Acta Numerica (1999), *pp.* 1–000

Numerical Methods in Tomography

Frank Natterer
Institut für Numerische
und instrumentelle Mathematik
Universität Münster
Einsteinstrasse 62
D-48149 Münster, Germany
E-mail: nattere@math.uni-muenster.de

CONTENTS

In this article we review the image reconstruction algorithms used in tomography. We restrict ourselves to the standard problems in the reconstruction of function from line or plane integrals as they occur in X-ray tomography, nuclear medicine, magnetic resonance imaging, and electron microscopy. Non-standard situations, such as incomplete data, unknown orientations, local tomography, discrete tomography are not dealt with. Nor do we treat nonlinear tomographic techniques such as impedance, ultrasound, and near infrared imaging.

1. Introduction

By "Tomography" we mean a technique for imaging 2D cross sections of 3D objects. It is derived from the greek word $\tau o\mu o s$ = slice. Tomographic techniques are used in radiology and in many branches of science and technology.

1.1. The basic example

In the simplest case, let us consider an object whose attenuation coefficient with respect to X-rays at the point x is $f(x)$. We scan the cross section by a thin X-ray beam L of unit intensity. The intensity past the object is

$$e^{-\int_L f(x)\,dx} \; .$$

This intensity is measured, providing us with the line integral

$$g(L) = \int_L f(x)dx \ . \tag{1.1}$$

The problem is to compute f from g.

In principle this problem has been solved by Radon (1917). Let L be the straight line $x \cdot \theta = s$ where $\theta = (\cos\varphi, \sin\varphi)^T$ and $s \in \mathbb{R}^1$. Then, (1.1) can be written as

$$g(\theta, s) = \int_{x\cdot\theta=s} f(x)dx = (Rf)(\theta, s) \ . \tag{1.2}$$

R is known as the Radon transform. Radon's inversion formula reads

$$f(x) = \frac{1}{4\pi^2} \int_0^{2\pi} \int_{\mathbb{R}^1} \frac{g'(\theta, s)}{x \cdot \theta - s} ds d\varphi \tag{1.3}$$

where g' is the derivative of g with respect to s and $\theta = (\cos\varphi, \sin\varphi)^T$. In principle, (1.3) solves our problem. So, why do we write an article on tomography?

First, inversion formulas such as (1.3) do not exist in all cases. For instance, in emission tomography, the mathematical model involves weighted line integrals, which in general do not admit explicit inversion. Also, even if explicit inversion is possible, it is not obvious how to turn an inversion formula such as (1.3) into an efficient and accurate algorithm. Many problems concerning sampling and discretization arise. Often not all of the data in an explicit inversion formula can be measured. Finally, (1.1) is a prime example for many imaging techniques, and a proper understanding of the inversion of (1.1) is a necessary prerequisite for the understanding of more complicated problems.

1.2. An annotated bibliography

As an introduction into the many applications of (1.1) we recommend Deans (1983). For the mathematical background see Natterer (1986). Herman (1980) gives a good introduction into the mathematical and algorithmic problems. For the practitioner we recommend Kak and Slaney (1987) which provides an in-depth coverage of the electrical engineering point of view.

1.3. Outline of the paper

In section 2 we describe the most important algorithm in tomography, namely the filtered backprojection algorithm. It not only is the working horse for present day's tomography, but also serves as model for the algorithms in future imaging devices such as the 3D algorithms described in section 3. Since many imaging problems can be described by large linear sparse systems of equations iterative methods suggest themselves, see section 4. Algorithms exploiting rotational symmetry of the imaging devices are described in section 5. In section 6 we deal with algorithms which work exclusively in Fourier space and which have the potential to outperform the filtered backprojection algorithm in speed.

1.4. Prerequisites

The reader needs only a rudimentary knowledge of numerical analysis (interpolation, quadrature, linear systems), functional analysis (linear operators, distributions), Fourier analysis (Fourier transform, Fourier series, inversion, convolution, fast Fourier transform (FFT)) and sampling theory (Shannon's sampling theorem for band-limited functions).

2. The filtered backprojection algorithm

In this section we give a detailed description of the most important algorithm in 2D tomography. The discrete implementation depends on the scanning geometry, i.e. the way the data is sampled.

This algorithm is essentially a numerical implementation of the Radon inversion formula (1.3). However, a different approach avoiding singular integrals is simpler. We describe this approach for the nD Radon transform

$$g(\theta, s) = \int\limits_{x\cdot\theta = s} f(x)dx = (Rf)(\theta, s)$$

where f is a function in $\mathbb{R}^n$ and $\theta \in S^{n-1}$, $s \in \mathbb{R}^1$. Let

$$(R^*g)(x) = \int\limits_{S^{n-1}} g(\theta, \theta \cdot x)d\theta$$

be the backprojection operator and let V, v be functions such that $V = R^*v$. Mathematically, R^* is simply the Hilbert space adjoint of the Radon transform R. It is easy to see that

$$V \star f = R^*(v \star g) \tag{2.1}$$

where the convolution on the left hand side is in R^n, while the convolution on the right hand side is a 1D convolution with respect to the second variable:

$$\int\limits_{\mathbb{R}^n} V(x-y)f(y)dy = \int\limits_{S^{n-1}} \int\limits_{\mathbb{R}^1} v(x\cdot\theta - s)g(\theta, s)dsd\theta\,. \tag{2.2}$$

The idea is to choose V as an approximation to Dirac's δ-function. Then,

$V \star f$ is close to f. The interrelation between V, v is easily described in terms of the Fourier transform. Denoting with the same symbol "$\wedge$" the nD Fourier transform

$$\hat{V}(\xi) = (2\pi)^{-n/2} \int\limits_{\mathbb{R}^n} e^{-ix\cdot\xi} V(x)dx \ , \quad \xi \in \mathbb{R}^n$$

and the 1D Fourier transform

$$\hat{v}(\theta, \sigma) = (2\pi)^{-1/2} \int\limits_{\mathbb{R}^1} e^{-is\sigma} v(\theta, s)ds \ , \quad \sigma \in \mathbb{R}^1$$

we have

$$\hat{V}(\xi) = 2(2\pi)^{(n-1)/2} |\xi|^{1-n} \hat{v}(\theta, |\xi|) \ ,$$

see Natterer (1986). By $|\xi|$ we mean the euclidean length of $\xi \in \mathbb{R}^n$.

The choice of V determines the spatial resolution of the reconstruction algorithm. We use the notion of resolution from sampling theory, see Jerry (1977). We give a short account of some basic facts of sampling theory. A function f in $\mathbb{R}^n$ is said to be band-limited with bandwidth Ω, or simply Ω-band-limited, if $\hat{f}(\xi) = 0$ for $|\xi| > \Omega$. An example for $n = 1$ is the sinc function

$$\mathrm{sinc}(x) = \frac{\sin x}{x}$$

which has bandwidth 1. Obiously, sinc (Ωx) has bandwidth Ω. Ω-band-limited functions are capable of representing details of size $2\pi/\Omega$ but no smaller ones. This becomes clear simply by looking at the graph of sinc.

