account. There has been a recent activity in 3D SPECT reconstructions that do not reduce to layer-by-layer 2D procedures, e.g. [89, 110, 151].

<sup>5</sup>This problem arises not only in medical imaging, but everywhere where one wants to reconstruct the interior of a self-radiating object, e.g. nuclear reactor, jet engine, etc. [125, 126].

In contrast with the standard Radon transform  $f \rightarrow \int_L f(x)dx$ , the resulting function depends on the orientation of the line  $L$ <sup>6</sup>.

It is often assumed that the attenuation  $\mu$  is constant inside the body and zero outside. If the body is convex and of a known shape, then it is easy to check (this was discovered first in [148]) that by a multiplication by a known function the attenuated transform can be reduced to a simpler one, called the **exponential Radon transform** of function  $f$ :

$$(7) \quad R_\mu f(\omega, s) = \int_{-\infty}^{\infty} f(s\omega + t\omega^\perp) e^{\mu t} dt$$

(here  $\mu$  is constant). For this integral to make sense, the function  $f(x)$  needs to have exponential decay at infinity sufficient to offset the effect of the exponential weight in the integral.

As before, the natural questions to ask about these two transforms are:

- **Injectivity:** Can a function of a natural class be reconstructed from its attenuated or exponential transforms?
- **Inversion formulae,** if injectivity is established.
- **Range.** Judging by the precedent of the standard Radon transform, these operators are unlikely to be surjective in any natural function spaces. So, what are the conditions the functions from the ranges of these operators must satisfy? As has been mentioned before, such knowledge is important for applications, as well as for understanding the analytic properties of these transforms.
- **Stability of inversion.**
- **Simultaneous reconstruction of the sources density  $f$  and attenuation  $\mu$ .** This is an unusual question, which does not arise for the standard Radon transforms. In most cases not only the value of the intensity distribution  $f(x)$ , but the attenuation coefficient  $\mu(x)$  as well is unknown. So, the question is whether it is possible to extract any information about both functions from the values of  $T_\mu f$  or  $R_\mu f$ ?

These problems will be briefly addressed below. In all cases we will describe first what is the situation with the simpler exponential transform, and then address the attenuated one. As it was mentioned, we will deal almost entirely with the  $2D$  case.

### 3.1. Uniqueness of reconstruction and inversion formulae.

3.1.1. *Exponential transform.* One of the useful properties of the exponential transform  $R_\mu$  is an analog of the projection-slice (also called Fourier-slice) theorem known for the Radon case. Namely, if  $f$  is compactly supported (or sufficiently fast exponentially decaying) function on the plane, a straightforward computation of Fourier transform leads to the formula

$$(8) \quad \widehat{R_\mu f}(\omega, \sigma) = \sqrt{2\pi} \widehat{f}(\sigma\omega + i\mu\omega^\perp).$$

Here the hat on the left is the  $1D$  Fourier transform with respect to  $s$ , while on the right it stands for the  $2D$  Fourier transform. I.e., projection data  $R_\mu f$  provides the values of the Fourier transform of the function  $f$  on the following surface in  $\mathbf{C}^2$ :

$$(9) \quad S_\mu = \{z = \sigma\omega + i\mu\omega^\perp \mid \sigma \in \mathbf{R}, \omega \in \mathbf{S}^1\} \subset \mathbf{C}^2.$$

---

<sup>6</sup>In practice one often averages over the two orientations, thus arriving to a function of non-oriented lines.

This is an indication that one can try to use methods of functions of several complex variables. Results of many studies (e.g., [3, 41, 46, 81, 82, 84, 98, 102, 117, 118]) show that the relation between two theories is indeed very deep. For instance, Paley-Wiener theorems that guarantee analyticity of  $\widehat{f}$ , together with the simple claim that the surface  $S_\mu$  is a uniqueness set for analytic functions, prove uniqueness of reconstruction (i.e., injectivity) for the exponential Radon transform.

The first explicit inversion formula for the exponential Radon transform in the plane was obtained in [149] (see also discussion in [105]). An inversion procedure was also provided in [15]. To describe the formula from [149], we introduce the dual exponential Radon transform (**exponential backprojection**)  $R_\mu^\#$ : applied to a function  $g(\omega, s)$ , it produces a function on the plane according to

$$(10) \quad (R_\mu^\# g)(x) = \int_{\mathbf{S}^1} e^{\mu x \cdot \omega^\perp} g(\omega, x \cdot \omega) d\omega.$$

Then a not very difficult calculation gives

$$(11) \quad R_{-\mu}^\# R_\mu f = \left( 2 \frac{\cosh \mu |x|}{|x|} \right) * f,$$

So, in order to reconstruct  $f$ , one needs to perform a de-convolution. Let

$$(12) \quad \zeta_\mu(\sigma) = \begin{cases} |\sigma| & \text{when } |\sigma| > |\mu| \\ 0 & \text{otherwise} \end{cases}$$

and  $I_\mu^{-1}$  (a generalized Riesz potential) be the Fourier multiplier by  $\zeta_\mu(\sigma)$ . Then the inversion formula from [149] reads as follows:

$$(13) \quad f = \frac{1}{4\pi} R_{-\mu}^\# I_\mu^{-1} R_\mu f.$$

There is, however, another way of looking at the inversion. Let us start with the standard Fourier inversion formula that involves integration of  $\widehat{f}$  over  $\mathbb{R}^2$ :

$$(14) \quad f(x) = (2\pi)^{-1/2} \int_{\mathbb{R}^2} \widehat{f}(\xi) e^{ix \cdot \xi} d\xi.$$

Rewriting it as

$$(15) \quad f(x) = (2\pi)^{-1/2} \int_{\mathbb{R}^2} \widehat{f}(z) e^{i(z_1 x_1 + z_2 x_2)} dz_1 \wedge dz_2,$$

one notices that the integration over  $\mathbb{R}^2$  is done of a holomorphic differential 2-form (we use here exponential decay of  $f$  and thus analiticity of  $\widehat{f}$ ). Since we know the values of  $\widehat{f}$  on  $S_\mu$ , the idea is to use Cauchy type argument to shift the integration from  $\mathbb{R}^2$  to  $S_\mu$ . This is not straightforward (since, in particular, the surface  $S_\mu$  has a hole in it and thus is not homological to  $\mathbb{R}^2$ ), but can be achieved [41, 46, 84, 98, 140, 141]. This lead in [84, 141] to a variety of inversion formulas. In particular, it was mentioned in [84] as a peculiar remark that one can invert an “obviously useless” (since such media apparently do not exist) exponential transform with the attenuation  $\mu$  depending on the direction vector  $\omega$ . It has turned out recently though, that such inversion formulas are important for some 3D SPECT scanning geometries [89, 151]. One should also notice recent results on exact inversion of the exponential transforms with “half-view” 180 degrees data

[110, 111, 135, 136]<sup>7</sup>. These can also be treated using the “useless” formula from [84].

3.1.2. *Attenuated transform.* Uniqueness of reconstruction and inversion problems are much harder for the full attenuated transform  $T_\mu$  (5) and have been resolved only very recently. Due to rather technical nature of these results, we will just try to give the reader a general idea of those, as well as main references.

The first results on uniqueness were the local ones. It was shown in [99] that if  $\mu \in C^2$ , then the transform (5) is injective on functions with a sufficiently small support. The idea is that when  $f$  is localized in a small neighborhood of a point, then the weight is almost constant on the support of the function, and thus the attenuated transform is very close to the usual Radon transform. Now injectivity follows just from simple operator perturbation argument (a bounded operator close to an injective semi-Fredholm one is injective). The next significant step was made in [44], where uniqueness was established under the condition that the diameter of the object was “not too large”. This result was sufficient for many practical situations, e.g. in medical applications it restricts the diameter of an object to 35.8cm. The proof was nontrivial and involved energy estimates. A breakthrough came in recent brilliant works [9, 114, 115], where uniqueness was proven under some mild smoothness condition on the attenuation and with no support size restrictions.

A similar uniqueness problem for attenuated  $X$ -ray transform in  $3D$  and higher dimensions happens to be trivial [44]. Indeed, let a compactly supported function be in the kernel of the transform. Taking into account only the rays that belong to a two-dimensional plane barely touching the support of the function, one deals with the small support  $2D$  situation and hence can conclude that the function must be zero. This allows one to “eat away” the whole support and to conclude that the function is in fact equal to zero.

The problem of uniqueness was also considered for transforms with more general positive weights  $w(x, \omega)$ :

$$(16) \quad \int f(y + t\omega)w(y + t\omega, \omega)dt$$

(e.g., [22, 127, 128, 130, 91]), and uniqueness results of different types were established. However, there exists a famous counterexample due to J. Boman [21] that shows that the condition of infinite smoothness of the weight function  $w$  alone does not guarantee uniqueness.

Let us now discuss **inversion**. An explicit inversion formula was found in ([114, 115]), while a less explicit procedure was discovered earlier in [9] (see also [24, 30, 45, 59, 87, 88, 106] for different derivations and implementations). Both approaches of [9, 114, 115] look at the more fundamental transport equation rather than the attenuated Radon transform itself, in order to obtain inversion formulas and procedures. We are not able to address here the details of these very interesting and illuminating techniques (see the recent surveys [24, 45]). Instead, we will just provide one of the incarnations of the inversion formula.

Let us denote by  $H$  the standard Hilbert transform and by  $R$  the standard Radon transform on the plane. Then the inversion formula of [114] can be written

---

<sup>7</sup>The reader should recall at this point that, unlike for the Radon transform, the exponential transform data are different at opposite locations.

as follows:

$$(17) \quad f(x) = -\frac{1}{4\pi} \operatorname{Re} \operatorname{div} \int_{S^1} \omega e^{(\mathcal{D}\mu)(x, \omega^\perp)} (e^{-h} H e^h T_\mu f)(\omega, x\omega) d\omega,$$

where  $h = \frac{1}{2}(Id + iH)R\mu$  and  $\mathcal{D}\mu$  is the so called divergent beam transform

$$\mathcal{D}\mu(\omega, y + t\omega) = \int_t^\infty \mu(y + \tau\omega) d\tau,$$

This formula was implemented numerically in [59, 87, 88, 106].

A short and illuminating derivation of the inversion formula has recently been obtained by A. Fokas [49, 50]

**3.2. Range conditions.** As before, we start with the simpler case of the exponential transform, which still leads to interesting analysis.

3.2.1. *Exponential transform.* The first appearance of the range conditions for  $R_\mu$  was in [15, 149]. Let  $f(x)$  be a smooth and compactly supported function on the plane and  $g(\omega, s, \mu)$  its exponential Radon transform with attenuation  $\mu$ . Representing  $\omega = (\cos \phi, \sin \phi)$  and expanding  $g(\omega, s, \mu)$  into the Fourier series with respect to  $\phi$ :

$$g(\omega, s, \mu) = \sum_l g_l(s, \mu) e^{il\phi},$$

one can establish necessary range conditions in terms of the Fourier transform  $\widehat{g}_l(\sigma, \mu)$  of  $g_l(s, \mu)$  with respect to  $s$ . It was observed in [15, 149] that the function

$$(\sigma + \mu)^l \widehat{g}_l(\sigma, \mu)$$

is even with respect to  $\sigma$  for any  $l \in \mathbf{Z}$ . It was shown in [82, 83] that this set of conditions is complete.

These range conditions do not have the usual momentum form. A complete momentum type set of conditions was also found in [82, 83]: if  $g(\omega, s) = R_\mu f$  for some  $f \in C_0^\infty(\mathbf{R}^2)$ , then the following identity is satisfied for any odd natural  $n$ :

$$(18) \quad \begin{aligned} & \sum_{k=0}^n \binom{n}{k} \frac{d}{d\phi} \circ \left( \frac{d}{d\phi} - i \right) \circ \dots \\ & \circ \left( \frac{d}{d\phi} - (k-1)i \right) \int_{-\infty}^\infty (\mu s)^{n-k} g(s, \omega) ds = 0. \end{aligned}$$

Here  $i$  is the imaginary unit,  $\omega = (\cos \phi, \sin \phi)$ , and  $\circ$  denotes composition of differential operators.

The condition (18) is not very intuitive and has been interpreted in several different ways [3, 4, 86, 117]. Even checking its necessity happens to be interesting, since a direct calculation shows that it is equivalent to the following series of identities for the usual sine function  $\sin \phi$ : for any odd natural  $n$

$$(19) \quad \begin{aligned} & \sum_{k=0}^n \binom{n}{k} \left( \frac{d}{d\phi} - \sin \phi \right) \circ \left( \frac{d}{d\phi} - \sin \phi + i \right) \circ \dots \\ & \circ \left( \frac{d}{d\phi} - \sin \phi + (k-1)i \right) \sin^{n-k} \phi = 0. \end{aligned}$$

The reader might want to try to establish these identities directly [82].

These conditions have also been studied in terms of complex analysis [3, 4, 117, 118]. It was shown in particular that they are essentially equivalent to some Bernstein-Hartogs' type theorems on extension of separately analytic functions [3, 4, 117, 118]. One of the amazing incarnations of the theorem is the following: let

$f$  be a function defined outside a disk in  $\mathbb{R}^2$  and such that its restriction to any tangent line to the disk extends to an entire function of one variable. Then function  $f$  extends from  $\mathbb{R}^2$  to an entire function on  $\mathbb{C}^2$  [3, 117, 118].

Range conditions for the exponential X-ray transforms in dimensions higher than two were obtained in [4]. A nice discussion can be found in [117, 118].

Let us mention briefly some applications of these range conditions. They have been used for detecting and correcting some data errors arising from hardware imperfection in SPECT [122] and for treatment of incomplete data problems [97]. Another interesting application is to radiation therapy planning, which deals with the operator dual to the exponential X-ray transform [36, 37]. Range conditions have proven to be useful in this area as well [36, 37, 80].

**3.2.2. Attenuated transform.** Even before the breakthrough in inverting the attenuated transform was achieved, an infinite (albeit still incomplete) set of range conditions was found [105]. We denote by  $H$  the Hilbert transform on the line:

$$Hp(x) = \frac{1}{\pi} v.p. \int_{-\infty}^{\infty} \frac{p(y)}{x-y} dy.$$

Here  $v.p.$  denotes the principal value of the integral. Let  $f$  and  $\mu$  belong to the Schwartz space  $\mathcal{S}(\mathbb{R}^2)$ . Then, for  $k > m \geq 0$  integers, we have

$$(20) \quad \int_{-\infty}^{\infty} \int_0^{2\pi} s^m e^{\pm ik\phi + 0.5(I \pm iH)R\mu(\omega, s)} T_{\mu} f(\omega, s) d\phi ds = 0,$$

where  $\omega = (\cos \phi, \sin \phi)$ ,  $I$  is the identity operator, and  $R\mu$  is the Radon transform of  $\mu$ .

These conditions have been used for the simultaneous recovery of the sources  $f(x)$  and attenuation  $\mu(x)$  (see [102]-[104] and discussion below).

The recent publication [116] contains a complete set of range conditions.

**3.3. Recovery of attenuation.** As we have discussed in the beginning, simultaneous recovery of the sources density  $f(x)$  and of the attenuation  $\mu(x)$  is an important applied issue. At the first glance, this problem looks hopeless: we are trying to recover two functions  $f(x)$  and  $\mu(x)$  of two variables with the data  $g = T_{\mu} f$  being a single function of two variables. This counting argument would be persuasive only if the operator  $T_{\mu}$  were close to a surjective one. However, we know that  $T_{\mu}$  has an infinite dimensional cokernel. Thus, when  $\mu$  changes, the range could in principle rotate so much that for different values of the attenuation the ranges would have zero (or a “very small”) intersection. If this were true, then both  $f$  and  $\mu$  would be recoverable or “almost recoverable”.

In the simplest case of the exponential X-ray transform, this problem was resolved in [73, 143, 144] (see also [8]). The range conditions were used to show that unless the function  $f(x)$  is radial, both  $f$  and  $\mu$  can be uniquely determined.

Recovery of a variable attenuation is definitely much more difficult. As in the exponential case, the range theorems are used to this end. The range conditions (20) have been used in order to do so [102, 103, 104]. Papers [26, 27, 9, 114, 115, 116] contain some additional indications on what techniques might be employed for that purpose. This issue, however, has not yet been resolved in a satisfactory way.

**3.4. Stability of reconstruction.** Reconstruction using attenuated or exponential Radon transform data is more unstable than in the usual Radon case, due

to the presence of exponentially growing factors in the direct transforms and back-projections (10). However, due to the infinite dimensionality of the co-kernel of the operator, one has a huge freedom in modifying inversion formulas. This freedom (in the exponential transform case) has been used to select the most stable inversion algorithms [64, 142]. This still needs to be done for the attenuated transform (see [88] for initial considerations).

**3.5. Other questions.** Here the author wants to briefly mention some other related developments.

Attenuated transforms with non-smooth attenuations were considered in [77, 78]. Such transforms arise naturally in medical imaging, since the attenuation coefficient has discontinuities along the tissue boundaries, which introduce artifacts into reconstruction. This effect was studied in the papers cited above.

Effects of non-perfect collimation of detectors were treated in [81].

Exponential Radon (rather than  $X$ -ray) transforms were studied in [140, 141].

An interesting “universal” transform that has no free parameters, but still incorporates all exponential  $X$ -ray transforms was introduced and studied in [41]. This transform has a lot of invariant structure built in. Its study in particular reveals relations between the F. John’s ultra-hyperbolic equation and boundary  $\bar{\partial}$ -operators.

#### 4. Electric impedance tomography and hyperbolic Radon transform

Electrical Impedance Tomography (EIT) is a promising and inexpensive method of medical diagnostics and of industrial nondestructive testing (e.g., [12, 13, 14, 25, 32, 33, 137]). Here one tries to recover the conductivity of the interior of an object (e.g., patient’s lungs and heart). The information about the electric conductivity is very important for medical diagnostics; it is also vital for some electrical procedures, such as defibrillation; it might also provide a cheap nondestructive evaluation technology. Here is the idea of EIT: one places electrodes around an object, creates through them various current patterns, and measures the corresponding boundary voltage drop responses (Fig. 5). Experimental and theoretical studies

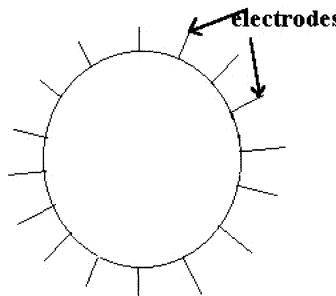


FIGURE 5. EIT

related to EIT are very active (see [25, 32, 33, 137] and references therein). Mathematically, the problem is much harder and less stable than the one of X-ray CT,

or MRI. In particular, in most approaches no Radon type transform arises. We will address here only one direction, which does involve a generalized Radon transform, and surprisingly enough, a non-Euclidean one!

Let us describe first the mathematical formulation of EIT, which is the so called inverse conductivity problem in  $2D$  (analogous formulations are available in higher dimensions as well). Let  $U \subset \mathbb{R}^2$  be a sufficiently smooth domain (say, a disk) with boundary  $\Gamma$ . The unknown conductivity function  $\beta(x)$  needs to be recovered from the following data. Given a known current function  $\psi$  on  $\Gamma$ , one measures the boundary value  $\phi$  of the potential. Mathematically speaking, one solves the Neumann boundary value problem

$$\begin{cases} \nabla \cdot (\beta \nabla u) = 0 & \text{in } U \\ \beta \frac{\partial u}{\partial \nu}|_{\Gamma} = \psi, \end{cases}$$

where  $\nu$  is the unit outer normal vector on  $\Gamma$  and  $\phi = u|_{\Gamma}$ . All the pairs  $(\psi, \phi)$  are assumed to be accessible. In other words, one knows the so called Dirichlet-to-Neumann operator  $\Lambda_{\beta} : \phi \rightarrow \psi$ . One needs to solve the nonlinear problem of recovery the conductivity  $\beta$  from this data. This happens to be a singularly hard inverse problem both analytically and numerically, not just (and not mainly) due to its nonlinearity, but mostly due to its severe instability. The main questions, as before are about uniqueness of reconstruction, its stability, and inversion procedures.

**4.1. Uniqueness.** After a long attempts and partial results, the uniqueness problem is essentially resolved positively (e.g., see [10, 28, 75, 101, 147, 150], and references therein), while the other questions (stability and reconstruction methods) are still under thorough investigation.

**4.2. Stability.** The general understanding is that the problem is highly unstable, so there is no hope to achieve the quality of reconstruction even close to the known for other common tomographic techniques. Indeed, as it will be in particular seen below, the problem is as unstable as the one of de-convolving a function with a Gaussian function. Saying this, we want to indicate that there are approaches that could possibly stabilize the problem. For instance, one could involve some additional available information about the image to be reconstructed, or one could try to reconstruct some useful functionals of the image rather than image's details, or finally one could try to change the physical set-up of EIT to improve stability of the reconstruction.

**4.3. Reconstruction algorithms and the hyperbolic integral geometry.** As we have already mentioned, the EIT problem (unlike the ones in X-ray, SPECT, PET, MRI, and TAT) is non-linear. Assuming that the unknown conductivity is a small variation of a constant, one can try to linearize the problem. This is exactly what the first practical algorithm of D. Barber and B. Brown [12]-[14] started with. Unfortunately, the linearized problem is still highly unstable. A study of this algorithm done in [138] lead in [18, 19] to the understanding that the linearized two-dimensional problem can be treated by means of hyperbolic geometry. Consider the  $2D$  unit disk  $U$ . We can view  $U$  as the Poincare model of the hyperbolic plane  $H^2$  (e.g., [16, 69]). There are some indications why the hyperbolic geometry might be relevant for the inverse conductivity problem. Indeed, if one creates a dipole current through a point on the boundary of  $U$ , then the

equipotential lines and the current lines form families of geodesics and horocycles in  $\mathbf{H}^2$  (geodesics are the circular arcs orthogonal to the boundary of the unit disc, while horocycles are the circles tangential to that boundary). Besides, the Laplace equation that arises in the linearized problem is invariant with respect to the group of Möbius transformations, which serve as motions of the hyperbolic plane. It was discovered in [18, 19] (following analysis of [138]) that the linearized inverse conductivity problem in  $U$  reduces to the following problem on  $\mathbf{H}^2$ : the available data enables one to find the function  $R_G(A * \beta)$ , where  $R_G$  is the geodesic Radon transform on  $\mathbf{H}^2$ ,  $A$  is an explicitly described radial function on  $\mathbf{H}^2$ :

$$A(r) = \text{const}(3 \cosh^{-4} r - \cosh^{-2} r),$$

and the star  $*$  denotes the (non-Euclidean) convolution on  $\mathbf{H}^2$ . Here the geodesic Radon transform integrates along geodesics in  $\mathbf{H}^2$  with respect to the measure induced by the Riemannian metric on  $\mathbf{H}^2$ . Methods of harmonic analysis (Fourier and Radon transforms and their inversions) on the hyperbolic plane are well developed (e.g., [17, 69, 70, 90, 92, 95]). One hopes to use them to invert the geodesic Radon transform, to de-convolve, and as the result recover  $\beta$ . In particular, the papers mentioned above contain explicit inversion formulas for the hyperbolic geodesic Radon transform. The formula obtained in [95] was numerically implemented in [51] and works as nicely and stably as the standard inversions of the regular Radon transform<sup>8</sup>.

As an illustration, we show in Fig. 6 below a numerical reconstruction from its geodesic Radon transform of a chessboard phantom.

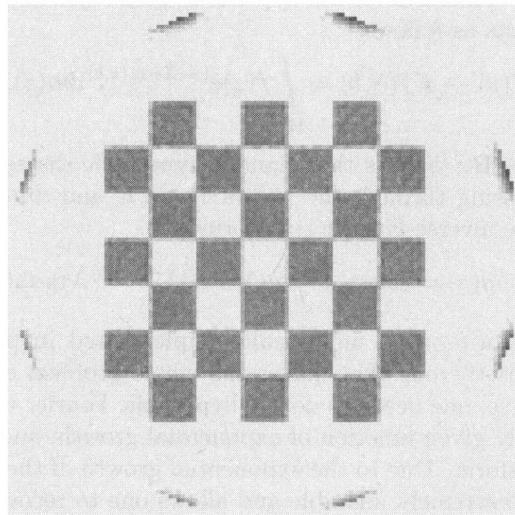


FIGURE 6. Hyperbolic reconstruction of a chessboard phantom.

The next Fig. 7 shows a similar reconstruction using a local tomography method ( $\Lambda$ -tomography, see the lectures by E. T. Quinto) that emphasizes boundaries.

---

<sup>8</sup>By an editorial error, all pictures have been omitted in [51]. They can be found at the URL <http://www.math.tamu.edu/~kuchment/hypnum.pdf>.

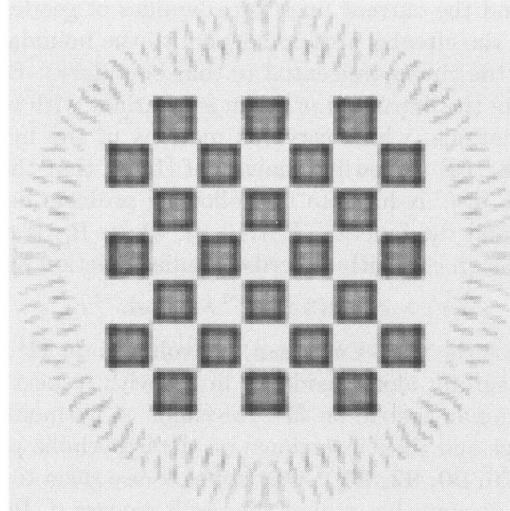


FIGURE 7. Local hyperbolic reconstruction of a chessboard phantom.

So, hyperbolic Radon transforms can be computed and inverted numerically. However, as it is discussed above, the next step of the linearized EIT inversion needs to be de-convolution. Its numerical implementation can be attempted by using the well studied Fourier transform on the hyperbolic plane and its inversion [69]. The Fourier transform acts as follows:

$$f(z) \rightarrow Ff(\lambda, b) = \int_{\mathbf{H}} f(z) e^{(-i\lambda+1)\langle z, b \rangle} dm(z),$$

where  $b \in \partial\mathbf{H}^2$ ,  $\lambda \in \mathbf{R}$ ,  $\langle z, b \rangle$  is the (signed) hyperbolic distance from the origin to the horocycle passing through the points  $z$  and  $b$ , and  $dm(z)$  is the invariant measure on  $\mathbf{H}^2$ . The inverse Fourier transform is

$$g(\lambda, b) \rightarrow F^{-1}g(z) = \text{const} \int \int g(\lambda, b) e^{(i\lambda+1)\langle z, b \rangle} \lambda \tanh(\pi\lambda/2) d\lambda db.$$

These Fourier transforms were numerically implemented in [51]. However, the deconvolution part is the one that makes the whole problem extremely unstable. Indeed, to de-convolve, one needs to do the hyperbolic Fourier transform, to multiply it by an explicitly given function of exponential growth, and then to apply the inverse Fourier transform. Due to the exponential growth of the Fourier multiplier, such a procedure is extremely unstable and allows one to recover stably only very low frequencies, and hence to get a strongly blurred image only. So, it should not be possible to get sharp resolution EIT, unless some radical additional information is incorporated (e.g., some a priori knowledge about the image), or the physical set-up of the technique is changed.

At the first glance, the relation of the linearized inverse conductivity problem with the hyperbolic integral geometry does not seem to work in dimensions higher than two, due to lack of hyperbolic invariance of the governing equations. It was a surprise then, when it was shown in [52] that a combination of Euclidean and hyperbolic integral geometries still does the trick.

**5. Acknowledgments**

This work was partly supported by the NSF Grants DMS 9971674 and 0002195. The author thanks the NSF for this support. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

The author also expresses his gratitude to the reviewer for the most helpful remarks.

## References

- [1] M. Agranovsky, C. A. Berenstein, and P. Kuchment, Approximation by spherical waves in  $L^p$  spaces, *J. Geom. Anal.*, **6**(1996), no. 3, 365–383.
- [2] M. L. Agranovsky, E. T. Quinto, Injectivity sets for the Radon transform over circles and complete systems of radial functions, *J. Funct. Anal.*, **139** (1996), 383–413.
- [3] V. Aguilar, L. Ehrenpreis, and P. Kuchment, Range conditions for the exponential Radon transform, *J. d'Analyse Mathematique*, **68**(1996), 1–13.
- [4] V. Aguilar and P. Kuchment, Range conditions for the multidimensional exponential X-ray transform, *Inverse Problems* **11**(1995), 977–982.
- [5] G. Ambartsoumian and P. Kuchment, On the injectivity of the circular Radon transform arising in thermoacoustic tomography, *Inverse Problems* **21** (2005), 473–485.
- [6] G. Ambartsoumian and P. Kuchment, A range description for the circular Radon transform on the plane, in preparation.
- [7] L.-E. Andersson, On the determination of a function from spherical averages, *SIAM J. Math. Anal.* **19** (1988), no. 1, 214–232.
- [8] Yu. E. Anikonov and I. Shneiberg, Radon transform with variable attenuation, *Doklady Akad. Nauk SSSR*, **316**(1991), no.1, 93–95. English translation in *Soviet. Math. Dokl.*
- [9] E.V. Arbuzov, A.L. Bukhgeim, and S.G. Kasantzev, Two-dimensional tomography problems and the theory of A-analytic functions, *Siberian Adv. in Math.* **8**(1998), no.4, 1–20.
- [10] K. Astala and L. Paivarinta, Calderon's inverse conductivity problem in the plane, ArXiv preprint math.AP/0401410, 2004.
- [11] G. Bal, On the attenuated Radon transform with full and partial measurements, *Inverse Problems* **20** (2004), 399–418.
- [12] D.C. Barber, B.H. Brown, Applied potential tomography, *J. Phys. E.: Sci. Instrum.* **17**(1984), 723–733.
- [13] D.C. Barber, B.H. Brown, Recent developments in applied potential tomography-APT, in: *Information Processing in Medical Imaging*, Nijhoff, Amsterdam, 1986, 106–121.
- [14] D.C. Barber, B.H. Brown, Progress in electrical impedance tomography, in *Inverse Problems in Partial Differential Equations*, SIAM, 1990, pp. 151–164.
- [15] S. Bellini, M. Piacentini, C. Caffario, and P. Rocca, Compensation of tissue absorption in emission tomography, *IEEE Trans. ASSP* **27**(1979), 213–218.
- [16] A. Berdon, *The Geometry of Discrete Groups*, Springer Verlag, New York- Heidelberg-Berlin, 1983.
- [17] C. Berenstein, E. Casadio Tarabusi, Inversion formulas for the k-dimensional Radon transform in real hyperbolic spaces, *Duke Math. J.* **62**(1991),no.3, 613–631.
- [18] C. Berenstein, E. Casadio Tarabusi, The inverse conductivity problem and the hyperbolic X-ray transform, pp. 39–44 in [57].
- [19] C. Berenstein, E. Casadio Tarabusi, Integral geometry in hyperbolic spaces and electrical impedance tomography, *SIAM J. Appl. Math.* **56**(1996), 755–764.
- [20] G. Beylkin, The inversion problem and applications of the generalized Radon transform, *Comm. Pure Appl. Math.* **37**(1984), 579–599.
- [21] J. Boman, An example of nonuniqueness for a generalized Radon transform. *J. Anal. Math.* **61**(1993), 395–401.
- [22] J. Boman and E. T. Quinto, Support theorems for real analytic Radon transforms, *Duke Math. J.* **55**(1987), no.4, 943–948.
- [23] J. Boman and E. T. Quinto, Support theorems for real analytic Radon transforms on line complexes, *Trans. Amer. Math. Soc* **335**(1993), 877–890.
- [24] J. Boman and J.-O. Strömberg, Novikov's inversion formula for the attenuated Radon transform – a new approach, *J. Geom. Anal.* **14** (2004), no. 2, 185–198.
- [25] L. Borcea, Electrical impedance tomography, *Inverse Problems* **18** (2002) R99–R136.
- [26] A. V. Bronnikov, Numerical solution of the identification problem for the attenuated Radon transform, *Inverse Problems*, **15** (1999), no. 5, 1315–1324.
- [27] A. V. Bronnikov and A. Kema, Reconstruction of attenuation map using discrete consistency conditions, *IEEE Trans. Med. Imaging*, **19** (2000), no. 5, 451–462.
- [28] R. Brown and G. Uhlmann, Uniqueness in the inverse conductivity problem for nonsmooth conductivities in two dimensions, *Comm. Part. Diff. Equat.* **22** (1997), no. 5–6, 1009–1027.

- [29] T.F. Budinger, G.T. Gullberg, and R.H. Huseman, Emission computed tomography, pp. 147–246 in [71].
- [30] A.A. Bukhgeim and S. G. Kazantsev, The attenuated Radon transform inversion formula for divergent beam geometry, Proc. IASTED Internat. Conf. “Automation, Control, and Information Technology”, June 10–13, Novosibirsk, Russia, 2002, 460–466.
- [31] A.L. Bukhgeim, Inversion formulas in inverse problems, a supplement to M. M. Lavrent’ev and L. Ya. Savel’ev, Linear Operators and Ill-posed Problems, Translated from the Russian, Consultants Bureau, New York; “Nauka”, Moscow, 1995.
- [32] M. Cheney, D. Isaacson, and J.C. Newell, Electrical Impedance Tomography, SIAM Review, 41, No. 1, (1999), 85–101.
- [33] B. Cipra, Shocking images from RPI, SIAM News, July 1994, 14–15.
- [34] A. Cormack, The Radon transform on a family of curves in the plane, Proc. Amer. Math. Soc. **83** (1981), no. 2, 325–330.
- [35] A. Cormack and E.T. Quinto, A Radon transform on spheres through the origin in  $\mathbb{R}^n$  and applications to the darboux equation, Trans. Amer. Math. Soc. **260** (1986), no. 2, 575–581.
- [36] A. Cormack and E.T. Quinto, A problem in radiotherapy: questions of non-negativity, Internat. J. Imaging Systems and Technology, 1(1989), 120–124.
- [37] A. Cormack and E.T. Quinto, The mathematics and physics of radiation dose planning, Contemporary Math. 113(1990), 41–55.
- [38] R. Courant and D. Hilbert, *Methods of Mathematical Physics, Volume II Partial Differential Equations*, Interscience, New York, 1962.
- [39] A. Denisiuk, Integral geometry on the family of semi-spheres. Fract. Calc. Appl. Anal. **2**(1999), no. 1, 31–46.
- [40] L. Ehrenpreis, *The Universality of the Radon Transform*, Oxford Univ. Press 2003.
- [41] L. Ehrenpreis, P. Kuchment, and A. Panchenko, The exponential X-ray transform and Fritz John’s equation. I. Range description, in *Analysis, Geometry, Number Theory: the Mathematics of Leon Ehrenpreis (Philadelphia, PA, 1998)*, 173–188, Contemporary Math., 251, Amer. Math. Soc., Providence, RI, 2000.
- [42] C. L. Epstein and. B. Kleiner, Spherical means in annular regions, Comm. Pure Appl. Math. **XLVI** (1993), 441–451.
- [43] J. A. Fawcett, Inversion of  $n$ -dimensional spherical averages, SIAM J. Appl. Math. **45**(1985), no. 2, 336–341.
- [44] D.V. Finch, Uniqueness for the X-ray transform in the physical range, Inverse Problems 2(1986), 197–203.
- [45] D.V. Finch, The attenuated X-ray transform: recent developments, in *Inside Out : Inverse Problems and Applications* , G. Uhlmann (Editor), 47–66, Cambridge Univ. Press 2003.
- [46] D.V. Finch and A. Hertle, The exponential Radon transform, Contemporary Math. 63(1987), 67–74.
- [47] D. Finch, Rakesh, and S. Patch, Determining a function from its mean values over a family of spheres, SIAM J. Math. Anal. **35** (2004), no. 5, 1213–1240.
- [48] L. Flatto, D. J. Newman, H. S. Shapiro, The level curves of harmonic functions, Trans. Amer. Math. Soc. **123** (1966), 425–436.
- [49] A. S. Fokas, Texas Notes, Lecture notes for the NSF CBMS Conference *New perspectives for boundary value problems and their asymptotics*, May 16-20, 2005 University of Texas - Pan American, Edinburg, Texas.
- [50] A. S. Fokas and L.-Y. Sung, Generalized Fourier transforms, their nonlinearization and the imaging of the brain, to appear in Notices of the AMS.
- [51] B. Fridman, P. Kuchment, K. Lancaster, S. Lissianoi, D. Ma, M. Mogilevsky, V. Papanicolaou, and I. Ponomarev, Numerical harmonic analysis on the hyperbolic plane. Appl. Anal. 76(2000), no. 3–4, 351–362.
- [52] B. Fridman, P. Kuchment, D. Ma, and V. Papanicolaou, Solution of the linearized conductivity problem in the half space via integral geometry, in *Voronezh Winter Mathematical Schools*, P. Kuchment (Editor), 85–95, Amer. Math. Soc. Transl. Ser. 2, 184, Amer. Math. Soc., Providence, RI, 1998.
- [53] I. Gelfand, S. Gindikin, and M. Graev, Integral geometry in affine and projective spaces, J. Sov. Math. 18(1980), 39–167.
- [54] I. Gelfand, S. Gindikin, and M. Graev, *Selected Topics in Integral Geometry*, Transl. Math. Monogr. v. 220, Amer. Math. Soc., Providence RI, 2003.

