TOWARDS QUANTITATIVE COMPUTED TOMOGRAPHY

S. A. Enright (Hons 1)

A thesis presented for the degree of
Doctor of Philosophy
in
Electrical and Electronic Engineering
at the
University of Canterbury,
Christchurch, New Zealand.

April 1992

ABSTRACT

Computed tomography is introduced along with an overview of its diverse applications in many scientific endeavours. A unified approach for the treatment of scattering from linear scalar wave motion is introduced. The assumptions under which wave motion within a medium can be characterised by concourses of rays are presented along with comment on the validity of these assumptions. Early and conventional theory applied for modelling the behaviour of rays, within media for which ray assumptions are valid, are reviewed.

A new computerised method is described for reconstruction of a refractive index distribution from time-of-flight measurements of radiation/waves passing through the distribution and taken on a known boundary surrounding it. The reconstruction method, aimed at solving the bent-ray computed tomography (CT) problem, is based on a novel ray description which doesn't require the ray paths to be known. This allows the refractive index to be found by iterative solution of a set of linear equations, rather than through the computationally intensive procedure of ray tracing, which normally accompanies iterative solutions to problems of this type. The preliminary results show that this method is capable of handling appreciable spatial refractive index variations in large bodies.

A review containing theory and techniques for image reconstruction from projections is presented, along with their historical development. The mathematical derivation of a recently-developed reconstruction technique, the *method of linograms* is considered.

An idea, termed the *plethora of views* idea, which aims to improve quantitative CT image reconstruction, is introduced. The theoretical foundation for this is the idea that when presented with a *plethora* of projections, by which is meant a number greater than that required to reconstruct the known region of support of an image, so that the permissible reconstruction region can be extended, then the intensity of the reconstructed distribution should be negligible throughout the extended region. Any reconstruction within the extended region, that departs from what would be termed negligible, is deduced to have been caused by imperfections of the projections. The implicit expectation of novel schemes which are presented for improving CT image reconstruction, is that contributions within the extended region can be utilised to ameliorate the effects of the imperfections on the reconstruction where the distribution is known to be contained. Preliminary experimental results are reported for an iterative algorithm proposed to correct a plethora of X-ray CT projection data containing imperfections. An extended definition is presented for the *consistency* of projections, termed *spatial consistency*, that incorporates the region with which the projection data is consistent. Using this definition and an associated definition, *spatial inconsistency*, an original technique is proposed and reported on for the recovery of inconsistencies that are contained in the projection data over a narrow range of angles.

ACKNOWLEDGEMENTS

The effort and time involved in the production of this thesis has been extensive but would have been much more onerous if it were not for the support and help of my supervisor, Dr K. L. Garden, who throughout has been cheerful, generous and whose friendship has certainly been appreciated. All this in spite of two young nippers in tow.

I would have sincerley wished to thank the late Professor R.H.T. Bates who did inspire me. I can only sit back in awe at his dedication within the many disciplines into which he delved. Without his help the original material in this thesis could not have been.

I greatly appreciate the love afforded to me by Margaret-Mary who, firstly as my friend, and then my wife has brought about the final motivation needed by me to bridge the uncertainties I had in this document. She has endured with her time, effort, support, understanding and a host of other affections with little thanks. I hope to repay over time the dedication and security given me.

Collaborators of whom I am most grateful are Dr Ross Murch, Dr Susan Dale and Vaughan Smith. Particularly to Ross and Susan who provided the best mix of collaboration that I could have asked for.

Paddy, Josie, Tony, Vicky, Wade, Stephanie, David David and Joe, an extra family providing an environment where I've always found fun and friendship ... and that touch of "St Bathans" madness.

I thank my colleagues (academic staff, technical staff and fellow students alike) at the University of Canterbury for their help or friendship, particularly Charles and Martin. A special thanks to Bob Young for his help under dire mechanical circumstances. Jan, Margaret, Maureen, Wilma and Bev for the many morning teas that have "made" the start of the day.

My flatmates Micheal, Bridget and Louise are commendable in their efforts to put up with me and their friendship is valued. I can be thankful, although not always, to those diversionaries such as Kev, Quinton, the Pisstons, Kirwee Seniors, who have all helped in keeping my feet on the ground. For those constantly expounding the virtues of the IRD, Tim & Bricky, may I frequent that department more than they.

Finally, to my father : simply thankyou. For the period of my PHD he has certainly motivated and helped behind the scenes in his usual manner, transparent to all others, but not transparent to me when it came to advice and support in achieving this thesis. For a father this of course is an understatement.

PREFACE

This thesis is composed of 8 chapters in all. The next 8 paragraphs overview the content and purpose of each chapter.

Chapter 1 serves to introduce the subject of computed tomography (CT). It does so by evoking for the reader an appreciation of some important applications, where CT techniques have been firmly embedded as routine practice because they have ensured, and continue to facilitate, the effectiveness of each application. Through this evocation it is anticipated that the reader may absorb the motivation for the development of computed tomography and the reasons why it is such an important field of study. Intuitively, such an important field must be understood well so that its full potential is realised and its quantitative information-yielding capacities are optimised. Computed tomography is an example of the more general class of problems known as inverse problems. Consequently, computed tomography utilises many of the techniques known in the context of inverse problems as well as direct problems. The philosophy of direct and inverse problems is therefore developed by introducing them in a relatively formal manner along with the terminology, notation and geometrical descriptions used in defining them. The philosophy behind an inverse problem includes the probing and sensing of objects in order to determine the properties of the object. The type of probing emanation used will determine the physical properties under study. A selection of probing emanations are listed along with the physical properties interacting with each and interpretations that are made as a result of observing scattered emanations. Two important CT imaging modalities that use emanations with very different properties, X-ray CT and ultrasound CT, are then described in more detail as the original contributions in this thesis are directly applicable to these modalities. Much of the motivation for the original research content follows in section 1.5 which aptly shares the title of this thesis.

Many physical processes manifest themselves under the regime of linear scalar wave motion in a number of scientifically and technologically important situations. The underlying equations which model linear scalar wave motion, known as wave equations, are given in chapter 2. These equations are in the form of linear partial differential equations that are regarded as being dependent upon either time or frequency. Both forms, thought to be theoretically equivalent and able to be related directly through the Fourier transform, each have their own advantages. The first form of wave equation given is known as the temporal wave equation since the wave motion is considered a function of time and is able to model transient wave motions. The second, regarded as the most popular form in which solutions for wave motion are derived, considers the wave motion as simple harmonic components. These physical models provide unified solutions to direct and inverse problems. A high-frequency approximation for wave motion is considered also, as much of the original work in this thesis is predicated upon such an approximation. Details of this high-frequency model for wave motion, known as ray theory, is deferred to chapter 3 to give emphasis to the ray model and classical ray tomography. Comment is made on the class of approximate techniques for solving the wave equation, known as *diffraction tomography*. Supplementary mathematics and definitions used to describe functions and images are also introduced.

As implied above, chapter 3 is concerned with the treatment of wave motion in the context of ray theory. The earliest models were those assumed for visible light where it was thought that light travelled as straight rays. Much later, radio/radar engineers, acousticians and quantum physicists developed methods whereby wave motions interacting with various types of media behave in the manner intuitively ascribed to rays. These rays obey Snell's law of refraction with respect to variations in the refractive index of media [Bates *et al.*, 1991]. Later the theory of rays was developed to account for traversing inhomogeneous media by relating the ray path explicitly to intrinsic properties of the medium, such as the refractive index or velocity distribution therein. Chapter 3 gives an account of these ray models as well as a solution to the direct problem, known as *ray tracing*, that is in prevalent use for inverse methods based on ray models. Lastly, the *geometrical theory of diffraction* is included to complete a review of classical theory developed on the ray model. Comment is made on the validity of the ray approach in the context of solving direct and inverse problems.

Chapter 4 introduces an original technique for solving the inverse problem known as the bent-ray CT problem. The general problem of bent-ray CT is presented for the quantitative reconstruction of a refractive index distribution. A reconstruction algorithm is proposed for bent-ray CT. The reconstruction method is based on a novel ray description which doesn't require the ray paths to be known. This allows the refractive index to be found by iterative solution of linear equations, rather than through the computationally intensive procedure of ray tracing, which normally accompanies iterative solutions to problems of this type. The more common problem of bent-ray transmission CT is considered and a more detailed algorithm is suggested for this problem. An approach for reconstructing circularly symmetric refractive index distributions is given along with results obtained by applying the algorithm for bent-ray transmission CT to reconstruct such distributions. Preliminary results show that this method is capable of handling appreciable spatial refractive index variations in large bodies.

Chapter 5 reviews much of X-ray CT image reconstruction. Starting with the original problem as given by Radon in 1917, it develops the various mathematical preliminaries relating to methods invented for the solution to the X-ray CT problem. The majority of solutions based on an inversion of the Radon transform are reviewed, along with their historical development. Of late, there have been few newly developed solutions for the problem of reconstruction from projections. One of the most recently developed solutions known as the *method of linograms* [Edholm *et al.*, 1987] is described in some detail. Edholm *et al.* [1988] found the method of linograms to be superior to that of the popular modified back-projection method from the point of view of computational speed and equivalent from the point of view of accuracy. This new method is a novel reconstruction technique and employs a scanning geometry of its own that is particularly convenient in the context of many industrial situations. The reconstruction method is developed in this chapter through the modified back-projection method and as such it can clearly be seen to be applicable to the reconstruction of data obtained from conventional scanning geometries. Although the method of linograms is a novel technique that can be introduced in an intuitive manner, its mathematical development from modified back-projection is not easily obtained and is therefore explained in some detail.

In many CT imaging systems there are often the facilities available to provide a plethora of projection data for the reconstruction of a spatial distribution. It has been established [Crowther *et al.*, 1970] for some time, the number of projections needed to reconstruct an image faithfully, to a desired resolution, is proportional to the radius circumscribing the known support of the distribution. Some unification of this fundamental principle for computed tomography reconstruction from finite projections is provided in section 6.3. If a plethora of projections are given, then the reconstruction radius can be extended beyond that of the support region. Since the true distribution is confined by definition to its support, when presented with a plethora of *ideal* projections, by which is meant a superfluity of projection data that have been obtained by emanations which can be accurately characterised by concourses of straight rays, then contributions outside the reconstruction region should be neglegible. However, when the wave motions employed exhibit appreciable curvature, as is the case for ultrasound, reconstructions need not obey such general rules as to be neglegible outside the region of support. Or even if projection data formed from ideal projections contains imperfections, such as those due to the finite accuracy of any measurement apparatus, then the reconstruction outside the support region may contain contributions that are recognised as being solely due to the imperfections in the data. It is the conjecture of chapter 6 that these contributions can be used to correct the given data, forcing them to be ideal or at least to ameliorate the effects of the imperfections, thereby reducing the contributions to be neglegible. Chapter 6 introduces this *plethora of views* idea as an open problem and examines an approach for correction of non-linear imperfections contained on X-ray CT projection data when reconstructing low-contrast tumours within a human phantom. It is first found necessary to investigate the contamination levels that can be expected from a single image detail when it is reconstructed from a finite number of projections. An expression is given for the normalised radius, from the image detail, from which contamination can be expected to be below a given level when reconstructing from a finite number, M, of projections. An iterative algorithm for image improvement is then developed and tested. Although successful in reducing the contribution of projection imperfections outside the support region, the algorithm is not effective in improving the image within its known support. As a result, an alternative approach to removing the imperfections was explored, and is the topic of chapter 7.

In an attempt to more intuitively reflect the relationship between imperfections of the projections, and a strategy for ameliorating their effects on the reconstructed image, the *consistency* of projection data is examined in chapter 7, and an extended definition presented, termed *spatial consistency*, which combines the *consistency* of projections and the region with which they are consistent. In conjunction a definition of *spatial inconsistency* is also presented. Using an inversion technique for the Radon transform and the aforementioned definition of spatial inconsistency, inconsistencies in the projection data are separated in the form of an expression from which they can be evaluated. An original approach is presented to recover inconsistencies from given imperfect projections, and is tested using computer-simulated data.

Finally, chapter 8 presents conclusions on, and suggests lines of further research into, the original contributions of this thesis.

Publications

The following publications were prepared during the course of this research :

S. A. Enright, S. M. Dale, V. A. Smith, R. D. Murch, and R. H. T. Bates. Towards solving the bent-ray tomographic problem. *Inverse Problems*, Vol. 8, No. 1, pp. 83–94, February 1992.

S. A. Enright, K. L. Garden, G. T. Herman, and R. H. T. Bates. Image reconstruction from projections XI : Reconstruction from a plethora of imperfect projections. to be submitted to *Optik*.

CONTENTS

CHAPTER 1

INTRODUCTION TO COMPUTED TOMOGRAPHY

In this chapter an overview is made of some applications that have been, or are now, important consequences of the development of *computed tomography* (CT). Once an appreciation of the applications for CT has been evoked, the reader may relate better to a more formal description, which follows, of the theoretical and algorithmic framework which supports these applications. The motivation for the specific work which is the subject of this thesis is then developed in section 1.5.

1.1 SETTING THE SCENE FOR CT

Of all the applications of CT, probably the greatest effect on the world at large has been in the area of diagnostic medicine [Herman, 1980].

Medical imaging techniques, of which CT is a major modality, allow the clinician to perceive the human body in a non-invasive manner. Without physically intervening with the human organs under study, CT allows observations to be made whilst the body is largely undisturbed from its normal morphology and physiology [Robb, 1985a]. The first X-ray tomographic machines built were for scanning the head [Hounsfield, 1973] which greatly decreased the need for the dangerous and uncomfortable procedures that were previously required to image the brain [Axel *et al.*, 1983].

From remote measurements taken around the body (these measurements are described in more detail in section 1.4) a digital computer employing a CT algorithm can compute detailed images of the body's internal structure. Once the images are computed and stored in digital format the computer can be programmed for display, or to perform a number of image processing algorithms to enhance the display. Typically, these include enhancement of low contrast structures or the segmentation of particular organs for independent display and analysis. Computer-displayed images (two or three dimensional, 2-D or 3-D) allow the clinician to delve into the anatomical structures and biophysical processes not normally accessible by the eye [Garden *et al.*, 1989].

Conventional X-ray CT scanners collect data associated with a single cross-section of an object, and provide an unambiguous and detailed view of the cross section. If required, 3-D structures can be reconstructed by "stacking" together adjacent cross-sectional slices each imaged separately.

However, it has been found that images formed from a collection of adjacent cross sections are not suitable for all imaging applications (such as the heart [McCann *et al.*, 1988]) and in some cases a fully 3-D scan is a more effective means of collecting data for a volume display. For instance, Robb *et al.* [1983] found that in the imaging of moving organs, such as the heart, from a collection of adjacent cross-sections scanned independently, images were blurred due to the heart's continual motion, which prompted the development of a high speed synchronous scanner capable of fast reconstruction of fully 3-D images. In the pre-operative planning for surgery on a human skull it can be difficult to comprehend the 3-D appearance of the skull from a sequence of 2-D cross-sections [Herman, 1986a] and so a 3-D representation is very useful (although in this case there is no requirement that the data for all sections be obtained simultaneously). Three-dimensional imaging allows surgeons to see what they will find before opening up a patient and has been used for pre-operative surgical planning as well as the simulation of surgical procedures [Peters *et al.*, 1989; Udupa, 1986]. Much work has been done in the area of transforming and displaying 3-D images created by CT scanners. Further comments on 3-D imaging is made in section 2.3.2. Cook *et al.* [1983] illustrate the use of 3-D techniques in the derivation of treatment plans for radiation therapy.

Another rapidly developing image technology in the area of 3-D imaging is that of multi-modal parameter [Jackson *et al.*, 1987; Wood *et al.*, 1979] image analysis. Multi-modal analysis is being used for precise location of physiological functions or pathological conditions. This is particularly necessary in the pre-surgical planning of neurosurgery. Peters *et al.* [1989] explain how multi-modal imaging has been used for the planning of neurosurgery on several hundreds of patients. By combining an imaging modality giving precise morphological information, such as X-ray CT, with an imaging modality that reveals pathological conditions, such as MRI *(magnetic resonance imaging)* [Young *et al.*, 1987], the regions responsible for these conditions can be more precisely defined. Evans *et al.* [1989] have made use of MRI and PET *(positron emission tomography)* images, taken simultaneously, in order to localise brain activity that is responsible for various physiological functions. PET can reveal the relative activity, at different sites, of a functional process by isolating it with a particular radionuclidic *tracer* injected into the blood stream. However, PET images suffer from poor spatial resolution due to a low signal-to-noise ratio in emission measurements. Conversely, brain morphology is revealed much more clearly from MRI images, so that superposition of PET and MRI images can enhance appreciation of the function revealed by the former [Garden *et al.*, 1989]. The ability to combine morphological information with that of ECT *(emission computed tomography)* [Budinger *et al.*, 1979] images, of which PET and SPECT *(single photon emission computed tomography)* are examples, present some exciting advances in understanding biochemical processes within the human body. Much work has been done to develop new radionuclidic tracers to image a variety of metabolic processes involved in human function [Garden *et al.*, 1989; Jaszczak, 1988]. Informative review of ECT can be found in [Jaszczak, 1988; Keyes, 1987]. Multi-parameter image analysis can also facilitate tissue characterisation, and therefore, segmentation of organs and the identification of pathological conditions by using more than one parameter where necessary.

With the need for representing the 3-D structure of bodies, newer data collection modalities have been developed which scan the entire 3-D volume rather than a cross-section. 3-D volume scanning is being used in X-ray CT imaging, magnetic resonance imaging [Udupa, 1986], ultrasound and microwave CT [Garden *et al.*, 1989]. A more recently developed form of gamma camera imaging, *ectomography*, is being used to reconstruct a number of cross-sections from a volume scan data set. Ectomography utilises its own reconstruction algorithm and is most advantageous when imaging sections close and parallel to the surface of the body [Dale *et al.*, 1985].

In conjunction with new data collection modalities, complementary CT algorithms must be developed to generate images of a region with data taken from the new scanning modality. Clearly the variation in scanning requirements of different applications has a direct influence in the development of methods for CT image reconstruction.

Outside the area of medical applications, one of the first practical reconstructions in radio astronomy was by Bracewell in 1956 where he applied the techniques of X-ray CT to determine regions of the sun emitting microwave radiation [Scudder, 1978]. Altschuler [1979] explains how 3-D CT techniques are useful in solar physics for the reconstruction of magnetic field and electron density distributions of the solar corona.

Geophysical computed tomography has become an important research topic because it is capable of determining subsurface structures [Lo *et al.*, 1988] contained within the earth's crust. The safety of highway and building structures depend on the surrounding material upon which they are constructed. Stability of underlying material under adverse conditions such as earthquake and water saturation is determinable if the constituents are known and their relative organisation. A considerable influence in the development of geophysical computed tomography has been the economic impact from oil prospecting and mineral exploration, particularly in seismic methods [Murch, 1990].

1.2 SENSING AN OBJECT REMOTELY

In order to sense an object a physical process must be provided that impinges upon it and some deviation in the process's behaviour detected which would not have occurred had the object been abscent. The physical properties of the object determine the nature of its interaction with impinging processes. Consequently, in sensing the resulting disturbances to the impinging processes it is possible to infer an object's physical properties. For instance, an object placed as a shield in front of the sun will be detected with the naked eye because it obscures the sun's impinging rays from reaching the eyes. Generally in computed tomography problems it is desirable to make quantitative measurements of some physical property the object possesses at points throughout the extent of the object. In order to do this, firstly, the disturbances must be measured quantitatively and have been at least partly caused by the physical property of which it is the intention to measure. Secondly, it must be understood how the physical property interacts with processes impinging upon it.

In sensing of an object remotely, all measurements are made physically apart from, or not invading, any component of the object being measured, as is satisfied by the eyes in the above example. Section 1.1 has outlined some of the advantages for using non-invasive measurement techniques. All problems discussed in this thesis are based upon measurement data that has been recorded "remotely" in the sense intimated above.

1.3 TERMINOLOGY, NOTATION, GEOMETRY AND PROBLEM DEFINITION

To facilitate the discussion of the class of problems covered throughout this thesis it is convenient here to make known a number of definitions common to the discussion of each problem. As is routinely the case for physical problem-solving, it is necessary to adopt here a mathematical framework in which to analyse the problem.

It is convenient to partition K-dimensional space as indicated in figure 1.1(a). When considering wave motion in K-dimensional space, the *scattering region* is denoted by Υ_-, which is finite and enclosed by the surface σ. The scattering region is partitioned by the surface σ_- centered on the origin O, inscribing Υ_-, into Υ_{--} and Υ_{-+}, interior and exterior, respectively to σ_-. The region Υ_+ exterior to Υ_- is also partitioned. The surface σ_+ centered on O, circumscribing Υ_-, partitions Υ_+ into Υ_{+-} and Υ_{++}, which are interior and exterior, respectively to σ_+. The regions Υ_{inc} and Υ_{obs} are where, respectively, the sources of the emanations reside and the emanations are observed. In terms of set notation

$$\Upsilon = \Upsilon_- \cup \sigma \cup \Upsilon_+ \quad \text{and} \quad \Upsilon_- \cap \sigma = \Upsilon_+ \cap \sigma = \emptyset, \tag{1.1}$$

where $\emptyset$ is the empty set, and

$$\Upsilon_- = \Upsilon_{--} \cup \sigma_- \cup \Upsilon_{-+} \quad \text{and} \quad \Upsilon_{--} \cap \sigma_- = \Upsilon_{-+} \cap \sigma_- = \emptyset, \tag{1.2}$$

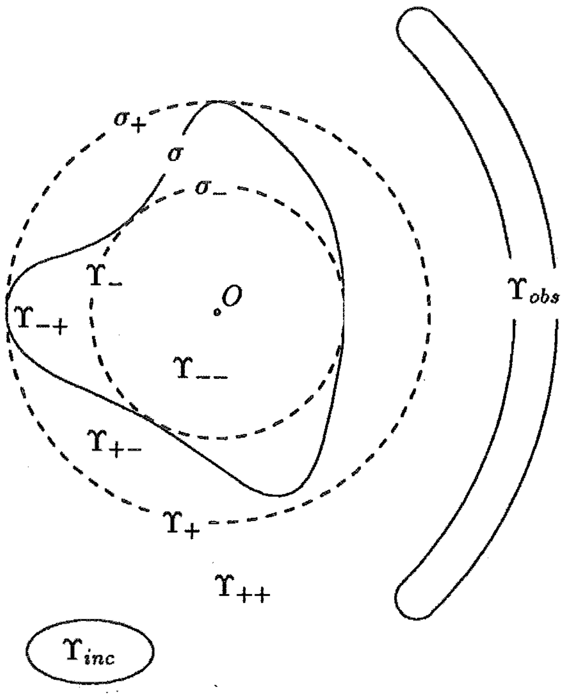
$$\Upsilon_+ = \Upsilon_{+-} \cup \sigma_+ \cup \Upsilon_{++} \quad \text{and} \quad \Upsilon_{+-} \cap \sigma_+ = \Upsilon_{++} \cap \sigma_+ = \emptyset, \tag{1.3}$$

$$\Upsilon_{inc} \subset \Upsilon_{++} \quad \text{and} \quad \Upsilon_{obs} \subset \Upsilon_{++}. \tag{1.4}$$

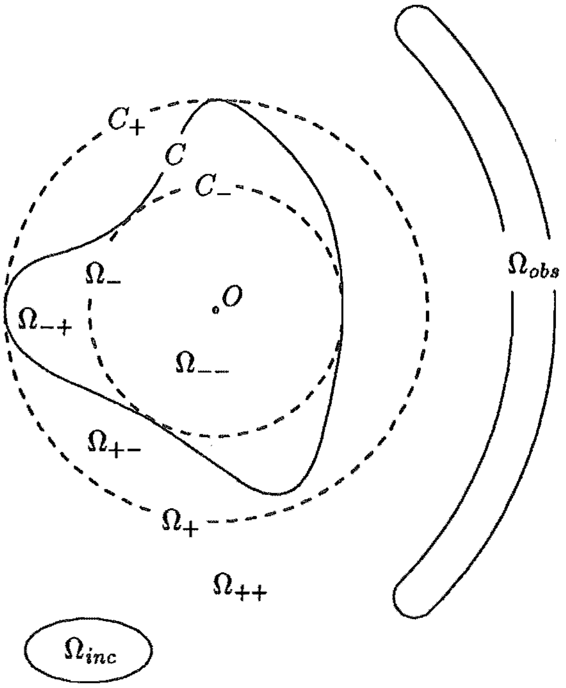
The terminology defined here follows that of Bates *et al.* [1983] and Tan [1988]; the notation and geometry (figure 1.1) are introduced in conjunction with terminology to clarify their physical relationships :

- The *experimenter* is anyone determining the physical properties of an object.
- The *body* represents any object whose physical properties are to be determined. It is alternatively referred to as the scattering region, Υ_-.
- The *emanations* represent any physical process used to study the body from remote positions. The emanations, ψ, are derived from sources contained in the region, Υ_{inc}, which emit emanations, ψ_{inc}, to impinge upon the body. The emanations resulting from the interaction of ψ_{inc} with the body in Υ_-, called the *scattered* emanations, ψ_{scat}, are observed as part of the total field, ψ, in the region Υ_{obs} :

$$\psi = \psi_{inc} + \psi_{scat}. \tag{1.5}$$



(a) K-dimensional space



(b) 2-dimensional space

Figure 1.1 Geometry and notation adopted for the description of the direct and inverse problems. The symbols Υ and σ refer to hyper-volumes and hyper-surfaces respectively in K-dimensions. Symbols Ω and C to planar regions and curves respectively in 2-D.

- The *system* represents the combination of apparatus, measurement scheme, and emanations employed by the experimenter.
- A *sensor* is a device that is employed in the measurement scheme of the system and is located in the observation region, Υ_{obs}, to quantify ψ.
- A *reconstruction* is an estimated spatial distribution of physical properties of the body, usually in Υ_- only, which is calculated from the system and its observations.
- The *generalised constitutive parameter*, χ, represents the true spatial distribution in Υ_- of whatever physical properties of the body the experimenter wishes to reconstruct.
- A *neighbourhood* is the smallest spatial element (volume or cross-section) of a body, that can be resolved by the system.
- A *clean image* of the generalised constitutive parameter is one for which the reconstructed value at any spatial point is (ideally) uncontaminated by values of the true parameter outside the neighbourhood of the point.

The problems examined in this thesis belong to two classes, *direct* problems and *inverse* (synonymous with *imaging* [McKinnon, 1980]) problems. Firstly a definition of these general problems is given, followed by a definition of CT.

The direct problem:
Given the generalised constitutive parameter (χ) in Υ_-, and incident emanations ψ_{inc}, solve for the emanations resulting from the interaction of ψ_{inc} with χ.

If Λ is chosen to model the interaction operator of χ with ψ_{inc}, then the direct problem may be stated as solving for ψ_{scat} from the equation

$$\psi_{scat} = \Lambda(\chi, \psi_{inc}). \tag{1.6}$$

The scattered field ψ_{scat} can be extracted from the total field ψ through (1.5).

The inverse problem:
Given the observations of the scattered emanations made within Υ_{obs}, and knowledge of ψ_{inc}, the inverse problem is to produce a clean image of χ. Two sub-problems can be defined under inverse problems depending on the class of emanation used. Firstly, when the source of emanations is exterior to Υ_- and is controlled by the experimenter, the problem is termed a *remote probing* problem. Secondly, when their source is contained within Υ_- the problem becomes known as a *remote sensing* problem.

The observations depend upon not only the generalised constitutive parameter but also the impinging ψ_{inc} and their interaction with χ, which is modelled by the operator Λ. This inverse problem may be stated as solving for χ the equation

$$\chi = \Lambda^{-1}(\psi_{inc}, \psi_{scat}), \tag{1.7}$$

where Λ^{-1} denotes the inverse operator of Λ.

Computed Tomography (following [Garden, 1984]); is the construction by digital computation of a clean image of the generalised constitutive parameter ($\mathcal{X}$) from observations made of emanations that have interacted with $\mathcal{X}$.

1.4 EMANATIONS IN PRACTICE

There are many types of emanations used in practice to explore objects. This section outlines some of those commonly used and how they are employed. Two imaging modalities, X-ray CT and ultrasound CT, are considered in more detail as they are important applications, especially in clinical use, but which employ emanations with very different characteristics.

In practice there are many imaging modalities available. Some have been developed and are accepted in regular industrial or clinical use, others are still developing or may be too costly, either in computation (or measurement) time or the technology is unaffordable at present [Jaszczak, 1988]. Garden *et al.* [1989] give detailed accounts of many of these. Table table 1.1 lists a selection of these, each having their advantages and disadvantages but because they employ different emanations, source positions that are internal (naturally or introduced) or external, or different reconstruction algorithms, then, they give a variety of imaging interpretations. The reconstruction algorithm implements a solution for the constitutive parameter based on an adopted inversion principle. Examples of inversion principles used for the various emanations can be found in [Bates *et al.*, 1991].

1.4.1 X-ray computed tomography

The discovery of this form of emanation came in 1895 by a German physicist, Wilhelm Conrad Röntgen while working with cathode ray tubes [Robb, 1985b]. "Röntgen's" rays (later nicknamed X-rays) are high energy rays that travel in straight line paths through the objects under analysis, diffraction effects are generally ignored. Throughout the traversal of the body the beam of X-ray photons are attenuated by the intervening material. The X-ray attenuation properties of the material reflect a combination of electron density (reflected in the Compton scattering contribution to attenuation) and atomic number information (reflected in the photoelectric absorbtion component) [Garden *et al.*, 1989]. The photoelectric absorbtion is thus closely related to a quantitative measure of the material's density along the X-ray path through it. Because of the close relationship between X-ray attenuation and density, X-rays are used extensively in conventional radiology [Garrett and Smithson, 1987; Moores, 1987] for the interpretation of organ morphology.

Later, through the independent work of researchers, such as Allan Macleod Cormack and Godfrey Newbold Hounsfield who realised the increase in quantitative information (over conventional radiographic methods) that can be obtained from computed tomograms of X-ray attenuation properties in the human body, numerically stable mathematical techniques as well as the electromechanical and X-ray technology was developed that heralded the formation of modern computed tomography. As a result of years of developing mathematical techniques (similar to those reported on in section 5.6.6), collecting experimental and clinical data, Hounsfield's efforts realised the installation of the first X-ray CT scanner for the head at

Emanations	Constitutive parameters	Interpretation	Imaging modality
X-ray	X-ray attenuation coefficient	Tissue density or Morphology, Pathology	X-ray CT or Radiography
Ultrasound; also Seismic or Acoustic waves	Density, Refractive Index, Attenuation, Compressibility, changes in Acoustic Impedance, Doppler Shift	Reflecting interfaces, Morphology, Blood flow	Ultrasound CT, B-scan, Seismic imaging
Magnetic and Radio frequency fields	Mobile proton density and Relaxation times	Chemical concentration, tissue Morphology, Pathology	Magnetic Resonance imaging
Magnetic or Electric fields	Electric current distribution	Physiology (brain, muscles and heart)	Biomagnetic or Bioelectric imaging
Gamma rays	Concentration of Radioisotopes (introduced)	Physiological processes, Morphology	Radionuclei imaging (PET, SPECT)
Electric potentials	Permittivity, Conductivity	Electrical properties	Electric impedance CT
Microwaves	Dielectric constant, Conductivity	Integrity of tissue, Cell membrane, Water content of cells	Microwave CT
Infra-red radiation	Thermal radiation (IR)	Temperature	Thermography
Spatially Incoherent electromagnetic radiation from microwaves to X-rays	Emission brightness	Concentration of emission generating processes	Astrological imaging

Table 1.1 Relationship between emanations, constitutive parameters and their interpretation for various imaging modalities.

Atkinson Morley's Hospital, England, in 1971 [Robb, 1985c]. In 1972, the first head scans from this machine were reported upon and published in [Hounsfield, 1973; Ambrose, 1973]. In 1979, Hounsfield was a co-recipient of the Nobel Prize with Cormack for their independent and unique contributions to the invention of computed tomography. A description of the events leading up to and including the employment of Hounsfield's work is included in his Nobel lecture, [Hounsfield, 1980]. Even earlier, Cormack had begun work on a mathematical method for uniquely reconstructing quantitative cross-sectional images from their X-ray projections. Cormack's formula for inverting the X-ray projections is detailed in [Cormack, 1963; Cormack, 1964] and section 5.6.2. Since these developments, X-ray CT has rapidly developed for medical and industrial use (as intimated in section 1.1). At present they are a conventional diagnostic tool for hospitals throughout the world.

The combined effects of scattering and absorbtion result in an exponential attenuation of the X-ray beam as it traverses the object. A monoenergetic beam of X-rays, of energy E, with intensity I_0 upon entering the homogeneous object, and passing through a length ds of the object, has an output intensity of

$$I = I_0 \, e^{-\mu \, ds}, \tag{1.8}$$

where $\mu = \mu(E)$ is the energy dependent *X-ray linear attenuation coefficent* of the homogeneous object [Scudder, 1978]. Consider an X-ray beam traversing an inhomogeneous object along a straight line path L, whose linear attenuation coefficient depends upon the position in the object, s, along the beam's path, $\mu = \mu(s, E)$. The attenuation of a monoenergetic beam, energy E, over L is

$$\frac{I(E)}{I_0(E)} = e^{-\int_L \mu(s,E)\, ds}, \tag{1.9}$$

and its projection over path L,

$$\begin{aligned} p_L(E) &= -\ln\left(\frac{I(E)}{I_0(E)}\right) \\ &= \int_L \mu(s, E)\, ds, \end{aligned} \tag{1.10}$$

where ln is the natural logarithm. From (1.10) it can be seen that the monochromatic projection gives the line integral of the objects X-ray linear attenuation coefficient over the path of the beam. In order to provide unambiguous images of the object the projection data needs to be gathered over many different paths throughout the object. The subject of projection requirements is considered in more detail in chapter 5 where the solution to reconstructing a distribution from its projections is reviewed in detail.

Consider now a polychromatic beam with energy dependent intensity $I_0(E)$, the attenuation of this beam is

$$\int \frac{I(E)}{I_0(E)}\, dE = \int \left[e^{-\int_L \mu(s,E) ds}\right] dE. \tag{1.11}$$

Since the attenuation coefficient is a non-linear function with respect to energy, the polychromatic projection data does not provide a line integral of the object's linear attenuation properties, and consequently, a distorted image is obtained. Because lower energy X-ray photons do not penetrate an object as well as those of higher energy, they are considered "softer". The beam then becomes composed of a larger proportion of high energy photons the further the beam penetrates the object. Hence this phenomenon is termed *beam-hardening*. To reduce polychromatic effects such as beam-hardening, the beam is, in cases, prehardened by passing it through a high-pass filter [Scudder, 1978].

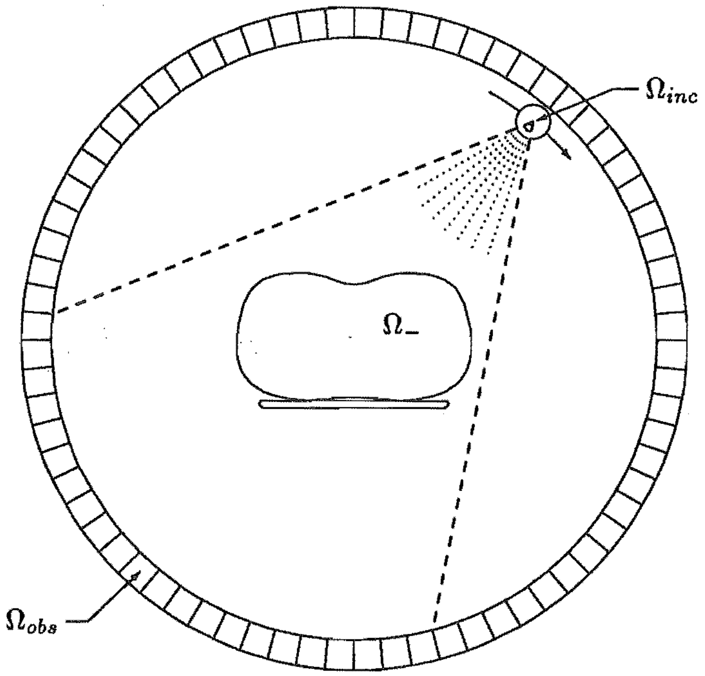


Figure 1.2 A CT scanner for the collection of cross-sectional information. The detectors remain stationary while the X-ray source rotates about the object under study.

Figure 1.2 illustrates a relatively modern geometry for high speed scanning of cross-sectional information. Here the stationary X-ray sensors (contained within Ω_{obs}) completely surround the cross-section being scanned. One or more sources in Ω_{inc} revolve at speed around the circumference. At regular angular positions the source emits a fan-beam of X-rays, illuminating the cross-section Ω_{-}, which are detected by adjacent sensors. A review of X-ray source and scanning geometries can be found in [Webb, 1987].

1.4.2 Ultrasound computed tomography

The exact historical events leading to the development of ultrasound are unclear, but perhaps its beginning should be marked by the discovery of the piezoelectric effect in 1880 and its utilisation in the construction of high frequency mechanical vibrations (ultrasound). Ultrasound was recognised as a suitable emanation for imaging in 1917 when sonar was being invented [Wells, 1977]. Later, in the 1920's it was developed for medical use. Further chronological account of developments relating to ultrasonic imaging can be found in [Erikson *et al.*, 1974;

Wells, 1977]. It was recognised that X-rays and other such ionising radiations can have adverse effects on patient tissue, the effects and their relation to dosage were not fully understood. Coupled with the low cost of ultrasonic imaging equipment, this provided substantial impetus for developing imaging methods with ultrasound as it is not an ionising radiation and has demonstrated relatively few adverse effects on the patient at power levels used for imaging [Herman, 1980]. Adverse effects have been investigated by many authors, details can be found in [Fry, 1979; Wells, 1977]. It should be noted that with the development of ultrasound, appliances are making use of beams with higher energy and/or reduced beamwidth and the effects of heating and cavitation in tissue is becoming of further concern. The remedial effects of heating with higher powered ultrasound are however harnessed by physiotherapists.

One of the most important clinical applications of ultrasound has been in obstetrics. Obstetricians currently make regular practice of monitoring foetal development during pregnancy with ultrasound scanning and newer techniques based on the Doppler principle for measuring blood flow through the umbilical chord make it possible to establish whether the foetus is not being sufficiently supported in this way. Mammographers utilised ultrasound for information in addition to that obtained from conventional X-ray scanning of the breast. The advantage of ultrasound was the low level of adverse effects it had on such a sensitive organ. This technique may prove to be promising for use in mass screening for carcinomas of the female breast [Garden *et al.*, 1989; Glover, 1979]. Various other organs are imaged using ultrasound, such as the brain, liver, kidney, pancreas and eye [Wells, 1981].

Ultrasonic waves are mechanical vibrations with frequencies, generally in the range 1-15 MHz, that lie above the range of human hearing [Wells, 1981]. This range of frequencies has been chosen as a compromise between the resolution of the image and the acceptable attenuation of the signal. The limiting resolution of images obtained is essentially set by the wavelength used. However, it is known that the attenuation of ultrasound increases (approximately linearly) with frequency, and so, limits the maximum frequency appropriate for use [Wells, 1981]. Ultrasound can be considered to be a propagating disturbance of the intrinsic properties of the medium, such as pressure, temperature and particle position, through which it travels. Unlike electromagnetic radiation which can exist in a vacuum, ultrasound must be supported by the medium through which it propagates. There is a complex interaction between these disturbances and the medium's intrinsic properties, such as density, compressibility, refractive index and attenuation [Greenleaf, 1983].

Ultrasound is currently used in several imaging modalities. The most widely used to obtain *qualitative* images is the two-dimensional *B-scan*, providing very useful morphological information [Garden *et al.*, 1989]. A narrow beam of ultrasound, normally generated by a piezoelectric transducer, is reflected from organs inside the body and conveniently received by the same transducer because the piezo-electric effect is reciprocal [Murch, 1990]. The backscatter data is interpreted under the principles of pulse-echo location. Thus, B-scan imaging assumes the ultrasound pulses travel along straight line paths to the reflective boundary and, as such, do not account for refraction and diffraction effects. A phased array of transducers allows the beam direction to be electronically changed at a speed sufficient to construct a 2-D image in real time. The transducers are contained within a hand-held scanning head that can be conveniently positioned around the body. As the images do not require computer processing, imaging is attained in real time. Figure 1.3 illustrates the typical use of B-scan instrumentation for the study of an organ within the body Ω_{-}. The ultrasound source/s and detector/s, which lie in the regions Ω_{inc} and Ω_{obs} respectively, are contained within a handheld unit that is positioned around the circumference of the body.

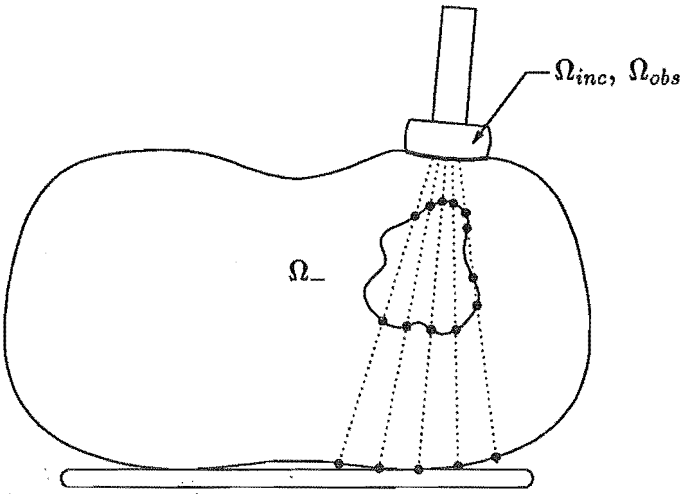


Figure 1.3 Scanning a simple body containing one organ with B-scan instrumentation. The echo image obtained is defined by the solid reflection boundary points lying along the pulse paths (paths are shown as dotted lines).

Quantitative imaging modalities using ultrasound have been confined mainly to transmission computed tomography [Greenleaf, 1983]. One of the first demonstrations of ultrasound transmission tomography was by Greenleaf *et al.* [1974]. Measurements are made of the time-of-flight for pulses to traverse the medium (equivalent to a line integral of the acoustic refractive index) and the attenuation of pulses (equivalent to a line integral of the ultrasound attenuation or absorbtion coefficient). Since the measurements are directly related to the distribution of a physical property of the medium, images obtained can quantify the distribution. From these measurements two separate reconstructions can be obtained : one producing an image of the acoustic refractive index distribution and the other of the acoustic attenuation distribution. In the transmission of the ultrasound wave to the receiver, the intensity is affected by back-scatter within the medium as well as the usual attenuation phenomena, such as absorbtion. Since the back-scatter effect is difficult to separate, ultrasound transmission tomography has often relied on time-of-flight as opposed to attenuation data. However, if back-scatter does not significantly influence measurements, then attenuation images can prove to be very useful. Additionally, the images of acoustic attenuation are affected by refraction and diffraction as the paths taken by the ultrasonic pulse within a medium of inhomogeneous refractive index are curved. Consequently, without knowing the refractive index distribution, accurate predictions of the attenuation path can be unreliable, and therefore, limiting of resolution. A new solution to ultrasonic transmission CT is suggested in chapter 4, for reconstruction of acoustic refractive index distributions which have the attendant problem of refraction in an inhomogeneous medium.

1.5 TOWARDS QUANTITATIVE COMPUTED TOMOGRAPHY

The first medical CT images obtained from X-ray measurements were images of the X-ray attenuation distribution. As intimated in section 1.4.1, this distribution is directly related to the density distribution throughout the body. From these images it is possible to determine the morphology and character of the various organs and their constituents. Although the X-ray attenuation coefficient is a dispersive parameter, and as such, the energy of the X-ray beam must be accounted for, this imaging modality has made it possible to document consistent quantitative properties for both normal and abnormal tissue in organs of the human body that are universally accepted. Since the distribution imaged reveals intrinsic properties of the body, which are independent of machine, operator and any qualitative instrument settings, the distribution is able to be reproduced. The value of obtaining quantitative measurements of intrinsic properties is easy to appreciate from other fields in which the fundamental properties such as density, refractive index and compressibility are used as intrinsic material characteristics, for example in physics, chemistry and non-destructive testing [Greenleaf and Bahn, 1981]. Quantitative computed tomography has the advantage of separating physical parameters allowing their inhomogeneous nature to be evaluated independently [Blackledge *et al.*, 1987a; Blackledge *et al.*, 1987b]. This also facilitates non-destructive testing [Bond and Reynolds, 1987] which is often concerned with the examination of independent physical properties such as density, elastic properties or refractive index.

The *system* (as defined in section 1.3) employed to image a *body* will determine the form and nature of the imaged parameters throughout the scattering region Υ_- (refer figure 1.1(a)). As indicated in table 1.1, the nature of the probing emanation will determine the types of interaction that can occur from a given body, and thus, the form and nature of the constitutive parameters. However, as shown in table 1.1, the interpreted imaged distribution does not necessarily relate directly to the constitutive parameters nor any physical properties the body is known to have, interpretations vary according to the system employed by the experimenter. Ultrasound images, such as those obtained from B-scan methods, are valued for their qualitative images but are lacking in consistent quantitative information. Quantitative information obtained from B-scans is highly dependent on the *system* employed to study the body, and on the calibration of the constituents in the system [Greenleaf and Bahn, 1981].

Morphology of human organs is of fundamental interest to clinicians. Therefore, segmentation of organs is a rudimentary instrument used to enhance their display by isolating the areas of interest. Tissue characterisation and organ segmentation can be delineated on the basis of their intrinsic quantities. Tissue characterisation by numerical delinearisation or thresholding is computationally simple and can be implemented with relatively minimal computational cost [Sahoo *et al.*, 1988]. For example, bone and soft tissue are easily discriminated from an X-ray CT image by their different densities. Much of human morphology that is of interest to the clinician can be determined from 2-D and 3-D X-ray CT images of density. Similarly, organs and organ structure can be delineated on the basis of other intrinsic properties such as the refractive index of the tissue [Greenleaf and Bahn, 1981; Scherzinger *et al.*, 1989].

Pathologists are able to classify disease and/or damage to tissue from intrinsic properties such as density and refractive index. Clinical studies have shown that speed of sound measurements are relatively precise representations of tissue properties [Scherzinger *et al.*, 1989]. Important variables for characterisation of malignant tissues in the breast are intrinsic properties such as speed of sound [Greenleaf and Bahn, 1981; Foster *et al.*, 1984; Scherzinger *et al.*, 1989].

There is considerable interest in computer aided classification of images [Newell, 1988]. In non-destructive testing, quantitative scanning modalities and algorithms are required for automated inspection to analyse the large quantities of data taken from structures [Smith and Martin, 1987]. To increase diagnostic efficiency of human tissue with regard to visual analysis, computer classification is used to identify much of the malignant tissue [Schreiman *et al.*, 1984]. However, in order to construct algorithms for automatic tissue classification that are acceptable to the clinician, a firm knowledge of tissue is required from images that relate well to intrinsic properties. Automated tissue classification has also become important in the area of radiotherapy. Treatment plans can be constructed from a knowledge of the 3-D tissue by reliably identifying regions containing malignant tissue. The treatment planner can then construct an optimal radiation treatment plan. Treatment is considered optimal on the basis of maximising radiation dosage to malignant regions while minimising dosage to normal or sensitive tissue areas.

One of the fundamental limitations to the spatial resolution of ultrasonic imaging (both B-scan, transmission CT imaging and diffraction tomography) has been its inability to cope with appreciable variations in refractive index [Bates *et al.*, 1991; Garden *et al.*, 1989]. For some time this has been a fundamental set-back for tomographic methods based on the presentation of wave motion as pencils of rays, and, where the medium exhibits appreciable refraction. Advocates of a ray model for wave motion have expressed the need for research into algorithms that cope with this elusive problem [Bates *et al.*, 1991; Enright *et al.*, 1992; Greenleaf and Bahn, 1981]. Insufficient knowledge of the wavefront curvature have an effect on imaging of other intrinsic properties. Diffraction tomography techniques, such as those discussed in section 2.2, that are based on extensions to the Born and Rytov approximations, would benefit by an independent method that better estimates the refractive index distribution throughout the scattering region so that the curvature of the wavefronts can be accounted for (see for example, the JWKB/Rayleigh-Gans inverse scattering approximation [Bates *et al.*, 1991; Bates, 1988b]). As intimated in section 1.4.2, ultrasound transmission computed tomography constructing images of the acoustic attenuation or absorbtion distribution would benefit by a better knowledge of the ultrasound path as it propagates through the medium. Therefore, any improvements to techniques for imaging the acoustic refractive index distribution, allowing the ultrasound paths to be estimated, would increase the resolution and accuracy obtained from acoustic attenuation images.

From the point of view of computational cost in solving the inverse problem, what is required are computationally manageable algorithms for reconstructing distributions of refractive index that give rise to appreciable spatial variations in the wavefront curvatures of wave motions traversing such distributions [Bates *et al.*, 1991]. The approach suggested in chapter 4, which is predicated on an iterative solution of non-linear equations by solving linear equations (for which efficient algorithms have existed for some time), should broach this requirement. However, it would be inappropriate to suggest it does so without first testing its computational requirements in reconstructing two-dimensional refractive index distributions. Section 4.7 presents evidence to suggest that there are circumstances where few iterations are required.

At present there remain several sources of image error in the field of X-ray CT. With the possibility of adverse effects on human tissue, it is desirable to reduce dosage levels to a minimum, but still maintain detail and image quality. However, image quality suffers with a reduction in the signal-to-noise ratio that occurs when the number of X-ray photons in the beam is reduced. This is because the reliability of projection measurements are statistically related to the number of photons in use. As intimated in section 1.4.1, X-ray projection data reflect the effects of both photoelectric absorbtion and Compton scattering. It is desirable to separate from attenuation measurements the effect of Compton scattering, and thus, make use of data that is more closely related to a quantitative measure of the material density through which the X-ray beam has traversed. Attempts have been made to resolve these effects by means of dual energy techniques [Garden *et al.*, 1989]. Currently the most serious cause of artefacts in X-ray CT is that of beam-hardening. The phenomenon of beam-hardening is described in section 1.4.1. When severe, it is responsible for "cupping" artefacts and pronounced "streaks" especially close to high density objects such as bones [Garden *et al.*, 1989].

It is suggested in chapter 6 that many effects on reconstruction quality due to measurement errors are inconsistent with known *a priori* information as to the medium under scrutiny. Often the only *a priori* information available is the support σ of Υ_-. Effects such as streaks appearing in the reconstruction and contributions outside Υ_- are known artefacts that are inconsistent with the known support. Algorithms have been devised to quantify inconsistencies on the projection data [Bates and McDonnell, 1989; Hanson, 1982; Ein-Gal *et al.*, 1974; Smith *et al.*, 1973] but do not consider reconstruction within Υ_+. However, these methods do not relate well to the physical process of data collection, nor the reconstruction steps that remerge the data to generate a consistent image from whence it came. Chapter 6 outlines a recovery scheme for inconsistencies that directly relates to the popular reconstruction method, *modified back-projection*, and the process for estimating the projection data, *forward projection*. There are a number of definitions available for *consistency* of projection data, each having quite different interpretations. In chapter 6, these are unified into one definition that more appropriately encompasses the known support of the medium, the process of data collection and that of reconstruction. Techniques based on these definitions are then given for the recovery of inconsistencies.

CHAPTER 2

MATHEMATICAL PRELIMINARIES

This chapter concerns details of various mathematical techniques employed in solving general inverse problems involving emanations of the kind described in section 1.4.

Section 2.1 introduces mathematical models for the emanations and media with which they interact. This is followed by a mathematical model for linear scalar wave motions as described by differential equations that are regarded as being dependent upon either time or frequency. *Diffraction tomography* is considered as a solution to inverse scattering problems in the context of forward scattering or weakly scattering objects.

Section 2.3 provides supplementary definitions of mathematical transforms and image representation used in this thesis, and which are commonly encountered in the analysis of various inverse problems and CT.

2.1 MATHEMATICAL PHYSICS OF SCATTERING

Section 2.1.1 lists the types of wave motion and properties of media considered in this thesis as well as their relationship to the constitutive parameters. Sections 2.1.2 and 2.1.3 outline conventional mathematical models for linear scalar wave motions within these media. Section 2.1.4 introduces a field solution for the special case of wave motion radiating from a point source within free-space. This is relevant to the development of diffraction tomography. These mathematical models are manipulated into a form suitable for the high frequency approximations in section 2.1.5.

2.1.1 Emanations and the medium

The emanations considered here for the study of any body from remote positions are electromagnetic, elastic or acoustic wave motions. These wave motions can in many situations of practical interest be described by common mathematical models such as those given in sections 2.1.2 and 2.1.3. Wave motion is represented, respectively, in each model by complex valued wavefunctions that are either time or frequency dependent, $\Psi(\mathrm{x}, t)$ or $\psi(\mathrm{x}, k)$, where t and k represent time and the free-space wavenumber, respectively. The two wavefunctions can be directly related through the Fourier transform [Bracewell, 1978] :

$$\Psi(\mathrm{x}, t) = \frac{c}{2\pi} \int_{-\infty}^{\infty} \psi(\mathrm{x}, k)\, e^{ickt}\, dk, \qquad (2.1)$$

where $i = \sqrt{-1}$ and c is the free-space wave speed. The wave motion, originating from sources that are contained solely within the region Υ_{inc} (refer figure 1.1(*a*)), produce a field that exists at all points in Υ and is observed within Υ_{obs}, where Υ_{obs} is distinct from the region containing the sources;

$$\Upsilon_{obs} \cap \Upsilon_{inc} = \emptyset. \tag{2.2}$$

The physical properties of a medium or scattering region that interact with the emanations, those properties which define the constitutive parameters to be imaged within Υ_-, and the notation used to represent these properties are shown in table 2.1. It is important to note in light of the developments to follow, that the physical properties of the medium under scrutiny are assumed here to be linear, time invariant and isotropic. Because of definition (1.4) introduced in section 1.3, it can also be seen that the scattering region is source free;

Physical Property	Notation
Free-space permeability	μ_0
Permittivity	ϵ
Free-space permittivity	ϵ_0
Electrical conductivity	σ
Density	ρ
Compression modulus	κ
Bulk modulus	$\hat{\lambda}$

Table 2.1 Physical properties of media and notation used to represent them.

$$\Upsilon_{inc} \cap \Upsilon_- = \emptyset. \tag{2.3}$$

It should be noted that there are scattering problems that arise in CT whereby the sources occur naturally within, or are introduced into, Υ_-. In these cases condition (2.3) does not hold. Examples include the various emission computed tomography techniques [Jaszczak, 1988] such as PET and SPECT mentioned in section 1.1.

2.1.2 Time domain wave equations

Effectively, the most general equation describing linear scalar wave motion in such a medium as that described in section 2.1.1 is [Bates *et al.*, 1991]

$$\nabla^2\Psi + \beta\dot{\Psi} - (\nu/c)^2\ddot{\Psi} + \mu\Psi = 0, \tag{2.4}$$

where $\dot{\Psi}$ and $\ddot{\Psi}$ represent Ψ differentiated with respect to the temporal parameter t once and twice, respectively, and $\beta = \beta(\mathrm{x})$, $\nu = \nu(\mathrm{x})$ and $\mu = \mu(\mathrm{x})$ are the set of scalar time invariant constitutive parameters (encompassed in section 1.3 by the parameter χ) characterising the scattering region Υ_-. The constitutive parameters are related to the physical properties of the medium, shown in table 2.1, in the manner described by table 2.2.

2.1.3 Frequency domain wave equations

It is assumed that the incident field, $\Psi_{inc}(\mathrm{x}, t)$, has the time dependence, e^{ickt}. Because the medium is linear, the total field, $\Psi(\mathrm{x}, t)$, has a time dependence that can also be described by a simple harmonic component with the same wavenumber k. If the response of the scattering region can be characterised for simple harmonic motion of any wavenumber k, then it is possible to characterise the response to an arbitrary function of time by use of Fourier transform techniques [Morse and Feshbach, 1953].

Replacing $\Psi(\mathrm{x}, t)$ by its temporal Fourier transform, $\psi(\mathrm{x}, k)$, substituting the time dependence e^{ickt} into (2.4) and eliminating this exponential factor gives the inhomogeneous Helmholtz wave equation :

$$\Big(\nabla^2 + k^2\nu^2 + (\mu + ick\beta)\Big)\,\psi(\mathrm{x}, k) = 0. \tag{2.5}$$

Equation (2.5) can be simplified and generalised into *canonical* form by including all dispersive and attenuation parameters, $\mu(\mathrm{x})$ and $\beta(\mathrm{x})$, respectively, within the one composite complex-valued and frequency-dependent parameter $\mu = \mu(\mathrm{x}, k)$:

$$\Big(\nabla^2 + k^2\nu^2 + \mu\Big)\,\psi(\mathrm{x}, k) = 0. \tag{2.6}$$

The non-dispersive constitutive parameter $\nu(\mathrm{x})$ has been historically referred to as the refractive index [Born and Wolf, 1980]. The velocity of a wave passing through a particular point depends upon $1/\nu(\mathrm{x})$ at that point and also the direction in which the wave is travelling (as determined by the laws of refraction, see section 3.3).

Free-space is characterised by the constitutive parameters $\nu(\mathrm{x}) = 1$ and $\mu(\mathrm{x}, k) = 0$. If these values are substituted into the canonical form of the Helmholtz equation, then (2.6)

Emanation	$\beta(\mathrm{x})$	$\mu(\mathrm{x})$	$\nu(\mathrm{x})$
Electromagnetic	$\mu_0\sigma$	0	$\sqrt{\frac{\epsilon}{\epsilon_0}}$
Elastic	0	$\frac{2\rho\nabla^2\rho - 3\nabla\rho\cdot\nabla\rho}{4\rho^2}$	$\sqrt{\frac{\rho}{\lambda}}$
Acoustic	0	$\frac{2\rho\nabla^2\rho - 3\nabla\rho\cdot\nabla\rho}{4\rho^2}$	$\sqrt{\frac{\rho}{\kappa}}$

Table 2.2 Relationships between the medium's physical properties (see table 2.1) and the constitutive parameters for electromagnetic, elastic and acoustic emanations.

becomes

$$\left(\nabla^2 + k^2\right)\psi(\mathbf{x},k) = 0, \tag{2.7}$$

which is known as the free-space wave equation.

2.1.4 Free-space Green's function

It is convenient at this point to introduce a solution to the free-space wave equation (2.7) for the field radiating from a monochromatic point source with unit strength (of wave number k) within free-space. Clearly, (2.7) must now account for the point source. Specifically, the free-space Green's function is defined as the particular solution of the equation

$$\left(\nabla^2 + k^2\right) g(\mathbf{x},\mathbf{x}') = -\delta(\mathbf{x}-\mathbf{x}'), \tag{2.8}$$

where $\delta(\mathbf{x}-\mathbf{x}')$ represents a point source placed at $\mathbf{x}'$. Physically, $g(\mathbf{x},\mathbf{x}')$, called the free-space Green's function, represents the field at $\mathbf{x}$ radiating from this point source in free-space. Explicit solutions for $g(\mathbf{x},\mathbf{x}')$ in two- and three-dimensions are [Jones, 1964]

$$g(\mathbf{x},\mathbf{x}') = \begin{cases} -\frac{i}{4} H_0^2(kr) & \mathbf{x} \in \mathbf{R}^2, \\ \\ \dfrac{e^{-ikr}}{4\pi r} & \mathbf{x} \in \mathbf{R}^3, \end{cases} \tag{2.9}$$

where $r = |\mathbf{x}-\mathbf{x}'|$, H_0^2 is the cylindrical Hankel function of the second kind of order zero, and $\mathbf{R}^K$ denotes the set of points in K-dimensional space.

2.1.5 High frequency approximations

When the wavelength of the field is so small that significant changes in the constitutive parameters of a medium occupy many wavelengths, then, in local regions, the field behaves as if it were a plane wave in a homogeneous medium; that is, [Jones, 1986],

$$\psi(\mathbf{x},\mathbf{k}) = e^{-i\mathbf{k}\cdot\mathbf{x}}, \tag{2.10}$$

where $\mathbf{k}$ is the direction of propagation and has unit vector $\hat{\mathbf{k}} = \mathbf{k}/k$. From (2.10), solutions to

$$\mathbf{k}\cdot\mathbf{x} = constant \tag{2.11}$$

represent surfaces of constant phase, termed *wavefronts*, for the field. These are planes for the field from a plane wave propagating in a homogeneous medium. When the refractive index

ν is inhomogeneous, the wavefronts are retarded surfaces and the field spreads in accordance with ν. Consequently, it is preferable to express the field as [Jones, 1986]

$$\psi(\mathrm{x}, \mathrm{k}) = A(\mathrm{x})\, e^{-ikH(\mathrm{x})}, \tag{2.12}$$

where A and H are real functions of position, x, in the medium. $A(\mathrm{x})$ is the amplitude and $k\, H(\mathrm{x})$ the phase of the field. The wavefront surfaces of constant phase are now represented more generally by

$$H(\mathrm{x}) = constant. \tag{2.13}$$

If this high frequency representation for the field in an inhomogeneous medium, (2.12), is substituted into the Helmholtz equation (2.6), then, it can be seen that

$$k^2 \left[(\nu^2 - \nabla H \cdot \nabla H) A \right] - ik \left[A \nabla^2 H + 2 \nabla A \cdot \nabla H \right] + \nabla^2 A = 0. \tag{2.14}$$

By definition of the high frequency field, k is large and significant variations in $A(\mathrm{x})$ only occur over many wavelengths so that ∇A is small. Consequently, the most significant term in (2.14) is the coefficient of k^2. Equating the significant term in the LHS of (2.14) to the RHS, gives

$$\nabla H(\mathrm{x}) \cdot \nabla H(\mathrm{x}) = \nu^2(\mathrm{x}). \tag{2.15}$$

Equation (2.15) is a differential equation for H, defining the wavefronts in terms of the refractive properties of a medium. H is termed the *eikonal* and (2.15) the *eikonal equation* (eikonal is Greek for "image").