In tomography the functions we are dealing with are usually of compact support. Such functions can't be strictly band-limited, unless being identically zero. Hence we require the functions only to be essentially Ω-band-limited, meaning that $\hat{f}(\xi) \sim 0$ for $|\xi| > \Omega$ in an appropriate sense, see Natterer (1986).

A reconstruction method in tomography is said to have resolution $2\pi/\Omega$ if it reconstructs reliably essentially Ω-band-limited functions.

For strictly Ω-band-limited functions we have the following propositions, which hold also, with very good accuracy, for essentially Ω-band-limited functions.

1. If f is Ω-band-limited and $h \le \pi/\Omega$ (Nyquist-condition), then f is uniquely determined by the values $f(hk)$, $k \in \mathbb{Z}$, and

$$f(x) = \sum_k f(hk) \mathrm{sinc}\ \Omega(x - hk) \ .$$

2. If f is Ω-band-limited and $h \leq \pi/\Omega$, then

$$\hat{f}(\xi) = h \sum_k e^{-i\xi hk} f(hk)$$

3. If f_1, f_2 are Ω-band-limited and $h \leq \pi/\Omega$, then

$$\int f_1(x) f_2(x) dx = h \sum_k f_1(hk) f_2(hk) \ .$$

Returning to the construction of a reconstruction algorithm with resolution $2\pi/\Omega$ we have to determine V such that

$$\hat{V}(\xi) = (2\pi)^{-n/2} \hat{\phi}\left(\frac{|\xi|}{\Omega}\right)$$

where $\hat{\phi}$ is a filter with the property

$$\hat{\phi}(\sigma) \begin{cases} \sim & 1 \quad , \quad |\sigma| \leq 1 \ , \\ = & 0 \quad , \quad |\sigma| > 1 \ . \end{cases}$$

This follows from the formula $\hat{\delta} = (2\pi)^{-n/2}$. This means that for the filter function v we must have

$$\hat{v}(\theta, \sigma) = \frac{1}{2} (2\pi)^{1/2-n} |\sigma|^{n-1} \hat{\phi}\left(\frac{|\sigma|}{\Omega}\right) \ . \tag{2.3}$$

Examples are the ideal low-pass

$$\hat{\phi}(\sigma) = \begin{cases} 1 & , \quad |\sigma| \geq \Omega \ , \\ 0 & , \quad |\sigma| > \Omega \ , \end{cases}$$

the cos filter

$$\hat{\phi}(\sigma) = \begin{cases} \cos \frac{\sigma\pi}{2} & , \quad |\sigma| \geq 1 \\ 0 & , \quad |\sigma| \geq 1 \end{cases}$$

and the filter

$$\hat{\phi}(\sigma) = \begin{cases} \mathrm{sinc}(\sigma\pi/2) & , \quad |\sigma| \leq 1 \\ 0 & , \quad |\sigma| > 1 \end{cases}$$

which has been introduced in tomography by Shepp and Logan (1974). The corresponding functions v for $n = 2$ are

$$v_\Omega(s) = \frac{\Omega^2}{4\pi^2} u(\Omega s) \ , \quad u(s) = \mathrm{sinc}(s) - \frac{1}{2}\left(\mathrm{sinc}(\frac{s}{2})\right)^2$$

for the ideal low pass,

$$v_\Omega(s) = \frac{\Omega^2}{8\pi^2}\left(u(s\Omega + \frac{\pi}{2}) + u(s\Omega - \frac{\pi}{2}\right)$$

with the same function u, and

$$v_\Omega(s) = \frac{\Omega^2}{2\pi^2} u(\Omega s) \ , \quad u(s) = \begin{cases} \pi/2 - s \sin s & , \quad s \neq \pm\frac{\pi}{2} \ , \\ \frac{1}{\pi} & , \quad s = \pm\frac{\pi}{2} \end{cases}$$

for the Shepp-Logan filter. More filters can be found in Chang and Herman (1980).

The integral on the right hand side of (2.2) has to be approximated by a quadrature rule. We have to distinguish between several ways of sampling $g = Rf$.

2.1. Parallel geometry in the plane

In this case the 2D Radon transform $g(\theta, s) = (Rf)(\theta, s)$ is sampled for $\theta = \theta_j = (\cos\varphi_j, \sin\varphi_j)^T$, $\varphi_j = \pi j/p$, $j = 0, \ldots, p-1$ and $s = s_\ell = \ell\rho/q$, $\ell = -q, \ldots, q$. Here ρ is the radius of the reconstruction region, i.e. we assume $f(x) = 0$ for $x \in \mathbb{R}^2$, $|x| \geq \rho$. This means that the measured rays come in p parallel bundles with directions evenly distributed over 180°, each bundle consisting of $2q+1$ equispaced lines. This was the scanning geometry of the first commercial scanner for which Hounsfield received the Nobel prize in 1979. This geometry has been replaced by more efficient ones in present day's scanners (see below), but it is still used in scientific and technical imaging.

We evaluate the integral in (2.2) by the trapezoidal rule:

$$(V \star f)(x) \sim \frac{2\pi\rho}{pq} \sum_{j=0}^{p-1} \sum_{\ell=-q}^{q} v_\Omega(x \cdot \theta_j - s_\ell) g(\theta_j, s_\ell) \tag{2.4}$$

The accuracy of this approximation can be assessed by sampling theory, according to which the trapezoidal rule for an inner product is exact provided the stepsize h satisfies the Nyquist criterion, i.e. $h \leq \pi/\Omega$ where Ω is the bandwidth of the factors in the inner product. In our case the first factor is $v_\Omega(x \cdot \theta_j - s)$ (as a function of s) which has bandwidth Ω. The second factor is $g(\theta_j, s)$ (again as a function of s). This is our data and does not, in general, have finite bandwidth. At this point we have to make an assumption.

We assume f to be essentially band-limited with essential bandwidth Ω. The nD Fourier transform of f and the 1D Fourier transform Rf (with respect to the second variable) are interrelated by

$$(Rf)^\wedge(\theta, \sigma) = (2\pi)^{(n-1)/2} \hat{f}(\sigma\theta) \ . \tag{2.5}$$

This is the famous (and easy to prove) "projection" or "central slice" theorem of computerized tomography. In the present context we need it only to deduce that f and $g = Rf$ have the same (essential) bandwidth. Thus the s-integral in (2.2) is accurately represented by the ℓ-sum in (2.4) provided that the stepsize $h = \rho/q$ in that sum satisfies the Nyquist criterion

$h \leq \pi/\Omega$. In other words,

$$q \geq \frac{1}{\pi}\rho\Omega \,. \tag{2.6}$$

The condition for the number p of directions which makes the j-sum in (2.4) a good approximation for the θ-integral in (2.2) is less obvious. Based on Debye's asymptotic relation for the Bessel functions one can show that the essential bandwidth of Rf as a function of φ, $\theta = (\cos\varphi, \sin\varphi)^T$, is $\Omega\rho$, see Natterer (1986). The stepsize h for the θ-integral being π/p the Nyquist criterion requires $\frac{\pi}{p} \leq \frac{\pi}{\Omega\rho}$, i.e.

$$p \geq \Omega\rho \,. \tag{2.7}$$

(2.6), (2.7) are the conditions for a good accuracy in (2.4), assuming f to be zero outside the ball of radius ρ and essentially band-limited with bandwidth Ω.