- [55] S. Gindikin (Editor), *Applied Problems of the Radon Transform*, AMS, Providence, RI, 1994.
- [56] S. Gindikin, Integral geometry on real quadrics, in *Lie groups and Lie algebras: E. B. Dynkin's Seminar*, 23–31, Amer. Math. Soc. Transl. Ser. 2, 169, Amer. Math. Soc., Providence, RI, 1995.
- [57] S. Gindikin and P. Michor (Editors), *75 Years of the Radon Transform*, Internat. Press 1994
- [58] A. Greenleaf and G. Uhlmann, Microlocal techniques in integral geometry, *Contemporary Math.* 113(1990), 149–155.
- [59] J.-P. Guillement, F. Jauberteau, L. Kunyansky, R. Novikov, and R. Trebessen, On SPECT imaging based on an exact formula for the nonuniform attenuation correction, *Inverse Problems*, 18 (2002) pp. L11–L19.
- [60] J.-P. Guillement and R. Novikov, A noise property analysis of single-photon emission computed tomography data, *Inverse Problems* 20 (2004), 175–198.
- [61] V. Guillemin, Fourier integral operators from the Radon transform point of view, *Proc. Symposia in Pure Math.*, 27(1975) 297–300.
- [62] V. Guillemin, On some results of Gelfand in integral geometry, *Proc. Symposia in Pure Math.*, 43(1985) 149–155.
- [63] V. Guillemin and S. Sternberg *Geometric Asymptotics*, Amer. Math. Soc., Providence, RI, 1977.
- [64] W. G. Hawkins, P. K. Leichner, and N. C. Yang, The circular harmonic transform for SPECT reconstruction and boundary conditions on the Fourier transform of the sinogram, *IEEE Trans. Med. Imag.* 7 (1988), 135–148.
- [65] I. Hazou and D. Solmon, Inversion of the exponential Radon transform I, *Analysis, Math. Methods in Appl. Sci.* 10(1988), 561–574.
- [66] I. Hazou and D. Solmon, Inversion of the exponential Radon transform II, *Numerics, Math. Methods Appl. Sci.* 13(1990), no. 3, 205–218.
- [67] I. Hazou and D. Solmon, Filtered-backprojection and the exponential Radon transform, *Math. Anal. Appl.* 141(1989), no.1, 109–119.
- [68] U. Heike, Single-photon emission computed tomography by inverting the attenuated Radon transform with least-squares collocation, *Inverse Problems* 2(1986), 307–330.
- [69] S. Helgason, *The Radon Transform*, Birkhäuser, Basel 1980.
- [70] S. Helgason, The totally-geodesic Radon transform on constant curvature spaces, *Contemporary Math.* 113(1990), 141–149.
- [71] G. Herman (Ed.), *Image Reconstruction from Projections*, Topics in Applied Physics, v. 32, Springer Verlag, Berlin, New York 1979.
- [72] A. Hertle, On the injectivity of the attenuated Radon transform, *Proc. Amer. Math. Soc.* 92(1984), 201–205.
- [73] A. Hertle, The identification problem for the constantly attenuated Radon transform, *Math. Z.* 197(1988), 13–19.
- [74] D. Isaacson and M. Cheney, Current problems in impedance imaging, in *Inverse Problems in Partial Differential Equations*, SIAM, 1990, pp. 141–149.
- [75] V. Isakov, *Inverse Problems for Partial Differential Equations*, Applied Mathematical Sciences, v. 127. Springer-Verlag, New York, 1998.
- [76] F. John, *Plane Waves and Spherical Means, Applied to Partial Differential Equations*, Dover 1971.
- [77] A. Katsevich, Local tomography for the generalized Radon transform, *SIAM J. Appl. Math.* 57(1997), 1128–1162.
- [78] A. Katsevich, Local tomography with nonsmooth attenuation, *Trans. Amer. Math. Soc.* 351(1999), no. 5, 1947–1974.
- [79] R. A. Kruger, P. Liu, Y. R. Fang, and C. R. Appledorn, Photoacoustic ultrasound (PAUS)–reconstruction tomography, *Med. Phys.* 22 (1995), 1605–1609.
- [80] P. Kuchment, On positivity problems for the Radon transform and some related transforms, *Contemporary Math.*, 140(1993), 87–95.
- [81] P. Kuchment, On inversion and range characterization of one transform arising in emission tomography, pp. 240–248 in [57].
- [82] P. Kuchment and S. Lvin, Paley-Wiener theorem for exponential Radon transform, *Acta Appl. Math.* 18(1990), 251–260
- [83] P. Kuchment and S. Lvin, The range of the exponential Radon transform, *Soviet Math. Dokl.* 42(1991), no.1, 183–184.

- [84] P. Kuchment and I. Shneiberg, Some inversion formulas in the single photon emission tomography, *Appl. Anal.* 53(1994), 221–231.
- [85] P. Kuchment, K. Lancaster, and L. Mogilevskaya, On local tomography, *Inverse Problems*, 11(1995), 571–589.
- [86] P. Kuchment and E. T. Quinto, Some problems of integral geometry arising in tomography, chapter XI in [40].
- [87] L. Kunyansky, A new SPECT reconstruction algorithm based on the Novikov's explicit inversion formula, *Inverse Problems* 17(2001), 293–306.
- [88] L. Kunyansky, Analytic reconstruction algorithms in emission tomography with variable attenuation, *J. of Computational Methods in Science and Engineering (JCMSE)*, 1 , Issue 2s-3s (2001), pp. 267–286.
- [89] L. Kunyansky, Inversion of the 3D exponential parallel-beam transform and the Radon transform with angle-dependent attenuation, *Inverse Problems* 20 (2004) 1455–1478.
- [90] A. Kurusa, The Radon transform on hyperbolic space, *Geometriae Dedicata* 40(1991), no.2, 325–339.
- [91] A. Kurusa, The invertibility of the Radon transform on abstract rotational manifolds of real type, *Math. Scand.* 70(1992), 112–126.
- [92] A. Kurusa, Support theorems for totally geodesic Radon transforms on constant curvature spaces, *Proc. AMS* 122(1994), no.2, 429–435
- [93] V. Ya. Lin and A. Pinkus, Fundamentality of ridge functions, *J. Approx. Theory*, 75 (1993), 295–311.
- [94] V. Ya. Lin and A. Pinkus, Approximation of multivariable functions, in *Advances in computational mathematics*, H. P. Dikshit and C. A. Micchelli, eds., World Sci. Publ., 1994, 1–9.
- [95] S. Lissianoi and I. Ponomarev, On the inversion of the geodesic Radon transform on the hyperbolic plane, *Inverse Problems* 13(1977), 1–10.
- [96] A. K. Louis and E. T. Quinto, Local tomographic methods in Sonar, in *Surveys on solution methods for inverse problems*, pp. 147–154, Springer, Vienna, 2000.
- [97] S. Lviv, Data correction and restoration in emission tomography, pp. 149–155 in [133].
- [98] A. Markoe, Fourier inversion of the attenuated X-ray transform, *SIAM J. Math. Anal.* 15(1984), no.4, 718–722.
- [99] A. Markoe and E. T. Quinto, An elementary proof of local invertibility for generalized and attenuated Radon transforms. *SIAM J. Math. Anal.* 16(1985), no. 5, 1114–1119.
- [100] C. Mennesier, F. Noo, R. Clackdoyle, G. Bal, and L. Desbat, Attenuation correction in SPECT using consistency conditions for the exponential ray transform, *Phys. Med. Biol.* 44 (1999), 2483–2510.
- [101] A. Nachman, Global uniqueness for a two-dimensional inverse boundary value problem, *Ann. Math.* 143(1996), 71–96.
- [102] F. Natterer, On the inversion of the attenuated Radon transform, *Numer. Math.* 32(1979), 431–438.
- [103] F. Natterer, Computerized tomography with unknown sources, *SIAM J. Appl. Math.* 43(1983), 1201–1212.
- [104] F. Natterer, Exploiting the range of Radon transform in tomography, in: Deuflhard P. and Hairer E. (Eds.), *Numerical treatment of inverse problems in differential and integral equations*, Birkhäuser Verlag, Basel 1983.
- [105] F. Natterer, *The mathematics of computerized tomography*, Wiley, New York, 1986.
- [106] F. Natterer, Inversion of the attenuated Radon transform, *Inverse Problems* 17(2001), no. 1, 113–119.
- [107] F. Natterer and F. Wübbeling, *Mathematical Methods in Image Reconstruction*, Monographs on Mathematical Modeling and Computation v. 5, SIAM, Philadelphia, PA 2001.
- [108] S. Nilsson, Application of fast backprojection techniques for some inverse problems of integral geometry, Linköping studies in science and technology, Dissertation 499, Dept. of Mathematics, Linköping university, Linköping, Sweden 1997.
- [109] C. J. Nolan and M. Cheney, Synthetic aperture inversion, *Inverse Problems* 18(2002), 221–235.
- [110] F. Noo, R. Clackdoyle, and J.-M. Wagner, Inversion of the 3D exponential X-ray transform for a half equatorial band and other semi-circular geometries, *Phys. Med. Biol.* 47 (2002), 2727–35.

- [111] F. Noo and J.-M. Wagner, Image reconstruction in 2D SPECT with 180° acquisition, *Inverse Problems*, 17(2001), 1357–1371.
- [112] S. J. Norton, Reconstruction of a two-dimensional reflecting medium over a circular domain: exact solution, *J. Acoust. Soc. Am.* **67** (1980), 1266–1273.
- [113] S. J. Norton and M. Linzer, Ultrasonic reflectivity imaging in three dimensions: exact inverse scattering solutions for plane, cylindrical, and spherical apertures, *IEEE Transactions on Biomedical Engineering*, 28(1981), 200–202.
- [114] R. Novikov, Une formule d'inversion pour la transformation d'un rayonnement  $X$  attenué, *C. R. Acad. Sci. Paris Ser. I Math* 332(2001), no. 12, 1059–1063.
- [115] R. Novikov, An inversion formula for the attenuated  $X$ -ray transformation, *Ark. Math.*, 40(2002), 145–167.
- [116] R. Novikov, On the range characterization for the two-dimensional attenuated  $X$ -ray transform, *Inverse Problems* 18(2002), 677–700.
- [117] O. Öktem, Comparing range characterizations of the exponential Radon transform, *Research reports in Math.*, no.17, 1996, Dept. Math., Stockholm Univ., Sweden.
- [118] O. Öktem, Extension of separately analytic functions and applications to range characterization of the exponential Radon transform, in *Complex Analysis and Applications (Warsaw, 1997)*, Ann. Polon. Math. 70(1998), 195–213.
- [119] V. P. Palamodov, An inversion method for an attenuated x-ray transform, *Inverse Problems* **12** (1996), 717–729.
- [120] V. P. Palamodov, Reconstruction from limited data of arc means, *J. Fourier Anal. Appl.* **6** (2000), no. 1, 25–42.
- [121] S. K. Patch, Thermoacoustic tomography - consistency conditions and the aptrial scan problem, *Phys. Med. Biol.* **49** (2004), 1–11.
- [122] I. Ponomarev, Correction of emission tomography data. Effects of detector displacement and non-constant sensitivity, *Inverse Problems*, 10(1995) 1–8.
- [123] D. A. Popov, The Generalized Radon Transform on the Plane, the Inverse Transform, and the Cavalieri Conditions, *Funct. Anal. and Its Appl.*, **35** (2001), no. 4, 270–283.
- [124] D. A. Popov, The Paley–Wiener Theorem for the Generalized Radon Transform on the Plane, *Funct. Anal. and Its Appl.*, **37** (2003), no. 3, 215–220.
- [125] N.G. Preobrazhensky and V.V. Pikalov, *Unstable Problems of Plasma Diagnostic*, Nauka, Novosibirsk, 1982.
- [126] N.G. Preobrazhensky and V.V. Pikalov, *Reconstructive Tomography in Gas Dynamics and Plasma Physics*, Nauka, Novosibirsk, 1987.
- [127] E. T. Quinto, The dependence of the generalized Radon transform on defining measures, *Trans. Amer. Math. Soc.* 257(1980), 331–346.
- [128] E.T. Quinto, The invertibility of rotation invariant Radon transforms, *J. Math. Anal. Appl.*, 91(1983), 510–522.
- [129] E. T. Quinto, Null spaces and ranges for the classical and spherical Radon transforms, *J. Math. Anal. Appl.* **90** (1982), no. 2, 408–420.
- [130] E. T. Quinto, Pompeiu transforms on geodesic spheres in real analytic manifolds, *Israel J. Math.* 84(1993), 353–363.
- [131] E. T. Quinto, Singularities of the X-ray transform and limited data tomography in  $\mathbf{R}^2$  and  $\mathbf{R}^3$ , *SIAM J. Math. Anal.* 24(1993), 1215–1225.
- [132] E. T. Quinto, Radon transforms on curves in the plane, in *Tomography, Impedance Imaging, and Integral Geometry*, 231–244, Lectures in Appl. Math., Vol. 30, Amer. Math. Soc. 1994.
- [133] E.T. Quinto, M. Cheney, and P. Kuchment (Editors), *Tomography, Impedance Imaging, and Integral Geometry*, Lectures in Appl. Math., vol. 30, AMS, Providence, RI 1994.
- [134] V. G. Romanov, Reconstructing functions from integrals over a family of curves, *Sib. Mat. Zh.* **7** (1967), 1206–1208.
- [135] H. Rullgård, An explicit inversion formula for the exponential Radon transform using data from 180 degrees, *Ark. Mat.* **42** (2004), 353–362.
- [136] H. Rullgård, Stability of the inverse problem for the attenuated Radon transform with 180 degrees data, *Inverse problems* **20** (2004), 781–797.
- [137] F. Santosa, Inverse problem holds key to safe, continuous imaging, *SIAM News*, July 1994, 1 and 16–18.
- [138] F. Santosa, M. Vogelius, A backprojection algorithm for electrical impedance imaging, *SIAM J. Appl. Math.*, 50(1990), 216–241.

- [139] V. A. Sharafutdinov, *Integral Geometry of Tensor Fields*, V.S.P. Intl Science 1994.
- [140] I. Shneiberg, Exponential Radon transform, Doklady Akad. Nauk SSSR, 320(1991), no.3, 567-571. English translation in Soviet. Math. Dokl.
- [141] I. Shneiberg, Exponential Radon transform, pp. 235–246 in [55].
- [142] I. Shneiberg, I. Ponomarev, V. Dmitrichenko, and S. Kalashnikov, On a new reconstruction algorithm in emission tomography, in [55], 247–255.
- [143] D. Solmon, Two inverse problems for the exponential Radon transform, in *Inverse Problems in Action*, (P.S. Sabatier, editor), 46–53, Springer Verlag, Berlin 1990.
- [144] D. Solmon, The identification problem for the exponential Radon transform, Math. Methods in the Applied Sciences, 18(1995), 687–695.
- [145] E. Somersalo, M. Cheney, D. Isaacson, E. Isaacson, Layer-stripping: a direct numerical method for impedance imaging, Inverse Problems, 7(1991), 899–926.
- [146] J. Sylvester, A Convergent Layer-Stripping Algorithm for the Radially Symmetric Impedance Tomography Problem, Comm. in Part. Diff. Equat., 17(1992), 1955–1994.
- [147] J. Sylvester and G. Uhlmann, The Dirichlet to Neumann map and its applications, in *Inverse Problems in Partial Differential Equations*, SIAM, 1990, pp. 101–139.
- [148] O.J. Tretiak and P. Delaney, The exponential convolution algorithm for emission computed axial tomography, Proc. Symp. Appl. Math., vol. 27, Amer. Math. Soc., Providence, R.I. 1982, pp. 25–33.
- [149] O.J. Tretiak and C. Metz, The exponential Radon transform, SIAM J.Appl.Math. 39(1980), 341–354.
- [150] G. Uhlmann, Inverse boundary value problems and applications, Asterisque 207(1992), 153–211
- [151] J.-M. Wagner, F. Noo, and R. Clackdoyle, Exact inversion of the exponential X-ray transform for rotating slant-hole (RSR) SPECT, Phys. Med. Biol. **47** (2002), 2713–26.
- [152] M. Xu and L.-H. V. Wang, Time-domain reconstruction for thermoacoustic tomography in a spherical geometry, IEEE Trans. Med. Imag. **21** (2002), 814–822.
- [153] Y. Xu, D. Feng, and L.-H. V. Wang, Exact frequency-domain reconstruction for thermoacoustic tomography: I. Planar geometry, IEEE Trans. Med. Imag. **21** (2002), 823–828.
- [154] Y. Xu, M. Xu, and L.-H. V. Wang, Exact frequency-domain reconstruction for thermoacoustic tomography: II. Cylindrical geometry, IEEE Trans. Med. Imag. **21** (2002), 829–833.
- [155] Y. Xu, L. Wang, G. Ambartsumian, and P. Kuchment, Reconstructions in limited view thermoacoustic tomography, Medical Physics 31(4) April 2004, 724–733.

MATHEMATICS DEPARTMENT, TEXAS A&M UNIVERSITY, COLLEGE STATION, TX 77845  
*E-mail address:* kuchment@math.tamu.edu

*This page intentionally left blank*

## Inverse Problems in Pipeline Inspection

Peter Massopust

**ABSTRACT.** One of the techniques to detect defects and anomalies in pipelines is based on the magnetization of the pipe wall. Anomalies and defects in the pipe cause changes in the induced magnetic field and these changes are measured by a finite set of sensors. From this set of measurements  $\{\mathbf{Y}\}$ , the geometry  $\Sigma$  of the defect needs to be inferred. Several factors, for instance, sensor sensitivity, probability of detection, and noise, can impede these measurements making the solution of the inverse problem  $\{\mathbf{Y}\} \rightarrow \Sigma$  more difficult. In this expository paper, we present a basic magnetic model used for pipeline inspection, set up the associated inverse problem, and show how B-spline and wavelet techniques can be successfully applied to overcome some of the aforementioned measurement impediments.

### 1. Introduction

One of the main tasks of the pipeline integrity industry is the detection and classification of defects in pipelines. There is a wide variety of defects but they can be put into essentially three major groups: (i) defects due to corrosion, (ii) defects generated by mechanical damage, and (iii) cracks created by stresses in the pipe wall. A general defect may belong to more than one group. This paper will concentrate on defects belonging only to group (i), i.e. only on corrosion defects, since the methodology for detecting and analyzing defects from the other two groups is different and involves additional techniques and models. Corrosion defects can occur in the interior or the exterior of a pipe, in both liquid and gas lines, and exhibit complex geometries. It is quintessential to grade the severity of corrosion defects by determining their size, in other words, the axial (length), circumferential (width) and radial (depth) extent. Deep defects<sup>1</sup> compromise the integrity of the pipe wall and can lead to catastrophic consequences.

There exist different tools and methods to detect defects, one of which is *in-line inspection*. A pipeline inspection device is inserted into the pipe, is moved by the medium across a large distance of pipe while collecting data, and is finally extracted. Once inside the pipe, there is virtually no communication between the

---

2000 *Mathematics Subject Classification*. Primary 34A55, 41A15, 42C40; Secondary 78A30.

*Key words and phrases*. Direct and Inverse Problem, Maxwell Equations, B-splines, Multi-wavelets, Wavelet-Based Denoising.

<sup>1</sup>In the following, when referring to *defect* we mean *corrosion defect*.

inspection device and the outside. (The pipe, which is made of steel, acts as a Faraday<sup>2</sup> Cage.) Inspection devices may employ ultrasonics, EMAT (electromagnetic acoustic transducer) technology, or the magnetic flux leakage (MFL) approach to detect defects. The MFL technique is the oldest and currently most applied method for the detection of defects through inline inspection. It is based on the properties of magnetic fields generated by ferromagnetic materials such as pipe steel. A *ferromagnet* is a metal with a crystalline structure that makes it easy to magnetize and behave like a magnet. A ferromagnet is made up of domains each of which has its own magnetization (sum of magnetic moments) generated by the spins of the electrons in this region. The distribution of these domains and their magnetic moments is random enough not to produce a noticeable net magnetization. If, however, a strong magnetic field  $\mathbf{H}$  is applied to a ferromagnet, the magnetic moments inside the domains tend to align in a common direction thus generating a magnetic field  $\mathbf{B}$ .<sup>3</sup> The strength of the induced field  $\mathbf{B}$  depends not only on  $\mathbf{H}$  but also on the material properties of the ferromagnet. One such property is *magnetic permeability*. The larger a ferromagnet's permeability the more powerful the induced magnetic field  $\mathbf{B}$ .

The relationship between  $\mathbf{H}$  and  $\mathbf{B}$  is in general expressible in terms of a non-linear multi-valued function, called *hysteresis function*. For a mathematical introduction to the underlying phenomenon of *hysteresis*, the reader is referred to [V].

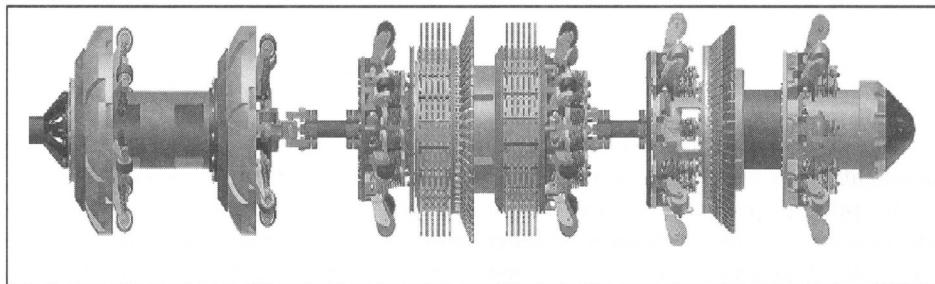


FIGURE 1. A pipeline inspection device.

A pipeline inspection device that employs the MFL technique is shown in Figure 1. This particular model is made up of three modules connected by U-joints. The first module is the drive section. There the batteries which power the onboard electronics are located. The center module consists of two strong ring magnets of opposite polarity one attached to each end. High permeability brushes surrounding the magnets are designed to bridge the gap between the magnets and the pipe wall and help to induce a more homogeneous magnetic field in the pipe. A ring of equally spaced groups of sensors, mounted between the two magnets, measures the magnetic field near the pipe. The third module contains the onboard electronics

<sup>2</sup>Michael Faraday, Sept 22, 1791 to Aug 25, 1867. British physicist and chemist who made major contributions to electromagnetism and electrochemistry. Considered to be have been one of the greatest scientists in history.

<sup>3</sup>This is a very simplified explanation of ferromagnetism and the interested reader is referred to [Bz, J, LL] for a more complete description of this phenomenon.

and in this model a ring of sensors to discriminate between interior and exterior defects.

The measurement of the magnetic field  $\mathbf{B}$  is done by thin semiconductive platelets, called *Hall*<sup>4</sup> sensors, making use of the *Hall effect*: An electric current  $I$  flowing through a (semi)conductor that moves perpendicular to a homogeneous magnetic field generates a voltage  $U_H$  that is transverse to  $I$  and perpendicular to  $\mathbf{B}$ . The situation is graphically displayed in Figure 2. If  $b$  is the width of the Hall

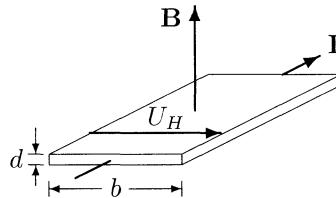


FIGURE 2. Schematics of the Hall effect.

sensor and  $d$  its thickness, then  $U_H$  is given by

$$(1.1) \quad U_H = R_H \frac{I \|\mathbf{B}\|}{d} =: c_H \|\mathbf{B}\|,$$

where  $R_H$  denotes the *Hall constant*, a characteristic of the (semi) conductor. Hence, Hall sensors measure the magnetic field through voltage changes. Depending on the orientation of the Hall sensor in the magnetic field, its radial, circumferential, or axial component may be measured. If two or more Hall sensors placed perpendicular to each other are used, more than one component can be measured. *For the purposes of this paper, we only consider measurements of the axial field component by one Hall device.* Due to the size of the measurement area of a Hall sensor, which is about 100 square micrometers, the measurements of the magnetic field component can be assumed to be point measurements. As the value of the magnetic field decreases with the square of the distance, the sensors are placed as closely as possible to the pipe wall. The magnetic field values are sampled uniformly along the axial direction, digitized and stored in a data array. Additionally, the sensors are equally spaced in the circumferential direction, thus producing a rectangular array  $\mathbf{Y}$  of bidirectionally uniform measurement samples of the magnetic field. The number of sensors on an inspection device runs from 8 to as many as 264, and each Hall sensor can make as many as  $2.5 \cdot 10^6$  measurements during an inspection run.

It must be pointed out that unlike in the case of the Radon<sup>5</sup> or X-ray transform, axially oriented Hall sensors measure only the magnitude of the magnetic field  $\mathbf{B}$  projected onto one plane perpendicular to the axial direction (which is the semiconductive platelet). Only an (infinite) plane bundle of sensors would give a complete representation of  $\mathbf{B}$ . Hence the situation here is much more restrictive

<sup>4</sup>Edward Herbert Hall, Nov 7, 1855 to Nov 20, 1938. American physicist who discovered the Hall effect.

<sup>5</sup>Johann Radon, Dec 16, 1887 to May 25, 1956. Austro-czech mathematician who is best known for his work in measure and integration theory and integral geometry.

than in the case of the Radon or X-ray transform and therefore the inversion or inverse problem much more complicated.

In the absence of any interior or exterior defects in the pipe, the response measured by the sensors is due to the approximately constant applied field  $\mathbf{H}$  (see Figure 3). (Noise and other measurement impediments do corrupt the data and cause deviations from constant.) The presence of a large enough defect however causes the induced magnetic field to leak out of the pipe. The shape and strength of this magnetic flux leakage (MFL) field is related to the size of the defect. Hall sensors will record this as a distinct deviation from constant. This response is called the *MFL signal*.<sup>6</sup> If the geometric width of a defect is large enough, then more than one Hall sensor will measure a non-zero voltage, i.e., a non-constant field, but with diminishing amplitude along the circumferential direction.

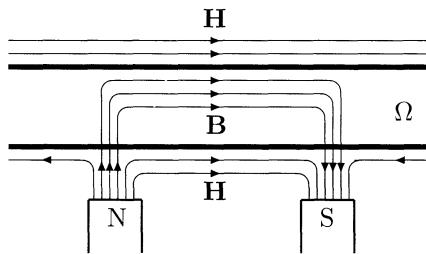


FIGURE 3. The magnetic fields around a magnetized pipe.

The problem of inferring the exact geometry of a defect in a hard ferromagnet from its MFL signal is challenging; within experimental constraints, different defect geometries may induce nearly identical MFL signals. The inverse problem *MFL signal  $\rightarrow$  geometry of defect* is still considered unsolved. In order to obtain a viable solution of this problem, it seems that a better understanding of the form and structure of an MFL signal needs to be achieved. As an MFL signal is, up to the proportionality constant  $c_H$ , equal to the magnetic field generated around a defect, it is natural to try to obtain exact analytical expressions for such fields based on rigorous mathematical models. The knowledge of the precise nature of the magnetic flux leakage field around a defect provides an exact relationship between the different parameters describing the geometry of a defect. This exact relationship needs to be used for the solution of the inverse problem.

The structure of this paper is as follows. In section 2, the basic equations governing magnetic fields in matter are presented and the *direct problem* is defined. A solution to the direct problem is derived for a simple model and its properties are discussed. A very brief review of the concept of an *inverse problem* is given in Section 3. Also in this section, the inverse problem associated with pipeline inspection is introduced and analyzed. Section 4 deals with signal impediments due to the uncontrollability of the measurement environment. Several types of such impediments are listed and discussed. The last section reviews some basic facts from multiwavelet theory before presenting the wavelet-based denoising algorithm developed by Donoho and Johnstone. This algorithm is then applied to remove

---

<sup>6</sup>Actually only *samples* are produced not the signal or function itself.

white Gaussian noise from MFL signals. In the Appendix some of the function spaces that appear in the main part of the text are defined for reference purposes.

## 2. Maxwell's Equations and Basic Magnetostatics

In this section the basic model for a hard ferromagnet is presented. Any second order effects such as the speed of the ferromagnet are ignored. The basic model is subject to modifications that describe the observed physical phenomena more closely but these modifications result in sets of solutions that are no longer expressible in simple closed analytical forms. For the purposes of this paper, however, the basic model contains all the important aspects of ferromagnetic theory.

Let  $\Omega \subset \mathbb{R}^3$  denote a hard ferromagnet<sup>7</sup> and  $\partial\Omega$  its piecewise smooth oriented surface. We assume that  $\Omega$  is a compact subset of  $\mathbb{R}^3$  and has nonempty simply-connected interior. Applying an exterior field  $\mathbf{H}$  to  $\Omega$  induces a *magnetization vector field*  $\mathbf{M} : \Omega \rightarrow \mathbb{R}^3$  in the interior  $\overset{\circ}{\Omega}$  of  $\Omega$ . It is advantageous to *assume* that  $\mathbf{M}$  vanishes on the boundary  $\partial\Omega$ :  $\mathbf{M}|_{\partial\Omega} \equiv 0$ . Note that this introduces a discontinuity of  $\mathbf{M}$  on the boundary, which is not observed in the physical reality but provides a convenient mathematical idealization. The vector fields  $\mathbf{H}$  and  $\mathbf{M}$  produce the *magnetic field*  $\mathbf{B}$  in  $\mathbb{R}^3$  according to the constitutive relationship<sup>8</sup>

$$(2.1) \quad \mathbf{B} = \mathbf{H} + 4\pi \mathbf{M}.$$

In addition, there exists also a constitutive relation between the vector fields  $\mathbf{H}$  and  $\mathbf{B}$  of the form

$$(2.2) \quad \mathbf{B} = \mathbf{f}(\mathbf{H}),$$

where  $\mathbf{f}$  is a multiple-valued function referred to as the *hysteresis function*. Figure 4 shows an example of a hysteresis function.

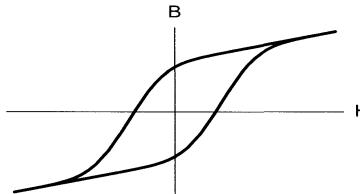


FIGURE 4. An hysteresis function.

In the neighborhood of any point  $\mathbf{H}_0$  where  $\mathbf{f}$  is single-valued, one may define the derivative

$$(2.3) \quad \mu(\mathbf{H}) := \frac{\partial \mathbf{B}}{\partial \mathbf{H}},$$

called the *differential magnetic permeability*. For an isotropic ferromagnet and for small values of  $\mathbf{H}$  and  $\mathbf{B}$ , there exists a *linear* relationship between the fields of the form

$$(2.4) \quad \mathbf{B} = \mu \mathbf{H}.$$

---

<sup>7</sup>A ferromagnet is called *hard* if its magnetization  $\mathbf{M}$  is essentially independent of any exterior applied field for moderate field strengths. One may treat such materials as having a fixed specific magnetization  $\mathbf{M}$ .

<sup>8</sup>In what follows, cgs units will be used.

The set of partial differential equations that describes the behavior of electric and magnetic fields in matter was derived by James Maxwell.<sup>9</sup> This set, called *Maxwell's Equations*, is based on theoretical and experimental results. The differential and integral forms of Maxwell's equations are as follows.<sup>10</sup>

$$(2.5) \quad \nabla \cdot (\epsilon \mathbf{E}) = 4\pi \varrho, \quad \oint_S \epsilon \mathbf{E} \cdot d\sigma = Q,$$

$$(2.6) \quad \nabla \cdot \mathbf{B} = 0, \quad \oint_S \mathbf{B} \cdot d\sigma = 0,$$

$$(2.7) \quad \nabla \times \mathbf{H} = \frac{4\pi}{c} \mathbf{J} + \frac{1}{c} \frac{\partial(\epsilon \mathbf{E})}{\partial t}, \quad \oint_{\gamma} \mathbf{H} \cdot ds = \frac{4\pi}{c} I + \frac{1}{c} \frac{\partial \Phi^e}{\partial t},$$

$$(2.8) \quad \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad \oint_{\gamma} \mathbf{E} \cdot ds = -\frac{1}{c} \frac{\partial \Phi^m}{\partial t}.$$

Here the vector function  $\mathbf{E} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  is called the *electric field* and  $\epsilon$  the *electric permittivity*. The constant  $c$  is the speed of light,  $\varrho : \mathbb{R}^3 \rightarrow \mathbb{R}$  the *electric charge density* and  $Q$  the total *electric charge* inside the closed surface  $S$ . The vector function  $\mathbf{J} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  denotes the *macroscopic current density* and  $I := \int_S \mathbf{J} \cdot d\sigma$  the (macroscopic) *current* encircled by the closed curve  $\gamma$  bounding the surface  $S$ . Given any closed orientable piecewise smooth curve  $\gamma$  bounding a surface  $S$ , the quantities  $\Phi^e, \Phi^m : \mathbb{R}^3 \rightarrow \mathbb{R}$  are called the *electric* and *magnetic flux*, respectively:

$$(2.9) \quad \Phi^e := \int_S \mathbf{E} \cdot d\sigma, \quad \Phi^m := \int_S \mathbf{B} \cdot d\sigma.$$

Here the orientation on  $S$  is induced by that of  $\gamma$ . (2.9) allows the interpretation of  $\mathbf{B}$  as a *magnetic flux density*, i.e., as the “number” of magnetic field lines enclosed by the curve  $\gamma$ , whereas  $\Phi^e$  can be interpreted as the “number” of electric field lines piercing the surface  $S$ . For similar interpretations of electrodynamic quantities and their modern differential-geometric interpretations in terms of differential forms, the reader is referred to [Sc, WSA].

(2.5) and (2.6) are referred to as Gauss's<sup>11</sup> law. In particular, (2.6) expresses the fact that there are no magnetic monopoles, i.e. that  $\mathbf{B}$  is an (axial) vector field or, more precisely, a 2-form:  $\mathbf{B} = B_x(\mathbf{x}) dy \wedge dz + B_y(\mathbf{x}) dz \wedge dx + B_z(\mathbf{x}) dx \wedge dy$ . The functions  $B_x, B_y, B_z : \mathbb{R}^3 \rightarrow \mathbb{R}$  are called the *x*, *y*, and *z-component of B*, respectively. (2.7) is commonly called *Ampère's*<sup>12</sup> law with Maxwell's extension and (2.8) *Faraday's Law of Induction*.

For the purposes of modelling the magnetic phenomena occurring in pipeline inspection, it is assumed that there are no electric fields present and that the

<sup>9</sup>James Clerk Maxwell, June 13, 1831 to Nov 5, 1879. Scottish physicist who completed the theory of electricity and magnetism to electrodynamics. He also worked in statistical mechanics.

<sup>10</sup>In the cgs system, the integration of the electric field over the surface of a sphere produces the factor  $4\pi$  (times the electric charge enclosed). This explains the ubiquitous occurrence of  $4\pi$ .

<sup>11</sup>Johann Carl Friedrich Gauss, April 30, 1777 to Feb 23, 1855. German mathematician, physicist and astronomer with major contributions to algebra, number theory, real and complex analysis, differential geometry, probability theory and statistics. Considered to be one of the greatest mathematicians of all time.

<sup>12</sup>André-Marie Ampère, Jan 22, 1775 to July 10, 1836. French mathematician and physicist who is generally considered as one of the discoverers of electromagnetism and the inventor of electric telegraphy. The measurement unit of electric current is named after Ampère.

magnetic field  $\mathbf{B}$  is static, i.e.,  $\partial\mathbf{B}/\partial t \equiv 0$ . Hence the set of Maxwell's equations reduces to

$$\begin{aligned} \nabla \cdot \mathbf{B} &= 0, & \oint_S \mathbf{B} \cdot d\sigma &= 0, \\ \nabla \times \mathbf{H} &= \frac{4\pi}{c} \mathbf{J}, & \oint_\gamma \mathbf{H} \cdot ds &= \frac{4\pi}{c} I. \end{aligned}$$

In the case when  $\mathbf{J}$  vanishes,  $\nabla \times \mathbf{H} = \mathbf{0}$  and thus there exists locally a continuously differentiable scalar function  $\varphi : \mathbb{R}^3 \rightarrow \mathbb{R}$ , called the *magnetic scalar potential*, so that

$$(2.10) \quad \mathbf{H} = -\nabla\varphi.$$

Using (2.10) together with (2.1) transforms  $\nabla \cdot \mathbf{B} = 0$  into the *Poisson*<sup>13</sup> equation

$$(2.11) \quad \Delta\varphi = 4\pi \nabla \cdot \mathbf{M}.$$

Here  $\Delta := \nabla \cdot \nabla$  denotes the Laplace<sup>14</sup> operator. The right-hand side of (2.11) can be thought of as a magnetic charge density  $\rho : \mathbb{R}^3 \rightarrow \mathbb{R}$ ,  $\mathbf{x} \mapsto -(\nabla \cdot \mathbf{M})(\mathbf{x})$ . Now choose a point  $p \in \partial\Omega$  and apply the Divergence Theorem to a cylinder of radius  $\varepsilon > 0$  and height  $2\varepsilon$  centered at  $p$  (see Figure 5) making use of the assumption that

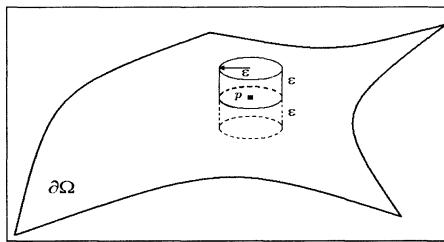


FIGURE 5

$\mathbf{M}$  vanishes on  $\partial\Omega$  to obtain that the quantity  $\mathbf{M} \cdot \mathbf{n}$  equals

$$\mathbf{M} \cdot \mathbf{n} = \frac{1}{4\pi} (\nabla\varphi) \Big|_{\partial\Omega} \cdot \mathbf{n}$$

and can therefore be thought of as a *magnetic surface charge density*. (Here  $\mathbf{n}$  denotes the unit (outward) normal to  $\partial\Omega$ .) A solution to Eqn. (2.11) is then given by (see for instance [CH])<sup>15</sup>

$$(2.12) \quad \boxed{\varphi(\mathbf{x}) = - \int_{\Omega} \frac{\nabla\xi \cdot \mathbf{M}(\xi)}{\|\mathbf{x} - \xi\|} dv + \oint_{\partial\Omega} \frac{\mathbf{M}(\xi) \cdot d\sigma(\xi)}{\|\mathbf{x} - \xi\|}}$$

<sup>13</sup>Siméon Denis Poisson, June 21, 1781 to April 25, 1840. French mathematician and physicist who made major contributions to the theory of heat and elasticity, acoustics, electricity, differential geometry, infinitesimal calculus, and probability theory.

<sup>14</sup>Pierre-Simon Laplace, March 28, 1749 to March 5, 1827. French mathematician and astronomer with major contributions to probability theory and differential equations.

<sup>15</sup>Here one needs to assume that  $\varphi \in C^2(\bar{\Omega}) \cap C(\Omega)$ .