The amplitude $A(\mathrm{x})$ can be estimated by equating the second term of (2.14), which is the coefficient of k, to zero;

$$A \nabla^2 H + 2 \nabla A \cdot \nabla H = 0 \tag{2.16}$$

is known as the *first transport* equation. In general (2.16) cannot be solved explicitly and approximate techniques such as ray tracing, geometrical optics or the geometrical theory of diffraction (see chapter 3 for a description of these) can be invoked to solve (2.15) and (2.16) for $H(\mathrm{x})$ and $A(\mathrm{x})$, respectively. Chapter 4 introduces a new algorithm based on (2.15) which may improve upon the techniques listed in the previous sentence to solve for $H(\mathrm{x})$, or equivalently $\nu(\mathrm{x})$.

As an alternative to the expression for $\psi(\mathrm{x}, k)$ in (2.12) for a more general medium whereby ∇A does not vanish identically, and consequently, (2.15) is not a true solution of (2.6), a more accurate solution can be obtained by expressing $\psi(\mathrm{x}, k)$ as an asymptotic expansion in inverse powers of k [Bates *et al.*, 1991; Born and Wolf, 1980; Kline, 1951] :

$$\psi(\mathrm{x}, k) = e^{-ikH(\mathrm{x})} \sum_{n=0}^{\infty} k^{-n} A_n(\mathrm{x}), \tag{2.17}$$

thereby introducing further transport equations for the A_n. Note that (2.12) is the first term of the asymptotic expansion (2.17). The dispersive constitutive parameter $\mu(\mathrm{x}, k)$ can also be expanded asymptotically in accord with (2.17) [Bates *et al.*, 1991].

2.2 DIFFRACTION TOMOGRAPHY

In this section formulations for forward-scatter imaging, known as *diffraction tomography* [Bates *et al.*, 1991] are introduced. The formulations considered here exclude wide-angle scattering. Those of forward-scatter imaging are more widely applicable to situations involving diffraction tomography (eg ultrasonic CT) [Bates *et al.*, 1991]. Diffraction tomography is considered in this section as an inverse problem for wave motion whose behaviour within a complex-valued diffractive medium is modelled by the Helmholtz equation (2.7).

Formulations for the solution of direct and inverse scattering problems are usually divided into two categories, *exact* and *approximate*. Exact formulations characterise diffraction as *full wave* as distinct from approximate solutions that introduce some dispensation to do with the medium under study or the scattered field. Exact solutions tend to be computationally unmanagable unless there are restrictions placed on the size of the scatterer, any variation in refractive index is minimal, or the scatterers are so idealised, as to be merely of academic interest. For these reasons exact solutions are not widely used in practice [Bates *et al.*, 1991].

A large number of approximate formulations have been used and it would not be appropriate to encompass all of them in this thesis. Perhaps the simplest and most commonly invoked mathematical method for describing wave motion scattering from small penetrable objects is the Born approximation [Bates *et al.*, 1991] (see section 2.2.3). The essence of the Born approach lies in the solution of an integral equation over the domain of the scattering object, such as the volume source formulation developed in section 2.2.2. Solutions to this integral equation comprise a range of mathematical approaches that have been reported on at length [Colton and Kress, 1983]. When the integral equation is solved by simple iteration using the Born approximation initially, the solution yields the Born series (shown in (2.27)). A second approximation, named after Rytov, although quite different from the Born approximation, possesses the same foundation, in the solution of a domain integral equation. This approximation, although not as simple, nor as widely adopted as a basis for extended algorithms, is an important improvement over the first order Born approximation because of the greater size of scatterers that it can cope with.

The Born and Rytov approximations are examples of quite different perturbation methods [Keller, 1969], and consequently, the domains of validity for each are quite different. Both are able to yield unique solutions for the constitutive parameters so long as the underlying assumptions are valid. Although the Born approximation is the most convenient to implement, the Rytov approximation is generally accepted as having a wider range of validity in forward scattering problems [Bates *et al.*, 1991; Keller, 1969]. The validity of both the Born and Rytov approximations have been investigated by many authors [Keller, 1969; Lo *et al.*, 1988; Slaney and Kak, 1985]. Attempts to extend the range of validity of the Born and Rytov approaches currently constitute much of the original research and theoretical advance in the area of diffraction tomography [Bates *et al.*, 1991; Kleinman *et al.*, 1990].

2.2.1 Partitioning of wave motion

A definition for the incident and scattered fields in terms of the total field is given in (1.5). It is convenient to create a new definition of the incident field, existing within all space apart from Υ_{inc} where the sources reside. The incident field is defined in the remaining space as being a solution to the free-space wave equation (2.7)

$$\left(\nabla^2 + k^2\right) \psi_{inc}(\mathrm{x}, k) = 0, \tag{2.18}$$

so that after substituting (1.5) and (2.18) into the canonical form of the inhomogeneous Helmholtz equation, (2.6) becomes

$$\left(\nabla^2 + k^2\right) \psi_{scat}(\mathrm{x}, k) = -\left[k^2\left(\nu^2 - 1\right) + \mu\right] \psi(\mathrm{x}, k), \tag{2.19}$$

where the left hand side of (2.19) is in the same form as the free-space wave equation (2.7). This model for ψ_{inc} assumes that the incident field continues unperturbed by the introduction of an object into its path.

2.2.2 Volume source formulation

In section 2.2.1, the incident field is treated as propagating in free-space everywhere except in Υ_{inc}. The essence of the volume source formulation is that the effect of introducing an object into the incident field can be analysed by considering that equivalent sources exist in the place of the object producing a re-radiating field that propagates in free-space to produce the same total field. The field due to the equivalent sources corresponds to ψ_{scat}. The form of (2.19) facilitates the treatment of ψ_{scat} in free-space as the left hand side containing the scattered field component is in the same form as the free-space wave equation (2.7). The right hand side can then be considered as the equivalent sources, $\tilde{\sigma} = \tilde{\sigma}(\mathrm{x}, k)$, yielding the re-radiated field in free-space :

$$\begin{aligned} \left(\nabla^2 + k^2\right) \psi_{scat}(\mathrm{x}, k) &= -\left[k^2\left(\nu^2 - 1\right) + \mu\right] \psi(\mathrm{x}, k) \\ &= -\tilde{\sigma}(\mathrm{x}, k). \end{aligned} \tag{2.20}$$

The equivalent sources can be thought of as an ensemble of point sources over the domain of the scatterer,

$$\tilde{\sigma}(\mathrm{x}, k) = \int_{\Upsilon_-} \tilde{\sigma}(\mathrm{x}', k)\, \delta(\mathrm{x} - \mathrm{x}')\, d\Upsilon(\mathrm{x}'). \tag{2.21}$$

The solution for the scattering field in free-space at x from a point source placed at x′ is known as the free-space Green's function $g(\mathrm{x}, \mathrm{x}')$ (see section 2.1.4 and equation (2.9) for the 2-D and 3-D Green's function). By assembling the scattered field from all point sources in the same manner ascribed to in (2.21), the scattered field can be written as the superposition of the Green's function for each point source

$$\begin{aligned}\psi_{scat}(\mathrm{x},k) &= \int_{\Upsilon_-} \tilde{\sigma}(\mathrm{x}',k)\, g(\mathrm{x}-\mathrm{x}')\, d\Upsilon(\mathrm{x}') \\ &= \int_{\Upsilon_-} \left[k^2 \left(\nu^2(\mathrm{x}') - 1\right) + \mu(\mathrm{x}',k) \right] \psi(\mathrm{x}',k)\, g(\mathrm{x}-\mathrm{x}')\, d\Upsilon(\mathrm{x}'),\end{aligned} \tag{2.22}$$

where the equivalent sources $\tilde{\sigma}$ are found in terms of the constitutive parameters and total field from (2.20). Since the incident field also assumes its free-space form and is not re-radiated, then from (1.5) and (2.22), the total field is

$$\psi(\mathrm{x},k) = \psi_{inc}(\mathrm{x},k) + \int_{\Upsilon_-} \left[k^2 \left(\nu^2(\mathrm{x}') - 1\right) + \mu(\mathrm{x}',k) \right] \psi(\mathrm{x}',k)\, g(\mathrm{x}-\mathrm{x}')\, d\Upsilon(\mathrm{x}'). \tag{2.23}$$

Although the development of (2.23) invoked a form for which the field travelled everywhere in free-space, which contradicts the behaviour of wave motion in refracting media, there are no approximations made. This is because wherever the propagation medium differs from free-space, equivalent re-radiating sources are introduced. These sources are referred to as polarisation sources since they represent the polarisation of the medium by the incident radiation [Bates and Ng, 1972]. Consequently, solutions to (2.23) for the direct and inverse problems are placed in the category of "exact" solutions, there being no approximation at this stage. The domain integral equation (2.23) is a Fredholm integral equation of the second kind [Morse and Feshbach, 1953]. While standard numerical techniques are available [Press *et al.*, 1987] for implementation by computer, care needs to be taken with regard to the stability of solutions and their sensitivity to errors in the data [Bates *et al.*, 1991].

2.2.3 Born (Rayleigh-Gans) approximation

The Born approximation, originally devised by Lord Rayleigh, was further developed by Gans and is sometimes known as the Rayleigh-Gans approximation. It was also developed by Born in the context of non-relativistic quantum mechanics where it is particularly useful for solving the Schrödinder equation, which assumes the refractive index everywhere to be unity and the constitutive parameters to be non-dispersive [Bates *et al.*, 1991; Jones, 1986; Murch, 1990].

In the Born approximation, the incident field is characterised, as described in section 2.2.1, as propagating everywhere (except Υ_{inc}) as though in free-space. The scattered field is then modelled by equivalent sources as shown in the volume source equation, (2.22), of the previous section. The Born approximation is predicated on the assumption that the total field within the scattering region can be approximated by the incident field. The volume source equation, solved for the scattered field within Υ_- using the first order Born approximation

$$\psi(\mathrm{x},k) \approx \psi_{inc}(\mathrm{x},k) \quad \text{for} \quad \mathrm{x} \in \Upsilon_-\,, \tag{2.24}$$

then becomes

$$\psi_{scat}(\mathbf{x},k)=\int_{\Upsilon_-}\left[k^2\left(\nu^2(\mathbf{x}'-1)\right)+\mu(\mathbf{x}',k)\right]\psi_{inc}(\mathbf{x},k)\,g(\mathbf{x}-\mathbf{x}')\,d\Upsilon(\mathbf{x}'). \quad (2.25)$$

This approximation is equivalent to assuming that the scattered field is negligible by comparison with the incident field within Υ_-;

$$|\psi_{scat}(\mathbf{x},k)| \ll |\psi_{inc}(\mathbf{x},k)| \quad \text{for} \quad \mathbf{x}\in\Upsilon_-. \quad (2.26)$$

It is important to understand that an appreciable deviation can also be said to occur when the phase of the wavefronts manifesting from the total field differ from that of the incident field by more than $\lambda/2$ [Goodman, 1968]. This severely limits the class of inhomogeneous media to which the Born approximation can be applied. Such an approximation cannot tolerate large variations in refractive index, and, since the wavefront curvature is cumulative over the entire propagation path in the medium, this severely limits the size of the medium that can be imaged.

Since the approximation in (2.24) can lead to unacceptable errors, (2.25) is sometimes used as a first order approximation to the scattered field in an iterative expansion. This *Born series* or *Neumann series* expansion, mentioned in section 2.2, for the total and scattered field is expressed as [Bates *et al.*, 1991; Kleinman *et al.*, 1990]

$$\psi(\mathbf{x},k)=\sum_{l=0}^{\infty}\psi_{scat,l}(\mathbf{x},k) \quad \text{and} \quad \psi_{scat}(\mathbf{x},k)=\sum_{l=1}^{\infty}\psi_{scat,l}(\mathbf{x},k), \quad (2.27)$$

where in accordance with the first order Born approximation $\psi_{scat,0}=\psi_{inc}$. Convergence of this series is limited. Kleinman *et al.* [1990] report a modification of (2.27) that significantly improves its convergence rate and domain of convergence.

2.2.4 Rytov approximation

Although the Born approximation provides a convenient form of solution to the volume source formulation, assuming the total field within Υ_- can be approximated by ψ_{inc} severely restricts the set of circumstances in which it can be applied. Following the form taken for the field within an inhomogeneous medium as given in (2.12), the field can be more fully represented by incorporating the wavefront curvature of the field as it propagates through the medium. The incident and total fields are thus represented as

$$\psi_{inc}(\mathbf{x},k) = e^{-ikS_{inc}(\mathbf{x})}, \quad (2.28)$$

$$\psi(\mathbf{x},k) = e^{-ik\left(S_{inc}(\mathbf{x})+S_{scat}(\mathbf{x})\right)}, \quad (2.29)$$

where $S_{inc}(\mathbf{x})$ and $S_{scat}(\mathbf{x})$ are here complex functions of position $\mathbf{x}$, whose real and imaginary parts can be considered to incorporate the wavefront curvature and amplitude variations, respectively, as the field propagates through the medium.

In the same manner as the Born approximation the incident field is taken to satisfy the free-space wave equation, as per (2.18). The complex valued function $S_{inc}(\mathrm{x})$ then satisfies

$$\nabla^2 S_{inc} - ik(\,\nabla S_{inc} \cdot \nabla S_{inc} - 1\,) \;=\; 0. \tag{2.30}$$

Substituting the total field (2.29) into the inhomogeneous Helmholtz equation, (2.6), and using the above condition on the incident field shows that $S_{scat}(\mathrm{x})$ satisfies

$$\nabla^2 S_{scat} + ik\left[\left(\nu^2 - 1\right) - \nabla S_{scat} \cdot \nabla S_{scat} - 2\,\nabla S_{inc} \cdot \nabla S_{scat}\right] + i\mu/k \;=\; 0. \tag{2.31}$$

On introducing Rytov's transformation [Chernov, 1967],

$$\tilde{\psi}(\mathrm{x}, k) \;=\; S_{scat}\,\psi_{inc}(\mathrm{x}, k) \;=\; S_{scat}\,e^{-ikS_{inc}}, \tag{2.32}$$

it can be seen, using (2.30), that

$$\left(\,\nabla^2 + k^2\,\right)\,\tilde{\psi} \;=\; \left(\nabla^2 S_{scat} - 2ik\nabla S_{inc} \cdot \nabla S_{scat}\right)\psi_{inc}, \tag{2.33}$$

and, with the aid of (2.31), can be rewritten as

$$\left(\,\nabla^2 + k^2\,\right)\,\tilde{\psi}(\mathrm{x}, k) \;=\; -ik\left[\left(\nu^2 - 1\right) + \mu/k^2 - \nabla S_{scat} \cdot \nabla S_{scat}\right]\,\psi_{inc}(\mathrm{x}, k). \tag{2.34}$$

Conventionally, the Rytov approximation is obtained by neglecting the term $\nabla S_{scat} \cdot \nabla S_{scat}$ leaving (2.34) in a familiar form of Helmholtz equation with equivalent sources (compare (2.20))

$$\left(\,\nabla^2 + k^2\,\right)\,\tilde{\psi}(\mathrm{x}, k) \;=\; -\frac{i}{k}\left[k^2\left(\nu^2 - 1\right) + \mu\right]\,\psi_{inc}(\mathrm{x}, k). \tag{2.35}$$

In the same manner that (2.22) follows on from (2.20) and (2.21), $\tilde{\psi}$ can be expressed in the form of a domain integral equation

$$\tilde{\psi}(\mathrm{x}, k) \;=\; \frac{i}{k}\int_{\Upsilon_-}\left[\,k^2\left(\nu^2(\mathrm{x}') - 1\right) + \mu(\mathrm{x}', k)\,\right]\psi_{inc}(\mathrm{x}', k)\,g(\mathrm{x} - \mathrm{x}')\,d\Upsilon(\mathrm{x}'). \tag{2.36}$$

As mentioned previously, the Rytov approximation is an improvement over that of the Born because it is valid over a wider range of media. The Rytov approximation also requires weak scattering but only requires inhomogeneities within the scatterer to exhibit gradual fluctuations so that the rate of change, in amplitude and phase ($\nabla S_{scat} \cdot \nabla S_{scat}$), of the scattered field is made small per wavelength. In contrast, the Born approximation requires that the accumulated change in phase across the entire medium must be small. In the Rytov case the scattering is mainly in the forward direction but may be considered to be substantially refractive [Robinson and Greenleaf, 1985]. This extends the size of the scatterers that can be examined. However, Keller [1969] has shown that the Rytov approximation exhibits superiority over the Born approximation for calculating scattering only when it is predominantly in the direction of the incident field (forward scattering). Bates *et al.* [1976] present an extension to the Rytov approximation that partially compensates for the neglected $\nabla S_{scat} \cdot \nabla S_{scat}$ which may better account for refraction. It was thought that in compensating for any distortions introduced by neglecting this term, the domain of validity of the Rytov approximation could be extended.

2.2.5 Final Remarks

Devaney [1982] introduces an efficient method of inverting scattering data whose interpretation is based on either the Born or Rytov approximations. It is an extension of the popular *modified back-projection* method used in straight ray or X-ray CT (see section 5.6.3.3 for a description of this inversion technique). As intimated in section 5.6.3.3, the modified back-projection method offered an alternative to Fourier methods of inverting X-ray CT projection data that were based on the *projection theorem* of section 5.2.

In analogy to the modified back-projection method, Devaney [1982] offers a technique known as *filtered back-propagation* for inverting scattering data as an alternative to Fourier techniques based on the *diffraction theorem.* The diffraction theorem [Devaney, 1982; Devaney, 1985] in 2-D derives the 1-D Fourier transform of the scattered field to be the Fourier transform of the constitutive parameters on circles in 2-D Fourier space. At a fixed frequency for the incident field, Fourier space can be constructed by employing incident waves from all angles. By inverting the Fourier transform space of the constitutive parameters, constructed from the circles, the distribution of the constitutive parameter can be found.

Since in practice only scattering data from a discrete number of angles is available, Fourier space cannot be completely estimated and as the inversion of discrete Fourier transform data requires points on a regular grid, the points are interpolated from the data known on the circular loci. As noted in section 5.6.3.1, the interpolation procedure in Fourier space introduces distortion. The filtered back-propagation algorithm proposed by Devaney [1982] eliminates the need for this interpolation in Fourier space by reconstructing the constitutive parameters in image space (see section 2.3.1 for an explanation of image and Fourier spaces in 2-D). The filtered back-propagation algorithm reconstructs the scattering data in a two-step process. The first step involves modifying the data by convolving with a filter function, analogous to the modifying process described in section 5.6.3.3. The second step is the superposition of the modified scattering data onto the image space of the constitutive parameter, analogous to back-projection described in section 5.4. The corresponding formulations for the first and second step in filtered back-propagation, modification by a filter and back-propagation, respectively, are given by Devaney [1982]. It should be noted that the value of this algorithm is its efficiency in reconstructing the scattering data and eliminating the distortion involved in Fourier domain interpolation. As noted in section 5.6.3.3 in the context of inverting X-ray CT data, modified back-projection leads to images that are more accurate than those produced by older Fourier techniques employing a straightforward approach to interpolation. More sophisticated interpolation methods have since been developed. However, the quality of any reconstruction is limited by the assumed processes, and so the filtered back-propagation algorithm can do no better than those developed from the Born or Rytov approximations.

Wang and Chew [1989] have provided one of the more recent promising approaches to diffraction tomography [Bates *et al.*, 1991] and have illustrated its application in imaging 2-D lossless permittivity profiles that are small compared to the wavelength. Their solution is an iterative algorithm for solving the inverse problem. At each iteration, the first step is to solve the direct problem for the scattered field within Υ_- and Υ_{obs}, from the known incident field and the current estimate of the constitutive parameters. The second step of the iteration is to solve the inverse problem for the next order estimate of the constitutive parameters, from a knowledge of the scattered field calculated in the first step. An initial estimate of the constitutive parameters is made via an approximation such as the first order Born (see section 2.2.3). Iterations are halted when the scattered field obtained from solving the forward scattering problem is consistent with that of the observed field. This technique of consistent reconstruction has been applied for some time in the field of X-ray computed tomography for estimating a 2-D distribution whose Radon transforms (or forward projections) are consistent with the measured data. Consistent reconstruction forms the foundation for the development of *Algebraic Reconstruction Techniques* [Gordon *et al.*, 1970] for the inversion of X-ray data in 3-D electron microscopy. Refer to section 5.6.6 for a discussion of algebraic reconstruction techniques. Consistent reconstruction for inversion of X-ray CT data is discussed further in section 7.1.

2.3 SUPPLEMENTARY DEFINITIONS AND MATHEMATICAL TECHNIQUES

2.3.1 Fourier transform and discrete Fourier transform

In this thesis, a function of the K-dimensional variable $\mathbf{x}$ is written in lower case letters, for example $f(\mathbf{x})$, while its *Fourier transform* (FT) is generally denoted by upper-case letters $F(\mathbf{u})$ where $\mathbf{u}$ is the FT variable that replaces $\mathbf{x}$. Here the multi-dimensional variables are denoted in bold-face, $\mathbf{x} = (x_1, ..., x_K)$, $\mathbf{u} = (u_1, ..., u_K)$. The operation of Fourier transformation on a function $f(\mathbf{x})$ is denoted by $\mathcal{F}\{\ f(\mathbf{x})\ \}$. Following Bracewell [1978] the FT of f with respect to the variable $\mathbf{x}$ is defined as

$$\begin{aligned} F(\mathbf{u}) &= \mathcal{F}\{\ f(\mathbf{x})\ \}(\mathbf{u}) \\ &= \int_{-\infty}^{\infty} f(\mathbf{x})\ e^{-i2\pi\mathbf{u}\cdot\mathbf{x}}\ d\mathbf{x}, \end{aligned} \tag{2.37}$$

and the reciprocal relation or *inverse Fourier transform* (IFT) is

$$\begin{aligned} f(\mathbf{x}) &= \mathcal{F}^{-1}\{\ F(\mathbf{u})\ \}(\mathbf{x}) \\ &= \int_{-\infty}^{\infty} F(\mathbf{u})\ e^{i2\pi\mathbf{x}\cdot\mathbf{u}}\ d\mathbf{u}. \end{aligned} \tag{2.38}$$

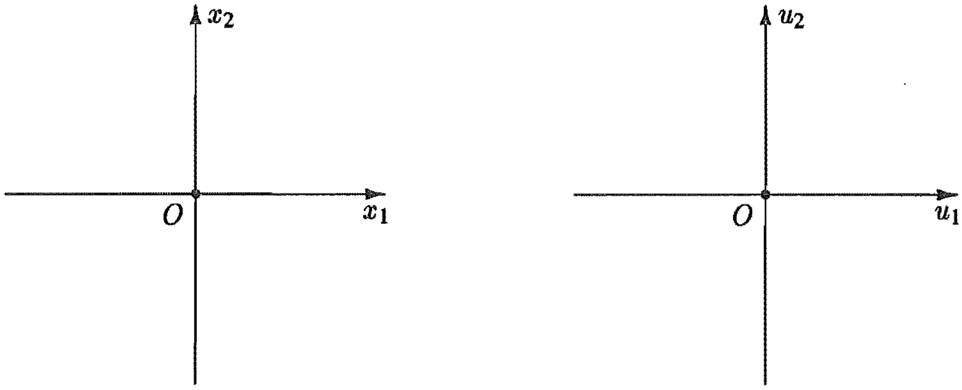


Image space
$f(x_1,x_2) = \mathcal{F}^{-1}\{\ F(u_1,u_2)\ \}(x_1,x_2)$

Fourier space
$F(u_1,u_2) = \mathcal{F}\{\ f(x_1,x_2)\ \}(u_1,u_2)$

Figure 2.1 Image space and Fourier space of a 2-D function of Cartesian coordinates (x_1, x_2). The Fourier Cartesian coordinates are (u_1, u_2).

Consider the FT of $f(\mathbf{x})$ with respect to a subset of the variables in $\mathbf{x}$. Let this subset be denoted by $\mathbf{x}_1 = (x_1, ..., x_I)$ (where $0 \leq I < K$), the remaining variables of $\mathbf{x}$ by $\mathbf{x}_2 = (x_{I+1}, ..., x_K)$, then in set notation $\mathbf{x} = \mathbf{x}_1 \cup \mathbf{x}_2$. The FT of f with respect to the subset of variables $\mathbf{x}_1$ is

$$\begin{aligned} F(\mathbf{u}_1,\mathbf{x}_2) &= \mathcal{F}_{(\mathbf{x}_1)}\{\ f(\mathbf{x})\ \}(\mathbf{u}_1,\mathbf{x}_2) \\ &= \int_{-\infty}^{\infty} f(\mathbf{x})\, e^{-i2\pi\mathbf{u}_1\cdot\mathbf{x}_1}\, d\mathbf{x}_1, \end{aligned} \tag{2.39}$$

and the reciprocal relation or IFT with this subset of Fourier variables is

$$\begin{aligned} f(\mathbf{x}) &= \mathcal{F}^{-1}_{(\mathbf{u}_1)}\{\ F(\mathbf{u}_1,\mathbf{x}_2)\ \}(\mathbf{x}) \\ &= \int_{-\infty}^{\infty} F(\mathbf{u}_1,\mathbf{x}_2)\, e^{i2\pi\mathbf{x}_1\cdot\mathbf{u}_1}\, d\mathbf{u}_1. \end{aligned} \tag{2.40}$$

The domain of $\mathbf{x}$ is referred to as image space while the domain of the Fourier variable $\mathbf{u}$ is termed as Fourier space. For example, consider a function of two Cartesian coordinates (x_1, x_2), namely $f(x_1, x_2)$, whose Fourier transform is $F(u_1, u_2)$. The image space and Fourier space of $f(x_1, x_2)$ are shown in figure 2.1. The Fourier transform pair for this Cartesian coordinate system can be written as

$$\begin{aligned} F(u_1,u_2) &= \int_{-\infty}^{\infty} f(x_1,x_2)\, e^{-i2\pi(u_1x_1+u_2x_2)}\, dx_1dx_2, \\ f(x_1,x_2) &= \int_{-\infty}^{\infty} F(u_1,u_2)\, e^{i2\pi(x_1u_1+x_2u_2)}\, du_1du_2. \end{aligned} \tag{2.41}$$

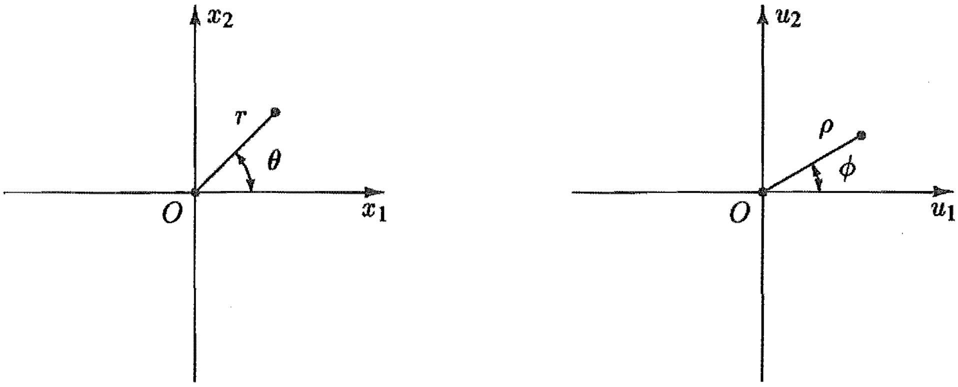


Image space
$f(r;\theta) = \mathcal{F}^{-1}\{\ F(\rho;\phi)\ \}(r;\theta)$

Fourier space
$F(\rho;\phi) = \mathcal{F}\{\ f(r;\theta)\ \}(\rho;\phi)$

Figure 2.2 Image space and Fourier space in polar coordinates.

Figure 2.2 shows the geometry of image space and Fourier space for a 2-D function and its Fourier transform, respectively, specified in polar coordinates. In order to express the 2-D Fourier transform in polar coordinates, the following identities are used (derived from the geometry given in figure 2.2) :

$$\begin{aligned} x_1 &= r\cos(\theta), \quad x_2 = r\sin(\theta), \\ u_1 &= \rho\cos(\phi), \quad u_2 = \rho\sin(\phi), \end{aligned} \tag{2.42}$$

and when substituted into (2.41) gives the required 2-D FT in polar coordinates :

$$\begin{aligned} F(\rho;\phi) &= \int_0^{2\pi}\int_0^{\infty} f(r;\theta)\, e^{-i2\pi\rho r\cos(\phi-\theta)}\ r\,dr\,d\theta, \\ f(r;\theta) &= \int_0^{2\pi}\int_0^{\infty} F(\rho;\phi)\, e^{i2\pi\rho r\cos(\phi-\theta)}\ \rho\,d\rho\,d\phi. \end{aligned} \tag{2.43}$$

2.3.1.1 Convolution theorem

The properties of the FT relationships given in (2.37) and (2.38) are extensively listed and detailed, see for example Bracewell [1978]. One such property of the FT that is important in the context of this thesis is referred to as the *convolution theorem*. The convolution of two functions p and h, of the same variable x, denoted $p(\mathrm{x})$ and $h(\mathrm{x})$, respectively, is defined here by

$$\begin{aligned} f(\mathrm{x}) &= p(\mathrm{x}) \odot h(\mathrm{x}) \\ &= \int_{-\infty}^{\infty} p(\mathrm{y})\, h(\mathrm{x}-\mathrm{y})\, d\mathrm{y} \\ &= h(\mathrm{x}) \odot p(\mathrm{x}), \end{aligned} \tag{2.44}$$

where $\odot$ denotes convolution. The convolution theorem relates the FT of this convolution to those of the individual convolving functions, $p(\mathrm{x})$ and $h(\mathrm{x})$, respectively. It states that

$$\begin{aligned} \mathcal{F}\{\, f(\mathrm{x}) \,\}(\mathrm{u}) &= \mathcal{F}\{\, p(\mathrm{x}) \,\}(\mathrm{u})\ \mathcal{F}\{\, h(\mathrm{x}) \,\}(\mathrm{u}) \\ &= P(\mathrm{u})\, H(\mathrm{u}), \end{aligned} \tag{2.45}$$

or alternatively, the Fourier transform of the convolved function is equal to the product of the convolving function's Fourier transforms.

2.3.1.2 Similarity theorem

Often it is the case that the variables of $f(\mathrm{x})$ are scaled by expansion or contraction factors $\mathbf{a}$, i.e., $f(\mathrm{x})$ becomes $f(\mathbf{a} \cdot \mathrm{x})$. The similarity theorem relates the Fourier transform of $f(\mathbf{a} \cdot \mathrm{x})$ to the Fourier transform $F(\mathrm{u})$ of $f(\mathrm{x})$. Consider the one-dimensional function $f(x)$ whose Fourier transform is $F(u)$. If the variable x is scaled by the scalar factor a, then

$$\mathcal{F}\{\, f(ax) \,\} = \frac{1}{|a|}\, F\left(\frac{u}{a}\right). \tag{2.46}$$

It can be seen from (2.46) that if a variable is expanded then the Fourier transform variable contracts (and *visa versa*) [Bracewell, 1978].

2.3.1.3 Discrete Fourier transform

As is usually the case in practice, a finite number of samples are the only representation for a function that is made available. Consequently, to take advantage from the properties of a function represented in the Fourier domain, a transformation is required that relates the samples to that of its Fourier transform. This transformation from discrete data to a trigonometric Fourier series representation is called the *discrete Fourier transform* (DFT). As is intuitively the case for all discrete representations of a continuous function in any dimensionality, sufficient samples must be available and at regular intervals such that these samples can adequately represent the information content required from the continuous function. Such concerns are the domain of *sampling theorems* [Bates and McDonnell, 1989; Bracewell, 1978] and are not considered here. Considering a 2-D function $f(x_1, x_2)$ as defined on the interval $-L_1/2 \le x_1 \le L_1/2$, $-L_2/2 \le x_2 \le L_2/2$ whose samples, equi-spaced in each linear dimension by a distance Δx_1 and Δx_2, respectively, are f_{m_1,m_2}. Then the discrete Fourier transform representation for $f(x_1, x_2)$ is defined as [Bates and McDonnell, 1989]

$$f(x_1, x_2) = \frac{1}{L_1 L_2} \sum_{m_1=-M_1}^{M_1} \sum_{m_2=-M_2}^{M_2} F_{m_1,m_2}\, e^{-i2\pi[\,(m_1\, x_1/L_1) + (m_2\, x_2/L_2)\,]}, \tag{2.47}$$

where the limits of M_1 and M_2 are in general infinite, but, for reasons of limited resources, are necessarily truncated to finite values. If the intervals L_1 and L_2 are chosen to be divisible by the sample spacing, i.e.,

$$\frac{L_1}{\Delta x_1} = 2M_1 + 1 \quad \text{and} \quad \frac{L_1}{\Delta x_2} = 2M_2 + 1, \tag{2.48}$$

then, the Fourier samples F_{m_1,m_2} can be expressed explicitly in terms of the f_{m_1,m_2}

$$F_{m_1,m_2} = \frac{1}{\Delta x_1\, \Delta x_2} \sum_{m_1'=-M_1}^{M_1} \sum_{m_2'=-M_2}^{M_2} f_{m_1',m_2'}\, e^{i2\pi\left[\,m_1 m_1'/(2M_1+1) + m_2 m_2'/(2M_2+1)\,\right]} \tag{2.49}$$

rendering the Fourier series coefficients F_{m_1,m_2} to be conveniently evaluated through (2.49) if (2.48) holds.

2.3.2 Image representation

As intimated in the previous section, only a finite storage space is available in which to represent any given image function f of whatever dimension. If the storage space is broken into discrete elements of finite storage, as in a digital storage device, then, inherently the distribution is stored in the manner of a finite number of these elements. Usually the representation of any function is taken to be in terms of appropriate basis functions. That is, they are appropriate because they are able and there is sufficient number to represent the required properties of f that need to be preserved in a finite representation. The discrete representation then becomes a combination of the basis functions, each with an appropriate weighting. The weightings of the basis functions are chosen in order to satisfy some optimal constraint appropriate to the image interpretation. The discrete Fourier series of section 2.3.1 is an example of such a representation where the basis functions are complex exponentials chosen to represent discrete spatial frequencies of the image and their weightings can be computed through (2.49).

A discrete representation of the continuous function $f(\mathrm{x})$ is denoted here as $\hat{f}(\mathrm{x})$ and the basis functions, of which there are a finite number J, are denoted by $\psi_j(\mathrm{x}),\ j = 1, \ldots, J$. If the transformation $\mathcal{T}$ from the $\psi_j(\mathrm{x})$ to $\hat{f}(\mathrm{x})$ is known, then, the relative weights $\hat{f}_j$ of each $\psi_j(\mathrm{x})$ only need be stored to compute

$$\hat{f}(\mathrm{x}) \;=\; \mathcal{T}\Big\{\, \hat{f}_j\, \psi_j(\mathrm{x});\ j = 1, \ldots, J \,\Big\}. \tag{2.50}$$

Usually the ψ_j are chosen so that they have a linear relationship to $\hat{f}$ and more commonly $\hat{f}$ is represented as a linear combination of the ψ_j, such as

$$\hat{f}(\mathrm{x}) = \sum_{j=1}^{J} \hat{f}_j \, \psi_j(\mathrm{x}), \tag{2.51}$$

where the coefficients of each basis function (weights), $\hat{f}_j$, are often samples of $\hat{f}$ and the basis functions normalised. If further transformations are to be made on the image, then more appropriate basis functions are chosen which facilitate the transformation. Lewitt [1990] presents Kaiser-Bessel functions as multi-dimensional basis functions that have explicit Fourier transformations and explicit formulas for Radon transformation, facilitating many of the processes in X-ray CT reconstruction algorithms such as those introduced in chapter 5.

Most applications of image reconstruction require the K-dimensional function to be estimated over a regular grid in a K-dimensional space. This image space is normally broken up into regular size shapes and within each it is assumed the image reconstruction $\hat{f}(\mathrm{x})$ is uniform. In this way the continuous distribution is approximated by non-overlapping linear sub-elements $\hat{f}_j$. The basis functions

$$\psi_j(\mathrm{x}) = \begin{cases} 1 & \text{if x belongs to } \hat{f}_j, \\ 0 & \text{otherwise.} \end{cases} \tag{2.52}$$

One of the advantages in using regular shapes is that they may be chosen to correspond well with the shapes used in rendering a display of the image. For example, volume images are usually represented by a set of adjacent volume elements known as *voxels* and cross-sections by 2-D elements known as *pixels* (see figure 2.3). In most cases the voxels are cubic elements and pixels are square and can be displayed as such.

A lot of interest has been shown in the rendering of CT images for clinical purpose and this has generated much research that currently continues. The trend for image display and visualisation has been towards an increase in the dimensionality of the display [Garden *et al.*, 1989]. Because objects are intrinsically three-dimensional, if displayed in a natural manner they may be better represented to a viewer in 3-D. It is becoming common to display surface- or volume-rendered images, calculated from the raw information contained in the voxel elements, and to allow the viewer to interact with the 3-D rendered images. Display techniques are rapidly developing with specialised computer hardware that allows rapid transformation and manipulation of large amounts of multi-dimensional data for display. Stereoscopic devices have developed and are accepted in clinical practice as a means of 3-D display [Peters *et al.*, 1989]. There is much literature available on the display and visualisation of 3-D CT images on 2-D displays [Chen *et al.*, 1985; Drebin *et al.*, 1988; Herman and Liu, 1979; Herman, 1986a; Herman, 1986b; Hull *et al.*, 1990; Robb and Barillot, 1989; Udupa, 1983].

Since the emphasis of this thesis is not on the display of images and because the image operations introduced (especially in chapter 5) can be readily extended to the 3-D voxel, the pixel is adopted as the basic 2-D element used for the purpose of image representation in this

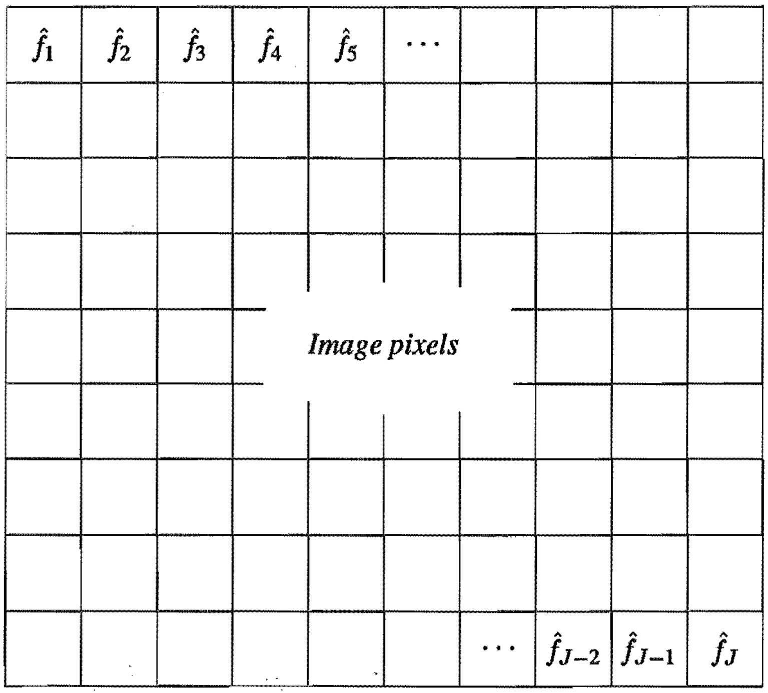


Figure 2.3 Pixelated 2-D image constructed of J square pixels.

thesis. Pixels have in the past been very suitable for internal representation of 2-D information as the display devices used small 2-D elements for displaying sub-elements, of varying colours or shades, and so no conversion from the internal representation was required for the display. However, these may not be optimal representations and many other 2-D shapes are available to represent the information. For example, Fuchs *et al.* [1977] use triangular tiles for surface reconstruction and Buonocore *et al.* [1981] suggest a natural pixel decomposition for a cross-sectional image. A natural pixel decomposition is a function of the technique used for physical generation of measurements and so facilitates further computation as do other basis functions mentioned in the previous paragraph.

CHAPTER 3

RAYS AND RAY TOMOGRAPHY

3.1 INTRODUCTION

An early model for the propagation of light was that of waves travelling along geometrical trajectories called *ray paths*. Even earlier, light was thought to travel only in straight paths. Perhaps one of the most influential factors in the growth of theory based on this model was the number of natural observations that could be accurately explained with this model. Later, Maxwell provided a unified theory for the fundamentals of the electromagnetic behaviour of light. With the derivation of the wave equation, and the acceptance of the wave nature of light, it was established that light itself was an electromagnetic phenomenon [Cornbleet, 1983]. Electromagnetic theory can thereby be used to explain all optical phenomena. The optical ray model has been found to be a first order solution to electromagnetic field theory.

Optical systems were, in earlier times, mainly designed from components consisting of homogeneous materials having constant refractive index, and subsequently, the ray paths, being along straight lines within each medium, were determined from simple geometrical rules applied to account for reflections and refractions at boundaries between media [Arnold, 1986]. Geometry thus became the natural discipline for the study of light travel and for the design of optical components. The theoretical foundations of which were developed under the discipline of *geometrical optics*. Because in general the refractive index distribution of a medium is not piecewise linear, or, alternatively stated, the medium may not simply be stratified, the ray paths are not in general straight, but are curved. More recently, the study of such inhomogeneous materials and their optical properties has become important. The effects of ray bending and diffraction have been explained and can in many cases be modelled accurately. Geometric optics accounts for the effects of ray bending within inhomogeneous media as well as the aforementioned reflection and refraction effects experienced at discontinuous boundaries. Keller [1962] extended the theory of geometrical optics to account for diffraction when he announced his *geometrical theory of diffraction* [James, 1986b]. Although there are no widely applicable inverse scattering algorithms based on the geometrical theory of diffraction, its development and extensions have confirmed that diffraction can be accounted for in ray optical terms in many interesting situations [Bates *et al.*, 1991]. An understanding of diffraction effects is considered essential in order to understand the validity of a ray approach.

There is considerable motivation to extend ray optical approaches to cope with more complicated phenomena [Felsen, 1985]. Ray theory is intuitive in nature and inverse scattering solutions having this foundation generally require simpler formulations than explicit solutions.

It can also provide a solution that is much less demanding computationally than exact inverse scattering solutions [Bates *et al.*, 1991; Enright *et al.*, 1992].

3.2 RAYS

A general representation, (2.12), for a high frequency field within an inhomogeneous medium was introduced in section 2.1.5, where it is assumed that the wavelength, λ, is much less than any significant spatial variation in properties of the medium. The field has all the properties locally of the plane wave. In accordance with (2.12), the field is defined as

$$\psi(\mathrm{x},\mathrm{k}) = A(\mathrm{x})\, e^{-ikH(\mathrm{x})}, \tag{3.1}$$

with velocity in the direction of $\nabla H(\mathrm{x})$ [Jones, 1986]. Energy propagates in the direction of $\nabla H(\mathrm{x})$, normal to the wavefronts defined by (2.13).

Under the ray approach, the rays are denoted as the curved paths along which energy is transported. Alternatively, a ray is defined as any curve with direction everywhere normal to the wavefronts, and whose direction corresponds to that of the propagation of energy. Each ray has an associated amplitude and polarisation that may vary along its path. For any pencil of rays, the intensity varies in proportion to their cross-section [Born and Wolf, 1980].

There is a direct correspondence between wave motion that is characterised by a high frequency approximation (described in the opening paragraph to this section) or in the wavelength limit, $\lambda \to 0$, and that which is characterised as rays. Wave motion characterised as rays has the same general character as that of a plane wave and penetrates an inhomogeneous region in such a way that its wavefronts obey geometrical principles, Snell's law of refraction (refer section 3.3), with respect to variations in refractive index.

The preceding remarks imply that, when the wavelength is small enough, the sum total of optical phenomena may be deduced from geometrical considerations, by determining the paths of the light rays and calculating the associated intensity and polarization [Born and Wolf, 1980]. What is required then, is an understanding of *geometrical optics* that embodies the classical tools for understanding the interaction of light, and more recently, electromagnetic propagation through inhomogeneous media.

3.3 GEOMETRICAL OPTICS

Classical *geometrical optics* (GO) is based entirely upon the *sine law of refraction* as derived by Snell [Cornbleet, 1983]. Snell's laws were derived for plane waves striking a plane interface. Geometrical optics can also accurately describe the propagation of electromagnetic radiation under certain conditions. Ideally GO is characterised by neglect of the wavelength, $\lambda \to 0$. However, the laws of geometrical optics can be applied with less stringent conditions on the wavelength, so long as it is small enough [Born and Wolf, 1980].

Snell's law of refraction introduces refracted and reflected waves whenever a plane wave is incident upon a point in the medium where there is an abrupt change or discontinuity in the refractive index. At high frequencies each portion of an interface can be regarded locally as a piece of tangent plane provided its principle radii of curvature are large compared with the wavelength [Jones, 1986; James, 1986a]. Hence if a ray falls on a discontinuous boundary it is split into a reflected and transmitted (refracted) ray, and the changes in the state of polarization as well as the reflectivity and transmissivity may be calculated from the corresponding formulae for plane waves [Born and Wolf, 1980].

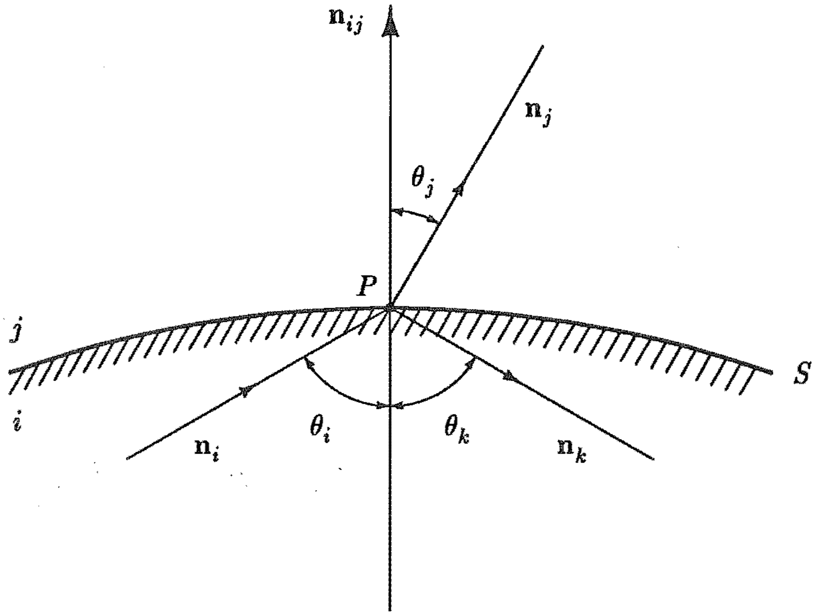


Figure 3.1 A ray incident upon a discontinuity in refractive index in the medium at point P. The incident, refracted and reflected rays having unit direction vectors $\mathbf{n}_i$, $\mathbf{n}_j$ and $\mathbf{n}_k$, respectively, are shown to all lie in the same plane. The unit normal to the surface at point P is $\mathbf{n}_{ij}$.

Consider a ray travelling through a medium whose refractive index is discontinuous at a boundary represented by a smooth surface S and is incident upon this surface at the point P (see figure 3.1). The direction of the incident ray, refracted ray, reflected ray and normal to the surface S at P all lie in the same plane and are represented by the unit vectors $\mathbf{n}_i$, $\mathbf{n}_j$, $\mathbf{n}_k$ and $\mathbf{n}_{ij}$ respectively. Snell's law of refraction and the law of reflection may then be stated, respectively, by

$$\begin{aligned} \nu_i\,(\mathbf{n}_{ij}\times\mathbf{n}_i) &= \nu_j\,(\mathbf{n}_{ij}\times\mathbf{n}_j) \\ \text{or,}\quad \nu_i\,\sin(\theta_i) &= \nu_j\,\sin(\theta_j) \end{aligned} \tag{3.2}$$

and

$$\theta_i = \theta_k, \tag{3.3}$$

where × denotes the vector product or cross product. According to Snell's law of refraction, the angle that the refracted ray makes with the normal to the surface at P can be found from (3.2). Similarly, Snell's law of reflection states that the angle in which the reflected ray makes with the normal at P is equal to that of the incident ray (equation (3.3)). The corresponding amplitudes of these rays are determined from refraction and reflection coefficients that assume continuity of the field across the surface S. Refraction and reflection (transmission) coefficients are detailed by Born and Wolf [1980] as well as their energy transfer coefficients (reflectivity and transmissivity).

3.4 FERMAT'S PRINCIPLE

A number of applications exist that require the study of electromagnetic radiation in media which is inhomogeneous. In these cases, straight paths cannot be assumed as the media may have a continuously varying refractive index. Fermat's principle, described below, characterises the ray paths through such media.

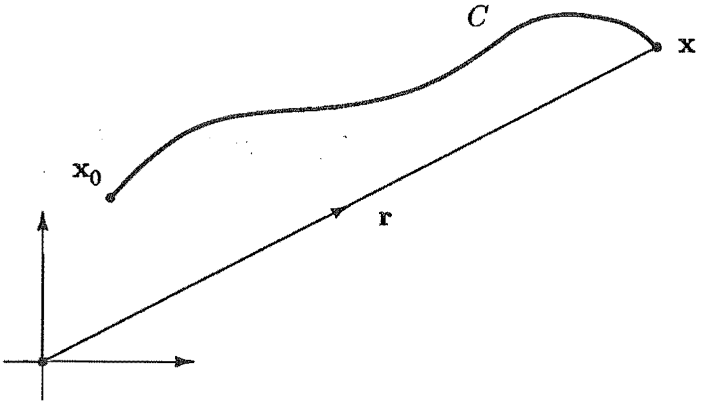


Figure 3.2 Ray path, C, between two points x_0 and x.

Consider a curve C joining the points x_0 and x in a medium whose refractive index at any position x is $\nu(x)$ (see figure 3.2). Then the curvilinear length of the curve C,

$$\begin{aligned} H(x_0, x) &= \int_C \nu \, ds \\ &= \int_{x_0}^{x} \nu(x) \, ds, \end{aligned} \tag{3.4}$$

is known as the *optical path length*, where ds is the element of arc along C. If C is the path of a ray, then Fermat's principle characterises C : *The rays between* x_0 *and* x *are those curves along which the optical path length is stationary with respect to infinitesimal variations in the path* C [Jones, 1986]. Fermat's solution for a stationary path in (3.4) is a problem which lends itself to the calculus of variations. As noted by [Jones, 1986], Fermat's principle does not necessarily require the optical path to be a minimum but to be stationary. If this point is forgotten significant rays may be lost.

3.5 RAY TRACING

Ray tracing is an example of the direct problem (refer to section 1.3) since the constitutive parameter of interest, $\nu(\mathrm{x})$, is given and the field distribution within the refractive medium is to be found. In calculating the field throughout a refractive index distribution, $\nu(\mathrm{x})$, the problem is broken into two steps. Firstly, the ray paths are calculated which determine the phase of the field independently from its amplitude [Deschamps, 1972]. Secondly, from these assumed paths the amplitude of the field can be estimated by invoking the principle of ***conservation of energy*** (refer section 3.6). This section outlines the mathematical foundation of algorithms for calculating the ray paths and section 3.6 the calculation of the field amplitude.

As defined in section 3.2, a ray is the trajectory orthogonal to the wavefronts $H(\mathrm{x}) = \mathit{constant}$ and has direction $\nabla H(\mathrm{x})$. Equation (2.15) reveals the magnitude of $\nabla H(\mathrm{x})$ to be $\nu(\mathrm{x})$, and so the differential equation for the path of a ray is

$$\frac{d\mathrm{r}(s)}{ds} = \frac{\nabla\, H(\mathrm{x})}{\nu(\mathrm{x})}, \qquad (3.5)$$

where $\mathrm{r}(s)$ is the position of the ray at an arbitrary point and s is the arc length along the ray [Anderson and Kak, 1982]. If the curvature of the ray vector $\mathrm{r}(s)$ is considered, it can be shown that a ray bends towards a region of higher refractive index [Jones, 1986].

As described in section 3.4, the quintessential concept describing valid ray paths is Fermat's principle. These are paths that render the optical path length, given by the variational integral (3.4), stationary. The Euler equation constitutes a necessary condition for the existence of an extremal value of this integral [Anderson and Kak, 1982]. This can be expressed by the differential equation

$$\nu\,\frac{d^2\mathrm{r}(s)}{ds^2} + \left(\nabla\nu\frac{d\mathrm{r}(s)}{ds}\right)\frac{d\mathrm{r}(s)}{ds} = \nabla\nu(\mathrm{x}). \qquad (3.6)$$

Algorithms for tracing rays through continuous media are based on one of the four equations (2.15), (3.4), (3.5), (3.6) [Murch, 1990]. Examples of these are described by [Anderson and Kak, 1982; Norton, 1987]. Once the ray paths are known, the amplitude of the rays can be estimated as per section 3.6.

When the desired path from a starting point x_0 is constrained by an end point x, the problem is known as a *ray-linking* problem because the ray that links x_0 and x must be found. In this case the initial direction of the ray that defines the path intercepting x is unknown. Often this problem is solved in 2-D by shooting rays in various directions and refining the launch angle by Newton's method. Another ray-linking approach that utilises a knowledge of both x_0 and x links rays that are launched from both points. Alternatively, rather than solving this problem by the numerically intensive problem of ray tracing, the ray equation can be transformed into an implicit integral equation and iteratively solved to satisfy the given boundary conditions. The advantage in doing so would be any computational advantage gained from using a numerical integration routine rather than differential equation solvers employed by the more traditional ray tracing methods [Norton, 1987].

The problem of tracing rays through spherically symmetric media has been known to seismologists for some time. Early models of the earth considered it to be symmetric about its centre and even stratified. When the refractive index distribution is spherically symmetric, $\nu = \nu(\rho)$, the solution of the bent-ray tomographic problem is known, and in seismology is referred to as the Herglotz-Wiechert formula [Aki and Richards, 1980]. However, these formula are not robust for all media. The condition of a *low velocity zone*,

$$\frac{d(\nu\rho)}{d\rho} < 0, \tag{3.7}$$

violates the applicability of the Herglotz-Wiechert approach as the solution is non-unique. In the case of a low velocity zone there exists a region on the boundary called a *shadow zone* where no rays reach. This makes it impossible to trace rays to the shadow zone.

3.6 INTENSITY LAW OF GEOMETRIC OPTICS

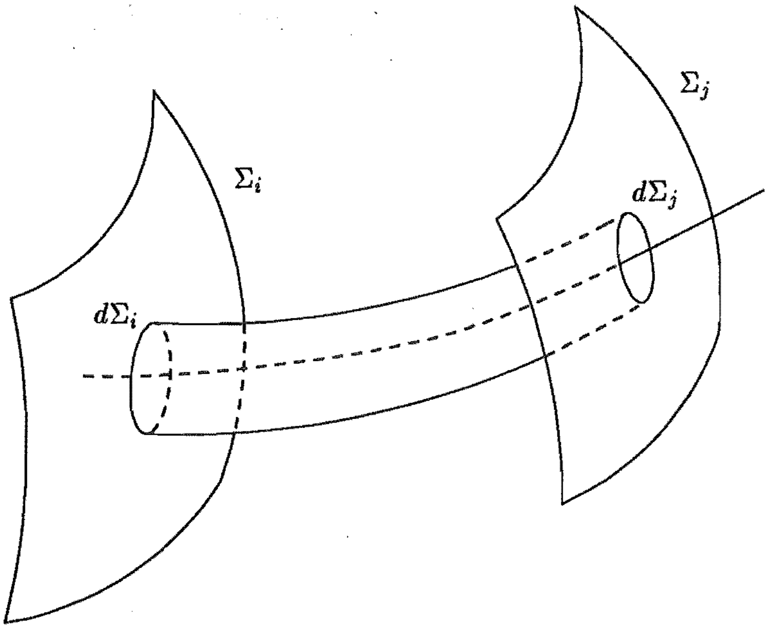


Figure 3.3 A tube of rays intersecting wavefronts Σ_i and Σ_j over the areas $d\Sigma_i$ and $d\Sigma_j$ respectively.

If the medium's attenuative properties can be neglected, then, the ray amplitudes can be calculated using the principle of *conservation of energy*. Consider a tube of rays intersecting the wavefront surface Σ_i over the region $d\Sigma_i$ (see figure 3.3). All rays emanating from $d\Sigma_i$ do so in the direction normal to the surface at their point of intersection and form a tube of rays intersecting the wavefront surface Σ_j over the region $d\Sigma_j$. From the law of conservation of energy, the intensity of the rays through the surfaces are related by

$$I_i \, d\Sigma_i = I_j \, d\Sigma_j$$

$$\text{or} \quad |A_i|^2 \, d\Sigma_i \;=\; |A_j|^2 \, d\Sigma_j, \tag{3.8}$$

where I_i and A_i are the intensity and the complex amplitude respectively of the wavefront at $d\Sigma_i$. The intensity law in geometrical optics is particularly convenient to apply when the rays travel through stratified media where the refractive index varies discontinuously [Keller, 1959]. Here the rays are straight in each homogeneous strata. The amplitude calculations at each wavefront are therefore considerably simplified. Consider a wavefront surface element $d\Sigma_i$ whose shape is characterised by two principle radii of curvature, ρ_1 and ρ_2 (see figure 3.4). Rays passing through $d\Sigma_i$, in a normal direction, form a tube of rays that intersect a second wavefront Σ_j over the area element $d\Sigma_j$. The principle radii are measured from the axial ray, AA', which connects the centre points of two wavefronts $d\Sigma_i$ and $d\Sigma_j$. Let r be the optical length of a ray between the two wavefronts, which as stated in section 3.4 is constant over the surfaces. The ratio in area between the two surface elements is given by [Jones, 1986]

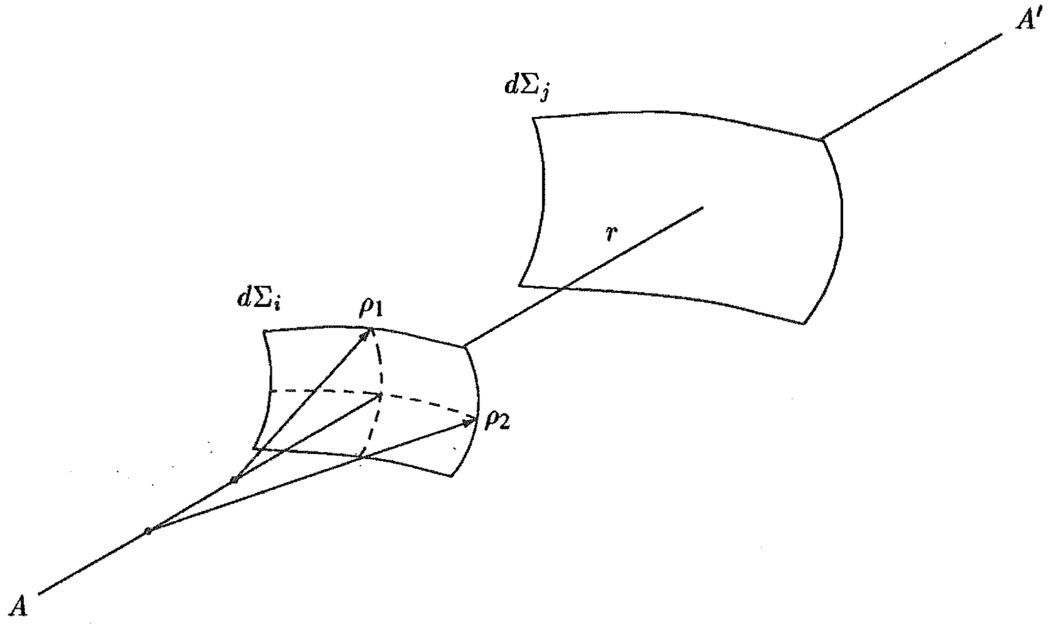


Figure 3.4 Ray propagation in a homogeneous region from wavefront surface $d\Sigma_i$ to $d\Sigma_j$.

$$\frac{d\Sigma_i}{d\Sigma_j} = \frac{\rho_1 \rho_2}{(\rho_1 + r)(\rho_2 + r)}. \tag{3.9}$$

Furthermore, a phase increment of $(-k\nu r)$ results when a ray travels an optical length r in a homogeneous medium of refractive index ν. Therefore, the complex amplitude of the field in a homogeneous medium at Σ_j, is related to the field at Σ_i through the intensity law (3.8) and through (3.9) by a phase factor $e^{-ik\nu r}$:

$$A_j \;=\; A_i \left[\frac{\rho_1 \rho_2}{(\rho_1 + r)(\rho_2 + r)}\right]^{\frac{1}{2}} e^{-ik\nu r}. \tag{3.10}$$

When the medium is stratified the ray will split into reflected and refracted rays when incident on the boundary between two media. Consequently, the formulation (3.10) must account for the split by introducing refraction and reflection coefficients which depend upon the boundary media and the angle of incidence.