The double sum in (2.4) has to be evaluated for each reconstruction point x. This leads to an unbearable complexity. This complexity can be reduced by introducing the function

$$h(\theta_j, s) = \frac{\rho}{q} \sum_{\ell=-q}^{q} v_\Omega(s - s_\ell) g(\theta_j, s_\ell) \,.$$

Then, (2.4) reads

$$(V \star f)(x) \sim \frac{2\pi}{p} \sum_{j=0}^{p-1} h(\theta_j, x \cdot \theta_j) \,.$$

This requires only a simple sum for each reconstruction point x, at the expense of an additional interpolation in the second argument of h. In most cases linear interpolation suffices (but nearest neighbour does not!). This leads us to the filtered backprojection algorithm.

Algorithm 1 (Filtered backprojection algorithm for standard parallel geometry.)

Data: The values $g_{j,\ell} = g(\theta_j, s_\ell)$, $j = 0, \ldots, p-1$, $\ell = -q, \ldots, q$, g is the 2D Radon transform of f.

Step 1: For $j = 0, \ldots, p-1$ carry out the discrete convolution

$$h_{j,k} = \frac{\rho}{q} \sum_{\ell=-q}^{q} v_\Omega(s_k - s_\ell) g_{j,\ell}, \; k = -q, \ldots, q \,.$$

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

$$f_{FB}(x) = \frac{2\pi}{p} \sum_{j=0}^{p-1} ((1-\vartheta)h_{j,k} + \vartheta h_{j,k+1})$$

where $k = k(j,x)$ and $\vartheta = \vartheta(j,x)$ are determined by

$$t = \frac{x \cdot \theta_j}{\rho/q} \ , \quad k = \lfloor t \rfloor \ , \ \vartheta = t - k \ .$$

Result: f_{FB} is an approximation to $f(x)$.

The algorithm depends on the parameters p, q and on the choice of v_Ω. It is designed to reconstruct a function f with support in $|x| < \rho$ and with essential bandwidth Ω i.e. the spatial resolution of the algorithm is $2\pi/\Omega$. The conditions (2.6), (2.7) should be satisfied. A few remarks are in order.

1. The condition (2.6) has to be strictly satisfied. Otherwise the s integral in (2.2) is not even approximately equal to the ℓ sum in (2.4), leading to unacceptable errors.
2. If (2.7) is not satisfied, the reconstruction is still accurate for $|x| \leq p/\Omega < \rho$.
3. Filter functions v_Ω whose "kernel sum"
$$\sum_\ell v_\Omega(s_\ell)$$
does not vanish should not be used, see Natterer and Faridani (1990).
4. Usually linear interpolation in step 2 is sufficient. However, for difficult functions f - e.g. functions containing large objects at the boundary of the reconstruction region - linear interpolation generates visible artefacts. In that case an oversampling procedure similar to the one of Algorithm 2 below is advisable. Alternatively one may use the circular harmonic algorithm from section 5.
5. The algorithm needs $\mathrm{O}(p)$ operations for each reconstruction point. Algorithms with lower complexity (such as $\mathrm{O}(\log p)$) can be obtained either by Fourier reconstruction, see section 6, or by the fast backprojection algorithm in section 2.5.
6. The conditions (2.6), (2.7) suggest to take $p = \pi q$. This much debated condition is usually not complied with in radiological applications, where p is chosen considerably smaller. This is due to the special requirements in radiological imaging.

2.2. The interlaced parallel geometry

It is well known (see e.g. Kruse (1989)) that the data in the standard parallel geometry is redundant: If p is even, then one can omit each $g(\theta_j, s_\ell)$ with

$\ell + j$ odd without impairing the resolution. Deriving algorithms which use only the remaining "interlaced" data (i.e. those $g(\theta_j, s_\ell)$ for which $j+\ell$ even) is fairly subtle. What happens is the following. If in the ℓ sum in (2.4) every second term is droped, the sum no longer approximates the corresponding s integral in (2.2). Miraculously the large quadrature error cancels when the j sum in (2.4) is computed. This means that success depends entirely on a subtle interplay between different directions. This interplay is disrupted by the interpolation procedure in step 2 of Algorithm 1. There are two ways out. The first one is to avoid interpolation alltogether by using circular harmonic algorithms, see section 5. The second one is to make the interpolation more accurate, for instance by oversampling. This leads to an algorithm which has the structure of a filtered backprojection algorithm.

Algorithm 2 (Filtered backprojection algorithm for parallel interlaced geometry.)

Data: The values g_{θ_j, s_ℓ}, $j = 0, \ldots, p-1$, $\ell = -q, \ldots, q-1$, $\ell + j$ even. g is the 2D Radon transform of f. p has to be even.

Step 1: Choose an integer $M > 0$ sufficiently large ($M = 16$ will do) and compute for $j = 0, \ldots, p-1$

$$h_{j,k} = 2\frac{\rho}{q} \sum_{\substack{\ell=-q \\ \ell+j \text{even}}}^{q} v_\Omega(s_k/M - s_\ell) g(\theta_j, s_\ell) \ , \quad k = -Mq, \ldots, Mq \ .$$

Step 2: For each reconstruction point x, compute

$$f_{FB}(x) = \frac{2\pi}{p} \sum_{j=0}^{p-1} ((1-\vartheta)h_{j,k} + \vartheta h_{j,k+1})$$

where $k = k(j,x)$, $\vartheta = \vartheta(j,x)$ are determined by

$$t = M\frac{x \cdot \theta_j}{\rho/q} \ , \quad k = \lfloor t \rfloor \ , \quad \vartheta = t - k \ .$$

Result: $f_{FB}(x)$ is an approximation to $f(x)$.

Note that the difference between this algorithm and Algorithm 1 is that it needs only one half of the data but produces the same image quality. We study the various assumptions underlying this algorithm.

1. The algorithm is designed to reconstruct a function f supported in $|x| < \rho$ with essential bandwidth Ω. The sampling conditions (2.6), (2.7) have to satisfied. In contrast to Algorithm 1, oversatisfying these conditions may lead to artefacts. Thus the algorithm should be used only if (2.6), (2.7) are satisfied with equality, i.e. for $p = \pi \cdot q$.

2. Only filters v with a smooth transition from non-zero to zero values should be used. The reason is that the additional filtering of the interpolation step is not present in Algorithm 2.