(Here  $\nabla_{\xi}$  refers to the gradient with respect to the variable  $\xi \in \Omega$ .) If we assume a linear relationship of the form (2.4), then one can compute the magnetic field  $\mathbf{B}$  from the above equation via

$$\mathbf{B} = -\nabla(\mu \varphi) = -\mu \nabla \varphi,$$

Assuming that the interchange of  $\nabla$  and  $\int$  is allowed, one obtains

$$(2.13) \quad \mathbf{B}(\mathbf{x}) = -\mu \int_{\Omega} \frac{(\mathbf{x} - \boldsymbol{\xi}) \nabla_{\boldsymbol{\xi}} \cdot \mathbf{M}(\boldsymbol{\xi})}{\|\mathbf{x} - \boldsymbol{\xi}\|^3} dv + \mu \oint_{\partial\Omega} \frac{(\mathbf{x} - \boldsymbol{\xi}) \mathbf{M}(\boldsymbol{\xi}) \cdot d\sigma(\boldsymbol{\xi})}{\|\mathbf{x} - \boldsymbol{\xi}\|^3}$$

From now on we specialize to the situation where  $\Omega$  is the region between two concentric cylinders given explicitly by

$$\Omega := A(r_i, r_e) \times [0, L], \quad 0 < r_i < r_e, \quad L \gg r_e,$$

where  $A(r_i, r_e) := \{(\xi, \eta) \mid r_i^2 \leq \xi^2 + \eta^2 \leq r_e^2\}$ . The quantities  $r_i$  and  $r_e$  are called the *interior* and *exterior radius*, respectively, and the difference  $r_e - r_i$  the *wall thickness of  $\Omega$* . (Clearly, this is the situation encountered in pipeline inspection. The ferromagnet  $\Omega$  is the cylindrically shaped pipe.) The field  $\mathbf{B}$  generated by such a ferromagnet resembles the field of a long bar magnet with most of the field lines remaining inside  $\Omega$ , leaving the ferromagnet at  $A(r_i, r_e) \times \{0\}$  and reentering it at  $A(r_i, r_e) \times \{L\}$ . For the above-described idealized situation of magnetizing  $\Omega$ , the magnetization field  $\mathbf{M}$  can be considered to be *uniform* and *constant*:  $\mathbf{M} = M_0 \mathbf{e}_z$ ,  $M_0 > 0$ ,  $\mathbf{e}_z$  unit vector in the  $z$ -direction. In this case, (2.13) simplifies to

$$(2.14) \quad \begin{aligned} \mathbf{B}(\mathbf{x}) &= \mu M_0 \oint_{\partial\Omega} \frac{(\mathbf{x} - \boldsymbol{\xi}) \mathbf{e}_z \cdot d\sigma(\boldsymbol{\xi})}{\|\mathbf{x} - \boldsymbol{\xi}\|^3} \\ &= \mu M_0 \left[ \int_{\Omega(r_i, r_e, L)} \frac{(\mathbf{x} - \boldsymbol{\xi})}{\|\mathbf{x} - \boldsymbol{\xi}\|^3} \Big|_{\zeta=L} d\xi d\eta \right. \\ &\quad \left. - \int_{\Omega(r_i, r_e, 0)} \frac{(\mathbf{x} - \boldsymbol{\xi})}{\|\mathbf{x} - \boldsymbol{\xi}\|^3} \Big|_{\zeta=0} d\xi d\eta \right] \end{aligned}$$

with  $\Omega(r_i, r_e, 0) := A(r_i, r_e) \times \{0\}$ ,  $\Omega(r_i, r_e, L) := A(r_i, r_e) \times \{L\}$ , and  $\boldsymbol{\xi} := (\xi, \eta, \zeta) \in \mathbb{R}^3$ .

The magnetic field  $\mathbf{B}$  tends to remain inside a hard ferromagnet since this is energetically more advantageous. If, however, there is defect  $\Sigma$  in  $\Omega$  then the field lines, even in the current idealization, will leave the ferromagnet. This phenomenon will now be described next.

To this end, a *defect*  $\Sigma$  is defined as missing mass that reduces the wall thickness of  $\Omega$ . A defect can occur on either the interior or exterior side of  $\Omega$ .  $\Sigma$  may be thought of as being contained in a cylindrical volume element of radial extent  $\Delta\rho$ , angular subtense  $\Delta\phi$ , and axial distance  $\Delta z$ . These three quantities are referred to as the *depth*, *width* and *length* of the defect  $\Sigma$ , respectively. If the ratio of the length of  $\Sigma$  is small compared to the inner or outer radius of  $\Omega$  (the vast majority of real world defects satisfies this criterion), then it is convenient to think of  $\Sigma$  as enclosed by a cartesian box of length  $\Delta z$ , width  $\Delta y$ , and depth  $\Delta x$ . In the following, we abuse notation and denote by  $\Sigma$  also the surface of a defect, which is assumed to be piecewise smooth and orientable.

Now if the dimensions of the defect  $\Sigma$  are big enough and the applied field  $\mathbf{H}$  is sufficiently large, then the magnetic field lines of  $\mathbf{B}$  will leave the interior of  $\Omega$ ,

penetrate its exterior, and re-enter  $\Omega$ . This process is called *magnetic flux leakage*. This situation is shown in Figure 6. As defects are local phenomena in regards to

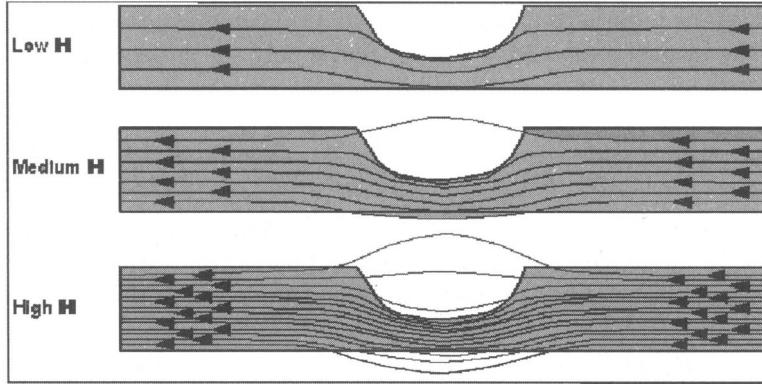


FIGURE 6. The principle of magnetic flux leakage.

the overall geometry, focus is entirely given to the magnetic field  $\mathbf{B}$  in the vicinity of  $\Sigma$ . This field,  $\mathbf{B}_\Sigma$ , is called the *magnetic flux leakage field* and its shape can be explained using *Ampère's Dipole Model*. One assumes that  $\mathbf{B}_\Sigma$  is generated by magnetic dipoles situated on opposite sides of the defect surface  $\Sigma$ . In other words, the distribution of magnetic dipoles on the surface  $\Sigma$ , regarded as a Radon measure on  $\Sigma$ , generates  $\mathbf{B}_\Sigma$ . An explicit expression for  $\mathbf{B}_\Sigma$  is given by (2.13) with only the slight modification of replacing  $\partial\Omega$  by  $\Sigma$ .

$$(2.15) \quad \mathbf{B}_\Sigma(\mathbf{x}) = - \int_{\Omega} \frac{(\mathbf{x} - \boldsymbol{\xi})(\nabla_{\boldsymbol{\xi}} \cdot \mathbf{M})(\boldsymbol{\xi})}{\|\mathbf{x} - \boldsymbol{\xi}\|^3} dv + \oint_{\Sigma} \frac{(\mathbf{x} - \boldsymbol{\xi})\mathbf{M}(\boldsymbol{\xi}) \cdot d\sigma(\boldsymbol{\xi})}{\|\mathbf{x} - \boldsymbol{\xi}\|^3}$$

Here we also used the fact that the magnetic permeability  $\mu$  is approximately equal to one outside the ferromagnet  $\Omega$ . The magnetic dipole distribution is given by the Radon measure  $\mu_\Sigma := \mathbf{M}(\boldsymbol{\xi}) \cdot d\sigma_\Sigma$  whose support is  $\Sigma$ . The distribution of the magnetic dipoles, i.e.,  $\mu_\Sigma$ , is determined by the magnetization  $\mathbf{M}$  in a neighborhood of  $\Sigma$  in  $\Omega$  and the specific material properties of the ferromagnet  $\Omega$ . Precise knowledge of these two properties together with exact information about the geometry of the defect  $\Sigma$  would uniquely determine the Radon measure  $\mu_\Sigma$ . On the other hand,  $\mu_\Sigma$  does determine the geometry of  $\Sigma$  once  $\mathbf{M}$  and the ferromagnet's material properties are known. In this case, one has the one-to-one correspondence  $\Sigma \longleftrightarrow \mu_\Sigma$ . Unfortunately,  $\mathbf{M}$  and the material properties of  $\Omega$  are only approximately, if at all, known.

**EXAMPLE 2.1.** To illustrate some of the issues, we next compute the magnetic flux leakage field  $\mathbf{B}_\Sigma$  for a simple idealized defect  $\Sigma$ . This is an example of a *forward* or *direct problem* in whose inversion we are ultimately interested. For this purpose, suppose the interior and exterior radius is large compared to the dimensions of the defect. In this case, we will work in cartesian coordinates  $(x, y, z)$ . Suppose then that  $\Sigma = [-d, d] \times [-w, w] \times [-\ell, \ell]$ , i.e.,  $\Sigma$  is an oriented parallelepiped<sup>16</sup> of

<sup>16</sup>This seemingly artificial defect can be used to give an exhaustion of a geometrically more complex  $\Sigma$  by a finite number of small parallelepipeds. The additivity property of the integral and the Radon measure  $\mu_\Sigma$  allows an approximation of the magnetic flux leakage field due to

length  $2\ell$ , width  $2w$ , and depth  $2d$  with the coordinate system chosen in such a way that the center of the box coincides with the origin. Moreover, assume that the pipe length  $L \gg \text{diam } \Sigma$  (This essentially means that the length of the pipe may be taken as infinite resulting in the neglect of any magnetic far field effects.) and that the magnetization  $\mathbf{M} = M_0 \mathbf{e}_z$ . Hence only the surface integral in (2.12) contributes to  $\varphi$ .

$$\varphi(\mathbf{x}) = \oint_{\Sigma} \frac{M_0 \mathbf{e}_z \cdot d\sigma(\xi)}{\|\mathbf{x} - \xi\|} = M_0 \left( \int_{-d}^d \int_{-w}^w \frac{d\xi d\eta}{r_-} - \int_{-d}^d \int_{-w}^w \frac{d\xi d\eta}{r_+} \right),$$

where  $r_{\pm} := 1/\sqrt{(x - \xi)^2 + (y - \eta)^2 + (z \mp \ell)^2}$ . These functions satisfy all the hypotheses for interchangeability of integral and partial derivative (see for instance [W], p. 59) and therefore (2.14) yields for the components of the magnetic field  $\mathbf{B}$

$$(2.16) \quad B_x(\mathbf{x}) = -\frac{\partial \varphi}{\partial x} = -M_0 \int_{-d}^d \int_{-w}^w \left( \frac{\partial r_-}{\partial x} - \frac{\partial r_+}{\partial x} \right) d\xi d\eta,$$

$$(2.17) \quad B_y(\mathbf{x}) = -\frac{\partial \varphi}{\partial y} = -M_0 \int_{-d}^d \int_{-w}^w \left( \frac{\partial r_-}{\partial y} - \frac{\partial r_+}{\partial y} \right) d\xi d\eta,$$

and

$$(2.18) \quad B_z(\mathbf{x}) = -\frac{\partial \varphi}{\partial z} = -M_0 \int_{-d}^d \int_{-w}^w \left( \frac{\partial r_-}{\partial z} - \frac{\partial r_+}{\partial z} \right) d\xi d\eta,$$

respectively.

Performing the partial differentiation in the expressions for  $B_x$ ,  $B_y$ , and  $B_z$  and integrating gives the following closed expressions.

$$\begin{aligned} B_x(x, y, z) &= M_0 [(\Theta(y - w, x - d, z + \ell) - \Theta(y - w, x + d, z + \ell)) \\ &\quad - (\Theta(y + w, x - d, z + \ell) - \Theta(y + w, x + d, z + \ell)) \\ &\quad - (\Theta(y - w, x - d, z - \ell) - \Theta(y - w, x + d, z - \ell)) \\ &\quad + (\Theta(y + w, x - d, z - \ell) - \Theta(y + w, x + d, z - \ell))] , \end{aligned}$$

$$\begin{aligned} B_y(x, y, z) &= M_0 [(\Theta(x - d, y - w, z + \ell) - \Theta(x + d, y - w, z + \ell)) \\ &\quad - (\Theta(x - d, y + w, z + \ell) - \Theta(x + d, y + w, z + \ell)) \\ &\quad - (\Theta(x - d, y - w, z - \ell) - \Theta(x + d, y - w, z - \ell)) \\ &\quad + (\Theta(x - d, y + w, z - \ell) - \Theta(x + d, y + w, z - \ell))] , \end{aligned}$$

and

$$\begin{aligned} B_z(x, y, z) &= M_0 [(\Gamma(x - d, y - w, z + \ell) - \Gamma(x + d, y - w, z + \ell)) \\ &\quad - (\Gamma(x - d, y + w, z + \ell) - \Gamma(x + d, y + w, z + \ell)) \\ &\quad - (\Gamma(x - d, y - w, z - \ell) - \Gamma(x + d, y - w, z - \ell)) \\ &\quad + (\Gamma(x - d, y + w, z - \ell) - \Gamma(x + d, y + w, z - \ell))] , \end{aligned}$$

with

$$\Theta(x, y, z) = \log(-x + \sqrt{x^2 + y^2 + z^2})$$

---

a complex defect  $\Sigma$  in terms of a finite sum of magnetic flux leakage fields generated by small enough parallelepipeds.

and

$$\Gamma(x, y, z) = y \operatorname{sgn}(z) \arcsin \left( \frac{y}{\sqrt{y^2 + z^2}} \frac{x}{\sqrt{x^2 + z^2}} \right),$$

where the sign function  $\operatorname{sgn}$  is defined as

$$\operatorname{sgn}(x) = \begin{cases} -1, & x < 0, \\ 0, & x = 0, \\ +1, & x > 0. \end{cases}$$

Figures 7, 8, and 9 depict the three components of the magnetic field  $\mathbf{B}$  in

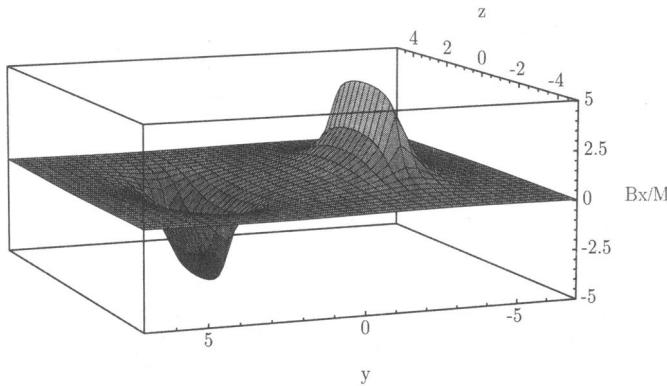


FIGURE 7. The  $x$ -component of the magnetic flux leakage field.

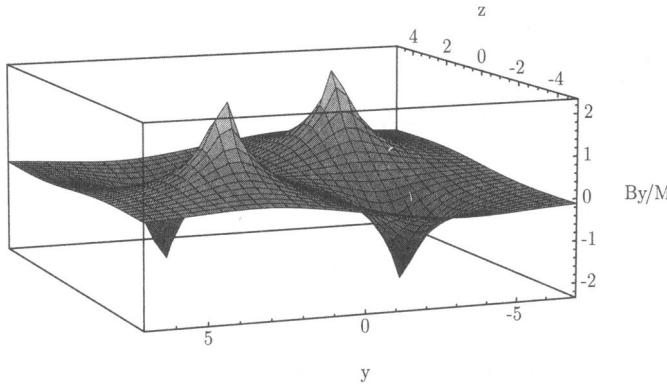
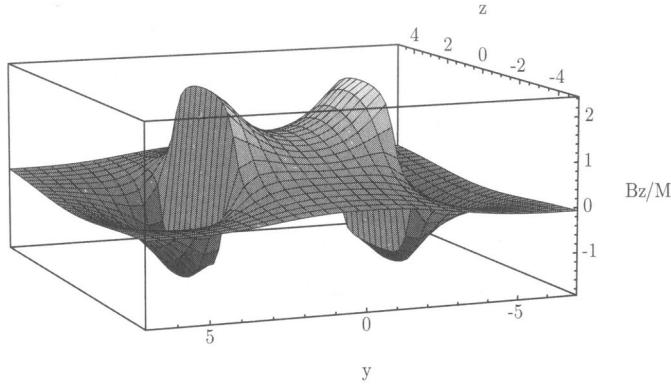
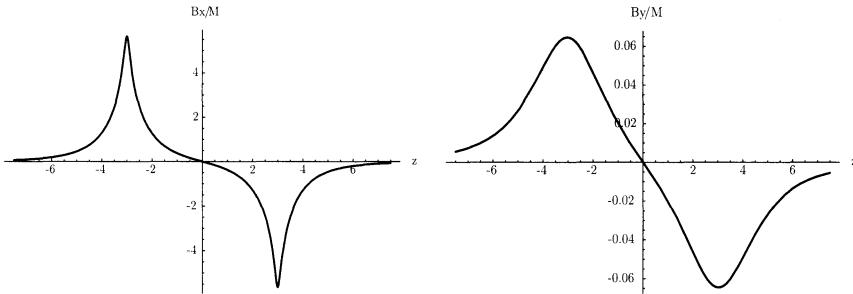
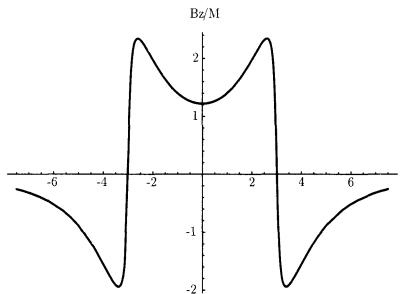


FIGURE 8. The  $y$ -component of the magnetic flux leakage field.

the case where  $\ell = 3$ ,  $w = 2$ ,  $d = 1$  and  $x = 1.1$ . Individual traces of these fields, as measured by a particular Hall sensor located at  $y = \text{const.}$ , are shown in Figures 10 and 11 for  $y = 0.1$ . It is worthwhile pointing out that the amplitude of the traces are quite different; the radial trace  $B_x(x = \text{const}_1, y = \text{const}_2, \cdot)$  and the axial trace  $B_z(x = \text{const}_1, y = \text{const}_2, \cdot)$  have the larger amplitudes, whereas the circumferential trace  $B_y(x = \text{const}_1, y = \text{const}_2, \cdot)$  is considerably smaller.

FIGURE 9. The  $z$ -component of the magnetic flux leakage field.FIGURE 10. Trace of  $B_x$  (left) and  $B_y$  (right).FIGURE 11. Trace of  $B_z$ .

The shapes of the depicted surfaces are characteristic of general magnetic flux leakage fields: The radial components  $B_x$  are bipolar and the circumferential components  $B_y$  are of quadrupole nature. The axial components  $B_z$  are symmetric about a plane perpendicular to the axis of the cylinder  $\Omega$  (for a symmetric defect.) One might conjecture that the shape of  $\mathbf{B}_\Sigma$ , in particular the intersection of graph  $\mathbf{B}_\Sigma$  with the  $yz$ -plane, outlines the defect or at least resembles it. Unfortunately, this is only true to a certain extent. Denote by  $P_{yz}\mathbf{B}_\Sigma$  the projection of graph  $\mathbf{B}_\Sigma$  onto the  $yz$ -plane. The maximum axial extend of the projection is a

number close to the length of  $\Sigma$  and can thus be used to approximate it. However, the maximum circumferential extend of  $P_{yz}\mathbf{B}_\Sigma$  is considerably larger than the actual geometric width of  $\Sigma$ . A precise relationship between what might be called the *magnetic width*  $w$  of a defect  $\Sigma$  and its geometric width  $w$  has not been found. Such a relationship would have to also involve the length and depth, as both numbers can be shown to affect  $w$ . The maximum value  $B_{\max}$  of  $\mathbf{B}_\Sigma$ , regarded as a function of length  $\ell$ , width  $w$ , and depth  $d$ , is an increasing function in each of the three variables when the other two are held constant. This property is most widely used in existing algorithms to grade the severity of a defect. The existence of a precise analytical relationship  $B_{\max} = f(\ell, w, d)$ , however, is still elusive.

It is possible to derive closed form expressions for a variety of geometrically attractive defects: cylindrical, (semi)spherical, elliptic parabolic, and combinations thereof. These closed expressions involve a large variety of special functions such as for instance Bessel<sup>17</sup> functions, Gegenbauer<sup>18</sup> polynomials, spherical harmonics, and Mathieu<sup>19</sup> functions. The complexity of the computations involving these types of defects, however, has impeded their usage in the solution of the associated inverse problem.

Besides the above-introduced simple model to compute the magnetic flux leakage field  $\mathbf{B}_\Sigma$ , there are several other models that improve upon the hypotheses and take into account some material properties. One obvious modification involves replacing the assumption that  $\mathbf{M}$  is constant by a more realistic one. The magnetic flux lines should be considered uniform and constant outside an (unknown) neighborhood of the defect  $\Sigma$  but within this neighborhood they resemble more the flux lines of fluids encountering an obstacle. The density of the magnetic flux lines is larger at the bottom of a defect than it is near the surface. In other words, the density of the Radon measure  $\mu_\Sigma$  is not uniform but rather place- and depth-dependent. This is exemplified in Figure 12. The computation of  $\mathbf{B}_\Sigma$  now includes the volume integral over this neighborhood of the defect (dashed lines in Figure 12). These modified models depend on a finite set of unknown parameters, such as  $\alpha$ ,  $\beta$ , and  $\gamma$ , that need to be determined from experimental or simulated data to obtain optimal agreement between the theoretical results and experimental measurements. Better agreement (see Figure 13) with actual magnetic flux measurements is achieved, in particular for the radial and circumferential components of  $\mathbf{B}_\Sigma$  but, unfortunately, the complexity of the ensuing analytical solutions makes these more realistic models difficult to deal with.

### 3. The Inverse Problem

In this section, the inverse to the problem of deducing the magnetic flux leakage field  $\mathbf{B}_\Sigma$  from a given defect  $\Sigma$  is introduced. This inverse problem consists of inferring the defect  $\Sigma$  or equivalently the Radon measure  $\mu_\Sigma$  from a finite set of indirect measurements of  $\mathbf{B}_\Sigma$ . Before addressing this particular inverse problem, a

<sup>17</sup>Friedrich Wilhelm Bessel, July 22, 1784 to March 17, 1846. German astronomer and mathematician who contributed to the theory of differential equations, astronomy, and geodesy.

<sup>18</sup>Leopold Gegenbauer, Feb 2, 1849 to June 3, 1903. Austrian mathematician who primarily work in algebra.

<sup>19</sup>Emile Léonard Mathieu, May 15, 1835 to Oct 19, 1890. French mathematician who contributed to mathematical physics and hypergeometric functions.

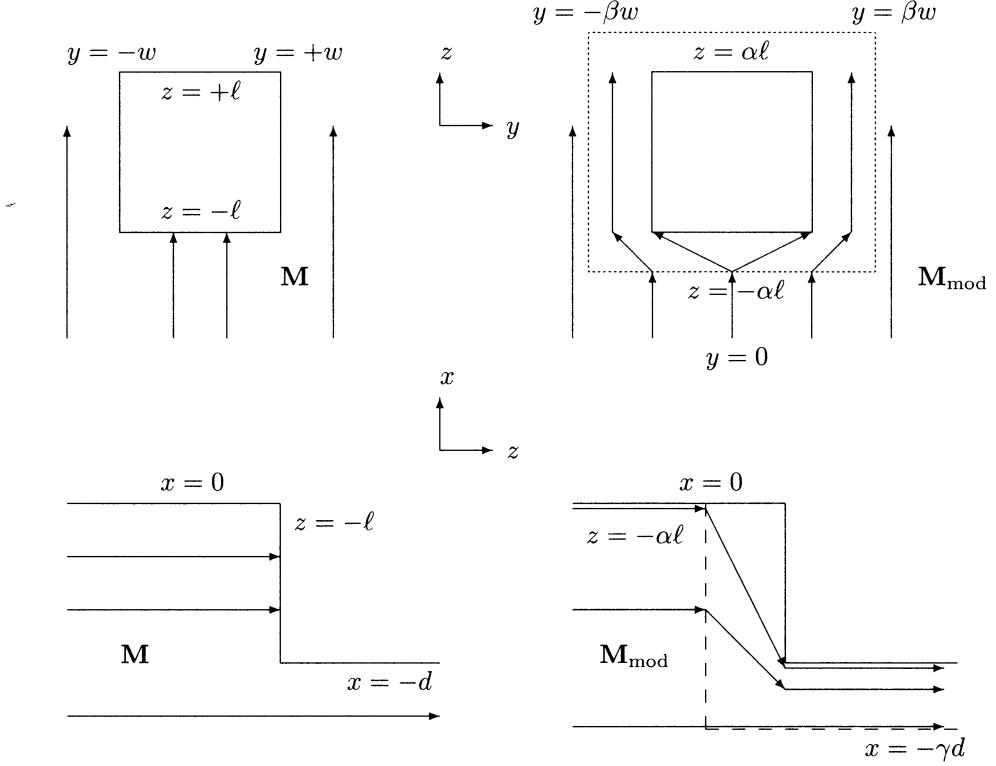


FIGURE 12. The magnetization  $\mathbf{M}$  for the dipole model (left) and a modification  $\mathbf{M}_{\text{mod}}$  (right).

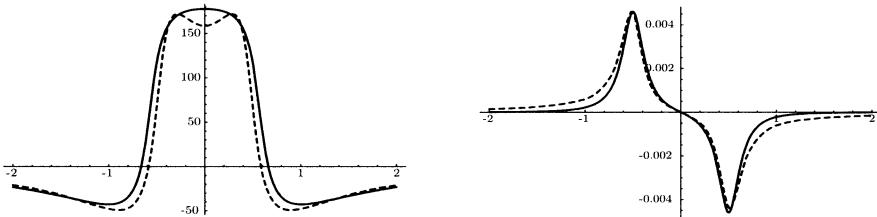


FIGURE 13. Comparison between the modified theoretical (solid) and experimental axial (left) and radial (right) MFL signals (dashed).

brief albeit incomplete review of inverse problem theory is presented. The interested reader is encouraged to consult some of the references for more details.

**3.1. Preliminaries.** For the definition of *direct* and *inverse problem*, [Ke] is often cited:

“We call two problems *inverses* of one another if the formulation of each involves all or part of the solution of the other. Often,

for historical reasons, one of the two problems has been studied extensively for some time, while the other is newer and not so well understood. In such cases, the former problem is called the *direct problem*, while the latter is called the *inverse problems*.”

For numerous applications, the aforementioned problems are not well-posed. The definition of a *well-posed problem* goes back to Hadamard<sup>20</sup> and is repeated here.

**DEFINITION 3.1** (Hadamard). A problem is called *well-posed* if it satisfies the following.

**Existence:** There exists a solution to the problem.

**Uniqueness:** There exists at most one solution to the problem.

**Stability:** The solution depends continuously on the data.

A problem that is not well-posed is called *ill-posed*. If for two problems that are inverses of one another, one of them is ill-posed, it is usually called the inverse problem and the other one the direct problem.

The existence of a solution for an inverse problem is in most cases secured by defining the data space to be the set of solutions to the direct problem. This approach may fail if the data is incomplete, perturbed, or noisy. If the uniqueness of a solution cannot be inferred from the given data, then one needs to either obtain additional data or restrict the set of admissible solutions using a-priori information about the solution. Of the three Hadamard requirements, stability is the most delicate to verify or obtain. If an inverse problem fails to be stable, then small round-off errors or noise in the data will amplify to a degree that renders a computed solution useless.

To illustrate some of the issues discussed above, the following example from [H] is employed.

**EXAMPLE 3.2.** It is well-known that differentiation and integration are two problems that are inverses to one another. For the purposes of this example, differentiation will be referred to as the inverse problem as it will turn out to be ill-posed.

Given any  $\varphi \in C[0, 1]$ , define an operator  $T$  acting on  $\varphi$  as follows.

$$(T\varphi)(x) := \int_0^x \varphi(t) dt, \quad x \in [0, 1].$$

The inverse problem consists of solving

$$(3.1) \quad T\varphi = f,$$

for given data  $f \in C[0, 1]$  satisfying  $f(0) = 0$ . Note that (3.1) is equivalent to computing  $\varphi = df/dx$ . Hence, (3.1) has a solution  $\varphi \in C[0, 1]$  iff  $f \in C^1[0, 1]$ . Therefore, the problem is well-posed if  $T$  maps  $C[0, 1]$  into  $C^1[0, 1]$ .

The situation changes, however, if for instance the data  $f$  is contaminated by noise and the noisy data  $f_\delta$  is only known to satisfy the estimate  $\|f - f_\delta\|_\infty \leq \delta$ ,  $0 < \delta \ll 1$ .

To be specific, suppose that

$$(3.2) \quad f_\delta^n(x) := f(x) + \delta \sin \frac{nx}{\delta}, \quad x \in [0, 1], \quad \delta \in (0, 1), \quad 2 \leq n \in \mathbb{N}.$$

---

<sup>20</sup>Jacques Solomon Hadamard, Dec 8, 1865 to Oct 17, 1963. French mathematician best known for his proof of the Prime Number Theorem. He also worked in the theory of partial differential equations.

Although

$$\|f_\delta^n - f\|_\infty \leq \delta, \quad \text{for all } n \in \mathbb{N},$$

the error in the solutions can become arbitrarily large:

$$\|df_\delta^n/dx - df/dx\|_\infty = \left\| n \cos \frac{nx}{\delta} \right\|_\infty = n.$$

A general setting for inverse problems is the following. Suppose that  $(X, \|\cdot\|_X)$  and  $(Y, \|\cdot\|_Y)$  are normed linear spaces. Let  $T : X \rightarrow Y$  be a (linear) operator and consider the operator equation

$$(3.3) \quad T\varphi = f.$$

The direct problem consists of determining  $f \in Y$  from given  $\varphi \in X$ , and the inverse problem of obtaining  $\varphi$  from the data  $f$ :  $\varphi = T^{-1}f$ . Hadamard's wellposedness conditions read now as follows.

- (1)  $T(X) = Y$ ;
- (2)  $T$  is injective;
- (3)  $T^{-1}$  is continuous.

For many applications, in particular those in this paper, the spaces  $X$  and  $Y$  are real Lebesgue<sup>21</sup> spaces  $L^p(\mathbb{R})$ ,  $1 \leq p < \infty$ . In this case, if  $T$  is a linear operator, then (3.3) can be written as an *integral equation of the first kind*:

$$(3.4) \quad (T\varphi)(x) = \int K(x, \xi) \varphi(\xi) d\xi,$$

for some kernel  $K$ .

**REMARK 3.3.** In case  $T$  is a compact<sup>22</sup> operator and  $X$  infinite-dimensional, then  $T^{-1}$  cannot be continuous (in the strong operator topology). For otherwise, the unit ball  $B = T^{-1}(T(B))$  in  $X$  would be relatively compact, which it is not.

**3.2. The inverse problem for pipeline inspection.** Next we discuss the inverse problem to (2.15): Given the finite set of indirect measurements  $\{\mathbf{Y}(\Sigma)\}$  corresponding to the magnetic flux leakage field  $\mathbf{B}_\Sigma$ , reconstruct the defect  $\Sigma$ .

To this end, recall that each Hall sensor  $H$  on the inspection device measures either the field  $\mathbf{H}$  (in the absence of a defect) or the magnetic flux leakage field  $\mathbf{B}_\Sigma$ . Each  $Y \in \mathbf{Y}$  thus represents a Hall voltage  $U_H$ , which by (1.1) is proportional to the magnitude of the magnetic field,  $\mathbf{H}$  or  $\mathbf{B}_\Sigma$ , measured by a particular Hall sensor  $H_j$ ,  $j = 1, \dots, n$ , at axial position  $z = z_i$ ,  $i = 1, \dots, m$ , where  $m$  is the number of measurements made and  $n$  the number of sensors on the pipe inspection device. If a cartesian coordinate system  $(x, y, z)$  is chosen in such a way that the  $z$ -axis coincides with the pipe axis, then as the inspection device moves down the pipeline, the trajectory of a sensor  $H_j$  is a line parallel to the pipe axis<sup>23</sup> given by the intersection of the planes  $x = x_j$  and  $y = y_j$ , where  $(x_j, y_j)$  denotes the fixed planar position of a Hall sensor  $H_j$  in the ring of sensors. Each Hall sensor  $H_j$ ,

---

<sup>21</sup>Henri Léon Lebesgue, June 28, 1875 to July 26, 1941. French mathematician famous for his theory of integration. He is one of the founders of measure theory.

<sup>22</sup>A linear operator  $T : X \rightarrow Y$  is called compact if it maps bounded sets in  $X$  to relatively compact sets (sets whose closure is compact) in  $Y$ .

<sup>23</sup>The inspection tool actually moves through the pipe along a helical path, but since the pitch of the helix is much larger than its radius, the effects due to it can be neglected.

$j = 1, \dots, n$ , measures the Hall voltage  $U_H$  at regularly sampled intervals along its trajectory. Thus,

$$(3.5) \quad Y = Y_{i,j} = c_H B_z^*(x_j, y_j, i \cdot h), \quad (i, j) \in I \times J.$$

Here we set  $I := \{1, \dots, m\}$ ,  $J := \{1, \dots, n\}$ ,  $\mathbf{B}^* = \mathbf{H}$  or  $\mathbf{B}_\Sigma$ , and assumed that the first measurement is taken at  $z = h$ . The quantity  $h$  is the reciprocal of the sample rate. The constant  $c_H$  may be obtained from the technical specifications of the Hall sensors and is therefore assumed to be known.

Equations (3.5) show that the magnetic flux leakage field is measured on the planes  $z = ih$ . Hence only the component of  $\mathbf{B}_\Sigma$  parallel to these planes, or equivalently, along the normals of these planes is sampled. In the language of the Radon or X-ray transform, it is sampling along a fixed direction only:  $\varphi = \text{constant}$ .

The  $m \times n$  system (3.5) is a collection of axially and circumferentially equally spaced samples of the magnetic field. Recall that for a fixed  $j \in \{1, \dots, n\}$  these samples are also called an MFL signal. Note that for all but a small number of pairs  $(i, j)$  the values  $Y_{i,j}$  are (approximately) constant. One may therefore normalize the  $Y_{i,j}$  in such a way that measurements of  $\mathbf{H}$  correspond to the value zero. We will adhere to this convention for the remainder of this paper.

The presence of a defect  $\Sigma$  causes the deviation of the  $Y_{i,j}$  from zero. Hence there exists a subset  $\mathfrak{S}$  of  $I \times J$  such that  $\mathbf{Y}(\Sigma) := \{Y_{i,j} : (i, j) \in \mathfrak{S}\}$  is the set of measurements of the magnetic flux leakage field  $\mathbf{B}_\Sigma$ . The cardinality of  $\mathfrak{S}$  is several orders of magnitude smaller than that of  $I \times J$ . Employing formula 2.15 yields

$$(3.6) \quad \begin{aligned} Y_{i,j}^\Sigma &= c_H \left[ - \int_{\Omega} \frac{(ih - \zeta) (\nabla_{\xi} \cdot \mathbf{M})(\xi)}{((x_j - \xi)^2 + (y_j - \eta)^2 + (ih - \zeta)^2)^{3/2}} dv \right. \\ &\quad \left. + \oint_{\Sigma} \frac{(ih - \zeta) \mathbf{M}(\xi) \cdot d\sigma(\xi)}{((x_j - \xi)^2 + (y_j - \eta)^2 + (ih - \zeta)^2)^{3/2}} \right], \quad (i, j) \in \mathfrak{S}. \end{aligned}$$

Here the superscript  $\Sigma$  was added to the  $Y_{i,j}$  to indicate these are values for a defect.

Let us reconsider the situation described in Example 2.1.

EXAMPLE 3.4. Assume again that  $\Sigma = [-d, d] \times [-w, w] \times [-\ell, \ell]$  and  $\mathbf{M} = M_0 \mathbf{e}_z$ . Then (3.6) reduces to

$$(3.7) \quad Y_{i,j}^\Sigma = c_H M_0 \int_{-d}^d \int_{-w}^w \frac{(ih - \zeta) d\xi d\eta}{((x_j - \xi)^2 + (y_j - \eta)^2 + (ih - \zeta)^2)^{3/2}},$$

for  $(i, j) \in \mathfrak{S}$ , with solution

$$(3.8) \quad \begin{aligned} Y_{i,j}^\Sigma &= c_H M_0 [(\Gamma(x_j - d, y_j - w, ih + \ell) - \Gamma(x_j + d, y_j - w, ih + \ell)) \\ &\quad - (\Gamma(x_j - d, y_j + w, ih + \ell) - \Gamma(x_j + d, y_j + w, ih + \ell)) \\ &\quad - (\Gamma(x_j - d, y_j - w, ih - \ell) - \Gamma(x_j + d, y_j - w, ih - \ell)) \\ &\quad + (\Gamma(x_j - d, y_j + w, ih - \ell) - \Gamma(x_j + d, y_j + w, ih - \ell))] , \end{aligned}$$

for  $(i, j) \in \mathfrak{S}$ . For a fixed sensor, say  $H_j$ , one may take three measurements and attempt to solve the system of three equations for the three unknowns  $\ell$ ,  $w$ , and  $d$ . For this purpose, the value  $M_0$  of the magnetization needs to be calculated. This can be done in principle by using (2.1).

Suppose then that the quantities  $Y_{i_k,j}/(c_H M_0)$  have been measured at  $z_k = i_k h$ ,  $k = 1, 2, 3$ . If we denote the expression in brackets in (3.8) evaluated at  $(x_j, y_j, z_k)$  by  $F_k(d, w, \ell)$  and consider the vector-valued function  $F : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  defined by

$$(3.9) \quad F(d, w, \ell) := \begin{pmatrix} F_1(d, w, \ell) - Y_{i_1,j}/(c_H M_0) \\ F_2(d, w, \ell) - Y_{i_2,j}/(c_H M_0) \\ F_3(d, w, \ell) - Y_{i_3,j}/(c_H M_0) \end{pmatrix},$$

then solving for the unknown depth, width, and length is paramount to finding the (unique positive) zero of  $F$ , if it exists. To solve nonlinear equations of the above type, one usually employs an iterative procedure and searches for a unique fixed point  $\mathbf{x}^* \in \mathbb{R}^3$  of  $F(\mathbf{x}) + \mathbf{x} = \mathbf{x}$ . In order to establish the existence of such a fixed point, conditions like the following need to be verified.

**THEOREM 3.5 ([SB], Theorem (5.2.3)).** *Let  $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be an iteration function,  $x_0 \in \mathbb{R}^n$  a starting point and  $x_{\nu+1} := \Phi(x_\nu)$ ,  $\nu \in \mathbb{Z}_0^+$ . Assume that there exists a neighborhood  $S_r(x_0) := \{x : \|x - x_0\| < r\}$  of  $x_0$  and a constant  $K$ ,  $0 < K < 1$ , such that*

- (1)  $\|\Phi(x) - \Phi(y)\| \leq K \|x - y\|$ ,  $\forall x, y \in \overline{S_r(x_0)} := \{x : \|x - x_0\| \leq r\}$ ;
- (2)  $\|x_1 - x_0\| = \|\Phi(x_0) - x_0\| \leq (1 - K) r < r$ .