Particular forms of the wavefronts can be realised by assigning the radii of curvature ρ_1 and ρ_2 certain values. If $\rho_1 = \rho_2$, the wavefronts are spherical. If one of the radii are infinite, then, the wavefronts are cylindrical and if both radii are infinite, the wavefronts are planar.

3.7 EIKONAL EQUATION

Recall from section 2.1.5 the eikonal equation

$$\nu^2(\mathrm{x}) = \nabla H(\mathrm{x}) \cdot \nabla H(\mathrm{x}), \tag{3.11}$$

where $H(\mathrm{x})$ defines the geometrical wavefront surface to the field in the medium of refractive index $\nu(\mathrm{x})$ at position x.

The eikonal equation is the basic equation of geometrical optics [Born and Wolf, 1980]. As mentioned in section 3.5, there are algorithms that are based on this equation for solving the direct problem of the field within a given refractive index distribution $\nu(\mathrm{x})$. Anderson and Kak [1982] introduce a ray tracing algorithm known as the method of characteristics (aptly named as the eikonal is sometimes referred to as the equation of the characteristics of the wave equation) based on the solution of differential equations. A formal solution to (3.11), if $\nu(\mathrm{x})$ is known, is the integral equation [Born and Wolf, 1980]

$$\begin{aligned} H(\mathrm{x}) &= \int_C \nu(\mathrm{x})\, ds \\ &= \int_{\mathrm{x}_0}^{\mathrm{x}} \nu(\mathrm{x})\, ds, \end{aligned} \tag{3.12}$$

where ds is the element of arc along a continuous (but not necessarily analytic) curve C, beginning at a point x_0 in Ω_{inc} and ending at the position x in the medium. These curves C have been referred to previously as ray paths and any such curve that is to yield a solution for $H(\mathrm{x})$ in (3.12) must satisfy Fermat's principle (refer section 3.4) [Born and Wolf, 1980].

Chapter 4 introduces a new technique based on the eikonal equation for solving the inverse problem (and not requiring a solution to a direct problem such as ray tracing) for a refractive index distribution $\nu(\mathrm{x})$ from measurements of the optical path length of rays passing through $\nu(\mathrm{x})$.

3.8 GEOMETRICAL THEORY OF DIFFRACTION

In 1962 Keller published the *geometrical theory of diffraction* (GTD) showing that diffracted rays were produced when a ray was incident on a boundary having small radii of curvature. He discovered this in a solution which he requested from a student of his, for wedge diffraction of a pulse. Upon comparing this solution to that obtained from Luneberg's theory, an extra circular wavefront was found present. Luneberg's theory was based on asymptotic expansions of the electromagnetic field as is geometrical optics theory and so Keller termed these extra rays *diffracted rays* [Keller, 1985].

Keller had discovered instances where diffracted rays were produced from sharp boundaries. New laws analogous to GO were deduced for the diffracted rays. In each instance, they were derived from a modified form of Fermat's principle.

The fundamental premise of the GTD is that the wavelength is so small that the wave propagation is entirely a local phenomenon. Thus, the diffracted rays are derived from the properties of the incident propagation and the diffracting media only within the immediate neighbourhood of the boundary, referred to as the *transition region*. Furthermore all fields, no matter how they are produced, must have the same local structure, namely, that of a plane wave. The diffracted field must therefore travel along rays also. These rays, once outside the immediate neighbourhood, behave like ordinary rays whose phase and amplitude along their paths is governed by GO theory [Keller, 1962].

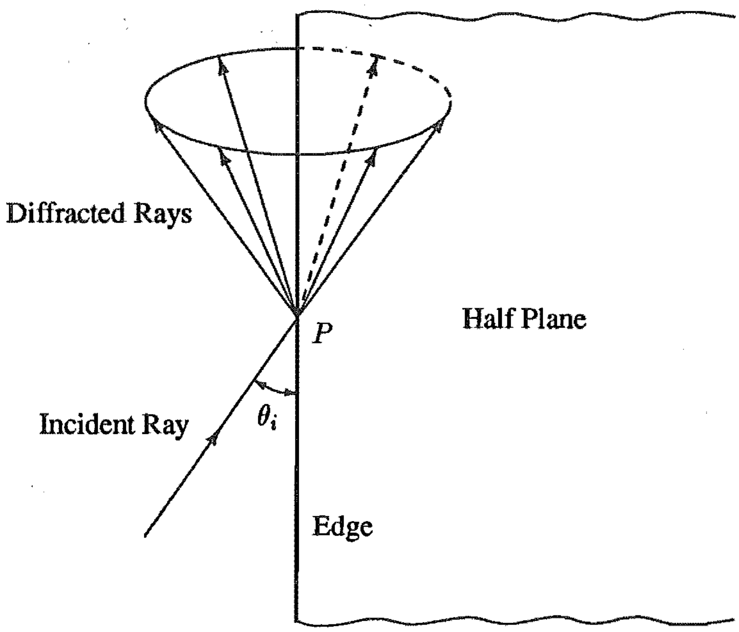


Figure 3.5 Diffraction of a ray incident upon an edge at point P. The angle of incidence is θ_i.

Consider a ray incident upon an edge at some point P as in figure 3.5. The law for diffraction at such an edge is a consequence of *Fermat's principle for edge diffraction*. An edge diffracted ray from one point in the media to another is a curve which has stationary length among all curves joining these points and having one point P on the edge. Resulting from this incident ray is a cone of *edge diffracted rays*. The rays lie on a cone whose semi-angle is equal to the angle of incidence θ_i. The incident and diffracted rays lie on opposite sides of the plane that is normal to the edge at P. When the two rays lie in different media the angle between the incident and diffracted rays and the normal plane is suggested by Snell's law of refraction (3.2).

Under GO, rays incident upon smooth discontinuities leave the discontinuity with a complex amplitude that can be obtained by multiplication of the incident field by a reflection or transmission coefficient. Keller [1962] expresses the complex amplitude of the diffracted field leaving a sharp discontinuity in terms of a diffraction coefficient D_{diff}. A review of available diffraction coefficients can be found in [Molinet, 1987a]. In homogeneous media the diffracted and incident amplitudes are related typically by

$$A_{diff} = D_{diff}\, A_{inc} \left[\frac{1}{r^{\alpha}}\right] e^{-ik\nu r}, \tag{3.13}$$

where A_{inc} and A_{diff} are the amplitudes of the incident and diffracted rays respectively, ν is the refractive index of the medium, r is the distance between the incident point P and the point where the diffracted field is observed, and α is a constant dependent on the type of sharp discontinuity. For edge diffraction where the wavefronts propagate cylindrically from the edge, the cross-sectional area of any tube of rays is proportional to r and so the amplitude is proportional to $r^{-1/2}$ as per (3.8) (in (3.13) $\alpha = 1/2$). For a ray incident upon an vertex, the diffracted rays travel out in all directions resulting in spherically propagating wavefronts. This is known as the law of *vertex diffraction*, the amplitude of these rays are proportional to r^{-1}. In the case of *surface diffraction*, where the incident ray grazes a smooth surface, it divides into two parts; one part of the incident energy continues in the same direction and the other part gives rise to a surface diffracted ray or creeping ray that travels along the surface with initial direction equal to the incident ray. At every point the surface ray sheds a diffracted ray tangentially as it propagates [Molinet, 1987a]. Surface diffracted rays obey a modified form of Fermat's principle for surface diffraction.

As suggested previously in this section, the diffracted field is defined as that part of the total field ψ whose behaviour cannot be explained under the GO approximations. This partition of the total field can thus be written

$$\psi = \psi_{go} + \psi_{diff}, \tag{3.14}$$

where ψ_{go} is that part explained by GO and ψ_{diff}, that part denoted the diffracted field contains the rest of ψ. Diffraction coefficients D_{diff} are derived to enable the diffracted field ψ_{diff} to be estimated. Functional forms of the diffraction coefficients can be found by subtracting from exact solutions using asymptotic expansions, that part accounted for by geometrical considerations [Molinet, 1987a]. Those problems where exact asymptotic solutions have been obtained are referred to as canonical diffraction problems. Since many complicated bodies are made up of simple shapes, an approximate solution to its diffraction pattern can be obtained by applying diffraction laws to the individual shapes (which go to make up the body) and then summing the individual contributions. This technique is based on the assumption that diffraction phenomena is localised around the neighbourhood of the discontinuities [James, 1986b; Molinet, 1987b]. Higher order diffraction terms can be incorporated by accounting for diffracted fields that are incident on other discontinuities.

The GTD extended the domain of optical and electromagnetic phenomena that could be explained in ray optical terms. However, there were still difficulties in the theory. GTD and GO yield fields that are infinite at focal points and caustic surfaces, including edges, corners and curved surfaces of diffracting bodies. These singularities arise because of the asymptotic nature of GO and GTD [Keller, 1985]. A significant limitation of the GTD is its inability to provide solutions for the field within transition regions. Keller's GTD gives invalid solutions for the diffraction coefficients within transition regions, giving a non-uniform solution for the diffracted field. Two main theories were developed within the context of GTD to give uniform solutions. These were the Uniform Geometrical Theory of Diffraction and the Uniform Asymptotic Theory [James, 1986b]. Many of the defects of GTD have been overcome by such extensions to it. Other methods such as boundary layer theory have been combined in solutions to cope with defects in GTD. Details of many of these can be found in [Felsen, 1986; James, 1986b; Ruan and Felsen, 1986; Molinet, 1987b].

3.9 VALIDITY OF THE RAY APPROACH

As intimated earlier, the assumptions of geometrical optics are valid when the wavelength λ is small compared to significant dimensions of the objects with which the field interacts. All lengths of interest (such as the length over which the index of refraction varies appreciably) are large compared to the wavelength, and therefore, the spacing between rays changes very little over a distance equal to λ. This is a general definition for the validity of applying a geometrical optics approach that has been accepted by many authors [Born and Wolf, 1980; Deschamps, 1972; Jones, 1986; James, 1986a]. As noted by Born and Wolf [1980], at optical wavelengths at least, the regions for which this simple geometrical model is inadequate are an exception rather than a rule; in fact for most optical problems geometrical optics furnishes at least a good starting point for more refined investigations.

One of the fundamental limitations of a ray model, where rays paths are governed by a changing refractive index, is that rays may not enter some regions of high refractive index obeying Fermat's principle which, generally speaking, minimises the optical path length of a ray [McKinnon and Bates, 1980]. Consequently, these regions are impossible to image and solutions obtained for the refractive index distribution are rendered partially non-unique [Berryman, 1990]. A second limitation is pointed out by Aki and Richards [1980] in the context of seismology, where the distribution examined has a low velocity zone (refer section 3.5, equation (3.7)). In this situation the low velocity zone creates a shadow zone on the boundary of the distribution that no rays reach and where no measurement data is available. This also introduces a region in which the solution for the refractive index is non-unique. Geometrical optics also fails to account for the existence of fields in the shadow region, normally attributed to the phenomena of diffraction [VanBladel, 1985].

When the wavelength is sufficiently small that any variation in the medium occupies many wavelengths except within transition regions of scattering centres, a reasonable hypothesis is that in local regions the field behaves as if it were in a homogeneous medium. Thus, locally, the field may be expected to look like a plane wave [Jones, 1986]. These conditions satisfy those required for application of the geometrical theory of diffraction, thereby including rays generated by scattering centres such as edges or corners (edge or corner diffracted rays), by critical refraction and shedding back into the host medium (lateral rays), by glancing incidence on smooth surfaces (surface diffraction) and so partly accounting for the existence of fields in the shadow region. However, as mentioned in section 3.1, the geometrical theory of diffraction has not found wide application in solving inverse problems. While this is so, there is general evidence to suggest that diffraction can be accounted for to high accuracy in ray optical terms [Bates *et al.*, 1991; Felsen, 1985].

Results obtained under the theory of geometrical optics and geometrical diffraction are questionable in transition regions or where rays diverge or converge rapidly and interact so violently as to destroy the independent local plane-wave character of each [Felsen, 1985]. This is particularly true at caustic points where the rays cross and the ray theory wrongly predicts an infinite field intensity.

In terms of applied ray theory, where the rays are validly interpreted as straight, there is an abundance of conventional theory developed for interpreting ray measurements that give useful information as to the body under study. Examples of such are; echo location principles used in radar [Cutrona, 1970] and sonar [Urick, 1975], industrial ultrasonic non-destructive testing [Bond and Reynolds, 1987; Thompson and Chementi, 1985], medical B-scan [Wells, 1977] and seismology [Aki and Richards, 1980; Schoenberger, 1984]; transmission (forward scattering) imaging used with X-ray CT [Bates *et al.*, 1983; Herman, 1980], ultrasonic transmission CT [Greenleaf, 1983; Robinson and Greenleaf, 1985] and radiopharmaceutical imaging (emission CT) [Jaszczak, 1988; Keyes, 1987].

Although extensively applied, there exists a number of desirable theoretical improvements to the ray model used by some of the aforementioned applications. What models based on straight ray theory can not presently do is reconstruct appreciable variations in refractive index distribution. Medical ultrasonic imaging (B-scan), ultrasonic transmission CT, ultrasonic non-destructive testing and seismology, among others, all can in general experience appreciable bending of the ray paths. These applications would benefit with an increase in their spatial resolution from a manageable model for bent rays.

CHAPTER 4

A NEW SOLUTION TO THE BENT-RAY PROBLEM

4.1 INTRODUCTION

A large number of scientifically and technologically important situations exist where wave motion can be sufficiently characterised by curved rays [Aki and Richards, 1980; Arnold, 1986; Bates *et al.*, 1991; Cutrona, 1970; Dines and Lytle, 1979; Keller, 1962; Santosa *et al.*, 1984; Sengbush, 1983; Wells, 1977]. In this chapter, an algorithm is introduced for the reconstruction of a refractive index or velocity distribution illuminated by wave motion that is characterised in this way. Reconstructions are derived from time-of-flight measurements made of wave emanations passing through the distribution. All measurements in this chapter are referred to as optical path lengths, as defined previously in section 3.4 for a refractive index distribution. Solutions to this type of inverse problem, known as bent-ray CT, have previously been proposed, particularly in the context of ultrasonic transmission tomography and seismic tomography. Iterative approaches have been developed for reconstruction of a refractive index distribution involving ray tracing [Berryman, 1990; Cha and Vest, 1981; Dines and Lytle, 1979; Lytle and Dines, 1980; McKinnon and Bates, 1980; Schomberg, 1978] (see section 3.5 for a description of ray tracing), which, based on Fermat's principle of minimum propagation time (section 3.4), is the estimation of the ray paths through the distribution. Perturbation solutions that do not require ray tracing have also been proposed [Bates and McKinnon, 1979; Norton and Linzer, 1982]. However, these are limited by the assumption that the rays do not deviate (bend) far from straight paths; that is, the distribution contains only small variations in refractive index.

Bent-ray CT reconstruction of objects with which wave motion is characterised by rays, would benefit by an approach that does not rely on tracing their paths and can still cope with appreciable spatial variations in the refractive index [Bates *et al.*, 1991]. In this chapter, such an approach is described and evidence presented to support its validity. The essence of this approach is a simple ray model for wave motion passing through the distribution. A novel ray description is adopted which doesn't require the ray paths to be known. This allows the refractive index to be found by iterative solution of linear equations and does not require the computer intensive procedure of ray tracing, which normally accompanies iterative solutions to problems of this type. Preliminary results show that this method is capable of handling appreciable spatial refractive index variations in large bodies.

As intimated in section 3.9 there are limitations on the ray approach to computed tomography. McKinnon and Bates [1980] have shown that there are distributions which contain regions through which no *minimum propagation time* rays will propagate. Rays selectively traverse that region which minimises their propagation time, and consequently there can be contrived distributions for which some regions are not traversed, and are therefore impossible to image. This can render the solution partially non-unique [Berryman, 1990].

Section 4.2 defines bent-ray CT as a general problem in reconstructing refractive index distributions and develops the theoretical foundation for the new reconstruction theory. Section 4.3 outlines an algorithmic approach for obtaining reconstructions of such distributions. Bent-ray transmission CT is then described in section 4.4 as a specific case of the general bent-ray CT problem. This is put into a framework where finite measurement data is available for a 2-dimensional refractive index distribution. In section 4.5, an algorithm is proposed for bent-ray transmission CT. The specific algorithm for the reconstruction of circularly symmetric distributions is given in section 4.6. Results of computer simulations are presented in section 4.7 for some circularly symmetric distributions. A discussion on physical relationships for wave motion in a refractive index distribution developed in section 4.2 and not utilised to obtain these results, but which are thought important for an efficient solution to the more general problem of reconstructing asymmetric distributions, is deferred to section 4.8.

4.2 BENT-RAY CT

As intimated in sections 3.2 and 3.9, wave motion in a refractive index distribution can be characterised in terms of bundles of rays. Through sections 3.2 and 3.7 it is argued that the rays can themselves be characterised satisfactorily by the eikonal equation (2.15) relating the refractive index and the ray paths. This relationship is now developed and enforced as the theoretical foundation of a reconstruction algorithm aimed at solving the bent-ray problem.

In order to define what is referred to here as the bent-ray CT problem, consider the adopted geometry of figure 4.1. It is convenient to denote a refractive index distribution, $\nu(P)$, to be reconstructed in two dimensions, where P is an arbitrary point in the medium Ω. The extension to three dimensions of the method developed hereafter is obvious. The region of interest has support which is entirely contained in the region Ω_{-}. It is also appropriate to introduce arbitrary points Q and Q' to indicate, respectively, points where the incident wave motion originate and is observed. In accordance with the notation introduced in section 1.3, these points are located within the regions Ω_{inc} and Ω_{obs}, respectively. The general bent-ray CT problem defines both these regions equivalently to contain all points in Ω. It is noted that no measurements are possible when Q and Q' coincide, thereby adhering to (2.2). By definition (4.2) any such time-of-flight (or optical path) measurements are zero. Gradient operators ∇_P and ∇_Q imply differentiation with respect to the coordinate axes of P and Q, respectively.

Rays emanating from an emitter Q spread out through Ω and the field is observed at various receiver points Q'. The real phase function or eikonal (as per section 2.1.5) of these rays at the observation point is here written as $H = H(Q', Q)$. The optical path length of a ray (the dashed line), propagating from Q to Q', at an arbitrary point P along the ray is, from (3.4),

$$H(P, Q) = \int_{Q}^{P} \nu \, ds, \tag{4.1}$$

where ds is the element of arc length along the ray path. With these definitions in mind, the

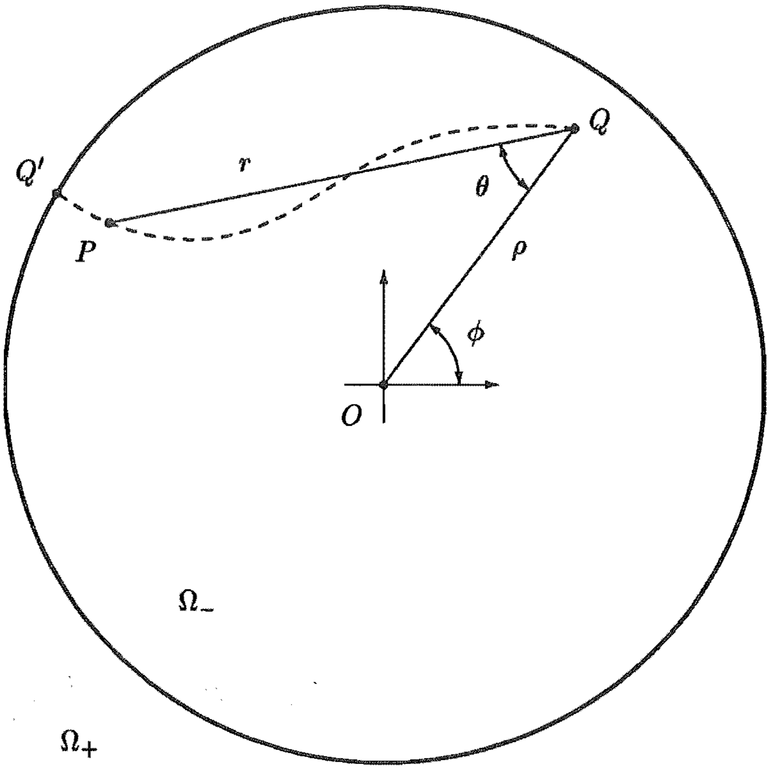


Figure 4.1 Coordinates for 2-D bent-ray CT. The dashed line represents the actual ray path from an emitter at Q to a receiver at Q' through some arbitrary point P on its path.

bent-ray CT problem is posed as

> Given $H(Q',Q)$ for sufficient pairs (Q',Q) of receiver-emitter points to permit reconstruction of ν to whatever spatial resolution is desired, estimate the ν for all points P contained within Ω_-.

The phase function $H(P,Q)$ is required to estimate ν from the eikonal equation (2.15), and can be further constrained. The optical path length from an emitter Q to an arbitrary point P in Ω must be zero as the separation between P and Q approaches zero :

$$\begin{aligned} H(P,Q) &= 0 \quad \text{as } P \to Q, \text{ or, alternatively} \\ H(Q,Q) &= 0. \end{aligned} \tag{4.2}$$

The path of a ray from the emitter placed at an arbitrary point Q in Ω to a point P satisfies Fermat's principle; it is stationary with respect to infinitesimal variations in its path (see section 3.4). It is also true that if the emitter and receiver points are exchanged, the reciprocal path to P satisfying Fermat's principle is unchanged. Alternatively stated, the path taken by energy propagating from one point to another in Ω is the same as that if the energy were propagating in the opposite direction. This symmetry can be expressed

$$H(P,Q) = H(Q,P). \tag{4.3}$$

As required by the eikonal (2.15), the phase function H at any point P is directly related to the refractive index at that point by

$$\nabla_P\, H(P,Q) \cdot \nabla_P\, H(P,Q) \;=\; \nu^2(P). \tag{4.4}$$

However, it is known that the refractive index at any point P in Ω, is not dependent on where the wave motion originated, implying that

$$\nabla_Q\, \left[\, \nu(P)\, \right] \;=\; 0 \tag{4.5}$$

Substituting (4.5) into (4.4) yields

$$\nabla_Q\, \left[\, \nabla_P\, H(P,Q) \cdot \nabla_P\, H(P,Q)\, \right] \;=\; 0. \tag{4.6}$$

Referring to the general position of an emitter point Q by the coordinates $(\rho;\, \phi)$ (see figure 4.1), the constraint equation (4.6) can be expressed in terms of each independent coordinate :

$$\frac{\partial}{\partial \rho}\, \left[\, \nabla_P\, H(P,Q) \cdot \nabla_P\, H(P,Q)\, \right] \;=\; 0, \tag{4.7}$$

$$\frac{\partial}{\partial \phi}\, \left[\, \nabla_P\, H(P,Q) \cdot \nabla_P\, H(P,Q)\, \right] \;=\; 0. \tag{4.8}$$

As a consequence of the given phase function measurements, $H(Q',Q)$, any solution for $H(P,Q)$ can be constrained to be consistent with faithful measurements;

$$H(P,Q) \;=\; H(Q',Q) \quad \text{as } P \rightarrow Q'. \tag{4.9}$$

If the refractive index of Ω_- is known on the region Ω_{obs} to be $\nu(Q')$, then this information can be incorporated as an additional set of constraints. It is surely likely to help stabilise an attempted solution, particularly at points close to Ω_{obs}.

$$\begin{gathered} \nabla_P\, H(P,Q) \cdot \nabla_P\, H(P,Q) \;=\; \nu^2(Q') \\ \text{as } P \rightarrow Q' \text{ for all known } \nu(Q') \text{ in } \Omega_{obs}. \end{gathered} \tag{4.10}$$

The basic equation for reconstructing $\nu(P)$ from $H(P,Q)$ is given in (4.4). The theoretical foundation of a solution for $H(P,Q)$ in 2-D are the known relationships (4.2), (4.3), (4.7), (4.8) and measurements (4.9), (4.10). An algorithm is required to solve for $H(P,Q)$ that is appropriate to the form of these relationships and incorporating the measurement data. Although this is clearly not a trivial task, there does not appear to be any fundamental difficulty, and such an algorithm is developed and described in section 4.3.

4.3 BENT-RAY CT RECONSTRUCTION ALGORITHM

As previously stated in section 4.2, the refractive index throughout Ω_- is to be characterised by a solution for the phase function $H(P,Q)$, of wave motion that has interacted with the region Ω_-. Clearly, any solution will be based on measurements (4.9), and if available (4.10). However, in practical problems only a limited number of measurements can be taken and the reliability of a solution for the refractive index distribution $\nu(P)$ in Ω_-, even though it is consistent with the finite measurements taken, is not in general guaranteed [Crowther and Klug, 1971]. The constraints (4.2), (4.3), (4.7), (4.8) help to reduce the solution space containing $\nu(P)$.

In order to solve for the phase function $H(P,Q)$, it is necessary to adopt a suitable representation of that function which is amenable to methods of numerical solution. The choice of representation is clearly important; it must not only be adequate to represent the spatial variation of $H(P,Q)$ but must also be able to represent those variations in refractive index given by $\nabla_P\ H(P,Q) \cdot \nabla_P\ H(P,Q)$ in (4.4). From a computational point of view, efficiency is also an issue to be addressed in considering the representation to be adopted. The fundamental efficaciousness of a chosen representation will depend on how well it accords with the physics of the situation.

A general coordinate system alluded to in figure 4.1 is adopted here for the representation of H, where the general position of an emitter point Q is $(\rho;\, \phi)$ and the position of a general point P in the medium is understood to have coordinates $(r;\, \theta)$ with respect to Q. Considering the case of a homogeneous refractive index distribution whereby $\nu(P)$ reduces to its average value $\bar{\nu}$ (throughout Ω_-), then $H(P,Q) = \bar{\nu}\, r$, where r is the length of the straight path to P from emitter Q (see figure 4.1). Because this form satisfies the four constraints, (4.2), (4.3), (4.7) and (4.8), it may be appropriate to base a general representation for H, valid for arbitrary variations of $\nu(P)$, on powers of r, for example,

$$H(P,Q) = \sum_{n=1}^{\infty} H_n(P,Q)\, r^n. \tag{4.11}$$

where suitable forms of the functions $H_n(P,Q)$ must be found. The representation of $H(P,Q)$ in (4.11) includes terms for $n > 0$ only so that it immediately satisfies constraint (4.2), since $r^n \to 0$ when $P \to Q$. Note, for example, the fact that

$$H_n(P,Q) = \sum_{l,m} A_{l,m,n}\ \left[\, R_l(P)\, S_m(Q) + R_l(Q)\, S_m(P)\,\right] \tag{4.12}$$

ensures $H(P,Q)$ satisfies (4.3), where the $A_{l,m,n}$ are constant expansion coefficients and the $R_l(P)$ and $S_m(P)$ are suitable basis functions.

With regard to satisfying the homogeneous constraint equations (4.7) and (4.8), that constrain any expression for the refractive index $(\nabla_P H \cdot \nabla_P H)$ to be independent of the orthogonal emitter coordinates, it would be advantageous to express $\nabla_P H \cdot \nabla_P H$ in terms of basis functions orthogonal in the emitter coordinates. If this was so, then satisfying (4.7) and (4.8) would reduce to ensuring that the coefficients of each orthogonal function sum to zero thereby ensuring that the eikonal is independent of each emitter coordinate. Computational gain may arise from a reduction in the number of terms involved in satisfying (4.7) and (4.8). The efficiency of this process depends largely on an appropriate choise of basis functions.

Clearly, what is required in a reconstruction algorithm is a robust method of solving for the $A_{l,m,n}$ in $H(P,Q)$ from the known relationships concerning H and ν, by which the refractive index distribution can then be estimated through equation (4.4). Recalling measurements (4.9), substituting these equations into the representation formed from (4.11) and (4.12) gives a set of linear relationships between the expansion coefficients, $A_{l,m,n}$, of the representation, and the measured path lengths between receiver-emitter pairs :

$$H(Q',Q) = \sum_{l,m,n=1} A_{l,m,n} \; [\, R_l(Q')\, S_m(Q) + R_l(Q)\, S_m(Q') \,] \; r^n. \qquad (4.13)$$

In solving for the $A_{l,m,n}$, equations (4.2), (4.3), (4.7), (4.8) and (4.13) can be invoked as well as (4.10) if measurements of ν are available. If the representation chosen is able to naturally satisfy one or other of the constraints (4.2), (4.3), (4.7) and (4.8), the number of equations to be solved is reduced. This is particularly advantageous if the set of equations overdetermine the coefficients, in which case reducing equation numbers can reduce the solution space. If (4.7) and (4.8) (and (4.10) if available) were linear in the $A_{l,m,n}$, the expansion coefficients could be solved by matrix inversion of the system of linear equations. They are, however, unavoidably quadratic in the $A_{l,m,n}$, and so an appropriate iterative procedure must be found. Again, the efficaciousness and efficiency of such a procedure is likely to depend strongly on the choice of basis functions $R_l(P)$, $S_m(P)$ and r^n and how well they accord with the physics of the situation.

4.4 BENT-RAY TRANSMISSION CT

In section 4.3 an algorithm has been described for the general bent-ray CT problem defined in section 4.2. In this section, a more specific problem is examined, namely that of reconstructing ν over a region Ω_-, shown in figure 4.2 to be of radius a, from wave propagation that has completely traversed Ω_-. Specifically the problem examined is one in which the region has been probed by emitters and receivers placed on the boundary of Ω_-. That is, the points Q and Q' lie on the coinciding regions Ω_{inc} and Ω_{obs}, respectively, as shown in figure 4.2. Again it is pointed out that no measurements are taken whereby the points Q and Q' are identical since these optical paths are known to be zero in agreement with (4.2). It is envisaged that a short enough pulse can be emitted such that $H(Q',Q)$ can be measured to a sufficient level of accuracy so that the quality of measurements is not responsible for limiting the reconstruction resolution. Practically speaking it is impossible to have anything other than a finite number of receiver-emitter pairs. The positions of this finite number of receiver and emitters points are denoted by Q'_m and Q_l, respectively, where the subscripts m and l refer to the location of each point on the boundary of Ω_-.

The bent-ray transmission CT problem is posed thus :

> Given $H(Q',Q)$ for wave motion that has traversed Ω_- from one point on its boundary to another, and for sufficient pairs (Q',Q) of receiver-emitter points

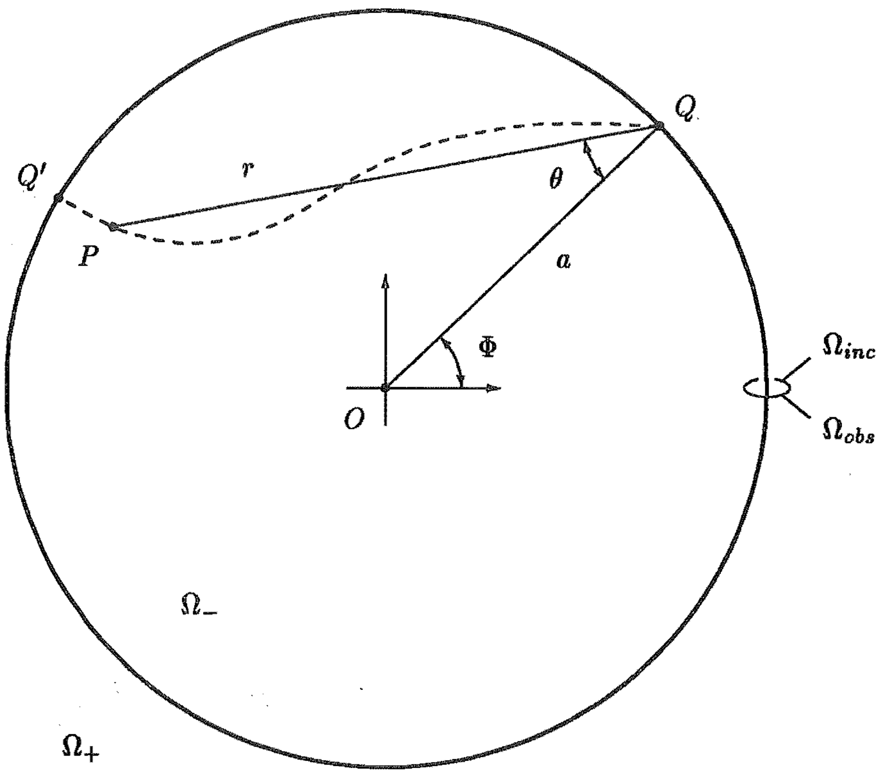


Figure 4.2 Coordinates for 2-D bent-ray transmission CT. The dashed line represents the actual ray path from an emitter Q to a receiver at Q' through some arbitrary point P on its path. Both regions Ω_{inc} and Ω_{obs} lie on the boundary of radius a.

> to permit reconstruction of the spatial refractive index to whatever resolution is desired, estimate the ν for all points P contained in Ω_-.

In addition, if ν is known at points on the boundary, then this information can be employed to add stability when searching for a solution.

In choosing a representation for the phase function H, the emitter is assumed to be at a point Q where ν is continuous and well behaved, implying that $H(P, Q)$ must be well behaved at P in the neighbourhood of Q, then both a Fourier and power series are analytically acceptable expansions of $H(P, Q)$. It is here postulated that a representation which is a combination of both of the above is adequate for reconstruction of non-trivial refractive index distributions. The representation is, therefore,

$$H(P,Q) = \left(\sum_{l=-\infty}^{\infty} A_l \, e^{il\Phi} \right) r + \sum_{l,m=-\infty}^{\infty} \sum_{n=2}^{\infty} B_{l,m,n} \, r^n \, e^{i(2m\theta + l\Phi)}, \tag{4.14}$$

where the coordinates $(r; \theta; \Phi)$ are shown in figure 4.2. The A_l and $B_{l,m,n}$ are complex-valued expansion coefficients required to satisfy $A_l^* = A_{-l}$ and $B_{l,m,n}^* = B_{-l,-m,n}$, where * denotes complex conjugation, to ensure that $H(P, Q)$ is real.

With regard to the representation chosen for $H(P,Q)$ given in (4.14); the power series in r is from (4.11) and is adopted for the reasons outlined in section 4.3. A general Fourier series expansion is adopted for angular variation of H in the coordinates θ and Φ. To ensure that the subsequent expression for the refractive index, $\nabla_P H \cdot \nabla_P H$ in (4.15), at emitter points on the boundary ($r = 0$) is a function of the position on the boundary Φ only and not θ, the power series term r^n, where $n = 1$, that accounts for the refractive index on the boundary has no summation over m in $e^{i2m\theta}$. It can be seen from (4.14) that the first term is independent of θ, and from (4.15), that this term is responsible for the refractive index at emitter points on the boundary ($r = 0$). This is pointed out later in (4.20). It can be clearly seen from figure 4.2 that the range of the coordinates $(r; \theta; \Phi)$ are $0 \leq r \leq 2a$, $-\pi/2 \leq \theta \leq \pi/2$ and $0 \leq \Phi < 2\pi$ respectively. Because of its Fourier series character, the representation (4.14) can accommodate discontinuities in $\nabla_P H(P,Q)$ along isolated curves inside the circle of radius a.

On substitution of (4.14) into (4.4), an expression for the square of the refractive index is obtained :

$$\begin{aligned}
\nu^2(P) &= \nabla_P H(P,Q) \cdot \nabla_P H(P,Q) \\
&= \left(\sum_{l=-\infty}^{\infty} A_l \, e^{il\Phi} \right)^2 \\
&\quad + 2 \left(\sum_{l=-\infty}^{\infty} A_l \, e^{il\Phi} \right) \sum_{l,m=-\infty}^{\infty} \sum_{n=2}^{\infty} B_{l,m,n} \, n \, r^{n-1} \, e^{i(2m\theta + l\Phi)} \\
&\quad + \sum_{l,l',m,m'=-\infty}^{\infty} \sum_{n,n'=2}^{\infty} B_{l,m,n} \, B_{l',m',n'} \, r^{n+n'-2} \, (nn' - 4mm') \, e^{i[2(m+m')\theta + (l+l')\Phi]}.
\end{aligned} \tag{4.15}$$

With the bent-ray transmission CT problem described here, there are three constraint equations applicable from section 4.2; (4.2), (4.3) and (4.8). The representation (4.14) naturally satisfies (4.2). Expression (4.15) can be rewritten as a linear expression in terms of the orthogonal functions chosen to represent the dependence of $H(P,Q)$ on emitter position Q (which varies only as a function of Φ) :

$$\nabla_P H(P,Q) \cdot \nabla_P H(P,Q) \; = \; \sum_{l=-\infty}^{\infty} C_l \, e^{i2\pi l\Phi}, \tag{4.16}$$

where the C_l are the coefficients of each orthogonal function $e^{i2\pi l\Phi}$. Constraint (4.8) will be satisfied if the coefficients C_l for all functions dependent on Φ are zero, or equivalently, the Fourier decomposition of (4.15) with respect to the position Φ is zero at all points P ;

$$\int_0^{2\pi} \left[\nabla_P H(P,Q) \cdot \nabla_P H(P,Q) \right] \, e^{-i2\pi\mu\Phi} \, d\Phi \; = \; 0 \quad \text{for all} \quad 0 < |\mu| < \infty, \tag{4.17}$$

where μ is a non-zero integer.

As for constraint (4.3), the symmetry in exchanging receiver-emitter points can be rewritten in terms of expression (4.14) by substituting $-\theta$ for θ and replacing Φ with the position Φ' of the receiver with respect to the origin, r remains unchanged. This can be written as

$$\begin{aligned}\left(\sum_{l=-\infty}^{\infty} A_l\, e^{il\Phi}\right) r + \sum_{l,m=-\infty}^{\infty} \sum_{n=2}^{\infty} B_{l,m,n}\, r^n\, e^{i(2m\theta+l\Phi)} \\ = \left(\sum_{l=-\infty}^{\infty} A_l\, e^{il\Phi'}\right) r + \sum_{l,m=-\infty}^{\infty} \sum_{n=2}^{\infty} B_{l,m,n}\, r^n\, e^{i(-2m\theta+l\Phi')}.\end{aligned} \tag{4.18}$$

This symmetry can be extended to any two points within Ω_- lying on the ray path. However, it is conceptually more difficult to ensure symmetry of ray paths when one of the points P lies within Ω_- as the path is unknown. By definition, the emitter point Q must lie on the boundary and not within Ω_-. This means Q and P can not be exchanged within the present representation.

From both a theoretical and a practical viewpoint, it is convenient to take the emitter points and the receiver points to be equally spaced. Consider $(2L+1)$ of each, where L is a finite positive integer. In terms of the notation established for emitter and receiver points in section 4.4, the data for this problem is taken to be the quantities

$$H_{l,m} = H(Q'_m, Q_l), \tag{4.19}$$

and for the boundary conditions it is assumed the refractive index to be known at all emitter points Q_l ;

$$\nu_l^2 = \nabla_P H(Q_l, Q_{l'}) \cdot \nabla_P H(Q_l, Q_{l'}). \tag{4.20}$$

The angular coordinates of Q'_m and Q_l are $\Phi'_m = 2m\pi/(2L+1)$ and $\Phi_l = 2l\pi/(2L+1)$, respectively, where the integers m and l span the ranges $0 \le |m| \le L$ and $0 \le |l| \le L$.

To aid algorithmic efficiency and numerical stability, a boundary series that is directly related to the known refractive index distribution on the boundary has been included in the representation (4.14). As mentioned, it is the first term in the power series, r^n where $n = 1$ in (4.14). This relationship is conveniently assured since at all emitter points $r = 0$, leaving the boundary series to represent the refractive indices at these points :

$$\begin{aligned}\nu_l^2 &= \nabla_P H(Q_l, Q_l) \cdot \nabla_P H(Q_l, Q_l) \\ &= \left(\sum_{l'=-\infty}^{\infty} A_{l'}\, e^{il'\Phi_l}\right)^2.\end{aligned} \tag{4.21}$$

It can then be seen, by appealing to the theory of the discrete Fourier transform (refer section 2.3.1.3), that the $A_{l'}$ can be found directly from the finite data, ν_l, through (2.49);

$$A_{l'} = \frac{1}{2L+1} \sum_{l=-L}^{L} \nu_l \, e^{-i2\pi l'\Phi_l} \quad \text{for} \quad -L \leq l' \leq L, \tag{4.22}$$

and the remaining problem is to solve for the $B_{l,m,n}$ in (4.14). The finite limits on l in (4.22) are in accordance with the finite data and the properties of the discrete Fourier transform. $H(P,Q)$ can now be approximated in (4.14) by replacing the infinite limits on the summation over l by $-L$ and L. Similarly, the infinite summations over m and n are replaced by finite integers M and N, respectively;

$$H(P,Q) \approx \left(\sum_{l=-L}^{L} A_l \, e^{il\Phi} \right) r + \sum_{l=-L}^{L} \sum_{m=-M}^{M} \sum_{n=2}^{N} B_{l,m,n} \, r^n \, e^{i(2m\theta + l\Phi)}. \tag{4.23}$$

The magnitude of the positive integers M and N are chosen so that (4.23) can sufficiently represent not only the spatial variation of $\nu(P)$, but also that of $H(P,Q)$. The question of an optimal representation is not treated here. However, it is noted that the available resolution of the representation, which depends on L, M and N in a convoluted manner, is bound by the finite data set, (4.19) and (4.20), made available to characterise $\nu(P)$. Similarly, the finite representation of $\nabla_P H(P,Q) \cdot \nabla_P H(P,Q)$ is

$$\begin{aligned} \nu^2(P) &= \nabla_P H(P,Q) \cdot \nabla_P H(P,Q) \\ &\approx \left(\sum_{l=-L}^{L} A_l \, e^{il\Phi} \right)^2 \\ &\quad + 2 \left(\sum_{l=-L}^{L} A_l \, e^{il\Phi} \right) \sum_{l=-L}^{L} \sum_{m=-M}^{M} \sum_{n=2}^{N} B_{l,m,n} \, n \, r^{n-1} \, e^{i(2m\theta + l\Phi)} \\ &\quad + \sum_{l,l'=-L}^{L} \sum_{m,m'=-M}^{M} \sum_{n,n'=2}^{N} B_{l,m,n} \, B_{l',m',n'} \, r^{n+n'-2} \, (nn' - 4mm') \, e^{i[2(m+m')\theta + (l+l')\Phi]}, \end{aligned} \tag{4.24}$$

and the discrete approximation of (4.17) is

$$\sum_{l=-L}^{L} \left[\nabla_P H(P,Q_l) \cdot \nabla_P H(P,Q_l) \right] \, e^{i2\pi\mu\Phi_l} = 0 \quad \text{for} \quad 0 < |\mu| \leq L, \tag{4.25}$$

where the integral of (4.17) has been approximated here by a *Riemann sum*. Conservation of the path symmetry between emitter-receiver pairs (4.18) can likewise be rewritten in terms of the finite representation and measurement points chosen, Φ_l and Φ'_m :

$$\begin{aligned} &\left(\sum_{l'=-L}^{L} A_{l'}\, e^{il'\Phi_l}\right) r + \sum_{l'=-L}^{L}\ \sum_{m'=-M}^{M}\ \sum_{n'=2}^{N} B_{l',m',n'}\, r^{n'}\, e^{i(2m'\theta+l'\Phi_l)} \\ &= \left(\sum_{l'=-L}^{\infty} A_{l'}\, e^{il'\Phi'_m}\right) r + \sum_{l'=-L}^{L}\ \sum_{m'=-M}^{M}\ \sum_{n'=2}^{N} B_{l',m',n'}\, r^{n'}\, e^{i(-2m'\theta+l'\Phi'_m)}. \end{aligned} \tag{4.26}$$

The theoretical foundation, or, set of relationships which leads to a solution for $H(P,Q)$, has been developed in terms of a chosen 2-D representation. What remains is an algorithmic treatment that is appropriate to solving the unknown coefficients from these relationships.

4.5 TOWARDS AN ALGORITHM FOR BENT-RAY TRANSMISSION CT

The bent-ray transmission CT problem can be considered solved when a number of the $B_{l,m,n}$, sufficient to characterise $\nu(P)$ to the desired resolution, have been determined to an adequate accuracy. The main obstacle to the development of a manageable algorithm is (4.25) being quadratic in the $B_{l,m,n}$. It seems mandatory to derive an appropriate linearisation of (4.25). This section introduces an iterative approach to a solution for the bent-ray transmission CT problem, based on the linearisation of (4.25).

The iterative approach proposed here rewrites the $B_{l,m,n}$ in terms of two components, $X_{l,m,n}$ and $y_{l,m,n}$, where

$$B_{l,m,n} = X_{l,m,n} - y_{l,m,n}, \tag{4.27}$$

on the understanding that, at each iteration, the $X_{l,m,n}$ are known and the $y_{l,m,n}$ are unknown and are to be estimated. The $X_{l,m,n}$ are taken to represent what is known of the $B_{l,m,n}$ that is gathered from subsequent iterations estimating the unknown $y_{l,m,n}$. The first iteration does not require an initial estimate for the $X_{l,m,n}$. The initial $X_{l,m,n}$ are taken to be zero. By substitution of (4.27) into (4.24) and neglecting terms quadratic in $y_{l,m,n}$, equation (4.25) is linearised in the unknown $y_{l,m,n}$. A standard numerical technique for solving linear equations can now be invoked to solve the entire set of equations which are linear in the $y_{l,m,n}$.

The entire set of linear equations in the $y_{l,m,n}$ include the above linearised equations from (4.25) as well as ray path symmetry constraints, (4.26), and those comprising the measurement data and boundary conditions. The linear equations arising out of the measurement data are formed by substituting (4.27) and (4.23) into (4.19). The boundary conditions, equation (4.20), have been used to solve for the A_l on the boundary where the receiver and emitter points coincide ($Q'_m = Q_l$), equations (4.21) and (4.22). To ensure the boundary conditions are satisfied when the receiver and emitter points do not coincide (where $Q'_m = Q_l$, but $l \neq l'$ in equation (4.20)), included is a set of linear equations formed by substituting (4.27) and (4.24) into (4.20) and neglecting terms quadratic in the $y_{l,m,n}$.

After solving for $y_{l,m,n}$, each $X_{l,m,n}$ is updated by subtracting $y_{l,m,n}$ from the present value of $X_{l,m,n}$ (as per equation (4.27)). These new $X_{l,m,n}$ can be substituted for the $B_{l,m,n}$ in (4.24) to yield a new estimate of $\nu^2(P)$. This completes an iteration, which can be repeated until the desired accuracy has been obtained. The development of the linearised equations mentioned is not included here for the general 2-D refractive index distribution. However, section 4.6 does so for the case of a circularly symmetric distribution.

4.6 ITERATIVE APPROACH FOR CIRCULARLY SYMMETRIC MEDIA

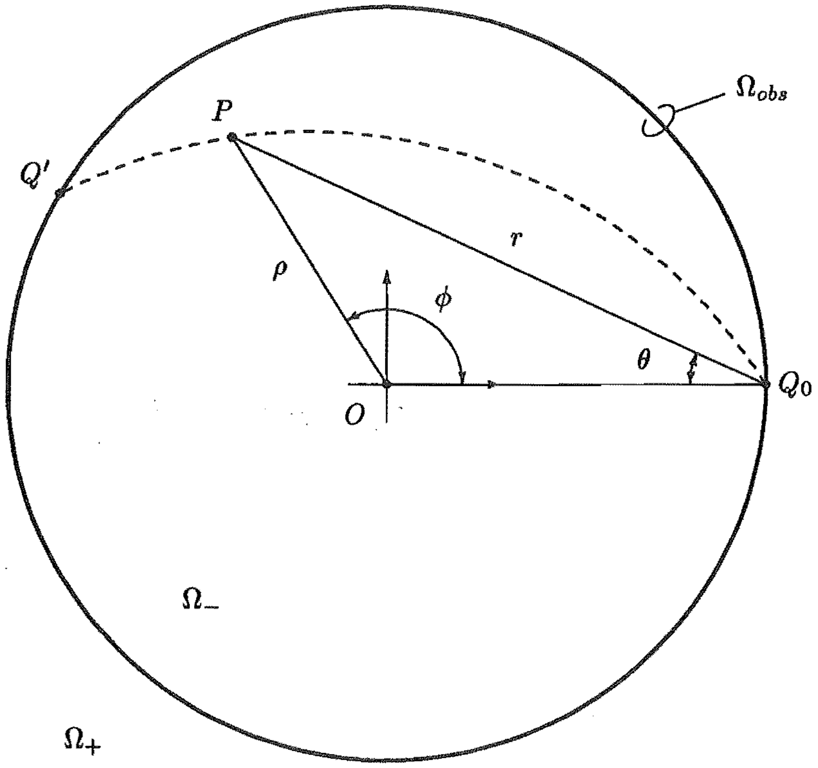


Figure 4.3 Coordinates for bent-ray transmission CT of a circularly symmetric refractive index distribution. The dashed line represents the actual ray path from an emitter Q to a receiver at Q' through some arbitrary point P on its path. Regions Ω_{inc} consists of the point Q_0 and Ω_{obs} lies on the boundary of radius a.

As mentioned in section 3.5, when the refractive index distribution is circularly symmetric, $\nu = \nu(\rho)$, the solution of the bent-ray tomographic problem is known, and in seismology is referred to as the Herglotz-Wiechert formula [Aki and Richards, 1980]. However, the circularly symmetric situation allows the feasibility of the approach described in this chapter to be determined before going to the trouble of developing algorithms for two and three dimensions.

Because of the symmetry only one emitter point, Q_0 say (see figure 4.3), need be considered and (4.14) can be replaced by

$$H(P,Q_0) = \nu_b\, r + \sum_{m=0}^{M} \sum_{n=2}^{N} B_{m,n}\, r^n \cos(2m\theta), \tag{4.28}$$

where, in accordance with (4.23), finite limits have been taken for the summations and where

ν_b is the known refractive index at the boundary Ω_{obs} (see figure 4.3). The data are the H_m measured on the boundary, defined by

$$H_m = H(Q'_m, Q_0) \quad \text{for} \quad 0 \leq m \leq L, \tag{4.29}$$

where the angular coordinates of Q'_m, Φ'_m, need only be defined within the range $0 \leq \Phi'_m \leq \pi$ because of the symmetry. Here the positions of receivers are taken as $\Phi'_m = (2m+1)\pi/(2L+1)$. The known refractive index ν_b on the boundary is satisfied by (4.28) at the receiver points if

$$\nu_b^2 = \nabla_P H(Q'_m, Q_0) \cdot \nabla_P H(Q'_m, Q_0) \quad \text{for} \quad 0 \leq m \leq L. \tag{4.30}$$

The circularly symmetric version of (4.24) is

$$\begin{aligned} \nabla_P H(P,Q_0) \cdot \nabla_P H(P,Q_0) &= \nu_b^2 + 2\,\nu_b \sum_{m=0}^{M} \sum_{n=2}^{N} B_{m,n}\, n\, r^{n-1} \cos(2m\theta) \\ &+ \sum_{m,m'=0}^{M} \sum_{n,n'=2}^{N} B_{m,n}\, B_{m',n'}\, r^{n+n'-2} \\ &\qquad (\, nn' \cos(2m\theta)\cos(2m'\theta) + 4mm' \sin(2m\theta)\sin(2m'\theta)\,)\,. \end{aligned} \tag{4.31}$$

Since $\nu = \nu(\rho)$ is independent of ϕ it can be said that for all $0 \leq \rho \leq a$, the Fourier decomposition of $\nabla_P H(P,Q_0) \cdot \nabla_P H(P,Q_0)$ into orthogonal components $\cos(2\pi\mu\phi)$ is zero for $0 < \mu \leq L$. In accord with (4.25), the Fourier integral over ϕ is approximated by a Riemann sum at a discrete number of equi-spaced angles, ϕ_m :

$$\sum_{m=0}^{L} [\, \nabla_P H(P(\rho;\phi_m),Q_0) \cdot \nabla_P H(P(\rho;\phi_m),Q_0)\,] \; \cos(2\pi\mu\phi_m) \;=\; 0 \quad \text{for} \quad 0 < \mu \leq L, \tag{4.32}$$

where the position of the point P is a function of ρ and ϕ_m as can be seen from figure 4.3.

Equation (4.32) can be solved at a sufficient number of discrete values for ρ, say ρ_j where $0 \leq j \leq J$, to yield a set of linear equations that is not underdetermined. Since there are $(M+1)(N-1)$ unknown coefficients, $(L+1)$ equations arising out of the measurement data, $(L+1)$ boundary equations and $(J+1)L$ equations arising from (4.32), the system of linear equations will not be underdetermined if

$$J \;\geq\; [\,(M+1)(N-1) - 3L - 2\,]\,/\,L, \tag{4.33}$$

and, the algorithm outlined in section 4.5 applied to solve for the $B_{m,n}$. As mentioned in section 4.4 the resolution of the reconstructed refractive index and the ray path length $H(P,Q_0)$ does not depend on L, M, or N in a simple fashion. This can be seen from the expression for the refractive index distribution (4.31). Although circularly symmetric about the origin O, it is expressed in terms involving the ray coordinates from the emitter position Q_0.

Aki and Richards [1980] have given conditions for a circularly symmetric velocity distribution whereby there exists a *shadow zone* on the boundary that no rays reach and where no measurement data, (4.29), is available. This is caused by the existence of a *low-velocity zone* (as it is known in seismology) in ν. The condition for a low-velocity zone in a circularly symmetric refractive index distribution is given in (3.7). If a low-velocity zone exists then the problem has a non-unique solution for ν in the low-velocity zone. This is distinct from the partially non-unique solution mentioned in section 4.1 which occurs when rays do not enter a region of ν.

4.7 COMPUTATIONAL RESULTS FOR CIRCULARLY SYMMETRIC REFRACTIVE INDEX DISTRIBUTIONS

The iterative approach proposed in Section 4.6 has been used to reconstruct circularly symmetric refractive index distributions from simulated time-of-flight measurements. This algorithm has been tested on two types of refractive index distributions. The results of these tests are presented in this section. The two distribution types are shown in figure 4.4 and it is found convenient to refer to the uni-modal distribution as type **A** and the bi-modal distribution as type **B**. The distributions have been plotted across their entire diameter to depict the circular symmetry. It should be noted that the distributions are normalised to unity on their boundaries and their amplitude variations are as yet unspecified. In figures 4.5 through 4.9, the reconstruction results are plotted over the distribution's radius in overlay fashion with the imaged distribution.

In order to ensure the distributions are sufficiently smooth and avoid the existence of a low-velocity zone, equation (3.7), the amplitude variation in ν for the distributions type **A** and **B** must lie within the ranges shown in table 4.1. The maximum deviation from the value of unity is represented as ν_{max}.

Distribution	Amplitude variation in ν
type **A**	$\nu_{max} \leq 1.45$
type **B**	$-0.59 \leq \nu_{max} \leq 0.49$

Table 4.1 Amplitude variation in ν such that a low-velocity zone does not exist.

Note also that because of the form of (4.28) and (4.31), which is due to the representation for H and ν respectively being based on ray parameters r and θ, although the distributions types **A** and **B** are simple functions of ρ the ray paths H, which are subject to solution here, are not such simple functions. Because of this, the distributions types **A** and **B** are in fact representative of a more general class of refractive index distributions and are a more severe test of the algorithm than may be initially apparent.

Time-of-flight measurement data, $H(Q'_m, Q_0)$, for each refractive index distribution, have been obtained from numerical simulations using a ray tracing algorithm (refer section 3.5). It is stressed here again that an initial estimate is not required for the $X_{l,m,n}$ in (4.27); they are assumed initially to be zero. The n^{th} resultant iterative reconstruction for $\nu(\rho)$ is denoted by I_n. For these simulations the solution points are taken at discrete radii ρ_j, that is, equi-spaced within the range $0 \leq \rho_j \leq a$.

To demonstrate the algorithm in reconstructing distributions of type A, with small and large amplitude variations, distributions are chosen with 5% variation in refractive index (figure 4.5), and, with a 50% variation (Figures 4.6 and 4.7). As the results indicate, already after one iteration the reconstruction, I_1, is close to the imaged distribution. After 5 iterations computations have converged to the imaged distribution (within a small percentage of the total spatial variation). Reconstructions of the type B distribution are shown in Figures 4.8 and 4.9 with 5% and 20% amplitude variations respectively. Again accurate results are obtained. In figure 4.9 the convergence properties of this algorithm are demonstrated by reconstructing a type B distribution with significant amplitude variation. The reconstruction from a number of consecutive iterations are displayed. The initial estimate I_0 is that of the homogeneous distribution $\nu = \nu_b$ (since initially the $X_{l,m,n}$ are zero), and the iterations converge towards the imaged distribution.

4.8 DISCUSSION

A general and novel bent-ray CT algorithm has been described and applied successfully to give quantitative reconstructions of circularly symmetric refractive index distributions. The reconstruction method is based on a novel ray description which doesn't require the ray paths to be known. This allows the refractive index to be found by iterative solution of linear equations, rather than through the computationally intensive procedure of ray tracing, which normally accompanies iterative solutions to problems of this type. The more common problem of bent-ray transmission CT has been considered and a more detailed algorithm suggested for this problem. An approach for reconstructing circularly symmetric refractive index distributions has been given along with results obtained by applying the algorithm for bent-ray transmission CT to reconstruct such distributions. Preliminary results have shown that this method is capable of handling appreciable spatial refractive index variations in large bodies. It is also able to handle appreciably larger variations in refractive index than the Born and Rytov approximations, while avoiding the computational consequences of an exact solution. It is worth noting that the accuracy of this algorithm is independent of the size of the region imaged, another limitation of methods based on the Born and Rytov approximations. These factors make it well suited to solving practical inverse problems.

As a result of investigations undertaken a number of interesting points and aspects have transpired that are important for an effective approach to the general problem. These are summarised in table 4.2 in the form of additional considerations and are discussed in more detail later in section 8.1 as appropriate areas for further investigation.

Representation for $H(P,Q)$	Appropriate representations for asymmetric distributions.
	Representations that automatically satisfy the reciprocity of ray paths within the distribution.
	Basis functions that allow orthogonality to be used in solving for the coefficients of each.
	Inclusion of terms in the representation that can be related directly to the measurement data.
Solving for the unknown coefficients of the basis functions	Making an initial estimate, or, use of any *a priori* information to make an initial estimate of the $X_{l,m,n}$.
	Appropriate linearisation algorithms or use of non-linear equation solving techniques.
Additional Constraints	Direct application of Fermat's principle (refer section 3.4) to constrain the phase function $H(P,Q)$.
Resolution of the Reconstruction	Estimation of the expansion limits L, M and N required for a desired resolution.

Table 4.2 Additional considerations for further research into a bent-ray CT algorithm for reconstructing general refractive index distributions.

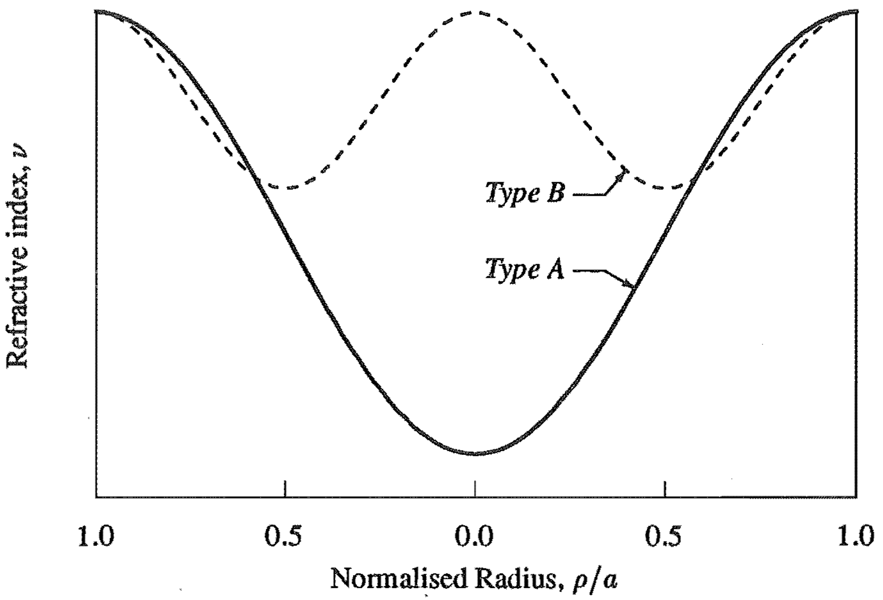


Figure 4.4 Two distributions to be reconstructed, denoted type A and type B, plotted across their diameter to show circular symmetry.

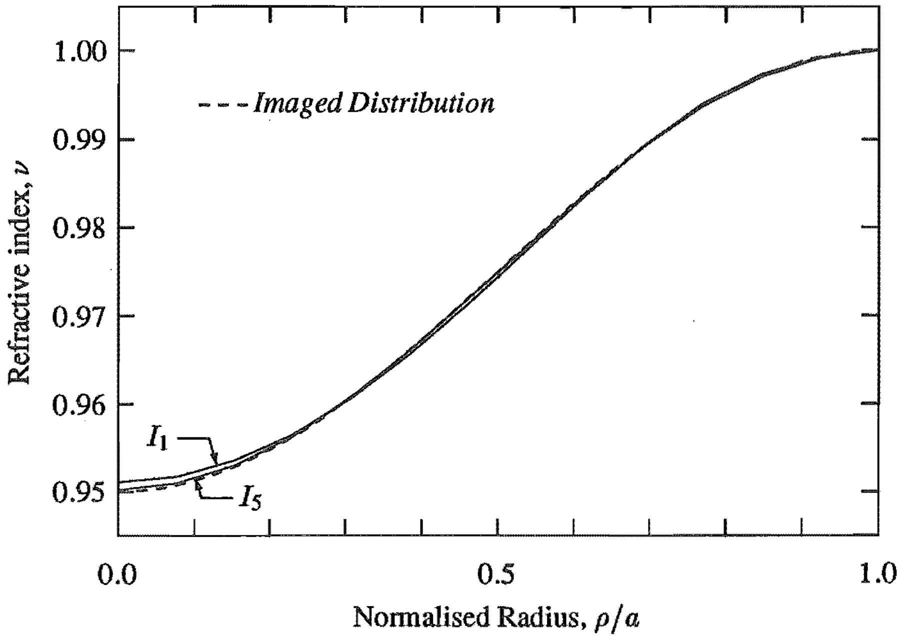


Figure 4.5 Reconstruction of a type A distribution with range $0.95 \leq \nu \leq 1.00$ and representation $L = 9$, $M = N = 11$, $J = 12$.