2.3. Standard fan beam geometry

This is the most widely used scanning geometry. It is generated by a source moving on a concentric circle of radius $r > \rho$ around the reconstruction region $|x| < \rho$, with opposite detectors being read out in small time intervals (third generation scanner). Equivalently we may have a fixed detector ring with only the source moving around (fourth generation scanner). Denoting the angular position of the source by β and the angle between a measured ray and the central ray by α ($\alpha > 0$ if the ray, viewed from the source, is left of the central ray), then fan beam scanning amounts to sampling the function

$$\begin{aligned} g(\beta,\alpha) &= (Rf)(\theta, r\sin\alpha)\ , \\ \theta &= \begin{pmatrix} \cos(\beta+\alpha) \\ \sin(\beta+\alpha) \end{pmatrix} \end{aligned} \tag{2.8}$$

at the points $\beta = \beta_j = j\Delta\beta$, $\Delta\beta = 2\pi/p$, $j = 0,\ldots,p-1$, $\alpha = \alpha_\ell = (\ell + d)/\Delta\alpha$, $\ell = -q,\ldots,q$. Here, q is chosen so as to cover the whole reconstruction region $|x| < \rho$ with rays. d is the detector offset which is either 0 or $\pm\Delta\alpha/4$.

First we derive the fan beam analogue of (2.1). We only have to put $\varphi = \beta + \alpha$, $s = r\sin\alpha$ to map fan beam coordinates to parallel coordinates as used in (2.1). The region $[0, 2\pi) \times (-\frac{\pi}{2}, \frac{\pi}{2})$ of the β-α-plane is mapped in a one-to-one fasion onto the domain $[0, 2\pi) \times (-r, r)$ in the φ-s-plane, and we have

$$\frac{\partial(\varphi, s)}{\partial(\beta,\alpha)} = \begin{vmatrix} 1 & 1 \\ 0 & r\cos\alpha \end{vmatrix} = r\cos\alpha\ .$$

Thus (2.2) in the new coordinates reads

$$(V \star f)(x) = r \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} v(x\cdot\theta - r\sin\alpha) g(\beta,\alpha)\cos\alpha d\alpha d\beta$$

with θ as in (2.8). Discretizing the integral by the trapezoidal rule yields

$$(V \star f)(x) \sim r\Delta\alpha\Delta\beta \sum_{j=0}^{p-1} \sum_{\ell=-q}^{q} v_\Omega(x\cdot\theta(\beta_j+\alpha_\ell) - r\sin\alpha_\ell) g(\beta_j,\alpha_\ell)\cos\alpha_\ell\ . \tag{2.9}$$

This is the fan beam analogue of (2.4) and defines a reconstruction algorithm for fan beam data. One can show that for this algorithm to have resolution $2\pi/\Omega$ one has to satisfy

$$\Delta\beta \leq \frac{r+\rho}{\rho}\frac{\pi}{\Omega r}\ , \qquad \Delta\alpha \leq \frac{\pi}{\Omega r}\ , \tag{2.10}$$

see Natterer (1993).

As in the parallel case, an algorithm based on (2.9) needs $O(pq)$ operations for each reconstruction point. Reducing this to $O(p)$ is possible here, too, but this is not as obvious as in the parallel case. We first establish a relation for the expression $x \cdot \theta(\varphi) - s$ in (2.2). Let $b = r\theta(\beta)$ be the source position, and let γ be the angle between $x - b$ and $-b$. We take γ positive if x, viewed from the source b, lies to the left of the central ray, i.e. we have

$$\cos\gamma = \frac{(b-x)\cdot b}{|b-x|\cdot|b|} \quad , \quad sgn(\gamma) = -sgn\left(x\cdot\theta(\varphi+\frac{\pi}{2})\right)$$

where $\theta(\varphi) = (\cos\varphi, \sin\varphi)^T$. Let y be the orthogonal projection of x onto the ray with fan beam coordinates β, α. Then, $|x\cdot\theta(\varphi) - s| = |x - y|$. Considering the rectangular triangle xyb we see that $|x - y| = |b - x|\sin|\gamma - \alpha|$, hence

$$|x\cdot\theta(\varphi) - s| = |b - x|\sin|\gamma-\alpha|\ .$$

Our filters v_Ω possess the homogeneity property

$$v_\Omega(t\sigma) = t^{-2}v_{t\Omega}(\sigma)\ . \tag{2.11}$$

Thus,

$$v_\Omega(x\cdot\theta(\varphi) - s) = |b-x|^{-2}v_{\Omega|b-x|}(\sin(\gamma-\alpha))\ .$$

Using this in (2.2) we obtain

$$(V\star f)(x) = r\int_0^{2\pi}|b-x|^{-2}\int_{-\pi/2}^{\pi/2} v_{\Omega|b-x|}(\sin(\gamma-\alpha))g(\beta,\alpha)\cos\alpha d\alpha d\beta\ .$$

Here, $b = r\theta(\beta)$, and γ is independent of α. Unfortunately, the α integral has to be evaluated for each x since the subscript $\Omega|b-x|$ depends on x. In order to avoid this we make an approximation: We replace $\Omega|b-x|$ by Ωr. This is not critical as long as $|b-x| \sim |b|$, i.e. as long as $\rho \ll r$. Fortunately, in most scanners $r \sim 3\rho$, and this is sufficient for the approximation to be satisfactory. However, if ρ is only slightly smaller than r, problems arise.

Upon the replacement of $v_{\Omega|b-x|}$ by $v_{\Omega r}$ we obtain

$$(V\star f)(x) \sim r\int_0^{2\pi}|b-x|^{-2}\int_{-\pi/2}^{\pi/2} v_{\Omega r}(\sin(\gamma-\alpha))g(\beta,\alpha)\cos\alpha d\alpha d\beta\ .$$

The α integral can now be precomputed as a function of γ and β, yielding an algorithm with the structure of a filtered backprojection algorithm.

Algorithm 3 (Filtered backprojection algorithm for parallel standard fan beam geometry.)

Data: The values $g_{j,\ell} = g(\beta_j, \alpha_\ell)$, $j = 0, \ldots, p-1$, $\ell = -q, \ldots, q$. g is the function in (2.8).

Step 1: For $j = 0, \ldots, p-1$ carry out the discrete convolutions

$$h_{j,k} = \Delta\alpha \sum_{\ell=-q}^{q} v_{\Omega r}(\sin(\alpha_k - \alpha_\ell)) g_{j\ell} \cos\alpha_\ell \ , \quad k = -q, \ldots, q \ .$$

Step 2: For each reconstruction point x, compute the discrete weighted backprojection

$$f_{FB}(x) = r\Delta\beta \sum_{j=0}^{p-1} |b_j - x|^{-2} \left((1-\vartheta)h_{j,k} + \vartheta h_{j,k+1}\right)$$

where $k = k(j,x)$ and $\vartheta = \vartheta(j,x)$ are determined by

$$\gamma = \pm \arccos \frac{(b_j - x)\cdot b_j}{|b_j - x||b_j|}$$

the sign being the one of $-x \cdot b_j$ and $b_j = r\theta(\beta_j)$,

$$t = \frac{\gamma}{\Delta\alpha} \ , \quad k = \lfloor t \rfloor \ , \quad \vartheta = t - k \ .$$

Result: $f_{FB}(x)$ is an approximation to $f(x)$.