Then it follows that for all  $\nu \in \mathbb{Z}_0^+$  one has:

- (1)  $x_\nu \in S_r(x_0)$ ;
- (2)  $\|x_{\nu+n} - x_\nu\| \leq K^\nu r$ , for all  $n \in \mathbb{N}$ ;
- (3)  $\Phi$  has exactly one fixed point  $x^*$  in  $\overline{S_r(x_0)}$  and

$$\lim_{\nu \rightarrow \infty} x_\nu = x^*.$$

For illustrative purposes and to be more specific, assume that we like to determine the dimensions of a defect  $\Sigma$ , which is the shape of a unit cube. Taking as an iteration function  $\Phi := F + \text{id}$ , and  $x_j := 0$ ,  $y_j := 0.51$ ,  $z_1 := 0$ ,  $z_2 := 0.75$ ,  $z_3 := 1$ , and working through the conditions of the above theorem, one can find a starting point  $(w_0, \ell_0, d_0) = (0.4, 0.4, 0.4)$  and an associated neighborhood  $S_{0.2}$  so that the iterates of  $(w_0, \ell_0, d_0)$  converge to a unique fixed point  $(0.495, 0.499, 0.502) \in S_{0.2}$ . Within numerical errors, these are clearly the dimensions of the “unknown” defect.

On the other hand, one can show that for certain values of  $x_j$  the hypotheses of Theorem 3.5 are not satisfiable. Hence, although one has an analytical solution for the direct problem and is able to formulate the inverse problem, no computationally efficient and robust algorithm has been found to give the correct dimensions for the simple defect  $\Sigma := [-1/2, 1/2] \times [-1/2, 1/2] \times [-1/2, 1/2]$ .

As the above example of a simple defect already indicates, the inverse problem for general complex defects is much more challenging. There are no analytical representations of such defects in terms of a finite number of characteristic parameters that can be obtained via some numerical procedure that is computationally stable and efficient. The following is the general formulation of the full inverse problem for complex defects.

The set of all defects  $\Sigma$  is a proper subset  $\mathfrak{D}$  of the set  $\mathbb{H}(\mathbb{R}^3)$  of non-empty compact subsets of  $\mathbb{R}^3$ . The latter set becomes a metric space when endowed with

the *Hausdorff*<sup>24</sup> metric  $\delta$  defined by

$$\delta(A, B) := \max \left\{ \max_{a \in A} \min_{b \in B} \|a - b\|, \max_{b \in B} \min_{a \in A} \|b - a\| \right\}.$$

Under this metric, two sets in  $\mathbb{H}(\mathbb{R}^3)$  have zero distance iff they coincide, i.e., are identical. The metric space  $(\mathbb{H}(\mathbb{R}^3), \delta)$  is called the *hyperspace of compact subsets of  $\mathbb{R}^3$* . The set  $\mathfrak{D}$ , which consists of compact surfaces represented by piecewise smooth functions, inherits the metric  $\delta$  and becomes a subspace of  $\mathbb{H}(\mathbb{R}^3)$ .

Recall that the direct problem is as follows. To a given defect  $\Sigma \in (\mathfrak{D}, \delta)$  in  $\Omega$  corresponds a magnetic flux leakage field  $\mathbf{B}_\Sigma \in \mathfrak{B} \subset C_0^1(\mathbb{R}^3, \mathbb{R}^3) := C^1(\mathbb{R}^3, \mathbb{R}^3) \cap \{\mathbf{f} : \mathbb{R}^3 \rightarrow \mathbb{R}^3 : \lim_{\|\mathbf{x}\| \rightarrow \infty} \mathbf{f}(\mathbf{x}) = \mathbf{0}\}$ . This correspondence is expressed through the operator  $T : \mathfrak{D} \rightarrow \mathfrak{B}$ ,

$$(3.10) \quad T\Sigma = \mathbf{B}_\Sigma = - \int_{\Omega} \frac{(\cdot - \xi) (\nabla_\xi \cdot \mathbf{M})(\xi)}{\|\cdot - \xi\|^3} dv + \oint_{\Sigma} \frac{(\cdot - \xi) \mathbf{M}(\xi) \cdot d\sigma(\xi)}{\|\cdot - \xi\|^3}.$$

The inverse problem asks to recover the geometry of a defect, given its magnetic flux leakage field  $\mathbf{B}_\Sigma$ . Formally this may be expressed as

$$\Sigma = T^{-1} \mathbf{B}_\Sigma.$$

One possible approach to solving the above inverse problem is to make use of the following observation that is based on the additivity of magnetic fields and integrals. Suppose that the contribution of the volume integral can be ignored and that it is possible to represent a defect as a formal sum of the form

$$(3.11) \quad \Sigma \sim \sum_{\nu \in N} c_\nu \Sigma_\nu,$$

where  $N$  is a finite index set,  $c_\nu \in \mathbb{Z}$ , and the  $\Sigma_\nu$  are known *basic defects*. As we assumed that defects are compact surfaces, simplicial homology theory shows that each such surface can be approximated arbitrarily close by a simplicial complex, i.e., a finite linear combination of 2-simplices with integer coefficients. More precisely, given any compact surface  $\mathcal{S} \subset \mathbb{R}^3$  and an  $\varepsilon > 0$ , there exists a finite number  $n$  and an associated *simplicial complex*  $K_n$  consisting of  $n$  2-simplices such that  $\delta(\mathcal{S}, K) < \varepsilon$ . (This is actually a theorem of Radó's<sup>25</sup> [R].) Hence, we could take

$\Sigma_\nu$  to be a 2-simplex  $\sigma_\nu^2 = \left\{ x \in \mathbb{R}^3 : x = \sum_{i=1}^3 \lambda_i v_i, \sum_{i=1}^3 \lambda_i = 1 \right\}$  with  $\{v_1, v_2, v_3\}$  a set of affinely independent points in  $\mathbb{R}^3$ . Then

$$\begin{aligned} \mathbf{B}_\Sigma &\approx \oint_{\sum_\nu c_\nu \sigma_\nu^2} \frac{(\cdot - \xi) \mathbf{M}(\xi) \cdot d\sigma(\xi)}{\|\cdot - \xi\|^3} \\ &= \sum_\nu c_\nu \oint_{\sigma_\nu^2} \frac{(\cdot - \xi) \mathbf{M}(\xi) \cdot d\sigma(\xi)}{\|\cdot - \xi\|^3} \\ &= \sum_\nu c_\nu \mathbf{B}_{\sigma_\nu^2}. \end{aligned}$$

---

<sup>24</sup>Felix Hausdorff, Nov 8, 1868 to Jan 26, 1942. German mathematician who is considered to be one of the founders of topology. He contributed considerably to set theory and functional analysis.

<sup>25</sup>Tibor Radó, June 2, 1895 to Dec 12, 1965. Hungarian mathematician who solved the Plateau Problem. He also worked in the areas of complex functions, variational calculus, and minimal surfaces.

Once an appropriate finite set of 2-simplices  $\{\sigma_\nu^2\}_{\nu=1}^N$  has been chosen, the magnetic flux leakage field  $\mathbf{B}_\Sigma$  is then expressed as a finite  $\mathbb{Z}$ -linear combination of the above form. This amounts to finding the coefficients  $c_\nu$  in this representation. Several options for obtaining the coefficients are available, for instance, neural nets, genetic algorithms, or support vector machines. Once the coefficients are found, the defect  $\Sigma$  is then approximately given by the linear combination (3.11). To present, however, a satisfactory reconstruction of a defect  $\Sigma$  from its measured magnetic flux leakage field  $\mathbf{B}_\Sigma$  is yet to be found.

#### 4. Signal Measurement Impediments

In the previous section, it was assumed that the measurements  $Y_{i,j}$  were *exact*, i.e., that there was no contamination of the data. In reality, however, the data samples are affected by numerous controllable and uncontrollable factors. These include the level of magnetization, sensor specifications and movement, tool velocity and performance, defect geometry, pipe material and its manufacturing process, electronic and material noise, pressure differences and temperature gradients in the medium, and debris in the medium.

As mentioned above, once the inspection device is inside the pipe, there is no communication taking place between the device and the outside. Measurements of the magnetic field take place in a highly uncontrolled environment and cannot be repeated. In order to solve the inverse problem, it is therefore imperative that data is as little as possible contaminated and that as many error variables as possible are removed or at least reduced to a level where they can be ignored. Instead of analyzing the samples  $Y_{i,j}$  of the surfaces  $B_{\Sigma,z}(x, y, z)$ , it is customary to only investigate its traces  $B_{\Sigma,z}(x_j, y_j, z)$ , i.e., the MFL signals, for each  $j = 1, \dots, n$ . In the following, the subscript  $j$  is dropped and it is understood that all equations and calculations refer to a fixed  $j$ .

The samples of the MFL signals that are collected and recorded by the inspection device and later retrieved are of the form

$$(4.1) \quad \mathfrak{Y}_i = \mathcal{D} Y_i + \lambda N_i, \quad (i, \cdot) \in \mathfrak{S}.$$

Here  $\mathcal{D}$  represents a *local degradation operator* affecting the quality of a measurement  $Y_i$ .<sup>26</sup> This situation includes missing or incorrect data due to sensor malfunctioning or limitations, velocity surges, wall thickness changes, and data acquisition. The set  $\{N_i : i = 1, \dots, m\}$  is a sequence of independent and identically distributed (i.i.d.) normal random variables with zero mean and unit standard deviation. The quantity  $\lambda$  represents the *noise level*. Thus, the term  $\lambda N_i$  represents *Gaussian White Noise* in the data samples and it can be successfully removed by employing a *wavelet-based denoising* technique due to Donoho and Johnstone [**Do2**, **DJ1**, **DJ2**, **DY**]. This technique and the associated wavelet-based denoising algorithm will be the content of Section 5.

**4.1. Signal degradation.** After the data has been denoised, it still may have to be enhanced in the sense that the effects of the degradation operator  $\mathcal{D}$  on the samples of the MFL signal need to be undone. Several degradations of the signal quality, such as changes in the wall thickness which reduce or increase the magnetization, movements of sensors away from the pipe wall resulting in signal amplitude

---

<sup>26</sup>Here we will not consider the case when two or more adjacent sensors fail to operate correctly and thus invalidate an entire set of measurements. This situation occurs very seldom.

reductions, or velocity surges by the tool due to pressure variations in the medium which results in an incorrect display of the samples, can be adjusted manually once the data is being analyzed. Malfunctioning of a sensor which results in recording a constant value may be remedied by replacing these constant sample values by an appropriately weighted average of the nearest neighbors of the malfunctioning sensor.

As already mentioned above, the maximum value  $B_{\max}$  of an MFL signal is used in existing algorithms to assess the severity of a defect. An incorrect measurement of  $B_{\max}$  or a missing maximum value can thus result in underestimating the severity. In order to remedy this type of signal degradation, one has to obtain an appropriate representation of the underlying continuous signal in terms of its sampled values.

**4.2. Signal sampling and representation.** When working with sampled values of a signal, one naturally has to ask to what degree the samples represent the signal and whether the entire signal can be reconstructed from its sampled values. An answer to these questions is given by the celebrated Whittaker-Shannon-Kotel'nikov (WSK) sampling theorem [Sh], which states that if a signal  $f$  does not contain frequencies above a certain value, say  $W$ , then it can be reconstructed in its entirety from an infinite number of samples  $f(n/2W)$ ,  $n \in \mathbb{Z}$ . More precisely, suppose  $f$  is such that it belongs to the space of *band-limited functions*, i.e.,

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-W}^{+W} e^{ix\xi} F(\xi) d\xi, \quad \text{for } F \in L^2[-W, W],$$

then it can be recovered from its samples  $\{f(n/2W) : n \in \mathbb{Z}\}$  via

$$(4.2) \quad f(x) = \sum_{n=-\infty}^{+\infty} f\left(\frac{n}{2W}\right) \frac{\sin \pi(2Wx - n)}{\pi(2Wx - n)}.$$

The condition that the signal be band-limited, i.e., its Fourier<sup>27</sup> transform compactly supported, can be quite restrictive for applications. Since MFL signals only have a finite number of zeros, they cannot be band-limited. (This follows from the fact that band-limited functions are associated via complex analysis with entire functions of exponential type.) Although there exist sampling theorems for not necessarily band-limited signals and error estimates involving inaccuracies in the amplitude and sample measurement and truncation of the infinite series [BBS], a representation of an MFL signal in terms of local basis functions would be more desirable. (Defects are local phenomena and variations of the MFL signal far from the defect should not affect its representation.)

A representation of an MFL signal that has been proven to be advantageous is given by *B-splines*. B-splines are special basis functions with remarkable properties [dB]. Moreover, they provide a computationally efficient and numerically stable framework for the evaluation of and approximation by splines. Sometimes, particularly in the presence of noise, it is advantageous to *approximate* rather than *interpolate* measured data. With this in mind, an MFL signal is represented in terms of fourth order B-splines employing what is called *Schoenberg*<sup>28</sup> variation

<sup>27</sup>Jean Baptiste Joseph Fourier, March 21, 1768 to May 16, 1830. French mathematician and physicist who is best known for his investigation of Fourier series and applications to heat flow.

<sup>28</sup>Isaac Jacob Schoenberg. April 21, 1903 to Feb 21, 1990. Romanian-born mathematician who is the founder of spline theory. He also worked in analysis and approximation theory.

*diminution.* This type of *quasi-interpolation*<sup>29</sup> preserves the shape (monotonicity and convexity) of the underlying samples and is therefore ideally suited for the purposes at hand. Employing this representation, several types of signal degradations have been reduced.

**4.3. Defect detection.** One immediate question that arises in the above context, involves the probability of detecting a defect, either on the interior or exterior of a pipe. Clearly, this probability depends on several parameters, some of which are not controllable, such as defect geometry and pipe material properties. Formulas can be derived for the probability of detection of a defect taking into account controllable and uncontrollable parameters. These formulas then give a lower bound for the width of an interior or exterior defect guaranteeing that it will be detected with probability one. Even if detected, a small defect may still only generate a magnetic flux leakage field  $\mathbf{B}_\Sigma$  whose magnitude is indistinguishable from the background noise level.

## 5. Wavelet-Based Denoising

Numerous sources exist that can contaminate the magnetic flux leakage (MFL) signal response from data gathered within a pipeline. These sources include

- System noise: This includes noise generated by the on-board electronics and the sensors. System noise can be modeled as additive white Gaussian noise, i.e., noise that has zero mean and unit standard deviation. It contributes the most to the high frequency noise in the data.
- Vibration-induced noise: The motion of the inspection tool in the pipe introduces a low frequency distortion of MFL data. As the inspection tool traverses the pipe, it reverberates between the pipe wall thus inducing slow variations in the data along the axial direction. These variations can mask small amplitude signals. Irregularities in the pipe surface can also contribute to this type of noise.
- Pipe noise: Pipe noise includes noise due to the grain structure of pipe material and seamless pipe noise. Seamless pipe noise is caused by variations in the pipe wall thickness introduced during the manufacturing process of seamless pipes.

Noise contributes to a reduction in the measurement and detection accuracy of an inspection device. This substantiates the need for methods to improve the signal-to-noise ratio (SNR) and thus the detectability of defects and accuracy of their characterization. At high magnetic field levels, noise is less prevalent than at lower magnetic field strengths, but smaller size defects can still be masked by noise. In this section, we present a technique that removes the unwanted (white) noise, thus improving the signal-to-noise ratio therefore enhancing feature detectability. A very successful denoising algorithm based on wavelets was proposed by Donoho and Johnstone [DJ1] and Donoho [Do1]. We follow the description of the denoising algorithm presented in [Do1]. Several other denoising methodologies have been designed to incorporate different types of noise such as Poisson [Ko] and

---

<sup>29</sup>A quasi-interpolant of order  $k$  is a linear mapping from the space of continuous functions into a space of splines reproducing polynomials up to order  $k$ .

Cauchy<sup>30</sup> [DY] or different types of fast wavelet algorithms such as the interpolatory [Do2] or redundant transforms [BW, CD]. The technique employed in the Donoho-Johnstone model removes (white) noise without much altering the frequency content of the uncorrupted MFL signal and is based on the mathematical theory of *wavelets*. Wavelets are special functions that are used to localize a given signal at different scales in both the time *and* frequency domain. Denoising is a procedure that attempts to remove any noise from a set of data and tries to restore the underlying function or signal while retaining its frequency content. Denoising should not be confused with smoothing, a process that indiscriminately removes all high frequencies and retains only low frequencies.

The basic idea behind a wavelet-based denoising algorithm is to describe the signal in terms of a *finite* number of coefficients, called the *wavelet coefficients*. The magnitude of these wavelet coefficients is indicative of the frequency content of the underlying signal. If a signal is contaminated by noise, its wavelet coefficients are large in those regions where the noise is prevalent. Thresholding the wavelet coefficients in an appropriate manner will remove the frequency contributions from the noise and leave the underlying signal intact with only a possibly minimal loss of information.

The advantage of wavelets over traditionally used denoising techniques lies in their ability to adaptively resolve a signal in both frequency *and* time domain and to provide fast algorithms, the so-called *discrete wavelet transform*, for the computation of the wavelet coefficients. The order of complexity of the wavelet transform is  $\mathcal{O}(N)$  as opposed to  $\mathcal{O}(N \log N)$  for the fast Fourier transform.

**5.1. Preliminaries on multiwavelets.** In this section, we give a brief introduction to the theory of multiwavelets and highlight some of their properties. As a specific example, the well-known Donovan-Geronimo-Hardin-Massopust (DGHM) multiwavelet is presented as it is employed in the denoising algorithm below. For a more detailed presentation of multiwavelets, the reader is referred to the references given in the bibliography [AL, DGHM, GHM, G, K, M, SN].

A finite collection of real-valued  $L^2$ -functions  $\Psi := (\psi_1, \dots, \psi_r)^T$  (The superscript  $T$  denotes the transpose operator.) is called a *multiwavelet* if the two-parameter family  $\{\Psi_{jk} := 2^{j/2}\Psi(2^j \cdot -k) : j, k \in \mathbb{Z}\}$  forms an orthonormal, or more generally, a Riesz<sup>31</sup> basis of  $L^2(\mathbb{R})$ . Recall that a system  $\{\psi_\nu\}$  is a Riesz basis for a separable Hilbert space  $\mathcal{H}$  iff there exists a bounded, linear, invertible operator  $U$  such that  $\{U\psi_\nu\}$  is an orthonormal basis of  $\mathcal{H}$ .

One way to construct a multiwavelet is through multiresolution analysis, which consists of a nested sequence  $V_j \subset V_{j+1}$ ,  $j \in \mathbb{Z}$ , of closed subspaces of  $L^2(\mathbb{R})$  with the property that the closure of their union is  $L^2(\mathbb{R})$  and their intersection is the trivial subspace  $\{0\}$ . Furthermore, each subspace  $V_j$  is spanned by the dyadic dilates and integer translates of a finite set of *scaling functions*  $\{\phi_i : i = 1, \dots, r\}$ , sometimes also called the *generators* of the multiresolution analysis. Typically, the *scaling vector*  $\Phi := (\phi_1, \dots, \phi_r)^T$  has compact support or decays rapidly enough at infinity. (Here, the support of a scaling vector  $\Phi$  is defined as the union of the

<sup>30</sup>Augustin Louis Cauchy, Aug 21, 1789 to May 23, 1857. French mathematician who was one of the pioneers of mathematical analysis formalizing and proving the foundations of the Calculus. He also made major contributions to complex function theory and complex analysis.

<sup>31</sup>Frigyes Riesz, Jan 22, 1880 to Feb 28, 1956. Hungarian mathematician who was one of the founders of functional analysis. He also contributed to measure and operator theory.

supports of its individual components.) For  $r = 1$  we obtain the classical wavelet systems as defined and discussed in, for instance, [C, D, RW].

The condition that the spaces  $V_j$  be nested implies that the scaling vector  $\Phi$  satisfies the following *two-scale matrix dilation equation* or *matrix refinement equation*

$$(5.1) \quad \Phi(x) = \sqrt{2} \sum_{k \in \mathbb{Z}} P(k) \Phi(2x - k),$$

where the *filter coefficient matrices*  $P(k)$  are  $r \times r$  matrices such that the sequence  $\{\|P(k)\|\ : k \in \mathbb{Z}\}$  is in  $\ell^2(\mathbb{Z})$ .

Denoting the  $L^2$ -orthogonal complement of  $V_j$  in  $V_{j+1}$  by  $W_j$ , it can be shown that there exists a multiwavelet  $\Psi$  such that  $W_j$  is spanned by the dyadic dilates and integer translates of  $\Psi$ . Moreover, a multiwavelet satisfies a two-scale matrix dilation equation of the form

$$(5.2) \quad \Psi(x) = \sqrt{2} \sum_{k \in \mathbb{Z}} Q(k) \Phi(2x - k),$$

where the  $r \times r$  matrices  $\{Q(k)\}_{k \in \mathbb{Z}}$  are in  $\ell^2(\mathbb{R}^{r \times r})$ .

The pair  $(\Phi, \Psi)$  will be called a *multiwavelet system*. As our emphasis is entirely on compactly supported and orthogonal multiwavelets, we assume that scaling vectors and multiwavelets satisfy the following conditions.

**Compact Support.** Both  $\Phi$  and  $\Psi$  have compact support. (This implies that the sums in (5.1) and (5.2) are finite).

**$L^2$ -Orthogonality.** The scaling vectors and multiwavelets are  $L^2$ -orthonormal in the following sense:

$$(5.3) \quad \begin{aligned} \langle \Phi_{jk}, \Phi_{j\ell} \rangle &= \delta_{k\ell} I_r, \\ \langle \Psi_{jk}, \Psi_{m\ell} \rangle &= \delta_{jm, k\ell} I_r, \\ \langle \Phi_{jk}, \Psi_{j\ell} \rangle &= O_r, \end{aligned}$$

where  $I$  and  $O$  denotes the identity and zero matrix, respectively. Here we defined the inner product of two vector-valued functions  $F$  and  $G$  by  $\langle F, G \rangle := \int_{\mathbb{R}} F(x) G^T(x) dx$ . Note that in the case of *complex-valued*  $L^2$  functions, the transpose operator  $T$  has to be replaced by the hermitian conjugate operator  $*$ .

In terms of the filter coefficient matrices the above orthogonality conditions read

$$\begin{aligned} \sum_{k \in \mathbb{Z}} P(k) P^T(k + 2\ell) &= \delta_{0\ell} I_r \\ \sum_{k \in \mathbb{Z}} Q(k) Q^T(k + 2\ell) &= \delta_{0\ell} I_r \\ \sum_{k \in \mathbb{Z}} P(k) Q^T(k + 2\ell) &= O_r, \quad \text{for all } \ell \in \mathbb{Z}. \end{aligned}$$

Since  $V_{j+1} = V_j \oplus W_j$ , every function  $f_{j+1} \in V_{j+1}$  can be *decomposed* into an “averaged” or “blurred” component  $f_j \in V_j$  and a “difference” or “fine-structure” component  $g_j \in W_j$ :

$$f_{j+1} = f_j + g_j.$$

This decomposition can be continued until  $f_{j+1}$  is decomposed into a coarsest component  $f_0$  and  $j$  difference components  $g_m$ ,  $m = 1, \dots, j$ :

$$(5.4) \quad f_{j+1} = f_0 + g_1 + \dots + g_j.$$

This *decomposition algorithm* can be reversed to give a *reconstruction algorithm*: Given the coarse components together with the fine structure components one reconstructs any  $f_j \in V_j$  via reversal of (5.4). Let us note that both algorithms are usually applied to the expansion coefficients (in terms of the underlying basis) of  $f$  and  $g$  and that they involve the matrices  $P(k)$  and  $Q(k)$ . More precisely, the *decomposition algorithm* applied to  $f \in V_{j+1}$  is defined by

$$(5.5) \quad \begin{aligned} f_{j+1} &= \sum_k \langle f, \Phi_{j+1,k} \rangle \Phi_{j+1,k} \\ &= \sum_k \langle f, \Phi_{jk} \rangle \Phi_{jk} + \sum_k \langle f, \Psi_{jk} \rangle \Psi_{jk}, \end{aligned}$$

where the inner products  $\langle f, \Phi_{j+1,k} \rangle$ ,  $\langle f, \Phi_{jk} \rangle$ , and  $\langle f, \Psi_{jk} \rangle$  (Note that the inner products are row vectors!) are related via

$$(5.6) \quad \langle f, \Phi_{jk} \rangle = \sum_m \langle f, \Phi_{j+1,m} \rangle P^T(m - 2k)$$

and

$$(5.7) \quad \langle f, \Psi_{jk} \rangle = \sum_m \langle f, \Psi_{j+1,m} \rangle Q^T(m - 2k).$$

Conversely, the *reconstruction algorithm* applied to a function  $f_j \in V_j$ ,  $f_j = \sum_k \langle f_j, \Phi_{jk} \rangle \Phi_{jk}$  and  $g_j \in W_j$ ,  $g_j = \sum_k \langle f_j, \Psi_{jk} \rangle \Psi_{jk}$  yields

$$(5.8) \quad \langle f_{j+1}, \Phi_{j+1,k} \rangle = \sum_m \langle f_j, \Phi_{jm} \rangle P(k - 2m) + \langle f_j, \Psi_{jm} \rangle Q(k - 2m).$$

Introducing the column vectors  $\mathbf{c}_{jk} := \langle f, \Phi_{jk} \rangle^T$  and  $\mathbf{d}_{jk} := \langle f, \Psi_{jk} \rangle^T$ , one can write (5.5) in the form

$$(5.9) \quad f_{j+1} = \sum_k \mathbf{c}_{jk}^T \Phi_{jk} + \mathbf{d}_{jk}^T \Psi_{jk},$$

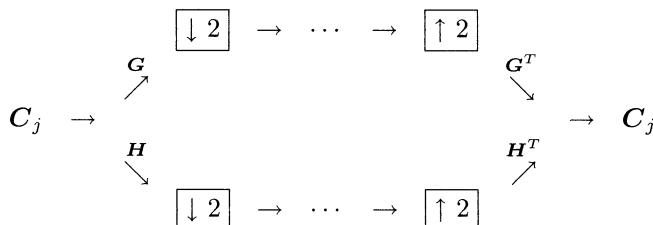
and (5.6) and (5.7) as

$$(5.10) \quad \mathbf{c}_{jk} = \sum_m P(m - 2k) \mathbf{c}_{j+1,m} \quad \text{and} \quad \mathbf{d}_{jk} = \sum_m Q(m - 2k) \mathbf{c}_{j+1,m},$$

while (5.8) is given by

$$(5.11) \quad \mathbf{c}_{j+1,k} = \sum_m P^T(k - 2m) \mathbf{c}_{jm} + Q^T(k - 2m) \mathbf{d}_{jm}.$$

Introducing the column vectors  $\mathbf{C}_j := (\mathbf{c}_{jk})$  and  $\mathbf{D}_j := (\mathbf{d}_{jk})$ , the decomposition and reconstruction algorithm may be schematically presented as follows.



where  $\mathbf{G}$  and  $\mathbf{H}$  are sparse *Toeplitz*<sup>32</sup> matrices with matrix entries ( $P$  and  $Q$ , respectively). One commonly refers to the matrices  $\mathbf{G}$  and  $\mathbf{H}$  as a *low pass* and *high pass filter*, respectively. The *downsampling operator*  $\downarrow$  uses only the even indices ( $2m$ ) at level  $j+1$  to obtain the coefficients at level  $j$ . The *upsampling operator*  $\uparrow$  inserts zero between consecutive indices at level  $j$  before  $\mathbf{G}$  and  $\mathbf{H}$  are applied to obtain the coefficients at level  $j+1$ .

As a consequence of the decomposition algorithm, any function  $f \in L^2(\mathbb{R})$  may be represented as a *multiwavelet series* of the form

$$(5.12) \quad f = \sum_{k \in \mathbb{Z}} \sum_{i=1}^r \langle f, \phi_{i,jk} \rangle \phi_{i,jk} + \sum_{m \geq j} \sum_{k \in \mathbb{Z}} \sum_{i=1}^r \langle f, \psi_{i,mk} \rangle \psi_{i,mk}$$

$$(5.13) \quad = \mathcal{P}^j[f] + \sum_{m \geq j} \mathcal{Q}^m[f],$$

where  $\mathcal{P}^j$  and  $\mathcal{Q}^j$  denote the orthogonal projectors of  $L^2(\mathbb{R})$  onto  $V_j$  and  $W_j$ ,  $j \in \mathbb{Z}$ , respectively. Here we defined  $\phi_{i,jk} = 2^{j/2} \phi_i(2^j \cdot -k)$  and  $\psi_{i,jk} = 2^{j/2} \psi_i(2^j \cdot -k)$ .

For approximation-theoretic purposes, the spaces  $V_j$  are usually required to reproduce polynomials up to a certain degree  $D-1$ . As the multiwavelet space  $W_j$  is orthogonal to  $V_j$ , the first  $D$  moments of the multiwavelet  $\Psi$  vanish

$$(5.14) \quad \int_{\mathbb{R}} x^p \Psi(x) dx = 0, \quad p = 0, \dots, D-1.$$

Such a multiwavelet system will be called a *multiwavelet system of order  $D$* . For the remainder of this paper, we assume that we always deal with a multiwavelet system of order  $D > 0$ . Note that if  $f$  is a polynomial of degree at most  $D-1$ , then its representation (5.13) reduces to  $f = \mathcal{P}^j[f]$ . In the case  $r=1$  this in particular implies that the span of  $\phi$  contains all polynomials of degree  $< D$ . For  $r > 1$ , the span of each individual scaling function  $\phi_i$  may in general not contain all such polynomials. (See [LV, S] for examples and details.)

In general, the projection  $\mathcal{P}^j[f]$  is at least as smooth as the most irregular component of the scaling vector  $\Phi$ . In particular, if  $\phi_i$  is in the Sobolev<sup>33</sup> space  $H^\alpha(\mathbb{R})$ ,  $\alpha > 0$ , for  $i = 1, \dots, r$ , then  $\mathcal{P}^j[f] \in H^\alpha(\mathbb{R})$  for each  $j \in \mathbb{Z}$ . (See the Appendix for definitions of function spaces used in this paper.)

It is well-known that a multiwavelet system of order  $D$  satisfies the following Jackson<sup>34</sup>-type inequality.

LEMMA 5.1. *Suppose that  $f \in C^n(\mathbb{R})$ ,  $1 \leq n \leq D$ , is compactly supported. Then*

$$(5.15) \quad \|f - \mathcal{P}^j[f]\|_{L^2} \leq C 2^{-jn},$$

for a positive constant  $C$  independent of  $j$  and  $n$ .

The exact relationship between the reproduction of polynomials by the integer shifts of  $\Phi$  and the  $L^p$ -approximation order of  $\mathcal{P}^j$  is discussed in [P].

<sup>32</sup>Otto Toeplitz, Aug 1, 1881 to Feb 15, 1940. German mathematician who worked on infinite linear and quadratic forms. He also was very interested in the history of mathematics.

<sup>33</sup>Sergei Lvovich Sobolev, Oct 6, 1908 to Jan 3, 1989. Russian mathematician who made major contributions to mathematical analysis and partial differential equations.

<sup>34</sup>Dunham Jackson, 1888 to 1946. American mathematician noted for his contributions to approximation theory.

Representing discretely sampled data in terms of multiwavelets requires special care since there is more than one generator for the spaces  $V_j$ . Suppose that  $\mathbf{f} \in \ell^2(\mathbb{Z})$  is a discrete scalar signal, representing the samples of a function  $f \in L^2(\mathbb{R})$ , and that the resolution of the samples is such that one has a representation of the form  $f = \sum_k \mathbf{c}_k^T \Phi_{jk}$ . Next, we show how the samples in  $\mathbf{f}$  are assigned to the coefficients  $\mathbf{c}$ . For this purpose, we consider the *polyphase form*  $\mathbf{F} \in (\ell^2(\mathbb{Z}))^r$  of  $\mathbf{f}$  defined by

$$(5.16) \quad \mathbf{F}(k) := \begin{pmatrix} f(rk) \\ \vdots \\ f(rk + r - 1) \end{pmatrix},$$

where  $\mathbf{f}(i)$  denotes the  $i$ th component of  $\mathbf{f} \in \ell^2(\mathbb{Z})$ .

Define a mapping  $\vartheta : (\ell^2(\mathbb{Z}))^r \rightarrow (\ell^2(\mathbb{Z}))^r$  by  $\mathbf{c}^T = \vartheta(\mathbf{F})$ . To proceed, we need the following result. (For a proof see, for instance, [F].)

**THEOREM 5.2.** *Suppose  $T : \ell^2(\mathbb{Z}) \rightarrow \ell^2(\mathbb{Z})$  is a bounded, shift-invariant linear transformation. Then there exists a  $q \in \ell^2(\mathbb{Z})$  such that*

$$(5.17) \quad T(\zeta) = q * \zeta, \quad \forall \zeta \in \ell^2(\mathbb{Z}).$$

Here  $* : \ell^2(\mathbb{Z}) \times \ell^2(\mathbb{Z}) \rightarrow \ell^2(\mathbb{Z})$  denotes the *convolution operator* defined by:

$$\{\zeta(\nu)\} * \{\theta(\nu)\} := \left\{ \sum_{\mu} \zeta(\mu) \theta(\nu - \mu) \right\}_{\nu=-\infty}^{+\infty}.$$

Thus, if  $\vartheta$  is a bounded linear and shift-invariant transformation with an inverse  $\vartheta^{-1}$  satisfying the same conditions, then both can be represented as a convolution:

$$\vartheta(\zeta) = q * \zeta, \quad \text{and} \quad \vartheta^{-1}(\zeta) = \tilde{q} * \zeta$$

where the sequences of  $r \times r$ -matrices  $q$  and  $\tilde{q}$  are called a *prefilter for  $\Phi$*  and *postfilter for  $\Phi$* , respectively.

In order to exploit the full power of filter banks, the filters  $q$  and  $\tilde{q}$  should be orthogonal (preserving the  $L^2$ -norm or energy of the signal) and preserve the approximation order  $D$  of the multiwavelet system. In [HR] such pre- and post-filters are constructed and applied to image compression. The construction of multiwavelet filters and the design for optimal orthogonal prefilters can be found in [HXGS] and [AHW], respectively.

**5.2. The GHM scaling vector and DGHM multiwavelet.** Next we consider a special scaling vector and associated multiwavelet that is being used in the denoising algorithm below. This so-called *GHM scaling vector* and *DGHM multiwavelet* were first introduced in [GHM, DGHM] and later in [M]. This multiwavelet system was the first example exhibiting wavelets that are compactly supported, continuous, orthogonal, and possess symmetry. Both the scaling vector and the multiwavelet are two-component vector functions  $\Phi = (\phi_1, \phi_2)^T$  and  $\Psi = (\psi_1, \psi_2)^T$  enjoying the following properties.

- $\text{supp } \phi_1 = [0, 1]$  and  $\text{supp } \phi_2 = \text{supp } \psi_1 = \text{supp } \psi_2 = [-1, 1]$ .
- The scaling vector  $\Phi$  and the associated multiwavelet  $\Psi$  satisfy (5.3) for  $L^2$ -orthonormality.
- The wavelets  $\psi_1$  and  $\psi_2$  are antisymmetric and symmetric, respectively.

- The multiwavelet system  $(\Phi, \Psi)$  is of order  $D = 2$ , i.e., has approximation order two: Constant and linear functions are reproduced by  $\Phi$ . This is equivalent to  $\Psi$  having two vanishing moments.
- $\Phi, \Psi \in (C^{0,1}(\mathbb{R}))^2$ . (See the Appendix for a definition of  $C^{0,1}(\mathbb{R})$ .) Hence all four component functions possess a weak first derivative.
- The GHM scaling vector is interpolatory: Given a set of interpolation points  $Z := \{Z_i\}$  supported on  $\frac{1}{2}\mathbb{Z}$ , there exists a set of vector coefficients  $\{\alpha_k\}$  such that  $\Theta(x) := \sum_k \alpha_k^T \Phi(x - k)$  interpolates  $Z$ . (Note that  $\phi_1(1/2) = 1 = \phi_2(0)$ .)
- The DGHM multiwavelet system can be easily modified to obtain a multiresolution analysis on  $L^2[0, 1]$  without the addition of boundary functions.

The GHM scaling vector and the DGHM multiwavelet are depicted in Figure 14.

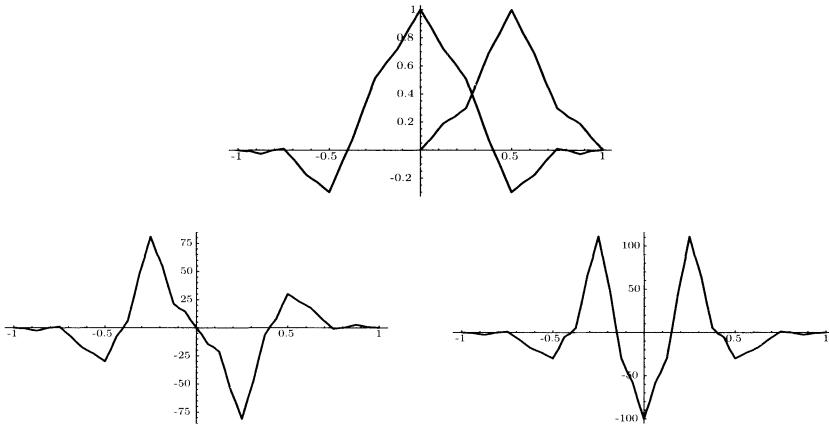


FIGURE 14. The orthogonal GHM scaling vector (top) and the orthogonal DGHM multiwavelet:  $\psi_1$  (bottom left) and  $\psi_2$  (bottom right).

**5.3. A multiresolution analysis on  $L^2[0, 1]$ .** As was mentioned above, it is possible to obtain a multiresolution analysis on an interval by modifying the DGHM multiwavelet system. Here we show how this may be accomplished on  $[0, 1]$ . The process involved in obtaining bases on  $[0, 1]$  without introducing additional boundary functions, as is the case for other wavelet constructions, only has to make use of the fact that the GHM scaling vector and the associated DGHM multiwavelet are *piecewise fractal functions* [DGHM, GHM, M]. The main idea is as follows. At any given level of approximation  $j \geq 0$ , take as a basis the restrictions to  $[0, 1]$  of all the translates of  $\phi_1$  and  $\phi_2$ , respectively,  $\psi_1$  and  $\psi_2$  at level  $j$  whose support has nonempty intersection with the open interval  $(0, 1)$ . More precisely, if

$$(5.18) \quad \phi_{i,jk}^* := \phi_{i,jk}|_{[0,1]} \quad \text{and} \quad \psi_{i,jk}^* := \psi_{i,jk}|_{[0,1]},$$

then the following, easily verified, theorem holds [DGHM, M].