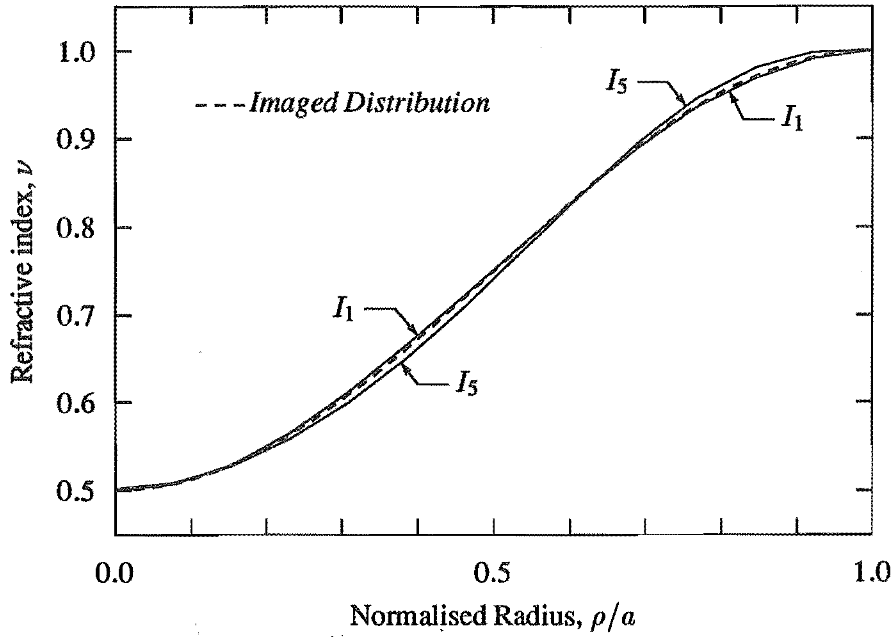


Figure 4.6 Reconstruction of a type A distribution with range $0.5 \leq \nu \leq 1.0$ and representation $L = 9$, $M = N = 11$, $J = 12$.

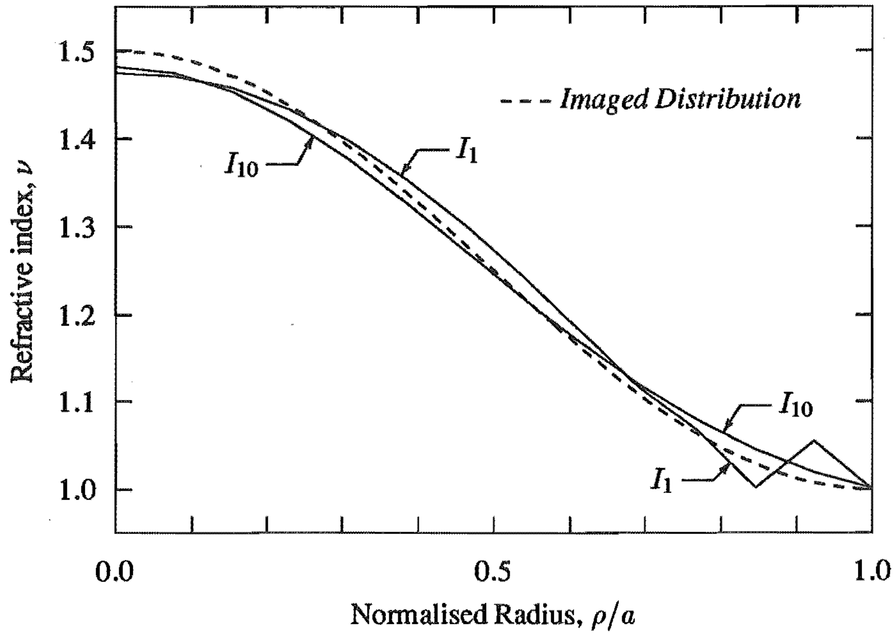


Figure 4.7 Reconstruction of a type A distribution with range $1.0 \leq \nu \leq 1.5$ and representation $L = 9$, $M = N = 11$, $J = 12$.

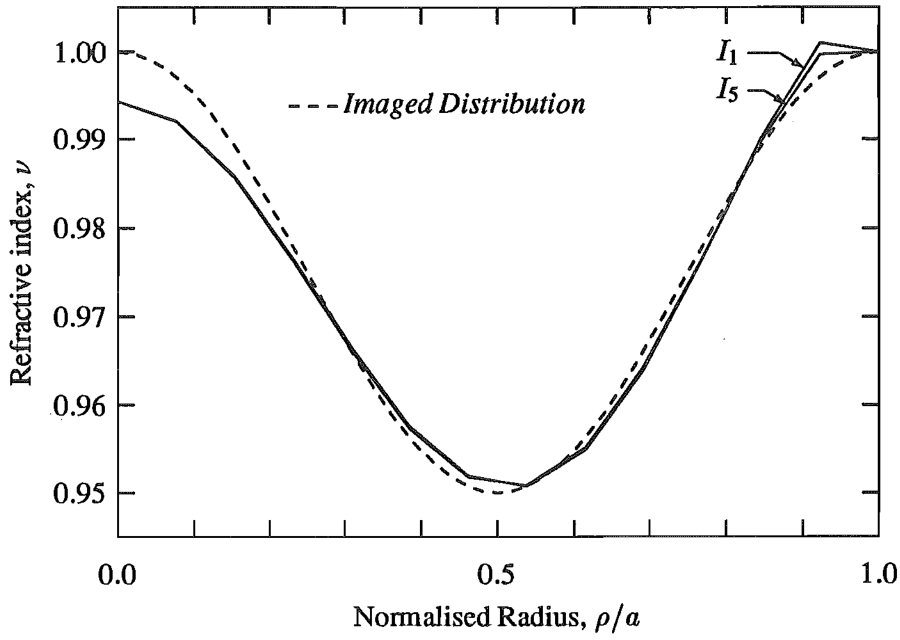


Figure 4.8 Reconstruction of a type B distribution with range $0.95 \leq \nu \leq 1.00$ and representation $L = 14$, $M = N = 14$, $J = 12$.

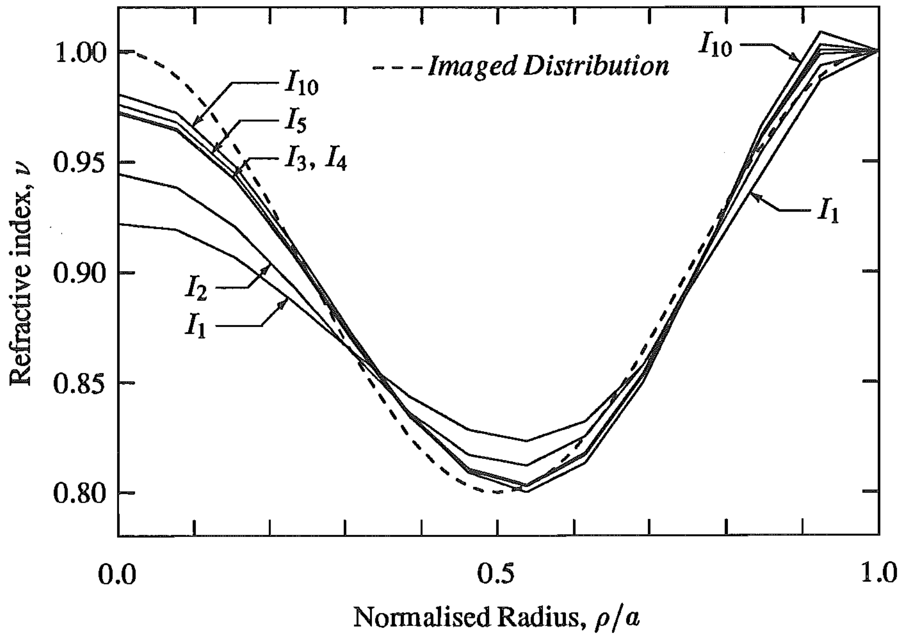


Figure 4.9 Reconstruction of a type B distribution with range $0.8 \leq \nu \leq 1.00$ and representation $L = 14$, $M = N = 14$, $J = 12$.

CHAPTER 5

COMPUTED TOMOGRAPHY WITH THE STRAIGHT-RAY APPROXIMATION

Computed tomography with the straight-ray approximation, sometimes phrased *image reconstruction from projections*, addresses the problem of producing an image of a two-dimensional distribution, $f(r;\theta)$, from integrals of $f(r;\theta)$ along lines in the $(r;\theta)$ plane. The transformation of $f(r;\theta)$ to line integrals is known here as the *Radon transform* and is detailed in section 5.1. The main thrust of this chapter is to describe the Radon transform, the predominant inversion formulas for the Radon transform, and common techniques applied to produce an estimate $\hat{f}(r;\theta)$ of the 2-D scattering distribution contained within the support Ω_- of $f(r;\theta)$ (refer figure 1.1).

5.1 RADON TRANSFORM

The Radon transform, sometimes referred to in the literature as the X-ray transform [Lewitt, 1977], was first introduced in a paper by Radon (1917), who outlined an inversion formula that would allow the reconstruction of a distribution f from its integrals along all possible lines through f (see section 5.6.1). Since then, this transform and its inversion formulas have been utilised throughout many scientific fields, none more so than medical X-ray computed tomography (CT). As intimated in section 1.4.1, in medical X-ray CT the basic data measured are the line integrals of the X-ray attenuation coefficient measured along straight ray paths through the body under examination. Complete inversion of the data reveals an image reconstruction or map of the linear attenuation coefficient for the cross-section examined. Since the linear X-ray attenuation coefficient is very closely related to the density of the material [Bates and Peters, 1971], the reconstruction is often interpreted as a density distribution.

The Radon transform of a distribution given in polar coordinates, $f(r;\theta)$, along a line characterised by the coordinates $(\xi;\phi)$ (see figure 5.1) is given by

$$p(\xi;\phi) = \int_{-\infty}^{\infty} f(r;\theta)\, d\eta, \tag{5.1}$$

where the infinite limits imply that the line integral traverses the entire domain of $f(r;\theta)$, and the coordinate system (ξ,η) rotates around the origin with the projection angle ϕ, from the fixed x-axis. The Radon transform of the function $f(r;\theta)$ is denoted as $\mathcal{R}\{\, f(r;\theta)\,\}(\xi;\phi)$.

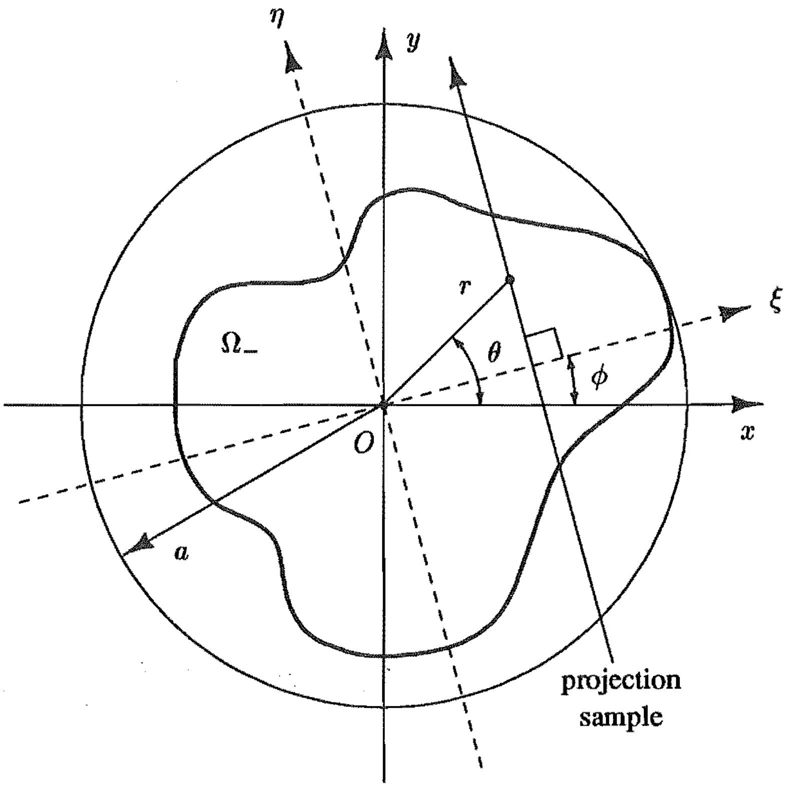


Figure 5.1 Coordinates $(\xi;\phi)$ for the Radon transform of a 2-D region Ω_-. A projection sample is specified by its perpendicular distance ξ from the origin and its orientation ϕ.

Equation (5.1) can be rewritten as

$$p(\xi;\phi) \;=\; \mathcal{R}\{\, f(r;\theta)\,\}(\xi;\phi). \tag{5.2}$$

It is convenient at this point to introduce the term *projection,* referring to the set of Radon transforms associated with any one projection angle ϕ. In order for the projections to be single-valued, the Radon transform must be symmetrical with respect to the direction of the line integral in (5.1). This symmetry in the projections can be written as

$$p((-1)^n\xi;\phi+n\pi) \;=\; p(\xi;\phi). \tag{5.3}$$

The total mass associated with the density distribution $f(r;\theta)$ is

$$\begin{aligned} \tau &= \int_0^{2\pi}\int_0^{\infty} f(r;\theta)\, r\, dr\, d\theta \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(r;\theta)\, d\eta\, d\xi, \end{aligned} \tag{5.4}$$

where the infinite limits imply that τ is the total mass of $f(r;\theta)$. Note, from (5.1) that

$$\int_{-\infty}^{\infty} p(\xi;\phi)\,d\xi = \tau. \tag{5.5}$$

As $f(r;\theta)$ is known to exist entirely within the radius a (refer figure 5.1), (5.5) becomes

$$\int_{-a}^{a} p(\xi;\phi)\,d\xi = \tau. \tag{5.6}$$

Transformation between the polar coordinates $(r;\theta)$ and that of the rotating coordinate system $(\xi;\eta)$ at an angle ϕ can be represented by the following identities derived from figure 5.1 :

$$\begin{aligned} \xi &= r\cos(\phi-\theta), & \eta &= -r\sin(\phi-\theta), \\ r &= (\eta^2+\xi^2)^{1/2}, & \theta &= \phi+\tan^{-1}(\eta/\xi), \end{aligned} \tag{5.7}$$

thereby allowing the Radon transform (5.1) to be rewritten in terms of the coordinates characterising the transform $(\xi;\phi)$,

$$p(\xi;\phi) = \int_{-\infty}^{\infty} f(\,(\eta^2+\xi^2)^{1/2};\phi+\tan^{-1}(\eta/\xi)\,)\,d\eta. \tag{5.8}$$

Similarly, the transformation for a function of Cartesian coordinates (x,y), to the coordinates of the Radon transform $(\xi;\eta)$ at an angle ϕ is

$$\begin{aligned} \xi &= x\cos(\phi)+y\sin(\phi), & \eta &= y\cos(\phi)-x\sin(\phi), \\ x &= \xi\cos(\phi)-\eta\sin(\phi), & y &= \xi\sin(\phi)+\eta\cos(\phi), \end{aligned} \tag{5.9}$$

and equivalently the Radon transform of $f(x,y)$ is from (5.8)

$$p(\xi;\phi) = \int_{-\infty}^{\infty} f(\,\xi\cos(\phi)-\eta\sin(\phi),\ \xi\sin(\phi)+\eta\cos(\phi)\,)\,d\eta. \tag{5.10}$$

As the distribution under examination is of finite extent, the limits of integration in (5.1) need only be taken over its region of support, Ω_-, giving

$$p(\xi;\phi) = \int_{\Omega_-} f(r;\theta)\,d\eta. \tag{5.11}$$

However, it is often the case in practice that the precise support of the distribution Ω_- is not known, but is known to lie within a circle of radius a, say, so that the Radon transform in (5.1) is taken to be

$$p(\xi;\phi) = \int_{-(a^2-\xi^2)^{1/2}}^{(a^2-\xi^2)^{1/2}} f(r;\theta)\, d\eta. \tag{5.12}$$

As discussed in section 5.5, it is impossible to measure single line integrals. Because of the finite size of sensor apertures, the measurements recorded of the Radon transforms are in practice strip integrals of the distribution, f, taken over the sensor's aperture whose response may be modelled by convolution with the sensor's instrument function (refer equation (5.35)).

5.2 PROJECTION THEOREM

In this section an important relationship is given between the 1-D Radon transforms of a distribution and its 2-D Fourier transform. Fourier methods for the inversion of the Radon transform are based on this relationship and are outlined in section 5.6.3.1. In this section a derivation is given for the relationship between the Radon transforms $p(\xi;\phi)$ of the distribution f taken at an orientation angle ϕ, the 1-D Fourièr transform of $p(\xi;\phi)$, and the Fourier transform of the distribution, $F(\rho;\phi)$.

Firstly, recall from section 2.3.1, equation (2.43), the 2-D Fourier transform of the 2-D distribution $f(r;\theta)$ in polar coordinates is

$$F(\rho;\phi) = \int_0^{2\pi} \int_0^{\infty} f(r;\theta)\, e^{-i2\pi\rho r\cos(\phi-\theta)}\, r\, dr\, d\theta. \tag{5.13}$$

Next consider the 1-D Fourier transform relationship for a single projection $p(\xi;\phi)$ at a particular orientation angle ϕ, whose Fourier transform is denoted $P_\phi(\rho)$,

$$P_\phi(\rho) = \int_{-\infty}^{\infty} p(\xi;\phi)\, e^{-i2\pi\rho\xi}\, d\xi, \tag{5.14}$$

$$p(\xi;\phi) = \int_{-\infty}^{\infty} P_\phi(\rho)\, e^{i2\pi\rho\xi}\, d\rho. \tag{5.15}$$

Recalling the Radon transform equation (5.1) and substituting this into (5.14) yields the Fourier transform of a single projection in terms of the image $f(r;\theta)$ and rotated projection coordinates $(\xi;\eta)$:

$$P_\phi(\rho) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(r;\theta)\, e^{-i2\pi\rho\xi}\, d\eta\, d\xi. \tag{5.16}$$

To make the transformation back to polar coordinates, the identities (5.7) are substituted into (5.16) giving $P_\phi(\rho)$ in terms of the coordinates $(r;\theta)$:

$$P_\phi(\rho) = \int_{-\pi/2}^{\pi/2} \int_{-\infty}^{\infty} f(r;\theta)\, e^{-i2\pi\rho r\cos(\phi-\theta)}\, r\, dr\, d\theta$$

$$= \int_0^{2\pi} \int_0^{\infty} f(r;\theta)\, e^{-i2\pi\rho r\cos(\phi-\theta)}\, r\, dr\, d\theta. \tag{5.17}$$

It can be seen from equations (5.13) and (5.17) that the 1-D Fourier transform of a projection at angle ϕ, $P_\phi(\rho)$, is equal to the values along a line passing through the origin (often referred to as a *central slice*) of the Fourier transform $F(\rho;\phi)$ at angle ϕ. This is conveniently expressed by

$$P_\phi(\rho) = F(\rho;\phi) \quad \text{for any} \quad 0 \leq \phi < \pi. \tag{5.18}$$

Equivalently, if the 2-D Fourier transform is represented in Cartesian coordinates (u, v) as $F(u, v)$, then

$$P_\phi(\rho) = F(\,\rho\cos(\phi),\ \rho\sin(\phi)\,) \quad \text{for any} \quad 0 \leq \phi < \pi. \tag{5.19}$$

Equation (5.18) or (5.19) is known as the *projection theorem* relating the 1-D Fourier transform of projections $P_\phi(\rho)$ to the 2-D Fourier transform of the distribution $F(\rho;\phi)$ or $F(u, v)$.

5.3 GEOMETRY OF DATA SAMPLING

In practice the measured Radon transform, of a distribution $f(r;\theta)$, is not a continuous one in either of the coordinates $(\xi;\phi)$. The transform can be measured for only a finite number of orientation angles ϕ_m, known as projection angles, or *views*. For each view a finite number of projection samples of varying ξ are taken, $p(\xi_n;\phi_m)$. The projection data available for reconstruction of the distribution $f(r;\theta)$ is then

$$p(\xi_n;\phi_m) = \int_{-\infty}^{\infty} f(\,(\xi_n^2+\eta^2)^{1/2};\ \phi_m + \tan^{-1}(\eta/\xi_n)\,)\, d\eta, \tag{5.20}$$

where the ξ_n and ϕ_m are usually equi-spaced. The mass τ of $f(r;\theta)$, equation (5.6), can then be approximated by the *Riemann sum*

$$\begin{aligned} \tau &= \int_{-a}^{a} p(\xi;\phi)\, d\xi \\ &\approx \sum_{n=-N}^{N} p(\xi_n;\phi_m)\, \Delta\xi. \end{aligned} \tag{5.21}$$

The sampling geometry for the Radon transform adopted throughout this thesis is that of the *parallel ray* geometry. This is illustrated in figure 5.3 where the projection is sampled at equi-spaced (spacing $\Delta\xi$) intervals, corresponding to the spacing of detectors, along the ξ-axis which is perpendicular to the rays. The ξ-axis is rotated by angle ϕ_m with respect to the x-axis. Most

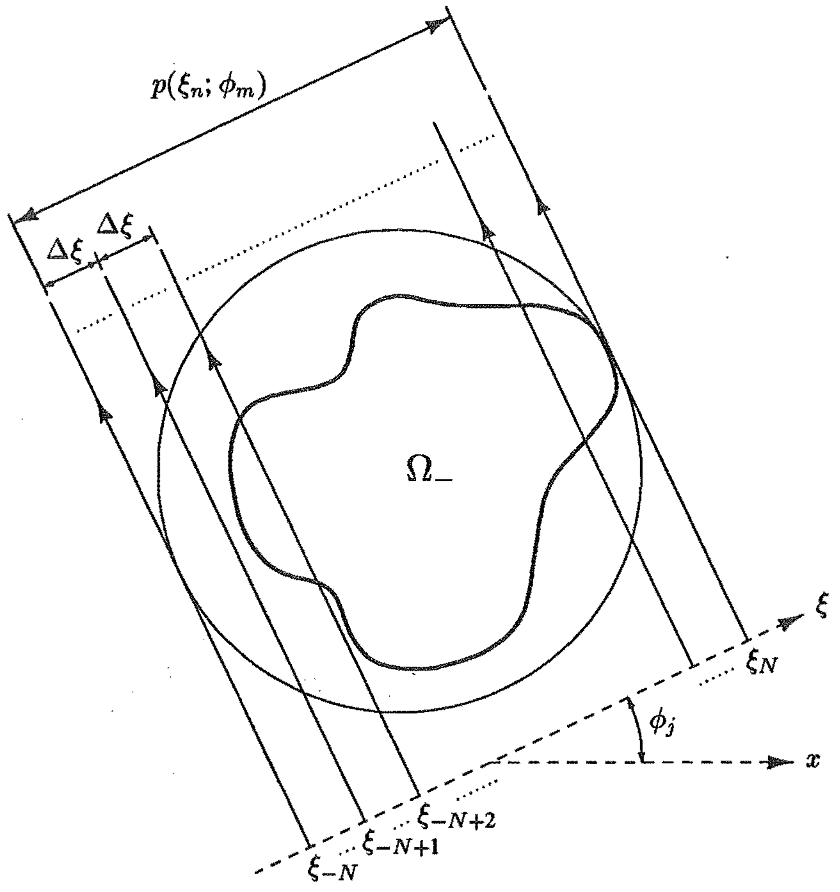


Figure 5.2 Parallel projection geometry. Projection $p(\xi_n; \phi_m)$ consists of projection samples taken at an orientation angle ϕ_m, equi-spaced in ξ by $\Delta\xi$.

of the original reconstruction algorithms [Bates and Peters, 1971; Bracewell and Riddle, 1967; Cormack, 1963; DeRosier and Klug, 1968; Gordon *et al.*, 1970] were derived based on this sampling geometry for the Radon transform. First and second generation X-ray CT scanners acquired line integrals of X-ray attenuation through a body and did so at equi-spaced intervals in both coordinates ξ and ϕ. However, early in their development scanners utilised a fan-beam geometry that makes more efficient use of the X-ray source, and this has become standard. Figure 5.3 shows a typical fan-beam geometry. With the fan-beam geometry a single source traverses the body under investigation (usually on a circular path surrounding the body) and illuminate it over a region that is the shape of a fan (see figure 5.3). Detectors placed opposite to the source, sample the line integrals at regular intervals in θ. The source then rotates to different positions on the surrounding path (usually at regularly spaced intervals in Φ), at each angle yielding new projection sets of line integrals.

Scanners with the fan-beam geometry were adopted quickly as this physical geometry facilitated faster scan times, simplified the construction by reducing the number of moving parts and reduced costs. Initially the reconstruction algorithms used to reconstruct the X-ray attenuation samples from fan-beam scanners were those developed for parallel geometry. Line integrals sampled regularly in $(\theta; \Phi)$ with the fan-beam geometry can be simply "rebinned"

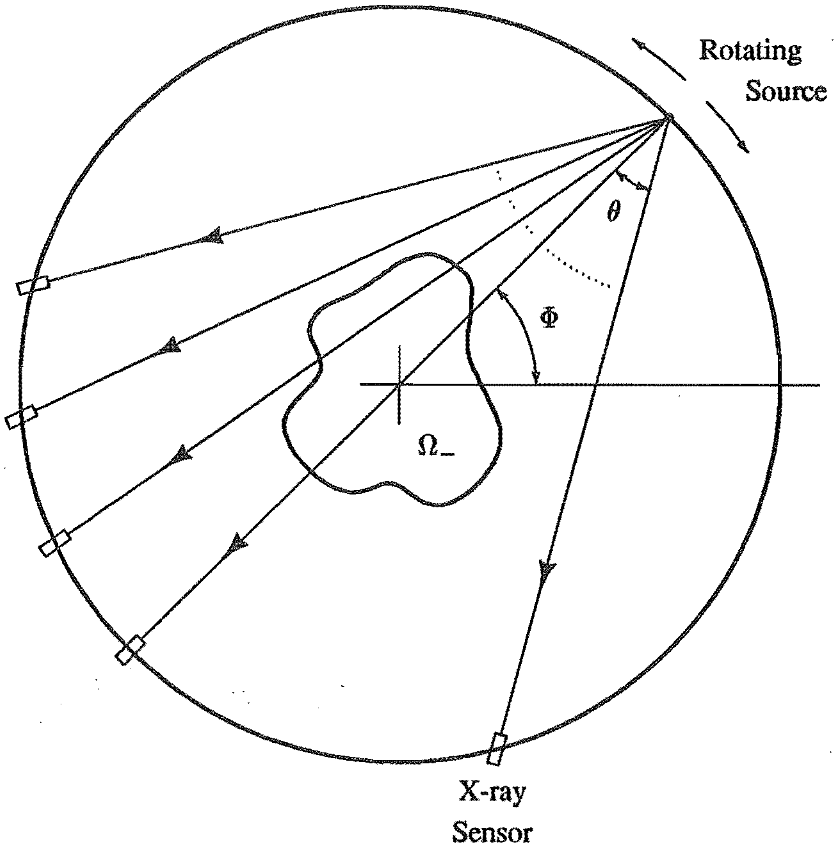


Figure 5.3 Fan-beam projection geometry with a single rotating source.

to regular samples in $(\xi; \phi)$, and, so long as sufficient samples are given to cover the required range of $(\xi; \phi)$, the parallel geometry algorithms can be applied to this data. The process of rebinning requires interpolation from regular samples in one coordinate system to the other and is described well by Peters and Lewitt [1977] and Herman [1980]. However, this process has become redundant as parallel geometry reconstruction methods were adapted [Herman *et al.*, 1976; Lewitt, 1983; Scudder, 1978] to the fan-beam geometry. More recently, reconstruction algorithms have been developed for direct inversion of data taken from a scanner implementing a fan-beam geometry [Horn, 1979].

5.4 BACK-PROJECTION

Back-projection is a simple method of reconstruction, which, before methods based on an inversion of the Radon transform, was used to produce a blurred image of f called a *tomogram*. The image is blurred because of an over-representation of low spatial frequencies in the projection data. However, modern methods use *back-projection* (BP) as a step in the reconstruction process but subject the data to a filtration of the lower frequencies first (see section 5.6.3.3). Recall that the Radon transform p along a line is the integral of f along all points on that line. Back-projection is somewhat the reverse of this process but is not a direct inversion of the Radon transform. In the back-projection process, the reconstruction value at a point $(r;\theta)$ is the integral of all Radon transforms whose line integral paths pass through this point. Back-projection has often been referred to in the literature as a *layergram*, reflecting the layering or superimposing of projections in image space to form an image, and, in this context the projections are known as *shadowgrams*. The BP of a function $p(\xi;\phi)$ is denoted here as $\mathcal{B}\{\,p(\xi;\phi)\,\}\,(r;\theta)$ and the image produced from the BP of simple projection data is denoted $b(r;\theta)$ to distinguish it from a true reconstruction of the image $f(r;\theta)$ since BP is not an inversion of the Radon transform.

If $p(\xi;\phi)$ are the line integral paths characterised by the coordinates $(\xi;\phi)$ as in section 5.1, then, the BP of $p(\xi;\phi)$ is defined here as

$$\begin{aligned} \mathcal{B}\{\,p(\xi;\phi)\,\}\,(r;\theta) &= \int_0^{\pi} p(\xi;\phi)\,d\phi \\ &= \int_0^{\pi} p(r\cos(\phi-\theta);\phi)\,d\phi. \end{aligned} \tag{5.22}$$

As mentioned in section 5.3 the data which are dealt with are a discrete number of Radon transform samples. Assuming there are M views corresponding to discrete values of ϕ, equi-spaced in angle by $\Delta\phi = \pi/M$ over the range $0 \le \phi < \pi$, then the integral in (5.22) may be approximated by the *Riemann sum* [Herman, 1980]

$$\mathcal{B}\{\,p(\xi;\phi_m)\,\}\,(r;\theta) = \Delta\phi \sum_{m=0}^{M-1} p(r\cos(\phi_m-\theta);\phi_m). \tag{5.23}$$

As there is a discrete number of projection samples for any view angle, the point $(r;\theta)$ may have no projection paths intersecting it. Interpolation can be used to estimate the projection passing through this point from those at discrete values ξ_n (refer to figure 5.3). Through the identities given in (5.7), the coordinate of the projection sample $\xi = r\cos(\phi_m-\theta)$ can be calculated for each projection angle ϕ_m. If the interpolation of $p(\xi_n;\phi_m)$ at $\xi = r\cos(\phi_m-\theta)$ for projection ϕ_m is denoted by $\mathcal{I}\{\,p(\xi_n;\phi_m)\,\}\,(\xi)$, i.e.,

$$p(\xi;\phi_m) \approx \sum_{n=-N}^{N} \mathcal{I}\{\,p(\xi_n;\phi_m)\,\}\,(\xi), \tag{5.24}$$

where the integer N is in general infinite, then the discrete BP can be rewritten in terms of the interpolation process;

$$\mathcal{B}\{\,p(\xi_n;\phi_m)\,\}\,(r;\theta) = \Delta\phi \sum_{m=0}^{M-1} \sum_{n=-N}^{N} \mathcal{I}\{\,p(\xi_n;\phi_m)\,\}\,(\xi), \tag{5.25}$$

where ξ is found from the image coordinates $(r;\theta)$. Linear interpolation has been extensively used for its numerical efficiency and is described in section 5.5 in the context of discrete back-projection. A number of interpolation methods and their effects on the quality of the reconstruction have been examined [Heffernan and Bates, 1982; Herman, 1980; Lewitt *et al.*, 1978; Rowland, 1979].

5.5 DISCRETE BACK-PROJECTION AND FORWARD-PROJECTION

Section 5.4 described the process of back-projection of $p(\xi;\phi)$ onto an image at the point $(r;\theta)$ as an analogy to reversing the Radon transformation. Conversely, *forward projection* (FP) is analogous with the Radon transform, but is distinguished by name here from the Radon transform because of the context from which it is usually derived. The Radon transform is assumed to describe measured data, as distinct from the FP in which projections are calculated from an estimate of the image $\hat{f}(r;\theta)$ that is derived initially from this measured data. In terms of the Radon transform given in section 5.1, the forward projection of the region Ω_- is (compare (5.11))

$$p(\xi;\phi) = \int_{\Omega_-} \hat{f}(r;\theta)\, d\eta. \tag{5.26}$$

Many reconstruction techniques have employed FP to ensure that the raw data is consistent with that calculated from the reconstruction. Iterative techniques have also employed FP to correct the measured data or replace missing data based on new estimates of the image. Examples of such methods as these are given in section 5.6.6. It should be noted here that special attention is being given here to FP because of its importance in chapter 6.

As pointed out in section 2.3.2, an image can only be represented at a finite set of points. Two-dimensional images are most often composed of a set of square pixels (see figure 2.3), which are assumed to be uniform over the domain of each pixel. This section deals with some of the techniques used to estimate BP and FP for this type of image representation. The techniques developed here are simply extended to a three-dimensional image whose representation is composed of voxels for its basis functions. Recall from section 5.4 that BP at a point is achieved by superimposing the contributions of each projection sample passing through that point. The *square pixel* method of BP integrates the projection sample over the path that it takes through the pixel. Note that in accord with the notation established in the previous section for the image obtained from back-projection of simple projection data, the image at the point $(r;\theta)$ is denoted $b(r;\theta)$. Because of the discrete nature of this image and the projection data the j^{th} pixel is denoted b_j and the i^{th} projection sample p_i corresponding to $p(\xi_n;\phi_m)$. Therefore, the BP contribution of p_i to b_j using the square pixel method,

$$\begin{aligned} \Delta b_{i,j} &= \int_{\eta_{i,j}} p(\xi_n;\phi_m)\, d\eta \\ &= \eta_{i,j}\, p_i \end{aligned} \tag{5.27}$$

since the pixel b_j is by definition uniform over its domain and $\eta_{i,j}$ is the path length through the pixel (refer figure 5.5). This method requires the projection path through image space to be calculated in order to determine the path lengths $\eta_{i,j}$. This ray path calculation, denoted as ray-tracing in section 3.5, is considerably simpler here than for the general case because the ray paths are known to be straight.

For the reverse process, forward projection, an estimate of the Radon transform is required from a given image representation f. Consider the 2-D image distribution represented as

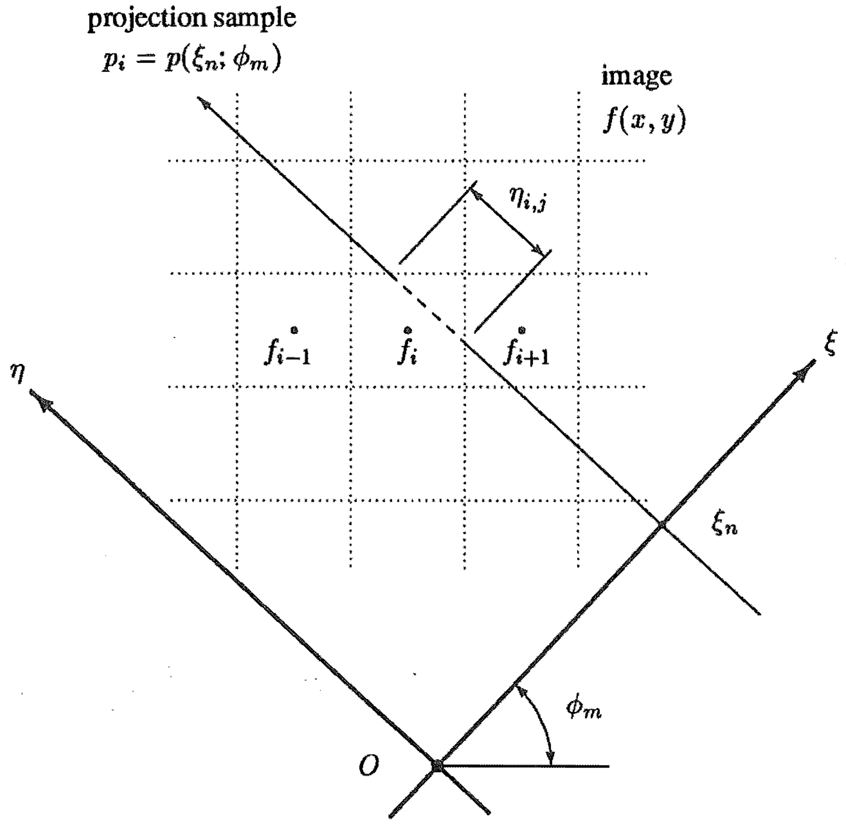


Figure 5.4 Projection through a discrete image represented as square pixels. The image is constant over each pixel and has the value f_j for the j^{th} pixel of the image. The length of the i^{th} projection path, having the projection sample coordinates $(\xi_n; \phi_m)$, through this pixel is $\eta_{i,j}$.

uniform pixels in figure 5.5. Following (5.11) where the support Ω_- is known and the notation of (5.27), the contribution of the image pixel f_j to the forward projection sample p_i is

$$\begin{aligned} \Delta p_{i,j} &= \int_{\eta_{i,j}} f_j \, d\eta \\ &= \eta_{i,j} \, f_j . \end{aligned} \tag{5.28}$$

Note that forward projections estimated from the square pixel method are by nature discontinuous because there is a discontinuity in the model as the rays cross pixel boundaries. Therefore when a ray crosses the boundary between pixels, a discontinuity is introduced into the projection data. This is especially apparent for projections that are horizontal or vertical [Joseph, 1982] where a number of pixel boundaries can be traversed at the one time.

To avoid the need for ray tracing, *pixel-driven* methods are available that require less computation, and through interpolation, avoid discontinuities across pixel boundaries. Pixel-driven methods do not require the ray paths to be traced through each pixel, only the location ξ of the projection sample that passes through the centre of the pixel. The central point of the pixel $(r;\theta)$ projected onto each view ϕ_m is simply calculated through (5.7), $\xi = r\cos(\phi_m - \theta)$. The projection coordinate ξ of the pixel generally lies within neighbouring discrete projection samples, $p(\xi_n;\phi_m)$ and $p(\xi_{n+1};\phi_m)$. Interpolation can be used to estimate $p(\xi;\phi_m)$ from discrete values (in agreement with (5.24)). In many CT applications, linear interpolation has been found to be of sufficient accuracy [Peters, 1981]. Referring to figure 5.5, and in agreement with (5.24), (5.27) can be rewritten for pixel-driven BP with linear interpolation :

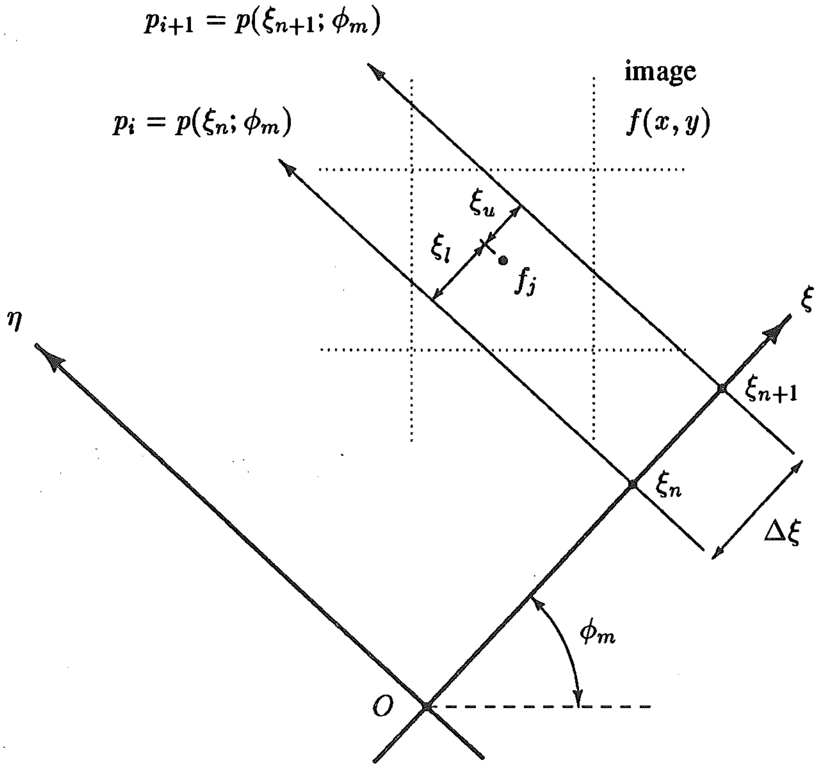


Figure 5.5 Linear interpolation of projections with rays lying on opposite sides from the centre of the image pixel. The spacing between rays at $(\xi_n;\phi_m)$ and $(\xi_{n+1};\phi_m)$ from the centre of the pixel are ξ_l and ξ_u, respectively.

$$\begin{aligned}
\Delta b_{i,j} &= \mathcal{I}\{\, p(\xi_n;\phi_m)\,\}(\xi) \;+\; \mathcal{I}\{\, p(\xi_{n+1};\phi_m)\,\}(\xi) \\
&= \left(1 - \frac{|\xi - \xi_n|}{\Delta\xi}\right) p(\xi_n;\phi_m) \;+\; \left(1 - \frac{|\xi - \xi_{n+1}|}{\Delta\xi}\right) p(\xi_{n+1};\phi_m) \\
&= \frac{|\xi_u|}{\Delta\xi}\, p_i \;+\; \frac{|\xi_l|}{\Delta\xi}\, p_{i+1} \\
&= a_{i,j}\, p_i \;+\; a_{i+1,j}\, p_{i+1},
\end{aligned} \tag{5.29}$$

where for linear interpolation it can be seen that

$$\mathcal{I}\{\, p(\xi_n; \phi_m) \,\}(\xi) = \begin{cases} \left(1 - \dfrac{|\xi - \xi_n|}{\Delta\xi}\right) p(\xi_n; \phi_m), & \text{for } \xi_{n-1} \leq \xi \leq \xi_{n+1}, \\ 0 & \text{otherwise.} \end{cases} \tag{5.30}$$

Back-projection employing interpolation suffers from reduced resolution of the reconstructed image because of the finite pixel spacing and number of projection angles. As can be seen from (5.30), linear interpolation spreads the contribution of each ray $p(\xi_n; \phi_m)$ by up to the distance of their spacing $\Delta\xi$. Similarly, in forward projection the contribution of pixels are spread out to projection samples which are spaced according to that of the original data given for a reconstruction. The FP of the image pixel f_j onto the projection with coordinates $(\xi_n; \phi_m)$, given by linear interpolation, is

$$\begin{aligned} \Delta p_{i,j} &= \left(1 - \frac{|\xi - \xi_n|}{\Delta\xi}\right) f_j \\ &= \frac{|\xi_u|}{\Delta\xi}\, f_j \\ &= a_{i,j}\, f_j . \end{aligned} \tag{5.31}$$

Joseph [1982] introduces a method which is a combination of the square pixel method and pixel-driven method with linear interpolation. Like the square pixel method, it uses the actual ray path length through the pixel to estimate the contribution of a pixel to a projection in the process of FP (and vice versa for BP), but a separate calculation involving the view angle ϕ is made for the $a_{i,j}$. This calculation, in effect, increases the density of the pixels. Increased resolution can be obtained compared to the pixel-driven method with linear interpolation and without suffering the artefacts incurred from the square pixel method because of the discontinuities.

It is apparent from equations (5.27) through (5.31) that FP and BP can be written as a set of linear equations. From (5.27), (5.28) and the notation of (5.29), (5.31),

$$\begin{aligned} p_i &= \sum_j \Delta p_{i,j} = \sum_j a_{i,j}\, f_j, \\ b_j &= \sum_i \Delta b_{i,j} = \sum_i a_{i,j}\, p_i, \end{aligned} \tag{5.32}$$

where the coefficients $a_{i,j}$ are obtained from the projection model of the j^{th} image element onto the i^{th} projection. The projection models for $a_{i,j}$ used in the square pixel method and the pixel-driven method with linear interpolation method have previously been given in equations (5.27), (5.28) and (5.29), (5.31), respectively. In matrix form, if $\mathbf{A}$ is the forward projection model matrix (*projection matrix*) consisting of the $a_{i,j}$ such that

$$\mathbf{P} = \mathbf{A}\,\mathbf{F}, \tag{5.33}$$

where $\mathbf{F}$ is a vector composed of the discrete representation f_j of the image f and $\mathbf{P}$ is the measurement vector composed of the discrete set p_i, then, from (5.32) it is clear that the same projection matrix can be used for the BP process;

$$\mathbf{B} = \mathbf{A}^T \mathbf{P}, \tag{5.34}$$

where $\mathbf{A}^T$ denotes the transpose of $\mathbf{A}$ and $\mathbf{B}$ contains the back-projection image elements b_j. It is important to note that the projection matrix $\mathbf{A}$ is dependent on the sampling geometry only, not on the distribution or data. In most fixed systems the sampling geometry does not vary from image to image. Once $\mathbf{A}$ has been pre-calculated and stored, this calculation need not be done again. The projection matrix $\mathbf{A}$ can be used for both BP and FP. This useful relationship between the processes of BP and FP is shown in equations (5.33) and (5.34).

The methods for BP and FP outlined in this section so far deal with approximate line integrals. Because in practice, source and sensor apertures are of finite width, the measured projection samples are closer to strip integrals. In systems for which projections are modelled by strips of finite width, the projections are no longer a discontinuous function of the ray position [Joseph, 1982]. Tan *et al.* [1986] have modelled the finite resolution of the detector (i.e., whatever detects the physical process which manifests as rays) by an instrument function denoted $i(\xi)$. The measured projections $\hat{p}$ are then the convolution of ideal projections with the instrument's point spread function (psf) $i(\xi)$,

$$\hat{p}(\xi;\phi) = p(\xi;\phi) \odot i(\xi), \tag{5.35}$$

where $\hat{p}(\xi;\phi)$ is referred to as a *blurred projection.* From (5.35) it can be seen that the instrument psf is what limits the resolution of measurements [Tan *et al.*, 1986; Gordon, 1974]. Without any attempts to deconvolve the instrument function $i(\xi)$ from $\hat{p}(\xi;\phi)$, the reconstructed image is also blurred.

5.6 INVERSION OF THE RADON TRANSFORM

There are a number of formulations that exist for the direct inversion of the Radon transform. That is, formulations that express a distribution $f(r;\theta)$ in terms of its Radon transform $p(\xi;\phi)$ (refer to section 5.1 for a description of the Radon transform). Radon's (1917) solution to the reconstruction problem is the earliest known solution for a 2-D distribution. However, this formulation was not widely known and its importance to fields such as radiological imaging was not recognised until the late 1960's [Barrett,]. As reported by Barrett [1983], a Russian scientist, S. I. Tetel'baum, published in 1957, a paper (Tetel'baum [1957]) containing an independent formulation for the inverse Radon transform which gave images that were found to be free of artefacts or other difficulties. These were overlooked in the English literature as there were no available English translations [Barrett, 1983]. Cormack [1963] gave an independent and mathematically rigorous formulation regarded as the beginning of modern CT [Barrett,]. This formulation is reported upon in section 5.6.2. The main thrust for later research on this topic has been the development of efficient and manageable algorithms that are computer-implemented to provide quality pictures of $f(r;\theta)$. A lot of attention has been given to the "robustness" of these methods in coping with limited data.

In this section some of the direct formulas for inversion, and a number of algorithms developed for inversion of the sampled Radon transform, are described.

5.6.1 Radon's inversion formula

The derivation of Radon's formulation is not given here but for historical interest his formula is stated. Radon's proof has been published [Radon, 1917] and Herman's treatment [Herman, 1980] follows this closely. Following Horn [1978], Radon's inversion formula can be written as

$$f(r;\theta) = \frac{1}{4\pi^2} \int_0^{2\pi} \int_{-\infty}^{\infty} - \left(\frac{1}{\xi - r\cos(\phi - \theta)} \right) \frac{\partial}{\partial \xi}\, p(\xi;\phi)\, d\xi\, d\phi, \qquad (5.36)$$

where $(\xi - r\cos(\phi - \theta))$ is the perpendicular distance t from any arbitrary point $(r;\theta)$ to the Radon integral path $(\xi;\phi)$. This integral with a singularity at $\xi - r\cos(\phi - \theta) = 0$ can be integrated in parts and interpreted as

$$f(r;\theta) = \frac{1}{4\pi^2} \int_0^{2\pi} \lim_{\epsilon \to 0} \int_{-\infty}^{\infty} h_\epsilon(\xi - r\cos(\phi - \theta))\, p(\xi;\phi)\, d\xi\, d\phi, \qquad (5.37)$$

where $h_\epsilon(\xi)$ is regarded as a modifying function that weights the contribution of a projection sample by a function of its perpendicular distance t from the point of reconstruction. It is important to note that each and every projection sample contributes to the reconstruction of a point according to this distance. It can be seen from a definition of the modifying function,

$$h_\epsilon(\xi) = \begin{cases} \dfrac{1}{\epsilon^2} & \text{for } |t| < \epsilon \\ -\dfrac{1}{t^2} & \text{for } |t| \geq \epsilon, \end{cases} \qquad (5.38)$$

that the weighting function for each projection sample is equal to the negative square of its distance t from the point of reconstruction so long as the ray path does not pass too close to the point of reconstruction. The modifying function $h_\epsilon(\xi)$ has the property that the total sum of all weights is zero (here referred to as the *zero-mean* property),

$$\int_{-\infty}^{\infty} h_\epsilon(\xi)\, d\xi = 0. \qquad (5.39)$$

This property is used for later developments in chapter 6.

5.6.2 Cormack's inversion formula

As indicated in section 1.4.1, Allan Macleod Cormack was co-recipient of the Nobel Prize for his contributions to X-ray CT in developing mathematical methods for uniquely reconstructing quantitative cross-sectional images of the human body from its X-ray projections. Cormack [Cormack, 1963; Cormack, 1964] proposed a method of reconstruction based on a circular harmonic expansion (CHE) of the 2-D distribution $f(r;\theta)$ and its Radon transform $p(\xi;\phi)$. This resulted in an inversion formulation for f that is in terms of the polar coordinates $(r;\theta)$. Similarly, a 3-D function and its planar projections can be represented by a spherical harmonic expansion. This 3-D counterpart of the 2-D inversion formula is called the Gegenbauer transform [Barrett, 1983].

If f is finite, single valued and continuous, except along a finite number of arcs in the circle of radius a, then it may be expanded in the following Fourier series or CHE

$$\begin{aligned} f(r;\theta) &= \sum_{n=-\infty}^{\infty} f_n(r)\, e^{in\theta} \\ \text{where} \quad f_n(r) &= \frac{1}{2\pi}\int_0^{2\pi} f(r;\theta)\, e^{-in\theta}\, d\theta \end{aligned} \tag{5.40}$$

since $f(r;\theta)$ is cyclic in θ with period 2π. Similarly a CHE of the Radon transform is given by

$$\begin{aligned} p(\xi;\phi) &= \sum_{n=-\infty}^{\infty} p_n(\xi)\, e^{in\phi} \\ \text{where} \quad p_n(\xi) &= \frac{1}{2\pi}\int_0^{2\pi} p(\xi;\phi)\, e^{-in\phi}\, d\phi \end{aligned} \tag{5.41}$$

and where $p(\xi;\phi)$ is cyclic in ϕ with period 2π as can be seen from (5.3). Referring to figure 5.1, recalling (5.12) and changing the variable of integration from η to r through the identity $r = (\eta^2 + \xi^2)^{1/2}$ (from (5.7)), yields

$$p(\xi;\phi) = \int_\xi^a \frac{[f(r;\theta) + f(r;2\phi-\theta)]\ r\,dr}{(r^2-\xi^2)^{1/2}}. \tag{5.42}$$

Using the CHE of f, (5.40), and from (5.7) the identity $\xi = r\cos(\phi-\theta)$, one can write

$$f(r;\theta) + f(r;2\phi-\theta) = 2\sum_{n=-\infty}^{\infty} f_n(r)\cos\left[n\cos^{-1}(\xi/r)\right]\ e^{in\phi} \tag{5.43}$$

and so obtain an expression for the Radon transform in terms of the CHE of f

$$p(\xi;\phi) = 2\sum_{n=-\infty}^{\infty}\int_\xi^a \frac{f_n(r)\cos\left[n\cos^{-1}(\xi/r)\right]\ r\,dr}{(r^2-\xi^2)^{1/2}}\ e^{in\phi}. \tag{5.44}$$

Since the functions $e^{in\phi}$ form a complete orthogonal set on the interval $0 \leq \phi < 2\pi$, (5.44) must hold for each term in ϕ of the expansion (5.41) :

$$p_n(\xi) = 2 \int_\xi^a \frac{f_n(r) \cos\left[n \cos^{-1}(\xi/r)\right] \, r\, dr}{(r^2 - \xi^2)^{1/2}}. \tag{5.45}$$

The function $\cos(n\cos^{-1}(x))$ is a polynomial of degree n in x and is known as a Tschebycheff (or Chebyshev) polynomial of the first kind [Abramowitz and Stegun, 1965]. It is denoted $T_n(x)$ so that (5.45) can be rewritten as

$$p_n(\xi) = 2 \int_\xi^a \frac{f_n(r)\, T_n(\xi/r)\, r\, dr}{(r^2 - \xi^2)^{1/2}}. \tag{5.46}$$

Equation (5.46) relates the circular harmonic expansions of the distribution and its Radon transform by an integral equation. What is required for an inversion formula is one that relates the CHE of f to an explicit formulation for the CHE of the Radon transform. This can be achieved by multiplying both sides of (5.46) by a common factor $T_n(\xi/z)(z/\xi)\,(\xi^2 - z^2)^{-1/2}$, integrating over the interval $z \leq \xi \leq a$ and interchanging the order of integration on the right hand side,

$$\begin{aligned} \int_z^a \frac{p_n(\xi)\, T_n(\xi/r)(z/\xi)\, d\xi}{(\xi^2 - z^2)^{1/2}} &= 2 \int_z^a \int_\xi^a \frac{f_n(r)\, T_n(\xi/r)\, T_n(\xi/r)(z/\xi)\, r\, dr\, d\xi}{(r^2 - \xi^2)^{1/2} (\xi^2 - z^2)^{1/2}} \\ &= 2 \int_z^a f_n(r) \left[\int_z^r \frac{z\, r\, T_n(\xi/r)\, T_n(\xi/r) \;\; d\xi}{\xi\, (r^2 - \xi^2)^{1/2} (\xi^2 - z^2)^{1/2}} \right] dr. \end{aligned} \tag{5.47}$$

Denoting the integral in square brackets by $I_n(r, z)$, where

$$I_n(r, z) = \left[\int_z^r \frac{z\, r\, T_n(\xi/r)\, T_n(\xi/r) \;\; d\xi}{\xi\, (r^2 - \xi^2)^{1/2} (\xi^2 - z^2)^{1/2}} \right], \tag{5.48}$$

then it can be shown that $I_{n+1}(r, z) = I_{n-1}(r, z)$ and $I_0 = I_1 = \pi/2$ [Cormack, 1963]. Substituting this result into (5.47) and exchanging the order of the equation, yields

$$\pi \int_z^a f_n(r)\, dr = z \int_z^a \frac{p_n(\xi)\, T_n(\xi/r)\, d\xi}{\xi\, (\xi^2 - z^2)^{1/2}}. \tag{5.49}$$

Differentiating the integral in the left hand side of (5.49) with respect to z,

$$\begin{aligned} \frac{d}{dz} \left[\int_z^a f_n(r)\, dr \right] &= \frac{d}{dz} \left[\int f_n(r)\, dr \right]_{r=a} - \frac{d}{dz} \left[\int f_n(r)\, dr \right]_{r=z} \\ &= -\frac{d}{dr} \left[\int f_n(r)\, dr \right] \\ &= -f_n(r), \end{aligned} \tag{5.50}$$

then substituting this result into (5.49) and differentiating the right hand side with respect to the same variable yields (with some minor rearrangements) Cormack's inversion formula

$$f_n(r) = -\frac{1}{\pi}\frac{d}{dr}\int_r^a \frac{r\, p_n(\xi)\, T_n(\xi/r)\, d\xi}{\xi\,(\xi^2 - r^2)^{1/2}}. \tag{5.51}$$

Cormack [1963] derived this inversion formula and goes further to show that it uniquely determines the $f_n(r)$ from $p_n(\xi)$. It is also indicated that this analysis may be generalised for projections samples that are taken along families of curves rather than just straight lines. A formal solution was given for line integrals taken along families of curves passing through the origin.

Equation (5.51) is sometimes referred to as an *external* inversion formula as it can uniquely determine $f(r;\theta)$ within an annulus $a \le r \le b$ from its hollow Radon transforms $\{\, p(\xi;\phi) : a \le \xi \le b \,\}$. Lewitt and Bates [1978] note that implementation of Cormack's external inversion formula for hollow projections is not straightforward in practice. Rose [1984] developed a formulation of the exterior inverse Radon transform for reconstruction of a 3-D body from its projections onto planes. Again a spherical harmonic expansion was used to represent f and p in 3-D. Additional representations have been investigated to facilitate the transformation from the projection data p to the distribution f. Many of these expansions are described well in the following texts [Lewitt, 1990; Smith *et al.*, 1973; Zeitler, 1974].

5.6.3 Transform methods

The class of techniques for inversion of Radon transform data, known as transform methods, are the dominant methods used in commercial CT systems [Garden, 1984]. Many of the transform methods are based on the projection theorem of section 5.2 and are referred to as Fourier techniques. This is explained in section 5.6.3.1. A number of other popular inversion techniques do not utilise an image's Fourier space but employ BP or a similar operation. Examples of such methods are the *rho-filtered back-projection* and *modified back-projection* methods given in sections 5.6.3.2 and 5.6.3.3, respectively. In section 5.6.4, a more recently developed inversion technique known as the method of *linograms* is described. This newer technique has been claimed to be computationally efficient compared to the popular modified back-projection method [Edholm *et al.*, 1988].

5.6.3.1 Fourier techniques

The first application of direct Fourier inversion to CT was in 1956, reported in the field of radio astronomy by Bracewell [1956] [Garden, 1984; Herman, 1980]. However, similar Fourier techniques were reported earlier by crystallographers and by O'Brien [1953] in 1953 to reconstruct the 2-D brightness distribution of the sun [Christiansen and Warburton, 1955]. DeRosier and Klug [1968] formulated a 3-D reconstruction algorithm using direct Fourier inversion for imaging biological specimens and particles from their electron microscope images (electron microscopy).

The Fourier method comprised much that was directly taken over from X-ray crystallography, and was therefore founded on more than two generations of experience [Correspondent, 1971]. Consequently, it gained much popularity creating rivalry for new methods of reconstruction being introduced, these were thoroughly scrutinised by supporters of Fourier inversion methods. An example of this rivalry is described in section 5.6.6 with the introduction of new techniques known as *Algebraic Reconstruction Techniques.*

The essence of Fourier techniques is the projection theorem of section 5.2, which relates a projection's Fourier transform to a central slice of the Fourier transform in the distribution f, from which the projection was obtained. The Fourier space of f is constructed from the Fourier transform of its projections taken from all view angles ϕ. According with the notation introduced for Fourier transformation in section 2.3.1 and the 1-D Fourier transform of a projection with view angle ϕ, $P_\phi(\rho)$, this can be expressed as

$$\begin{aligned} F(\rho;\phi) &= P_\phi(\rho) && \text{for all } 0 \le \phi < \pi \\ &= \mathcal{F}_{(\xi)}\{\, p(\xi;\phi)\,\}(\rho;\phi) && \text{for all } 0 \le \phi < \pi. \end{aligned} \tag{5.52}$$

Using the properties of the inverse Fourier transform, the underlying equation for Fourier techniques can be expressed as

$$\begin{aligned} f(r;\theta) &= \mathcal{F}^{-1}\{\, F(\rho;\phi)\,\}(r;\theta) \\ &= \mathcal{F}^{-1}\{\, P_\phi(\rho)\,\}(r;\theta) && \text{for all } 0 \le \phi < \pi \\ &= \mathcal{F}^{-1}\Big\{\, \mathcal{F}_{(\xi)}\{\, p(\xi;\phi)\,\}\Big\}(r;\theta) && \text{for all } 0 \le \phi < \pi. \end{aligned} \tag{5.53}$$

As mentioned in section 5.3, only a finite number of samples of the Radon transform for any distribution are available from which to reconstruct $f(r;\theta)$. If sufficient samples of the Radon transform $p(\xi_n;\phi_m)$ are available, then through the projection theorem (5.18) it is possible to construct samples of Fourier space $F(\rho_n;\phi_m)$ from the DFT (refer section 2.3.1.3) of the projections. An example of the Fourier space samples for projections taken with a parallel geometry is shown in figure 5.6. The Fourier samples $F(\rho_n;\phi_m)$ lie equi-spaced along central slices or spokes that are separated in angle by $\Delta\phi$. However, if the IFT algorithm used to produce the image space representation of $F(\rho_n;\phi_m)$ through (5.53) requires the samples to be on a regular grid of coordinates that are not polar coordinates, then these points may be estimated by interpolation from the polar coordinate grid.