The algorithm as it stands is disigned to reconstruct a function f with support in $|x| < \rho$ which essentially band-limited with bandwidth Ω from fan beam data with the source on a circle of radius $r > \rho$. The remarks following Algorithm 1 apply by analogy. In particular the conditions (2.10) have to be satisfied. For $r \sim \rho$ and with dense parts of the object close to the boundary of the reconstruction region, problems are likely to occur.

2.4. Linear fan beam geometry

Here, the detector positions within a fan with vertex b are evenly spaced on the line perpendicular to b. We need the explicit form of the inversion formula mainly for the derivation of the FDK algorithm in 3D cone beam tomography in the next section. With g the function in (2.8), the sampled data is

$$g_{j,\ell} \ = \ g(\beta_j, \alpha_\ell) \ , \quad j = 0, \ldots, p-1 \ , \quad \ell = -q, \ldots, q$$

$$\beta_j = \frac{2\pi}{p} j \,, \quad \alpha_\ell = \tan(y_\ell / r) \,, \quad y_\ell = (\ell + d)\Delta y \,.$$

The coordinates β, y are related to the parallel coordinates φ, s in the representation $x \cdot \theta(\varphi) = s$ of the rays by

$$\varphi = \beta + \arctan \frac{y}{r} \,, \qquad s = \frac{ry}{(r^2 + y^2)^{1/2}} \,. \tag{2.12}$$

Hence,

$$\frac{\partial(\varphi, s)}{\partial(\beta, y)} = \frac{r^3}{(r^2 + y^2)^{3/2}} \,.$$

Substituting β, y for φ, s in (2.2) leads to

$$(V \star f)(x) = \int_0^{2\pi} \int_{\mathbb{R}^1} v(x \cdot \theta(\varphi) - s) g(\beta, y) \frac{r^3 dy d\beta}{(r^2 + y^2)^{3/2}}$$

where (2.12) has to be inserted for φ, s. As in the standard fan beam case a direct implementation of this formula results in an algorithm whose complexity is not competitive. Again we can circumvent this problem by exploiting the homogeneity properties of v. A lengthy but elementary computation shows that

$$x \cdot \theta(\varphi) - s = c(z - y) \,,$$

$$c = \frac{r - x \cdot \theta(\beta)}{(y^2 + r^2)^{1/2}} \,, \quad z = \frac{r x \cdot \theta(\beta + \pi/2)}{r - x \cdot \theta(\beta)} \,.$$

From (2.11) it follows that

$$v_\Omega(c(z - y)) = c^{-2} v_{c\Omega}(z - y) \,,$$

yielding

$$\begin{aligned} (V \star f) &= \int_0^{2\pi} \int_{\mathbb{R}^1} c^{-2} v_{c\Omega}(z - y) g(\beta, y) \frac{r^3 dy d\beta}{(r^2 + y^2)^{1/2}} \\ &= r^3 \int_0^{2\pi} \frac{1}{(r - x \cdot \theta(\beta))^2} \int_{\mathbb{R}^1} v_{c\Omega}(z - y) g(\beta, y) \frac{dy}{(r^2 + y^2)^{1/2}} d\beta \,. \end{aligned}$$

As in the standard fan beam case we make the approximation $c \sim 1$. Again this is justified if $\rho \ll r$, e.g. $r \geq 3\rho$. Then,

$$(V \star f)(x) \sim \int_{S^1} \frac{r^2}{(r - x \cdot \theta)^2} \int_{\mathbb{R}^1} v_\Omega \left(\frac{r x \cdot \theta_\perp}{r - x \cdot \theta} - y \right) g(\theta, y) \frac{r dy}{(r^2 + y^2)^{1/2}} d\theta \,.$$

where $\theta_\perp(\varphi) = \theta(\varphi + \pi/2)$. Defining

$$h(\theta, z) = \int_{\mathbb{R}^1} v_\Omega(z-y) g(\theta, y) \frac{r dy}{(r^2+y^2)^{1/2}} \tag{2.13}$$

this can be written as

$$(V \star f)(x) \sim \int_{S^1} h\left(\theta, \frac{r x \cdot \theta^\perp}{r - x\cdot\theta}\right) d\theta \ . \tag{2.14}$$

The implementation of (2.13), (2.14) can now be done exactly as in the standard case, leading to a filtered backprojection algorithm which needs $O(p)$ operations for each reconstruction point x.

2.5. *Fast backprojection*

The backprojection (i.e. Step 2 in Algorithms 1-3) is the most time consuming part of the filtered backprojection algorithm. The filtering or convolution step (Step 1 in Algorithms 1-3) requires in principle the same number of operations, but this can easily be reduced drastically either by cutting off the filter v_Ω or by using the fast Fourier transform (FFT).

The backprojection consists is the evaluation of the sums

$$f(x) = \sum_{j=0}^{p-1} g(\theta_j, x\cdot\theta_j)$$

on a $p \times p$ grid. This is the simplest case of Algorithm 1, the resolution of the image being adjusted to the number of views p according to the sampling theorem. Nilsson (1997) suggested a devide and conquer strategy for doing this with $O(p^2 \log p)$ operations, as opposed to the $O(p^3)$ operations of a direct evaluation. Suppose $p = 2^m$.

Step 1: For $j = 0, 1, 2, \ldots, p-1$ compute

$$f_j^1(x) = g(\theta_j, x\cdot\theta_j) \ .$$

Since f_j^1 is constant along the lines $x\cdot\theta_j = s$ it suffices to compute $f_j^1(x)$ at $2p$ points x. We need $p\cdot 2p$ operations.

Step 2: For $j = 0, 2, 4, \ldots, p-2$ compute

$$f_j^2(x) = f_j^1(x) + f_{j+1}^1(x) \ .$$

Since f_j^1, f_{j+1}^1 are constant along the lines $x\cdot\theta_j = s$, $x\cdot\theta_{j+1} = s$, respectively, f_j^2 is almost constant along the lines $x\cdot\tilde\theta_j = s$ where $\tilde\theta_j = (\theta_j + \theta_{j+1})/\sqrt{2}$. Hence it suffices to compute f_j^2 only for a few, say 2, points on each of those lines. This means that we have to evaluate f_j^2 at $4p$ points, requiring $\frac{p}{2} \cdot 4p$ operations.

Step 3: For $j = 0, 4, 8, \ldots, p-4$ compute

$$f_j^3(x) = f_j^2(x) + f_{j+2}^2(x) \ .$$

With the same reasoning as in step 1 and 2 we find that it suffices to compute $f_j^3(x)$ for only for $8p$ points, requiring $\frac{p}{4} \cdot 8p$ operations.