**THEOREM 5.3.** *For all  $j \in \mathbb{Z}_0^+$ , the set  $\mathcal{B}_{\Phi,j}^* := \{\phi_{i,jk}^* : i = 1, 2; k = 0, 1, \dots, 2^j - 2 + i\}$  is an orthonormal basis for  $V_j^* := V_j \cap L^2[0, 1]$  and  $\mathcal{B}_{\Psi,j}^* := \{\psi_{i,jk}^* :$*

$i = 1, 2; k = i - 1, \dots, 2^j + 1 - i\}$  constitutes an orthonormal bases for  $W_j^* := W_j \cap L^2[0, 1]$ . Moreover,  $\text{card } \mathcal{B}_{\Phi,j}^* = 2^{j+1} + 1$ ,  $\text{card } \mathcal{B}_{\Psi,j}^* = 2^{j+1}$  and  $L^2[0, 1] = V_0^* \bigcup_{j \geq 0} W_j^*$ .

We note that the elements in  $V_j^*$  provide interpolation on the lattice  $2^{-(j+1)}\mathbb{Z}$ : The scaling function  $\phi_{2,jk}$  interpolates at  $2^{-j}\mathbb{Z}$ , whereas the function  $\phi_{1,jk}$  interpolates in-between, i.e., on  $2^{-(j+1)}\mathbb{Z}$ .

**5.4. Function sampling on  $[0, 1]$ .** In many applications one deals with a finite amount of data that needs to be analyzed or stored in a buffer for later retrieval. In order to employ a multiscale decomposition of the type introduced above, one chooses a finest level of approximation, say  $J > 0$ , and takes  $2^{J+1} + 1$  data points or samples. (This is the number of GHM scaling functions on  $[0, 1]$  at level  $J > 0$  with data supported on  $2^{-(J+1)}\mathbb{Z}$ .) In this case, we denote the collection of samples by  $\mathbf{f}_J = (f_i)_{i=0}^{2^{J+1}}$ . Using the elements in  $\mathcal{B}_{\Phi,j}^*$ , which we express in the form

$$\boldsymbol{\phi}^* := (\phi_{2,J,0}^*, \phi_{1,J,0}^*, \phi_{2,J,1}^*, \dots, \phi_{1,J,2^J}^*, \phi_{2,J,2^J+1}^*)^T,$$

we need to assign a data vector  $\mathbf{c}_J$  to this collection of samples. This is done via the polyphase representation applied now to the case  $r = 2$ . In [HR], orthogonal pre- and post filters that preserve the approximation order  $D = 2$  of the DGHM multiwavelet system were constructed. Employing these filters yields the required assignment  $\mathbf{f}_J \mapsto \mathbf{c}_J$ .

Applying the decomposition and reconstruction algorithm to a finite set of data such as  $\mathbf{c}_J$  is now straightforward. The length of the data vector  $\mathbf{c}_J$  equals  $2^{J+1} + 1$  and application of the matrices  $\mathbf{G}$  and  $\mathbf{H}$ , followed by downsampling  $\downarrow 2$ , produces two data vectors  $\mathbf{c}_{J-1}$  and  $\mathbf{d}_{J-1}$  of length  $2^J + 1$  and  $2^J$ , respectively. The data vector  $\mathbf{c}_{J-1}$  may be regarded as a weighted average with respect to the filter coefficients in  $\mathbf{G}$  of the original data vector  $\mathbf{c}_J$ , and the vector  $\mathbf{d}_{J-1}$  carries the information that was lost in the averaging procedure  $\mathbf{c}_J \mapsto \mathbf{c}_{J-1}$ . Thus, the data vector  $\mathbf{d}_{J-1}$  contains the detail or fine structure of the original data  $\mathbf{f}$ . The data vector  $\mathbf{c}_{J-1}$  may further be decomposed according to the scheme

$$\begin{array}{ccccccc} \mathbf{c}_J & \xrightarrow{\quad} & \mathbf{c}_{J-1} & \xrightarrow{\quad} & \cdots & \xrightarrow{\quad} & \mathbf{c}_L \\ & \searrow & & \searrow & \cdots & \searrow & \\ & & \mathbf{d}_{J-1} & & & & \mathbf{d}_L \end{array}$$

The mapping  $\mathcal{W} : \ell^2(\mathbb{Z}) \rightarrow \ell^2(\mathbb{Z})$ ,  $\mathcal{W}(\mathbf{c}_J) := (\mathbf{d}_{J-1}, \mathbf{c}_{J-1})$ , is called the discrete (*multi*)wavelet transform. Repetitively applying  $\mathcal{W}$  until a coarsest level  $L < J$  is reached, yields a multiscale representation of the original data vector  $\mathbf{c}_J$  in the form

$$\mathbf{c}_J = (\mathbf{d}_{J-1}, \mathbf{d}_{J-2}, \dots, \mathbf{d}_L, \mathbf{c}_L)$$

where the lengths of the multiscale components are  $(2^J, 2^{J-1}, \dots, 2^{L+1}, 2^{L+1} + 1)$ .

Reconstruction proceeds according to the scheme

$$\begin{array}{ccccccc} \mathbf{c}_L & \xrightarrow{\quad} & \mathbf{c}_{L+1} & \xrightarrow{\quad} & \cdots & \xrightarrow{\quad} & \mathbf{c}_{J-1} & \xrightarrow{\quad} & \mathbf{c}_J \\ & \nearrow & & \nearrow & \cdots & & \nearrow & & \\ & & \mathbf{d}_L & & \mathbf{d}_{L+1} & & \cdots & & \mathbf{d}_{J-1} \end{array}$$

Note that for the reconstruction, the data vectors  $\mathbf{c}_j$  and  $\mathbf{d}_j$  need to be upsampled,  $\uparrow 2$ , in order to generate  $\mathbf{c}_{j+1}$ ,  $L \leq j < J$ .

**5.5. The stochastic denoising model.** We consider samples of a continuous function  $Y$  defined on the interval  $[0, 1]$  which are corrupted by Gaussian white noise. The basic stochastic model for the measured or observed noisy data  $\mathfrak{Y}_i$  is of the form

$$(5.19) \quad \mathfrak{Y}_i = Y_i + \lambda N_i, \quad i = 0, 1, \dots, n - 1.$$

where  $\{Y_i = Y(i/n)\}_{i=0}^{n-1}$  are the samples of  $Y$  and the set  $\{N_i\}$  is a sequence of independent and identically distributed (i.i.d.) random variables whose probability distribution functions (pdfs)  $p(t)$  are all normal with zero mean and unit standard deviation:

$$p(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right).$$

Such pdfs are also denoted by  $N(0, 1)$ . The quantity  $\lambda$  denotes the *noise level*.

The goal of denoising is to find an estimator  $\hat{Y}$  of  $Y$  such that the mean square error between  $\hat{Y}$  and  $Y$  is minimal,

$$n^{-1} \mathbb{E} \|\hat{Y} - Y\|_{\ell^2}^2 = n^{-1} \sum_{i=0}^{n-1} \mathbb{E} (\hat{Y}(i/n) - Y(i/n))^2 = \min.$$

and that *with high probability*  $\hat{Y}$  is at least as smooth as  $Y$ . The latter requirement guarantees that undesirable, noise-induced structures such as ripples, blips, and spurious oscillations disappear. These undesirable features are generally present if one only tries to optimize the mean square error between  $\hat{Y}$  and  $Y$  and may lead to interpretation errors.

The main idea of the denoising algorithm developed by Donoho and Johnstone is to compute the wavelet coefficients of the noisy data  $\{\mathfrak{Y}_i\}$ , characterize their magnitude, and to find a threshold that removes the noise almost surely as  $n \rightarrow \infty$ .<sup>35</sup> The now famous expression for the threshold  $\tau = \lambda\sqrt{2\log n}$  is a consequence of the following theorem from probability theory [**LLR**].

**THEOREM 5.4.** *Suppose  $\{N_i\}_{i=1}^n$  is an i.i.d. sequence of  $N(0, 1)$  random variables. Then*

$$(5.20) \quad \text{Prob} \left\{ \max_{i=1, \dots, n} N_i \leq \sqrt{2\log n} \right\} \rightarrow 1 \quad \text{as } n \rightarrow \infty.$$

In our context the above theorem means that if we choose for the threshold  $\tau$  the value  $\lambda\sqrt{2\log n}$ , then with probability one all noise in the wavelet coefficients will be removed as  $n \rightarrow \infty$ .

There are two thresholding operations: hard and soft. The *hard-thresholding operator*  $\eta_\tau^{\text{hard}}$  assigns to an input  $t$  the output

$$(5.21) \quad \eta_\tau^{\text{hard}}(t) = \begin{cases} t, & |t| \geq \tau, \\ 0, & \text{otherwise.} \end{cases}$$

whereas the *soft-thresholding operator*  $\eta_\tau^{\text{soft}}$  first shifts all coefficients uniformly to zero by the amount  $\tau$ , and then sets them equal to zero if they are less than zero:

$$(5.22) \quad \eta_\tau^{\text{soft}}(t) = \text{sgn}(t)(|t| - \tau)_+,$$

---

<sup>35</sup>The discrete wavelet transform  $\mathcal{W}$  applied to the data vector  $\mathfrak{Y} = (\mathfrak{Y}_i)$  produces  $\mathcal{W}(\mathfrak{Y}) = \mathcal{W}(\mathbf{f}) + \lambda\mathcal{W}(\mathbf{N})$ ,  $\mathbf{N} := (N_i)$ . Since we are dealing with orthogonal scaling functions and wavelets,  $\mathcal{W}(\mathbf{N})$  is again an i.i.d. sequence of  $N(0, 1)$  random variables.

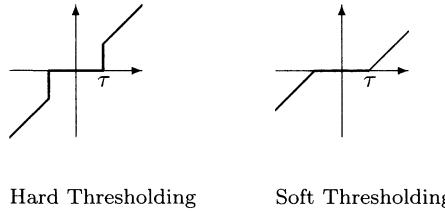
where  $\text{sgn}$  denotes the signum function defined by

$$\text{sgn}(y) = \begin{cases} +1, & t > 0, \\ -1, & t < 0. \end{cases}$$

and the notation  $(\cdot)_+$  means

$$t_+ = \begin{cases} t, & \text{if } t > 0, \\ 0, & \text{if } t \leq 0. \end{cases}$$

The following figure displays the action of these two different thresholding operators on an input.



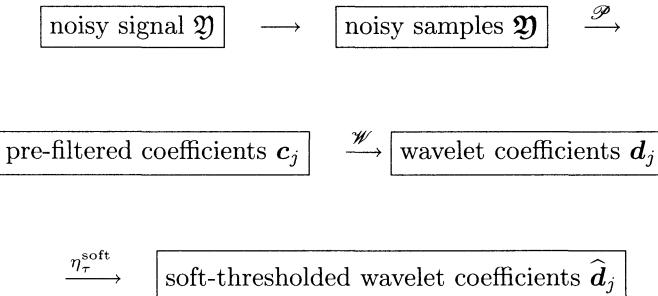
The Donoho – Johnstone denoising model uses soft thresholding for the removal of the noise. Soft thresholding guarantees the optimal recovery, in the sense of smoothness, of the function or signal  $Y$  (Cf. [Do1]). In what follows, we also use this procedure without explicit mention every time.

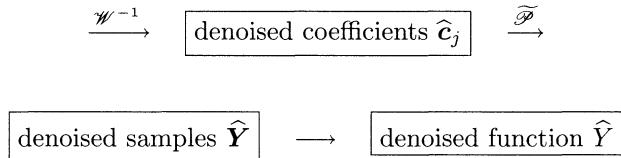
**5.6. Denoising algorithm.** Given the noisy samples  $\mathfrak{Y}$  (5.19), apply the pre-filtering operation  $\mathcal{P}$  to obtain the wavelet coefficients  $\mathbf{c}_J$  for the data array  $\mathfrak{Y}$ . Then, the wavelet transform  $\mathcal{W}$  is successively applied to  $\mathbf{c}_J$  to obtain its multiscale decomposition

$$\mathbf{c}_J = (\mathbf{d}_{J-1}, \mathbf{d}_{J-2}, \dots, \mathbf{d}_L, \mathbf{c}_L)$$

and the Donoho – Johnstone denoising algorithm is applied to each set of detail coefficients  $\mathbf{d}_j$ ,  $L \leq j < J$  using the soft-threshold operator  $\eta_\tau^{\text{soft}}$ , where  $\tau = \lambda\sqrt{2\log n}$ . (For each level  $L \leq j < J$ , the number  $n$  is the length of the data vector  $\mathbf{c}_j$ .) This produces the denoised coefficients  $\hat{\mathbf{d}}_j$ . Now apply the inverse wavelet transform  $\mathcal{W}^{-1}$  (the reconstruction algorithm) followed by the post-filtering operation  $\widetilde{\mathcal{P}}$  to obtain the denoised samples  $\widehat{\mathbf{Y}}$ , yielding  $\widehat{\mathbf{Y}}$  the, in the sense of smoothness, optimal estimator for the function or signal  $Y$ .

The aforementioned algorithm can be represented schematically as follows.





A good and more detailed description of multiwavelet denoising and compression of signals is [HSSTH].

In Figures 15 and 16 we show examples of the above-mentioned denoising algorithm using actual data obtained from a pipe inspection device.

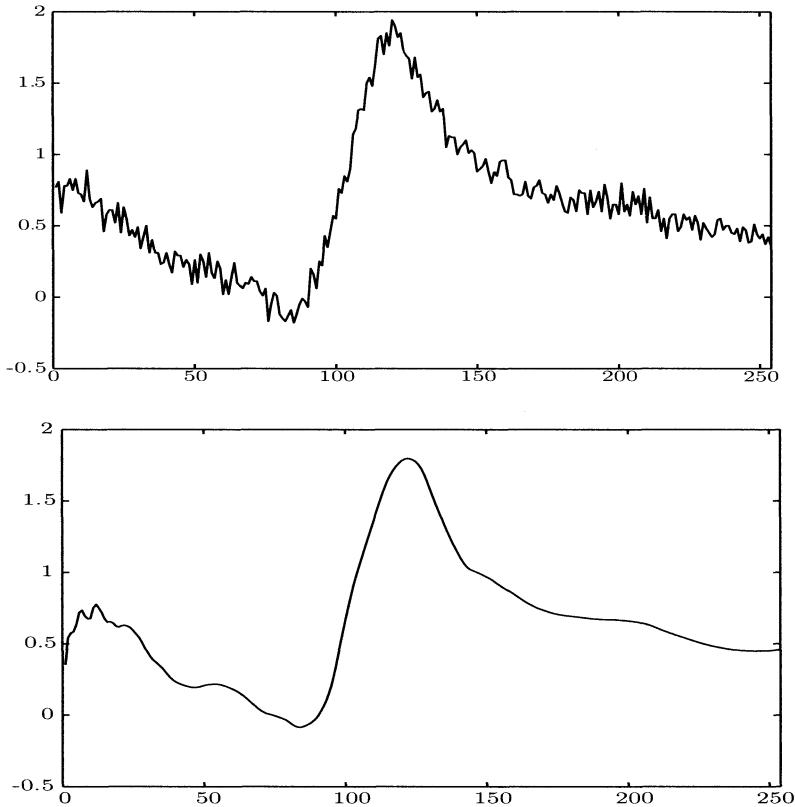


FIGURE 15. Noisy (top) and denoised (bottom) MFL signal.

**REMARK 5.5.** The noise level  $\lambda$  may be unknown for some practical applications and has therefore to be estimated. One such estimate is given by

$$\hat{\lambda} = \mu / 0.6745,$$

where  $\mu$  is the median absolute value of the appropriately normalized wavelet coefficients  $d_j$ ,  $L \leq j < J$ . For a derivation of this estimate, we refer the reader to [Do1].



FIGURE 16. Noisy (top) and denoised (bottom) MFL signals.

## 6. Appendix

Here we provide the definitions and some of the basic properties of the function spaces introduced in this paper. More details can be found in [T].

**Hölder Spaces.** Let  $\Omega \subset \mathbb{R}^n$  be an open set. A real-valued function  $f$  on  $\Omega$  is called *Hölder continuous with exponent*  $0 < s \leq 1$  if there exists a positive constant  $c$  so that

$$\frac{|f(x) - f(x')|}{|x - x'|^s} \leq c < \infty$$

holds for all  $x, x' \in \Omega$ ,  $x \neq x'$ . We note that if  $f$  is Hölder continuous with exponent  $s > 1$ , then  $f$  is constant. Functions with Hölder exponent  $s = 1$  are also called *Lipschitz functions*.

The *Hölder space*  $C^{n,s}(\Omega)$  consists of all  $n$ -times continuously differentiable and bounded functions  $f$  on  $\Omega$  for which the  $n$ -th derivative is Hölder continuous with exponent  $0 < s \leq 1$ . The linear space  $C^{n,s}(\Omega)$  becomes a Banach<sup>38</sup> space when endowed with the (inhomogeneous) norm

$$\|f\|_{C^{n,s}} := \sup_{\substack{|\nu| \leq n \\ x \in \Omega}} |D^\nu f(x)| + \sup_{\substack{|\nu|=n \\ x, x' \in \Omega \\ x \neq x'}} \frac{|D^\nu f(x) - D^\nu f(x')|}{|x - x'|^s}.$$

(Here  $\nu := (\nu_1, \dots, \nu_n)$  is a multi-index with length  $|\nu| := \nu_1 + \dots + \nu_n$ .) Note that  $C^{n,0}(\Omega)$  is isomorphic to the space  $C^n(\Omega)$  of all  $n$ -times continuously differentiable functions on  $\Omega$ . Setting  $C^{0,0}(\Omega) := C^0(\Omega)$ , we obtain a ladder of function spaces  $C^{n,s}(\Omega)$ :

$$C^n(\Omega) \subset C^{n,s}(\Omega) \subset C^{n,s'}(\Omega) \subset C^{n+1}(\Omega), \quad \text{for } 0 < s < s' \leq 1.$$

**Sobolev Spaces.** The collection of all  $q$ -times differentiable functions  $f$  on  $\Omega$  that are in  $L^p(\Omega)$  and whose derivatives  $D^\nu$ ,  $|\nu| \leq q$ , are also in  $L^p(\Omega)$  constitutes the *Sobolev space*  $W^{p,q}(\Omega)$ ,  $1 \leq p < \infty$ ;  $q \in \mathbb{Z}_0^+$ . Under the norm

$$\|f\|_{W^{p,q}(\Omega)} := \left( \sum_{|\nu| \leq q} \|D^\nu f\|_{L^p}^p \right)^{1/p}$$

the spaces  $W^{p,q}(\Omega)$  become Banach spaces, for  $p = 2$ , Hilbert<sup>39</sup> spaces. ( $1 \leq p < \infty$ ;  $q \in \mathbb{Z}_0^+$ ) We remark that  $W^{p,0}(\Omega)$  is isomorphic to  $L^p(\Omega)$ .

Let  $\mathcal{D}$  denote the linear space of all infinitely differentiable functions with compact support. Then we have the following inclusions.

$$\mathcal{D} \subset \dots \subset W^{p,q+1}(\Omega) \subset W^{2,p}(\Omega) \subset \dots \subset W^{p,0}(\Omega) = L^p(\Omega).$$

<sup>36</sup>Otto Ludwig Hölder, Dec 22, 1859 to Aug 29, 1937. German mathematician who contributed primarily to group theory.

<sup>37</sup>Rudolph Lipschitz, May 14, 1832 to Oct 7, 1903. German mathematician who worked in number theory, algebra, analysis, differential geometry, and classical mechanics.

<sup>38</sup>Stefan Banach, March 30, 1892 to Aug 31, 1945. Polish mathematician who is one of the founders of functional analysis. He also made contributions to vector spaces, measure theory, and set theory.

<sup>39</sup>David Hilbert, Jan 23, 1862 to Feb 14, 1943. German mathematician considered to be one of the outstanding researchers of the 20th century. He made contributions to invariant theory, the theory of algebraic numbers, functional analysis and axiomatized Euclidean geometry.

The Hilbert spaces  $H^q(\Omega) := W^{2,q}(\Omega)$  can also be defined for *real*  $q$  via the Fourier transform  $\mathcal{F}$ :

$$H^q(\Omega) := \left\{ f \in L^2(\Omega) : \|f\|_{H^q(\Omega)}^2 := \int_{\mathbb{R}^n} |\mathcal{F}(f)(\xi)|^2 (1 + |\xi|^2)^q d\xi < \infty \right\}.$$

## References

- [AL] A. Aldroubi and E. Lin (Eds.), *Wavelets, Multiwavelets, and Their Applications*, Contemporary Mathematics, Vol. 216, American Mathematical Society, Providence, 1998.
- [AHW] K. Attakitmongcol, D. Hardin and D. Wilkes, *Multiwavelet Prefilters II: Optimal orthogonal prefilters*, IEEE Trans. Image Proc. **10** (2001), 1476 – 1487.
- [B] A. Bossavit, *Computational Electrodynamics: Variational Formulations, Complementarity, Edge Elements*, Academic Press, San Diego, 1998.
- [Bz] R. Bozorth, *Ferrromagnetism*, Wiley – IEEE Press, Reissue Ed., New York, 1993.
- [BBS] P. L. Butzer, W. Splettstößer and R. L. Stens, *The sampling theorem and linear prediction in signal analysis*, Jber. d. Dt. Math.-Verein. **90** (1988), 1 – 70.
- [BW] K. Berkner and R. O. Wells, Jr., *Smoothness estimates for soft-thresholded denoising via invariant wavelet transforms*, Technical Report CML TR 98-01, Computational Mathematics Laboratory, Rice University, 1998.
- [C] C. Chui, *An Introduction to Wavelets*, Academic Press, San Diego, 1992.
- [CD] R. R. Coifman and D. L. Donoho, *Translation invariant denoising*, in A. Antoniades, editor, *Wavelets and Statistics*, Springer Lecture Notes, Springer Verlag, 1995.
- [CH] R. Courant and D. Hilbert, *Methods of Mathematical Physics*, Vol. I and II, Wiley - Interscience, New York, 1989.
- [D] I. Daubechies, *Ten Lectures on Wavelets*, CBMS, SIAM, Philadelphia, 1992.
- [dB] C. de Boor, *A Practical Guide to Splines*, revised ed., Springer Verlag, New York, 2001.
- [DGHM] G. Donavan, J. Geronimo, D. Hardin, and P. Massopust, *Construction of orthogonal wavelets using fractal interpolation functions*, SIAM J. Math. Anal. **27**(4) (1996), 1158 – 1192.
- [Do1] D. L. Donoho, *Denoising by soft thresholding*, IEEE Transactions on Information Theory, Vol. 41, No. 3 (1995), 613 – 627.
- [Do2] D. L. Donoho, *Interpolating wavelet transforms*, Technical Report, Department of Statistics, Stanford University, 1992.
- [DJ1] D. L. Donoho and I. M. Johnstone, *Ideal spatial adaption via wavelet shrinkage*, Technical Report, Department of Statistics, Stanford University, 1992.
- [DJ2] D. L. Donoho and I. M. Johnstone, *Adapting to unknown smoothness via wavelet shrinkage*, Technical Report, Department of Statistics, Stanford University, 1994.
- [DY] D. L. Donoho and T. P. - Y. Yu, *Nonlinear wavelet transform via median interpolation*, Technical Report, Department of Statistics, Stanford University, 1997.
- [F] M. Frazier, *An Introduction to Wavelets Through Linear Algebra*, Springer Verlag, New York, 1999.
- [G] A. Garcia, *Orthogonal sampling formulas: A unified approach*, SIAM Review, Vol. 42, No. 3 (2000), 499–512.
- [GHM] J. S. Geronimo, D. P. Hardin, and P. R. Massopust, *Fractal functions and wavelet expansions based upon several scaling functions*, J. Approx. Th. **78**(3) (1994), 373–401.
- [G] T. N. T. Goodman and S. L. Lee, *Wavelets of multiplicity r*, Trans. Amer. Math. Soc. **342**(1) (1994), 307–324.
- [H] T. Hohage, *Lecture Notes on Inverse Problems*, Vorlesungskript, University of Göttingen, Germany, 2002.
- [HR] D. P. Hardin and D. Roach, *Multiwavelet prefilters I: Orthogonal prefilters preserving approximation order  $p \leq 2$* , IEEE Trans. Circ. and Sys. II: Anal. and Dig. Sign. Proc. **45**(8) (1998), 1106 – 1112.
- [HSSTH] P. Heller, G. Strang, V. Strela, P. Topiwala and C. Heil, *Application of multiwavelets to signal compression and denoising*, Proc. UK Symposium on Applications of Time-Frequency and Time-Scale Methods, University of Warwick, 1995.
- [HXGS] D. P. Hardin X.-G. Xia, J. Geronimo and B. Suter, *Design of prefilters for discrete multiwavelet transforms*, IEEE Trans. on Signal Processing **44** (1996), 25 – 35.

- [I] V. Isakov, *Inverse Problems for Partial Differential Equations*, Springer Verlag, New York, 1998.
- [J] J. D. Jackson, *Classical Electrodynamics*, John Wiley & Sons, 3rd ed. New York, 1975.
- [K] F. Keinert, *Wavelets and Multiwavelets*, Studies in Advanced Mathematics, Chapman & Hall/CRC, New York, 2003.
- [Ke] J. B. Keller, *Inverse Problems*, Am. Math. Mon., **83** (1976), 107 – 118.
- [Ki] A. Kirsch, *An Introduction to the Mathematical Theory of Inverse Problems*, Springer Verlag, New York, 1996.
- [Ke] E. D. Kolaczyk, *Estimation of intensities of burst-like Poisson processes using Haar wavelets*, The Astrophysical Journal, Vol. 483 (1997), 340 – 349.
- [LL] L. Landau and E. Lifschitz, *Lehrbuch der Theoretischen Physik, Vol. XIII, Elektrodynamik der Kontinua*, Akademie Verlag, Berlin, Germany, 1985.
- [L] A. K. Louis, *Inverse und schlechtgestellte Probleme*, Teubner Verlag, Stuttgart, Germany, 1989.
- [LLR] M. R. Leadbetter, G. Lindgren, and G. Rootzen, *Extremes and Related Properties of Random Sequences and Processes*, Springer Verlag, New York, 1983.
- [LV] J. Lebrun and M. Vetterli, *Higher order balanced multiwavelets*, IEEE ICASSP, 1998.
- [M] P. Massopust, *Fractal Functions, Fractal Surfaces, and Wavelets*, Academic Press, New York, 1995.
- [P] G. Plonka, *Approximation order provided by refinable function vectors*, Constr. Approx. **13** (1997), 221 – 244.
- [R] T. Radó, *Über den Begriff der Riemannschen Fläche*, Acta. Litt. Sci. Szeged. **2** (1925), 101 – 121.
- [RW] H. L. Resnikoff and R. O. Wells, Jr, *Wavelet Analysis and Scalable Structure of Information*, Springer-Verlag, New York, 1998.
- [S] I. Selesnik, *Multiwavelet bases with extra approximation properties*, Technical Report, Department of Electrical and Computer Engineering, Rice University, 1997.
- [Sc] B. Schutz, *Geometrical Methods of Mathematical Physics*, Cambridge University Press, UK, 1980.
- [Sh] C. E. Shannon, *A mathematical theory of communication*, The Bell System Technical Journal **27** (1948), 379 – 423, 623 – 656.
- [SB] J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis*, Texts in Applied Mathematics, Vol. 12, 2nd ed., Springer Verlag, New York, 1993.
- [SN] G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, 1996.
- [T] H. Triebel, *Theory of Function Spaces II*, Monographs in Mathematics, Vol. 84, Springer Verlag, 1992.
- [V] A. Visintin, *Differential Models of Hysteresis*, Springer Verlag, New York, 1995.
- [W] W. Walter, *Potentialtheorie*, BI Hochschulskripten, 765/765a, Bibliographisches Institut, Mannheim, 1971.
- [WSA] K. Warnick, R. Selfridge and D. Arnold, *Teaching electromagnetic field theory using differential forms*, IEEE Trans. on Education, Vol. 40, No.1 (1997), 53 – 68.

ENGINEERING RESEARCH AND DEVELOPMENT, TUBOSCOPE PIPELINE SERVICES, HOUSTON,  
TEXAS 77051

*E-mail address:* pmassopust@varco.com

## Robust Interferometric Imaging in Random Media

Liliana Borcea

**ABSTRACT.** We review a coherent interferometric approach to imaging in random media, developed in [15, 14, 13]. We discuss the resolution of this method and compare it with deterministic ones, such as Kirchhoff migration. Coherent interferometry is introduced and analyzed in [15, 13, 24] for isotropic random media, and in [14] for randomly layered media. For didactic purposes, we describe in more detail the latter case. Explicitly, we review wave propagation in finely layered media, as developed in [55, 22, 21, 57, 1, 27, 66, 41, 40] and then, we analyze coherent interferometry in such media, as introduced in [14].

### 1. Introduction and Overview

In an important class of inverse problems for the wave equation, we seek to identify reflecting surfaces or compactly supported sources or reflectors, from waves recorded at a remote array of receivers. This is a very active field, covering a variety of topics such as: (a) reflection seismology [26, 71, 62, 7, 8, 54]; (b) synthetic aperture radar imaging [30, 18, 25, 53, 36, 35]; (c) interferometric radar imaging [73, 52, 51, 43, 64]; (d) inverse scattering of time harmonic acoustic or electromagnetic waves [29]. The literature is vast and the list of references mentioned here is by no means exhaustive.

The common theme of most available results is that imaging occurs in a smooth background that is either known or it can be estimated. However, in applications such as geophysical wave propagation and ground or foliage penetrating radar, besides the smooth, large scale variations of the background, we have small scale fluctuations, as shown in Figures 1 and 2. Estimating the smoothly varying part of the medium is known as *velocity analysis* and it can be done, for example, with the differential semblance approach developed in [23, 70, 44, 69, 68]. Naturally, we cannot hope to know the fast fluctuations in a precise manner, but we can get an idea about their scale and magnitude from direct measurements that can be made, for example, in wells drilled in the earth. This information can then be incorporated in the modeling of our uncertainty of the medium as a random process.

The inverse problems that we consider seek to determine the location of sources or strong scatterers buried in clutter, as shown in Figure 2, from the recorded

---

1991 *Mathematics Subject Classification*. Primary 86A22, 60H30; Secondary 35L05.

*Key words and phrases*. random media, imaging, interferometry.

The author was supported in part by NSF Grants DMS-0305056, DMS 0354658 and by the ONR grants N00014-02-1-0088, N00014-05-10699.

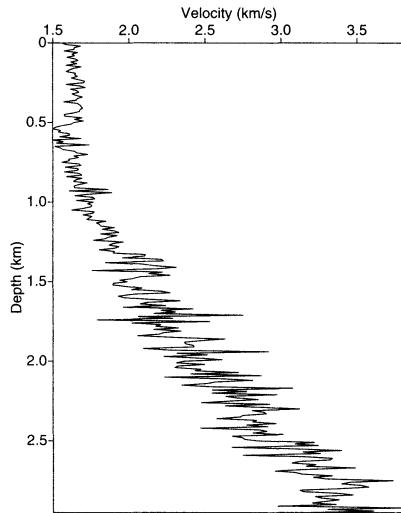


FIGURE 1. Example of an acoustic velocity profile measured in a well drilled in the earth.

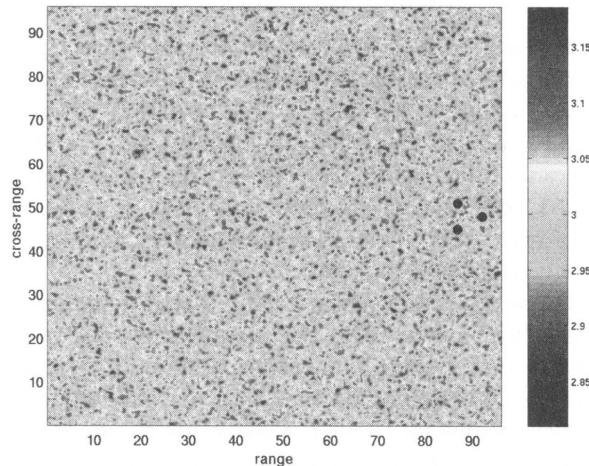


FIGURE 2. An illustration of an isotropic cluttered medium containing three sources or scatterers shown with black dots. The velocity of the waves in the clutter varies randomly about the constant value of 3km/s. The axes stand for range (distance from the array that is supposed planar, in the direction normal to it) and cross-range (coordinate in the direction orthogonal to the range).

signals at a remote array. We are particularly interested in regimes where there is significant multipathing of the waves, by the inhomogeneities in clutter. In this case, the coherent signals coming from the sources or reflectors that we wish to image are typically weak and the data shows significant, incoherent fluctuations, due to multiple scattering of the waves by the clutter. The incoherent fluctuations in the data, that can be spread out over a long time interval, are referred to as delay spread or coda and they impede the imaging process. The challenge is then to come up with inversion methods that give reliable images with such noisy data, without being certain of the clutter that we model as a random process. Explicitly, we seek statistically stable (self-averaging) imaging functions that work well regardless of the realization of the clutter.

The analysis and computation of wave propagation in random media raises many difficult and interesting mathematical questions. It is beyond the scope of this paper to review the literature on wave propagation in random media, but we refer the reader to a few important references [47, 48, 45, 55, 22, 21, 57, 1, 6, 27, 66, 46].

Recently, there has been an outburst of interesting theoretical and experimental developments in *time reversal* in random media, where signals emitted by a source are recorded at an array of receivers, time reversed and sent back in the same medium. If there is no dissipation, the time reversibility of the wave equation implies diffraction limited refocusing at the source. This is easily seen in a homogeneous medium, where waves of carrier wavelength  $\lambda_0$ , captured on an array of aperture  $a$ , refocus near the source that is at distance  $L$ , with a spot size  $\lambda_0 L/a$  [19]. Intuitively, we expect that in random media, the refocusing should be much worse, but amazingly, it is significantly better. This has been demonstrated experimentally in [37, 38, 31, 39, 50, 67] and it has been proved rigorously, at least for some scaling regimes, in the theoretical and computational studies [28, 10, 72, 58, 3, 4, 41, 65, 40]. The improved refocusing of time reversed waves in random media is referred to as *super resolution* and it is due to random inhomogeneities which distribute the waves over a larger part of the medium, as if they came from a wider array, of large effective aperture.

Super-resolution in the mean, for ensemble averaged time reversed fields, has been studied in [32, 33], but this does not explain why it is observed in all the experiments [37, 38, 31, 39, 50, 67]. It is the work in [10] that shows that, for super-resolution to be robustly observed, the time-reversed field must be *self-averaging* (i.e. it should be equal to its statistical expectation and not depend on the realization of the clutter). The question of when self-averaging occurs is addressed in [28, 10, 58, 3, 4, 41, 65, 40]. These studies consider various scaling regimes and media, but their general conclusion is that self-averaging is a broad band phenomenon.

It is common knowledge that any array imaging method involves some form of time reversal, back propagation or migration to search locations, where we believe that our targets lie. As a result, one might be misled into thinking that imaging and time reversal are the same thing. In fact, there is a major difference between the two. The back propagation in imaging is done numerically, in a fictitious medium, that is our best guess of the background, whereas in time reversal, the waves are re-emitted into the real medium. In the presence of fast fluctuations (clutter) in

the environment, the fictitious medium used in imaging lacks the detailed information about the inhomogeneities and, consequently, the desired properties of super-resolution and self-averaging do not occur. This is why methods such as synthetic aperture imaging [30, 18, 25] and Kirchhoff migration [26, 9] that work well in known media, fail to image reliably in clutter.

The question is then how to develop imaging methods that are robust in the sense that they are insensitive to our uncertainty of the clutter. In particular, is there a way of taking advantage of the super-resolution and stability properties of time reversal in our design of imaging algorithms? It is shown in [15, 12, 11, 16, 17] that, to make wave imaging work in clutter, one should do *interferometric migration*, rather than direct migration as usual. The coherent interferometric approach developed in [15, 13, 14, 24] gives good, robust images in clutter, by exploiting the broadband nature of the signal and the desirable properties of time reversal. The latter is because the mathematical expressions that are involved in coherent interferometry are specially constructed to resemble those of time reversal.

Coherent interferometry works in a variety of regimes, with significant multipathing of the waves, by the clutter. For example, the method was originally developed in [15, 13] for imaging in isotropic, weakly fluctuating media, where most scattering occurs in the forward direction. Later, the method was extended in [14], to randomly layered media that create significant back scattering and long and noisy cudas of the traces received at the array. Although these regimes are significantly different, coherent interferometry works equally well, which is an indication of its robustness.

Coherent interferometry can be viewed as differential Kirchhoff migration on the lag of the interferograms, which are correlations of traces computed over specific time windows. The idea is that for appropriate time segments and separations between the receivers at which we correlate traces, we achieve random phase cancellations that lead to self-averaging. Coherent interferometry can also be viewed (see [13]) as a nontrivial extension of classic statistical smoothing techniques used in the spectral estimation of random time series [61]. Roughly said, by segmenting the time and space, cross-correlating our recorded data over these segments (windows) and then migrating it, we obtain approximately the convolution of the Fourier transform of the windows with the Kirchhoff migration function (actually the square of it). The Kirchhoff imaging function is randomly fluctuating but, due to the convolution with the windows, we obtain a stable result.

It is shown in [15, 13, 14] that there is a delicate trade-off between having enough smoothing and achieving good resolution of the images. If we smooth too much, by for example taking very long time segments in our correlations, our images will be stable but very blurry. On the other hand, if we do not smooth enough, by taking for example too short time segments, the image is not stable. There is therefore an optimal choice of the smoothing parameters that depends, naturally, on the random medium. These parameters can be estimated in principle from the recorded data at the array, which is noisy. However, the smoothing parameters are intrinsically related with the quality of the image, which is stable, so it is better to estimate them by looking at the image itself. This is done in [13] (see also [12]).

This paper is organized as follows: In Section 2 we formulate the inverse problem for imaging sources in a cluttered medium. In Section 3 we describe briefly the Kirchhoff migration imaging approach and explain why it fails in random

media. In Section 4 we introduce the coherent interferometric imaging function and describe qualitatively its resolution. For didactic purposes, we give some details of the analysis of the method in finely layered random media, for the case of a single point source. We begin in Section 5 with the forward model for the acoustic pressure waves recorded at the array and originating from a point source buried in a randomly layered medium. Here, we describe the scaling and the machinery used in the calculation of wave fields transmitted through randomly layered media. The techniques are well established and have been developed in [55, 22, 21, 57, 1, 27, 66, 41, 40], but we review them for the convenience of the reader. In Section 6 we describe the statistics of the transmitted acoustic pressure through the random medium, in the asymptotic regime of high frequency waves and long distances of propagation, implied by the scaling of Section 5. Imaging in layered random media is discussed in Section 7.