Methods of interpolation on to a rectangular grid from samples given on a polar grid have been examined by a number of authors [Stark, 1981; Stark *et al.*, 1981]. However, interpolation errors in Fourier space tend to be amplified when the inverse Fourier transform is taken [Bates and McDonnell, 1989; Goodman, 1968]. Straightforward approaches to Fourier domain interpolation have in the past resulted in reconstructions that are inferior to the method of modified back-projection, section 5.6.3.3, as artefacts are introduced which are reminiscent of laser speckle [Garden *et al.*, 1989]. More recent approaches have been developed that apply a better understanding of interpolation techniques [Edholm *et al.*, 1988]. For example, Stark [1981] gives a method of exact interpolation in Fourier space for a function that is space-limited and angularly band-limited.

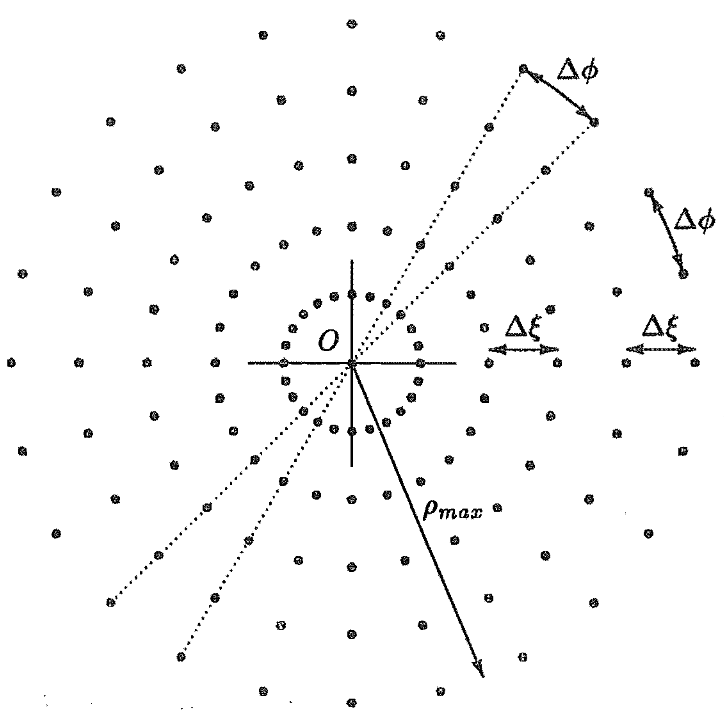


Figure 5.6 Discrete Fourier transform (DFT) samples of F from projections taken with a parallel geometry. Views are equi-spaced in angle by $\Delta\phi$. Projection samples on each view are equi-spaced in image space by $\Delta\xi$ and the corresponding DFT samples of these are equi-spaced by $\Delta\rho$. The radial extent of samples is limited to ρ_{max}.

Crowther *et al.* [1970] proposed an algorithm for direct Fourier inversion based on a Fourier-Bessel transform that directly utilises cylindrical coordinates in both Fourier and image space. Interpolation is then required in image space, as opposed to Fourier space, to obtain the image in Cartesian coordinates. The authors have applied this algorithm to the reconstruction of 3-D images from electron micrographs.

Finally, the method of direct Fourier inversion is thought to be more efficient than convolution methods such as modified back-projection (discussed in section 5.6.3.3) when the matrix being reconstructed is large [Cheung and Lewitt, 1991; Cho *et al.*, 1975; Edholm *et al.*, 1988]. The speed of these techniques has been facilitated by their ability to make use of the efficient FFT algorithms available.

5.6.3.2 Rho-filtered back-projection

Bates and Peters [1971] proposed an algorithm for the reconstruction of $f(r;\theta)$ from its projections $p(\xi;\phi)$ which was referred to later as the *rho-filtered layergram*. At the time of its development, conventional CT machines used the method of BP to produce a blurred image or layergram of f (section 5.4). Rho-filtered back-projection (RFBP) was proposed to eliminate the blurring function within the reconstruction. This section outlines the procedure.

Consider the back-projection $b(r;\theta)$ of $p(\xi;\phi)$ from equation (5.22), section 5.4. Substituting into this the Fourier domain representation for the $P_\phi(\rho)$ from (5.15)

$$\begin{aligned} b(r;\theta) &= \mathcal{B}\{\, p(\xi;\phi)\,\}(r;\theta) \\ &= \int_0^\pi p(r\cos(\phi-\theta);\phi)\, d\phi \\ &= \int_0^\pi \int_{-\infty}^{\infty} P_\phi(\rho)\, e^{i2\pi\rho r\cos(\phi-\theta)}\, d\rho\, d\phi, \end{aligned} \tag{5.54}$$

where in accordance with the notation of section 5.4, the BP image formed from straightforward projections is denoted by $b(r;\theta)$. Using the projection theorem (5.18) in (5.54), $b(r;\theta)$ is related to the Fourier transform of the distribution, $F(\rho;\phi)$ by

$$b(r;\theta) = \int_0^\pi \int_{-\infty}^{\infty} F(\rho;\phi)\, e^{i2\pi\rho r\cos(\phi-\theta)}\, d\rho\, d\phi. \tag{5.55}$$

Comparing (5.55) with the FT relationships (2.43) and (2.44) it can be deduced that

$$\begin{aligned} B(\rho;\phi) &= F(\rho;\phi)\ \frac{1}{\rho}, \\ b(r;\theta) &= f(r;\theta) \odot \frac{1}{r}, \end{aligned} \tag{5.56}$$

where the functions $1/\rho$ and $1/r$ are referred to as blurring functions in Fourier and image space, respectively. This is essentially the reconstruction error of BP referred to in section 5.4. Bates and Peters [1971] method of RFBP suggests deblurring the BP in the Fourier space to make use of the efficient FFT algorithms available. Alternatively, in image space $b(r;\theta)$ can be corrected by deconvolving the radial blurring function $1/r$. Thus, rho-filtered back-projection is a two step process, firstly constructing the back-projection image, and secondly, filtering this with the rho-filter function $1/\rho$ (hence its name RFBP).

Rowland [1979] concluded that this method requires a significantly larger computation time than the modified back-projection method of section 5.6.3.3. It should be noted that this method does not permit the computation of a single arbitrary point in the reconstruction to be made. Instead the complete reconstruction of image space is required.

5.6.3.3 Modified back-projection

In 1967, Bracewell and Riddle [1967] derived an alternative formulation that is now more extensively used in CT than Fourier methods for the reconstruction of images from projections. *Modified back-projection* (MBP) is one of the most popular image reconstruction techniques [Edholm *et al.*, 1988]. A major reason for its popularity and acceptance after earlier Fourier techniques being that a straightforward implementation of this method leads to images that are more accurate than those produced from a straightforward implementation of the basic Fourier method [Cheung and Lewitt, 1991]. Later, Ramachandran and Lakshminarayanan [1971] suggested just such a method as an alternative for the Fourier technique of DeRosier and Klug [1968] (reported in section 5.6.3.1) used in the reconstruction of 3-D structures from electron micrographs. It involved the building of a 3-D structure from adjacent 2-D cross-sections. Ramachandran and Lakshminarayanan [1971] proposed the MBP method, supposedly independent from the derivation of Bracewell and Riddle [1967], for the reconstruction of the 2-D cross-sections and gave a simple expression for a discrete filter to be used in modifying sampled projections prior to BP.

Bracewell and Riddle [1967] interpreted his formulation in two parts. The first was the modification of the projections he called "line scans" (the Radon transform samples for each view angle), and secondly, the superposition of these modified projections across the image space to be reconstructed. This superposition is now more commonly referred to as back-projection. This method of reconstruction is therefore called modified back-projection although it is often referred to in the literature as filtered or convolution back-projection. Bracewell and Riddle [1967] in this work also considered the number of projections required to reconstruct an image whose spatial frequencies are limited.

To derive the MBP formulation, recall from section 2.3.1 that the function f can be expressed as an inverse transform of its Fourier representation in polar coordinates (equation (2.43)). For $p(\xi;\phi)$ and $F(\rho;\phi)$ to be single valued, it is seen, respectively, that [Tan *et al.*, 1986]

$$p(-\xi;\phi) = p(\xi;\phi+\pi) \quad \text{and} \quad F(-\rho;\phi) = F(\rho;\phi+\pi). \tag{5.57}$$

Substituting the identity (5.57) into (2.43) yields

$$f(r;\theta) = \int_0^\pi \int_{-\infty}^{\infty} F(\rho;\phi)\, e^{i2\pi\rho r\cos(\phi-\theta)}\, |\rho|\, d\rho\, d\phi. \tag{5.58}$$

By invoking the projection theorem (equations (5.18) and (5.14)), $f(r;\theta)$ can be expressed in terms of its projections :

$$\begin{aligned} f(r;\theta) &= \int_0^\pi \int_{-\infty}^{\infty} P_\phi(\rho)\, e^{i2\pi\rho r\cos(\phi-\theta)}\, |\rho|\, d\rho\, d\phi \\ &= \int_0^\pi \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(\xi';\phi)\, e^{i2\pi\rho(r\cos(\phi-\theta)-\xi')}\, |\rho|\, d\xi' d\rho\, d\phi. \end{aligned} \tag{5.59}$$

Denoting $h(\xi)$ as the modifying function referred to in section 5.6.1, here defined by

$$h(\xi) = \int_{-\infty}^{\infty} |\rho|\, e^{i2\pi\xi\rho}\, d\rho, \tag{5.60}$$

and substituting this definition into (5.59) above gives the inner integral as a convolution of the projections p and the modifying function h. By making use of the definition for the modifying function (5.60), then recalling from (5.7) that $\xi = r\cos(\phi - \theta)$ and (2.44), (5.59) can be re-expressed as

$$\begin{aligned} f(r;\theta) &= \int_0^{\pi}\int_{-\infty}^{\infty} p(\xi';\phi)\, h(r\cos(\phi-\theta)-\xi')\, d\xi' d\phi \\ &= \int_0^{\pi}\int_{-\infty}^{\infty} p(\xi';\phi)\, h(\xi-\xi')\, d\xi' d\phi \\ &= \int_0^{\pi} [\, p(\xi;\phi)\odot h(\xi)\,]\; d\phi \\ &= \mathcal{B}\{\, p(\xi;\phi)\odot h(\xi)\,\}(r;\theta). \end{aligned} \tag{5.61}$$

Equation (5.61) involves two mathematical transforms; firstly the convolution of the projections with $h(\xi)$, and secondly, the BP of these modified projections. These correspond to the two steps required for MBP. Thus, modified back-projection can be written as

$$f(r;\theta) = \mathcal{B}\{\, \tilde{p}(\xi;\phi)\,\}(r;\theta), \tag{5.62}$$

where $\tilde{p}(\xi;\phi)$ are the modified projections defined by

$$\tilde{p}(\xi;\phi) = p(\xi;\phi) \,\odot\, h(\xi). \tag{5.63}$$

It can be seen from (5.60) that the modifying function when convolved with the projections amplifies their frequency content by the factor $|\rho|$. Having infinite bandwidth, the modifying function will amplify any high frequency noise to unacceptable levels thereby producing artefact in the reconstructed image. Consequently, the modifying function is bandlimited first by a windowing function $W(\rho)$ of finite width. Equation (5.60) can be rewritten as

$$\begin{aligned} h(\xi) &= \int_{-\infty}^{\infty} W(\rho)\,|\rho|\, e^{i2\pi\xi\rho}\, d\rho \\ &= \mathcal{F}^{-1}\{\, W(\rho)\,|\rho|\,\}(\xi). \end{aligned} \tag{5.64}$$

For MBP the first process for reconstruction is to modify the projections $p(\xi;\phi)$ by convolving them with $h(\xi)$ through (5.63) for each view angle ϕ. Since there are only a discrete number of samples for each view angle and a discrete number of view angles ϕ_m, the convolution integral of (5.63) is approximated by the Riemann sum (assuming equi-spaced projection samples)

$$\tilde{p}(\xi_n;\phi_m) = \Delta\xi \sum_{n'=-N}^{N} p(\xi_{n'};\phi_m)\, h(\xi_{n-n'}) \tag{5.65}$$

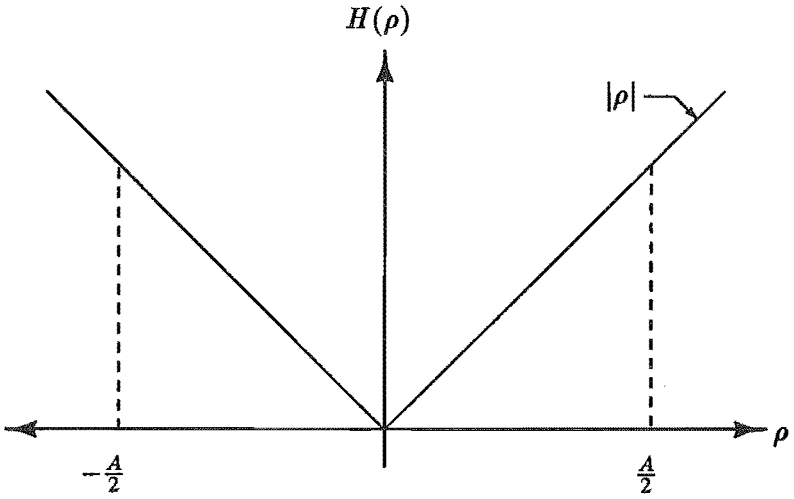


Figure 5.7 Modifying function or rho-filter $h(\xi)$ shown in Fourier space as $H(\rho)$. The dotted lines show the cut-off frequency $A/2$ that $H(\rho)$ is bandlimited to.

for each view angle where $\Delta\xi$ is the spacing between projection samples. Secondly, the projections are mapped onto image space by the discrete back-projection formula of (5.25) :

$$\begin{aligned} \hat{f}(r;\theta) &= \Delta\phi \sum_{m=0}^{M-1} \sum_{n=-N}^{N} \mathcal{I}\{\, \tilde{p}(\xi_n;\phi_m)\,\}(\xi) \\ &= \Delta\phi\,\Delta\xi \sum_{m=0}^{M-1} \sum_{n=-N}^{N} \mathcal{I}\left\{ \sum_{n'=-N}^{N} p(\xi_{n'};\phi_m)h(\xi_{n-n'}) \right\}(\xi). \end{aligned} \tag{5.66}$$

For successful reconstruction with MBP, an appropriate discrete form of the modifying filter must be chosen. A number of these have been developed [Bracewell and Riddle, 1967; Kwoh *et al.*, 1977; Lewitt, 1979b; Ramachandran and Lakshminarayanan, 1971; Shepp and Logan, 1974]. The filter given by Bracewell and Riddle [1967] was band-limited by an ideal low-pass filter with cut-off frequency $A/2$ (see figure 5.7) giving the following window

$$W(\rho) = \begin{cases} 1 & |\rho| \leq A/2 \\ 0 & |\rho| > A/2. \end{cases} \tag{5.67}$$

Ramachandran and Lakshminarayanan [1971] used the same window function (5.67) but also gave a discrete representation of $h(\xi)$ for modifying of equi-spaced projection samples by discrete convolution with

$$h(n\Delta\xi) = \begin{cases} 1/4\Delta\xi^2 & n = 0 \\ -1/\pi^2 n^2 \Delta\xi^2 & n \text{ odd} \\ 0 & n \text{ even.} \end{cases} \tag{5.68}$$

The effect of a sharply truncating window, as per (5.67), is that the discrete modifying filter has a discontinuous derivative in the region of the folding frequency harmonics. This leads to "overshoot" in the reconstruction of sharp transitions in the distribution $f(r;\theta)$ (Gibbs' phenomenon) [Lewitt *et al.*, 1978]. Shepp and Logan [1974] also reported that they found this filter function (5.68) led to oscillations in the reconstruction. Shepp and Logan [1974] and Tan *et al.* [1986] suggest a filter without the sharp truncation in Fourier space. Shepp and Logan [1974] gave the following discrete filter representation without a sharp transition in Fourier space :

$$h(n\,\Delta\xi) = \frac{-2}{(\pi\,\Delta\xi)^2\,(4n^2-1)}, \tag{5.69}$$

having the Fourier domain representation

$$H(\rho) = \left|\frac{1}{\Delta\xi}\sin(\pi\,\Delta\xi\,\rho)\right| \left(\frac{\sin(\pi\,\Delta\xi\,\rho)}{\pi\,\Delta\xi\,\rho}\right)^2. \tag{5.70}$$

The *generalised hamming* window has the form

$$H(\rho) = [\alpha + (1-\alpha)\cos(2\pi\,\Delta\xi\rho)] \quad \text{for} \quad 0 \le \alpha \le 1. \tag{5.71}$$

and the discrete representation can be estimated by taking the following three-point running average a discrete representation of the ideal modifying filter (suc as given in equation (5.60)) [Rowland, 1979] :

$$w(n\Delta\xi) = \begin{cases} \alpha/\Delta\xi & n=0 \\ (1-\alpha)/\Delta\xi & |n|=1 \\ 0 & \text{otherwise.} \end{cases} \tag{5.72}$$

Rowland [1979] investigated the relative merits of different filter and modifying functions. This includes evaluation of their sensitivity to noise, overshoot and reconstruction resolution.

It is also worth noting here that the method of MBP is thought to be more computationally efficient than direct Fourier inversion when the matrix being reconstructed is not large, and, has been found to be "robust" when reconstructing images from incomplete projection data [Lewitt, 1979a].

5.6.3.4 Discrete modified back-projection

It is convenient at this point, for the benefit of later developments in chapter 6, to express the linear modified back-projection operation (MBP) in terms of matrix notation. As mentioned in section 5.6.3.3, MBP can be formulated as a two-part process. The second part, back-projection, has already been considered as a discrete process in section 5.5 where it was formulated in terms of the projection matrix, A, shown in equation (5.34). This section describes the first part of MBP, the modification of the discrete projection data p_i, as a matrix operation so that the entire MBP operation can be formulated in terms of matrices.

Consider the modification of the discrete projection data $p(\xi_n; \phi_m)$ for $-N \leq n \leq N$ and $0 \leq m \leq M-1$ by the discrete modifying function $h(\xi_n)$. Let $\mathbf{P}_m$ be a column vector consisting of the discrete set of $p(\xi_n; \phi_m)$ for $-N \leq n \leq N$ from one projection angle ϕ_m, then the column vector representing the modified projection data is

$$\tilde{\mathbf{P}}_m = \mathbf{h}\,\mathbf{P}_m, \tag{5.73}$$

where $\mathbf{h}$ is a modifying matrix. The modifying matrix, for any projection ϕ_m, is $\mathbf{h}$, such that

$$\mathbf{h} = \begin{bmatrix} h_{11} & \cdots & h_{1(2N+1)} \\ \vdots & \ddots & \vdots \\ h_{(2N+1)1} & \cdots & h_{(2N+1)(2N+1)} \end{bmatrix} \tag{5.74}$$

where in accord with (5.65) the elements of $\mathbf{h}$ along each row $\left[h_{n'1} \cdots h_{n'(2N+1)} \right]$ are

$$h_{n'n} = \Delta\, h(\xi_{n-n'}). \tag{5.75}$$

It can be seen from (5.75) that each row of $\mathbf{h}$ is equal to the coefficients of the filter shifted right by one position for each row, so that

$$h_{n'n} = h_{(n'+1)(n+1)}. \tag{5.76}$$

and $\mathbf{h}$ is therefore diagonally constant. Since the modifying function $h(\xi)$ is most often symmetrical about $\xi = 0$, as is the case for those given in (5.68) and (5.69), $h_{n'n} = h_{nn'}$, or

$$\mathbf{h} = \mathbf{h}^T. \tag{5.77}$$

The properties given in (5.76) and (5.77) are those of a Toeplitz [Bitmead and Anderson, 1981] matrix as $\mathbf{h}$ is symmetric and diagonally constant. The full measurement vector

$$\mathbf{P} = \begin{bmatrix} \mathbf{p}_0 \\ \mathbf{p}_1 \\ \vdots \\ \mathbf{p}_{M-1} \end{bmatrix} \tag{5.78}$$

can thereby be modified by the modifying matrix $\mathbf{H}$ composed of M matrices $\mathbf{h}$ lying on the diagonal :

$$\mathbf{H} = \begin{bmatrix} \mathbf{h} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{h} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{h} \end{bmatrix}, \tag{5.79}$$

where $\mathbf{0}$ is the zero matrix. The modified measurement vector can now be expressed as

$$\tilde{\mathbf{P}} = \mathbf{H}\,\mathbf{P}. \tag{5.80}$$

Combining (5.80) with (5.62) and (5.34), modified back-projection can be expressed in the following matrix form

$$\begin{aligned} \mathbf{F} &= \mathbf{A}^T \tilde{\mathbf{P}} \\ &= \mathbf{A}^T \{\mathbf{H}\,\mathbf{P}\} \\ &= \left\{\mathbf{A}^T \mathbf{H}\right\} \mathbf{P}. \end{aligned} \tag{5.81}$$

If both parts of the MBP process are represented by the one *reconstruction matrix*, $\mathbf{M}$, then (5.81) can be rewritten as

$$\mathbf{F} = \mathbf{M}\,\mathbf{P}, \tag{5.82}$$

and it can clearly be seen, on comparing equations (5.81) and (5.82), that the reconstruction matrix

$$\mathbf{M} = \mathbf{A}^T \mathbf{H}. \tag{5.83}$$

5.6.4 Linograms

This section details one of the most recently developed techniques for image reconstruction from projections, the *method of linograms*. The method of linograms, first introduced by Edholm and Herman [1987], is a novel reconstruction technique, the foundation of which can be developed through the modified back-projection method (see section 5.6.3.3) by making use of the Radon transform and the projection theorem (both of whose properties are well known and which are detailed in sections 5.1 and 5.2 respectively). However, although it is a novel technique that can be introduced in an intuitive manner, the mathematical development from modified back-projection to the method of linograms is not easily obtained and is therefore explained in some detail. The aim of this section is to bring together many of the intuitive aspects published on the method of linograms [Edholm and Herman, 1987; Edholm, 1987; Edholm *et al.*, 1987; Edholm *et al.*, 1988] and elucidate its mathematical development by additional explanation and detail.

If the Radon transforms or projections $p(\xi;\phi)$ are arranged in a 2-D mapping with projection sample path coordinates ξ and ϕ taken as Cartesian coordinates, then this arrangement of $p(\xi;\phi)$ is known as *projection space* (see figure 5.8). The process of BP, as described in section 5.4 is to integrate the contributions of all projections whose paths $(\xi;\phi)$ intersect the distribution $f(r;\theta)$ at the point $(r;\theta)$. The path of this integral in projection space for each point is a sinusoid described in (5.7) by $\xi = r\cos(\phi-\theta)$, shown in figure 5.8. This arrangement of $p(\xi;\phi)$ is well known and often referred to as a *sinogram* [Edholm, 1977].

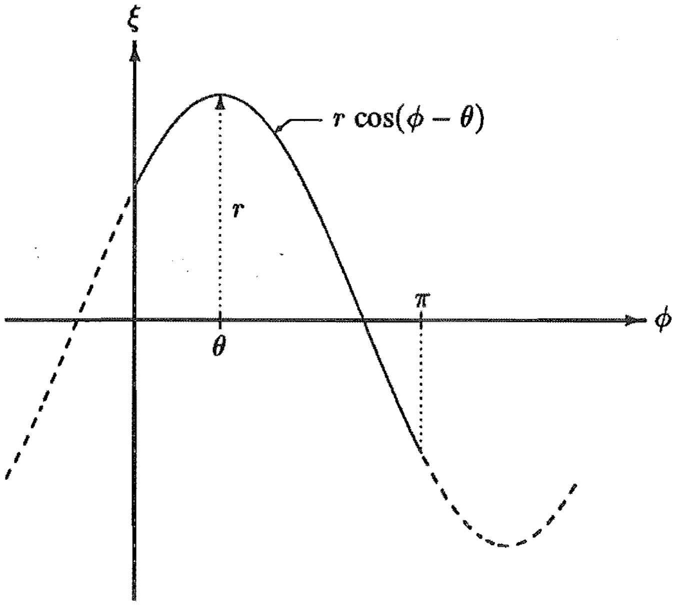


Figure 5.8 Projection space $(\xi;\phi)$ and a sinogram.

In 1987, Edholm and Herman [1987] introduced coordinate transformations that rearranged projection space in such a way that the BP process could be expressed as integrals along straight lines. The corresponding mapping of $p(\xi;\phi)$ to Cartesian coordinates (s,t) is denoted here as *linogram space* and this arrangement of $p(\xi;\phi)$ in (s,t) for BP was referred to by Edholm and Herman [1987] as a *linogram* in analogy with the sinogram.

It was recognised that the BP operation being transformed to line integrals of $p(\xi;\phi)$ in linogram space, could be re-expressed in terms of the Radon transform (5.1). Through the properties of the Radon transform a new reconstruction algorithm was proposed based on these coordinate transformations, the projection theorem (5.18) and Fourier transforms. This reconstruction algorithm was implemented, evaluated further, and the results published [Edholm *et al.*, 1988]. Edholm *et al.* [1988] found the method of linograms to be superior to that of modified back-projection from the point of view of computational speed and equivalent from the point of view of accuracy. It achieves this speed advantage by making use of computationally efficient FFT algorithms and unlike other basic Fourier techniques based on the projection theorem, does not require explicit interpolations. Without explicit interpolations the reconstruction accuracy is not compromised.

Back-projection from linograms is examined in this section first as BP is an essential step in many reconstruction methods (such as those introduced in sections 5.6.3.2 and 5.6.3.3) and the development of this leads to the reconstruction method proposed by Edholm and Herman [1987] using linograms. Firstly, a derivation giving the BP in terms of linograms is presented. The BP, then becoming a series of line integrals, is therefore able to be put into terms involving the Radon transform (5.1). Each linogram represents the transformed projections over a different interval of view angle ϕ. The BP over $0 \leq \phi < \pi$, is broken into sub-intervals as such to ensure the linogram integral along straight lines in linogram space (s, t) are of finite extent. This separation into sub-intervals is considered in more detail later in this section.

For convenience of the derivation, the position of a point in a 2-D reconstruction is represented equivalently by either polar $(r; \theta)$ by its Cartesian coordinates (x, y). The transformation between each representation is simple and is given in (2.42).

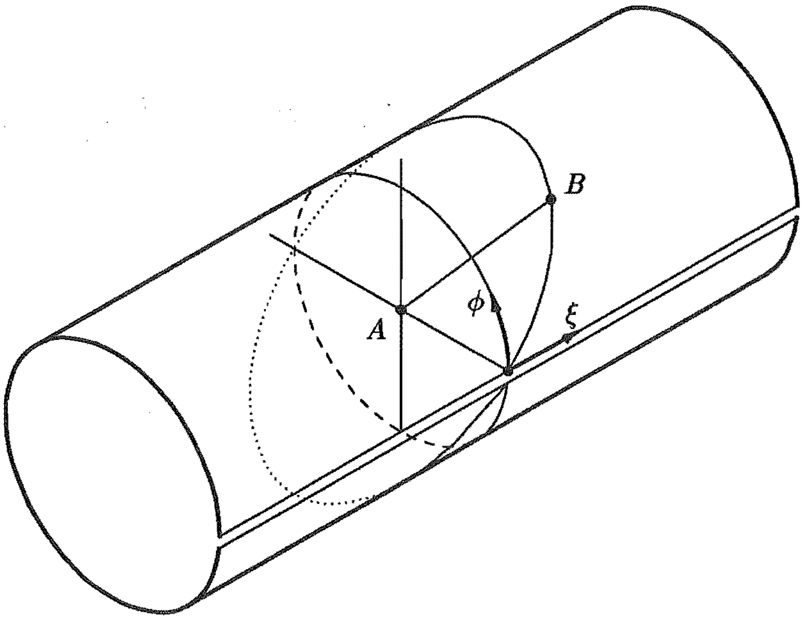


Figure 5.9 A sinogram with the angular range $0 \leq \phi < 2\pi$. The sinusoid $\xi = r\cos(\phi - \theta)$ lies on a plane which intersects the central axis of the cylinder at the point A.

Consider first the sinogram given in figure 5.8 of projection data whose projection path passes through the point $(r; \theta)$. Because the sinogram is periodic with period 2π, it is intuitively attractive to imagine the sinogram as a cylindrical surface (cylinder of revolution) with the ϕ-axis over the range of one period, $0 \leq \phi < 2\pi$, curved around the circumference of the cylinder and the ξ-axis parallel to the cylinder's central axis (refer to figure 5.9). It can be shown that any sinusoid $\xi = r\cos(\phi - \theta)$ will lie on a plane which intersects the central axis of the cylinder at the point A where $\xi = 0$. If, with this point as the centre of the projection, the sinusoid is projected along the line AB onto any plane, then the projection will be a straight line [Edholm, 1987]. This arrangement of the projection data onto a plane is known as linogram space and the projections, when weighted by a suitable factor, correspond to a linogram. Since the aforementioned projection along the line AB of any sinusoid over the full angular range $0 \leq \phi < \pi$ is a line of infinite extent, more than one projection plane (more than one linogram) is required to contain the full range of ϕ, i.e., at least two finite linograms are required.

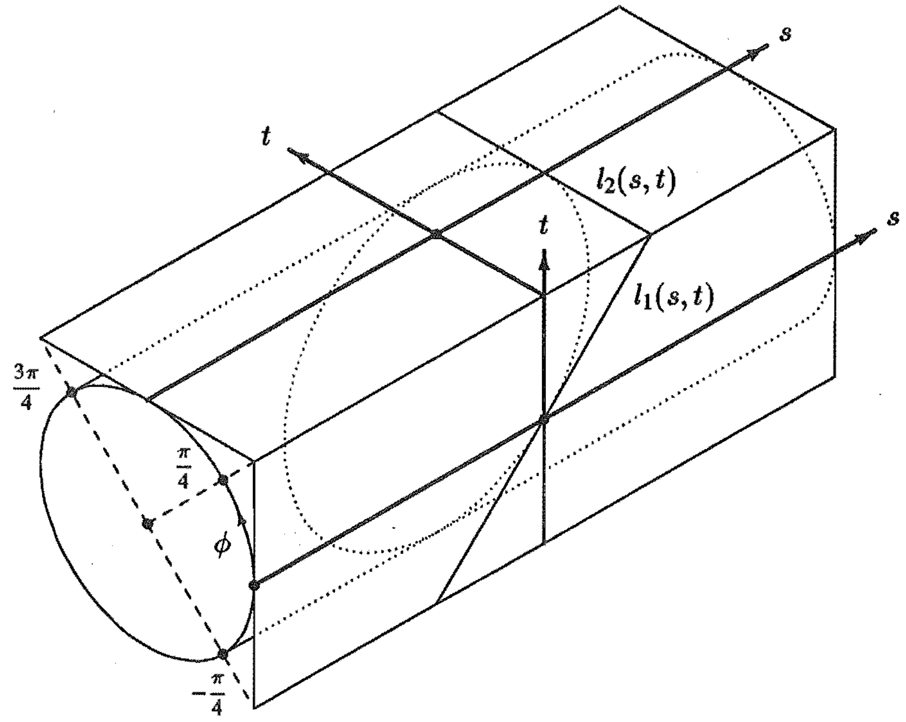


Figure 5.10 The cylindrical sinogram scaled and projected onto two projection planes tangential to the cylinder along the lines $\phi = 0$ and $\phi = \pi/2$ producing the linograms $l_1(s,t)$ and $l_2(s,t)$, respectively. The coordinates of each linogram are (s,t) where the s-axis lies along the tangent line the projection plane makes with the cylinder and the t-axis along the projection of the ϕ-axis.

Consider two linograms, l_1 and l_2, corresponding to two projection planes both of which are tangential to the cylinder at points on the ϕ-axis, $\phi = 0$ and $\phi = \pi/2$, respectively (refer to figure 5.10). The Cartesian coordinate system of each linogram has its origin at the point where the tangent plane intersects the ϕ-axis. The coordinates of each linogram are (s,t) where the s-axis coincides with the tangential line of the corresponding projection plane and the t-axis is the projection of the ϕ-axis onto the plane. It can be seen from figure 5.10 that $l_1(s,t)$ and $l_2(s,t)$ represent the range of projections for $-\pi/4 < \phi < \pi/4$ and $\pi/4 < \phi < 3\pi/4$, respectively.

In the derivation given by Edholm and Herman [1987], BP is considered as the sum of two linograms. It is also appropriate at this point to introduce the two coordinate transformations and the associated linogram transformations of $p(\xi;\phi)$ for BP, denoted $\mathcal{J}_n^{[\kappa,\lambda]}$ and $\mathcal{C}_n^{[\kappa,\lambda]}$ by [Edholm and Herman, 1987] for $n = 1, 2$. These are to be used later in the derivation of the method of linograms. The two coordinate and linogram transformations are, firstly,

$$\xi = s\ \cos(\phi) \quad \text{and} \quad \phi = \tan^{-1}(t) \tag{5.84}$$

and

$$\mathcal{J}_n^{[\kappa,\lambda]}\{\, p(\xi;\phi)\,\}(s,t) = \begin{cases} \left(\dfrac{1}{\sqrt{1+t^2}}\right)^n p\left(\dfrac{s}{\sqrt{1+t^2}};\ \tan^{-1}((-1)^n t)\right) & \text{if } \kappa < t < \lambda \\ 0 & \text{otherwise,} \end{cases} \tag{5.85}$$

and secondly,

$$\xi = t\,\sin(\phi) \quad \text{and} \quad \phi = -\cot^{-1}(s) \tag{5.86}$$

and

$$\mathcal{C}_n^{[\kappa,\lambda]}\{\, p(\xi;\phi)\,\}(s,t) = \begin{cases} \left(\dfrac{1}{\sqrt{1+s^2}}\right)^n p\left(\dfrac{t}{\sqrt{1+s^2}};\ -\cot^{-1}((-1)^n s)\right) & \text{if } \kappa < s < \lambda \\ 0 & \text{otherwise.} \end{cases} \tag{5.87}$$

The geometrical interpretation of these transformations for $n = 2$ is shown in figure 5.10 as the weighted projection of the sinogram onto two planes giving the linograms $l_1(s,t)$ and $l_2(s,t)$. The linograms $l_1(s,t)$ and $l_2(s,t)$ are given in terms of these transformations in equations (5.94) and (5.101), respectively, through the derivation that follows. It can be seen from (5.85) and (5.87) that the interval $[\kappa,\lambda]$ is that which $\mathcal{J}_n^{[\kappa,\lambda]}$ and $\mathcal{C}_n^{[\kappa,\lambda]}$ are defined over the coordinates t and s, respectively.

As intimated earlier, the BP process of (5.22) can be rewritten in terms of the coordinates (s,t) and the linogram transformations $\mathcal{J}_n^{[\kappa,\lambda]}$ and $\mathcal{C}_n^{[\kappa,\lambda]}$. Because of the symmetry of the Radon transform (5.3), the BP integral can be rewritten as

$$\begin{aligned} \mathcal{B}\{\, p(\xi;\phi)\,\}(r;\theta) &= \int_0^{\pi} p(\xi;\phi)\,d\phi \\ &= \int_{-\pi/4}^{3\pi/4} p(\xi;\phi)\,d\phi \\ &= \int_{-\pi/4}^{\pi/4} p(\xi;\phi)\,d\phi \;+\; \int_{\pi/4}^{3\pi/4} p(\xi;\phi)\,d\phi. \end{aligned} \tag{5.88}$$

Both BP terms in (5.88) can be transformed into integrals involving linograms by the coordinate transformations of (5.84) and (5.86), respectively. The linograms resulting from each coordinate transformation over the intervals $-\pi/4 < \phi < \pi/4$ and $\pi/4 < \phi < 3\pi/4$ shown in figure 5.10 are represented by $\ell_1(s,t)$ and $\ell_2(s,t)$, respectively. The reverse coordinate transformation of that given in (5.84), that is the projection coordinates $(\xi;\phi)$ from linogram coordinates (s, t), can be derived with the help of the trigonometric identities

$$\begin{aligned} t &= \tan(\phi) = \frac{\sqrt{1-\cos^2(\phi)}}{\cos(\phi)}, \\ \text{or } \cos(\phi) &= \frac{1}{\sqrt{1+t^2}}, \\ \text{and so} \qquad \xi &= s\cos(\phi) = \frac{s}{\sqrt{1+t^2}}. \end{aligned} \tag{5.89}$$

Projection coordinates $(\xi; \phi)$ are given in terms of coordinates (s, t) for the first linogram transformation (5.84) through the following reverse coordinate transformation

$$\xi = \frac{s}{\sqrt{1+t^2}} \qquad \text{and} \qquad \phi = \tan^{-1}(t). \tag{5.90}$$

Through a similar derivation as (5.89) the reverse coordinate transformation of (5.86) is

$$\xi = \frac{t}{\sqrt{1+s^2}} \qquad \text{and} \qquad \phi = -\cot^{-1}(t). \tag{5.91}$$

Substituting the coordinate transformations (5.84) and (5.90) into the first term of the BP equation, (5.88), yields this operation in linogram space

$$\begin{aligned} \int_{-\pi/4}^{\pi/4} p(\xi; \phi)\, d\phi &= \int_{-1}^{1} \frac{1}{1+t^2}\, p\left(\frac{s}{\sqrt{1+t^2}};\ \tan^{-1}(t)\right) dt \\ &= \int_{-1}^{1} \ell_1(s,t)\, dt, \end{aligned} \tag{5.92}$$

where the first linogram

$$\ell_1(s,t) = \frac{1}{1+t^2}\, p\left(\frac{s}{\sqrt{1+t^2}};\ \tan^{-1}(t)\right) \tag{5.93}$$

and is defined here on the interval $-1 \le t \le 1$. In terms of the transformation (5.85) defined previously,

$$\ell_1(s,t) = \mathcal{J}_2^{[-1,1]}\{\, p(\xi; \phi)\,\}(s,t), \tag{5.94}$$

and the first BP term (5.92) can be rewritten as

$$\int_{-\pi/4}^{\pi/4} p(\xi; \phi)\, d\phi = \int_{-\infty}^{\infty} \mathcal{J}_2^{[-1,1]}\{\, p(\xi; \phi)\,\}(s,t)\, dt. \tag{5.95}$$

In (5.95) the back-projection term has been expressed as an integral of $\ell_1(s,t)$. In projection space the path of this integral is a sinusoid given by $\xi = r\cos(\phi-\theta)$. With the transformation of variables given in (5.84), this path can be seen to have been changed to a straight line in linogram space (s,t) (as intimated earlier) through the following derivation

$$\begin{aligned} \xi &= r\cos(\phi-\theta) \;=\; r\cos(\phi)\,\cos(\theta) \;+\; r\sin(\phi)\,\sin(\theta), \\ \text{or}\quad \frac{\xi}{\cos(\phi)} &= r\cos(\theta) \;+\; r\sin(\theta)\,\tan(\phi), \\ \text{and so}\quad s &= r\cos(\theta) \;+\; r\sin(\theta)\,t. \end{aligned} \tag{5.96}$$

The path $s = r\cos(\theta) + r\sin(\theta)\,t$ is clearly the line with intercept x and slope y in linogram space (s,t)

$$s = x + y\,t, \tag{5.97}$$

showing that the integral path over t for the BP in (5.92) is in fact the line integral

$$\int_{-\pi/4}^{\pi/4} p(\xi;\phi)\,d\phi \;=\; \int_{-1}^{1} \ell_1(x+y\,t,\,t)\,dt. \tag{5.98}$$

A similar derivation shows that the second term of the BP (5.88) can be rewritten as an integral of the second linogram, $\ell_2(s,t)$. Substituting (5.86) and (5.91) into the second term of (5.88) yields

$$\begin{aligned} \int_{\pi/4}^{3\pi/4} p(\xi;\phi)\,d\phi &= \int_{-1}^{1} \frac{1}{1+s^2}\; p\left(\frac{t}{\sqrt{1+s^2}};\; \text{-}\cot^{-1}(s)\right)\,ds \\ &= \int_{-1}^{1} \ell_2(s,t)\,ds, \end{aligned} \tag{5.99}$$

where the second linogram

$$\ell_2(s,t) \;=\; \frac{1}{1+s^2}\; p\left(\frac{t}{\sqrt{1+s^2}};\; \text{-}\cot^{-1}(s)\right) \tag{5.100}$$

and is defined here on the interval $-1 \le s \le 1$. In terms of the transformation (5.87) defined previously,

$$\ell_2(s,t) \;=\; \mathcal{C}_2^{[-1,1]}\{\,p(\xi;\phi)\,\}(s,t), \tag{5.101}$$

and the second BP term (5.99) can be rewritten as

$$\int_{\pi/4}^{3\pi/4} p(\xi;\phi)\,d\phi \;=\; \int_{-\infty}^{\infty} \mathcal{C}_2^{[-1,1]}\{\,p(\xi;\phi)\,\}(s,t)\,ds. \tag{5.102}$$

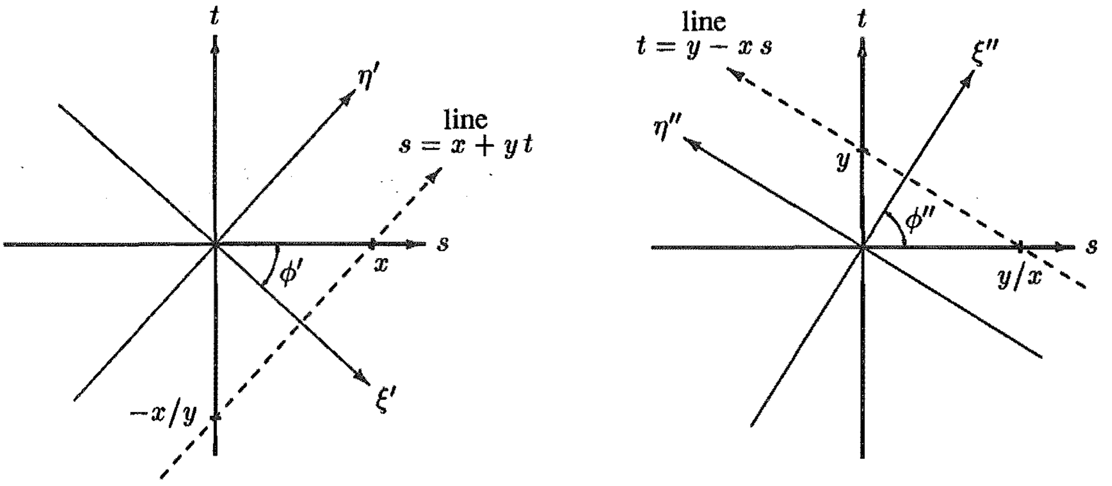
Through a similar derivation as that used in (5.96), for this second coordinate transformation, the coordinates (s, t) are related by the line

$$t = y - x\,s, \tag{5.103}$$

showing that the BP integral over s in (5.99) is also a line integral

$$\int_{\pi/4}^{3\pi/4} p(\xi;\phi)\,d\phi \;=\; \int_{-1}^{1} \ell_2(s,\, y - x\,s)\,ds. \tag{5.104}$$

Both line integral paths (5.97) and (5.103) are shown in linogram space in figures 5.11(*a*) and 5.11(*b*), respectively.



(*a*) $\ell_1(s,t)$ (*b*) $\ell_2(s,t)$

Figure 5.11 Back-projection line integral paths in linogram spaces $l_1(s,t)$ and $l_2(s,t)$ with their projection coordinates $(\xi';\phi')$ and $(\xi'';\phi'')$, respectively.

From (5.92), (5.97), (5.99) and (5.103) it has been shown that the BP image at a point (x, y) is transformed into two line integrals in linogram space;

$$\begin{aligned} \mathcal{B}\{\,p(\xi;\phi)\,\}(x,y) &= \int_{-1}^{1} \ell_1(x + y\,t,\, t)\,dt \;+\; \int_{-1}^{1} \ell_2(s\,, y - x\,s)\,ds \\ &= \int_{-\infty}^{\infty} \mathcal{J}_2^{[-1,1]}\{\,p(\xi;\phi)\,\}\,(x + y\,t,\, t)\,dt \\ &\quad + \int_{-\infty}^{\infty} \mathcal{C}_2^{[-1,1]}\{\,p(\xi;\phi)\,\}\,(s\,, y - x\,s)\,ds. \end{aligned} \tag{5.105}$$

Because only line integrals are involved, it seems intuitive to express them in terms of the Radon transform, from which the raw data is derived. In figure 5.11(*a*) and figure 5.11(*b*) the line integral paths of (5.97) and (5.103), respectively, are shown in linogram space. Also shown are projection coordinates for each line integral path. It can be seen from figure 5.11(a) that for the first linogram $\xi' = x\cos(\phi')$ and $\phi' = \tan^{-1}(y)$. Through a similar derivation as (5.89), the projection coordinate transformation from (s,t) to $(\xi';\phi')$ is

$$\xi' = \frac{x}{\sqrt{1+y^2}} \quad \text{and} \quad \phi' = -\tan^{-1}(y), \tag{5.106}$$

and the integration variable in the first term of (5.105) is

$$dt = \frac{1}{\sqrt{1+y^2}}\, d\eta'. \tag{5.107}$$

For the second linogram the projection coordinate transformation is

$$\xi'' = \frac{y}{\sqrt{1+x^2}} \quad \text{and} \quad \phi'' = \cot^{-1}(x), \tag{5.108}$$

and the integration variable in the second term of (5.105) is

$$ds = -\frac{1}{\sqrt{1+x^2}}\, d\eta''. \tag{5.109}$$

If the integration over t and s in each term of (5.105) is transformed to projection coordinates η' and η'' through (5.107) and (5.109), respectively, then

$$\begin{aligned} \mathcal{B}\{\, p(\xi;\phi)\,\}(x,y) \;&=\; \int_{-\infty}^{\infty} \mathcal{J}_2^{[-1,1]}\{\, p(\xi;\phi)\,\}(s,t)\, \frac{d\eta'}{\sqrt{1+y^2}} \\ &+\; \int_{-\infty}^{\infty} \mathcal{C}_2^{[-1,1]}\{\, p(\xi;\phi)\,\}(s,t)\, \frac{d\eta''}{\sqrt{1+x^2}}, \end{aligned} \tag{5.110}$$

where the sign in (5.109) disappears since the coordinate transformation reverses the direction of integration along the line, Consider the Radon transform of the linograms $\ell_1(s,t)$ and $\ell_2(s,t)$.

$$\begin{aligned} \mathcal{R}\{\, \ell_1(s,t)\,\}(\xi';\phi') \;&=\; \int_{-\infty}^{\infty} \ell_1(s,t)\, d\eta' \\ &=\; \int_{-\infty}^{\infty} \mathcal{J}_2^{[-1,1]}\{\, p(\xi;\phi)\,\}(s,t)\, d\eta', \end{aligned} \tag{5.111}$$

and

$$\begin{aligned}\mathcal{R}\{\,\ell_2(s,t)\,\}(\xi'';\phi'') &= \int_{-\infty}^{\infty} \ell_2(s,t)\,d\eta'' \\ &= \int_{-\infty}^{\infty} \mathcal{C}_2^{[-1,1]}\{\,p(\xi;\phi)\,\}(s,t)\,d\eta''.\end{aligned} \tag{5.112}$$

If (5.111) and (5.112) are compared with each term in (5.110) it can readily be seen that these terms can be written as Radon transforms

$$\begin{aligned}\mathcal{B}\{\,p(\xi;\phi)\,\}(x,y) &= \frac{1}{\sqrt{1+y^2}}\,\mathcal{R}\,\mathcal{J}_2^{[-1,1]}\{\,p(\xi;\phi)\,\}(\xi';\phi') \\ &+ \frac{1}{\sqrt{1+x^2}}\,\mathcal{R}\,\mathcal{C}_2^{[-1,1]}\{\,p(\xi;\phi)\,\}(\xi'';\phi''),\end{aligned} \tag{5.113}$$

where $(\xi';\phi')$ and $(\xi'';\phi'')$ are given in terms of the BP coordinates (x,y) by (5.106) and (5.108), respectively. In changing back to the coordinates of the reconstruction point (x,y) by substituting (5.106) and (5.108) into (5.113), BP is given in terms of the Radon transform and linogram transformations defined in (5.85) and (5.87)

$$\begin{aligned}\mathcal{B}\{\,p(\xi;\phi)\,\}(x,y) &= \frac{1}{\sqrt{1+y^2}}\,\mathcal{R}\,\mathcal{J}_2^{[-1,1]}\{\,p(\xi;\phi)\,\}\left(\frac{x}{\sqrt{1+y^2}};-\tan^{-1}(y)\right) \\ &+ \frac{1}{\sqrt{1+x^2}}\,\mathcal{R}\,\mathcal{C}_2^{[-1,1]}\{\,p(\xi;\phi)\,\}\left(\frac{y}{\sqrt{1+x^2}};\cot^{-1}(x)\right)\end{aligned} \tag{5.114}$$

$$\begin{aligned}&= \mathcal{J}_1^{[-\infty,\infty]}\,\mathcal{R}\,\mathcal{J}_2^{[-1,1]}\{\,p(\xi;\phi)\,\}(x,y) \\ &+ \mathcal{C}_1^{[-\infty,\infty]}\,\mathcal{R}\,\mathcal{C}_2^{[-1,1]}\{\,p(\xi;\phi)\,\}(x,y)\end{aligned} \tag{5.115}$$

$$\begin{aligned}&= \mathcal{J}_1^{[-\infty,\infty]}\,\mathcal{R}\{\,l_1(s,t)\,\}(x,y) \\ &+ \mathcal{C}_1^{[-\infty,\infty]}\,\mathcal{R}\{\,l_2(s,t)\,\}(x,y).\end{aligned} \tag{5.116}$$

Equation (5.116) concludes the initial subject examined in this section, back-projection from linograms. As intimated earlier, in the opening paragraphs of this section, BP from linograms leads to the reconstruction method proposed by Edholm and Herman [1987]. However, as is shown in section 5.6.3.2, BP provides only a blurred image of f. In agreement with the method of modified back-projection, equation (5.62), the reconstruction method examined here must encompass the modification of the projections $p(\xi;\phi)$ as per (5.63) and (5.64), that is,

$$f(x,y) = \mathcal{B}\{\,\tilde{p}(\xi;\phi)\,\}(x,y), \tag{5.117}$$

where

$$\tilde{P}(\rho;\phi) = P(\rho;\phi)\,H(\rho) \tag{5.118}$$

and

$$H(\rho) = W(\rho)\,|\rho|. \tag{5.119}$$

Substituting (5.114), then (5.94) and (5.101) into (5.117) yields

$$\begin{aligned} f(x,y) &= \frac{1}{\sqrt{1+y^2}}\,\mathcal{R}\,\mathcal{J}_2^{[-1,1]}\{\,\tilde{p}(\xi;\phi)\,\}\left(\frac{x}{\sqrt{1+y^2}};-\tan^{-1}(y)\right) \\ &\quad + \frac{1}{\sqrt{1+x^2}}\,\mathcal{R}\,\mathcal{C}_2^{[-1,1]}\{\,\tilde{p}(\xi;\phi)\,\}\left(\frac{y}{\sqrt{1+x^2}};\cot^{-1}(x)\right) \qquad (5.120) \\ &= \frac{1}{\sqrt{1+y^2}}\,\mathcal{R}\Big\{\,\tilde{l}_1(s,t)\,\Big\}\left(\frac{x}{\sqrt{1+y^2}};-\tan^{-1}(y)\right) \\ &\quad + \frac{1}{\sqrt{1+x^2}}\,\mathcal{R}\Big\{\,\tilde{l}_2(s,t)\,\Big\}\left(\frac{y}{\sqrt{1+x^2}};\cot^{-1}(x)\right) \qquad (5.121) \\ &= f_1(x,y) + f_2(x,y), \qquad (5.122) \end{aligned}$$

where $\tilde{l}_1(s,t)$ and $\tilde{l}_2(s,t)$ are the linograms of the modified projections, or alternatively, the modified linograms. It can be seen from (5.122) that the reconstruction can be decomposed into two images, $f_1(x,y)$ and $f_2(x,y)$, which are derived from the modified linograms, $\tilde{l}_1(s,t)$ and $\tilde{l}_2(s,t)$, respectively, where

$$\tilde{l}_1(s,t) = \frac{1}{1+t^2}\;\tilde{p}\left(\frac{s}{\sqrt{1+t^2}};\;\tan^{-1}(t)\right) \tag{5.123}$$

$$\text{and}\quad \tilde{l}_2(s,t) = \frac{1}{1+s^2}\;\tilde{p}\left(\frac{t}{\sqrt{1+s^2}};\;-\cot^{-1}(s)\right). \tag{5.124}$$

Upon examination of (5.121) and (5.122), the two images can be seen to be given by

$$f_1(x,y) = \frac{1}{\sqrt{1+y^2}}\,\mathcal{R}\Big\{\,\tilde{l}_1(s,t)\,\Big\}\left(\frac{x}{\sqrt{1+y^2}};-\tan^{-1}(y)\right), \tag{5.125}$$

$$f_2(x,y) = \frac{1}{\sqrt{1+x^2}}\,\mathcal{R}\Big\{\,\tilde{l}_2(s,t)\,\Big\}\left(\frac{y}{\sqrt{1+x^2}};\;\cot^{-1}(x)\right). \tag{5.126}$$

Clearly, the method of reconstruction from linograms can be considered as the reconstruction of the two images $f_1(x,y)$ and $f_2(x,y)$ which when superimposed as per (5.122) yields $f(x,y)$.

Since the derivation of the method of linograms for reconstruction of $f_1(x, y)$ and $f_2(x, y)$ are closely related, only the derivation for $f_1(x, y)$ is shown here.

Upon examination of (5.125) it can be seen that the image $f_1(x, y)$ is expressed in terms of the Radon transform of the modified linogram $\tilde{l}_1(s, t)$. Consider first the modification of $l_1(s, t)$ to give the modified linogram $\tilde{l}_1(s, t)$ given in (5.123). Taking the 1-D Fourier transform of each side with respect to the first variable s, and using the similarity theorem (equation (2.46)),

$$\tilde{L}_1(S,t) = \frac{1}{\sqrt{1+t^2}}\, \tilde{P}\Big(S\sqrt{1+t^2};\, \tan^{-1}(t)\Big), \tag{5.127}$$

where the Fourier variables of s and t are denoted S and T, respectively. Substituting the modified projections $\tilde{P}$ from (5.118) and (5.119) into (5.127) yields

$$\tilde{L}_1(S,t) = \frac{1}{\sqrt{1+t^2}}\, |S|\; W\Big(S\sqrt{1+t^2}\Big)\; P\Big(S\sqrt{1+t^2};\, \tan^{-1}(t)\Big), \tag{5.128}$$

a Fourier expression for the modified linogram which involves a Fourier transform of the projections. What is required here is an expression involving the linogram only. Taking the Fourier transform of $l_1(s, t)$ in (5.93) using the similarity theorem (2.46) for the right hand side,

$$\begin{aligned} L_1(S,t) &= \frac{1}{\sqrt{1+t^2}}\, P\Big(S\sqrt{1+t^2};\, \tan^{-1}(t)\Big) \\ \text{or}\quad P\Big(S\sqrt{1+t^2};\, \tan^{-1}(t)\Big) &= \frac{1}{\sqrt{1+t^2}}\, L_1(S,t). \end{aligned} \tag{5.129}$$

Inserting the right hand side of (5.129) into (5.128) gives the modified linogram in terms of the modifying function (with window W) and the linogram itself,

$$\tilde{L}_1(S,t) = \Big|S\sqrt{1+t^2}\Big|\; W\Big(S\sqrt{1+t^2}\Big)\; L_1(S,t). \tag{5.130}$$

As intimated in the beginning of this section in regard to the properties of the Radon transform (such as that given in (5.125)), it is intuitive to introduce into the derivation here the projection theorem (refer section 5.2) which relates the 1-D Fourier transform of a function's Radon transform to its 2-D Fourier transform. Taking the 1-D FT of (5.125) with respect to the first variable x and making use of the similarity theorem for the right hand side, gives in the Fourier notation of section 2.3.1

$$F_1(X,y) = \mathcal{F}_{(x)}\mathcal{R}\Big\{\, \tilde{l}_1(s,t) \,\Big\}\Big(X\sqrt{1+y^2};\, \phi'\Big), \tag{5.131}$$

where ϕ' is given in (5.106) and the Fourier variables of x and y are here denoted X and Y, respectively. The projection theorem, equation (5.19), applied to (5.131) states that

$$\begin{aligned}
&\mathcal{F}_{(x)}\mathcal{R}\left\{\,\tilde{l}_1(s,t)\,\right\}\left(X\sqrt{1+y^2};\,\phi'\right)\\
&\qquad = \mathcal{F}\left\{\,\tilde{l}_1(s,t)\,\right\}\left(X\sqrt{1+y^2}\cos(\phi'),\,X\sqrt{1+y^2}\sin(\phi')\right)\\
&\qquad = \mathcal{F}\left\{\,\tilde{l}_1(s,t)\,\right\}(X,\,-yX), \qquad (5.132)
\end{aligned}$$

since in accord with (5.106)

$$\cos(\phi') = \frac{1}{\sqrt{1+y^2}} \quad \text{and} \quad \sin(\phi') = \frac{-y}{\sqrt{1+y^2}}. \qquad (5.133)$$

Returning to the image in (5.131),

$$\begin{aligned}
F_1(X,y) &= \mathcal{F}\left\{\,\tilde{l}_1(s,t)\,\right\}(X,\,-yX)\\
&= \tilde{L}_1(S,T) \quad \text{where} \quad S = X \quad \text{and} \quad T = -yX\\
&= \tilde{L}_1(X,-yX) \qquad (5.134)\\
\text{or} \quad f_1(x,y) &= \mathcal{F}^{-1}_{(X)}\left\{\,\tilde{L}_1(X,-yX)\,\right\}(x,y). \qquad (5.135)
\end{aligned}$$

The line $T = -yX$ and $S = X$ passes through the origin of the modified linogram $\tilde{L}_1(S,T)$ with slope dependent on y. Equation (5.135) therefore says that the image $f_1(x,y)$, for any fixed y, can be found through a 1-D inverse FT of $\tilde{L}_1(S,T)$ along this line through the origin.

The evaluation of the modified linogram $\tilde{L}_1(S,t)$ has been given in (5.130) in terms of a 1-D FT with respect to the first variable s of the linogram $l_1(s,t)$. Similarly, $\tilde{L}_1(S,T)$ can be evaluated with a 1-D FT with respect to the second variable t

$$\tilde{L}_1(S,T) = \mathcal{F}_{(t)}\left\{\,\tilde{L}_1(S,t)\,\right\}(S,T), \qquad (5.136)$$

allowing $\tilde{L}_1(X,-yX)$ to be evaluated for all fixed y.

Table 5.1 summaries the reconstruction algorithms for obtaining both images $f_1(x,y)$ and $f_2(x,y)$ separately starting with the linograms $l_1(s,t)$ and $l_2(s,t)$, respectively. The final step combines both images to give the reconstruction $f(x,y)$. It is apparent from table 5.1 that reconstruction by the method of linograms consists of a series of one-dimensional Fourier transforms (Steps 1, 3 and 4) as well as pointwise multiplications (Step 2), and finally, the addition of the two reconstructed images (Step 5). It should be noted that the implementation

Step 1	A 1-D Fourier transform of $l_1(s,t)$ and $l_2(s,t)$ with respect to the variables s and t, respectively $L_1(S,t) \;=\; \mathcal{F}_{(s)}\{\, l_1(s,t) \,\}\,(S,t))$ $L_2(S,t) \;=\; \mathcal{F}_{(t)}\{\, l_2(s,t) \,\}\,(s,T))$
Step 2	Modification of the linograms in Fourier space by pointwise multiplication with filter functions $\tilde{L}_1(S,t) \;=\; \left\|S\sqrt{1+t^2}\right\|\; W\!\left(S\sqrt{1+t^2}\right)\; L_1(S,t)$ $\tilde{L}_2(s,T) \;=\; \left\|T\sqrt{1+s^2}\right\|\; W\!\left(T\sqrt{1+s^2}\right)\; L_2(s,T)$
Step 3	A 1-D Fourier transform of $\tilde{L}_1(S,t)$ and $\tilde{L}_2(s,T)$ with respect to the other variable t and s, respectively $\tilde{L}_1(S,T) \;=\; \mathcal{F}_{(t)}\left\{\, \tilde{L}_1(S,t) \,\right\}(S,T)$ $\tilde{L}_2(S,T) \;=\; \mathcal{F}_{(s)}\left\{\, \tilde{L}_2(s,T) \,\right\}(S,T)$
Step 4	A 1-D inverse Fourier transform of $\tilde{L}_1(X,-yX)$ and $\tilde{L}_2(xY,Y)$ for all fixed values of y and x, respectively $f_1(x,y) \;=\; \mathcal{F}^{-1}_{(X)}\left\{\, \tilde{L}_1(X,-yX) \,\right\}(x,y)$ $f_2(x,y) \;=\; \mathcal{F}^{-1}_{(Y)}\left\{\, \tilde{L}_2(xY,Y) \,\right\}(x,y)$
Step 5	Superposition of both images $f(x,y) \;=\; f_1(x,y) \;+\; f_2(x,y)$

Table 5.1 Steps for reconstruction of $f(x,y)$ from linograms $l_1(s,t)$ and $l_2(s,t)$ by the method of linograms.

of this method in a discrete manner requires additional developments which is the subject of [Edholm *et al.*, 1988].