Proceeding in this fashion we arrive in step m at the approximation f_j^m to f, which has to be evaluated at $2^m p$ points. Hence the number of operations in step 1 to m is

$$p \cdot 2p + \frac{p}{2} \cdot 4p + \frac{p}{4} \cdot 8p + \cdots + \frac{p}{2^{m-1}} \cdot 2^m p = mp^2$$

and this is $\mathrm{O}(p^2 \log p)$. Of course this derivation is heuristic, and we simply ignored the necessary interpolations and approximations. But practical experience demonstrates that such an algorithm can be made work.

3. 3D reconstruction formulas

Algorithms for 3D tomography are still under development. The main problem is that the data entering explicit inversion formulas are usually not available. Thus the main task in 3D tomography is the derivation of inversion formulas which use only the data measured by a specific imaging device. It is clear that these formulas are taylored to the imaging device. In this section we restrict ourselves to the derivation of exact or approximate inversion formulas. The implementation in a discrete setting can be done along the lines of the 2D algorithms.

3.1. Inversion of the 3D Radon transform

Let $g = Rf$, R the 3D Radon transform, be given on $S^2 \times \mathbb{R}^1$. Using (2.1) for $n = 3$ leads directly to a filtered backprojection algorithm, exactly as in the 2D case. Introducing spherical coordinates φ, ψ on S^2, i.e.

$$\theta = \theta(\varphi, \psi) = \begin{pmatrix} \sin\psi & \cos\varphi \\ \sin\psi & \sin\varphi \\ \multicolumn{2}{c}{\cos\psi} \end{pmatrix} , \qquad 0 \le \varphi < 2\pi \ , \quad 0 \le \psi \le \pi \ ,$$

(2.1) reads

$$(V \star f)(x) = \int_0^{2\pi} \int_0^{\pi} h(\theta, x \cdot \theta) \sin\psi d\psi d\varphi \tag{3.1}$$

where $\theta = \theta(\varphi, \psi)$ and

$$h(\theta, t) = \int_{\mathbb{R}^1} g(\theta, s) v(t - s) ds \, .$$

Once h is computed, the evaluation of (3.1) requires the computation of a 2D integral for each reconstruction point. This is prohibitive in real world applications.

Fortunately we can exploit the structure of the 3D Radon transform as the composition of two 2D Radon transforms. Putting

$$k_\varphi(s, t) = \int_0^\pi h(\theta(\vartheta, \psi), s\cos\psi + t\sin\psi)\sin\psi d\psi$$

we can rewrite (3.1) as

$$(V \star f)(x) = \int_0^{2\pi} k_\varphi(x_3, x_1\cos\varphi + x_2\sin\varphi) d\varphi \, . \tag{3.2}$$

The last two formulas are essentially 2D backprojections. They can be evaluated exactly as described in the previous section. After having precomputed h and k the final reconstruction step (3.2) requires only a 1D integral for each reconstruction point. This algorithm is reminescent of the two-stage algorithm of Marr et al. (1981) which has been developed for magnetic resonance imaging (MIR), except that the convolution steps are not present.

3.2. The FDK approximate formula

This is the most widely used algorithm for cone beam tomography with the source running on a circle. It is well known that this inversion problem is highly unstable. But practical experience with the FDK formula is nevertheless quite encouraging.

The function which is sampled in cone beam tomography with the source on a circle is

$$g(\theta, y) = \int_{\mathbb{R}^1} f(r\theta + ty) dt \, , \qquad y \in \theta^\perp$$

where θ is a direction vector in the $x_1 - x_2$-plane, $\theta = (\cos\varphi, \sin\varphi, 0)^T$. $\theta^\perp$ is the subspace orthogonal to θ, while $\theta_\perp$ (see below) is the vector $(-\sin\varphi, \cos\varphi, 0)^T$ perpendicular to θ. As usual we assume $f = 0$ outside $|x| \leq \rho$ where $\rho < r$.

The FDK formula is an ingenious adaption of the 2D inversion formula of section 2.4 to 3D. Consider the plane $\pi(\theta, x)$ through $r\theta$ and x which intersects $\theta^\perp$ in a line parallel to the $x_1 - x_2$-plane. Compute in this plane for each θ the contribution to (2.14). Finally, integrate all these contributions over S^1, disregarding that those contributions come from different planes.

The necessary computations are unpleasant, but the result is fairly simple. Based on (2.14),

$$(V \star f)(x) \sim \int\limits_{S^1} \frac{r^2}{(r - x\cdot)^2} \int\limits_{R^1} v_\Omega(u - u') g(\theta, u', z) \frac{r du' d\theta}{\sqrt{r^2 + u'^2 + z^2}} \tag{3.3}$$

where

$$u = \frac{r}{r - x \cdot \theta} x \cdot \theta_\perp \quad , \qquad z = \frac{r}{r - x \cdot \theta} x_3$$

and u', z are coordinates in $\theta^\perp$, i.e. $g(\theta, u', z)$ stands for $g(\theta, y)$ with $y = (-u' \sin\varphi, u' \cos\varphi, z)^T$. The implementation of (3.3) leads to a reconstruction algorithm of the filtered backprojection type. The reconstructions computed with the FDK formula (3.3) are - understandably - quite good for flat objects, i.e. if f is non-zero only close to the $x_1 - x_2$-plane in which the source runs. If this is not the case then exact formula using more data such as Grangeat's formula, see below, have to be used.

3.3. Grangeat's formula

Grangeat's formula requires sources on a curve C with the following property: Each plane meeting $\text{supp}(f)$ contains at least one source. This condition is obviously not satisfied for C a plane circle for which the FDK approximation has been derived.

The data for Grangeat's formula is

$$g(c, \theta) = \int\limits_{\mathbb{R}^1} f(c + t\theta) dt \ , \quad c \in C \ , \quad \theta \in S^2 \ .$$

The condition on the source curve means that for each x with $f(x) \neq 0$ and each $\theta \in S^2$ there exists a source $c = c(x, \theta) \in C$ such that $x \cdot \theta = c(x, \theta) \cdot \theta$.

The gist of Grangeat's inversion is a relation between g and the 3D Radon transform Rf of f. This relation reads (Grangeat (1991))

$$\frac{\partial}{\partial s}(Rf)(\theta, s)|_{s = x\cdot\theta} = \int\limits_{S^2 \cap \theta^\perp} \frac{\partial}{\partial \theta} g(c(x, \theta), \omega) d\omega \tag{3.4}$$

where $\frac{\partial}{\partial\theta}$ stands for the derivative in the direction $\theta \in S^2$, acting on the second argument. For this to make sense we have to extend g to all of $C \times \mathbb{R}^3$ by using the above definition not only for $\theta \in S^2$, but for all of $\mathbb{R}^3$. This is equivalent to extending g by homogeneity of degree -1 in the second argument. With help of the 3D inversion formula

$$f(x) = -\frac{1}{8\pi^2}\int\limits_{S^2} \frac{\partial^2}{\partial s^2}(Rf)(\theta f)(\theta, s)|_{s=x\cdot\theta}\, d\theta$$

for the 3D Radon transform, Grangeat's formula leads immediately to an inversion procedure for the data g. Related inversion formulas for cone beam tomography have been derived by Tuy (1983), Gelfand and Goncharov (1987). For the details of the implementation see Defrise and Clack (1995).