## 2. Formulation of the problem

Our forward model is based on the linear acoustic equations, which are appropriate for the description of compressional (P) waves and neglect entirely the shear (S) waves. The linear equations of acoustics are

$$(2.1) \quad \begin{aligned} \rho(\vec{x}) \frac{\partial \vec{u}}{\partial t}(t, \vec{x}) + \nabla p(t, \vec{x}) &= \vec{0}, \\ v^{-2}(\vec{x}) \frac{\partial p}{\partial t}(t, \vec{x}) + \rho(\vec{x}) \nabla \cdot \vec{u}(t, \vec{x}) &= 0, \end{aligned}$$

for  $\vec{x} \in \mathbb{R}^3 \setminus \mathcal{D}$ , and  $t > 0$ , with initial conditions

$$(2.2) \quad \vec{u}(t, \vec{x}) = \vec{0}, \quad p(t, \vec{x}) = 0, \quad t < 0.$$

Here,  $\vec{u}$  is the velocity of the material particle located at  $\vec{x}$ ,  $p$  is the acoustic pressure,  $v$  is the wave speed,  $\rho$  is the mass density and  $\mathcal{D}$  is the support of the sources that emit signals which propagate through the medium and are captured at the array. For simplicity, we suppose that the same pulse

$$(2.3) \quad f(t) = e^{-i\omega_o t} f_B(t)$$

is emitted simultaneously, by all sources, where  $B$  is the bandwidth and  $\omega_o$  is the carrier frequency. The sources could be a collection of points, at locations  $\vec{y}_j$ ,  $j = 1, \dots, N_S$ , or they can be continuously distributed in  $\mathcal{D}$ , in which case  $f(t)$  may be emitted with different intensities  $w(\vec{y})$ , from  $\vec{y} \in \mathcal{D}$ .

We denote by  $\vec{x}_r \in \mathcal{A}$ ,  $r = 1, \dots, N$ , the locations of the  $N$  receivers in the array  $\mathcal{A}$  and we suppose that we record the traces

$$(2.4) \quad p(t, \vec{x}_r), \quad \text{for } t \in [t_1, t_2] \quad \text{and } r = 1, \dots, N,$$

for a fixed time window  $[t_1, t_2]$ . See Figure 3 for the traces recorded in the configuration of Figure 2, at an array of  $N = 185$  transducers, placed equidistantly at range zero, in the whole cross-range interval. The top picture corresponds to the traces recorded in a homogeneous medium, with speed of sound of 3km/s and the bottom picture corresponds to the cluttered medium shown in Figure 2. The fluctuations of  $v(\vec{x})$  are modeled using a random Fourier series with mean 3km/s and Gaussian correlation function. The correlation length is 1.5m, and the standard deviation is 3%. Finally, the pulse is the time derivative of a Gaussian, with central frequency  $2\pi\omega_0 = 1\text{kHz}$  and bandwidth 0.6 – 1.3kHz, measured at 6dB. For more details of the numerical simulations, see [15, 13].

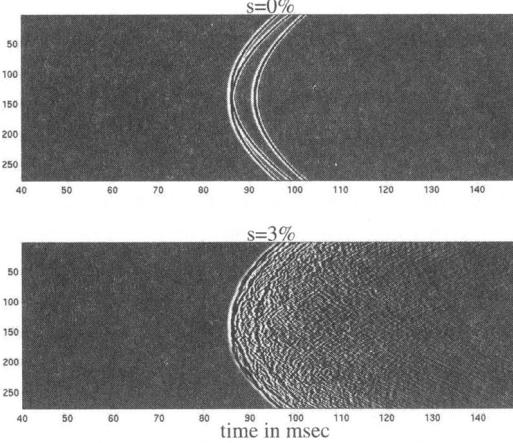


FIGURE 3. Time traces recorded on the array, due to the sources shown as black dots in Figure 2. Top: homogeneous medium. Bottom: random medium. The horizontal axis is time and the vertical is array transducer location.

The inverse problem is to determine the locations  $\vec{y}_j$ ,  $j = 1, \dots, N_S$ , of the sources, or the distributed source intensity  $w(\vec{y})$  (i.e. its support  $\mathcal{D}$ ) from the recorded traces (2.4) at the array.

### 3. Kirchhoff migration

Kirchhoff migration [26, 63, 9, 7, 71] is widely used in geophysics inversion and it is very successful in imaging the support of sources or scatterers in *known* and *smooth* backgrounds. We review briefly this method for the case of homogeneous backgrounds, with speed of sound  $v(\vec{x}) = c$  and for isotropic, distributed sources that emit pulse  $f(t)$  with intensity  $w(\vec{y})$  compactly supported in  $\mathcal{D}$ . Then, we explain why Kirchhoff migration does not work in clutter.

**3.1. Kirchhoff migration in homogeneous media.** Letting  $G_o(\vec{x}_r, \vec{y}, t)$  be the Green's function of the acoustic wave equation in the homogeneous medium, we have the forward model

$$(3.1) \quad p(t, \vec{x}_r) = \int_{\mathcal{D}} w(\vec{y}) e^{-i\omega_o t} f_B(t) * G_o(t, \vec{x}_r, \vec{y}) d\vec{y}, \quad r = 1, \dots, N,$$

for the pressure recorded at the array, where  $*$  denotes convolution in time. In the Fourier domain,

$$(3.2) \quad \hat{p}(\omega, \vec{x}_r) = \hat{f}_B(\omega - \omega_o) \int_{\mathcal{D}} w(\vec{y}) \hat{G}_o(\omega, \vec{x}_r, \vec{y}) d\vec{y},$$

where

$$(3.3) \quad \hat{G}_o(\omega, \vec{x}_r, \vec{y}) = \frac{1}{4\pi|\vec{x}_r - \vec{y}|} e^{i\omega\tau(\vec{x}_r, \vec{y})}$$

and

$$(3.4) \quad \tau(\vec{x}_r, \vec{y}) = |\vec{x}_r - \vec{y}|/c$$

is the travel time.

The Kirchoff migration function is

$$(3.5) \quad \mathcal{I}^{\text{KM}}(\vec{y}^s) = \sum_{\vec{x}_r \in \mathcal{A}} p(\tau(\vec{x}_r, \vec{y}^s), \vec{x}_r),$$

for a search point  $\vec{y}^s$  to which we migrate  $p(\vec{x}_r, t)$  by evaluating it at the travel time  $\tau(\vec{x}_r, \vec{y}^s)$ . Assuming a dense array of transducers, we can replace the sum in (3.5) by an integral over the surface of the array parametrized by  $\mathbf{r} \in \mathbb{R}^2$  and write

$$(3.6) \quad \mathcal{I}^{\text{KM}}(\vec{y}^s) \sim \int_{\mathcal{A}} d\mathbf{r} \int_{\mathcal{D}} d\vec{y} w(\vec{y}) \int_{|\omega - \omega_o| \leq B} d\omega \frac{\hat{f}_B(\omega - \omega_o)}{|\vec{x}(\mathbf{r}) - \vec{y}|} e^{i\omega[\tau(\vec{x}(\mathbf{r}), \vec{y}) - \tau(\vec{x}(\mathbf{r}), \vec{y}^s)]},$$

where symbol  $\sim$  stands for approximate, up to a multiplicative constant.

The underlying assumption in Kirchhoff migration is the existence of a length scale separation [9, 71]. We have a long scale over which the smooth (constant in our case) medium varies and a short scale, given by a typical wavelength  $\lambda_o = 2\pi c/\omega_o$ . Thus, we are in a high frequency asymptotic regime and we can approximate (3.6) with the method of stationary phase [5], which gives that the leading order term of the  $\omega$  and  $\mathbf{r}$  integrals comes from the vicinity of the stationary points satisfying

$$(3.7) \quad \begin{aligned} |\vec{x}(\mathbf{r}) - \vec{y}| - |\vec{x}(\mathbf{r}) - \vec{y}^s| &= 0, \\ \nabla_{\mathbf{r}}(|\vec{x}(\mathbf{r}) - \vec{y}| - |\vec{x}(\mathbf{r}) - \vec{y}^s|) &= \mathbf{0}. \end{aligned}$$

These equations hold, without any restriction on  $\mathbf{r}$  and  $\omega$ , if  $\vec{y} = \vec{y}^s$ , so only the vicinity of the search point  $\vec{y}^s$  counts. The stationary phase calculation is carried out in detail in [71] and it is shown there that, in the ideal situation of an infinite bandwidth and aperture, the Kirchhoff migration imaging function is given by the action of a pseudo-differential operator on the unknown intensity function  $w(\vec{y})$ . This implies that we can image very well the singularities of  $w(\vec{y})$ , i.e. its support  $\mathcal{D}$ . In the simple setups considered in this paper, the sources are small (pointlike) so we actually image their location.

When the bandwidth and the array are finite, we can still image the source, as shown by the stationary phase result (3.7), but at the expense of some loss in resolution. To see this, let us recall that  $\mathcal{I}^{\text{KM}}$  peaks for  $\vec{y}^s$  near the source, and approximate the phase in (3.6) by

$$(3.8) \quad \omega [\tau(\vec{x}(\mathbf{r}), \vec{y}) - \tau(\vec{x}(\mathbf{r}), \vec{y}^s)] \approx \zeta \cdot (\vec{y} - \vec{y}^s),$$

where

$$(3.9) \quad \zeta = \omega \nabla_{\vec{y}} \tau(\vec{x}(\mathbf{r}), \vec{y})|_{\vec{y}=\vec{y}^s}.$$

In the high frequency regime considered here, the amplitudes play a negligible role in inversion, so we can modify  $\mathcal{I}^{\text{KM}}$  for the sake of the argument as

$$(3.10) \quad \begin{aligned} \mathcal{I}^{\tilde{\text{KM}}}(\vec{y}^s) &\sim \int_{\mathcal{A}} d\mathbf{r} \mathcal{M}(\mathbf{r}, \vec{y}^s) \int_{|\omega - \omega_o| \leq B} d\omega \hat{p}(\tau(\vec{x}(\mathbf{r}), \vec{y}^s), \vec{x}(\mathbf{r})) \\ &= \int \frac{\hat{f}_B(\omega(\zeta) - \omega_0)}{\omega^2(\zeta)} e^{i\zeta \cdot (\vec{y} - \vec{y}^s)} d\zeta \end{aligned}$$

with

$$(3.11) \quad \omega(\zeta) = \frac{\zeta \cdot \nabla_{\vec{y}} \tau(\vec{x}(\mathbf{r}) - \vec{y})|_{\vec{y}=\vec{y}^s}}{\left| \nabla_{\vec{y}} \tau(\vec{x}(\mathbf{r}) - \vec{y})|_{\vec{y}=\vec{y}^s} \right|^2}$$

and amplitude factor  $\mathcal{M}$  chosen to absorb the determinant of the Jacobian arising in change of variables (3.9).

The domain of integration in (3.10) depends on the bandwidth and aperture in a complicated way and it determines the resolution of the image. We can obtain a qualitative idea about the resolution limits by considering a simple set-up, where the array is planar, of aperture  $a$  that is much smaller than the range of the source. Using the notation  $\vec{y} = (y, z)$  that distinguishes between the range  $z$  and cross-range  $y \in \mathbb{R}^2$ , we let  $\vec{x}(\mathbf{r}) = (\mathbf{x}(\mathbf{r}), 0) \in \mathcal{A}$  and obtain

$$(3.12) \quad \begin{aligned} \zeta \cdot (\vec{y} - \vec{y}^s) &= \frac{\omega}{c} \left( \frac{\mathbf{x}(\mathbf{r}) - \mathbf{x}^s}{|\vec{x}(\mathbf{r}) - \mathbf{x}^s|}, \frac{z^s}{|\vec{x}(\mathbf{r}) - \mathbf{x}^s|} \right) \cdot (\mathbf{y} - \mathbf{y}^s, z - z^s) \\ &\approx \frac{\omega}{c} \frac{(\mathbf{x}(\mathbf{r}) - \mathbf{x}^s)}{z^s} \cdot (\mathbf{y} - \mathbf{y}^s) + \frac{\omega}{c} (z - z^s). \end{aligned}$$

This shows that the component of  $\zeta$  in the range direction is essentially  $\omega/c$  and it is restricted to the interval  $\frac{1}{c}(\omega_o - B, \omega_o + B)$ , which makes the range resolution

$$(3.13) \quad |z - z^s| \leq O\left(\frac{c}{B}\right).$$

Finally, the component of  $\zeta$  in cross-range belongs to a square domain of side  $\frac{aB}{cz^s}$ , so the cross-range resolution is

$$(3.14) \quad |\mathbf{x} - \mathbf{x}^s| \leq O\left(\frac{cz^s}{Ba}\right).$$

For more details on the resolution of Kirchhoff migration images we point the reader to [71, 9, 7].

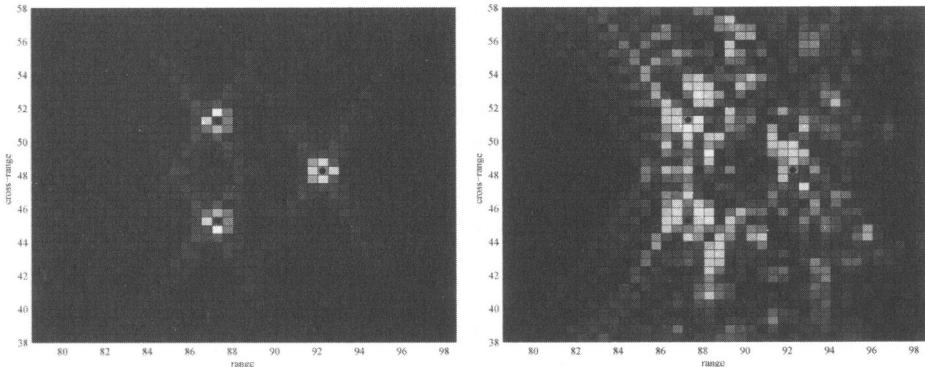


FIGURE 4. Kirchhoff migration images of the sources in Figure 2. Left: image in a homogeneous background. Right: image in clutter. The true source locations are indicated correctly in the picture on the left.

**3.2. Kirchhoff migration in clutter.** When imaging in clutter, Kirchhoff migration does not work well because we do not know the medium in detail, just

its smooth part that we assume for simplicity as uniform. To image, we migrate traces (2.4) in the fictitious uniform medium, to a search point  $\vec{y}^s$ , by evaluating  $p(t, \vec{x}_r)$  at travel time (3.4). Since the traces are noisy (random), the result is an unstable image that changes unpredictably with the realization of the clutter. This is illustrated in Figure 4, where we plot the Kirchhoff migration image of the sources in Figure 2, when the background is homogeneous and cluttered, respectively. In the latter case, the image is noisy and the identification of the sources is not reliable. Moreover, the picture changes with the realization of the clutter, as shown in [15, 13, 12].

We shall see in the next section how to stabilize the results, by cross-correlating traces before migrating them. These cross-correlations lead to random phase cancellations in the imaging function that becomes therefore self-averaging.

#### 4. Imaging in clutter

**4.1. Time reversal.** If we knew the random medium between the array and the sources, we could migrate the traces to the search point by means of the random Green's function  $\widehat{G}(\omega, \vec{x}_r, \vec{y}^s)$ . The result would be

$$(4.1) \quad \mathcal{I}^{\text{TR}}(\vec{y}^s) = \sum_{\vec{x}_r \in \mathcal{A}} \int_{|\omega - \omega_0| \leq B} d\omega \widehat{p}(\omega, \vec{x}_r) \overline{\widehat{G}(\omega, \vec{x}_r, \vec{y}^s)},$$

where the bar indicates complex conjugation. Since (4.1) represents the field at  $\vec{y}^s$ , when the array time-reverses the signals received and re-emits them in the random medium, it is called the time reversal functional.

$\mathcal{I}^{\text{TR}}$  is the ideal imaging function for two reasons: (1) it is usually self-averaging, especially in broad-band regimes [28, 10, 58, 3, 4, 41, 65, 40] and (2) it focuses much better in clutter [37, 38, 31, 39, 67, 10, 3, 4, 15]. The latter property is called super-resolution and it is due to the random inhomogeneities which distribute the waves over a larger part of the medium, as if they came from a wider array, of effective aperture  $a_e \gg a$ . Then, instead of (3.14), we have

$$(4.2) \quad |\mathbf{x} - \mathbf{x}^s| \leq O\left(\frac{cz^s}{Ba_e}\right) \ll O\left(\frac{cz^s}{Ba}\right),$$

so we get very sharp images even with small arrays. The self-averaging of  $\mathcal{I}^{\text{TR}}$  is due to significant phase cancellation occurring in the products  $\widehat{G}\overline{\widehat{G}}$  in

$$(4.3) \quad \widehat{p}(\omega, \vec{x}_r) \overline{\widehat{G}(\omega, \vec{x}_r, \vec{y}^s)} = \widehat{f}_B(\omega - \omega_o) \int_{\mathcal{D}} d\vec{y} w(\vec{y}) \widehat{G}(\omega, \vec{x}_r, \vec{y}) \overline{\widehat{G}(\omega, \vec{x}_r, \vec{y}^s)},$$

for  $\vec{y}^s$  near (or in)  $\mathcal{D}$ , and to the averaging of the small remaining phases over the bandwidth.

**4.2. Interferometric imaging.** Of course, the problem with  $\mathcal{I}^{\text{TR}}$  is that we do not know the clutter (i.e.  $\widehat{G}(\omega, \vec{x}_r, \vec{y}^s)$ ) in detail. As a matter of fact, it is because we do not know the clutter that we model it as a random process. Nevertheless, we can use  $\mathcal{I}^{\text{TR}}$  as a guide in creating imaging functions that allow random phase cancellations. For example, we can cross-correlate the data traces to create the so

called interferograms which we then migrate to  $\vec{\mathbf{y}}^s$ , as in

$$(4.4) \quad \mathcal{I}^{\text{MF}}(\vec{\mathbf{y}}^s) = \int d\omega \sum_{\vec{\mathbf{x}}_r \in \mathcal{A}} \sum_{\substack{\vec{\mathbf{x}}_{r'} \in \mathcal{A} \\ |\vec{\mathbf{x}}_r - \vec{\mathbf{x}}_{r'}| \leq X_d}} \widehat{p}(\vec{\mathbf{x}}_r, \omega) \overline{\widehat{p}(\vec{\mathbf{x}}_{r'}, \omega)} e^{-i\omega[\tau(\vec{\mathbf{x}}_r, \vec{\mathbf{y}}^s) - \tau(\vec{\mathbf{x}}_{r'}, \vec{\mathbf{y}}^s)]},$$

where  $X_d$  is the decoherence length that is typically much smaller than  $a$ . This is the matched field function [20, 2, 49, 34] that is stable, as shown in [12, 16, 17], but has no range resolution because we take cross-correlations over all time, so that there is no absolute arrival time information. Matched field works with differential arrival times

$$(4.5) \quad \tau(\vec{\mathbf{x}}_r, \vec{\mathbf{y}}^s) - \tau(\vec{\mathbf{x}}_{r'}, \vec{\mathbf{y}}^s) \approx \nabla_{\vec{\mathbf{x}}_r} \tau(\vec{\mathbf{x}}_r, \vec{\mathbf{y}}^s) \cdot (\vec{\mathbf{x}}_r - \vec{\mathbf{x}}_{r'})$$

which means that in the usual situation of decoherence length  $X_d$  much smaller than the distance of propagation, we can estimate just the cross-range of the source. See [12] for a detailed analysis of matched field in clutter.

**4.3. Coherent interferometry.** Recently, it has been shown in [15, 13, 14, 24] that both the range and direction of arrival can be robustly estimated with a new, coherent interferometric imaging approach. The key idea in coherent interferometry is that, to obtain good images in clutter, we should exploit the spatial and frequency coherence in the data recorded at the array. Explicitly, we construct

$$(4.6) \quad \begin{aligned} \mathcal{I}^{\text{CINT}}(\vec{\mathbf{y}}^s) = & \int_{|\omega - \omega_o| \leq B} d\omega \int_{\substack{|\omega' - \omega_o| \leq B \\ |\omega - \omega'| \leq \Omega_d}} d\omega' \sum_{\vec{\mathbf{x}}_r \in \mathcal{A}} \sum_{\substack{\vec{\mathbf{x}}_{r'} \in \mathcal{A} \\ |\vec{\mathbf{x}}_r - \vec{\mathbf{x}}_{r'}| \leq X_d}} \widehat{p}(\omega, \vec{\mathbf{x}}_r) \\ & \times \overline{\widehat{p}(\omega', \vec{\mathbf{x}}_{r'})} \exp \{-i[\omega \tau(\vec{\mathbf{x}}_r, \vec{\mathbf{y}}^s) - \omega' \tau(\vec{\mathbf{x}}_{r'}, \vec{\mathbf{y}}^s)]\}, \end{aligned}$$

where  $\Omega_d$  is the decoherence frequency, that is inverse proportional to the delay spread  $T_d$  in the traces.

Let us suppose for simplicity that our array is planar and set  $\vec{\mathbf{x}}_r = (\mathbf{x}_r, 0)$ , for all  $r = 1, \dots, N$ . Simplify the notation in the travel times and traces as

$$\tau((\mathbf{x}_r, 0), \vec{\mathbf{y}}^s) \sim \tau(\mathbf{x}_r, \vec{\mathbf{y}}^s), \quad p(t, (\mathbf{x}_r, 0)) \sim p(t, \mathbf{x}_r)$$

and introduce the midpoint and offset variables

$$(4.7) \quad \mathbf{x}_r = \bar{\mathbf{x}} - \frac{\tilde{\mathbf{x}}}{2}, \quad \mathbf{x}_{r'} = \bar{\mathbf{x}} + \frac{\tilde{\mathbf{x}}}{2}$$

as well as the center frequency and difference frequency variables

$$(4.8) \quad \omega = \bar{\omega} - \frac{\tilde{\omega}}{2}, \quad \omega' = \bar{\omega} + \frac{\tilde{\omega}}{2}.$$

The coherent interferometric function becomes

$$(4.9) \quad \begin{aligned} \mathcal{I}^{\text{CINT}}(\vec{\mathbf{y}}^s) = & \int_{|\bar{\omega} - \omega_o| \leq B} d\bar{\omega} \int_{-\Omega_d}^{\Omega_d} d\tilde{\omega} \sum_{\tilde{\mathbf{x}} \in \mathcal{A}} \sum_{|\tilde{\mathbf{x}}| \leq X_d} \widehat{p}\left(\bar{\omega} - \frac{\tilde{\omega}}{2}, \bar{\mathbf{x}} - \frac{\tilde{\mathbf{x}}}{2}\right) \\ & \times \overline{\widehat{p}\left(\bar{\omega} + \frac{\tilde{\omega}}{2}, \bar{\mathbf{x}} + \frac{\tilde{\mathbf{x}}}{2}\right)} \exp \{i\bar{\omega} [\tau\left(\bar{\mathbf{x}} + \frac{\tilde{\mathbf{x}}}{2}, \vec{\mathbf{y}}^s\right) - \tau\left(\bar{\mathbf{x}} - \frac{\tilde{\mathbf{x}}}{2}, \vec{\mathbf{y}}^s\right)]\} \\ & \times \exp \left\{ i\frac{\tilde{\omega}}{2} [\tau\left(\bar{\mathbf{x}} + \frac{\tilde{\mathbf{x}}}{2}, \vec{\mathbf{y}}^s\right) + \tau\left(\bar{\mathbf{x}} - \frac{\tilde{\mathbf{x}}}{2}, \vec{\mathbf{y}}^s\right)] \right\}, \end{aligned}$$

and, under the assumption that  $X_d$  is small<sup>1</sup>, we have

$$(4.10) \quad \begin{aligned} \tau(\bar{\mathbf{x}} + \frac{\tilde{\mathbf{x}}}{2}, \vec{\mathbf{y}}^s) - \tau(\bar{\mathbf{x}} - \frac{\tilde{\mathbf{x}}}{2}, \vec{\mathbf{y}}^s) &\approx \nabla_{\bar{\mathbf{x}}} \tau(\bar{\mathbf{x}}, \vec{\mathbf{y}}^s) \cdot \tilde{\mathbf{x}}, \\ \frac{1}{2}[\tau(\bar{\mathbf{x}} - \frac{\tilde{\mathbf{x}}}{2}, \vec{\mathbf{y}}^s) + \tau(\bar{\mathbf{x}} + \frac{\tilde{\mathbf{x}}}{2}, \vec{\mathbf{y}}^s)] &\approx \tau(\bar{\mathbf{x}}, \vec{\mathbf{y}}^s). \end{aligned}$$

Finally, we can write the time domain analog of (4.9) (see [15])

$$(4.11) \quad \begin{aligned} \mathcal{I}^{\text{CINT}}(\vec{\mathbf{y}}^s) &\sim \sum_{\bar{\mathbf{x}} \in A} \sum_{|\tilde{\mathbf{x}}| \leq X_d} \int_{|\tilde{t} - \tau(\bar{\mathbf{x}}, \vec{\mathbf{y}}^s)| \leq T_d} d\tilde{t} \int_{|\tilde{t} - \nabla_{\bar{\mathbf{x}}} \tau(\bar{\mathbf{x}}, \vec{\mathbf{y}}^s) \cdot \tilde{\mathbf{x}}| \leq T_p} d\tilde{t} \\ &\quad p(\tilde{t} - \frac{\tilde{t}}{2}, \bar{\mathbf{x}} - \frac{\tilde{\mathbf{x}}}{2}) p(\tilde{t} + \frac{\tilde{t}}{2}, \bar{\mathbf{x}} + \frac{\tilde{\mathbf{x}}}{2}), \end{aligned}$$

where  $T_d = \pi/\Omega_d$  is the decoherence time and  $T_p = \pi/B$  is the pulse width.

Now, to analyze (4.9) or its time domain analog (4.11), we need to estimate the moments of  $\mathcal{I}^{\text{CINT}}$ . While  $p(t, \bar{\mathbf{x}})$  is random, in  $\mathcal{I}^{\text{CINT}}$  we smooth it by averaging over the decoherence length and frequency, the bandwidth and the array, and the result is self-averaging, as discussed in [15, 13]. Thus,  $\mathcal{I}^{\text{CINT}}$  is essentially equal to its expectation  $E\{\mathcal{I}^{\text{CINT}}\}$ . The expression of

$$\begin{aligned} E\left\{\hat{p}\left(\bar{\omega} - \frac{\tilde{\omega}}{2}, \bar{\mathbf{x}} - \frac{\tilde{\mathbf{x}}}{2}\right) \hat{p}\left(\bar{\omega} + \frac{\tilde{\omega}}{2}, \bar{\mathbf{x}} + \frac{\tilde{\mathbf{x}}}{2}\right)\right\} &= \widehat{f}_B\left(\bar{\omega} - \frac{\tilde{\omega}}{2} - \omega_o\right) \overline{\widehat{f}_B\left(\bar{\omega} + \frac{\tilde{\omega}}{2} - \omega_o\right)} \\ &\times \int_{\mathcal{D}} d\vec{\mathbf{y}} w(\vec{\mathbf{y}}) \int_{\mathcal{D}} d\vec{\mathbf{y}}' w(\vec{\mathbf{y}}') E\left\{\widehat{G}\left(\bar{\omega} - \frac{\tilde{\omega}}{2}, \bar{\mathbf{x}} - \frac{\tilde{\mathbf{x}}}{2}, \vec{\mathbf{y}}\right) \overline{\widehat{G}\left(\bar{\omega} + \frac{\tilde{\omega}}{2}, \bar{\mathbf{x}} + \frac{\tilde{\mathbf{x}}}{2}, \vec{\mathbf{y}}'\right)}\right\} \end{aligned}$$

depends on the clutter and the scaling regime. For example, it is proved in [15] that in weakly fluctuating, isotropic cluttered media and in some particular scaling regime, we have

$$(4.12) \quad \begin{aligned} E\left\{\widehat{G}\left(\bar{\omega} - \frac{\tilde{\omega}}{2}, \bar{\mathbf{x}} - \frac{\tilde{\mathbf{x}}}{2}\right) \overline{\widehat{G}\left(\bar{\omega} + \frac{\tilde{\omega}}{2}, \bar{\mathbf{x}} + \frac{\tilde{\mathbf{x}}}{2}\right)}\right\} &\approx \widehat{G}_o\left(\bar{\omega} - \frac{\tilde{\omega}}{2}, \bar{\mathbf{x}} - \frac{\tilde{\mathbf{x}}}{2}, \vec{\mathbf{y}}\right) \\ &\times \overline{\widehat{G}_o\left(\bar{\omega} + \frac{\tilde{\omega}}{2}, \bar{\mathbf{x}} + \frac{\tilde{\mathbf{x}}}{2}, \vec{\mathbf{y}}'\right)} \exp\left\{-\frac{\tilde{\omega}^2 D_f |z' - z|}{2\theta^2 c^2}\right\} \\ &\times \exp\left\{-\frac{\tilde{\omega}^2 D_f z \wedge z'}{2c^2} - \frac{\tilde{\omega}^2 D_p z \wedge z'}{6c^2} (|\tilde{\mathbf{x}}|^2 + \tilde{\mathbf{x}} \cdot \tilde{\mathbf{y}} + |\tilde{\mathbf{y}}|^2)\right\}, \end{aligned}$$

where

$$(4.13) \quad \theta = \frac{\lambda_o L}{2\pi a^2}$$

is the Fresnel number satisfying  $\theta \ll 1$ , in the high frequency regime considered in [15].  $L$  is the range scale,  $D_f$  and  $D_p$  are random medium parameters related to the decoherence length and frequency as

$$(4.14) \quad \Omega_d = \frac{1}{\sqrt{D_f L}}, \quad X_d(\bar{\omega}) = \frac{c}{\bar{\omega}} \sqrt{\frac{3}{D_p L}}$$

and

$$(4.15) \quad \vec{\mathbf{y}} = (\mathbf{y}, z), \quad \vec{\mathbf{y}}' = (\mathbf{y}', z'), \quad \tilde{\mathbf{y}} = \mathbf{y}' - \mathbf{y}, \quad z \wedge z' = \min\{z, z'\}.$$

---

<sup>1</sup>Usually this assumption is satisfied, after proper scaling of the length variables by the range of the source.

Then, the analysis of  $\mathcal{I}^{\text{CINT}}$  in the scaling regime where (4.12) applies reduces to a tedious, but straightforward evaluation of the integrals in (4.9).

Obviously, different scalings and assumptions on the clutter lead to different moment formulae, but in the end, the coherent interferometric method seems to behave the same, in a variety of regimes. For example, the analysis in [15] based on moment formula (4.12) and the one in [14], applying to randomly layered media (see also Section 7), leads to very similar resolution limits. Moreover, these limits can be heuristically estimated from (4.11), without using any moment formula, as follows: The uncertainty in the range comes from the width of the time intervals over which we cross-correlate the traces,  $|\bar{t} - \tau(\bar{\mathbf{x}}, \vec{\mathbf{y}}^s)| \leq T_d$ , so the range resolution is roughly

$$(4.16) \quad |z - z^s| \leq cT_d = O\left(\frac{c}{\Omega_d}\right).$$

Similarly, the cross-range uncertainty comes from the domain of integration of the time lag  $\bar{t}$ . Since the pulse width is  $T_p = \frac{\pi}{B}$ ,  $|\bar{\mathbf{x}}| \leq X_d$  and

$$(4.17) \quad \nabla_{\bar{\mathbf{x}}} \tau(\bar{\mathbf{x}}, \vec{\mathbf{y}}^s) = \frac{\bar{\mathbf{x}} - \mathbf{y}^s}{c|(\bar{\mathbf{x}}, 0) - \vec{\mathbf{y}}^s|},$$

we get that the uncertainty in the direction of arrival is

$$(4.18) \quad \left| \frac{\bar{\mathbf{x}} - \mathbf{y}^s}{|(\bar{\mathbf{x}}, 0) - \vec{\mathbf{y}}^s|} - \frac{\bar{\mathbf{x}} - \mathbf{y}}{|(\bar{\mathbf{x}}, 0) - \vec{\mathbf{y}}|} \right| = O\left(\frac{c}{BX_d}\right)$$

which means that

$$(4.19) \quad |\mathbf{y} - \mathbf{y}^s| \leq O\left(\frac{cz^s}{BX_d}\right)$$

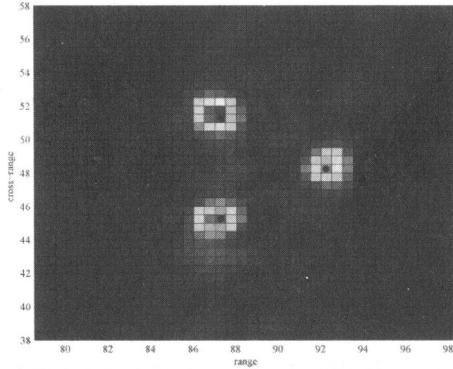


FIGURE 5. Coherent interferometric image of the sources in Figure 2. The true location of the sources is indicated with black dots.

Comparing formulae (4.17)-(4.19) with (3.13)-(3.14), we note that the resolution of the coherent interferometric method in clutter is quite similar to that of Kirchhoff migration in homogeneous media, except for the fact that the bandwidth  $B$  and array aperture  $a$  are replaced by the decoherence parameters  $\Omega_d$  and  $X_d$ . These parameters depend on the random medium, but they can be reliably estimated from the image itself, as explained in [15, 13, 12]. As an illustration of the

performance of the method, we show in Figure 5 the image of the sources displayed in Figure 2, in the same realization of the clutter as in the right picture of Figure 4. Note that the image is slightly blurrier than that in a homogeneous medium, but it is stable, as it is not noisy and it does not change with the realizations of the clutter [15, 13].

In the following sections we review in more detail the coherent interferometric approach, as it applies to randomly layered media. For simplicity, we take the case of a single point source, buried in the finely layered medium. We begin with the derivation of the transmitted acoustic pressure through the medium, as it is obtained in [55, 22, 21, 57, 1, 27, 66, 41, 40]. Then, we review the coherent interferometric point spread function, analyzed in detail in [14].

### 5. The forward model in a randomly layered slab

Consider the setup in Figure 6, where a point source, located at  $\vec{y}$  emits a short pulse  $f(t)$  at  $t = 0$ , in the vertical direction  $e_3$ . We are interested in the inverse problem of estimating the location  $\vec{y}$  of the source from the recorded traces  $p(\vec{x}, t)$ , at various locations  $\vec{x}$  in an array of transducers, over some time window  $\psi(t)$ . The pressure  $p(t, \vec{x})$  is obviously given by the time convolution of the pulse with the Green's function of the wave equation. In homogeneous media, this is simple to do, but in randomly fluctuating media, the modeling of  $p(t, \vec{x})$  is much more complicated. In the remainder of this section, we derive the expression of  $p(t, \vec{x})$  for the case of randomly layered media. This derivation is obtained in [40, 42], but we repeat it here, for the convenience of the reader.

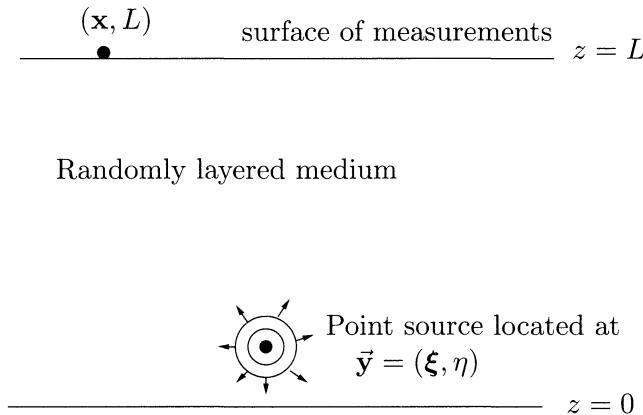


FIGURE 6. A point source located at  $\vec{y} = (\xi, \eta)$ , inside a randomly layered slab of depth  $L$ , emits a pulse that is captured at receiver location  $(\mathbf{x}, L)$ , on the top of the slab.

The equations are

$$(5.1) \quad \begin{aligned} \rho \frac{\partial \vec{u}}{\partial t}(t, \vec{x}) + \nabla p(t, \vec{x}) &= \vec{F}(t, \vec{x}), \quad \vec{x} \in \mathbb{R}^3, \quad t > 0, \\ v^{-2}(z) \frac{\partial p}{\partial t}(t, \vec{x}) + \rho \nabla \cdot \vec{u}(t, \vec{x}) &= 0, \quad \vec{x} \in \mathbb{R}^3, \quad t > 0, \\ \vec{u}(t, \vec{x}) &= \vec{0}, \quad p(t, \vec{x}) = 0, \quad t < 0, \end{aligned}$$

where we assume a constant density  $\rho$  and the wave speed  $v$  varies with depth  $z$ , in a finite slab of depth  $L = O(1)$ , with the source sitting near the bottom and the array on top, as shown in Figure 6. We introduce the notation

$$(5.2) \quad \vec{x} = (\mathbf{x}, z),$$

for arbitrary points  $\vec{x} \in \mathbb{R}^3$ , where  $z$  is measured from the bottom of the slab, and  $\mathbf{x}$  is the two-dimensional vector of lateral coordinates. The location of the source is denoted by  $\vec{y} = (\xi, \eta)$  and we record  $p$  at locations  $(\mathbf{x}, L)$  on the surface  $z = L$ . Using a high frequency regime where the carrier wavelength is much smaller than the propagation distance  $|\vec{x} - \vec{y}| = O(1)$ , we set the forcing term in (5.1) to

$$(5.3) \quad \vec{\mathbf{F}}(t, \vec{x}) = e^{-i\frac{\omega_0}{\epsilon}t} f_{\frac{B}{\epsilon}}\left(\frac{t}{\epsilon}\right) \delta(\vec{x} - \vec{y}) \mathbf{e}_3, \quad \text{for } \epsilon \ll 1,$$

and we note that in our scaling, the bandwidth of the pulse is  $B/\epsilon$ .

The derivation of the pressure field at  $\vec{x} = (\mathbf{x}, L)$  consists of three steps: (1) We begin with the decomposition of the wave field into a family of oblique, plane waves in section 5.2. (2) The oblique plane waves are then decomposed into up and down going waves, whose amplitudes satisfy a two point boundary value problem derived in section 5.3. (3) Finally, we write an integral representation of the pressure field that we record at the array. Before taking all these steps, we define the random model in Section 5.1.