The use of Fourier transforms to achieve reconstructions is a property that the linogram method shares with Fourier techniques (refer section 5.6.3.1). The method of linograms also makes use of the computationally efficient FFT algorithms that require operations of the order of $N^2 \log N$. However, unlike basic Fourier techniques, no explicit interpolation is required. As the MBP method (refer section 5.6.3.3) requires computations of the order N^3 for appreciable reconstructions, the method of linograms is computationally efficient compared to back-projection methods and is of comparable accuracy since no interpolation is required [Edholm *et al.*, 1987; Edholm *et al.*, 1988].

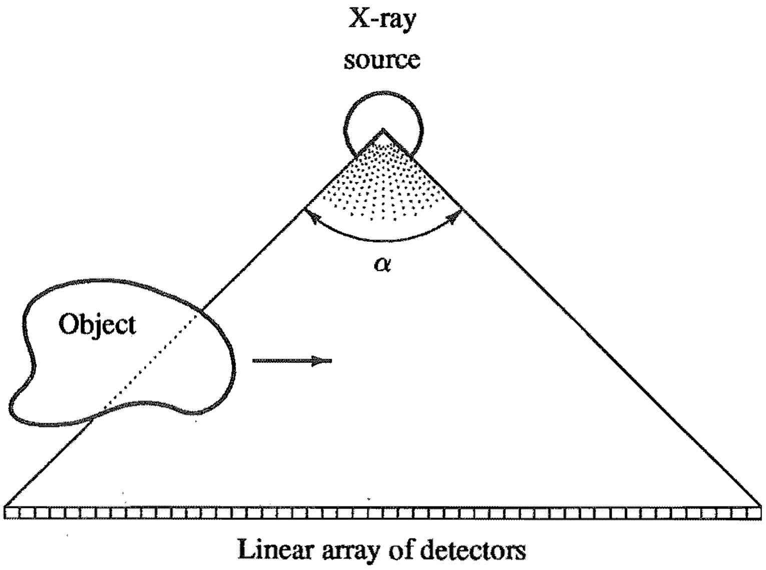


Figure 5.12 Scanner geometry for generating linograms directly. The object moves at a constant velocity parallel to the linear detector array. Multiple linograms are generated by rotating the object before rescanning.

Although linograms are presented here as a transformation of the projections $p(\xi;\,\phi)$, they can in fact be produced directly from a number of scanning geometries. A linogram scanner has been proposed by Edholm [1987] which is simple in construction and particularly appropriate for use in industrial applications. Figure 5.12 depicts this linogram scanner. The object is moved throughout the fan-beam of an X-ray source at a constant velocity parallel to a linear array of detectors. The object's movement through the X-ray beam can be simply realised by a mechanical conveyor. Since the beam-width of the X-ray source is finite, as indicated by the angle $\alpha < \pi$, the object must be rotated in order to obtain multiple linograms. For the case where $\alpha = \pi/2$, two linograms are required (as is the case for the derivation given), $l_1(s,t)$ and $l_2(s,t)$. The second linogram is obtained by a 90 degree rotation of the object with respect to the given apparatus in figure 5.12 and a second scan taken. Similarly, the object can be constructed from three linograms with a fan-beam width $\alpha = \pi/3$. The second and third linograms are obtained after subsequent rotations of the object by 60 degrees.

5.6.5 Finite series expansion reconstruction methods

Series expansion methods (SEM) are basically different from transform methods for inversion of the Radon transform. In SEMs, the problem is discretised from the very beginning whereas transform methods treat it as a continuous problem until the very end when the final formulae are "discretised" for computational implementation [Censor, 1983; Herman, 1980]. One of the first SEMs to be developed is that of the *Algebraic Reconstruction Technique* (ART) by Gordon *et al.* [1970] (refer section 5.6.6).

The discretisation starts with the image representation. It is assumed the image can be represented by a linear combination of basis functions (see also section 2.3.2). In the reconstruction of cross-sectional information, pixels $\hat{f}_j$ are normally chosen as basis functions. Refer to figure 2.3 and equation (2.51) for a description of the pixelation of an image. If the basis functions are chosen so that they are linearly independent (that is neither can be described by a linear combination of the others), then the image representation (2.51) is unique in the $\hat{f}_j$ [Herman, 1980].

If the individual measurement process for the discrete data p_i taken from f is modelled by the functionals $\mathcal{M}_i$, where $\mathcal{M}_i$ are a set of linear and continuous functionals, then the problem at hand is to estimate f from the measured data and measurement model. Since the discrete representation $\hat{f}$ is adequate to represent f by definition,

$$\begin{aligned} p_i &= \mathcal{M}_i\{\, f \,\} \\ &\approx \mathcal{M}_i\{\, \hat{f} \,\}, \end{aligned} \tag{5.137}$$

and because the $\mathcal{M}_i$ are linear, (2.51) can be used to define the measurements in terms of the basis functions ψ_j for $\hat{f}$;

$$\begin{aligned} p_i \approx \mathcal{M}_i\{\, \hat{f} \,\} &= \mathcal{M}_i\left\{ \sum_j \hat{f}_j \, \psi_j \right\} \\ &= \sum_j \hat{f}_j \, \mathcal{M}_i\{\, \psi_j \,\} \\ &= \sum_j \hat{f}_j \, a_{i,j}. \end{aligned} \tag{5.138}$$

As can be seen from (5.138), the $a_{i,j}$ model the interaction of the measurement functionals $\mathcal{M}_i$ with the basis functions ψ_j and, as pointed out in section 5.5, are independent of their weighting $\hat{f}_j$ because of the linearity of $\mathcal{M}_i$. In the case of X-ray CT, the functionals $\mathcal{M}_i$ represent a line integral of the distribution f or Radon transform. Section 5.5 gives two types of projection models for the $a_{i,j}$, the square pixel method and pixel-driven method, equations (5.28) and (5.31), respectively. However, the $\mathcal{M}_i$ are general and may represent more complicated measurements such as integrals along curved ray paths.

Solutions to the linear equations in (5.138) are not simple in general. Because of the desired resolution and size of the imaged region, (5.138) usually consists of a large set of linear equations. When the functionals $\mathcal{M}_i$ are taken along curved paths, the solution may be ill-conditioned [Berryman, 1990]. Often there are insufficient measurements rendering the solution indeterminate (more unknowns than measurements). Several criteria [Herman, 1980] have been proposed to define an optimal solution. These criteria are an artificial means of introducing more equations in order to render the solution less indeterminate and reduce the solution space. If any *a priori* information is known, such as positivity of f, then this can be incorporated to restrain the solution space. Examples of some of these methods using image variance or entropy optimisation and as applied to CT are given in section 5.6.6. Such techniques, falling into the general class known as *regularisation* techniques, are common to the more general class of inverse problems because solutions are often ill-conditioned [Bates *et al.*, 1991]. A review of regularisation techniques can be found in [Bertero *et al.*, 1988].

5.6.6 Algebraic reconstruction techniques

Another class of image reconstruction techniques that were born with an eye to the 3-D reconstruction of biological structures from electron micrographs is that of *Algebraic Reconstruction Techniques* (ART). Gordon *et al.* [1970] proposed ART as an improvement on the popular and newly developed Fourier techniques proposed by DeRosier and Klug [1968] that had gained considerable interest at that time (refer section 5.6.3.1). The instrumentation for measurement and computational methods in use at that time had perhaps given ART advantages when view numbers were small, in computation time and (more controversially) better resolution. In particular ART accomodates reconstruction with small or incomplete data sets, giving a solution, but one that is not necessarily unique. Bender *et al.* [1970] immediately published the results of its application to electron microscopy with the reconstruction of an asymmetric biological structure, ribosome (rat liver). Crowther and Klug [1971] quickly responded to defend their Fourier techniques and a critical account was given of claims made by [Gordon *et al.*, 1970] on their findings. So began a controversial start for the new technique, ART, which is now much better understood and developed. However, it should be noted for historical interest, coincidently, Hounsfield [1972] had developed a reconstruction technique that was essentially the same as the ART, originally filed within a patent specification in 1968 [Herman, 1980].

Early ART was concerned with producing reconstructions by inverting the projection model from which the measurement data was obtained. By doing this, the reconstruction would be consistent with the data from which it was derived. Bender *et al.* [1970] felt that this was a proof of the logical correctness of the data aquisition and reconstruction. Crowther and Klug [1971] pointed out that the mere *consistency* of the reconstruction with a limited number of projections (which was the case in which this method was being applied [Bender *et al.*, 1970]) does not in general guarantee its reliability.

ART is a solution to a general discretised problem, an example of a SEM (refer section 5.6.5) where the reconstruction $\hat{f}$ of a distribution f is divided into a discrete number of basis functions whose weightings $\hat{f}_j$ are here encompassed in the vector $\mathbf{F}$. This discretization is described in section 5.5. Samples of the projections of f are available as data and a model is assumed for their derivation from a discrete representation $\mathbf{F}$. If $\mathbf{P}$ is a vector containing the discrete projection measurement data and $\mathbf{A}$ is the projection model matrix assumed for $\mathbf{P}$, then the initial aim of ART was to obtain a discrete reconstruction F such that

$$\mathbf{A}\,\mathbf{F} = \mathbf{P}, \tag{5.139}$$

thereby, having a reconstruction consistent with the data P and measurement matrix A. The reconstruction F that satisfies (5.139) can clearly be found by inverting the projection matrix

$$\mathbf{F} = \mathbf{A}^{-1}\,\mathbf{P}. \tag{5.140}$$

Gordon *et al.* [1970] gave the initial ART solutions for (5.140) that were based on iteratively refining the elements of F. Two methods were proposed, both began with an initial estimate of the elements $\hat{f}_j$ equal to the average density of the image. Since there are J elements in F, the initial estimate for the image

$$\hat{f}_j^0 = \frac{1}{J}\sum_{j=1}^{J} \hat{f}_j, \tag{5.141}$$

where the superscript on $\hat{f}_j^0$ denotes the initial iteration. The total mass τ of f can be found from any projection as per (5.6), and estimated from any discrete projection as per (5.21). Then (5.141) can be replaced with a calculation for the initial value of the image elements involving the known discrete projections

$$\hat{f}_j^0 = \frac{1}{J}\sum_{n=-N}^{N} p(\xi_n;\phi_m)\,\Delta\xi \qquad \text{for any } m = 1,..,M. \tag{5.142}$$

The first method given, the *direct multiplicative method*, iteratively refines the density at all points in the image along each ray path in succession. It does so multiplying the image values by a factor equal to the ratio of the measured projection to that calculated from the current image. The adjusted density for all image points on the ray path is

$$\hat{f}_j^{q+1} = \left(\frac{p_i}{p_i^q}\right)\hat{f}_j^q \quad \text{for all } \left\{j : \hat{f}_j \in \text{path of } p_i\right\}, \tag{5.143}$$

thereby temporarily restoring each projection sample to the measured projection because after applying (5.143) to image points on the ray path of p_i,

$$\sum_{\{j:\, \hat{f}_j \in \text{path of } p_i\}} a_{i,j}\,\hat{f}_j^{q+1} = p_i. \tag{5.144}$$

After considering the other projections, (5.144) will no longer hold for previous p_i [Gordon *et al.*, 1970]. Thus (5.143) must be applied iteratively until a stable solution for the $\hat{f}_j$ has been found.

In a second method described by Gordon *et al.* [1970], which they denoted the *direct additive method*, the $\hat{f}_j^q$ are iteratively refined by the difference between the measured projection samples p_i intersecting $\hat{f}_j$ and that calculated from the current estimate of $\hat{f}_j^q$. The additive method also enforced positivity. The iterative refinement of $\hat{f}_j^q$ is

$$\hat{f}_j^{q+1} = \max\left[\hat{f}_j^q + (p_i - p_i^q)/N_i, 0\right] \hat{f}_j^q, \tag{5.145}$$

where N_i is the number of elements in $\hat{f}_j$ which belong to the ray path of p_i.

The corrections to densities $\hat{f}_j^q$ made in the above methods are applied in a sequential manner for each projection. Gilbert [1972] proposed a method that changes the density $\hat{f}_j^q$ at each iteration by using data from all the projections simultaneously. It also accounts for the length of the projection sample path L_i through the distribution, which varies with view angle ϕ. This was denoted the *Simultaneous Iterative Reconstruction Technique* (SIRT). New direct additive and multiplicative methods were devised that reportedly [Gilbert, 1972] reduced the deviation of the reconstruction $\hat{f}$ from the original f :

$$\begin{aligned} \hat{f}_j^{q+1} &= \left(\frac{\sum_i p_i \sum_i N_i}{\sum_i L_i \sum_i p_i^q}\right) \hat{f}_j^q && \text{multiplicative method,} \\ \hat{f}_j^{q+1} &= \max\left[\hat{f}_j^q + \left(\frac{\sum_i p_i}{\sum_i L_i} - \frac{\sum_i p_i^q}{\sum_i N_i}\right), 0\right] && \text{additive method,} \end{aligned} \tag{5.146}$$

where L_i is the total length of the projection sample path of p_i through $\hat{f}$. From figure 5.5

$$L_i = \sum \eta_{i,j} \quad \text{for all } \left\{j : \hat{f}_j \in \text{path of } p_i\right\}. \tag{5.147}$$

Solutions to the ART problem (5.140) are in general non-unique [Freider and Herman, 1971]. With limited data p_i a solution is required from an underdetermined set of equations. As mentioned, the early constraints on the solution were the consistency of the measured projection data and reconstructed image, with a chosen projection model. Later, other constraints were suggested to optimise the solution or quality of the reconstruction. [Gaarder and Herman, 1972] proposed a solution that gave a reconstruction of greatest uniformity. Previously solutions obtained from ART were considered to be "peppered". Smoothing could be attained by minimising the variance V of the image. The variance of the image was calculated as follows

$$V^q = \sum_j (\hat{f}_j^q - \bar{f}_j^q)^2, \tag{5.148}$$

where $\bar{f}_j$ refers to the mean value of $\hat{f}_j$. If the projection data is without noise, this is equivalent to solving

$$\min \sum_j (\hat{f}_j^q)^2. \tag{5.149}$$

Gordon *et al.* [1970] suggested that a "best" solution would be one that maximised entropy, or alternatively, the solution that gave no more "information" than that contained in the data. The solution that maximised entropy S would "minimise bias". Thus, with this approach, the introduction of extraneous information or artificial structure can be avoided. The entropy of the image was calculated as follows

$$S^q = -\frac{1}{\ln J} \sum_{j=1}^{J} \left(\frac{\hat{f}_j^q}{\bar{f}_j}\right) \ln \left(\frac{\hat{f}_j^q}{\bar{f}_j}\right). \tag{5.150}$$

The image reconstructed $\hat{f}_j$ can be constrained in a number of ways in order to reduce the solution space. Three common methods of constraint are categorised by Herman *et al.* [1973] and referred to by the following terminology :

$$\begin{aligned}
&\text{Unconstrained ART} && \hat{f}_j = \hat{f}_j^q, \\
&\text{Partially constrained ART} && \hat{f}_j = \begin{cases} 0 & \hat{f}_j^q < 0 \\ \hat{f}_j^q & \text{otherwise,} \end{cases} \\
&\text{Fully constrained ART} && \hat{f}_j = \begin{cases} 0 & \hat{f}_j^q < 0 \\ \hat{f}_j^q & 0 \leq \hat{f}_j^q \leq 1 \\ 1 & \hat{f}_j^q > 1. \end{cases}
\end{aligned} \tag{5.151}$$

As noted earlier, ART algorithms are in the category of series expansion methods. All ART algorithms use the same basis set, the division of the reconstruction space into non-overlapping elements. All share the property of trying to satisfy the consistency of reconstruction with the given data. As such, ART algorithms are intuitively simple and have scope for a variety of solutions.

CHAPTER 6

RECONSTRUCTION FROM A PLETHORA OF VIEWS

6.1 INTRODUCTION

This chapter discusses an idea termed the *plethora of views* idea, first proposed by Professor R.H.T. Bates of the University of Canterbury in 1984 and developed further by Bates and Herman in March 1988 [Bates, 1988b; Herman, 1988]. The theoretical foundation for the studies reported herein is the well-established idea [Crowther *et al.*, 1970] that the number, denoted here by N, of projections needed to reconstruct an image faithfully, to a specified resolution, is proportional to the radius a of the circle C_- circumscribing the cross-section Ω_- (refer to figure 6.1(a)). Since the true image is confined by definition to Ω_-, when presented with what is here defined to be a *plethora* of projections, by which is meant a number $M > N$, so that the permissible reconstruction region can be extended to a radius

$$b = \frac{M}{N}\,a, \tag{6.1}$$

then the intensity (or density as it has been termed in section 5.1) of the reconstructed image should be negligible throughout the region, denoted Ω_+, between the circles C_- and C_+ shown in figure 6.1(a).

The following motivation inspired the research which led to the results presented in this chapter. If each projection, belonging to a given plethora, differs inescapably from what it would *ideally* be if formed from emanations that can be accurately characterised by concourses of straight rays, then it may be possible to iteratively correct the given projections, forcing them to be ideal, by cancelling the reconstructed density within Ω_+. While it is agreed that X-rays can be accurately represented as concourses of straight rays, at least to the accuracies appropriate for computed tomography (refer section 1.4.1), the same cannot be said for macroscopic wave motions either employed or proposed for medical imaging, such as ultrasound (section 1.4.2) or microwaves [Bolomey *et al.*, 1982; Slaney *et al.*, 1984] respectively. The approximations involved in representing ultrasonic wavefronts as concourses of rays are rarely objectionable, from the point of view of CT at any rate [Bates *et al.*, 1991], but the rays tend to be appreciably curved, to accommodate spatial variations of refractive index within the bodies being imaged. It is thought that plethoras of ultrasonic projections, representing either delay or attenuation through a cross-section (refer section 1.4.2), and plethoras of X-ray projections, afflicted by the effects of beam-hardening, are prime candidates for correction in the manner intimated above.

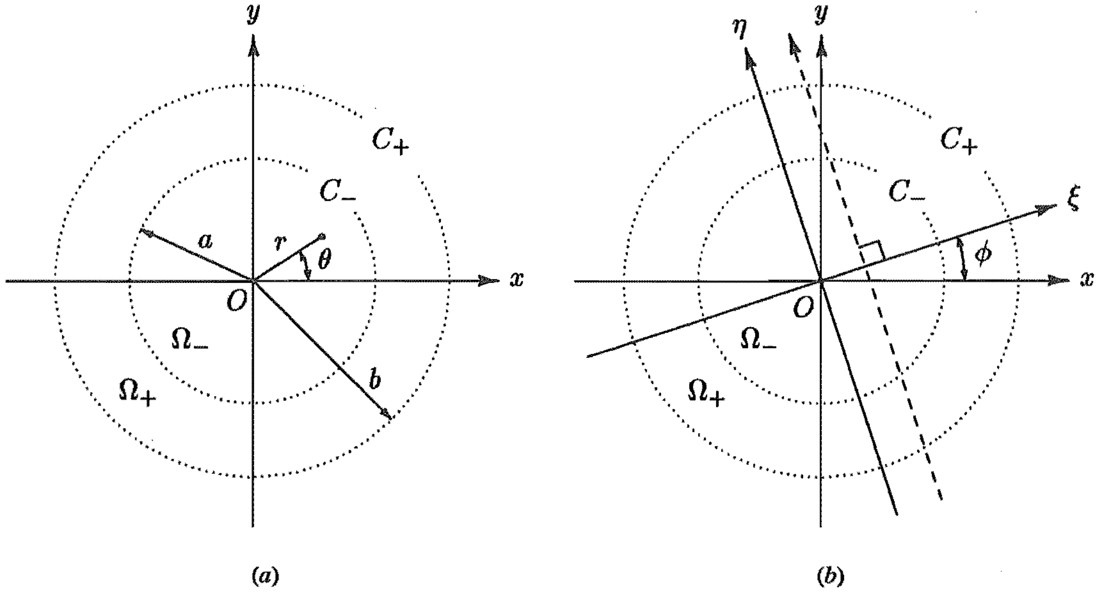


Figure 6.1 Parallel projection geometry for a 2-D distribution contained entirely within the region Ω_- circumscribed by the circle C_- of radius a. The region between C_- and the circle C_+, of radius b, is Ω_+.

When developing an algorithm designed to compensate for unavoidable imperfections of data, it is essential to begin with computer simulations, in order to be able to assess the significance of preliminary results. It is also important to model the imperfections realistically. The latter task is far from trivial with regard to projection data from macroscopic wave motion, because of the enormous computational effort required to compute the passage of wavefronts through bodies having linear dimensions large enough in terms of the wavelength to make the simulations of practical interest. Consequently, the work presented in this chapter starts out with the computationally more manageable, and only slightly less important (from a practical viewpoint), problem of attempting to ameliorate the effects of noise on plethoras of ideal (i.e., straight ray) projections.

Clearly, it can be deduced, from the discussion in the opening paragraph, that any reconstruction within Ω_+, from a plethora of imperfect projections, that departs from what would be termed negligible has been caused by the imperfections of the projections. It is the conjecture of this chapter that contributions within Ω_+ can be utilised to ameliorate the effects of the imperfections on the reconstruction within Ω_- where the distribution is known to be contained.

Section 6.2 presents the plethora of views idea proposed by Bates [1988b] as an open problem. Section 6.3 unifies some of the fundamental theory, on which this problem is based, for the finite number of projections required to reconstruct an image within a region of known radius. It has been necessary to examine, in section 6.4, the level of contamination that can be expected when reconstructing image detail from a finite number of projections. An iterative solution proposed by Herman [1988] for the correction of X-ray projections containing imperfections which is based on the open problem of section 6.2 is given in section 6.5. Presented in section 6.6 through section 6.9 is the initial implementation of this iterative solution aimed at improving the quality in quantitative reconstructions of low-contrast tumours embedded in soft-tissue thereby increasing their contrast. As a consequence of the somewhat anti-climatic results obtained from this iterative solution an alternative approach to correcting imperfections on the projections is explored in chapter 7.

6.2 OPEN PROBLEM

Suppose that there exists a model $\mathcal{M}$ of the physical measurement process generating the measured (or given) imperfect projections :

$$p_g(\xi;\phi_m) \;=\; \mathcal{M}\{\, f(x,y)\,\}\,, \tag{6.2}$$

where the perfect X-ray projections $p(\xi;\phi_m)$ have been modelled in section 5.1 by the Radon transform, represented by $\mathcal{R}$ in accord with (5.2).

Bates [1988a] proposed an open problem : Construct the class of transformations (expected, in general, to be iterative, or at least recursive),

$$\begin{gathered} p_g(\xi;\phi_m) \;\longrightarrow\; \hat{p}(\xi;\phi_m) \\ \text{for} \quad -b<\xi<b \quad \text{and} \quad m=0,1,\ldots,M-1, \end{gathered} \tag{6.3}$$

such that, when all $\hat{p}(\xi;\phi_m)$ are subjected to modified back-projection (refer section 5.6.3.3) $\hat{f}(x,y)$ is negligible within Ω_+, i.e.,

$$\left|\hat{f}(x,y)\right| < \epsilon \quad \text{for all} \quad (x,y)\in\Omega_+. \tag{6.4}$$

Also, find the particular transformation consistent with $\mathcal{M}$ such that

$$\hat{p}(\xi;\phi_m) \;\equiv\; p(\xi;\phi_m). \tag{6.5}$$

An additional constraint on the solution for $\hat{p}(\xi;\phi_m)$ can also be invoked as a means of minimising the loss of information, through (6.3), about the actual density within Ω_- contained in the given data, $p_g(\xi;\phi_m)$. It is suggested that a solution for $\hat{p}(\xi;\phi_m)$ should be that for which

$$\|\, p_g(\xi;\phi_m) - \hat{p}(\xi;\phi_m)\,\| \quad \text{is minimised.} \tag{6.6}$$

6.3 PLETHORA OF VIEWS

As noted in the introductory section, the foundation for the studies reported here is that there are sufficient projections $p_g(\xi;\theta)$ given to reconstruct the smallest detail required, and so that the reconstruction of $f(r;\theta)$ does not contaminate the region Ω_+ if $p_g(\xi;\phi)$ are perfect ideal projections of Ω_-. It is therefore important in the context of this study to understand the relationship between the maximum contamination present in Ω_+ when reconstructing image features of finite width σ, and the number of projections used to achieve that reconstruction. This section deals with the first issue concerning the number of projections required to reconstruct the desired detail. The relationship between the number of projections and the contamination in Ω_+ is deferred to section 6.4.

Crowther *et al.* [1970] were among the first to formulate an expression for the number N of projections required to reconstruct an image of radius a, in terms of the desired resolution. The image distribution is expanded in terms of a finite number of cylindrical functions in ϕ whose coefficients are determined from the N projections (equi-spaced in angle ϕ by $\Delta\phi = \pi/N$). The resolution of this representation is of course limited by the finite number of terms included in the expansion, which in turn is bounded (in order to determine all the expansion coefficients) by the finite number of projections. Crowther *et al.* [1970] gave the minimum number of projections required to reconstruct a distribution as that required to determine the coefficients of all significant terms in the expansion :

$$N \approx \frac{2\pi a}{\sigma}, \tag{6.7}$$

where σ is the required resolution. A similar formulation was given by Bracewell and Riddle [1967] in terms of the extent of spatial frequencies contained in the image $f(r;\theta)$. As mentioned in section 5.2, the 1-D Fourier transforms of the projections at angles ϕ_m are equal to the values along lines passing through the origin of the distribution's Fourier transform $F(\rho;\phi)$ at the respective angles $\phi = \phi_m$. As can be seen from figure 5.6, the maximum radial spacing between samples of $F(\rho;\phi)$ is $\pi\rho_{max}/N$, where ρ_{max} is the radial extent of $F(\rho;\phi)$. Bracewell and Riddle [1967] give a limit for the radial spacing, determined by N, based on the sampling theorem [Bates and McDonnell, 1989] for the maximum allowable distance between Fourier samples $F(u,v)$ on a square lattice that permits unique recovery of the image $f(x,y)$ from its Fourier samples :

$$\frac{\pi\rho_{max}}{N} \leq \frac{1}{2b}. \tag{6.8}$$

If the resolution is taken to be $\sigma = 1/\rho_{max}$, as is done by Crowther *et al.* [1970], then it can be seen that (6.8) is equivalent to (6.7). However, as noted by Bates and McDonnell [1989], because the radial sample spacing of the projection points in $F(\rho;\phi)$ is proportional to ρ, it has become accepted that the "effective" overall sample spacing is $\pi\rho_{max}/2N$. This implies that

$$N \approx \frac{\pi a}{\sigma}. \tag{6.9}$$

6.4 CONTAMINATION FROM FINITE NUMBERS OF PROJECTIONS

One of the characteristics of images reconstructed from projections is that features within the object being imaged cause errors over the whole of the image. Basically the absolute value of these errors increases with distance from the feature responsible, and with the difference in density between the feature and the surrounding material. Naturally these errors are reduced by increasing the number of projections [McKinnon and Bates, 1980]. As noted in section 6.3, when reconstructing the smallest detail required there must be sufficient numbers of projections so that the contamination in Ω_+ is not caused by the finite number of projections.

A single feature in an image is represented here by the gaussian "blob"

$$f(r;\theta) \;=\; A\,e^{-r^2/2\sigma^2} \qquad (6.10)$$

of amplitude A and where σ is the "effective radius" of the "blob" of density constituting the image detail. In reconstructing the image from the perfect projections of $f(r;\theta)$ issue is not taken here with the fact that the projections are sampled at discrete ξ and consequently are by nature of finite resolution. The effects of such on reconstructing gaussian blobs are detailed by Tan *et al.* [1986]. Suffice to say that in this study all samples are equi-spaced with resolution

$$\Delta\xi \;\ll\; \sigma \qquad (6.11)$$

so that contamination from the reconstruction process is attributable to the finite number of projections. The maximum contamination at any point in the reconstruction $\hat{f}(r;\theta)$ as a function of radius from the centre of the image feature was estimated by reconstructing M equi-spaced perfect projections

$$p(\xi;\phi_n) \qquad \text{where} \qquad \phi_m = \frac{m\pi}{M} \quad \text{for} \quad m = 0, 1, \ldots, M-1. \qquad (6.12)$$

Contamination was measured as absolute difference between the true detail distribution $f(r;\theta)$ and that reconstructed, $\hat{f}(r;\theta)$ (but normalised with respect to the amplitude A of this feature). The radius from the centre of the image feature was normalised with respect to the size σ of the original gaussian feature. Figures 6.2(*a*) and 6.2(*b*) show examples of the variation in maximum contamination as a function of the normalised radius in reconstructing from $M = 50$ and $M = 100$ projections, respectively, a single gaussian feature.

It is apparent from figures 6.2(*a*) and 6.2(*b*), that the asymptotic and maximum normalised error that can be encountered when reconstructing a single gaussian feature is

$$\lim_{r/\sigma\to\infty} \text{error} \;=\; \frac{1}{M}. \qquad (6.13)$$

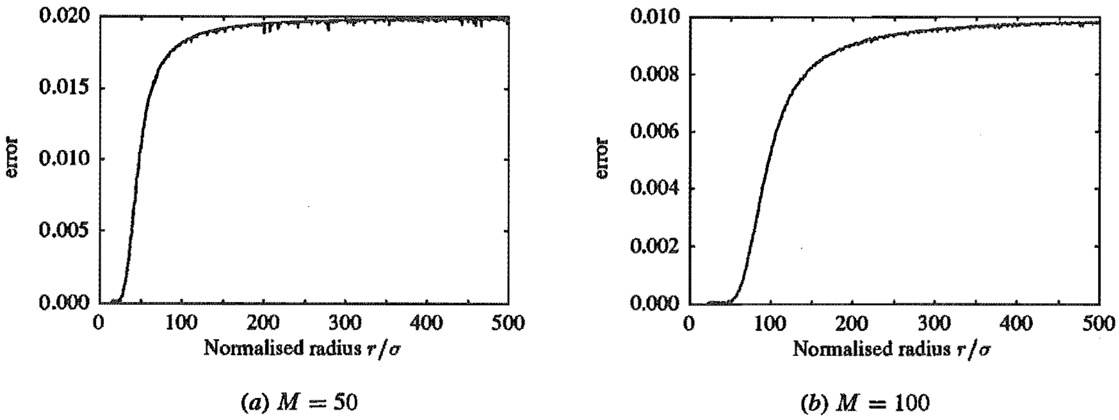
This was noted by Lewitt *et al.* [1978], and Rowland [1979] upon examining the point spread function for modified back-projection concludes that it contains angular and radial oscillations which depend on the choice of the projection sampling method, modifying function and interpolating function, but that the limiting value for the normalised error is independent of these and has the value given in (6.13). This error limit is attained only at a radius so great that the number of projections are completely insufficient to produce any meaningful image and is not of interest here. More pertinent is the reconstruction radius for which contamination begins to exceed an acceptable level, and the relationship this radius has to the number of projections. Accordingly, figure 6.3 presents plots of $\ln(R_\alpha)$, where R_α is the normalised radius above which the contamination from a single feature may be larger than a pre-defined level α, as a function of the logarithm of the number of projections $\ln(M)$. Thus, figure 6.3 permits one to estimate the reconstruction radius within which contamination from a single feature will remain below the specified α level, when M projections are given. It is also apparent from figure 6.3 that R_α has a log-linear relationship with the number of projections M :

$$\ln(R_\alpha) = r_\alpha + k_\alpha \ln(M), \tag{6.14}$$

where r_α is an intercept ($M = 1$) and k_α is the slope of a line fit. The relationship (6.14) allows direct calculation of R_α. However, as observed in figure 6.3, the value of R_α "blows up" as M approaches $1/\alpha$ since α becomes greater than the maximum error limit given by (6.13). Two such asymptotes for $M = 100$ and $M = 200$, respectively, are shown in figure 6.3. Smith *et al.* [1973] have presented, along a similar theme, a plot of reconstruction contamination, α, as a function of the number of projections and the normalised radius from the gaussian feature. The contamination has been evaluated over this 2-D space within the range of interest for the number of projections.

6.5 SIMULTANEOUS ITERATIVE REFINEMENT ALGORITHM

Herman [1988] proposed a specific algorithm for solution of the problem of section 6.2 subject to the constraint given in (6.6). This transformation is given here as an algorithm for the simultaneous iterative refinement (SIR) of all projection data. The specific problem for which this iterative algorithm is designed to solve is given here in matrix notation for later convenience :



(*a*) $M = 50$ (*b*) $M = 100$

Figure 6.2 Maximum normalised error or contamination in reconstructing a gaussian blob from different projection numbers M as a function of the normalised radius from its centre.

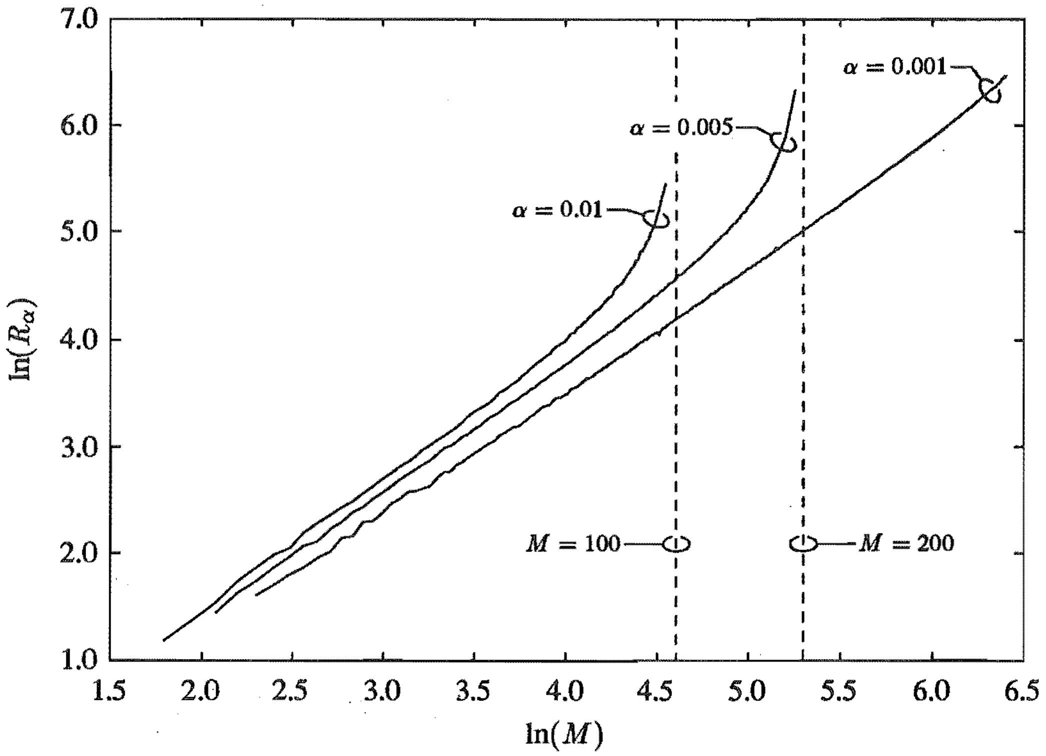


Figure 6.3 The reconstruction radius R_α at which the maximum normalised contamination from a single feature begins to exceed the specified level, α.

> Given the vector containing the discrete set of measurement data $\mathbf{P}_g$ and the reconstruction matrix M (refer to section 5.6.3.4), find a solution vector $\hat{\mathbf{P}}$, such that the norm $\left\| \mathbf{P}_g - \hat{\mathbf{P}} \right\|$ is minimised subject to negligible reconstruction within Ω_+, i.e., $\mathbf{M}\,\hat{\mathbf{P}} = \mathbf{0}$ in Ω_+.

As mentioned in section 5.6.3.4, the reconstruction matrix M for modified back-projection (MBP) consists of a modifying matrix H followed by a back-projection matrix $\mathbf{A}^T$ (equation (5.83)).

Herman [1988] proposed a quadratic optimisation approach to solve this specific problem, subject to the said constraint of minimising $\left\| \mathbf{P}_g - \hat{\mathbf{P}} \right\|$, based on an iterative method, called in numerical analysis a *Richardson's method* [Herman, 1980], for solving systems of linear equations. The initial estimate for $\hat{\mathbf{P}}$ is taken to be the given data $\mathbf{P}_g$ and Richardson's method is invoked to make subsequent estimates that simultaneously correct for imperfections in all the data :

$$\hat{\mathbf{P}}^0 = \mathbf{P}_g, \tag{6.15}$$

$$\hat{\mathbf{P}}^{k+1} = \hat{\mathbf{P}}^k - \lambda \left\{ \mathbf{M}^T \mathbf{M}_{\Omega_+} \right\} \hat{\mathbf{P}}^k, \tag{6.16}$$

where λ is a relaxation factor for the iterations and the subscript Ω_+ implies that the reconstruction matrix is confined to the region Ω_+ only. On examination of (6.16) it can be seen that at each iteration the current estimate $\hat{\mathbf{P}}^k$ is reconstructed in the region Ω_+ by multiplication with $\mathbf{M}_{\Omega_+}$. The coefficients of the matrix $\mathbf{M}_{\Omega_+}$ are formed from those of modifying and back-projection, $\mathbf{H}$ and $\mathbf{A}^T$ respectively (refer to (5.83)). Correction data are estimated from the reconstruction within Ω_+, using the same coefficients as used in the reconstruction process, by multiplication with $\mathbf{M}^T$. Since $\mathbf{M}^T = \mathbf{H}\,\mathbf{A}$, the correction data can be considered as calculating the forward projections of the region Ω_+ and then modifying them. Since the operation $\mathbf{M}^T$ makes use of the same standard computational components used in the reconstruction method, $\mathbf{M}$, the iterative algorithm in (6.16) is neatly expressed in a closed form. This approach minimises $\left\| \mathbf{P}_g - \hat{\mathbf{P}} \right\|$ if and only if

$$\left\{ \mathbf{M}^T\, \mathbf{M}_{\Omega_+} \right\} \hat{\mathbf{P}} = \mathbf{0}. \tag{6.17}$$

Clearly, (6.17) is satisfied if the reconstruction within Ω_+ is zero, i.e.,

$$\mathbf{M}_{\Omega_+}\, \hat{\mathbf{P}} = \mathbf{0}. \tag{6.18}$$

The modified forward projections of Ω_+ are then applied for correction of the current estimate $\hat{\mathbf{P}}^k$ via a relaxation factor λ. The iterative approach of (6.16) converges to a solution $\hat{\mathbf{P}}$ provided λ is chosen so that [Herman, 1980]

$$0 < \lambda < \frac{2}{\lambda_{max}\, \mathbf{M}^T\, \mathbf{M}_{\Omega_+}}, \tag{6.19}$$

where $\lambda_{max}\, \mathbf{M}^T\, \mathbf{M}_{\Omega_+}$ and $\lambda_{min}\, \mathbf{M}^T\, \mathbf{M}_{\Omega_+}$ denote, respectively, the largest and smallest eigenvalues of $\mathbf{M}^T\, \mathbf{M}_{\Omega_+}$. Furthermore an "optimal" value for λ is given by

$$\lambda = \frac{2}{\lambda_{max}\, \mathbf{M}^T\, \mathbf{M}_{\Omega_+} + \lambda_{min}\, \mathbf{M}^T\, \mathbf{M}_{\Omega_+}}. \tag{6.20}$$

Table 6.1 shows the iterative steps proposed in an algorithm by Herman [1988].

6.6 TEST PHANTOM

As intimated in section 6.1, in order to develop a proposed solution, such as that described in section 6.5, it is essential to begin with computer simulations, in order to evaluate its preliminary performance. This section describes a mathematically defined phantom that allows reconstructions to be compared with a precisely-known original. The nature of the phantom also allows precise calculation of the perfect projections.

Consider the phantom represented diagrammatically in figure 6.4. The phantom, chosen to approximate a cross-section of the human thorax, contains six bone structures (a vertebra,

Step 1	Initialise the iterate : $\hat{\mathbf{P}}^0 = \mathbf{P}_g.$
Step 2	Reconstruct $\hat{\mathbf{P}}^k$ within Ω_+ : $\hat{\mathbf{F}}_{\Omega_+} = \mathbf{M}_{\Omega_+} \hat{\mathbf{P}}^k.$
Step 3	Check the reconstruction within Ω_+ : If $\left\Vert \hat{\mathbf{F}}_{\Omega_+} \right\Vert < \epsilon$ set $\hat{\mathbf{P}} = \hat{\mathbf{P}}^k$ and stop iterations.
Step 4	Calculate a new estimate for the iterate : $\hat{\mathbf{P}}^{k+1} = \hat{\mathbf{P}}^k - \lambda \left\{ \mathbf{M}^T \mathbf{M}_{\Omega_+} \right\} \hat{\mathbf{P}}^k.$
Step 5	Increase the iteration index, $k \rightarrow k+1$ and goto step 2.

Table 6.1 Steps of the iterative algorithm proposed by Herman [1988]. Given $\mathbf{P}_g$, solve for the $\hat{\mathbf{P}}$ that minimises $\left\Vert \mathbf{P}_g - \hat{\mathbf{P}} \right\Vert$ subject to the constraint that the reconstruction within Ω_+, $\hat{\mathbf{F}}_{\Omega_+}$, is negligible. The reconstruction matrix for the region Ω_+ only is represented by $\mathbf{M}_{\Omega_+}$.

sternum and four ribs) and ten tumours superimposed onto a background of soft-tissue. The motivation for such a test phantom is to evaluate the algorithm of section 6.5, in the context of improving tumour detection, to improve the reconstruction of low-contrast tumours from a plethora of imperfect projections.

Each of the structures in the phantom are elliptical objects of constant X-ray attenuation coefficient. The precise projections through elliptic structures are simply calculated from a knowledge of their X-ray attenuation coefficient, major and minor axes, position and orientation of the major axis. These are shown in table 6.2 for the test phantom of figure 6.4. Note that the X-ray attenuation coefficient of each structure is that superimposed over the back-ground soft-tissue. Calculation of the full projection data is then the superposition of that calculated for the individual structures. Soft-tissue is assumed to have an X-ray attenuation coefficient of $0.200 cm^{-1}$, bone a total attenuation of $0.400 cm^{-1}$, while the low-tumours are assumed to have an X-ray attenuation 2.5 percent greater than that of the surrounding soft-tissue ($0.205 cm^{-1}$ in total). The tumours clearly consist of low-contrast material in the soft-tissue.

The ten tumours have been placed at random positions in the proximity of the lungs. Five out of ten potential tumour locations lie in the left lung while the other five lie in the right lung. Also shown in figure 6.4 are ten control locations symmetric in the y-axis from where tumours are known not to exist.

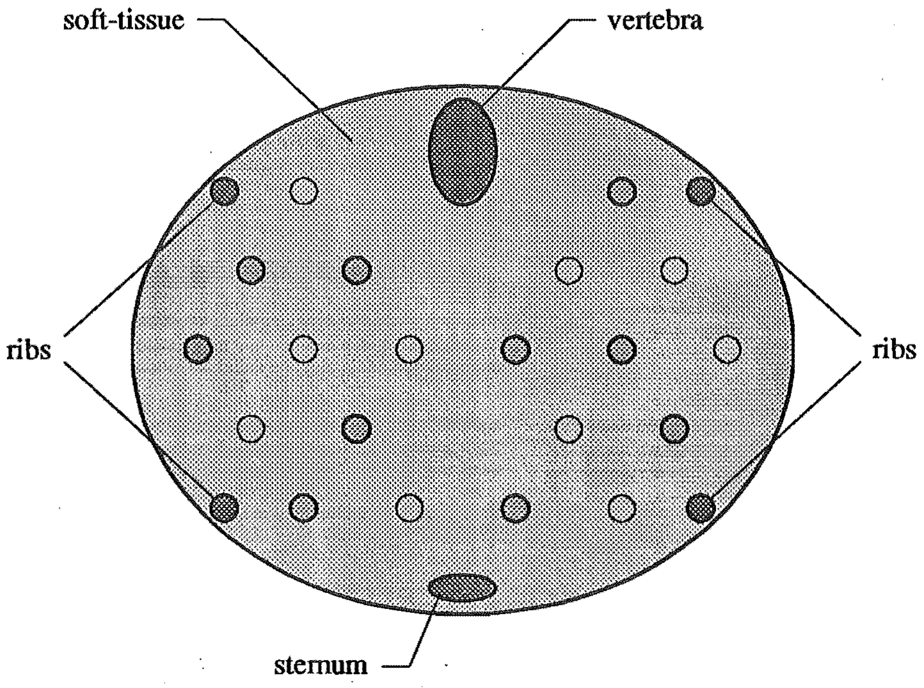


Figure 6.4 Test phantom.

The reconstructed image is composed of a 225 x 225 grid of pixels, with sides of length $0.5cm$. The radii of the circles circumscribing the regions Ω_- and Ω_+, a and b, respectively, have been chosen to be $28cm$ and $56cm$. The perfect projection data have been numerically collected for a parallel projection geometry (refer to section 5.3) using the mathematical description of the phantom in table 6.2. Spacing between projection samples on projections should be no larger than the smallest detail that is to be reconstructed. The sample spacing is chosen to equal the length of a pixel side, i.e., $\Delta\xi = 0.5cm$. The number of projection samples, per projection, is 227 to encompass the complete reconstruction grid. In accordance with (6.9), the number of views taken for the reconstruction, for a radius $b = 56cm$ and resolution $\sigma = 0.5cm$, is $M = 360$.

The reconstruction method chosen is the modified back-projection method described in section 5.6.3.3 using the discrete modifying filter of Shepp and Logan [1974], equation (5.69), and linear interpolation for the discrete back-projection process. The complete formulation for the reconstruction $\hat{f}(r;\theta)$ can be obtained by substituting the linear interpolation procedure of equation (5.30) into the discrete modified back-projection formulation of equation (5.66).

Clearly the reconstruction must be evaluated by some measure of "goodness", and since the motivation for this phantom was to evaluate an algorithm's ability to improve the detectability of tumours, the essence of the evaluation chosen is tumour contrast. Tumour contrast is measured by a *Hotelling trace* [Coggins, 1992], the separation between the classes of tumour reconstruction and the background reconstruction surrounding the tumour. The average reconstruction of the tumour, $\hat{f}^t$, is measured by the average of nine central pixels, $\hat{f}^t_j$ for $j = 1, \ldots 9$, contained within the tumour region (refer figure 6.5) :

$$\hat{f}^t = \frac{1}{9} \sum_{j=1}^{9} \hat{f}^t_j. \tag{6.21}$$

The average background reconstruction surrounding the tumour, $\hat{f}^b$, is measured by the average of eight similar constellations (involving nine pixels) surrounding this tumour :

$$\hat{f}^b = \frac{1}{8} \sum_{k=1}^{8} \hat{f}^b_k, \tag{6.22}$$

Structure	Attenuation cm^{-1}	Position (x, y)	Axes (minor,major)	Orientation (degrees)
Soft-tissue	0.200	(0,0)	(40,50)	0
Vertebra	0.200	(0,15)	(5,8)	90
Sternum	0.200	(0,-18)	(2,5)	0
Rib	0.200	(18,12)	(2,2)	0
Rib	0.200	(-18,12)	(2,2)	0
Rib	0.200	(18,-12)	(2,2)	0
Rib	0.200	(-18,-12)	(2,2)	0
Tumour 1	0.005	(-20,0)	(2,2)	0
Tumour 2	0.005	(-16,6)	(2,2)	0
Tumour 3	0.005	(-12,-12)	(2,2)	0
Tumour 4	0.005	(-8,-6)	(2,2)	0
Tumour 5	0.005	(-8,6)	(2,2)	0
Tumour 6	0.005	(4,-12)	(2,2)	0
Tumour 7	0.005	(4,0)	(2,2)	0
Tumour 8	0.005	(12,0)	(2,2)	0
Tumour 9	0.005	(12,12)	(2,2)	0
Tumour 10	0.005	(16,-6)	(2,2)	0

Table 6.2 Properties of the structures contained within the phantom of figure 6.4. The phantom is composed of six bone structures (a vertebra, sternum and four ribs) and ten tumours that are superimposed onto soft-tissue. Each structure is modelled by an ellipse of constant X-ray attenuation whose orientation is specified as the angle the major axis makes with the x-axis.

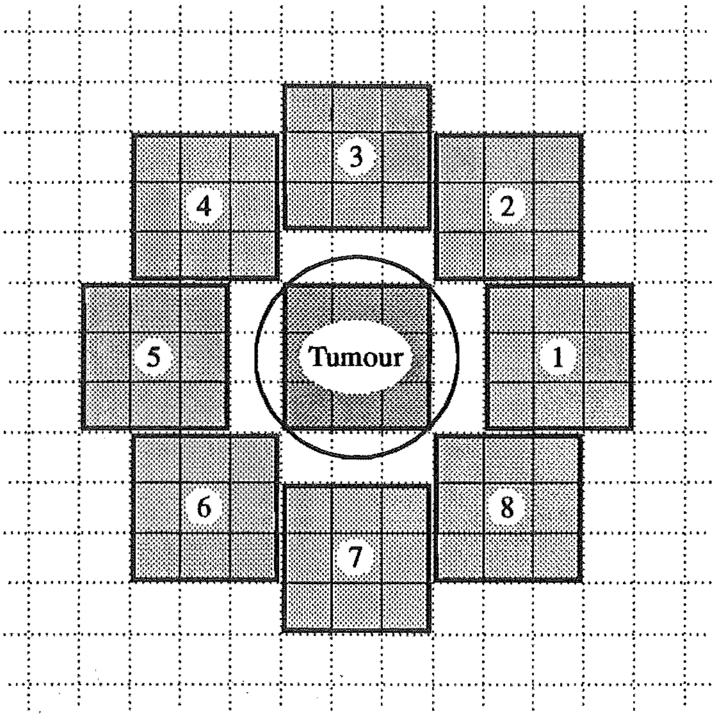


Figure 6.5 A tumour region containing nine pixel elements surrounded by eight similar background configurations.

where in accordance with (6.21), if a background pixel of the k^{th} constellation is $\hat{f}^b_{j,k}$, then

$$\hat{f}^b_k = \frac{1}{9}\sum_{j=1}^{9} \hat{f}^b_{j,k}. \tag{6.23}$$

The deviation of the background reconstruction is measured between these eight constellations,

$$\sigma^b = \frac{1}{8}\sqrt{\sum_{k=1}^{8}\left(\hat{f}^b_k - \hat{f}^b\right)^2}, \tag{6.24}$$

and the tumour contrast

$$c^t = \frac{\hat{f}^t - \hat{f}^b}{\sigma^b}. \tag{6.25}$$

At the ten tumour locations the contrast of each tumour is defined through (6.25) as c^t_i (for $1 \leq i \leq 10$). Similarly, the control set of non-tumours located at positions symmetric in the y-axis have contrast c^n_i (for $1 \leq i \leq 10$). The average tumour and non-tumour contrasts are, respectively,

$$\bar{c}^t = \frac{1}{10} \sum_{i=1}^{10} c_i^t \tag{6.26}$$

and
$$\bar{c}^n = \frac{1}{10} \sum_{i=1}^{10} c_i^n. \tag{6.27}$$

Figure 6.6 shows the reconstruction from perfect projections of the phantom. For the purpose of displaying the reconstruction of the low contrast tumours, the image has been thresholded and quantised to discrete integer levels within the range $[0, 255]$. Attenuation levels equal to or below $0.190cm^{-1}$ and equal to or above $0.210cm^{-1}$ have been assigned to 0 and 255 respectively. Table 6.3 shows contrasts calculated from the reconstruction for both tumour and non-tumour locations using (6.25), the sum of the squares of the reconstruction within the region Ω_+ and the maximum absolute reconstruction within Ω_+.

The reconstruction in figure 6.6 shows streaking effects from sharp bodies within the image due to reconstruction from a finite number of projections. Since the phantom is constructed from sharp bodies, contributing significantly to attenuation value, the perfect projections are highly discontinuous and thereby exceed the maximum radial extent of Fourier components assumed by equation (6.9). There is therefore an insufficient number of projections available to reconstruct the detail of sharp discontinuities. It is also apparent from table 6.3 that the reconstruction within Ω_+ is significant.

Rowland [1979] has investigated these streaking effects, or overshoot, as it is termed in section 5.6.3.3, and their dependence on the type of modifying filter when using linear interpolation in the back-projection process. Four window functions were examined. Of greatest interest is the ideal low-pass filter of (5.67), a modification of the filter given by Shepp and Logan [1974], and a generalised hamming low-pass window such as that given in (5.71). Rowland [1979] concluded that the ideal low-pass filter produced the best resolution, the generalised hamming low-pass window the least overshoot and the filter based on that of Shepp and Logan [1974] produced the best combined resolution and overshoot characteristics.

It was decided to adopt the modifying filter of Shepp and Logan [1974] for experimental work reported in this chapter so that resolution is not severely sacrificed. However, to reduce overshoot the detector (or sensor) response was modelled in a more realistic manner. In this way it was felt the simulated data would better represent those measured in practice while maintaining to a reasonable degree of the resolution that can be obtained from measured data. As mentioned in section 5.5, the detector, in practice, has finite width. The detector response can be modelled by convolving the perfect projections with the point spread function (psf) of the detector, as shown in (5.35), giving blurred perfect projections. Tan *et al.* [1986] suggest the following gaussian psf, $\imath(\xi)$, sufficiently characterises the main finite resolution effects of detectors :

$$\imath(\xi) = e^{-\xi^2/\gamma^2}, \tag{6.28}$$

where γ is known as the "effective width" of $\imath(\xi)$. It should be noted here that to evaluate the

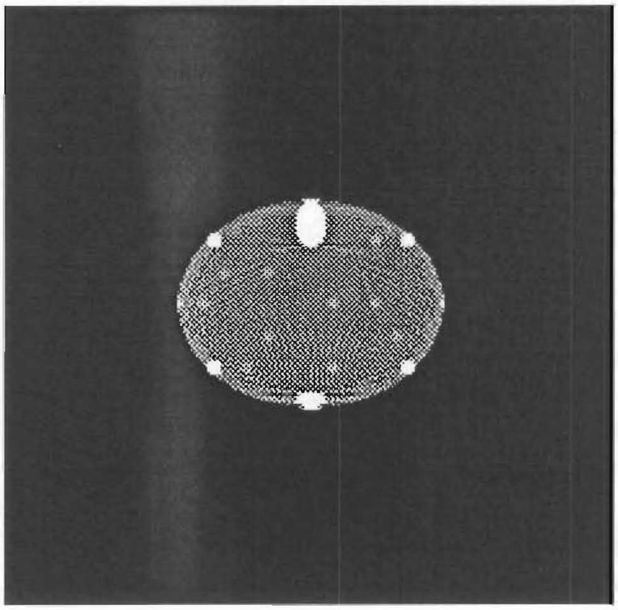


Figure 6.6 Reconstruction of the phantom, figure 6.4, on a 225 x 225 grid of square pixels with side length $0.5cm$ from 360 perfect projections with a sample spacing of $\Delta\xi = 0.5cm$.

Location number	Contrast tumour	Contrast non-tumour
1	60.2	3.4
2	118.2	0.9
3	200.1	7.3
4	107.9	-2.3
5	37.9	3.8
6	124.3	-6.4
7	52.0	-2.9
8	47.2	-13.6
9	80.2	-4.2
10	31.1	3.3
Mean Contrast	85.9	-1.1

Sum of squares of the reconstruction within Ω_+	0.36×10^{-1}
Maximum absolute reconstruction within Ω_+	0.77×10^{-2}

Table 6.3 Results of tumour reconstruction calculated on the reconstructed image, from perfect projections, shown in figure 6.6.

blurred perfect projections at the points ξ_n by convolution of $i(\xi)$ with $p(\xi; \phi_m)$, the perfect projections must be evaluated at regular points within the detector spacings so that a numerical integration technique (such as a Riemann sum) can be invoked to provide an accurate estimate of the blurred projections.

Figure 6.7 shows the reconstruction from the blurred perfect projections and table 6.4 evaluation of the reconstruction of each tumour. It can be seen by comparing table 6.4 to table 6.3 that the overshoot has been reduced to a much lower level by modelling the detector response with (6.28). The sum of squares of the reconstruction in Ω_+, shown in table 6.4, indicates a reduction (from the levels of table 6.3) to a level closer to what would be expected and indicates that there are now sufficient projections to reconstruct Ω_+. Tumour contrast has increased and the mean contrast nearly doubled. It should be noted that the sum of squares of the reconstruction within Ω_+, and the maximum absolute reconstruction within Ω_+, given in table 6.4, are valid estimates for reconstruction levels that can be termed *negligible* since these are the levels observed when reconstructing from an adequate number of perfect projections. That is to say, the maximum absolute reconstruction within Ω_+ gives an indication of ϵ in constraint (6.4).

6.7 PROJECTION NOISE

The ability of the simultaneous iterative refinement algorithm suggested in section 6.5 to correct for non-linear noise on the projections was evaluated. The noise was modelled to simulate that generated from an X-ray source emitting a mean number of X-ray photons equal to ζ within any measurement interval. Basic limitations to the accuracy of X-ray measurements are the statistical natures of X-ray photon production, photon interaction with attenuating material, and photon detection. At this stage the statistical nature of the interaction of photons with attenuating material and the statistical nature of the detector is not considered, in the manner of Herman [1980]. It is here assumed that all photons reaching the detector are counted and that if I_0 photons are generated from a monochromatic source over the measurement interval, then the number I reaching the detector, through the projection path L with coordinates $(\xi; \phi)$ (refer figure 5.1), is governed by the relationship given in (1.9) :

$$\begin{aligned} I &= I_0 \, e^{-\int_L \mu(s)\, ds} \\ &= I_0 \, e^{-p(\xi;\phi)}, \end{aligned} \tag{6.29}$$

where μ is the X-ray linear attenuation distribution of the material and $p(\xi; \phi)$ is referred to as the perfect projection of the X-ray linear attenuation, at a given energy for the X-ray photons, along the path defined by $(\xi; \phi)$.

Following Herman [1980], the probability of I_0 photons generated by an X-ray source can be modelled as a discrete random variable having a Poisson probability distribution with mean ζ and standard deviation $\sqrt{\zeta}$:

$$P(I_0) = \frac{e^{-\zeta} \, \zeta^{I_0}}{I_0!}. \tag{6.30}$$

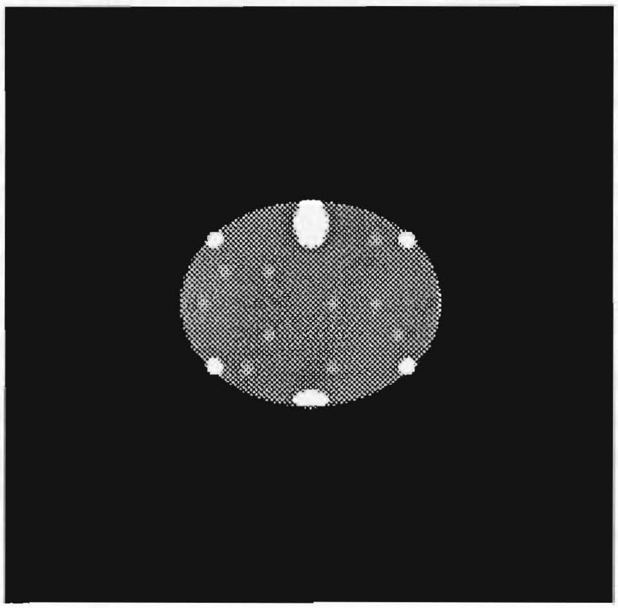


Figure 6.7 Reconstruction of the phantom, figure 6.4, on a 225 x 225 grid of square pixels with side length $0.5cm$ from 360 blurred perfect projections with a sample spacing of $\Delta\xi = 0.5cm$.

Location number	Contrast tumour	Contrast non-tumour
1	325.9	8.4
2	91.6	0.5
3	269.1	2.7
4	154.1	2.2
5	74.4	6.2
6	166.8	-0.4
7	154.8	0.2
8	60.8	-10.8
9	166.0	-3.5
10	153.9	4.0
Mean Contrast	161.8	1.0

Sum of squares of the reconstruction within Ω_+	0.63×10^{-3}
Maximum absolute reconstruction within Ω_+	0.66×10^{-3}

Table 6.4 Results of tumour reconstruction calculated on the reconstructed image, from blurred perfect projections, shown in figure 6.7.