3.4. Orlov's inversion formula

This formula inverts the ray transform

$$(Pf)(\theta, y) = \int\limits_{\mathbb{R}^1} f(y + t\theta)dt\ , \quad y \in \theta^\perp\ , \quad \theta \in S^2 \tag{3.5}$$

which comes up e.g. in 3D emission tomography (PET, Defrise et al. (1989)). If θ is restricted to a plane, then we have simply the Radon transform in this plane, and we can reconstruct f in that plane by any of the methods in the previous section. In practice $g = Pf$ is measured for $\theta \in S_0^2$ where $S_0^2 \subseteq S^2$. In Orlov's formula (Orlov (1976)), S_0^2 is a spherical zone around the equator, i.e.

$$S_0^2 = \{\theta(\varphi, \vartheta) : \vartheta_-(\varphi) \le \vartheta_+(\varphi)\ , \quad 0 \le \varphi \le 2\pi\}$$

where φ, ϑ are the spherical coordinates of $\theta(\varphi, \vartheta) = (\cos\varphi\cos\vartheta, \sin\varphi\cos\vartheta, \sin\vartheta)^T$ and $-\frac{\pi}{2} < \vartheta_- < \vartheta_+ < \frac{\pi}{2}$. Then,

$$\begin{aligned} f(x) &= \Delta \int\limits_{S_0^2} h(\theta, x - (x\cdot\theta)\theta)d\theta\ , \\ h(\theta, x) &= \frac{1}{4\pi^2}\int\limits_{\theta^\perp} \frac{g(\theta, x-y)}{|y|\ell(\theta, y)}dy \end{aligned} \tag{3.6}$$

where Δ is the Laplacian acting on x and $\ell(\theta, y)$ is the length of the intersection of S_0^2 with the plane spanned by $0, \theta, y \in \mathbb{R}^3$. The first formula of (3.6) is - up to Δ - a backprojection, while the second one a convolution in $\theta^\perp$. Thus an implementation of (3.6) is again a filtered backprojection algorithm.

P can also be inverted by the Fourier transform. We have

$$(Pf)^\wedge(\theta, \xi) = (2\pi)^{1/2}\hat{f}(\xi)\ , \quad \xi \in \theta^\perp \tag{3.7}$$

where "$\wedge$" denotes the $(n-1)$-dimensional Fourier transform in $\theta^\perp$ on the left hand side and the Fourier transform in $\mathbb{R}^n$ on the right hand side.

Assume that $S_0^2 \subseteq S^2$ satisfies the Orlov condition: Every equatorial circle of S^2 meets S_0^2. Note that the set S_0^2 - the spherical zone - we used above in Orlov's formula satisfies this condition. From (3.7) it follows that f is uniquely determined by $(Pf)(\theta, \cdot)$ for $\theta \in S_0^2$ under the Orlov condition. Namely if $\xi \in \mathbb{R}^n$ is arbitrary, then Orlov's condition says that there exists $\theta \in S_0^2 \cap \xi^\perp$, and $\hat{f}(\xi)$ is determined from (3.7).

3.5. Colsher's inversion formula

Assume again that $g = Pf$ is known for $\theta \in S_0^2 \subseteq S^2$.

We want to derive an inversion procedure similar to the one in section 2.1. With the backprojection

$$(P^*g)(x) = \int_{S_0^2} g(\theta, x - (x \cdot \theta)\theta) d\theta$$

we have again

$$V * f = P^*(v * g)$$

provided that $V = P^* v$. Again the convolutions on the left hand right hand side have different meanings. Explicitly this reads

$$(V * f)(x) = \int_{S_0^2} \int_{\theta^\perp} v(\theta, x - (\theta \cdot \theta)\theta - y) g(\theta, y) dy d\theta \tag{3.8}$$

which corresponds to (2.2). As in (2.2) we express the relationship $V = P^* v$ in Fourier space, obtaining

$$\hat{V}(\xi) = (2\pi)^{1/2} \frac{1}{|\xi|} \int_{S_0^2 \cap \xi^\perp} \hat{v}(\theta, \xi) d\theta \ ,$$

see Colsher (1980). In order to get an inversion formula for P we have to determine v such that $V = \delta$ or $\hat{V} = (2\pi)^{-3/2}$, i.e.

$$\int_{S_0^2 \cap \xi^\perp} \hat{v}(\theta, \xi) d\theta = (2\pi)^{-2} |\xi| \ . \tag{3.9}$$

A solution $\hat{v}$ independent of θ is

$$\hat{v}(\theta, \xi) = \frac{|\xi|}{(2\pi)^2 |S_0^2 \cap \xi^\perp|}$$

where $|S_0^2 \cap \xi^\perp|$ is the length of $S_0^2 \cap \xi^\perp$. For the spherical zone S_0^2 from section 3.4 with $\vartheta_- = \vartheta_0$, $\vartheta_+ = \vartheta_0$, ϑ_0 a constant with $0 < \vartheta_0 < \frac{\pi}{2}$, Colsher computed $\hat{h}$ explicitly. With $\xi_3 = |\xi| \cos\vartheta$ be obtained

$$\hat{v}(\theta,\xi) = \frac{|\xi|}{(2\pi)^2} \begin{cases} \frac{1}{2\pi} \,, & |\vartheta| \geq \vartheta_0 \,, \\ \frac{1}{4 \arcsin(\sin\vartheta_0 / \sin\vartheta)} \,, & |\vartheta| \leq \vartheta_0 \,. \end{cases} \tag{3.10}$$

Filters such as the Colsher filter (3.10) do not have small support. This means that g in (3.8) has to be known in all of $\theta^\perp$. Often g is only available in part of $\theta^\perp$ (truncated projections). Let us choose

$$\hat{v}(\theta,\xi) = (2\pi)^{-2} \frac{a(\theta)\cdot\xi}{4\vartheta_0} \tag{3.11}$$

where $\theta = \theta(\varphi,\vartheta)$ from section (3.4) and $a(\theta) = (-\sin\varphi, \cos\varphi, 0)$ is a horizontal unit vector. Since (3.11) is constant in the vertical direction, v is a δ-function in the vertical variable. Hence the integral on the right hand side of (3.8) reduces to an integral over horizontal lines in $\theta^\perp$, making it possible to handle truncated projections. Unfortunately, (3.11) does not quite satisfy $\hat{V} = (2\pi)^{-3/2}$, i.e. it does not provide an exact inversion. Instead we only have

$$\hat{V}(\xi) = (2\pi)^{-3/2} \begin{cases} \vartheta/\vartheta_0 \,, & 0 \leq \vartheta \leq \vartheta_0 \,, \\ 1 \,, & \vartheta_0 \leq \vartheta \leq \pi/2 \end{cases}$$

where $|\xi_3| = |\xi'| \tan\vartheta$, $|\xi'| = \sqrt{\xi_1^2 + \xi_2^2}$. This is close to $\hat{V} = (2\pi)^{-3/2}$ if ϑ_0 is small. In this case reconstruction from truncated projections is possible, at least approximately.