**5.1. The random model of the fluctuations and scaling.** We model our uncertainty of the fluctuating part of the medium as a random process

$$(5.4) \quad v^{-2}(z) = \begin{cases} c^{-2}, & z \in (-\infty, 0) \cup (L, \infty), \\ c^{-2} [1 + \nu(\frac{z}{\epsilon^2\ell})], & z \in [0, L], \end{cases}$$

where  $c$  is the constant speed of sound in the homogeneous, average background and  $\nu$  models inhomogeneities of size  $O(\epsilon^2)$  and strength  $O(1)$ . We suppose that  $\nu$  is stationary (statistically homogeneous), with mean

$$(5.5) \quad E\{\nu(z)\} = 0,$$

and mixing, so there are no long range correlations. In particular, the covariance

$$(5.6) \quad C(z_1 - z_2) = E\{\nu(z_1)\nu(z_2)\}$$

has compact support and

$$(5.7) \quad 1 = \int_0^\infty E\{\nu(0)\nu(s)\} ds = \int_0^\infty C(s) ds$$

so that

$$(5.8) \quad \int_0^\infty E\left\{\nu\left(\frac{z}{\epsilon^2\ell}\right)\nu\left(\frac{z+s}{\epsilon^2\ell}\right)\right\} ds = \int_0^\infty C\left(\frac{s}{\epsilon^2\ell}\right) ds = \epsilon^2\ell$$

and  $\epsilon^2\ell$  is the correlation length.

Now, we see that we have three important scales in our problem: The propagation distance  $L = O(1)$ , which is much larger than the typical wavelength  $O(\epsilon)$ , so we deal with a high frequency regime. The microscale is  $O(\epsilon^2)$  so there is averaging over a wavelength, as explained in section 6.

Finally, we note that this is not the only scaling in which the analysis can be done. For example, we can have a weakly fluctuating regime [66, 1], where the strength of the fluctuations is small,  $O(\epsilon)$ , but the wavelength is on the same scale

$O(\epsilon^2)$  as the inhomogeneities, so there is a lot of interaction of the waves, with the medium. It turns out that the results reviewed in this paper extend with almost no change to this scaling, as explained in [66, 1].

**5.2. The plane wave decomposition.** Because the background is one dimensional, we can decompose the wave field in a family of plane waves, by Fourier transforming the equations with respect to  $\mathbf{x}$ .

Let us define the scaled Fourier transform in  $t$  and  $\mathbf{x}$  of the acoustic pressure as

$$(5.9) \quad \widehat{P}(\omega, \boldsymbol{\kappa}, z) = \int dt \int d\mathbf{x} p(t, \vec{\mathbf{x}}; \vec{\mathbf{y}}) e^{i\frac{\omega}{\epsilon}(t - \boldsymbol{\kappa} \cdot \mathbf{x})},$$

and its inverse

$$(5.10) \quad p(t, \vec{\mathbf{x}}) = \frac{1}{(2\pi\epsilon)^3} \int d\omega \omega^2 \int d\boldsymbol{\kappa} \widehat{P}(\omega, \boldsymbol{\kappa}, z) e^{-i\frac{\omega}{\epsilon}(t - \boldsymbol{\kappa} \cdot \mathbf{x})},$$

where  $\boldsymbol{\kappa}$  is a two dimensional slowness vector<sup>2</sup>. Similarly, we transform the components  $u_j$  of the particle velocity field, for  $j = 1, 2, 3$ . Equations (5.1) become

$$\begin{aligned} -i\frac{\omega}{\epsilon}\rho\widehat{U}_j(\omega, \boldsymbol{\kappa}, z) + i\frac{\omega}{\epsilon}\kappa_j\widehat{P}(\omega, \boldsymbol{\kappa}, z) &= 0, \quad j = 1, 2, \\ -i\frac{\omega}{\epsilon}\rho\widehat{U}_3(\omega, \boldsymbol{\kappa}, z) + \frac{\partial\widehat{P}(\omega, \boldsymbol{\kappa}, z)}{\partial z} &= \epsilon\widehat{f}_{\frac{B}{\epsilon}}(\omega - \omega_o)e^{-i\frac{\omega}{\epsilon}\boldsymbol{\kappa} \cdot \boldsymbol{\xi}}\delta(z - \eta), \\ -i\frac{\omega}{\epsilon}\frac{\widehat{P}(\omega, \boldsymbol{\kappa}, z)}{\rho v^2(z)} + i\frac{\omega}{\epsilon}\sum_{j=1}^2 \widehat{U}_j(\omega, \boldsymbol{\kappa}, z)k_j &= -\frac{\partial\widehat{u}_3(\omega, \boldsymbol{\kappa}, z)}{\partial z} \end{aligned}$$

and, after eliminating  $\widehat{U}_1$ ,  $\widehat{U}_2$  and using (5.4), we have

$$(5.11) \quad \begin{aligned} -i\frac{\omega}{\epsilon}\rho\widehat{U}_3(\omega, \boldsymbol{\kappa}, z) + \frac{\partial\widehat{P}(\omega, \boldsymbol{\kappa}, z)}{\partial z} &= \epsilon\widehat{f}_{\frac{B}{\epsilon}}(\omega - \omega_o)e^{-i\frac{\omega}{\epsilon}\boldsymbol{\kappa} \cdot \boldsymbol{\xi}}\delta(z - \eta), \\ i\frac{\omega}{\epsilon c^2\rho} [1 - c^2k^2 + \nu\left(\frac{z}{\epsilon^2}\right)] \widehat{P}(\omega, \boldsymbol{\kappa}, z) &= \frac{\partial\widehat{U}_3(\omega, \boldsymbol{\kappa}, z)}{\partial z}, \end{aligned}$$

for  $z \in (0, L)$  and  $k = |\boldsymbol{\kappa}|$ . Thus, we have reduced our three dimensional problem to an infinite system of decoupled, one dimensional equations, for slowness vectors  $\boldsymbol{\kappa} \in \mathbb{R}^2$  and frequency  $\omega$  in the bandwidth of the pulse.

**5.3. The up and down going waves.** If the medium were homogeneous, the plane waves obtained in section 5.2 would move undisturbed. However, because of the fluctuations, these waves are scattered and they can be decomposed in up and down going waves, with amplitudes  $U(\omega, \boldsymbol{\kappa}, z)$  and  $D(\omega, \boldsymbol{\kappa}, z)$ , respectively.

The up and down wave decomposition is given by [1]

$$(5.12) \quad \begin{aligned} \widehat{P} &= \zeta^{\frac{1}{2}}(\boldsymbol{\kappa}) (Ue^{i\frac{\omega}{\epsilon}\kappa_z z} - De^{-i\frac{\omega}{\epsilon}\kappa_z z}), \\ \widehat{U}_3 &= \zeta^{-\frac{1}{2}}(\boldsymbol{\kappa}) (Ue^{i\frac{\omega}{\epsilon}\kappa_z z} + De^{-i\frac{\omega}{\epsilon}\kappa_z z}), \end{aligned}$$

where

$$(5.13) \quad |\boldsymbol{\kappa}|^2 + \kappa_z^2 = c^{-2} \implies \kappa_z = \frac{\sqrt{1 - c^2\kappa^2}}{c}, \quad \kappa = |\boldsymbol{\kappa}|$$

---

<sup>2</sup>By taking the Fourier transform in both  $\mathbf{x}$  and  $t$ , we are in fact decomposing  $p$  into plane waves, which travel along the direction  $(\boldsymbol{\kappa}, \kappa_z)$ .

and

$$(5.14) \quad \zeta(\kappa) = \frac{\rho}{\kappa_z}$$

is the acoustic impedance of the average background of slowness  $c^{-1}$ . The amplitudes  $U$  and  $D$  change with depth according to differential equations

$$(5.15) \quad \frac{d}{dz} \begin{pmatrix} U \\ D \end{pmatrix} = \frac{1}{\epsilon} \mathcal{H}_{\omega, \kappa} \left( \frac{z}{\epsilon}, \nu \left( \frac{z}{\epsilon^2 \ell} \right) \right) \begin{pmatrix} U \\ D \end{pmatrix}, \quad z \neq \eta,$$

derived from (5.11), with matrix

$$(5.16) \quad \mathcal{H}_{\omega, \kappa} \left( \frac{z}{\epsilon}, \nu \left( \frac{z}{\epsilon^2 \ell} \right) \right) = \frac{i\omega \nu \left( \frac{z}{\epsilon^2 \ell} \right)}{2c^2 \kappa_z} \begin{pmatrix} 1 & -e^{-\frac{2i\omega}{\epsilon} \kappa_z z} \\ e^{\frac{2i\omega}{\epsilon} \kappa_z z} & -1 \end{pmatrix}.$$

We also have jump conditions

$$(5.17) \quad \begin{aligned} U(\omega, \kappa, \eta^+) - U(\omega, \kappa, \eta^-) &= \frac{\epsilon \zeta^{-\frac{1}{2}}(\kappa)}{2} \widehat{f}_{\frac{B}{\epsilon}}(\omega - \omega_o) e^{-i\frac{\omega}{\epsilon}(\kappa, \kappa_z) \cdot (\xi, \eta)}, \\ D(\omega, \kappa, \eta^+) - D(\omega, \kappa, \eta^-) &= -\frac{\epsilon \zeta^{-\frac{1}{2}}(\kappa)}{2} \widehat{f}_{\frac{B}{\epsilon}}(\omega - \omega_o) e^{-i\frac{\omega}{\epsilon}(\kappa, \kappa_z) \cdot (\xi, -\eta)} \end{aligned}$$

at the source depth  $z = \eta$  and, by the continuity of the velocity field at the top and bottom of the slab, we can set boundary conditions

$$(5.18) \quad U(\omega, \kappa, 0) = 0, \quad D(\omega, \kappa, L) = 0.$$

We wish to solve (5.15)-(5.18) for  $z \in (0, L)$  and obtain the up going amplitude  $U$  at the surface  $z = L$  of measurements. This can be done with a classic ordinary differential equations technique, in terms of a “propagator” matrix  $\mathcal{P}^\epsilon(\omega, \kappa, z; z_0)$  satisfying initial value problem

$$(5.19) \quad \begin{aligned} d\mathcal{P}^\epsilon(\omega, \kappa, z; z_0) &= \frac{1}{\epsilon} \mathcal{H}_{\omega, \kappa} \left( \frac{z}{\epsilon}, \nu \left( \frac{z}{\epsilon^2 \ell} \right) \right) \mathcal{P}^\epsilon(\omega, \kappa, z; z_0), \quad z \neq z_0, \\ \mathcal{P}^\epsilon(\omega, \kappa, z = z_0; z_0) &= I_{2 \times 2}, \end{aligned}$$

where  $I_{2 \times 2}$  is the identity matrix in  $\mathbb{R}^2$ . Obviously, the columns of  $\mathcal{P}^\epsilon$  stand for two linearly independent solutions of differential equation

$$\frac{d\mathbf{y}(z)}{dz} = \frac{1}{\epsilon} \mathcal{H}_{\omega, \kappa} \left( \frac{z}{\epsilon}, \nu \left( \frac{z}{\epsilon^2 \ell} \right) \right) \mathbf{y}(z),$$

so we can write, using linearity,

$$(5.20) \quad \begin{pmatrix} U(\omega, \kappa, \eta^-) \\ D(\omega, \kappa, \eta^-) \end{pmatrix} = \mathcal{P}^\epsilon(\omega, \kappa, \eta; 0) \begin{pmatrix} 0 \\ D(\omega, \kappa, 0) \end{pmatrix}$$

and

$$(5.21) \quad \begin{pmatrix} U(\omega, \kappa, L) \\ 0 \end{pmatrix} = \mathcal{P}^\epsilon(\omega, \kappa, L; \eta) \begin{pmatrix} U(\omega, \kappa, \eta^+) \\ D(\omega, \kappa, \eta^+) \end{pmatrix}.$$

Thus, to get the amplitude  $A$  at the surface  $z = L$  of measurements, it suffices to calculate the propagator  $\mathcal{P}^\epsilon$ .

It turns out that, because of the special structure of  $\mathcal{H}_{\omega, \kappa}$ , the propagator is of the form [41, 40, 42]

$$(5.22) \quad \mathcal{P}^\epsilon = \begin{pmatrix} \alpha^\epsilon & \overline{\beta^\epsilon} \\ \beta^\epsilon & \overline{\alpha^\epsilon} \end{pmatrix}$$

with entries  $\alpha^\epsilon$  and  $\beta^\epsilon$  satisfying the identity

$$(5.23) \quad |\alpha^\epsilon|^2 - |\beta^\epsilon|^2 = 1, \quad \forall z.$$

Then, we use (5.17), (5.18), (5.20) and (5.21) to obtain after some straightforward algebra [41, 40, 42] that

$$(5.24) \quad U(\omega, \kappa, L) = \frac{\epsilon \zeta^{-\frac{1}{2}}(\kappa)}{2} \widehat{f}_{\frac{B}{\epsilon}}(\omega - \omega_o) \left[ \mathcal{T} e^{-i\frac{\omega}{\epsilon}(\kappa, \kappa_z) \cdot (\xi, \eta)} + \mathcal{R} e^{-i\frac{\omega}{\epsilon}(\kappa, \kappa_z) \cdot (\xi, -\eta)} \right],$$

where

$$(5.25) \quad \mathcal{T} = \frac{T^\epsilon}{1 - \tilde{R}^\epsilon R^\epsilon}, \quad \mathcal{R} = \tilde{R}^\epsilon \mathcal{T},$$

$R^\epsilon, T^\epsilon$  are the reflection and transmission coefficients of the random medium above the source [1, 40, 42]

$$(5.26) \quad T^\epsilon = \frac{1}{\alpha^\epsilon(\omega, \kappa, L; \eta)}, \quad R^\epsilon = -\frac{\beta^\epsilon(\omega, \kappa, L; \eta)}{\alpha^\epsilon(\omega, \kappa, L; \eta)},$$

and

$$(5.27) \quad \tilde{R}^\epsilon = \frac{\overline{\beta^\epsilon}(\omega, \kappa, \eta; 0)}{\overline{\alpha^\epsilon}(\omega, \kappa, \eta; 0)}$$

is the reflection coefficient of the random medium below the source.

Finally, we note that identity (5.23) implies that

$$(5.28) \quad |T^\epsilon| \leq 1, \quad |R^\epsilon| < 1, \quad |\tilde{R}^\epsilon| < 1,$$

so we can write  $\mathcal{T}$  and  $\mathcal{R}$  as the convergent power series

$$(5.29a) \quad \mathcal{T} = \sum_{n=0}^{\infty} T^\epsilon (R^\epsilon)^n \left( \tilde{R}^\epsilon \right)^n,$$

$$(5.29b) \quad \mathcal{R} = \sum_{n=0}^{\infty} T^\epsilon (R^\epsilon)^n \left( \tilde{R}^\epsilon \right)^{n+1}.$$

Now, we gather all our results and obtain by Fourier synthesis the acoustic pressure at receiver location  $(\mathbf{x}, L)$

$$(5.30) \quad p(t, \mathbf{x}, L) = \frac{1}{2\pi\epsilon} \int d\omega \widehat{p}(\omega, \mathbf{x}, L) e^{-i\frac{\omega}{\epsilon}t},$$

where

$$(5.31) \quad \begin{aligned} \widehat{p}(\omega, \mathbf{x}, L) &= \frac{\omega^2}{4\pi^2\epsilon^2} \int d\kappa \widehat{P}(\omega, \kappa, L) e^{i\frac{\omega}{\epsilon}\kappa \cdot \mathbf{x}} = \frac{\omega^2}{8\pi^2\epsilon} \widehat{f}_{\frac{B}{\epsilon}}(\omega - \omega_o) \times \\ &\quad \int d\kappa \left[ \mathcal{T} e^{i\frac{\omega}{\epsilon}(\kappa, \kappa_z) \cdot (\mathbf{x} - \xi, L - \eta)} + \mathcal{R} e^{i\frac{\omega}{\epsilon}(\kappa, \kappa_z) \cdot (\mathbf{x} - \xi, L + \eta)} \right]. \end{aligned}$$

## 6. Statistics of the pressure field

Expression (5.30)-(5.31) is the plane wave decomposition of the acoustic pressure field originating from the source at  $\vec{y} = (\xi, \eta)$  and measured at location  $\vec{x} = (\mathbf{x}, L)$ . The oblique plane waves travel along direction  $\vec{\kappa} = (\kappa, \kappa_z)$ , where  $|\vec{\kappa}|$  is the slowness  $c^{-1}$  in the average background. The effect of the random medium is encoded in coefficients  $\mathcal{T}$  and  $\mathcal{R}$  that account for multiple reflections and transmissions in the random medium. For example, the first term in series (5.29a) corresponds to the waves going directly from the source to the surface. The first term in (5.29b) accounts for the waves that are reflected by the random medium bellow the source, before being transmitted to the surface, and so on. Finally, we

note that  $\mathcal{R}$  differs from  $\mathcal{T}$  by an extra reflection by the random medium below the source, so the second term in (5.31) can be interpreted as the contribution of an image source at  $(\xi, -\eta)$ , the reflection of  $\vec{y}$  with respect to the bottom  $z = 0$  of the strip.

Now, let us use the fact that  $\epsilon \ll 1$  and average the stochastic equations in Section 5 over the microscale. This is done in Section 6.1 by taking the *diffusion limit* [59, 60]. Then, we deal with the high frequency nature of the problem by using the method of stationary phase, and we analyze in Section 6.2 the coherent part of the transmitted pressure.

**6.1. The diffusion limit.** To characterize the random field (5.31), we wish to calculate the limit of its moments. Take for example the first and second moments

$$(6.1) \quad \lim_{\epsilon \rightarrow 0} E \{ \widehat{p}(\omega, \kappa, L) \} \quad \text{and} \quad \lim_{\epsilon \rightarrow 0} E \left\{ \widehat{p}(\omega_1, \kappa_1, L) \overline{\widehat{p}(\omega_2, \kappa_2, L)} \right\},$$

where the bar indicates complex conjugate.

Now, we recall (5.29a), (5.29b) and (5.31) to see that the first moment of the acoustic pressure  $p(t, (\mathbf{x}, L))$  requires the calculation of

$$(6.2a) \quad \lim_{\epsilon \rightarrow 0} E \left\{ T^\epsilon(\omega, \kappa) (R^\epsilon(\omega, \kappa))^n \left( \tilde{R}^\epsilon \right)^n \right\},$$

$$(6.2b) \quad \lim_{\epsilon \rightarrow 0} E \left\{ T^\epsilon(\omega, \kappa) (R^\epsilon(\omega, \kappa))^n \left( \tilde{R}^\epsilon(\omega, \kappa) \right)^{n+1} \right\},$$

for  $n \geq 0$ , whereas the second moment involves

$$(6.3a) \quad \lim_{\epsilon \rightarrow 0} E \left\{ T^\epsilon(\omega_1, \kappa_1) \overline{T^\epsilon(\omega_2, \kappa_2)} (R^\epsilon(\omega_1, \kappa_1))^n \overline{(R^\epsilon(\omega_2, \kappa_2))^m} \times \left( \tilde{R}^\epsilon(\omega_1, \kappa_1) \right)^n \overline{\left( \tilde{R}^\epsilon(\omega_2, \kappa_2) \right)^m} \right\},$$

$$(6.3b) \quad \lim_{\epsilon \rightarrow 0} E \left\{ T^\epsilon(\omega_1, \kappa_1) \overline{T^\epsilon(\omega_2, \kappa_2)} (R^\epsilon(\omega_1, \kappa_1))^n \overline{(R^\epsilon(\omega_2, \kappa_2))^m} \times \left( \tilde{R}^\epsilon(\omega_1, \kappa_1) \right)^{n+1} \overline{\left( \tilde{R}^\epsilon(\omega_2, \kappa_2) \right)^{m+1}} \right\},$$

$$(6.3c) \quad \lim_{\epsilon \rightarrow 0} E \left\{ T^\epsilon(\omega_1, \kappa_1) \overline{T^\epsilon(\omega_2, \kappa_2)} (R^\epsilon(\omega_1, \kappa_1))^n \overline{(R^\epsilon(\omega_2, \kappa_2))^m} \times \left( \tilde{R}^\epsilon(\omega_1, \kappa_1) \right)^n \overline{\left( \tilde{R}^\epsilon(\omega_2, \kappa_2) \right)^{m+1}} \right\},$$

for  $n, m \geq 0$ . We also recall from (5.26) that  $T^\epsilon$  and  $R^\epsilon$  are defined by the propagator  $\mathcal{P}^\epsilon(\omega, \kappa, L; \eta)$ , while  $\tilde{R}^\epsilon$  is determined by  $\mathcal{P}^\epsilon(\omega, \kappa, \eta; 0)$ . These two propagators describe the waves in two different regions of the random medium and, in the limit

$\epsilon \rightarrow 0$ , they decorrelate [40, 42], so we can rewrite (6.3) as

$$(6.4a) \quad \lim_{\epsilon \rightarrow 0} E \left\{ T^\epsilon(\omega_1, \kappa_1) \overline{T^\epsilon(\omega_2, \kappa_2)} (R^\epsilon(\omega_1, \kappa_1))^n \overline{(R^\epsilon(\omega_2, \kappa_2))^m} \right\} \times \\ E \left\{ \left( \tilde{R}^\epsilon(\omega_1, \kappa_1) \right)^n \overline{\left( \tilde{R}^\epsilon(\omega_2, \kappa_2) \right)^m} \right\},$$

$$(6.4b) \quad \lim_{\epsilon \rightarrow 0} E \left\{ T^\epsilon(\omega_1, \kappa_1) \overline{T^\epsilon(\omega_2, \kappa_2)} (R^\epsilon(\omega_1, \kappa_1))^n \overline{(R^\epsilon(\omega_2, \kappa_2))^m} \right\} \times \\ E \left\{ \left( \tilde{R}^\epsilon(\omega_1, \kappa_1) \right)^{n+1} \overline{\left( \tilde{R}^\epsilon(\omega_2, \kappa_2) \right)^{m+1}} \right\},$$

$$(6.4c) \quad \lim_{\epsilon \rightarrow 0} E \left\{ T^\epsilon(\omega_1, \kappa_1) \overline{T^\epsilon(\omega_2, \kappa_2)} (R^\epsilon(\omega_1, \kappa_1))^n \overline{(R^\epsilon(\omega_2, \kappa_2))^m} \right\} \times \\ E \left\{ \left( \tilde{R}^\epsilon(\omega_1, \kappa_1) \right)^n \overline{\left( \tilde{R}^\epsilon(\omega_2, \kappa_2) \right)^{m+1}} \right\}.$$

Limits (6.2) and (6.4) are calculated in [59, 60, 1, 27] and we summarize the results in the following remarks:

REMARK 6.1. In the limit  $\epsilon \rightarrow 0$ , we have the following results:

**Case (a):**

$$(6.5) \quad E \left\{ \left( \tilde{R}^\epsilon(\omega_1, \kappa_1) \right)^n \overline{\left( \tilde{R}^\epsilon(\omega_2, \kappa_2) \right)^m} \right\} \rightarrow 0,$$

if either of the following is true:

$$n \neq m \text{ or } \omega_1 \sqrt{1 - c^2 |\kappa_1|^2} - \omega_2 \sqrt{1 - c^2 |\kappa_2|^2} > O(\epsilon) \text{ and } n > 0.$$

**Case (b):** For  $n = m \geq 1$  and nearby frequencies and slowness vectors,

$$(6.6a) \quad \omega_1 = \omega + \frac{\epsilon}{2} \tilde{\omega}, \quad \omega_2 = \omega - \frac{\epsilon}{2} \tilde{\omega},$$

$$(6.6b) \quad k_1 = k + \frac{\epsilon}{2} \tilde{k}, \quad k_2 = k - \frac{\epsilon}{2} \tilde{k},$$

we have

$$(6.7) \quad E \left\{ \left( \tilde{R}^\epsilon(\omega_1, \kappa_1) \right)^n \overline{\left( \tilde{R}^\epsilon(\omega_2, \kappa_2) \right)^n} \right\} \rightarrow \int dt W_n(t, \omega, k, \eta; 0) \times \\ \exp \left\{ itc^2 \omega k \tilde{k} - it\tilde{\omega}(1 - c^2 k^2) - 2in\eta\tilde{\omega} \frac{\sqrt{1 - c^2 \kappa^2}}{c} - 2in\eta k \tilde{k} \frac{c}{\sqrt{1 - c^2 \kappa^2}} \right\},$$

where  $W_n$  solve the infinite system of transport equations

$$(6.8) \quad \frac{\partial W_n}{\partial z} + 2n \frac{\partial W_n}{\partial t} = \frac{\ell \omega^2 n^2}{2c^2(1 - c^2 k^2)} (W_{n+1} - 2W_n + W_{n-1}), \quad z \in (0, \eta),$$

$$W_n = \delta_{0,n} \delta(t), \quad \text{at } z = 0,$$

and  $\delta_{0,n}$  is the Kronecker delta symbol.

Similarly, the results in [59, 60, 1, 27] give that the factors in (6.4), depending on the propagator between  $\eta$  and  $L$ , satisfy:

REMARK 6.2. In the limit  $\epsilon \rightarrow 0$ , we have the following results:

**Case (a):**

$$(6.9) \quad E \left\{ T^\epsilon(\omega_1, \kappa_1) \overline{T^\epsilon(\omega_2, \kappa_2)} (R^\epsilon(\omega_1, \kappa_1))^n \overline{(R^\epsilon(\omega_2, \kappa_2))^m} \right\} \rightarrow 0,$$

if either of the following is true:

$$n \neq m \text{ or } \omega_1 \sqrt{1 - c^2 |\kappa_1|^2} - \omega_2 \sqrt{1 - c^2 |\kappa_2|^2} > O(\epsilon) \text{ and } n > 0.$$

**Case (b):** For  $n = m \geq 0$  and  $\omega_1, \omega_2, k_1$  and  $k_2$  defined by (6.6), we have

$$(6.10) \quad \begin{aligned} E \left\{ T^\epsilon(\omega_1, \kappa_1) \overline{T^\epsilon(\omega_2, \kappa_2)} (R^\epsilon(\omega_1, \kappa_1))^n \overline{(R^\epsilon(\omega_2, \kappa_2))^m} \right\} \rightarrow \\ \int dt W_n(t, \omega, k, L; \eta) \exp \left\{ itc^2 \omega k \tilde{k} - it\tilde{\omega}(1 - c^2 k^2) + \right. \\ \left. 2in\eta\tilde{\omega} \frac{\sqrt{1 - c^2 \kappa^2}}{c} + 2in\eta k \tilde{k} \frac{c}{\sqrt{1 - c^2 \kappa^2}} \right\}, \end{aligned}$$

where  $W_n$  solve the infinite system of transport equations

$$(6.11) \quad \begin{aligned} \frac{\partial W_n}{\partial z} + 2n \frac{\partial W_n}{\partial t} &= \frac{\ell \omega^2}{2c^2(1 - c^2 k^2)} [(n+1)^2 W_{n+1} + n^2 W_{n-1} - \\ &\quad (2n^2 + 2n + 1) W_n], \quad z \in (\eta, L), \\ W_n &= \delta_{0,n} \delta(t), \quad \text{at } z = \eta. \end{aligned}$$

**Case (c):** The expected transmission coefficient is

$$(6.12) \quad E \{T^\epsilon(\omega, \kappa)\} \rightarrow E \{T(\omega, \kappa)\}$$

and, if either  $|\omega_1 - \omega_2| > O(\epsilon)$  and/or  $|\kappa_1 - \kappa_2| > O(\epsilon)$ , we obtain

$$(6.13) \quad E \left\{ T^\epsilon(\omega_1, \kappa_1) \overline{T^\epsilon(\omega_2, \kappa_2)} \right\} \rightarrow E \left\{ K_{\text{ODA}}(\omega_1, \kappa_1) \overline{K_{\text{ODA}}(\omega_2, \kappa_2)} \right\},$$

where  $K_{\text{ODA}}$  is the so-called ODA (O'Doherty Anstey) kernel

$$(6.14) \quad K_{\text{ODA}}(\omega, \kappa) = \exp \left[ \frac{i\omega}{c\sqrt{1 - c^2 \kappa^2}} \sqrt{\frac{\ell}{2}} W(L - \eta) - \frac{\omega^2 \ell (L - \eta)}{4c^2(1 - c^2 k^2)} \right],$$

and  $W$  is standard Brownian motion [56].

These results seem quite complicated but we shall see that all we need for our coherent interferometric function are case (a) in Remark 6.1 and cases (a) and (c) in Remark (6.2).

**6.2. The coherent transmitted pressure field.** Remarks 6.1 and 6.2 show that the coherent (i.e. expected) pressure field at the receiver location  $(\mathbf{x}, L)$  is given by

$$(6.15) \quad E \{p(t, \mathbf{x}, L)\} \approx \int \frac{d\omega \omega^2 \widehat{f}_B(\omega - \omega_o)}{2(2\pi)^3 \epsilon^2} \int d\kappa e^{i\frac{\omega}{\epsilon}[(\kappa, \kappa_z) \cdot (\mathbf{x} - \boldsymbol{\xi}, L - \eta) - t]} E \{T^\epsilon(\omega, k)\},$$

where

$$(6.16) \quad \lim_{\epsilon \rightarrow 0} E \{T^\epsilon(\omega, k)\} \approx E \left\{ e^{\frac{i\omega}{c\sqrt{1 - c^2 \kappa^2}} \sqrt{\frac{\ell}{2}} W(L - \eta) - \frac{\omega^2 \ell (L - \eta)}{4c^2(1 - c^2 k^2)}} \right\} = e^{-\frac{\omega^2 \ell (L - \eta)}{2c^2(1 - c^2 k^2)}}.$$

Now, let us use the high frequency aspect of the problem and approximate (6.15) using the method of stationary phase [5]. The main contribution to the integral over  $\kappa$  in (6.15) comes from the slowness vector

$$(6.17) \quad \kappa^* = \frac{\mathbf{x} - \boldsymbol{\xi}}{c r(\vec{\mathbf{y}}, \mathbf{x})}, \quad r(\vec{\mathbf{y}}, \mathbf{x}) = |\vec{\mathbf{y}} - (\mathbf{x}, L)| = \sqrt{(L - \eta)^2 + |\mathbf{x} - \boldsymbol{\xi}|^2},$$

and the stationary phase formula gives

$$(6.18) \quad E\{p(t, \mathbf{x}, L)\} \approx \frac{L - \eta}{8\pi^2 c\epsilon r^2(\vec{\mathbf{y}}, \mathbf{x})} \int d\omega (-i\omega) \hat{f}_{\frac{B}{\epsilon}}(\omega - \omega_o) e^{i\frac{\omega}{\epsilon}(\frac{r(\vec{\mathbf{y}}, \mathbf{x})}{c} - t) - \frac{\omega^2 \ell r^2(\vec{\mathbf{y}}, \mathbf{x})}{2c^2(L - \eta)}}.$$

In the absence of the fluctuations, the pressure recorded at the array would be

$$(6.19) \quad \begin{aligned} p_0(t, \mathbf{x}, L) &= \frac{L - \eta}{8\pi^2 c\epsilon r^2(\vec{\mathbf{y}}, \mathbf{x})} \int d\omega (-i\omega) \hat{f}_{\frac{B}{\epsilon}}(\omega - \omega_o) e^{i\frac{\omega}{\epsilon}(\frac{r(\vec{\mathbf{y}}, \mathbf{x})}{c} - t)} \\ &= \frac{\partial}{\partial z} \left\{ \frac{f\left(\frac{ct - r(\vec{\mathbf{y}}, \mathbf{x})}{ce}\right)}{4\pi r(\vec{\mathbf{y}}, \mathbf{x})} \right\}_{z=L}. \end{aligned}$$

Thus, we conclude from (6.18) that the presence of the fluctuations changes the coherent pressure by an apparent broadening of the pulse, due to the convolution with a Gaussian. Explicitly, we have

$$(6.20) \quad E\{p(t, \mathbf{x}, L)\} \approx \frac{\partial}{\partial z} \left\{ \frac{(f * \mathcal{N})\left(\frac{ct - r(\vec{\mathbf{y}}, \mathbf{x})}{ce}\right)}{4\pi r(\vec{\mathbf{y}}, \mathbf{x})} \right\}_{z=L},$$

where

$$(6.21) \quad \mathcal{N}(t) = \frac{c}{r(\vec{\mathbf{y}}, \mathbf{x})} \sqrt{\frac{L - \eta}{2\pi\ell}} e^{-\frac{t^2 c^2 (L - \eta)}{2\ell r^2(\vec{\mathbf{y}}, \mathbf{x})}}.$$

While (6.20) gives just the expectation of the pressure field at the receiver, it is shown in [1, 27, 66, 22, 6] that, for a small time window centered at the expected arrival time  $r(\vec{\mathbf{y}}, \mathbf{x})/c$ , we have that  $p$  is given by

$$(6.22) \quad \begin{aligned} p\left(\frac{r(\vec{\mathbf{y}}, \mathbf{x})}{c} + \epsilon\sigma, \mathbf{x}, L; \vec{\mathbf{y}}\right) &\approx \frac{L - \eta}{8\pi^2 c\epsilon r^2(\vec{\mathbf{y}}, \mathbf{x})} \int d\omega (-i\omega) \hat{f}_{\frac{B}{\epsilon}}(\omega - \omega_o) e^{i\frac{\omega}{c}\sigma} \times \\ &\quad e^{i\sqrt{\frac{\ell}{2}} \frac{\omega r(\vec{\mathbf{y}}, \mathbf{x})}{c(L - \eta)} W(L - \eta) - \frac{\omega^2 \ell r^2(\vec{\mathbf{y}}, \mathbf{x})}{4c^2(L - \eta)}}. \end{aligned}$$

This is the O'Doherty Anstey (ODA) formula [55], that shows that, if we focus attention on the coherent (ballistic) part of the signal, then the presence of the inhomogeneities creates a pulse broadening by a Gaussian with *half* the variance of that in (6.21) and a random shift of the arrival time. If we knew this random shift and centered our window at

$$r(\vec{\mathbf{y}}, \mathbf{x})/c + \epsilon c \sqrt{1 - c^2 \kappa^2} \sqrt{\frac{\ell}{2}} W(L - \eta),$$

then our coherent part of the signal would be deterministic and we say that we have *pulse stabilization*.

Of course, description (6.22) does not hold for times  $t > r(\vec{\mathbf{y}}, \mathbf{x})/c$ , when the pressure is randomly fluctuating about mean zero, and it is observed in our measurements as a long coda. We shall take care of this coda in our coherent interferometric approach, by doing some smoothing over appropriately chosen time windows.

## 7. Imaging

We use the asymptotic description of the acoustic pressure at the surface of measurements to identify the source location. We begin in section 7.1 with an imaging approach that uses the coherent (ballistic) part of the recorded pressure

to find the source. Then, we discuss in section 7.2 the coherent interferometric method, that is more robust and can be extended easily to distributed sources and to imaging of reflectors with scattering data.

### 7.1. Imaging with the coherent front of the recorded pressure field.

The ODA formula (6.22) tells us that, in a time window of size  $O(\epsilon)$  around the coherent arrival time, the pressure recorded at receiver location  $(\mathbf{x}, L)$  is similar to what we would measure in a homogeneous medium, except for a slight random arrival time shift and a pulse spreading. This fact leads to the following imaging approach:

Assuming that we have a simple pulse  $f(t)$ , of known shape, track the front of the signal, at a couple of receiver locations  $(\mathbf{x}_i, L)$ ,  $i = 1, \dots, n$ , and simply pick the time at which the maximum occurs. This time is, according to (6.22),

$$(7.1) \quad \tau_i = \frac{r(\vec{\mathbf{y}}, \mathbf{x}_i)}{c} + \epsilon \sqrt{\frac{\ell}{2}} \frac{r(\vec{\mathbf{y}}, \mathbf{x}_i)}{c(L - \eta)} W(L - \eta), \quad i = 1, \dots, n,$$

where the Brownian motion  $W(L - \eta)$  is *the same* for all measurements. Now, use (7.1) to find the source location. This can be done with:

- (a) A deterministic imaging approach that uses three measurements to estimate  $\xi, \eta$  with an error of the order of the wavelength  $\epsilon$ , due to the random shift.
- (b) An improved, random approach that finds  $\xi, \eta, W(L - \eta)$ , from (7.1), with four measurements (i.e.  $n = 4$ ). This is actually done in [42] and it works well for the simple problem that we consider here.

However, this imaging method has two drawbacks: First, it relies on our ability to track the coherent part of the signal and to pick its maximum. This is rather tricky if the pulse is complicated in shape and unknown. Second, this imaging method is too model dependent in the sense that it works only for perfectly layered random media and for imaging small (point like), well separated sources.