This probability distribution approximates that of a normal distribution with mean ζ and standard deviation $\sqrt{\zeta}$ provided ζ is large (e.g., $\zeta > 100$). The number of photons generated from the source, when this number is large, can therefore be modelled as

$$I_0 = \mathcal{N}\left(\zeta, \sqrt{\zeta}\right), \tag{6.31}$$

where $\mathcal{N}(\zeta, \sqrt{\zeta})$ represents here a normal distribution with mean ζ and standard deviation $\sqrt{\zeta}$. Since the number of photons reaching the detector are directly related to the number generated through (6.29), the statistical nature of photons reaching the detector can be incorporated into the interaction with the attenuating material. It is transparent to the detector whether the statistical nature is due to the source or due to the interaction of photons with the intervening material. Hence if the source is assumed to have a constant output $I_0 = \zeta$ of photons over a given measurement interval, then the number I of photons reaching the detector can be modelled by a normal distribution with mean $I_0 e^{-p(\xi;\phi)}$ and standard deviation $\sqrt{e^{-p(\xi;\phi)}/I_0}$, if the mean number out is sufficiently large, i.e.,

$$\begin{aligned} I &= \mathcal{N}\left(I_0 e^{-p(\xi;\phi)}, \sqrt{I_0 e^{-p(\xi;\phi)}} \right) & (6.32) \\ &= I_0 \, \mathcal{N}\left(e^{-p(\xi;\phi)}, \sqrt{e^{-p(\xi;\phi)}/I_0} \right), & (6.33) \end{aligned}$$

and the attenuation as seen from the detector is

$$\begin{aligned} \frac{I}{I_0} &= \mathcal{N}\left(e^{-p(\xi;\phi)}, \sqrt{e^{-p(\xi;\phi)}/I_0} \right) & (6.34) \\ &= e^{-p(\xi;\phi)} + \mathcal{N}\left(0, \sqrt{e^{-p(\xi;\phi)}/I_0} \right) & (6.35) \\ &= e^{-p(\xi;\phi)} + \mathcal{N}(0,1) \sqrt{\frac{e^{-p(\xi;\phi)}}{I_0}}. & (6.36) \end{aligned}$$

Equation (6.36) gives the attenuation of photons as seen from the detector when I_0 photons are generated at the source. The projections measured by the detector in terms of the perfect projections, and incorporating the statistical nature of the source are

$$p_{\mathcal{N}}(\xi;\phi) = -\ln\left[e^{-p(\xi;\phi)} + \mathcal{N}(0,1) \sqrt{\frac{e^{-p(\xi;\phi)}}{I_0}} \right]. \tag{6.37}$$

It can be seen from (6.37) that the closeness of measured projections to the perfect projections $p(\xi;\phi)$ is limited by the statistical nature of the source (as mentioned above).

6.8 RESULTS OF EXPERIMENTS USING THE SIMULTANEOUS ITERATIVE REFINEMENT ALGORITHM

This section reports the results of experiments in which the algorithm, detailed in section 6.5, was applied to improve the "detectability" of tumours reconstructed from projections which contain imperfections of the type described in section 6.7. The "detectability" of tumours was determined from contrast calculations as detailed in section 6.6. Two noisy projection sets, modelling the statistical nature of the source, were generated from the "perfect" (in the sense of section 6.6 where the detector response has been modelled) projections and used for reconstruction of the test phantom specified in table 6.2. Low- and high-noise projection sets were generated to model the statistical nature of the source when the mean number of photons generated over a measurement interval is $\zeta = 10^8$ and $\zeta = 10^7$ respectively. The low- and high-noise projection data sets were, respectively, from (6.37)

$$l(\xi;\phi) = e^{-p(\xi;\phi)} + \mathcal{N}(0,1)\sqrt{\frac{e^{-p(\xi;\phi)}}{10^8}} \tag{6.38}$$

and

$$h(\xi;\phi) = e^{-p(\xi;\phi)} + \mathcal{N}(0,1)\sqrt{\frac{e^{-p(\xi;\phi)}}{10^7}}, \tag{6.39}$$

and where the mean number of photons output from the X-ray attenuation phantom of section 6.6 is always greater than 100. Figure 6.8 and table 6.5 show the reconstruction from the low-noise projections and the tumour reconstruction evaluations respectively, figure 6.9 and table 6.6 show those for the high-noise projections.

The algorithm of section 6.5 was applied to both noisy projection sets. Values for the sum of squares of the reconstruction within Ω_+ and the maximum absolute reconstruction within Ω_+ determined the termination criteria for the iterations described in table 6.1. Iterations were stopped once these values were exceeded by those obtained from the perfect projections (refer table 6.4). The maximum eigenvalue $\lambda_{max}\,\mathrm{M}^T\,\mathrm{M}_{\Omega_+}$ of the matrix $\mathrm{M}^T\,\mathrm{M}_{\Omega_+}$, used in (6.19) to estimate the relaxation factor for the iterative procedure, is dependent upon the projection and modifying matrices. Therefore the eigenvalue is independent of the projection data and need only be estimated once for a given reconstruction matrix. The maximum eigenvalue estimated for the reconstruction method described in section 6.6 is $\lambda_{max}\,\mathrm{M}^T\,\mathrm{M}_{\Omega_+} = 0.54$. Using (6.19) as a guideline for the relaxation factor, $\lambda = 2.0$ was chosen for correction of the noisy projection sets in (6.38) and (6.39).

Figure 6.10 and table 6.7 show the reconstruction and tumour reconstruction evaluations, respectively, from projections obtained by correcting the low-noise projections, given in (6.38), in the above manner. Figure 6.11 and table 6.8 show those results obtained from correction of the high-noise projections given in (6.39).

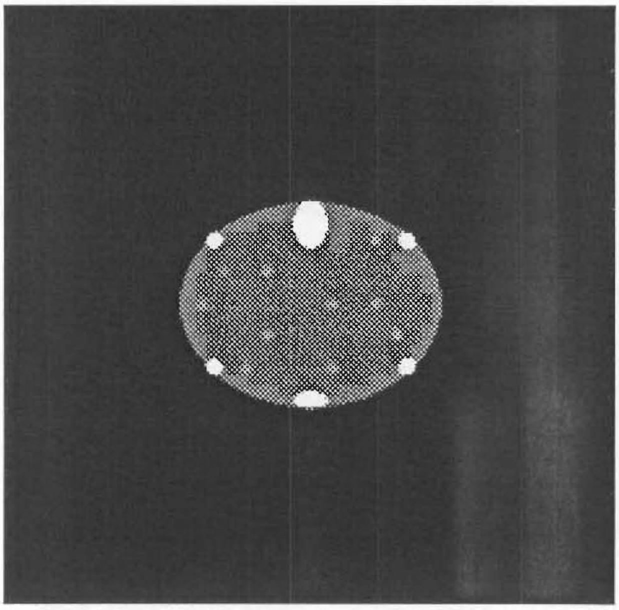


Figure 6.8 Reconstruction of the phantom, figure 6.4, on a 225 x 225 grid of square pixels with side length $0.5cm$ from 360 noisy projections using a source delivering a mean 10^8 photons per measurement interval. Projection sample spacing is $\Delta\xi = 0.5cm$.

Location number	Contrast tumour	Contrast non-tumour
1	98.7	8.9
2	77.5	2.5
3	77.1	6.2
4	126.4	1.4
5	45.4	2.7
6	58.8	1.6
7	75.6	-1.6
8	67.3	-10.2
9	134.4	2.9
10	88.1	6.1
Mean Contrast	84.9	2.0

Sum of squares of the reconstruction within Ω_+	0.19×10^{-2}
Maximum absolute reconstruction within Ω_+	0.11×10^{-2}

Table 6.5 Results of tumour reconstruction calculated on the reconstructed image, from projections taken with a source delivering a mean of 10^8 photons per measurement, shown in figure 6.8.

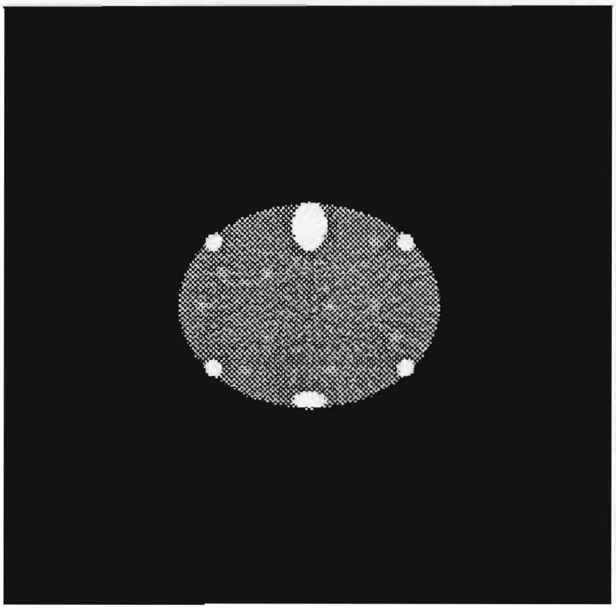


Figure 6.9 Reconstruction of the phantom, figure 6.4, on a 225 x 225 grid of square pixels with side length $0.5cm$ from 360 noisy projections using a source delivering a mean 10^7 photons per measurement interval. Projection sample spacing is $\Delta\xi = 0.5cm$.

Location number	Contrast tumour	Contrast non-tumour
1	28.5	8.6
2	19.4	-1.5
3	13.2	0.6
4	19.9	3.4
5	14.6	-0.3
6	27.1	2.6
7	20.1	2.8
8	15.6	-2.2
9	22.7	-0.1
10	34.1	2.3
Mean Contrast	21.5	1.6

Sum of squares of the reconstruction within Ω_+	0.14×10^{-1}
Maximum absolute reconstruction within Ω_+	0.34×10^{-2}

Table 6.6 Results of tumour reconstruction calculated on the reconstructed image, from projections taken with a source delivering a mean of 10^7 photons per measurement, shown in figure 6.8.

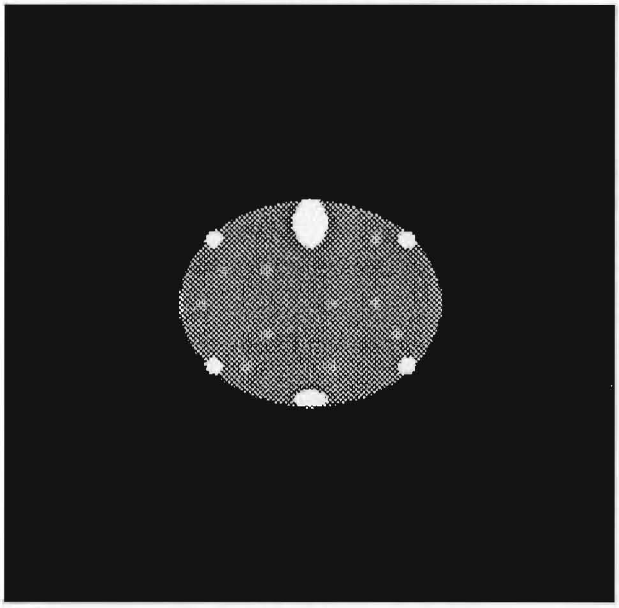


Figure 6.10 Reconstruction of the phantom, figure 6.4, from the corrected low-noise projections after 1 iteration.

Location number	Contrast tumour	Contrast non-tumour
1	99.7	8.6
2	78.6	2.4
3	74.4	6.3
4	125.0	1.3
5	45.8	2.8
6	60.1	1.5
7	75.5	-1.8
8	67.4	-10.2
9	137.8	2.8
10	89.2	5.9
Mean Contrast	85.4	1.9

Sum of squares of the reconstruction within Ω_+	0.46×10^{-3}
Maximum absolute reconstruction within Ω_+	0.62×10^{-3}

Table 6.7 Results of tumour reconstruction calculated on the reconstructed image shown in figure 6.10, from the corrected low-noise projections after 1 iteration.

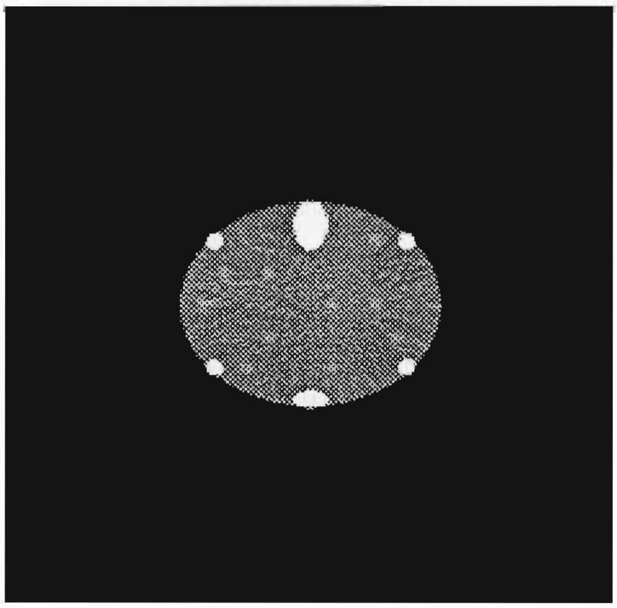


Figure 6.11 Reconstruction of the phantom, figure 6.4, from the corrected high-noise projections after 20 iterations.

Location number	Contrast tumour	Contrast non-tumour
1	29.9	9.6
2	19.8	-1.5
3	13.1	0.5
4	20.1	3.7
5	14.7	-0.3
6	28.4	2.5
7	19.8	2.8
8	15.7	-2.4
9	22.1	0.1
10	33.3	2.4
Mean Contrast	21.7	1.7

Sum of squares of the reconstruction within Ω_+	0.31×10^{-3}
Maximum absolute reconstruction within Ω_+	0.66×10^{-3}

Table 6.8 Results of tumour reconstruction calculated on the reconstructed image shown in figure 6.11, from the corrected high-noise projections after 20 iterations.

6.9 DISCUSSION : LIMITATIONS OF THE SIMULTANEOUS ITERATIVE REFINEMENT ALGORITHM

Upon comparing the reconstruction levels within Ω_+ from the perfect projections with the reconstruction from both noisy projection sets, it can be seen that there is an increase attributable solely to the imperfections. This indicates that there is in fact information within Ω_+ which may allow correction of the imperfections. It is also apparent from comparisons made of the reconstruction from blurred perfect projections, figure 6.7, with those from the noisy projections sets, figures 6.8 and 6.9, that the quality of the reconstruction with respect to the tumours has diminished. This result is reflected in the reduction of tumour contrast for these reconstructions, shown in tables 6.5 and 6.6. In both cases the imperfections in the projection data were responsible for alterations to the reconstruction within the known support, Ω_-. However, there has been no significant change in the contrast of the control tumours at non-tumour locations.

Upon applying iterative correction to the low-noise projection set with a relaxation factor $\lambda = 2.0$, after only 1 iteration the termination criteria was satisfied. The algorithm was able to adjust the projection data so that the reconstruction within Ω_+ was negligible (according to the criteria given in section 6.8). However, as is apparent from figure 6.10 and table 6.7, application of this algorithm to this projection data has not brought about any marked improvement in the contrast (i.e., detectability) of the tumours. A similar result was obtained from the application of this algorithm to the high-noise projection set. Figure 6.11 and table 6.8 show the reconstruction and reconstruction results from the corrected projections after 20 iterations.

It should be noted that for the correction of both noisy projection sets, iterations were continued beyond the minimum number required to meet the termination criteria and the solution remained convergent.

6.10 SUMMARY

Section 6.1 introduced the plethora of views idea for correction of projection data measured from wave motion that can be accurately represented as concourses of rays. The idea is based on the facility of extra projections which allow the reconstruction region to be extended past the known support of the distribution of interest, and the ability to take advantage of any reconstruction within the extended region for the correction of imperfections. This idea led to the proposal of an open problem in section 6.2 for the correction of projection data measured from wave motions that can be accurately represented by concourses of rays. The remainder of the chapter examined the problem in the context of rays whose paths are straight. Section 6.3 unifies some of the fundamental theory, on which this problem is based, for the finite number of projections required to reconstruct an image within a region of known radius.

Section 6.4 examined the reconstruction contamination levels resulting from the reconstruction of a single feature from a finite number of projections. A linear relationship is detailed that provides an estimate of the normalised radius, from any single feature, for which contamination levels from the reconstruction are below a required level, given a finite number of projections.

In section 6.6 it was noted that the reconstruction from a finite number of perfect projections of a sharp body, such as the phantom introduced in the section, produced considerable overshoot within the reconstruction region. The formulations given by Crowther *et al.* [1970] and Bates and McDonnell [1989], equations (6.7) and (6.9) respectively, in terms of the required resolution σ are valid if the projections are bandlimited to a radial limit of $\rho_{max} = 1/\sigma$. It is often the case that the required resolution is specified in terms of the minimum size of pixels used to represent the reconstructed image. If so, then the spectral content of the projections should be limited according to this resolution to avoid violation of the assumed bandlimit constraint. This bandlimit is most often incorporated into the window used for the modifying function (refer section 5.6.3.3). As intimated in section 6.6, from the type of window used for the modifying function there is a trade-off between the resolution obtained in the reconstruction and the amount of overshoot that can occur. To ensure that overshoot is not excessive for sharp bodies, thereby ensuring that non-negligible levels of reconstruction within Ω_+ are due to imperfections in the projection data, an appropriate window that reduces overshoot, such as the generalised hamming window, should be applied as a windowing function.

A simultaneous iterative refinement algorithm has been given in section 6.5 and applied for the correction of noisy projection data, containing non-linear imperfections, in the manner outlined in table 6.1. This algorithm utilises the reconstruction contributions within Ω_+. The reconstruction within the region Ω_+ has been established as contributions that are attributed to the imperfections. From this it can be deduced that there is information contained in Ω_+, for the types of non-linear imperfections given in section 6.7, that may be utilised to ameliorate the imperfections. At the same time these imperfections have seriously degraded the reconstruction (in the sense of section 6.6) within the known support region of the distribution of interest, Ω_-. The algorithm has been shown to converge when reconstructing projections of the phantom described in section 6.6 containing non-linear distortions of the type given in section 6.7. The solutions to which it converged have rendered the reconstruction within Ω_+ negligible. In this sense the primary aims of the algorithm have been satisfied as the algorithm is able to successfully correct the projection data to zero reconstruction in Ω_+. However, the projection data has been adjusted in a manner that does not relate well to the original imperfections. This is apparent from the lack of improvement in the quality of the reconstruction obtained within the region Ω_-.

CHAPTER 7

CONSISTENCY-BASED RECOVERY OF PROJECTION IMPERFECTIONS

This chapter follows the notation and terminology for the *plethora of views* idea established in chapter 6. *Consistency* of projections is examined in section 7.1 and in section 7.2 this definition is extended to incorporate a region with which the projections are consistent. The new definition, termed *spatial consistency*, in conjunction with a definition of *spatial inconsistency* forms the basis for an approach to recovering inconsistencies within projection data, the motivation being that if the inconsistencies can be recovered effectively, they may be compensated for in the projection data, thereby improving the reconstructed image within the known region of its support. The approach for recovering inconsistencies in projection data is detailed in section 7.4 along with the results of its application to inconsistencies contained on projections within a limited range of angles.

7.1 CONSISTENCY CONSIDERATIONS

As noted in section 6.9, application of the algorithm defined in section 6.5 adjusts the imperfect projection data on the noisy data sets, defined in equations (6.38) and (6.39), so that the reconstruction within Ω_+ is negligible. However, this algorithm has little effect on the reconstruction within the support region Ω_- and subsequently, there was little change observed in contrast at the tumour and non-tumour locations in the image.

It is here argued that the manner of correction outlined in section 6.5 does not directly relate the reconstruction contamination to the imperfections contained in the projection data, but does so indirectly and without regard to the physical reconstruction process introducing the contamination. For these reasons, interpretation of the results is made difficult by the lack of intuitive knowledge linking the correction process and the imperfections. Surely a more intuitive correction process would facilitate, from the point of view of interpreting results at least, the correction of the imperfections. The latter reasoning provided the motivation for investigating an alternative correction strategy.

A number of correction schemes have been suggested, based on ensuring the *consistency* of the projections, that utilise the difference $p_d(\xi;\phi)$ between the given projections $p_g(\xi;\phi)$ and those of the forward projections $p_f(\xi;\phi)$ calculated from a reconstruction (refer to section 5.5 for an explanation of forward-projection) [Orun-Ozturk *et al.*, 1987; Oskoui-Fard and Stark, 1988; Watt and Vest, 1989]. The difference projections $p_d(\xi;\phi)$ are used in an iterative manner for correction of the given data. However, the difference projections are not necessarily a true representation of the shape of the inconsistencies, as will become apparent in section 7.4.2. Although the philosophy of these methods implies that the solution for the projections is a *consistent* one, it is argued that these schemes also correct the projections in a manner that is not necessarily related well to the original inconsistencies, as was the case for the algorithm developed and tested in section 6.5 through to section 6.9.

Section 7.2 examines the relationship between the projections and the region from whence they came. That is, if the projection data is inverted to give a reconstruction, then ideally there should be no reconstruction outside the support region unless the data contains imperfections. This raises the question of whether of not projection data are *consistent* in some sense with a given region.

7.2 INCONSISTENCIES ON THE PROJECTIONS

The foundation for the developments reported within this section is the well-established idea [Herman, 1980; Louis, 1981; Radon, 1917] that the perfect projections $p(\xi;\phi)$ uniquely determine $f(r;\theta)$ if known for all $(\xi;\phi)$. In terms of a finite number of projection data, if sufficient of these data are subject to a "faithful" reconstruction process producing the reconstructed image $\hat{f}(r;\theta)$, then, $\hat{f}(r;\theta)$ is indistinguishable from $f(r;\theta)$ in terms of the desired resolution, as was discussed in section 6.3.

As mentioned in section 5.6.6, *consistency* of the projections was used as a basis for the first Algebraic Reconstruction Techniques [Bender *et al.*, 1970; Gordon *et al.*, 1970] where consistency of the projections is attained when

$$p_f(\xi;\phi) = p_g(\xi;\phi), \tag{7.1}$$

and where it is understood that $p_g(\xi;\phi)$ are the projections used to form $\hat{f}(r;\theta)$ and the forward projections $p_f(\xi;\phi)$ were defined in section 5.5. For the region $\Omega_+ \cup \Omega_-$ (refer figure 6.1) the $p_f(\xi;\phi)$ are :

$$p_f(\xi;\phi) = \int_{\Omega_+ \cup \Omega_-} \hat{f}(r;\theta)\, d\eta. \tag{7.2}$$

Other definitions of consistency have been given [Bates and McDonnell, 1989; Hanson, 1982; Smith *et al.*, 1973; Watt and Vest, 1989]. Smith *et al.* [1973] define a consistency relationship based on constraining the reconstruction $\hat{f}(r;\theta)$ to the known support Ω_-;

$$\hat{f}(r;\theta) = 0 \quad \text{for} \quad (r;\theta) \notin \Omega_-. \tag{7.3}$$

To facilitate further development, it is here found useful to create a definition of consistency, denoted *spatial consistency*, which is a combination of both (7.1) and (7.3), associating projection data with a *region* with which they are consistent.

> Projection data $p_g(\xi;\phi)$ are *spatially consistent* with a region Ω_- if, when subjected to a reconstruction process that produces a "faithful" reconstruction throughout Ω,

> all contributions are contained solely within Ω_-. The distribution $\hat{f}(r;\theta)$ is then confined to Ω_- and the forward projections of the reconstruction are consistent with the projection data (i.e., (7.1) holds).

As a further consequence of this definition it is possible to define regions with which projection data are spatially inconsistent.

> Projection data are *spatially inconsistent* with a region Ω_- if, when subjected to a reconstruction process that produces a faithful image throughout Ω, there are contributions outside Ω_-. The distribution $\hat{f}(r;\theta)$ is not confined to Ω_- and the forward projections of Ω_- are not consistent with the given data (i.e., (7.1) is violated).

Since in practice there are only ever data which are finite in number available for any reconstruction process, so that a faithful reconstruction cannot be obtained for all Ω (for reasons described in sections 6.3 and 6.4), it seems more useful, then, to develop a definition of *spatial inconsistency* which is meaningful in the context of reconstruction from a finite amount of data.

> Projection data are *spatially inconsistent* with a region Ω_- if, when subjected to a reconstruction process that produces a faithful image throughout Ω_-, the contributions within Ω_- do not form an image whose forward projections are consistent with the given data (i.e., (7.1) is violated).

This definition includes a reconstruction that has no contributions within Ω_-.

The above definition of spatial consistency relies on the foundations outlined in the opening paragraph to this section. (It should be noted, that there are distributions that can be contrived in such a way that in taking a finite number of projections, the projection data are all zero. Such distributions, referred to as "Ghosts" by Louis [1981] can exist outside Ω_-.)

7.3 RESTATEMENT OF THE PROBLEM

With the above terminology defined it is now possible to describe the given (measured) projection data. A plethora of imperfect projections $p_g(\xi;\phi)$ are given to reconstruct an unknown distribution $f(r;\theta)$ contained within the region Ω_- of radius a. The given projections may be considered as comprising two components :

$$p_g(\xi;\phi) = p'(\xi;\phi) + e(\xi;\phi), \tag{7.4}$$

where $p'(\xi;\phi)$ are projections spatially consistent with the region Ω_- and $e(\xi;\phi)$ are contributions to the given projections that arise due to imperfections, from measurements perhaps, and are spatially inconsistent with Ω_-. The projection data $p'(\xi;\phi)$ may vary inexplicably from $p(\xi;\phi)$ as a result of imperfections which give rise to projection contributions that are spatially consistent with the region of interest Ω_-. These variations $p'(\xi;\phi) - p(\xi;\phi)$ are inseparable from $p'(\xi;\phi)$ without other *a priori* knowledge about $f(r;\theta)$ since they comply with the only *a priori* information known of $f(r;\theta)$, its support. It is the conjecture here that there are few cases in practice whereby the imperfections form a perfect reconstruction in Ω_- or are spatially consistent with this region and which are a significant part of the total distortion. As a consequence of this conjecture, even though the $p'(\xi;\phi)$ are inseparable from $p(\xi;\phi)$, it is assumed that the $p'(\xi;\phi)$ are equal to the perfect projections $p(\xi;\phi)$ of $f(r;\theta)$, so that

$$p_g(\xi;\phi) = p(\xi;\phi) + e(\xi;\phi). \tag{7.5}$$

Any attempt to ameliorate the effects of imperfections, based on a knowledge of the support region, can only hope to separate the inconsistent projection data $e(\xi;\phi)$ from the given projections $p_g(\xi;\phi)$.

7.4 ALTERNATIVE APPROACH TO RECOVERING INCONSISTENCIES

It is apparent from equations (5.11) and (5.26) that the Radon transform of the distribution $\hat{f}(r;\theta)$ (known as the forward projections $p_f(\xi;\phi)$) obtained from a "faithful" reconstruction method is equal to the perfect projections $p(\xi;\phi)$ of the actual distribution $f(r;\theta)$, and any deviation of $p_f(\xi;\phi)$ from $p(\xi;\phi)$ is assumed here to have been caused by imperfections existing on the data. That is to say, the reconstruction method itself gives a "faithful" reconstruction. Since the given projections $p_g(\xi;\phi)$ can be separated into the perfect projections $p(\xi;\phi)$ added to the imperfections $e(\xi;\phi)$ (in the sense of section 7.3), it is conjectured here that the unknown perfect projections may be separated out by the difference between $p_f(\xi;\phi)$ and $p_g(\xi;\phi)$ leaving $p_d(\xi;\phi)$ equal to an expression containing the inconsistencies in terms of this difference. This section examines a new approach of estimating the inconsistencies $e(\xi;\phi)$ in the projections. The techniques developed are not new in the sense of separating out $p(\xi;\phi)$ through the difference projections but original in the method of recovery of $e(\xi;\phi)$.

Section 7.4.1 develops the mathematical basis by which the unknown inconsistencies $e(\xi;\phi)$ are separted from the unknown perfect projections $p(\xi;\phi)$. This development relies on those of section 7.2 and section 7.3. Section 7.4.2 examines the effects of a single point inconsistency, placed on one projection, on the difference projections $p_d(\xi;\phi)$ and the recovery of this inconsistency. In evaluating the effects of a single point the impulse response is recovered. This allows the recovery of more general inconsistencies to be assessed since the processes involved are linear. The impulse response developed in section 7.4.1 is used to recover the inconsistency by deconvolution. Section 7.4.3 proposes an alternative deconvolution based on an elimination of assumptions made in section 7.4.1 for the impulse response on $p_d(\xi;\phi)$.

7.4.1 Projection error impulse responses

Consider the reconstruction of the distribution $f(r;\theta)$ whose region of support is Ω_- (refer to figure 6.1), the given projections measured from Ω_-, in agreement with (7.5), are

$$\begin{aligned} p_g(\xi;\phi) &= p(\xi;\phi) + e(\xi;\phi) \\ &= 0 \quad \text{for} \quad |\xi| > a. \end{aligned} \tag{7.6}$$

In this section the effects are examined of a single point error (i.e., inconsistency), superimposed onto one projection, on the difference projections

$$p_d(\xi;\phi) = p_f(\xi;\phi) - p_g(\xi;\phi). \tag{7.7}$$

This is done by considering the inconsistency $e(\xi;\phi)$ to be a single dirac-delta function with coordinates $(\xi_1;\phi_1)$,

$$e(\xi;\phi) = \delta(\xi-\xi_1)\,\delta(\phi-\phi_1) \quad \text{and} \quad |\xi_1| \le a, \tag{7.8}$$

where the δ are 1-D dirac-delta functions. The contributions of this delta-function to $p_d(\xi;\phi)$ are here referred to as the *projection error impulse response*. It should be noted that the inconsistency given in (7.8) is spatially inconsistent with any finite region, in the sense described in section 7.2. In substituting the inconsistencies, (7.8), into (5.63) it can be seen that the modified projections of $e(\xi;\phi)$ then have the shape of the modifying function $h(\xi)$ (refer section 5.6.3.3) displaced ξ_1 along the ξ-axis for one view angle ϕ_1 :

$$\begin{aligned} \tilde{e}(\xi;\phi) &= e(\xi;\phi) \odot h(\xi) \\ &= h(\xi-\xi_1)\,\delta(\phi-\phi_1). \end{aligned} \tag{7.9}$$

Consider first the reconstruction in the region $\Omega_+ \cup \Omega_-$ (refer figure 6.1) from the given projections $p_g(\xi;\phi)$ (described in section 7.3). The reconstruction method adopted here to give a "faithful" reconstruction is the modified back-projection method described in section 5.6.3.3. Upon substituting the given projections, equation (7.6), into (5.62), the reconstruction is obtained as

$$\begin{aligned} \hat{f}(r;\theta) &= \mathcal{B}\{\,\tilde{p}_g(\xi;\phi)\,\}(r;\theta) \\ &= \mathcal{B}\{\,\tilde{p}(\xi;\phi)\,\}(r;\theta) + \mathcal{B}\{\,\tilde{e}(\xi;\phi)\,\}(r;\theta), \end{aligned} \tag{7.10}$$

and upon recalling (5.62) this can be written in terms of the true distribution $f(r;\theta)$:

$$\hat{f}(r;\theta) = f(r;\theta) + \mathcal{B}\{\,\tilde{e}(\xi;\phi)\,\}(r;\theta). \tag{7.11}$$

The forward projections of the reconstruction region are by definition the Radon transform of the reconstruction $\hat{f}(r;\theta)$ (refer section 5.5) :

$$\begin{aligned} p_f(\xi;\phi) &= \int_{\Omega_+\cup\Omega_-} \hat{f}(r;\theta)\,d\eta \\ &= \int_{\Omega_+\cup\Omega_-} f(r;\theta)\,d\eta + \int_{\Omega_+\cup\Omega_-} \mathcal{B}\{\,\tilde{e}(\xi;\phi)\,\}(r;\theta)\,d\eta. \end{aligned} \tag{7.12}$$

Since the true distribution is confined to the region of support Ω_- then through the definition for the perfect projections given in equation (5.11)

$$p_f(\xi;\phi) = p(\xi;\phi) + \int_{\Omega_+\cup\Omega_-} \mathcal{B}\{\, \tilde{e}(\xi;\phi) \,\}(r;\theta)\, d\eta. \tag{7.13}$$

The difference projections can now be written, substituting (7.6) and (7.13) into (7.7), as :

$$p_d(\xi;\phi) = -e(\xi;\phi) + \int_{\Omega_+\cup\Omega_-} \mathcal{B}\{\, \tilde{e}(\xi;\phi) \,\}(r;\theta)\, d\eta. \tag{7.14}$$

Upon examination of (7.14), there does not seem to be any apparent difficulty in solving for the $e(\xi;\phi)$ since $p_d(\xi;\phi)$ can be found. Substituting (5.63) into (7.14), then replacing the convolution and back-projection operators by their integral representations, (5.22) and (2.44) respectively, and rearranging the order of integration, yields

$$\begin{aligned}
p_d(\xi;\phi) &= -e(\xi;\phi) + \int_{\Omega_+\cup\Omega_-} \mathcal{B}\{\, e(\xi;\phi)\odot h(\xi) \,\}(r;\theta)\, d\eta \\
&= -e(\xi;\phi) + \int_{\Omega_+\cup\Omega_-} \mathcal{B}\left\{ \int_{-\infty}^{\infty} e(\xi';\phi')\, h(\xi-\xi')\, d\xi' \right\}(r;\theta)\, d\eta \\
&= -e(\xi;\phi) + \int_{\Omega_+\cup\Omega_-} \int_0^{\pi} \int_{-\infty}^{\infty} e(\xi';\phi')\, h(r\cos(\phi'-\theta)-\xi')\, d\xi'\, d\phi'\, d\eta \\
&= -e(\xi;\phi) + \int_{-\sqrt{b^2-\xi^2}}^{\sqrt{b^2-\xi^2}} \int_0^{\pi} \int_{-\infty}^{\infty} e(\xi';\phi')\, h(r\cos(\phi'-\theta)-\xi')\, d\xi'\, d\phi'\, d\eta.
\end{aligned} \tag{7.15}$$

Since the extent of the given projections (in the ξ' direction) are limited to the region of support as shown in equation (7.6), i.e., $e(\xi';\phi') = 0$ for $|\xi'| > a$, then the convolution integral in (7.15) need only be evaluated over the range $[-a,a]$ for ξ'. The difference projections can now be written as

$$\begin{aligned}
p_d(\xi;\phi) &= -e(\xi;\phi) + \int_{-\sqrt{b^2-\xi^2}}^{\sqrt{b^2-\xi^2}} \int_0^{\pi} \int_{-a}^{a} e(\xi';\phi')\, h(r\cos(\phi'-\theta)-\xi')\, d\xi'\, d\phi'\, d\eta \\
&= -e(\xi;\phi) + \int_0^{\pi} \int_{-a}^{a} e(\xi';\phi') \left[\int_{-\sqrt{b^2-\xi^2}}^{\sqrt{b^2-\xi^2}} h(r\cos(\phi'-\theta)-\xi')\, d\eta \right] d\xi'\, d\phi'.
\end{aligned} \tag{7.16}$$

Equation (7.16) can now be put into the form of a Fredholm integral of the second kind

[Morse and Feshbach, 1953], involving the imperfections $e(\xi;\phi)$:

$$e(\xi;\phi) = -p_d(\xi;\phi) + \int_0^{\pi}\int_{-a}^{a} e(\xi';\phi')\, h'(\xi';\phi';\xi;\phi)\, d\xi'\, d\phi' \tag{7.17}$$

where the kernel $h'(\xi';\phi';\xi;\phi)$ is given by

$$h'(\xi';\phi';\xi;\phi) = \int_{-\sqrt{b^2-\xi^2}}^{\sqrt{b^2-\xi^2}} h(r\cos(\phi'-\theta)-\xi')\, d\eta \tag{7.18}$$

and is referred to from this point as the *error kernel*. At this stage the inconsistencies have been expressed solely in terms of the known difference projections $p_d(\xi;\phi)$ but in a form that does not allow direct evaluation of $e(\xi;\phi)$. Although techniques are well developed for solving Fredholm integral equations, of the type shown in (7.17), a solution is not invoked here. Instead the situation is simplified through a careful examination of the error kernel given in (7.18).

The position $\xi = r\cos(\phi'-\theta)-\xi'$ along the modifying function $h(\xi)$ can be rewritten in a form varying linearly with the projection coordinate η (refer to figure 6.1) :

$$r\cos(\phi'-\theta)-\xi' = c + m\,\eta, \tag{7.19}$$

where the slope and intercept can be expressed by expanding $\cos(\phi'-\theta)$ as $\cos\left((\phi'-\phi)-(\theta-\phi)\right)$ using trigonometric identities, and through the identities given in (5.7) :

$$\begin{aligned} c &= \xi\cos(\phi-\phi')-\xi' \\ m &= \sin(\phi-\phi'). \end{aligned} \tag{7.20}$$

The error kernel can then be rewritten as the following integral over the line of slope m :

$$h'(\xi';\phi';\xi;\phi) = \int_{-\sqrt{b^2-\xi^2}}^{\sqrt{b^2-\xi^2}} h(c+m\,\eta)\, d\eta. \tag{7.21}$$

The projections are limited to the range $|\xi| \leq a$ since the true distribution is known to have the support region Ω_- of radius a (as given in (7.6)). Likewise, it is desired to correct for inconsistencies within this range of ξ since the projections are known to be zero outside. If b, the radius of Ω_+, is chosen so that $b^2 \gg a^2$, then

$$\sqrt{b^2-\xi^2} \approx b \quad \text{for } b^2 \gg a^2 \text{ and } |\xi| \leq a, \tag{7.22}$$

and (7.21) can be rewritten

$$h'(\xi';\phi';\xi;\phi) \approx \int_{-b}^{b} h(c+m\,\eta)\,d\eta. \tag{7.23}$$

Consider from (7.23) the case of $h'(\xi';\phi';\xi;\phi)$ for $\phi' = \phi$. In this situation the back-projection angle ϕ' is equal to the angle of the forward-projection ϕ. Substituting $\phi' = \phi$ into (7.20) then gives the error kernel

$$\begin{aligned} h'(\xi';\phi;\xi;\phi) &= \int_{-b}^{b} h(\xi-\xi')\,d\eta \\ &= 2\,b\,h(\xi-\xi'). \end{aligned} \tag{7.24}$$

When these two angles are not the same

$$h'(\xi';\phi';\xi;\phi) = \int_{-b}^{b} h(c+m\,\eta)\,d\eta \quad \text{and} \quad m \neq 0. \tag{7.25}$$

As noted in section 5.6.1, the modifying function $h(\xi)$ has a zero-mean property, expressed in (5.39), so that if the radius b is made sufficiently large and $\phi' \neq \phi$, as is the case in (7.25), then the kernel approaches the zero-mean property given in (5.39), i.e.,

$$\lim_{b\to\infty} \int_{-b}^{b} h(c+m\,\eta)\,d\eta = 0 \quad \text{for} \quad m \neq 0. \tag{7.26}$$

Practically speaking, the reconstruction radius cannot be made infinite. However, it is noted here that the modifying function $h(\xi)$ converges to zero as $1/\xi^2$ as shown in (5.38). This can also be seen to be the case for the discrete modifying functions given in (5.68) and (5.69). Assuming that the reconstruction radius b is sufficiently large such that

$$\begin{aligned} h'(\xi';\phi';\xi;\phi) &= \int_{-b}^{b} h(c+m\,\eta)\,d\eta \\ &\approx 0 \quad \text{for large } b \quad \text{and} \quad m \neq 0, \end{aligned} \tag{7.27}$$

then the kernel can be rewritten as

$$h'(\xi';\phi';\xi;\phi) = \begin{cases} 2\,b\,h(\xi-\xi') & \phi' = \phi \\ 0 & \phi' \neq \phi. \end{cases} \tag{7.28}$$

Equation (7.17) can then be rewritten as

$$\begin{aligned} e(\xi;\phi) &= -p_d(\xi;\phi) + 2\,b \int_{-a}^{a} e(\xi';\phi)\,h(\xi-\xi')d\xi' \\ &= -p_d(\xi;\phi) + 2\,b\,\left[e(\xi';\phi) \odot h(\xi)\right], \end{aligned} \qquad (7.29)$$

and the difference projections $p_d(\xi;\phi)$ can be expressed in terms of the inconsistencies $e(\xi;\phi)$:

$$\begin{aligned} p_d(\xi;\phi) &= 2\,b\,\left[e(\xi;\phi) \odot h(\xi)\right] - e(\xi;\phi) \\ &= e(\xi;\phi) \odot \left[2\,b\,h(\xi) - \delta(\xi)\right]. \end{aligned} \qquad (7.30)$$

Equation (7.30) implies that the inconsistencies can be obtained from the difference projections by a series of 1-D deconvolutions of the 1-D error kernel

$$h'(\xi) = 2\,b\,h(\xi) - \delta(\xi) \qquad (7.31)$$

for each projection angle ϕ. Once they have been recovered they can be subtracted from the given projections, thereby ameliorating their effects on the reconstruction within the known support Ω_-. Table 7.1 shows the steps proposed to recover $e(\xi;\phi)$.

7.4.2 Results of 1-D deconvolutions for estimating inconsistencies

Consider the inconsistencies given in (7.8). Upon substituting these into (7.30), the difference projections are

$$p_d(\xi;\phi) = \left[2\,b\,h(\xi-\xi_1) - \delta(\xi-\xi_1)\right]\;\delta(\phi-\phi_1). \qquad (7.32)$$

Equation (7.32) describes the impulse response manifested in the difference projections. This section describes the results obtained when the algorithm outlined in table 7.1 was used for deconvolving the inconsistencies given by (7.8), and assesses how accuractely (7.32) describes the impulse response when reconstructing from a finite number of projections.

The support of the actual distribution $f(r;\theta)$, Ω_-, is taken to be of radius $a = 28cm$ and the radius of the reconstruction region Ω_+ has been chosen to be $b = 4a$. The image is reconstructed on a 225 x 225 grid of pixels (of which only the pixels within the radius b are reconstructed), with sides of length $1cm$. Sample spacing for projection samples are chosen to equal the length of the pixel side, $\Delta\xi = 1cm$. In accordance with (6.9), the number of projections taken for the reconstruction, for a radius $b = 112cm$ and resolution $\sigma = 1cm$, was $M = 360$. Discrete "inconsistencies" consisting of single points with value unity are added to otherwise zero-valued projections.

Step 1	Reconstruct the given projections within the region $\Omega_+ \cup \Omega_-$ of radius b : $\hat{f}(r;\theta) = \mathcal{B}\{\,\tilde{p}_g(\xi;\phi)\,\}(r;\theta) \quad \text{for} \quad (r;\theta) \in \Omega_+ \cup \Omega_-.$
Step 2	Calculate the forward projections : $p_f(\xi;\phi) = \int_{\Omega_+ \cup \Omega_-} \hat{f}(r;\theta)\, d\eta.$
Step 3	Calculate the difference projections by subtracting the given projections from $p_f(\xi;\phi)$: $p_d(\xi;\phi) = p_f(\xi;\phi) - p_g(\xi;\phi).$
Step 4	Estimate the inconsistencies $e(\xi;\phi)$ by a 1-D deconvolution from the difference projections of the kernel : $h'(\xi) = 2\,b\,h(\xi) - \delta(\xi).$
Step 5	Correct the given projection data by subtracting the estimated inconsistencies $\hat{e}(\xi;\phi)$: $p(\xi;\phi) \approx p_g(\xi;\phi) - \hat{e}(\xi;\phi).$

Table 7.1 Steps involved in an algorithm for estimating the inconsistencies in the given projections through a 1-D deconvolution of the error kernel $h'(\xi)$ given in (7.31).

As intimated in section 7.4.1 the method of modified back-projection was chosen for the analysis. The discrete modifying function of Shepp and Logan [1974], equation (5.69), was used and the square pixel method of section 5.5 was implemented for the back- and forward-projection; both of these choices were made for resolution purposes. For the adopted modifying function of equation (5.69), the discrete representation of the 1-D error kernel in (7.31) is

$$\begin{aligned} h'(n\,\Delta\xi) &= \frac{-4\,b\,\Delta\phi}{\pi^2\,\Delta\xi\,(4n^2-1)} - \delta(n) \\ &= \frac{-4\,b}{(\pi\,\Delta\xi\,M)\,(4n^2-1)} - \delta(n), \end{aligned} \tag{7.33}$$

where $\delta(n)$ is the impulse function used here and has a value of unity for $n = 0$ and zero for $n \neq 0$. It should be noted that in (7.33) the discrete convolution and back-projection integration intervals, $\Delta\xi$ and $\Delta\phi = \pi/M$ respectively, have been incorporated into the 1-D error kernel $h'(n\,\Delta\xi)$.

Figures 7.1 (*a*) and (*b*) shown two inconsistent projections each containing a single point of value unity placed at the points $(\xi_1; \phi_1) = (0,0)$ and $(\xi_1; \phi_1) = (27; 0)$, respectively, that have been added to the otherwise zero-valued projections. The aim of this is to give insight into the results of the 1-D deconvolution algorithm outlined in table 7.1. The given projections $p_g(\xi; \phi)$ have been reconstructed by the method of modified back-projection, as mentioned above, and the forward projections of the reconstruction $p_f(\xi; \phi)$ are shown respectively in figures 7.1 (*c*) and (*d*). The difference projections, calculated according to (7.7), are shown in figures 7.1 (*e*) and (*f*).

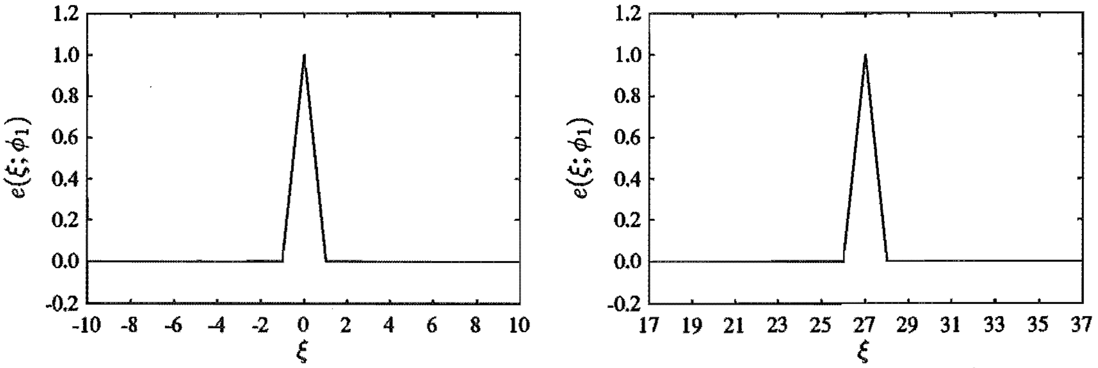
Figures 7.2 (*c*) and (*d*) show the inconsistencies estimated from the algorithm in table 7.1 by deconvolving the discrete kernel given in (7.33) for $\phi = \phi_1$. Figures 7.2 (*e*) and (*f*) show the impulse response of the estimated inconsistencies for a range of ϕ adjacent to ϕ_1.

It should be noted from figures 7.2 (*a*) and (*b*) that the difference projections are significantly distorted from the actual inconsistencies. This affirms the motivation outlined in section 7.1 for deconvolving the kernel in (7.33) to account for the shape of the modifying function. It is apparent from figures 7.2 (*c*) and (*d*) that the inconsistency has been well recovered from the 1-D deconvolution for $\phi = \phi_1$. However, if the impulse response is examined, shown in figure 7.2 (*c*), it can be seen that residuals are evident on the projection angles adjacent to $\phi = \phi_1$ and that these residuals decrease as the deviations $|\phi - \phi_1|$ and $|\xi - \xi_1|$ are increased. The same can be observed in figure 7.2 (*f*) and it can be seen upon comparing the two impulse responses that the residuals are relatively invariant for impulses placed across the range of $\xi' = \xi_1$ spanning the support of region Ω_-.

Upon examination of figure 7.3 it can be seen that there are significant residues on the difference projections adjacent to that containing the impulse inconsistency. Ideally the difference projection for $\phi = \phi_1$ has the shape of the kernel given in (7.33) and there would be no residues on other projections $\phi \neq \phi_1$ in accordance with (7.28), i.e., the impulse response would be one-dimensional (as is the assumption made in (7.28)). However, as the reconstruction radius is of finite extent this is not the case and the impulse response has "leaked" into the projections $\phi \neq \phi_1$. As intimated above, the impulse response appears to be relatively invariant across the range of $\xi' = \xi_1$ spanning the support of region Ω_-. Subsequently, is seems intuitive to use a 2-D deconvolution to recover the inconsistencies. This is done by considering the error kernel to be the 2-D function $h'(\xi; \phi)$.

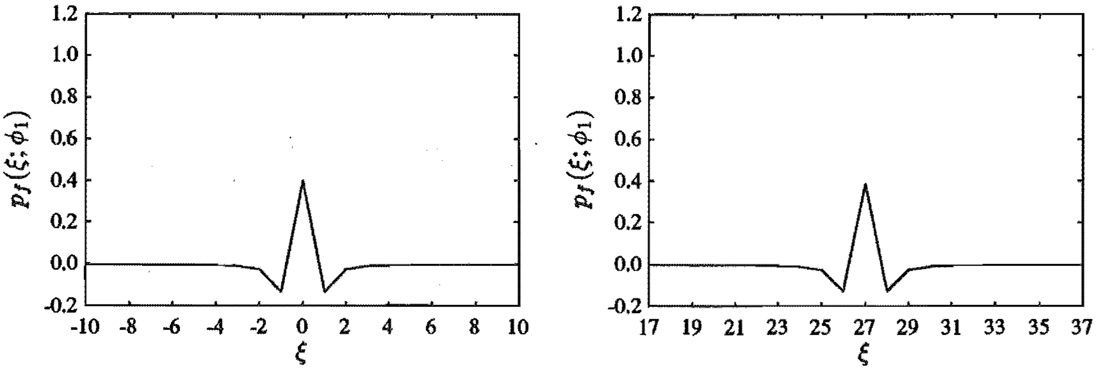
7.4.3 Results of 2-D deconvolutions for estimating inconsistencies

As indicated in section 7.4.2, the assumptions made in section 7.4.1 for the recovery of inconsistencies through deconvolution of a 1-D error kernel, that is the contributions to the difference projections of a single point inconsistency placed on a projection of angle $\phi = \phi_1$ should not appear in significant values on projections at other angles $\phi \neq \phi_1$, are violated when the reconstruction radius is for practical purposes finite, i.e., equation (7.27) is violated when $b = 128\,\Delta\xi$ and $\Delta\phi = \pi/360$. This is apparent from figure 7.3 which shows the residues residing on the projections adjacent to $\phi = \phi_1$. It is also apparent from figures 7.2 (*c*) and (*f*) that the 2-D error kernel is relatively invariant for single errors placed over the range of ξ', $-a \leq \xi' \leq a$. As intimated in section 5.5, the back-projection process is not invariant with the back-projection angle ϕ'. This is because of the discrete nature and shape of the basis functions used in the image representation. In order to estimate the 2-D impulse response on the difference projections the impulse response, for a single point error placed at



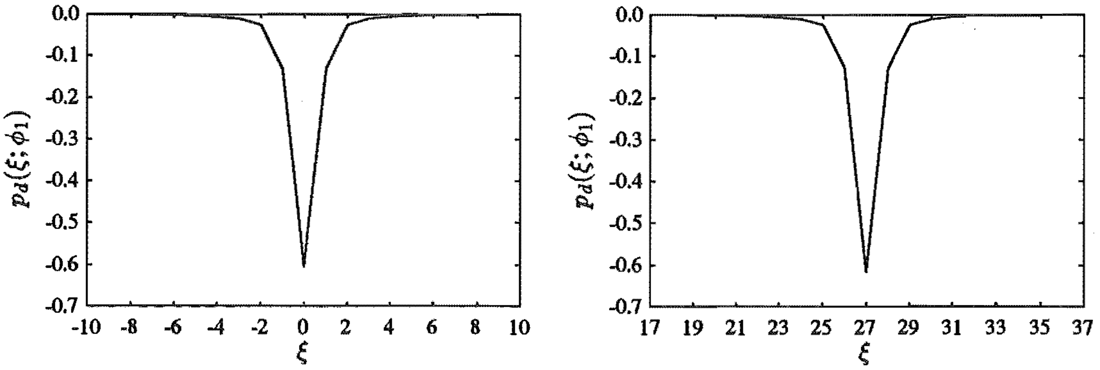
(*a*) Inconsistent projection $e(\xi;\phi_1)$ where $(\xi_1;\phi_1) = (0;0)$.

(*b*) Inconsistent projection $e(\xi;\phi_1)$ where $(\xi_1;\phi_1) = (27;0)$.



(*c*) Forward projection $p_f(\xi;\phi_1)$ where $(\xi_1;\phi_1) = (0;0)$.

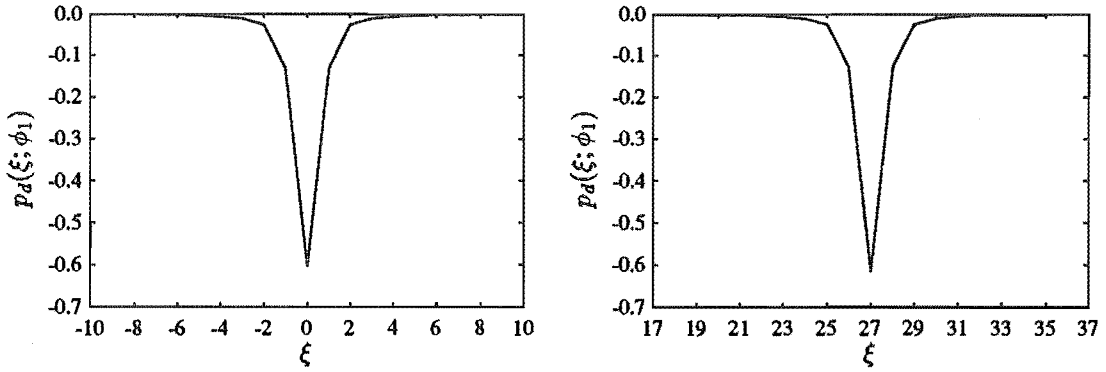
(*d*) Forward projection $p_f(\xi;\phi_1)$ where $(\xi_1;\phi_1) = (27;0)$.



(*e*) Difference projection $p_d(\xi;\phi_1)$ where $(\xi_1;\phi_1) = (0;0)$.

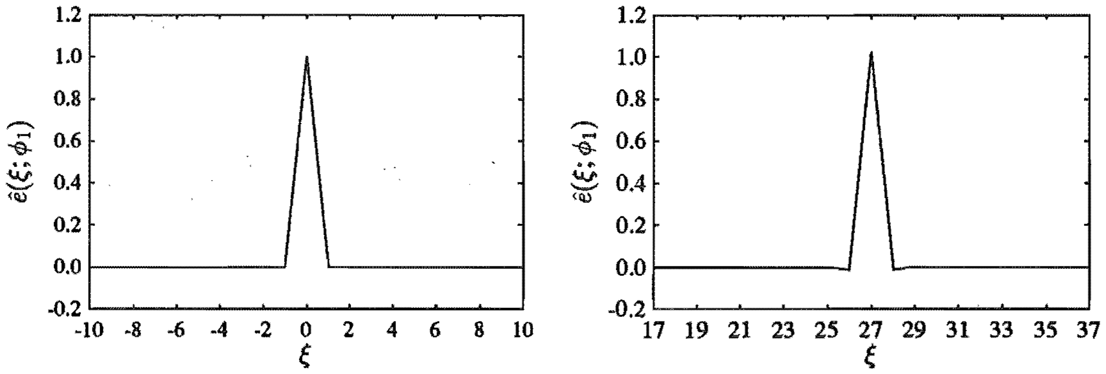
(*f*) Difference projection $p_d(\xi;\phi_1)$ where $(\xi_1;\phi_1) = (27;0)$.

Figure 7.1 The inconsistent projection, forward projection and difference projection for single inconsistencies of value unity placed at $(\xi_1;\phi_1) = (0;0)$ and $(\xi_1;\phi_1) = (27;0)$ are shown, respectively, in figures (*a*),(*c*),(*e*) and (*b*),(*d*),(*f*).



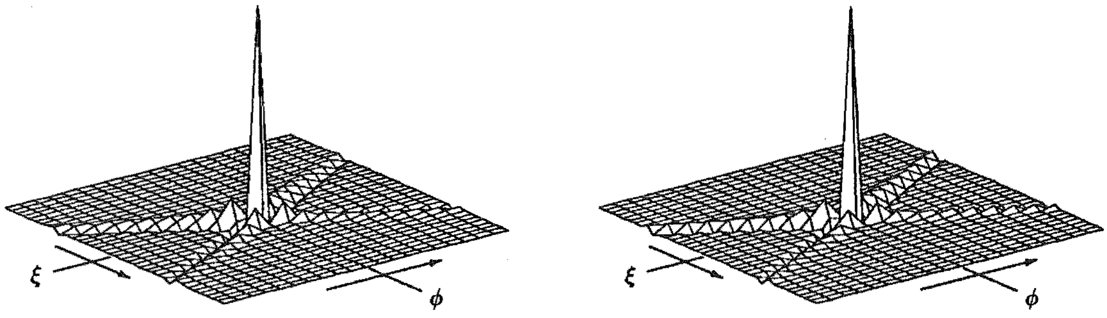
(*a*) Difference projection $p_f(\xi;\phi_1)$ where $(\xi_1;\phi_1) = (0;0)$.

(*b*) Difference projection $p_f(\xi;\phi_1)$ where $(\xi_1;\phi_1) = (27;0)$.



(*c*) Estimated inconsistencies $\hat{e}(\xi;\phi_1)$ where $(\xi_1;\phi_1) = (0;0)$.

(*d*) Estimated inconsistencies $\hat{e}(\xi;\phi_1)$ where $(\xi_1;\phi_1) = (27;0)$.



(*e*) Estimated inconsistencies $\hat{e}(\xi;\phi)$ when $(\xi_1;\phi_1) = (0;0)$ for $-18 \le \xi \le 18$ and $-\frac{\pi}{36} \le \phi \le \frac{\pi}{36}$.

(*f*) Estimated inconsistencies $\hat{e}(\xi;\phi)$ when $(\xi_1;\phi_1) = (27;0)$ for $9 \le \xi \le 45$ and $-\frac{\pi}{36} \le \phi \le \frac{\pi}{36}$.

Figure 7.2 Application of the algorithm detailed in table 7.1 to estimate inconsistencies $e(\xi;\phi)$. The results of the difference projection, estimated inconsistencies for the same projection, estimated inconsistencies over a range of adjacent projections for inconsistencies, of unity value, placed at $(\xi_1;\phi_1) = (0;0)$ and $(\xi_1;\phi_1) = (27;0)$ are shown, respectively, in (*a*), (*c*), (*e*) and (*b*), (*d*), (*f*).

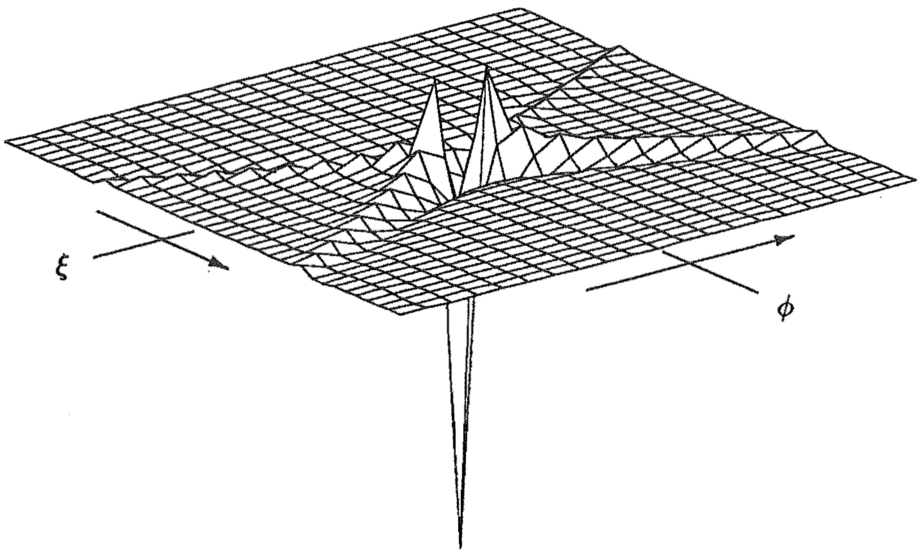


Figure 7.3 The difference projections $p_d(\xi;\phi)$ calculated from an inconsistency of value unity placed at $(\xi_1;\phi_1) = (0;0)$. The projections $p_d(\xi;\phi)$ are shown for $-18 \le \xi \le 18$ and $-\frac{\pi}{36} \le \phi \le \frac{\pi}{36}$.

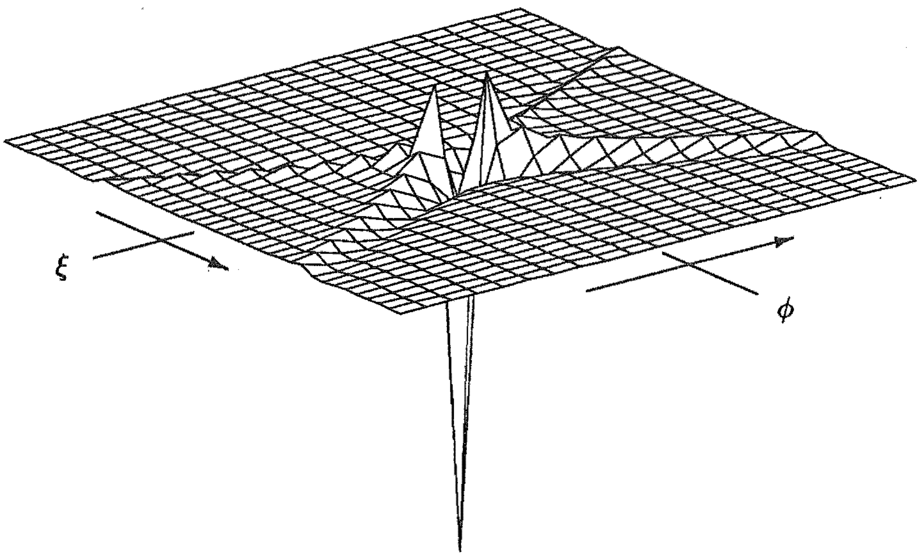


Figure 7.4 The average 2-D response function for an inconsistency of value unity. The average has been calculated from the difference projections when the inconsistency was placed at $(\xi_1;\phi_1) = (0,\phi_m)$ for all $m = 0, 1, \ldots, M-1$.

$(\xi_1;\phi_1) = (0;\phi'_m)$, has been averaged over all $\phi'_m = m\pi/M$. Figure 7.4 shows the average 2-D impulse response. It can be seen upon comparison with figure 7.3, that there is a small amount of variation between the averaged 2-D error kernel and that for $\phi' = 0$.

The 2-D error kernel shown in figure 7.4 has been used to recover inconsistencies through a two-dimensional deconvolution from the difference projections. A number of deconvolution techniques are available in 2-D [Bates and McDonnell, 1989; Davey, 1989]. However, for simplification of the analysis the *inverse filtering* method is adopted here. If $P_d(u,v)$ and

$H'(u,v)$ are the 2-D Fourier transforms (refer section 2.3.1) of the difference projections and 2-D error kernel ($p_d(\xi;\phi)$ and $h'(\xi;\phi)$), respectively, then the deconvolution of $e(\xi;\phi)$ from $p_d(\xi;\phi)$ to give an estimate the inconsistencies, $\hat{e}(\xi;\phi)$, by the method of inverse filtering, is [Bates and McDonnell, 1989] :

$$\hat{E}(u,v) \;=\; P_d(u,v)\,\frac{1}{H'(u,v)}, \qquad (7.34)$$

where $\hat{E}(u,v)$ is the 2-D Fourier transform of $\hat{e}(\xi;\phi)$. Although (7.34) represents one of the most simple conventional deconvolution procedures it is well known to be unforgiving when there is contamination present in the $P_d(u,v)$. This is because the contamination is amplified when $H'(u,v)$ is small. The degradation is often manifest as ringing in $\hat{E}(u,v)$ from the frequencies for which $H'(u,v)$ is small. For this purpose, it was chosen to deal partially with the inadequacies of this method by simply eliminating frequencies whereby the magnitude of $H'(u,v)$ was less than 3.5 percent of its maximum Fourier component.

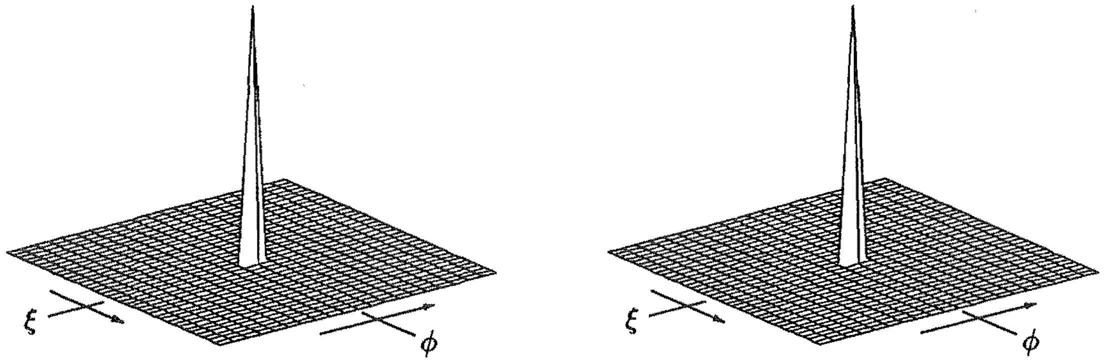
Four sets of inconsistent projections $e(\xi;\phi)$ have been estimated by the aforementioned 2-D deconvolution from the difference projections. The first three contain single point inconsistencies placed at $(\xi_1;\phi_1) = (0,0)$, $(\xi_1;\phi_1) = (14,0)$ and $(\xi_1;\phi_1) = (0,\frac{\pi}{4})$. These are shown, respectively, in figures 7.5 (*a*), 7.5 (*b*) and 7.6 (*a*). Figure 7.6 (*b*) shows the forth inconsistent projection set containing six points, each of value unity and placed at the coordinates given in table 7.2. Also shown in figure 7.5 and figure 7.6 are the difference projections $p_d(\xi;\phi)$ for these four inconsistent projection sets and the estimated inconsistencies $\hat{e}(\xi;\phi)$ obtained from a 2-D deconvolution of $h'(\xi;\phi)$ given in figure 7.4. It is apparent from these results that the main detail of the inconsistencies is recoverable for the range of single point inconsistencies placed over $-a \leq \xi' \leq a$ and ϕ'_m. In particular, it can be seen from figure 7.6 (*f*) that a group of inconsistencies placed together is recoverable. However, the residues from the deconvolution have reinforced here. These residues may be largely caused by ringing effects introduced from the inverse filtering method, mentioned above, used for deconvolution.

ϕ'	ξ'
$\frac{\pi}{4} - \frac{\pi}{M}$	13, 14
$\frac{\pi}{4}$	12, 13, 14, 15
$\frac{\pi}{4} + \frac{\pi}{M}$	13, 14

Table 7.2 Coordinates of the single point inconsistencies, of value unity, shown in figure 7.6 (*b*). Four points have been placed on the projection $\phi' = \pi/4$, two points on the projections immediately adjacent.

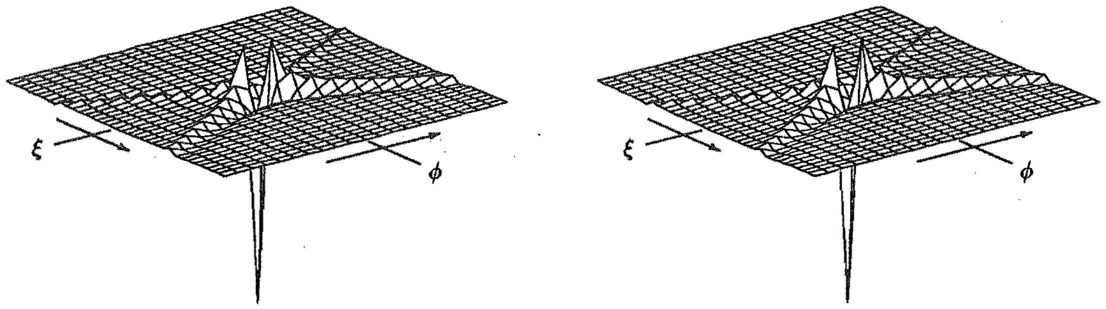
7.5 CONCLUSIONS

In chapter 6 a simultaneous iterative algorithm was proposed and tested to exploit the plethora of views idea, given as an open problem in section 6.2, for improvement of images reconstructed



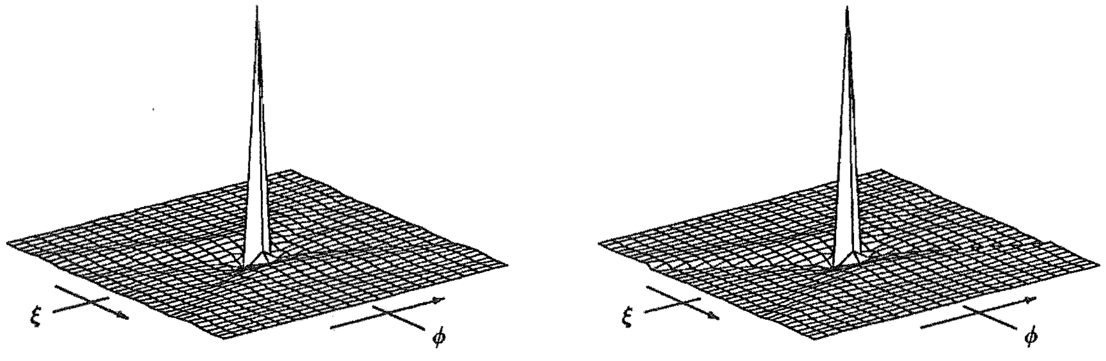
(*a*) Inconsistent projections $e(\xi;\phi)$ when $(\xi_1;\phi_1) = (0;0)$ for $-18 \le \xi \le 18$ and $-\frac{\pi}{36} \le \phi \le \frac{\pi}{36}$.

(*b*) Inconsistent projections $e(\xi;\phi)$ when $(\xi_1;\phi_1) = (14;0)$ for $9 \le \xi \le 45$ and $-\frac{\pi}{36} \le \phi \le \frac{\pi}{36}$.



(*c*) Difference projections $p_d(\xi;\phi)$ when $(\xi_1;\phi_1) = (0;0)$ for $-18 \le \xi \le 18$ and $-\frac{\pi}{36} \le \phi \le \frac{\pi}{36}$.

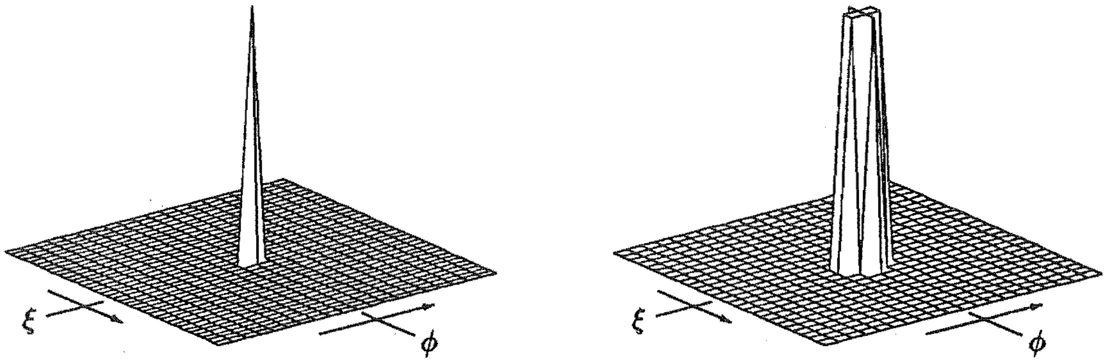
(*d*) Difference projections $p_d(\xi;\phi)$ when $(\xi_1;\phi_1) = (14;0)$ for $9 \le \xi \le 45$ and $-\frac{\pi}{36} \le \phi \le \frac{\pi}{36}$.



(*e*) Estimated inconsistencies $\hat{e}(\xi;\phi)$ when $(\xi_1;\phi_1) = (0;0)$ for $-18 \le \xi \le 18$ and $-\frac{\pi}{36} \le \phi \le \frac{\pi}{36}$.

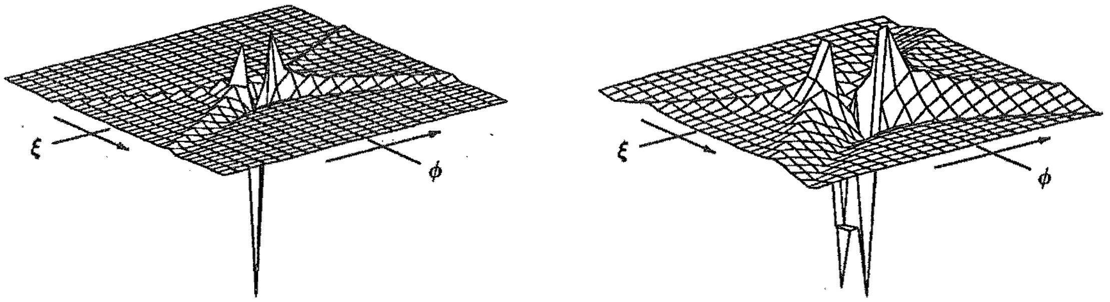
(*f*) Estimated inconsistencies $\hat{e}(\xi;\phi)$ when $(\xi_1;\phi_1) = (14;0)$ for $9 \le \xi \le 45$ and $-\frac{\pi}{36} \le \phi \le \frac{\pi}{36}$.

Figure 7.5 Two-dimensional deconvolution of the kernel shown in figure 7.4 from the difference projections in (*c*) and (*d*) to give an estimate, shown in (*e*) and (*f*) respectively, of the inconsistent projections, also shown in (*a*) and (*b*).



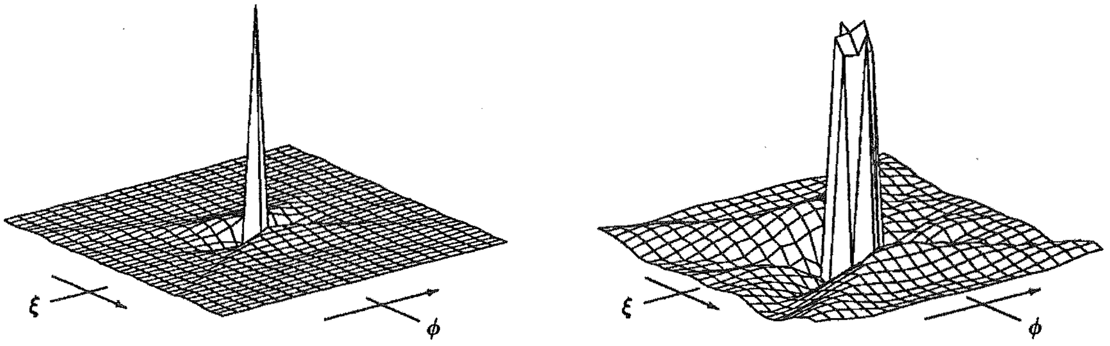
(*a*) Inconsistent projections $e(\xi;\phi)$
when $(\xi_1;\phi_1) = (0;\frac{\pi}{4})$
for $-18 \le \xi \le 18$ and $(\frac{\pi}{4} - \frac{\pi}{36}) \le \phi \le (\frac{\pi}{4} + \frac{\pi}{36})$.

(*b*) Inconsistent projections $e(\xi;\phi)$
for $0 \le \xi \le 36$ and $(\frac{\pi}{4} - \frac{\pi}{36}) \le \phi \le (\frac{\pi}{4} + \frac{\pi}{36})$.



(*c*) Difference projections $p_d(\xi;\phi)$
when $(\xi_1;\phi_1) = (0;\frac{\pi}{4})$
for $-18 \le \xi \le 18$ and $(\frac{\pi}{4} - \frac{\pi}{36}) \le \phi \le (\frac{\pi}{4} + \frac{\pi}{36})$.

(*d*) Difference projections $p_d(\xi;\phi)$
for $0 \le \xi \le 36$ and $(\frac{\pi}{4} - \frac{\pi}{36}) \le \phi \le (\frac{\pi}{4} + \frac{\pi}{36})$.



(*e*) Estimated inconsistencies $\hat{e}(\xi;\phi)$
when $(\xi_1;\phi_1) = (0;\frac{\pi}{4})$
for $-18 \le \xi \le 18$ and $(\frac{\pi}{4} - \frac{\pi}{36}) \le \phi \le (\frac{\pi}{4} + \frac{\pi}{36})$.

(*f*) Estimated inconsistencies $\hat{e}(\xi;\phi)$
for $0 \le \xi \le 36$ and $(\frac{\pi}{4} - \frac{\pi}{36}) \le \phi \le (\frac{\pi}{4} + \frac{\pi}{36})$.

Figure 7.6 Two-dimensional deconvolution of the kernel shown in figure 7.4 from the difference projections in (*c*) and (*d*) to give an estimate, shown in (*e*) and (*f*) respectively, of the inconsistent projections, also shown in (*a*) and (*b*).

from imperfect projections. However, it was found that this algorithm was not successful in improving significantly the quality of the reconstruction within the known support of the image distribution although it was successful in correcting the projection data to zero the reconstruction in Ω_+.

What is required for correction of imperfections is a technique for their recovery that can relate, in a more "intuitive" fashion, the reconstruction distortion with the actual imperfections. Such an approach has been the topic of chapter 7. In section 7.2 an extended definition for the consistency of projections, termed spatial consistency, is given that relates a set of consistent projections with a region with which they are consistent. In conjunction with this definition an alternative definition, spatial inconsistency, has been given which has allowed the projections to be considered as comprising two components, a set of projections spatially consistent with the reconstruction region Ω_- and a set spatially inconsistent with Ω_-.

In section 7.4 an approach has been developed for recovering the set of spatially inconsistent projections. The approach is a direct method for the recovery of the inconsistent projections based on a deconvolution, from the difference projections, of the inconsistencies using a 1-D error kernel developed in section 7.4.1. The 1-D error kernel is developed by considering the reconstruction distortion and its forward projections. The given projections are then subtracted from the forward projections to give an expression, a Fredholm integral equation, that has separated out the inconsistencies from the perfect projection data. This expression is then simplified by a careful examination of the response of the difference projections to single point inconsistencies. By examining the response of the difference projections to a single point inconsistency, it has been demonstrated that its effect is spread over other projection angles apart from that projection angle in which it was originally contained. As a result, it is seen to be necessary to deconvolve the effects of inconsistencies over two dimensions. Algorithms employing the difference projections, in their raw form, for correction of projection data imperfections will therefore have their effects blurred over the projections. An impulse response function has been proposed for a 2-D deconvolution, from the difference projections, of the inconsistencies to reduce the blurring over projections. Both of these deconvolution approaches has been implemented and shown to be successful for the recovery of sharp inconsistencies confined to a narrow range of projections.

CHAPTER 8

CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH

The aim of this thesis has been the investigation of aspects of computed tomography and to devise new and improved methods for the quantitative reconstruction of computed tomographic images.

Chapter 1 introduces the motivation for studying computed tomography by describing some of its important applications. The problem of computing a tomographic image is set within the framework of the more general class of inverse problems. Section 1.5 provides much of the motivation for the original contributions contained in chapters 4, 6 and 7. Chapter 2 introduces equations describing the behaviour of wavefields exhibiting linear scalar wave motion. The equations describing wavefields are examined under a high frequency approximation. Solutions to the wave equation under the regime of diffraction tomography are given. Certain supplementary material is also presented. Chapter 3 examines the ray as a model for wavefields. Conventional physical theory regarding the behaviour of rays has been presented with comment on the validity of the ray model. Chapter 5 reviews solutions to X-ray computed tomography reconstruction along with much of their historical development.

This chapter summarises important aspects regarding the original contributions presented in this thesis. Conclusions and suggestions for further research on the original contributions documented in chapters 4, 6 and 7 are collected into sections 8.1, 8.2 and 8.3 respectively.

8.1 BENT-RAY CT

Section 4.3 introduces a novel algorithm for solving the bent-ray CT problem of reconstructing a refractive index distribution. It has been formulated for asymmetric and circularly symmetric (two- and one-dimensional) refractive index distributions. The basis for this formulation is a simple ray model of wave motion which allows the refractive index distribution to be found by solution of linear equations, rather than through the computationally expensive task of ray tracing associated with other proposed solutions. The algorithm has been used to give quantitative reconstructions of circularly symmetric refractive index distributions containing quite large variations. It is also able to handle appreciably larger variations in refractive index than the Born and Rytov approximations, while avoiding the computational consequences of an exact solution. It is worth noting that the accuracy of this algorithm is independent of the size of the region imaged, another limitation of methods based on the Born and Rytov approximations. These factors make it well suited to solving practical inverse problems.

While numerical results have been obtained for refractive index distributions that are circularly symmetric, a major limitation to the appraisal of the algorithm in its current form is that it remains untested for asymmetric distributions. Several aspects should be investigated to further develop its performance for circularly symmetric distributions and these developments extended to the reconstruction of asymmetric distributions. A number of aspects to be considered are outlined in table 4.2. These aspects are now considered in more detail below.

A representation has been proposed in section 4.6 and tested for the phase function from circularly symmetric refractive index distributions. Although the representation contains relatively simple basis functions, other more sophisticated representations have been investigated without the degree of success attributable to that finally chosen. This has led to the belief that the efficaciousness of the representation is crucial not only for the quality of the reconstruction but also for the efficiency of any solution. At present an optimal representation has not been investigated. There are however a number of points outlined in table 4.2, revealed through the investigations that have provided the results set out in section 4.7, that need to be considered when optimising the representation for an efficient solution. These points should be particularly important for providing an effective approach to reconstructing asymmetrical distributions.

The representation chosen includes terms that can be related directly to the known refractive index on the boundary. This has helped stabilise a numerical solution near the boundary because it provides known reference points. A representation that can, at least in part, be directly related to the measured phase data should do likewise. The efficaciousness of the representation also depends on how well it accords with the physics of the situation. The path taken by energy propagating from one point to another inside the medium is the same as that if the energy were travelling in the opposite direction. Basis functions could be chosen that comply with this reciprocity condition, thereby enforcing the known ray physics. Equation (4.12) suggests an expansion to accommodate the reciprocal nature of the ray path. Additionally, computational advantage could be gained in evaluating the eikonal's independence with respect to the transmitter coordinate, Φ (equation (4.17)). If $\nabla_P H(P,Q) \cdot \nabla_P H(P,Q)$ was composed of basis functions orthogonal in Φ, this integral over Φ would not be necessary.

A representation has been suggested in section 4.4 for the phase function from an asymmetric distribution. This representation has been adequate to represent the phase function from circularly symmetric distributions. However, further investigation of the algorithm, introduced in chapter 4, requires that this representation be tested for its adequacy with respect to representing an asymmetric distribution, and if not, that a suitable replacement be developed. Not only should it be able to represent an asymmetric distribution but it should provide an efficient representation in terms of the method of solution. This raises the question again of its efficaciousness.

The suitability of any finite representation raises a number of other questions when considering the size of the expansion required. Of considerable importance for the application of this algorithm is the reliability of reconstructions. In particular, the question of resolution is fundamental in understanding the detail revealed in a reconstruction. The resolution of the refractive index distribution reconstructed from a solution for the phase function is limited foremost by the finite number of terms involved in the expansion. Secondly, the basis functions chosen for the representation dictate the type of detail that can be resolved from a finite number of the basis functions. On the whole, once a finite expansion has been chosen to represent the phase function it is important that the resolution achievable can be related to data requirements and the expansion limits. One point has become apparent from the reconstruction of circularly symmetric distributions. An obvious method of improving resolution is to make available further measurements, or an abundance of them. This would reduce the number of equations introduced, rather "artificially" through (4.33), to ensure the entire system of equations is not underdetermined. Equation (4.32), enforcing the known circular symmetry of the distribution is introduced to the system at any number of radii, given by (4.33).

Additionally, as the physics of the phase function is further understood, new constraint equations can be introduced into the system of equations. Direct application of Fermat's principle, section 3.4, to constrain the phase function may lead to an improved solution for the refractive index distribution. However, formulating Fermat's principle to constrain the coefficients of the expansion for the phase function is not thought to be straightforward.

As noted in section 4.5, the main obstacle to formulating a manageable algorithm in solving for the refractive index distribution through the eikonal equation is that not all of the equations are linear in the expansion coefficients. The algorithm proposed in section 4.5 linearises the equations and solves for the non-linear components in an iterative fashion. The numerical linearisation technique used to obtain the solution is an important aspect that must be considered. There are a number of additional options available for the solution of non-linear equations that may be more apt in this application.

At present the algorithm proposed in section 4.5 assumes no *a priori* information regarding the distribution and no initial estimate of the expansion coefficients in the representation. As a result of assigning the coefficients initially to zero, the estimate assumed for the first iteration is that of a homogeneous distribution. If any *a priori* information is known about the distribution under examination, convergence may be speeded by making an initial estimate of the expansion coefficients. This initial estimate of the expansion coefficients could be calculated based on the same linearisation of (4.24) as outlined in section 4.5.

8.2 PLETHORA OF VIEWS: SIMULTANEOUS ITERATIVE REFINEMENT

Section 6.1 introduces the *plethora of views* idea (proposed by Bates [1988b]) aimed at improving quantitative CT image reconstruction. The theoretical foundation for this is the well-established idea [Crowther *et al.*, 1970] that when presented with a *plethora* of projections, by which is meant a number greater than that required to reconstruct the known region of support Ω_-, so that the permissible reconstruction region can be extended by the region Ω_+ (refer to figure 6.1) then the intensity of the reconstructed distribution should be negligible throughout the region denoted Ω_+. Any reconstruction within Ω_+, from a plethora of imperfect projections, that departs from what would be termed negligible has been caused by the imperfections of the projections. It is the conjecture of chapter 6 that contributions within Ω_+ can be utilised to ameliorate the effects of the imperfections on the reconstruction within Ω_- where the distribution is known to be contained.

The initial implementation of the plethora of views idea in chapter 6 has been in the context of wave motion that can be characterised as concourses of rays travelling along straight paths. An iterative algorithm (proposed by Herman [1988]) which uses the information contained within Ω_+, for the correction of imperfections contained on projection data, is detailed in section 6.5. The algorithm was designed to nullify the reconstruction (due to imperfections) within Ω_+, the implicit expectation being that this would also diminish the effect of the imperfections within Ω_-. The motivation for the experimental work adopted (detailed in section 6.6) to test this algorithm was for improvement in the quality of quantitative reconstructions of cross-sections of the human body-specifically, to improve the contrast of low-contrast tumours embedded in soft tissue. It was established in section 6.8 that there was information contained in the reconstruction within Ω_+ that was related to non-linear imperfections on the projection data, of the type described in section 6.7. The algorithm was successful in that a solution was converged upon whereby contributions within the region Ω_+ were negligible (in the sense of section 6.6). However, the solution for which the algorithm converged did not provide a significant improvement in tumour contrast within Ω_-.

8.3 PLETHORA OF VIEWS: CONSISTENCY-BASED RECOVERY OF IMPERFECTIONS

As the starting point for an original approach to exploiting the plethora of views idea, consistency of projections is examined in section 7.1 and in section 7.2 an extended definition is given, termed *spatial consistency*, for consistency of projections that incorporates the region of reconstruction. Included is an associated definition of *spatial inconsistency*. Based on these definitions, and *a priori* information regarding the region of support for the distribution of interest, techniques have been developed in section 7.4 for the recovery of inconsistencies contained on a limited range of projection angles. Once the inconsistencies have been recovered they can be removed from the given projection data to reduce their adverse effects on the quality of the reconstruction within the known support. It should be noted that the techniques presented in section 7.4 are not thought amenable for the recovery of the general class of inconsistencies found in projection data.

It was mentioned in section 1.5 that although X-ray CT presently provides detailed quantitative images of very high resolution, which for most intents and purposes are highly adequate, there remain several sources of error. One of the most serious is that of beam-hardening (described in section 1.4.1). Beam-hardening is known to produce pronounced "streaks" in the reconstruction close to high density objects [Garden *et al.*, 1989]. These effects are particularly severe when high-density objects are encountered that are elongated in a direction parallel to that of the X-ray projection paths. At these angles the effects of beam-hardening are most pronounced because of the quantity of high density material traversed by the X-rays. The techniques developed in section 7.4 lend themselves to the recovery of these type of inconsistencies introduced by such phenomena since the severest of the inconsistencies are confined to a narrow range of angles. Once the inconsistencies are recovered they may be compensated for in the measured data, thereby reducing the cause of distortion in quantitative and qualitative information contained in the reconstruction.

If missing data over a limited range of projection angles are viewed in the light of inconsistencies, then this method may be suitable for their partial recovery. There are a number of X-ray CT applications [Rangayyan *et al.*, 1985] for which data are not available over a limited range of view angles. These applications should be investigated as they may be prime candidates for the recovery techniques in section 7.4.

Inconsistencies in projection data are spread over a range of projection angles when considering the difference projections. The effective recovery of inconsistencies relies on the deconvolution of a 2-D error kernel. The performance of this deconvolution for the recovery of the inconsistencies is dependent on the 2-D error kernel used and the method of deconvolution. The simple method of deconvolution used here, the method of inverse filtering, is known to be noise sensitive. The algorithm outlined in section 7.4.3 would benefit from a more appropriate (with respect to noise sensitivity) 2-D deconvolution technique. Wiener filtering [Bates and McDonnell, 1989] is a prime candidate to replace the inverse filtering used but requires that the noise characteristics of the 2-D error kernel be known. Alternatives to the Wiener filter (which operates in Fourier space) are *subtractive deconvolution* techniques such as those based on the iterative "cleaning" of Högbom [1974]. These techniques [Tsao and Steinberg, 1988] operate in image space and therefore do not have the same sensitivity to noise.

A 2-D error kernel has been given in section 7.4.3 for the deconvolution of inconsistencies of the type demonstrated. The error kernel has been taken to be an average of the difference projections, when the inconsistencies reconstructed is a single point placed at $\xi' = 0$ on otherwise zero-valued projections, for all projection angles. What is required for effective recovery is an optimum 2-D error kernel that provides best performance in the recovery of the type of inconsistencies for which it will be used. Once an optimum error kernel is constructed for a particular reconstruction configuration, and with regard to the type of inconsistencies encountered, it need not be re-estimated.

REFERENCES

M. Abramowitz and I. A. Stegun. *Handbook of mathematical functions.* Dover Publications Inc., New York, 1965.

Keiiti Aki and Paul G. Richards. *Quantitative Seismology: Theory and Methods*, Vol. 1-2. Freeman, San Francisco, 1980.

M. D. Altschuler. Reconstruction of the global-scale three-dimensional solar corona. In G. T. Herman, editor, *Topics in applied physics. Volume 32. Image reconstruction from projections: implementation and applications*, pp. 105–145. Springer-Verlag, New York, 1979.

J. Ambrose. Computerised transverse axial scanning (tomography). II. Clinical applications. *British Journal of Radiology*, Vol. 46, pp. 1034, 1973.

A. H. Anderson and A. C. Kak. Digital ray tracing in two-dimensional refractive fields. *Journal of the Acoustic Society of America*, Vol. 72, No. 5, pp. 1593–1606, November 1982.

J. M. Arnold. Geometrical theories of wave propagation: a contemporary review. *Proceedings of the IEE*, Vol. 133, No. 2, pp. 165–188, April 1986.

Leon Axel, Peter H. Arger, and Robert A. Zimmerman. Applications of computerized tomography to diagnostic radiology. *Proceedings of the IEEE*, Vol. 71, No. 3, pp. 293–297, March 1983.

H. H. Barrett. The radon transform and its applications.

H. H. Barrett. Historical note on computed tomography. *Radiology*, Vol. 147, No. 1, pp. 172, April 1983.

R. H. T. Bates. Image reconstruction from a plethora of imperfect projections. Private communication, March 1988.

R. H. T. Bates. JWKB/Rayleigh-Gans (Born) inverse scattering approximation and reconstruction algorithm. *Inverse Problems*, Vol. 4, pp. 129–132, 1988.

R. H. T. Bates and M. J. McDonnell. *Image Restoration and Reconstruction.* Oxford University Press, Walton Street, Oxford OX2 6DP, 2nd edition, 1989.

R. H. T. Bates and G. C. McKinnon. Towards improving images in ultrasonic transmission tomography. *Australasian Physical Sciences in Medicine*, Vol. 2-3, No. 80, pp. 134–140, April 1979.

R. H. T. Bates and F. L. Ng. Polarisation-source formulation of electromagnetism and dielectric loaded waveguides. *Proceedings of the IEE*, Vol. 119, No. 11, pp. 1568–1574, November 1972.

R. H. T. Bates and T. M. Peters. Towards improvements in tomography. *New Zealand Journal of Science*, Vol. 14, pp. 883–896, December 1971.

R. H. T. Bates, W. M. Boerner, and R. G. Dunlop. An extended Rytov approximation and its significance for remote sensing and inverse scattering. *Optics Communications*, Vol. 18, No. 4, pp. 421–423, November 1976.

R. H. T. Bates, K. L. Garden, and T. M. Peters. Overview of computerized tomography with emphasis on future developments. *Proceedings of the IEEE*, Vol. 71, No. 3, pp. 356–372, March 1983.

R. H. T. Bates, V. A. Smith, and R. D. Murch. Managable multidimensional inverse scattering theory. *Physics Reports*, Vol. 201, No. 4, pp. 185–277, April 1991.

R. Bender, S. H. Bellman, and R. Gordon. Art and the ribosome: A preliminary report on the three-dimensional structure of individual ribosomes determined by an algebraic reconstruction technique. *Journal of Theoretical Biology*, Vol. 29, pp. 483–487, 1970.

J. G. Berryman. Stable iterative reconstruction algorithm for non-linear traveltime tomography. *Inverse Problems*, Vol. 6, No. 1, pp. 21–42, February 1990.

M. Bertero, C. De Mol, and E. R. Pike. Linear inverse problems with discrete data: II. Stability and regularisation. *Inverse Problems*, Vol. 4, pp. 573–594, 1988.

R. R. Bitmead and B. D. O. Anderson. Asymptotically fast solution of Toeplitz and related systems of linear equations. In *Large scale matrix problems*, pp. 103–116. North Holland, New York, 1981.

J. M. Blackledge, R. E. Burge, K. I. Hopcraft, and R. J. Wombell. Quantitative diffraction tomography: Part I. *Journal of Physics D: Applied Physics*, Vol. 20, No. 1, pp. 1–10, January 1987.

J. M. Blackledge, R. E. Burge, K. I. Hopcraft, and R. J. Wombell. Quantitative diffraction tomography: Part II. *Journal of Physics D: Applied Physics*, Vol. 20, No. 1, pp. 11–17, January 1987.

J. C. Bolomey, A. Izadnegahdar, L. Jofre, C. Pichot, G. Peronnet, and M. Solaimani. Microwave diffraction tomography for biomedical applications. *IEEE Transactions on Microwave Theory and Techniques*, Vol. MTT-30, No. 11, pp. 1798–2000, 1982.

L. J. Bond and W. N. Reynolds. Editorial, NDT: an integral part of engineering. *IEE Proceedings Part A*, Vol. 134, pp. 237–238, March 1987.

M. Born and E. Wolf. *Principles of optics*. Pergamon Press, Oxford, 6th edition, 1980.

R. N. Bracewell. Strip integration in radio astronomy. *Australian Journal of Physics*, Vol. 9, No. 2, pp. 198–217, 1956.

Ronald N. Bracewell. *The Fourier Transform and its Applications*. Electrical and Electronic Engineering. McGraw-Hill, London, 2nd edition, 1978.

R. N. Bracewell and A. C. Riddle. Inversion of fan-beam scans in radio astronomy. *The astrophysical journal*, Vol. 150, pp. 427–434, 1967.

T. F. Budinger, G. T. Gullberg, and R. H. Huesman. Emission computed tomography. In G. T. Herman, editor, *Topics in applied physics. Volume 32. Image reconstruction from projections: implementation and applications*, pp. 147–246. Springer-Verlag, New York, 1979.

M. H. Buonocore, W. R. Brady, and A. Macovski. A natural pixel decomposition for two-dimensional image reconstruction. *IEEE Transactions on Biomedical Engineering*, Vol. BME-28, pp. 69–78, 1981.

Y. Censor. Finite series-expansion reconstruction methods. *Proceedings of the IEEE*, Vol. 71, No. 3, pp. 409–419, March 1983.

Soyoung Cha and C. M. Vest. Tomographic reconstruction of strongly refracting fields and its application to interferometric measurement of boundary layers. *Applied Optics*, Vol. 20, No. 16, pp. 2787–2794, August 1981.

Lih Shyang Chen, G. T. Herman, R. A. Reynolds, and J. K. Udapa. Surface shading in the cuberille environment. *IEEE Transactions on Computer Graphics and Applications*, Vol. 5, No. 12, pp. 33–43, December 1985.

L. Chernov. *Wave propagation in a random medium*. Dover, New York, 1967.

W. K. Cheung and R. M. Lewitt. Modified Fourier reconstruction using shifted transform samples. *Physics in Medicine and Biology*, Vol. 36, No. 2, pp. 269–277, 1991.

Z. H. Cho, J. K. Chan, E. L. Hall, R. P. Kruger, and D. G. McCaughey. A comparative study of 3-D reconstruction algorithms with reference to number of projections and noise filtering. *IEEE Transactions on Nuclear Science*, Vol. 22, pp. 344–358, February 1975.

W. N. Christiansen and J. A. Warburton. The distribution of radio brightness over the solar disk at a wavelength of 21 centimetres. *Australian Journal of Physics*, Vol. 8, No. 4, pp. 474–486, 1955.

J. M. Coggins. Statistical investigations of multiscale image structure. *Proceedings of the SPIE*, Vol. 1808, pp. 145–158, 1992.

D. Colton and R. Kress. *Integral equation methods in inverse scattering theory*. Wiley, New York, 1983.

L. T. Cook, S. J. Dwyer, Solomon Batnitzky, and Kyo Rak Lee. A three-dimensional display system for diagnostic imaging applications. *IEEE Transactions on Computer Graphics and Applications*, pp. 13–19, August 1983.

A. M. Cormack. Representation of a function by its line integrals, with some radiological applications. *Journal of Applied Physics*, Vol. 34, No. 9, pp. 2722–2727, September 1963.

A. M. Cormack. Representation of a function by its line integrals, with some radiological applications. II. *Journal of Applied Physics*, Vol. 35, No. 10, pp. 2908–2912, October 1964.

S. Cornbleet. Geometrical optics reviewed: A new light on an old subject. *Proceedings of the IEEE*, Vol. 71, No. 4, pp. 471–502, April 1983.

Correspondent. The ART of the possible. *Nature New Biology*, Vol. 232, pp. 131, August 1971.

R. A. Crowther and A. Klug. ART and science or conditions for three-dimensional reconstruction from electron microscope images. *Journal of Theoretical Biology*, Vol. 32, pp. 199–203, 1971.

R. A. Crowther, D. J. DeRosier, and A. Klug. The reconstruction of a three-dimensional structure from projections and its application to electron microscopy. *Proceedings of the Royal Society of London*, Vol. 317, pp. 319–340, 1970.

L. J. Cutrona. Sythetic aperture radar. In P. S. Green, editor, *Radar Handbook*, chapter 23. McGraw-Hill, New York, 1970.

S. Dale, P. E. Edholm, L. G. Hellström, and S. Larson. Ectomography–a tomographic method for gamma camera imaging. *Physics in Medicine and Biology*, Vol. 30, No. 11, pp. 1237–1249, 1985.

B. L. K. Davey. *Advances in blind deconvolution*. PhD thesis, University of Canterbury, Christchurch, New Zealand, 1989.

D. J. DeRosier and A. Klug. Reconstruction of three dimensional structures from electron micrographs. *Nature*, Vol. 217, pp. 130–134, January 1968.

G. A. Deschamps. Ray techniques in electromagnetics. *Proceedings of the IEEE*, Vol. 60, No. 9, pp. 1022–1035, September 1972.

A. J. Devaney. A filtered backpropagation algorithm for diffraction tomography. *Ultrasonic Imaging*, Vol. 4, pp. 336–350, 1982.

A. J. Devaney. Generalized projection-slice theorem for fan beam diffraction tomography. *Ultrasonic Imaging*, Vol. 7, No. 3, pp. 264–275, July 1985.

Kris A. Dines and R. Jeffrey Lytle. Computerized geophysical tomography. *Proceedings of the IEEE*, Vol. 67, No. 7, pp. 1065–1073, July 1979.

R. A. Drebin, Loren Carpenter, and Pat Hanrahan. Volume rendering. *Computer Graphics*, Vol. 22, No. 4, pp. 65–74, August 1988.

P. Edholm. Tomogram reconstruction using an opticophotographic method. *Acta Radiologica Diagnosis*, Vol. 18, pp. 125–144, January 1977.

P. Edholm. *Linograms*. Medical Imaging Section, Department of Radiology, University of Pennsylvania, Philadelphia, 1987. Medical image processing group technical report.

P. Edholm and G. T. Herman. Linograms in image reconstruction from projections. *IEEE Transactions on Medical Imaging*, Vol. 6, No. 4, pp. 301–307, December 1987.

P. Edholm, G. T. Herman, and D. A. Roberts. *Implementation and evaluation of Image reconstruction from linograms*. Medical Imaging Section, Department of Radiology, University of Pennsylvania, Philadelphia, December 1987. Medical image processing group technical report, No. MIPG125.

P. Edholm, G. T. Herman, and D. A. Roberts. Image reconstruction from linograms : Implementation and evaluation. *IEEE Transactions on Medical Imaging*, Vol. 7, No. 3, pp. 239–246, September 1988.

M. Ein-Gal, D. Rosenfeld, and A. Macovski. The consistency of the shadow: An approach to preprocessing in computerized tomography. *Proceedings of the Stanford Conference on Computerized Tomography*, 1974.

S. A. Enright, S. M. Dale, V. A. Smith, R. D. Murch, and R. H. T. Bates. Towards solving the bent-ray tomographic problem. *Inverse Problems*, Vol. 8, No. 1, pp. 83–94, February 1992.

K. R. Erikson, F. J. Fry, and J. P. Jones. Ultrasound in medicine - A review. *IEEE Transactions on Sonics and Ultrasonics*, Vol. SU-21, No. 3, pp. 144, July 1974.

A. C. Evans, S. Marrett, L. Collins, and T. M. Peters. Anatomical-functional correlative analysis of the human brain using three-dimensional imaging systems. *Proceedings of the SPIE*, Vol. 1092, pp. 264–274, 1989.

L. B. Felsen. Novel ways for tracking rays. *Journal of the Optical Society of America A*, Vol. 2, No. 6, pp. 954–963, June 1985.

L. B. Felsen. Real spectra, complex spectra and compact spectra. *Journal of the Optical Society of America A*, Vol. 3, No. 4, pp. 486–496, April 1986.

F. S. Foster, M. Strban, and G. Austin. The ultrasound macroscope: Initial studies of breast tissue. *Ultrasonic Imaging*, Vol. 6, pp. 243–261, 1984.

G. Freider and G. T. Herman. Resolution in reconstructing objects from electron micrographs. *Journal of Theoretical Biology*, Vol. 33, pp. 189–211, October 1971.

F. J. Fry. Biological effects of ultrasound - A reveiw. *Proceedings of the IEEE*, Vol. 67, No. 12, pp. 604–619, December 1979.

H. Fuchs, Z. M. Kedem, and S. P. Uselton. Optimal surface reconstruction from planar contours. *Communications of the ACM*, Vol. 20, No. 10, pp. 693–702, October 1977.

N. T. Gaarder and G. T. Herman. Algorithm for reproducing objects from their X-rays. *Computer Graphics and Image Processing*, Vol. 1, No. 1, pp. 97–106, April 1972.

K. L. Garden. *An overview of computed tomography*. PhD thesis, University of Canterbury, Christchurch, New Zealand, 1984.

K. L. Garden, P. J. Bones, and R. H. T. Bates. From living being to medical image - bridging the dimensionality gap. *Australasian Physical & Engineering Sciences in Medicine*, Vol. 12, No. 4, pp. 186–204, 1989.

J. A. Garrett and P. H. Smithson. Conventional x-ray imaging. *IEE Proceedings Part A*, Vol. 134, No. 2, pp. 107–114, February 1987.

P. Gilbert. Iterative methods for the three-dimensional reconstruction of an object from projections. *Journal of Theoretical Biology*, Vol. 36, pp. 105–117, 1972.

G. H. Glover. Characterization of *in vivo* breast tissue by time-of-flight computed tomography. In M. Linzer, editor, *Ultrasonic Tissue Characterization II*, pp. 221–233. National Bureau of Standards, Spec. Publ. 525, Washington, DC 20234, April 1979.

J. W. Goodman. *Introduction to Fourier optics*. McGraw-Hill, San Franscisco, 1968.

R. Gordon. A tutorial on ART (Algebraic Reconstruction Techniques). *IEEE Transactions on Nuclear Science*, Vol. 21, pp. 78–93, 1974.

R. Gordon, R. Bender, and G. T. Herman. Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography. *Journal of Theoretical Biology*, Vol. 29, pp. 471–481, 1970.

J. F. Greenleaf. Computerized tomography with ultrasound. *Proceedings of the IEEE*, Vol. 71, No. 3, pp. 330–337, March 1983.

J. F. Greenleaf and R. C. Bahn. Clinical imaging with transmissive ultrasonic computerized tomography. *IEEE Transactions on Biomedical Engineering*, Vol. BME-28, No. 2, pp. 177–184, February 1981.

J. F. Greenleaf, S. A. Johnson, S. L. Lee, G. T. Herman, and E. H. Wood. Algebraic reconstruction of spatial distributions of acoustic absorption of tissues from their two-dimensional acoustic projections. *Acoustical Holography*, Vol. 5, pp. 591–603, 1974.

K. M. Hanson. CT reconstruction from limited projection angles. In *Proc. Applications of Optical Instrumentation in Medicine X, New Orleans, Proc. SPIE*, Vol. 347, New Orleans, May 19-23 1982.

P. B. Heffernan and R. H. T. Bates. Image reconstruction from projections. VI: Comparison of interpolation methods. *Optik*, Vol. 60, No. 2, pp. 129–142, 1982.

Gabor T. Herman. *Image reconstruction from projections*. Computer science and applied mathematics. Academic Press, London, 1980.

Gabor T. Herman. *Computerized reconstruction and 3-D imaging in medicine*. Medical Imaging Section, Department of Radiology, University of Pennsylvania, Philadelphia, February 1986. Medical image processing group technical report, No. MIPG108.

Gabor T. Herman. The MIPG perspective, in pictorial information systems in medicine. In K. H. Hoehne, editor, *Three-dimensional computer graphic display in medicine*, pp. 181–210. Springer-Verlag, Berlin, 1986.

G. T. Herman. Image reconstruction from a plethora of imperfect projections. Private communication, March 1988.

G. T. Herman and Hsun Kao Liu. Three-dimensional display of human organs from computed tomograms. *Computer Graphics and Image Processing*, Vol. 9, pp. 1–21, 1979.

G. T. Herman, A. Lent, and S. W. Rowland. ART: Mathematics and applications. A report on the mathematical foundations and on the applicability to real data of the algebraic reconstruction techniques. *Journal of Theoretical Biology*, Vol. 42, pp. 1–32, 1973.

G. T. Herman, A. V. Lakshminarayanan, and A. Naparstek. Convolution reconstruction techniques for divergent beams. *Comput. Biol. Med.*, Vol. 6, pp. 259–271, 1976.

J. A. Högbom. Aperture synthesis with a non-regular distribution of interferometer baselines. *Astronomy and Astrophysics Supplementary Series*, Vol. 15, No. 3, pp. 417–426, June 1974.

B. K. P. Horn. Density reconstruction using arbitrary ray-sampling schemes. *Proceedings of the IEEE*, Vol. 66, No. 5, pp. 551–562, May 1978.

B. K. P. Horn. Fan-beam reconstruction methods. *Proceedings of the IEEE*, Vol. 67, No. 12, pp. 1616–1623, December 1979.

G. N. Hounsfield. *A method and apparatus for examination of a body by radiation such as X or gamma radiation.* The Patent Office, London, England, 1972.

G. N. Hounsfield. Computerised transverse axial scanning (tomography). I. Description of system. *British Journal of Radiology*, Vol. 46, pp. 1016, 1973.

G. N. Hounsfield. Computed medical imaging. *Science*, Vol. 210, pp. 22–28, 1980.

M. E. C. Hull, J. H. Frazer, and R. J. Millar. Octree-based modelling of computed-tomography images. *Proceedings of the IEEE*, Vol. 137, No. 3, pp. 118, June 1990.

P. C. Jackson, H. Key, and E. M. Pitcher. A review of the display of combined images. *IEE Proceedings Part A*, Vol. 134, No. 2, pp. 175–178, February 1987.

B. D. James. Vector diffraction tomography using radon transform techniques in computer assisted electromagnetic imaging. Master's thesis, Department of Electrical Engineering and Computer Science, University of Illinois, Chicago, July 1986.

G. L. James. *Geometrical theory of diffraction for electromagnetic waves.* Peter Peregrinus Ltd, London, UK, 3nd edition, 1986.

R. J. Jaszczak. Tomographic radiopharmaceutical imaging. *Proceedings of the IEEE*, Vol. 76, No. 9, pp. 1079–1094, September 1988.

D. S. Jones. *The theory of electromagnetism.* Pergamon Press, New York, 1964.

D. S. Jones. *Acoustic and electromagnetic waves.* Clarendon Press, Oxford, 1986.

P. M. Joseph. An improved algorithm for reprojecting rays through pixel images. *IEEE Transactions on Medical Imaging*, Vol. 1, No. 3, pp. 192–186, 1982.

J. B. Keller. The inverse scattering problem in geometrical optics and the design of reflectors. *IRE Transactions on Antennas and Propagation*, pp. 146–149, April 1959.

J. B. Keller. Geometrical theory of diffraction. *Journal of the Optical Society of America*, Vol. 52, No. 2, pp. 116–130, February 1962.

J. B. Keller. Accuracy and validity of the Born and Rytov approximations. *Journal of the Optical Society of America*, Vol. 59, pp. 1003–1004, 1969.

J. B. Keller. One hundred years of diffraction theory. *IEEE Transactions on Antennas and Propagation*, Vol. 33, No. 2, pp. 1003–1004, February 1985.

W. I. Keyes. Radionuclide imaging. *IEE Proceedings Part A*, Vol. 134, No. 2, pp. 161–170, February 1987.

R. E. Kleinman, G. F. Roach, and P. M. van den Berg. Convergent Born series for large refractive indices. *Journal of the Optical Society of America A*, Vol. 7, No. 5, pp. 890–897, May 1990.

Morris Kline. An asymptotic solution of Maxwell's equations. *Communications on Pure and Applied Mathematics*, Vol. 5, pp. 225–262, 1951.

Y. S. Kwoh, I. S. Reed, and T. K. Truong. A generalized $|w|$-filter for 3-D reconstruction. *IEEE Transactions on Nuclear Science*, Vol. 24, pp. 1990–1998, 1977.

R. M. Lewitt. *Contributions to image reconstruction.* PhD thesis, University of Canterbury, Christchurch, New Zealand, 1977.

R. M. Lewitt. Aspects of the convolution method for image reconstruction from projections. *SPIE: Application of Optical Instrumentation in Medicine VII*, Vol. 173, pp. 271–278, 1979.

R. M. Lewitt. Ultra - fast convolution approximations for computerised tomography. *IEEE Transactions on Nuclear Science*, Vol. 26, pp. 2678–2681, 1979.

R. M. Lewitt. Reconstruction algorithms: Transform methods. *Proceedings of the IEEE*, Vol. 71, No. 3, pp. 390–408, 1983.

R. M. Lewitt. Multidimensional image representations using generalized Kaiser-Bessel window functions. *Journal of the Optical Society of America A*, Vol. 7, No. 10, pp. 1834, 1990.

R. M. Lewitt and R.H.T. Bates. Image reconstruction from projections: I: General theoretical considerations. *Optik*, Vol. 50, pp. 19–33, March 1978.

R. M. Lewitt, R. H. T. Bates, and T. M. Peters. Image reconstruction from projections: II: Modified back-projection methods. *Optik*, Vol. 50, No. 2, pp. 85–109, 1978.

T. Lo, M. N. Toksöz, S. Xu, and R. Wu. Ultrasonic laboratory tests of geophysical tomographic reconstruction. *Geophysics*, Vol. 53, No. 7, pp. 947–955, 1988.

A. K. Louis. Ghosts in tomography - the null space of the Radon transform. *Mathematical Methods in Applied Science*, Vol. 3, pp. 1–10, 1981.

R. Jeffrey Lytle and Kris A. Dines. Iterative ray tracing between boreholes for underground image reconstruction. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. GE-18, No. 3, pp. 234–240, July 1980.

H. A. McCann, J. C. Sharp, T. M. Kinter, C. N. McEwan, C. Barillot, and J. F. Greenleaf. Multidimensional ultrasonic imaging for cardiology. *Proceedings of the IEEE*, Vol. 76, No. 9, pp. 1063–1073, September 1988.

G. C. McKinnon. *Contributions to imaging*. PhD thesis, University of Canterbury, Christchurch, New Zealand, 1980.

G. C. McKinnon and R. H. T. Bates. A limitation on ultrasonic transmission tomography. *Ultrasonic Imaging*, Vol. 2, pp. 48–54, 1980.

F. A. Molinet. Geometrical theory of diffraction (GTD) Part I: Foundation of the theory. *IEEE Antennas and Propagation Society*, Vol. 29, No. 4, pp. 6–17, August 1987.

F. A. Molinet. Geometrical theory of diffraction (GTD) Part II: Extensions and future trends of the theory. *IEEE Antennas and Propagation Society*, Vol. 29, No. 4, pp. 6–17, October 1987.

B. M. Moores. Digital X-ray imaging. *IEE Proceedings Part A*, Vol. 134, No. 2, pp. 115–125, February 1987.

P. M. Morse and H. Feshbach. *Methods of theoretical physics*. McGraw-Hill, New York, 1953.

Ross Murch. *Inverse scattering and shape reconstruction*. PhD thesis, University of Canterbury, Christchurch, New Zealand, April 1990.

J. A. Newell. Medical images and automated interpretation. *Journal of Biomedical Engineering*, Vol. 10, pp. 555–561, November 1988.

Stephen J. Norton. Computing ray trajectories between two points: A solution to the ray linking problem. *Journal of the Optical Society of America A*, Vol. 4, pp. 1919–1922, 1987.

Stephen J. Norton and Melvin Linzer. Correcting for ray refraction in velocity and attenuation tomography : A perturbation approach. *Ultrasonic Imaging*, Vol. 4, pp. 201–233, 1982.

P. A. O'Brien. *Mon. Not. R. Astr. Soc.*, Vol. 113, pp. 597, 1953.

Hatice Orun-Ozturk, H. Joel Trussell, and M. Reha Civanlar. Detection of artifacts in CT images using the consistency of the projection data. *Medical Imaging*, Vol. 767, pp. 361–366, 1987.

Peyma Oskoui-Fard and Henry Stark. Tomographic image reconstruction using the theory of convex projections. *IEEE Transactions on Medical Imaging*, Vol. 7, No. 1, pp. 45–58, March 1988.

T. M. Peters. Algorithms for fast back- and re-projection in computed tomography. *IEEE Transactions on Nuclear Science*, Vol. 28, No. 4, pp. 3641–3647, August 1981.

T. M. Peters and R. M. Lewitt. Computed tomography with fan-beam geometry. *Journal of Computer Assisted Tomography*, Vol. 1, pp. 429–436, 1977.

T. M. Peters, J. A. Clark, G. B. Pike, C. Henri, L. Collins, D. Leksell, and O. Jeppsson. Stereotactic neurosurgery planning on a personal-computer-based work station. *Journal of Digital Imaging*, Vol. 2, No. 2, pp. 75–81, 1989.

W. H. Press, B. P. Flanney, S. A. Teukolsky, and W. T. Vetterling. *Numerical recipes: The art of scientific computing*. Cambridge Univ. Press, Cambridge, 1987.

J. Radon. Uber die bestimmung von function durch ihre integralwerte langsgwisser mannigfaltigkeiten (on the determination of functions from their integrals along certain manifolds). *Berichte Sachsische Akadamie der Wissenschaften (Leipzig) Mathematische-Physische Klasse*, Vol. 69, pp. 262–277, 1917.

G. N. Ramachandran and A. V. Lakshminarayanan. Three-dimensional reconstruction from radiographs and electron micrographs: Application of convolutions instead of Fourier transforms. *Proceedings of the National Academy of Science USA*, Vol. 68, No. 9, pp. 2236–2240, September 1971.

Rangaraj Rangayyan, Atam Prakash Dhawan, and Richard Gordon. Algorithms for limited-view computed tomography : an anotated bibliography and a challenge. *Applied Optics*, Vol. 24, No. 23, pp. 4000–4012, December 1985.

R. A. Robb. Introduction to three-dimensional biomedical imaging. In R. A. Robb, editor, *Three-dimensional biomedical imaging*, Vol. I. CRC Press, Boca Raton, Florida, 1985.

R. A. Robb. X-ray computed tomography: Advanced systems and applications in biomedical research and diagnosis. In R. A. Robb, editor, *Three-dimensional biomedical imaging*, Vol. I. CRC Press, Boca Raton, Florida, 1985.

R. A. Robb. X-ray computed tomography: Implementation and applications. In R. A. Robb, editor, *Three-dimensional biomedical imaging*, Vol. I. CRC Press, Boca Raton, Florida, 1985.

R. A. Robb and C. Barillot. Interactive display and analysis of 3-D medical images. *IEEE Transactions on Medical Imaging*, Vol. 8, No. 3, pp. 217, September 1989.

R. A. Robb, E. A. Hoffman, L. J. Sinak, L. D. Harris, and E. L. Ritman. High speed three-dimensional x-ray computed tomography: The spatial reconstructor. *Proceedings of the IEEE*, Vol. 71, No. 3, pp. 308–319, March 1983.

B. S. Robinson and J. F. Greenleaf. Computerized ultrasound tomography. In R. A. Robb, editor, *Three-dimensional biomedical imaging*, Vol. II. CRC Press, Boca Raton, Florida, 1985.

J. H. Rose. Exterior reconstruction of a three dimensional scatterer. *Wave Motion*, Vol. 6, pp. 149–154, 1984.

S. W. Rowland. Computer implementation of image reconstruction formulas. In G. T. Herman, editor, *Topics in applied physics. Volume 32. Image reconstruction from projections: implementation and applications*, pp. 9–79. Springer-Verlag, New York, 1979.

Y. Z. Ruan and L. B. Felsen. Reflection and transmission of beams at a curved surface. *Journal of the Optical Society of America A*, Vol. 3, No. 4, pp. 566–579, April 1986.

P. K. Sahoo, S. Soltani, and A. K. C. Wong. A survey of thresholding techniques. *Computer Vision, Graphics, and Image Processing*, Vol. 41, pp. 233–260, 1988.

F. Santosa, Y. H. Pao, W. W. Symes, and C. Holland, editors. *Proceedings of International Conference on Inverse Problems of Acoustic and Elastic Waves, Cornell University*, Philadelphia, 1984. SIAM.

A. L. Scherzinger, R. A. Belgam, P. L. Carson, C. R. Meyer, J. V. Sutherland, F. L. Bookstein, and T. M. Silver. Assessment of ultrasonic computed tomography in symptomatic breast patients by discriminant analysis. *Ultrasound in Medicine and Biology*, Vol. 15, No. 1, pp. 21–28, 1989.

M. Schoenberger. Special issue on seismic signal processing. *Proceedings of the IEEE*, Vol. 72, pp. 1235–1412, 1984.

Hermann Schomberg. An improved approach to reconstructive ultrasound tomography. *Journal of Physics D: Applied Physics*, Vol. 11, pp. L181–L185, 1978.

J. S. Schreiman, J. J. Gisvold, J. F. Greenleaf, and R. C. Bahn. Ultrasound transmission tomography of the breast. *Radiology*, Vol. 150, pp. 523–530, 1984.

Henry J. Scudder. Introduction to computer aided tomography. *Proceedings of the IEEE*, Vol. 66, No. 6, pp. 628–637, June 1978.

R. L. Sengbush. *Seismic Exploration Methods*. International Human Resources Development Corporation, Boston, 1983.

L. A. Shepp and B. F. Logan. The Fourier reconstruction of a head section. *IEEE Transactions on Nuclear Science*, Vol. 21, pp. 21–43, June 1974.

Malcolm Slaney and A. C. Kak. *Imaging with diffraction tomography*, Vol. TR-EE 85-5. School of Electrical Engineering, Purdue University, West Lafayette, Indiana 47907, 1985.

Malcolm Slaney, A. C. Kak, and L. E. Larson. Limitations of imaging with first-order diffraction tomography. *IEEE Transactions on Microwave Theory and Techniques*, Vol. MTT-32, No. 8, pp. 860–874, 1984.

B. J. Smith and R. Martin. Computers in ultrasonic NDT. *IEE Proceedings Part A*, Vol. 134, No. 3, pp. 239–247, March 1987.

P. R. Smith, T. M. Peters, and R. H. T. Bates. Image reconstruction from finite numbers of projections. *Journal of Physics A: Math., Nucl. Gen.*, Vol. 6, pp. 361–382, March 1973.

Henry J. Stark. Direct Fourier reconstruction in computer tomography. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 29, No. 2, pp. 237–245, April 1981.

Henry J. Stark, J. W. Woods, Indraneel Paul, and Rajesh Hingorani. An investigation of computerized tomography by direct Fourier inversion and optimum interpolation. *IEEE Transactions on Biomedical Engineering*, Vol. BME-28, No. 7, pp. 496–505, July 1981.

David Guan Hock Tan. *Implementable multi-dimensional inverse scattering theory*. PhD thesis, University of Canterbury, Christchurch, New Zealand, February 1988.

D. G. H. Tan, J. X. Qu, and R. H. T. Bates. Image reconstruction from projections: VIII: Effects of finite resolution and sampling of individual projections. *Optik*, Vol. 73, No. 1, pp. 25–29, 1986.

S. I. Tetel'baum. About a method of obtaining volume images with the help of x-rays. *Bulletin of Kiev Polytechnic Institute*, Vol. 22, pp. 154–160, 1957.

D. O. Thompson and D. E. Chementi. *Review of progress in quantitative nondestructive evaluation*, Vol. 5A. Plenum, New York, 1985.

J. Tsao and B. D. Steinberg. Reduction of sidelobe and speckle artefacts in microwave imaging : The CLEAN technique. *IEEEtAP*, Vol. 36, No. 4, pp. 543, April 1988.

Jayaram K. Udupa. Display of 3D information in discrete 3D scenes produced by computerized tomography. *Proceedings of the IEEE*, Vol. 71, No. 3, pp. 420–431, 1983.

Jayaram K. Udupa. Simulation of surgical procedures using computer graphics. In *Applications of computer graphics to radiology, surgery, and neuropsychology*, pp. 3–15. Medical Imaging Section, Department of Radiology, University of Pennsylvania, Philadelphia, February 1986. Medical image processing group technical report, No. MIPG109.

R. J. Urick. *Underwater sound for engineers*. McGraw-Hill, New York, 1975.

J. VanBladel. *Electromagnetic Fields*. Hemisphere Publishing Corp., London, 1985.