**7.2. The coherent interferometric point spread function.** Suppose that we record the pressure field at receiver location  $(\mathbf{x}, L)$ , over a time window  $\psi(\frac{t}{\alpha})$  of size  $O(\alpha)$ . If we were interested in just the coherent part of the signal, we would center  $\psi$  at the deterministic arrival time  $r(\vec{\mathbf{y}}, \mathbf{x})/c$  and we would set  $\alpha = \epsilon$ . Here, we take  $\alpha = 1$  and record both the ballistic part of  $p$  and the coda. The windowed pressure field is (recall (5.31))

$$(7.2) \quad \begin{aligned} p_w(t, \mathbf{x}, L) &= \frac{1}{2\pi\epsilon} \int d\omega \widehat{p}_w(\omega, \mathbf{x}, L) e^{-i\frac{\omega}{\epsilon}t}, \\ \widehat{p}_w(\omega, \mathbf{x}, L) &= -\frac{1}{2\pi} \int du \widehat{\psi}(u) \widehat{p}(\omega - \epsilon u, \mathbf{x}, L; \vec{\mathbf{y}}). \end{aligned}$$

Now, imagine that we record the traces at symmetric locations with respect to a fixed location  $(\mathbf{x}, L)$  on the surface. This is called a common mid-point gather data acquisition. Other acquisition geometries are possible, but to fix ideas, we concentrate on this one. Suppose that our array size is small in comparison with the propagation distance but large in comparison with the wavelength, so we set the aperture to  $\delta a$ , for  $\epsilon \ll \delta \ll 1$ . The coherent interferometric imaging function

is

$$(7.3) \quad \mathcal{I}^{\text{CINT}}(\vec{\mathbf{y}}^s) \sim \int_{|\tilde{\mathbf{x}}| \leq \delta a/2} d\tilde{\mathbf{x}} \int_{\mathcal{B}} d\omega \int_{-\Omega_d}^{\Omega_d} d\tilde{\omega} \widehat{p}_w \left( \omega + \frac{\epsilon \tilde{\omega}}{2}, \mathbf{x} + \frac{\tilde{\mathbf{x}}}{2}, L; \vec{\mathbf{y}} \right) \times e^{-i\left(\frac{\omega}{\epsilon} + \frac{\tilde{\omega}}{2}\right) \frac{r(\vec{\mathbf{y}}^s, \mathbf{x} + \frac{\tilde{\mathbf{x}}}{2})}{c}} \widehat{p}_w \left( \omega - \frac{\epsilon \tilde{\omega}}{2}, \mathbf{x} - \frac{\tilde{\mathbf{x}}}{2}, L; \vec{\mathbf{y}} \right) e^{i\left(\frac{\omega}{\epsilon} - \frac{\tilde{\omega}}{2}\right) \frac{r(\vec{\mathbf{y}}^s, \mathbf{x} - \frac{\tilde{\mathbf{x}}}{2})}{c}},$$

where  $\Omega_d$  is the inverse of  $T_d$ , the time window over which we smooth the traces, and it is chosen such that<sup>3</sup>

$$O(\epsilon) \ll \epsilon \Omega_d \ll B.$$

Next, let us use (7.2) and (5.31) in (7.3) and obtain

$$(7.4) \quad \begin{aligned} \mathcal{I}^{\text{CINT}}(\vec{\mathbf{y}}^s) &\sim \int_{|\tilde{\mathbf{x}}| \leq \delta a/2} d\tilde{\mathbf{x}} \int_{\mathcal{B}} d\omega \omega^4 |\widehat{f}(\omega)|^2 \int_{\Omega_d}^{\Omega_d} d\tilde{\omega} \int du \int du' \widehat{\psi}(u) \overline{\widehat{\psi}(u')} \\ &\quad \times \int d\kappa \int d\kappa' \mathcal{T} \left( \omega + \frac{\epsilon \tilde{\omega}}{2} - \epsilon u, \kappa \right) \overline{\mathcal{T} \left( \omega - \frac{\epsilon \tilde{\omega}}{2} - \epsilon u', \kappa' \right)} \\ &\quad \exp \left\{ i \left( \frac{\omega}{\epsilon} + \frac{\tilde{\omega}}{2} - u \right) (\kappa, \kappa_z) \cdot (\mathbf{x} + \frac{\tilde{\mathbf{x}}}{2} - \boldsymbol{\xi} L - \eta) \right\} - \\ &\quad i \left( \frac{\omega}{\epsilon} - \frac{\tilde{\omega}}{2} - u' \right) (\kappa', \kappa'_z) \cdot (\mathbf{x} - \frac{\tilde{\mathbf{x}}}{2} - \boldsymbol{\xi}, L - \eta) \Big] - \\ &\quad i \left( \frac{\omega}{\epsilon} + \frac{\tilde{\omega}}{2} \right) \frac{r(\vec{\mathbf{y}}^s, \mathbf{x} + \frac{\tilde{\mathbf{x}}}{2})}{c} + i \left( \frac{\omega}{\epsilon} - \frac{\tilde{\omega}}{2} \right) \frac{r(\vec{\mathbf{y}}^s, \mathbf{x} - \frac{\tilde{\mathbf{x}}}{2})}{c} \Big\} + \text{other terms}, \end{aligned}$$

where symbol  $\sim$  stands for approximate, up to a multiplicative constant and by “other terms” we mean those that contain the products  $\mathcal{T}\mathcal{R}$ ,  $\mathcal{R}\mathcal{T}$  and  $\mathcal{R}\mathcal{R}$ , respectively.

Recall from [27, 28, 42] and Remarks 6.1 and 6.2 that the reflection coefficients become decorrelated at different frequencies and slownesses. The transmission coefficients remain correlated, but they can be replaced by the ODA ones, given by (6.14). Because of this, and because of the integration in (7.4), over different slowness vectors  $\kappa$  and  $\kappa'$ , we obtain in [14] that as  $\epsilon \rightarrow 0$ ,  $\mathcal{I}^{\text{CINT}}$  converges in distribution to

$$(7.5) \quad \begin{aligned} \mathcal{I}_{\text{ODA}}^{\text{CINT}}(\vec{\mathbf{y}}^s) &\sim \int_{|\tilde{\mathbf{x}}| \leq \delta a/2} d\tilde{\mathbf{x}} \int_{\mathcal{B}} d\omega \omega^4 |\widehat{f}(\omega)|^2 \int_{\Omega_d}^{\Omega_d} d\tilde{\omega} \int du \int du' \widehat{\psi}(u) \overline{\widehat{\psi}(u')} \times \\ &\quad \int d\kappa \int d\kappa' K_{\text{ODA}}(\omega, \kappa) \overline{K_{\text{ODA}}(\omega, \kappa')} e^{i\left(\frac{\omega}{\epsilon} + \frac{\tilde{\omega}}{2} - u\right)(\kappa, \kappa_z) \cdot (\mathbf{x} + \frac{\tilde{\mathbf{x}}}{2} - \boldsymbol{\xi}, L - \eta)} \times \\ &\quad e^{i\left(\frac{\omega}{\epsilon} - \frac{\tilde{\omega}}{2} - u'\right)(\kappa', \kappa'_z) \cdot (\mathbf{x} - \frac{\tilde{\mathbf{x}}}{2} - \boldsymbol{\xi}, L - \eta) - i\left(\frac{\omega}{\epsilon} + \frac{\tilde{\omega}}{2}\right) \frac{r(\vec{\mathbf{y}}^s, \mathbf{x} + \frac{\tilde{\mathbf{x}}}{2})}{c} + i\left(\frac{\omega}{\epsilon} - \frac{\tilde{\omega}}{2}\right) \frac{r(\vec{\mathbf{y}}^s, \mathbf{x} - \frac{\tilde{\mathbf{x}}}{2})}{c}}. \end{aligned}$$

Thus, we reduced our problem to an essentially deterministic one and all that is left to do is to evaluate the integrals in (7.5) with the method of stationary phase.

---

<sup>3</sup>Note that here and actually, in most wave scattering regimes, it is very difficult to specify the proper  $\Omega_d$  directly from the theory. In practice, we deal with this issue by adaptively choosing the  $\Omega_d$  that gives the best quality of an image, as explained in [13].

The rapidly fluctuating phase in (7.5) is

$$(7.6) \quad \begin{aligned} \frac{\omega}{\epsilon} \Phi &= \frac{\omega}{\epsilon} \left\{ \boldsymbol{\kappa} \cdot \left( \mathbf{x} + \frac{\tilde{\mathbf{x}}}{2} - \boldsymbol{\xi} \right) + \frac{\sqrt{1-c^2\kappa^2}}{c} (L-\eta) - \frac{r(\vec{\mathbf{y}}^s, \mathbf{x} + \frac{\tilde{\mathbf{x}}}{2})}{c} - \right. \\ &\quad \left. \boldsymbol{\kappa}' \cdot \left( \mathbf{x} - \frac{\tilde{\mathbf{x}}}{2} - \boldsymbol{\xi} \right) - \frac{\sqrt{1-c^2(\kappa')^2}}{c} (L-\eta) + \frac{r(\vec{\mathbf{y}}^s, \mathbf{x} - \frac{\tilde{\mathbf{x}}}{2})}{c} \right\}. \end{aligned}$$

Then, when integrating over  $\boldsymbol{\kappa}, \boldsymbol{\kappa}'$  we have that most contribution comes from the vicinity of the stationary points satisfying

$$(7.7a) \quad \nabla_{\boldsymbol{\kappa}} \Phi = \mathbf{0} \implies \boldsymbol{\kappa} = \frac{\mathbf{x} + \frac{\tilde{\mathbf{x}}}{2} - \boldsymbol{\xi}}{cr(\vec{\mathbf{y}}, \mathbf{x} + \frac{\tilde{\mathbf{x}}}{2})},$$

$$(7.7b) \quad \nabla_{\boldsymbol{\kappa}'} \Phi = \mathbf{0} \implies \boldsymbol{\kappa}' = \frac{\mathbf{x} - \frac{\tilde{\mathbf{x}}}{2} - \boldsymbol{\xi}}{cr(\vec{\mathbf{y}}, \mathbf{x} - \frac{\tilde{\mathbf{x}}}{2})},$$

so after integration, the phase becomes

$$(7.8) \quad \frac{\omega}{\epsilon} \Phi = \frac{\omega}{\epsilon c} [r(\vec{\mathbf{y}}, \mathbf{x} + \frac{\tilde{\mathbf{x}}}{2}) - r(\vec{\mathbf{y}}^s, \mathbf{x} + \frac{\tilde{\mathbf{x}}}{2}) - r(\vec{\mathbf{y}}, \mathbf{x} - \frac{\tilde{\mathbf{x}}}{2}) + r(\vec{\mathbf{y}}^s, \mathbf{x} + \frac{\tilde{\mathbf{x}}}{2})]$$

and the integral over  $\tilde{\omega}$  reduces to

$$(7.9) \quad \int_{-\Omega_d}^{\Omega_d} d\tilde{\omega} e^{i\frac{\tilde{\omega}}{c}[r(\vec{\mathbf{y}}, \mathbf{x} + \frac{\tilde{\mathbf{x}}}{2}) - r(\vec{\mathbf{y}}^s, \mathbf{x} + \frac{\tilde{\mathbf{x}}}{2}) + r(\vec{\mathbf{y}}, \mathbf{x} - \frac{\tilde{\mathbf{x}}}{2}) - r(\vec{\mathbf{y}}^s, \mathbf{x} + \frac{\tilde{\mathbf{x}}}{2})]}$$

We use further the method of stationary phase, to integrate over  $\tilde{\mathbf{x}}$  and  $\omega$  and find from the stationary conditions  $\nabla_{\tilde{\mathbf{x}}} \Phi = \mathbf{0}$  and  $\nabla_{\omega} \Phi = 0$  that

$$(7.10a) \quad \frac{\mathbf{x} + \frac{\tilde{\mathbf{x}}}{2} - \boldsymbol{\xi}}{r(\vec{\mathbf{y}}, \mathbf{x} + \frac{\tilde{\mathbf{x}}}{2})} + \frac{\mathbf{x} - \frac{\tilde{\mathbf{x}}}{2} - \boldsymbol{\xi}}{r(\vec{\mathbf{y}}, \mathbf{x} - \frac{\tilde{\mathbf{x}}}{2})} = \frac{\mathbf{x} + \frac{\tilde{\mathbf{x}}}{2} - \boldsymbol{\xi}^s}{r(\vec{\mathbf{y}}^s, \mathbf{x} + \frac{\tilde{\mathbf{x}}}{2})} + \frac{\mathbf{x} - \frac{\tilde{\mathbf{x}}}{2} - \boldsymbol{\xi}^s}{r(\vec{\mathbf{y}}^s, \mathbf{x} - \frac{\tilde{\mathbf{x}}}{2})},$$

$$(7.10b) \quad r\left(\vec{\mathbf{y}}, \mathbf{x} + \frac{\tilde{\mathbf{x}}}{2}\right) - r\left(\vec{\mathbf{y}}^s, \mathbf{x} + \frac{\tilde{\mathbf{x}}}{2}\right) = r\left(\vec{\mathbf{y}}, \mathbf{x} - \frac{\tilde{\mathbf{x}}}{2}\right) - r\left(\vec{\mathbf{y}}^s, \mathbf{x} - \frac{\tilde{\mathbf{x}}}{2}\right).$$

The integral over  $\tilde{\omega}$  gives

$$(7.11) \quad r\left(\vec{\mathbf{y}}, \mathbf{x} + \frac{\tilde{\mathbf{x}}}{2}\right) + r\left(\vec{\mathbf{y}}^s, \mathbf{x} + \frac{\tilde{\mathbf{x}}}{2}\right) = r\left(\vec{\mathbf{y}}, \mathbf{x} - \frac{\tilde{\mathbf{x}}}{2}\right) + r\left(\vec{\mathbf{y}}^s, \mathbf{x} - \frac{\tilde{\mathbf{x}}}{2}\right) + O\left(\frac{c}{\Omega_d}\right)$$

and, finally, we have

$$(7.12) \quad \begin{aligned} K_{\text{ODA}}(\omega, \boldsymbol{\kappa}) \overline{K_{\text{ODA}}(\omega, \boldsymbol{\kappa}')} &\approx e^{\frac{i\omega W(L-\eta)}{c(L-\eta)} \sqrt{\frac{\ell}{2}} [r(\vec{\mathbf{y}}, \mathbf{x} + \frac{\tilde{\mathbf{x}}^*}{2}) - r(\vec{\mathbf{y}}, \mathbf{x} - \frac{\tilde{\mathbf{x}}^*}{2})]} \\ &\times e^{-\frac{\omega^2 \ell}{4c^2(L-\eta)} [r^2(\vec{\mathbf{y}}, \mathbf{x} + \frac{\tilde{\mathbf{x}}^*}{2}) + r^2(\vec{\mathbf{y}}, \mathbf{x} - \frac{\tilde{\mathbf{x}}^*}{2})]} \\ &= e^{-\frac{\omega^2 \ell r^2(\vec{\mathbf{y}}, \mathbf{x})}{2c^2(L-\eta)}} (1 + O(\delta)). \end{aligned}$$

Note that, when taking expectations in (7.12), the correction becomes  $O(\delta^2)$  instead of  $O(\delta)$ , because

$$E \left\{ \exp \left\{ i\omega \frac{W(L-\eta)}{c(L-\eta)} \sqrt{\frac{\ell}{2}} [r(\vec{\mathbf{y}}, \mathbf{x} + \frac{\tilde{\mathbf{x}}^*}{2}) - r(\vec{\mathbf{y}}, \mathbf{x} - \frac{\tilde{\mathbf{x}}^*}{2})] \right\} \right\} =$$

$$\exp \left\{ -\omega^2 \frac{\ell}{4c(L-\eta)} [r(\vec{\mathbf{y}}, \mathbf{x} + \frac{\tilde{\mathbf{x}}^*}{2}) - r(\vec{\mathbf{y}}, \mathbf{x} - \frac{\tilde{\mathbf{x}}^*}{2})]^2 \right\} = O(\delta^2),$$

so, due to phase cancellations, the result is deterministic, to leading order.

It is easy to see from (7.10) and (7.11) that  $\mathcal{I}_{\text{ODA}}^{\text{CINT}}(\vec{\mathbf{y}}^s)$  peaks at  $\vec{\mathbf{y}}^s \approx \vec{\mathbf{y}}$ . In fact, in [14] we carry out a detailed analysis of (7.5) and find that the source is determined up to the resolution limits

$$(7.13) \quad r(\vec{\mathbf{y}}, \mathbf{x}) - r(\vec{\mathbf{y}}^s, \mathbf{x}) = O\left(\frac{c}{\Omega_d}\right),$$

$$(7.14) \quad \left| \frac{\mathbf{x} - \boldsymbol{\xi}}{r(\vec{\mathbf{y}}, \vec{\mathbf{y}})} - \frac{\mathbf{x} - \boldsymbol{\xi}^s}{r(\vec{\mathbf{y}}, \vec{\mathbf{y}}^s)} \right| = O\left(\frac{\epsilon c}{B\delta a}\right).$$

Thus, the resolution of the image in random media is similar to what we would get in a homogeneous medium, except that the bandwidth is replaced by the decoherence frequency in the uncertainty of the range. The decoherence frequency is typically much smaller than the bandwidth, which means that the image is blurrier than in homogeneous media, but  $\Omega_d$  is still much larger than one, so our estimate of the source is useful. The important question that arises is how to choose  $\Omega_d$ ? It is very difficult to find  $\Omega_d$  from the theory, but it can be estimated with some clever signal processing, at the same time as we form the image, as explained in [15, 13].

### Acknowledgment

The coherent interferometric imaging approach discussed in this paper has been obtained in collaboration with George Papanicolaou and Chrysoula Tsogka. A detailed analysis of this approach can be found in [15, 13, 14]. We also used the theory of wave propagation in randomly layered media, developed in a sequence of studies by: M. Asch, R. Burridge, J. P. Fouque, J. Garnier, W. Kohler, A. Nachbin, G. Papanicolaou, M. Postel, K. Solna and B. White. Citations of these studies can be found throughout the paper.

### References

1. M. Asch, W. Kohler, G. Papanicolaou, M. Postel, and B. White. Frequency content of randomly scattered signals. *Siam Review*, 33:519–626, 1991.
2. A. B. Baggeroer, W. A. Kuperman, and P. N. Mikhalevsky. An overview of matched field methods in ocean acoustics. *IEEE J. Ocean Eng.*, 18:401–424, 1993.
3. G. Bal, G. Papanicolaou, and L. Ryzhik. Self-averaging in time reversal for the parabolic wave equation. *Stochastics and Dynamics*, 2:507–531, 2002.
4. G. Bal and L. Ryzhik. Time reversal and refocusing in random media. *SIAM J. Appl. Math.*, 63(5):1475–1498, 2003.
5. C. M. Bender and S. A. Orszag. *Advanced mathematical methods for scientists and engineers*. Springer-Verlag, New York, 1999. reprint of the 1978 original.
6. L. Berlyand and R. Burridge. The accuracy of the O'Doherty Anstey approximation for waves propagating in highly disordered stratified media. *Wave Motion*, 21:357–373, 1995.
7. G. Beylkin. Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform. *J. Math. Phys.*, 26(1):99–108, 1985.
8. G. Beylkin and R. Burridge. Linearized inverse scattering problems in acoustics and elasticity. *Wave Motion*, 12(1):15–52, 1990.
9. N. Bleistein, J.K. Cohen, and J.W. Stockwell Jr. *Mathematics of multidimensional seismic imaging, migration, and inversion*. Springer, New York, 2001.
10. P. Blomgren, G. Papanicolaou, and H. Zhao. Super-resolution in time-reversal acoustics. *Journal of the Acoustical Society of America*, 111:238–248, 2002.
11. L. Borcea, G. Papanicolaou, and C. Tsogka. A resolution study for imaging and time reversal in random media. *Contemporary Math*, 333:63–77, 2003.
12. L. Borcea, G. Papanicolaou, and C. Tsogka. Theory and applications of time reversal and interferometric imaging. *Inverse Problems*, 19:S134–164, 2003.

13. L. Borcea, G. Papanicolaou, and C. Tsogka. Adaptive interferometric imaging in clutter. Preprint, 2005.
14. L. Borcea, G. Papanicolaou, and C. Tsogka. Coherent interferometry in finely layered random media. submitted to SIAM Multiscale Modeling and Simulations, 2005.
15. L. Borcea, G. Papanicolaou, and C. Tsogka. Interferometric array imaging in clutter. Inverse Problems, in press, 2005.
16. L. Borcea, G. Papanicolaou, C. Tsogka, and J. Berryman. Imaging and time reversal in random media. *Inverse Problems*, 18:1247–1279, 2002.
17. L. Borcea, G. Papanicolaou, C. Tsogka, and J. Berryman. Statistically stable ultrasonic imaging in random media. *JASA*, 112:1509–1522, 2002.
18. B. Borden. Mathematical problems in radar inverse scattering. *Inverse Problems*, 19:R1–R28, 2002.
19. M. Born and E. Wolf. *Principles of optics*. Academic Press, New York, 1970.
20. H. P. Bucker. Use of calculated sound field and matched-field detection to locate sound sources in shallow water. *J. Acoust. Soc. Am.*, 59:368–373, 1976.
21. R. Burridge, G. Papanicolaou, P. Sheng, and B. White. Probing a random medium with a pulse. *SIAM J. Appl. Math.*, 49:582–607, 1989.
22. R. Burridge, G. Papanicolaou, and B. White. Statistics for pulse reflection from a randomly layered medium. *SIAM J. Appl. Math.*, 47:146–168, 1987.
23. J. Carazzone and W. Symes. Velocity inversion by differential semblance optimization. *Geophysics*, 56, 1991.
24. T. K. Chan, Y. Kuga, and A. Ishimaru. Experimental studies on circular sar imaging in clutter using angular correlation function technique. *IEEE Trans. Geoscience and Remote Sensing*, 37:2192–2197, 1999.
25. M. Cheney. A mathematical tutorial on synthetic aperture radar. *SIAM Rev.*, 43(2):301–312, 2001.
26. J. F. Claerbout. *Fundamentals of geophysical data processing : with applications to petroleum prospecting*. CA : Blackwell Scientific Publications, Palo Alto, 1985.
27. J. F. Clouet and J. P. Fouque. Spreading of a pulse travelling in random media. *Annals of Applied Probability*, 4:1083–1097, 1994.
28. J. F. Clouet and J. P. Fouque. A time reversal method for an acoustical pulse propagating in randomly layered media. *Wave Motion*, 25:361–368, 1997.
29. D. Colton and R. Cress. *Inverse acoustic and electromagnetic scattering theory*, volume 93 of *Applied Math. Sc.* Springer Verlag, Berlin, Heidelberg, second edition edition, 1998.
30. J.C. Curlander and R.N. McDonough. *Synthetic Aperture Radar*. Wiley, New York, 1991.
31. A. Derode, P. Roux, and M. Fink. Robust acoustic time reversal with high-order multiple scattering. *Physical Review Letters*, 75:4206–4209, 1995.
32. D. R. Dowling and D. R. Jackson. Phase conjugation in underwater acoustics. *J. Acoust. Soc. Am.*, 89:171–181, 1990.
33. D. R. Dowling and D. R. Jackson. Narrow band performance of phase conjugate arrays in dynamic random media. *J. Acoust. Soc. Am.*, 91:3257–3277, 1992.
34. G. L. D'Spain, J. J. Murray, W. S. Hodkiss, N. O. Booth, and P. W. Schey. Mirages in shallow water matched field processing. *J. Acoust. Soc. Am.*, 105:3245–3265, 1999.
35. B. Edde. *Radar: Principles, Technology, Applications*. Prentice-Hall, New York, 1993.
36. C. Elachi. *Spaceborne Radar Remote Sensing: Applications and Techniques*. IEEE, New York, 1987.
37. M. Fink. Time reversal mirrors. *J. Phys. D*, 26:1330–1350, 1993.
38. M. Fink. Time reversed acoustics. *Phys. Today*, 50:34–40, 1997.
39. M. Fink, D. Cassereau, A. Derode, C. Prada, P. Roux, and M. Tanter. Time-reversed acoustics. *Rep. Prog. Phys.*, 63:1933–1994, 2000.
40. J. P. Fouque, S. Garnier, A. Nachbin, and K. Solna. Time reversal refocusing for a point source in randomly layered media. submitted.
41. J. P. Fouque and K. Solna. Time reversal aperture enhancement. *SIAM Multiscale Model. Simul.*, 1(2):239–259, 2003.
42. J. Garnier. Imaging in randomly layered media by cross-correlating noisy signals. submitted.
43. R. Gens and J. L. van Gendewren. SAR interferometry - issues, techniques, applications. *International Journal of Remote Sensing*, 17(10):1803–1835, 1996.

44. M.S. Gockenbach. *An abstract analysis of differential semblance optimization*. PhD thesis, Rice University, 1994.
45. A. Ishimaru. *Wave Propagation and Scattering in Random Media*. Academic Press, New York, 1978.
46. A. Ishimaru. *Wave propagation and scattering in random media*. Academic Press, New York, 1978.
47. F. C. Karal Jr. and J. B. Keller. Elastic, electromagnetic, and other waves in a random medium. *J. Math. Phys.*, 5:537–547, 1964.
48. J. B. Keller. Stochastic equations and wave propagation in random media. In *16th Symposium in Applied Mathematics, New York, 1963*, pages 145–170, Providence, Rhode Island, 1964. American Mathematical Society.
49. J. L. Krolik. Matched field minimum-variance beamforming in random ocean channel. *J. Acoust. Soc. Am.*, 92:1408–1419, 1992.
50. W. A. Kuperman, W. S. Hodgkiss, H. C. Song, T. Akal, C. Ferla, and D. R. Jackson. Phase conjugation in the ocean : Experimental demonstration of an acoustic time reversal mirror. *J. Acoust. Soc. Am.*, 103:25–40, 1998.
51. D. Massonnet. Satellite radar interferometry. *Scientific American*, 36:46–53, February 1997.
52. D. Massonnet and K. L. Feigl. Radar interferometry and its application to changes in the earth's surface. *Reviews of Geophysics*, 36:441–500, 1998.
53. C. Nolan and M. Cheney. Synthetic aperture inversion. *Inverse Problems*, 18(1):221–235, 2002.
54. C. Nolan and W. Symes. Global solution of a linearized inverse problem for the wave equation. *Comm. Partial Differential Equations*, 22(5-6):919–952, 1997.
55. R. F. O'Doherty and N. A. Anstey. Reflections on amplitudes. *Geophys. Prospecting*, 19:430–458, 1971.
56. B. Oksendal. *Stochastic differential equations. An introduction with applications*. Springer-Verlag, Berlin, second edition, 2000.
57. G. Papanicolaou, M. Postel, P. Sheng, and B. White. Frequency content of randomly scattered signals ii: Inversion. *Wave Motion*, 12:527–549, 1990.
58. G. Papanicolaou, L. Ryzhik, and K. Solna. Statistical stability in time reversal. *SIAM J. Applied Mathematics*, 64(4):1133–1155, 2004.
59. G. C. Papanicolaou, D. Stroock, and S. R. S. Varadhan. Martingale approach to some limit theorems. In D. Ruelle, editor, *Statistical Mechanics, Dynamical Systems and the Duke Turbulence Conference*, volume 3 of *Duke Univ. Math. Series*, pages ii+120, Durham, NC, 1977.
60. G. C. Papanicolaou and S. R. S. Varadhan. A limit theorem with strong mixing in Banach space and two applications to stochastic differential equations. *Comm. Pure Appl. Math.*, 1973:497–524, 26.
61. M. B. Priestley. *Spectral analysis and time series*. Academic Press, London, New York, 1981.
62. Rakesh. A linearised inverse problem for the wave equation. *Comm. Partial Differential Equations*, 13(5):573–601, 1988.
63. W. A. Schneider. Integral formulation for migration in two and three dimensions. *Geophysics*, 43:49–76, 1978.
64. G. T. Schuster, J. Yu, J. Sheng, and J. Rickett. Interferometric/daylight seismic imaging. *Geophys. J. Int.*, 157:838–852, 2004.
65. K. Solna. Focusing of time-reversed reflections. *Waves Random Media*, 12(3):365–385, 2002.
66. K. Solna and G. Papanicolaou. Ray theory for a locally layered random medium. *Waves Random Media*, 10:151–198, 2000.
67. H. C. Song, W. A. Kuperman, and W. S. Hodgkiss. Iterative time reversal in the ocean. *J. Acoust. Soc. Am.*, 105:3176–3184, 1999.
68. C.C. Stolk and W. Symes. Smooth objective functionals for seismic velocity inversion. *Inverse Problems*, 19(1):73–89, 2003.
69. W. Symes. Layered velocity inversion: a model problem from reflection seismology. *SIAM J. Math. Anal.*, 22(3):680–716, 1991.
70. W. Symes. A differential semblance criterion for inversion of multioffset seismic reflection data. *J. Geophys. Res.*, 88, 1993.
71. W. Symes. Lecture notes in seismic imaging. Mathematical Geophysics Summer School, Stanford, available at [www.trip.caam.rice.edu](http://www.trip.caam.rice.edu), 1998.

72. C. Tsogka and G. Papanicolaou. Time reversal through a solid-liquid interface and super-resolution. *Inverse Problems*, 18:1639–1657, 2002.
73. H. Zebker. Studying the earth with interferometric radar. *Computing in Science and Engineering*, 2(3):52–60, 2000.

COMPUTATIONAL AND APPLIED MATHEMATICS, MS 134, RICE UNIVERSITY, 6100 MAIN ST., HOUSTON, TX 77005-1892

*E-mail address:* borcea@caam.rice.edu

# Index

- Algorithm
  - Decomposition, 117
  - Donoho and Johnstone denoising, 122
  - Electrical impedance reconstruction algorithm, 82
  - Filtered backprojection algorithm, 29, 60
  - Reconstruction, 117
- Ampère's law, 98
- Approximate Inverse, 26
- Attenuated Radon transform ( $T_\mu f(\omega, s)$ ), 75
  - Inversion, 79
  - Range conditions, 80
  - Recovery of attenuation, 80
- Band-limited function, 113
- Circular Radon transform ( $Rf(p, r)$ ), 68
  - Inversion, 72
  - Range, 73
  - Set of injectivity, 69
  - Stability of reconstruction, 72
- Coherent interferometric function ( $\mathcal{I}^{\text{CINT}}$ ), 138
- Cone beam transform ( $Df(a, \theta)$ ), 38
  - Inversion formula, 39
- Decoherent
  - Frequency ( $\Omega_d$ ), 139
  - Length ( $X_d$ ), 139
- Discrete (multi)wavelet transform ( $\mathcal{W}$ ), 121
- Electrical Impedance Tomography, 81
  - Reconstruction algorithm, 82
  - Stability, 82
- Emission tomography, 75
- Exponential Radon transform ( $R_\mu f(\omega, s)$ ), 76
  - Inversion formula, 77
  - Range conditions, 79
- Faraday's law of induction, 98
- Filter coefficient matrices ( $\mathbf{P}(k)$ ), 116
- Filtered back projection
  - Algorithm, 29, 60
  - and approximate inverse, 29
  - Inversion formula, 6
- Fourier transform ( $\mathcal{F}f$  or  $\hat{f}$ ), 4
  - Partial Fourier transform ( $\mathcal{F}_s g$ ), 4
- Gauss law, 98
- Hausdorff metric ( $\delta$ ), 111
- $\mathcal{H}_{\omega, \kappa}$ , 144
- Hölder Space ( $C^{n,s}(\Omega)$ ), 126
- Integral equation of the first kind, 108
- $K(\theta, b)$ , 50
- Kirchhoff migration
  - Function ( $\mathcal{I}^{\text{KM}}$ ), 135
  - In homogeneous medium, 134
  - In clutter, 137
- Lambda tomography, 13
- Lattice
  - Sampling ( $L(N, P, Q)$ ), 47, 48
  - Reciprocal lattice ( $\mathbf{L}_S^\perp$ ), 50
- Limited angle Radon transform ( $R_\phi(\theta, s)$ ), 32
  - Limited angle reconstruction kernel, 36
  - Limited angle region of interest tomography, 14
- Limited data tomography, 10
  - Exterior tomography, 11
  - Region of interest tomography, 13
- Line ( $L(\varphi, s)$ ), 3
- Linear acoustic equation, 133
- Lipschitz function, 126
- Magnetic flux leakage, 101
- Magnetic flux leakage field ( $\mathbf{B}_\Sigma$ ), 101
- Matrix dilation equation, 116
- Maxwell's Equations, 98
- Microlocal analysis, 7, 15
- Multiresolution analysis, 115
- Multiwavelet, 115
- Multiwavelet system, 116

- O'Doherty Anstey,  
Formula, 149  
Kernel ( $K_{\text{ODA}}$ ), 148
- Poisson equation, 99
- Poisson summation formula
  - For the line, 46
  - For the 2-dim torus, 49
- Problem
  - ill-posed, 107
  - inverse, 106
  - well-posed, 107
- Propagator matrix ( $\mathcal{P}^\epsilon(\omega, \kappa, z; z_0)$ ), 144
- Radon Transform
  - as a Fourier integral operator, 17
  - Dual Radon transform ( $R^*g(x)$ ), 3
  - Fourier slice theorem, 4
  - in  $\mathbb{R}^2$  ( $Rf(\varphi, s)$ ), 3
  - in  $\mathbb{R}^n$  ( $\mathbf{R}f(\theta, s)$ ), 28
  - Limited angle Radon transform  
( $R_\phi(\theta, s)$ ), 28
  - Microlocal regularity theorem, 8
  - Projection slice theorem, 4, 45
  - Range theorem, 5
- Sampling
  - Classical
    - (Whittaker-Shannon-Kotel'nikov)
    - Theorem, 46, 113
  - Fan-beam geometry, 45
  - Parallel-beam geometry, 45
  - Periodic sampling set, 62
- Scaling vector, 115
- Sobolev space ( $H^\alpha$  or  $W^{p,q}(\Omega)$ ), 7, 126
- Sobolev Wavefront set ( $WF^\alpha$ ), 7
- Time reversal functional ( $\mathcal{I}^{\text{TR}}$ ), 137
- Thermoacoustic Tomography, 68
- Undersampling, 57
- X-ray tomography, 2, 44

## **Titles in This Series**

- 63 **Gestur Ólafsson and Eric Todd Quinto, Editors**, The radon transform, inverse problems, and tomography (Atlanta, Georgia, January 2005)
- 62 **Paul Garrett and Daniel Lieman, Editors**, Public-key cryptography (Baltimore, Maryland, January 2003)
- 61 **Serkan Hoşten, Jon Lee, and Rekha R. Thomas, Editors**, Trends in optimization (Phoenix, Arizona, January 2004)
- 60 **Susan G. Williams, Editor**, Symbolic dynamics and its applications (San Diego, California, January 2002)
- 59 **James Sneyd, Editor**, An introduction to mathematical modeling in physiology, cell biology, and immunology (New Orleans, Louisiana, January 2001)
- 58 **Samuel J. Lomonaco, Jr., Editor**, Quantum computation: A grand mathematical challenge for the twenty-first century and the millennium (Washington, DC, January 2000)
- 57 **David C. Heath and Glen Swindle, Editors**, Introduction to mathematical finance (San Diego, California, January 1997)
- 56 **Jane Cronin and Robert E. O'Malley, Jr., Editors**, Analyzing multiscale phenomena using singular perturbation methods (Baltimore, Maryland, January 1998)
- 55 **Frederick Hoffman, Editor**, Mathematical aspects of artificial intelligence (Orlando, Florida, January 1996)
- 54 **Renato Spigler and Stephanos Venakides, Editors**, Recent advances in partial differential equations (Venice, Italy, June 1996)
- 53 **David A. Cox and Bernd Sturmfels, Editors**, Applications of computational algebraic geometry (San Diego, California, January 1997)
- 52 **V. Mandrekar and P. R. Masani, Editors**, Proceedings of the Norbert Wiener Centenary Congress, 1994 (East Lansing, Michigan, 1994)
- 51 **Louis H. Kauffman, Editor**, The interface of knots and physics (San Francisco, California, January 1995)
- 50 **Robert Calderbank, Editor**, Different aspects of coding theory (San Francisco, California, January 1995)
- 49 **Robert L. Devaney, Editor**, Complex dynamical systems: The mathematics behind the Mandelbrot and Julia sets (Cincinnati, Ohio, January 1994)
- 48 **Walter Gautschi, Editor**, Mathematics of Computation 1943–1993: A half century of computational mathematics (Vancouver, British Columbia, August 1993)
- 47 **Ingrid Daubechies, Editor**, Different perspectives on wavelets (San Antonio, Texas, January 1993)
- 46 **Stefan A. Burr, Editor**, The unreasonable effectiveness of number theory (Orono, Maine, August 1991)
- 45 **De Witt L. Sumners, Editor**, New scientific applications of geometry and topology (Baltimore, Maryland, January 1992)
- 44 **Béla Bollobás, Editor**, Probabilistic combinatorics and its applications (San Francisco, California, January 1991)
- 43 **Richard K. Guy, Editor**, Combinatorial games (Columbus, Ohio, August 1990)
- 42 **C. Pomerance, Editor**, Cryptology and computational number theory (Boulder, Colorado, August 1989)
- 41 **R. W. Brockett, Editor**, Robotics (Louisville, Kentucky, January 1990)
- 40 **Charles R. Johnson, Editor**, Matrix theory and applications (Phoenix, Arizona, January 1989)
- 39 **Robert L. Devaney and Linda Keen, Editors**, Chaos and fractals: The mathematics behind the computer graphics (Providence, Rhode Island, August 1988)
- 38 **Juris Hartmanis, Editor**, Computational complexity theory (Atlanta, Georgia, January 1988)

TITLES IN THIS SERIES

- 37 **Henry J. Landau, Editor**, Moments in mathematics (San Antonio, Texas, January 1987)
- 36 **Carl de Boor, Editor**, Approximation theory (New Orleans, Louisiana, January 1986)
- 35 **Harry H. Panjer, Editor**, Actuarial mathematics (Laramie, Wyoming, August 1985)
- 34 **Michael Anshel and William Gewirtz, Editors**, Mathematics of information processing (Louisville, Kentucky, January 1984)
- 33 **H. Peyton Young, Editor**, Fair allocation (Anaheim, California, January 1985)
- 32 **R. W. McKelvey, Editor**, Environmental and natural resource mathematics (Eugene, Oregon, August 1984)
- 31 **B. Gopinath, Editor**, Computer communications (Denver, Colorado, January 1983)
- 30 **Simon A. Levin, Editor**, Population biology (Albany, New York, August 1983)
- 29 **R. A. DeMillo, G. I. Davida, D. P. Dobkin, M. A. Harrison, and R. J. Lipton**, Applied cryptology, cryptographic protocols, and computer security models (San Francisco, California, January 1981)
- 28 **R. Gnanadesikan, Editor**, Statistical data analysis (Toronto, Ontario, August 1982)
- 27 **L. A. Shepp, Editor**, Computed tomography (Cincinnati, Ohio, January 1982)
- 26 **S. A. Burr, Editor**, The mathematics of networks (Pittsburgh, Pennsylvania, August 1981)
- 25 **S. I. Gass, Editor**, Operations research: mathematics and models (Duluth, Minnesota, August 1979)
- 24 **W. F. Lucas, Editor**, Game theory and its applications (Biloxi, Mississippi, January 1979)
- 23 **R. V. Hogg, Editor**, Modern statistics: Methods and applications (San Antonio, Texas, January 1980)
- 22 **G. H. Golub and J. Oliker, Editors**, Numerical analysis (Atlanta, Georgia, January 1978)
- 21 **P. D. Lax, Editor**, Mathematical aspects of production and distribution of energy (San Antonio, Texas, January 1976)
- 20 **J. P. LaSalle, Editor**, The influence of computing on mathematical research and education (University of Montana, August 1973)
- 19 **J. T. Schwartz, Editor**, Mathematical aspects of computer science (New York City, April 1966)
- 18 **H. Grad, Editor**, Magneto-fluid and plasma dynamics (New York City, April 1965)
- 17 **R. Finn, Editor**, Applications of nonlinear partial differential equations in mathematical physics (New York City, April 1964)
- 16 **R. Bellman, Editor**, Stochastic processes in mathematical physics and engineering (New York City, April 1963)
- 15 **N. C. Metropolis, A. H. Taub, J. Todd, and C. B. Tompkins, Editors**, Experimental arithmetic, high speed computing, and mathematics (Atlantic City and Chicago, April 1962)
- 14 **R. Bellman, Editor**, Mathematical problems in the biological sciences (New York City, April 1961)
- 13 **R. Bellman, G. Birkhoff, and C. C. Lin, Editors**, Hydrodynamic instability (New York City, April 1960)
- 12 **R. Jakobson, Editor**, Structure of language and its mathematical aspects (New York City, April 1960)

For a complete list of titles in this series, visit the  
AMS Bookstore at [www.ams.org/bookstore/](http://www.ams.org/bookstore/).