3.6. Conical tilt geometry

In electron microscopy (Frank (1992)) one is faced with the case

$$S_0^2 = \{\theta(\varphi,\psi) : \psi = \psi_0\}$$

for some ψ_0 where $\theta(\varphi,\psi) = (\cos\varphi \sin\psi, \sin\varphi \sin\psi, \cos\psi)$. S_0^2 does not satisfy Orlov's condition, and (3.9) cannot be satisfied since $S_0^2 \cap \xi^\perp = \phi$ for some ξ. In that case we put

$$\hat{v}(\theta,\xi) = \begin{cases} \frac{|\xi|}{(2\pi)^2 |S_0^2 \cap \xi^\perp|} \,, & |S_0^2 \cap \xi^\perp| \neq 0 \,, \\ 0 \,, & \text{otherwise} \,. \end{cases}$$

With this choice of v, (3.8) is the minimal norm solution of $Pf = g$. A proper discretization along the lines of section 2.1 leads to the weighted backprojection algorithm of electron microscopy, see Frank (1992).

4. Iterative methods

If exact inversion formulas are not available, iterative methods are the algorithms of choice. But even if exact inversion is possible, iterative methods may be preferable due to their simplicity, versatility and ability to handle constraints and noise.

Iterative methods are usually applied to discrete versions of the reconstruction problem. These discrete versions are obtained either by starting out from discrete models - as in the EM algorithm below - or by a projection method - in the tomographic community "series expansion method", Censor (1981). This means that the unknown function f is written as

$$f = \sum_{\ell=1}^{N} f_\ell B_\ell$$

with certain basis functions B_ℓ. With g_i the i-th measurement, the measurement process being linear, we obtain the linear system

$$g_i = \sum_{\ell=1}^{N} a_{i\ell} f_\ell \ , \quad i = 1, \ldots, M \tag{4.1}$$

for the expansion coefficients f_ℓ, the matrix elements $a_{i\ell}$ being the i-th measurement for the Basis function B_ℓ. In tomography we always have $a_{i\ell} \geq 0$. Also, the matrix $(a_{i\ell})$ is typically sparse. Often B_ℓ is the characteristic function of pixels or voxels. Recently smooth radially symmetric functions with small support (the "blobs" of Lewitt (1992), Marabini et al. (1998)) have been used. Blobs have several advantages over pixel or voxel based functions. Due to the radial symmetriy it is easier to apply the Radon transform (or any of the other integral transforms) to B_ℓ, making it easier to set up the linear system (4.1). The smoothness of the B_ℓ prevents the "checkerboard" effect (i.e. the visual appearance of the pixels or voxels in the reconstruction) and does part of the necessary filtering a smoothing.

The linear system (4.1) may be overdetermined ($M > N$) or underdetermined ($M < N$), consistent or inconsistent. Useful iterative methods must be able to handle all these cases.

4.1. ART (algebraic reconstruction technique)

This is an extension of the Kaczmarz (1937) method for solving linear systems. It has been introduced in imaging by Gordon et al. (1970). We describe it in a more general context. We consider the linear system

$$A_j f = g_j \ , \qquad j = 0, \ldots, p-1 \tag{4.2}$$

where $A_j : H \to H_j$ are bounded linear operators from the Hilbert space H into the Hilbertspace H_j. With $C_j : H_j \to H_j$ a positive definite operator we define an iteration step $f^k \to f^{k+1}$ as follows:

$$\begin{aligned} f^{k,0} &= f^k \ , \\ f^{k,j+1} &= f^{k,j} + \omega A_j^* C_j^{-1}(g_j - A_j f^{k,j}) \ , \quad j = 0, \ldots, p-1 \\ f^{k+1} &= f^{k,p} \ . \end{aligned} \tag{4.3}$$

If $C_j = A_j^* A_j$ (provided A_j is onto) and $\omega = 1$, then $f^{k,j}$ is the orthogonal projection of $f^{k,j-1}$ onto the affine subspace $A_j f = g_j$. For $\dim(H_j) = 1$ this is the original Kaczmarz method. Other special cases are the Landweber method ($p = 1$, $C_1 = I$) and fixed-block ART ($\dim(H_j)$ finite, C_j diagonal, see Censor and Zenios (1997)). It is clear that there are many ways to apply (4.3) to (4.1), and we will make use of this freedom to our advantage.

One can show (Censor et al. (1983), Natterer (1986)) that (4.3) converges provided that

$$f^0 \in \sum_{j=0}^{p-1} \text{range } (A_j)$$

and $0 < \omega < 2$. This is reminiscent of the SOR-theory of numerical analysis. In fact we have $f^k = Au^k$ where u^k is the k-th SOR iterative for the linear system $AA^*u = g$ with

$$A = \begin{pmatrix} A_0 \\ \vdots \\ A_{p-1} \end{pmatrix} \quad , \quad g = \begin{pmatrix} g_0 \\ \vdots \\ g_{p-1} \end{pmatrix} \ .$$

If (4.2) is consistent, ART converges to the solution of (4.2) with minimal norm in H.

Plain convergence is useful, but we can say much more about the qualitative behaviour and the speed of convergence by exploiting the special structure of the image reconstruction problems at hand. With R the Radon transform in $\mathbb{R}^n$ we can put

$$H = L_2(|x| < 1) \ , \quad H_j = L_2(-1, +1; w^{1-n})$$

$$(A_j f)(s) = (Rf)(\theta_j, s)$$

where w is the weight function $(1 - s^2)^{1/2}$ and $\theta_j \in S^{n-1}$. One can show that the subspaces

$$\mathcal{C}_m = \langle C_{m,1}, \ldots, C_{m,p} \rangle \ ,$$

$$C_{m,j}(x) = C_m^{n/2}(x \cdot \theta_j) \ ,$$

C_m^λ the Gegenbauer polynomials of degree m (Abramowitz and Stegun

(1970)) are invariant subspaces of the iteration (4.3). This has been discovered by Hamaker and Solmon (1978). Thus it suffices to study the convergence on each subspace $\mathcal{C}_m$ separately. The speed of convergence depends drastically on ω and - surprisingly - on the way the directions θ_j are ordered. We summerize the findings of Hamaker and Solmon (1978), Natterer (1986) for the 2D case where $\theta_j = (\cos\varphi_j, \sin\varphi_j)$.

1. Let the φ_j be ordered consecutively, i.e. $\varphi_j = j\pi/p$, $j = 0, \ldots, p-1$. For ω large (e.g. $\omega = 1$) convergence on $\mathcal{C}_m$ is fast for $m < p$ large and slow for m small. This means that the high frequency parts of f (such as noise) are recovered first, while the overall features of f become visible only in the later iterations. For ω small (e.g. $\omega = 0.05$) the opposite is the case.

This explains why the first ART iterates for $\omega = 1$ look terrible and why surprisingly small relaxation factors (e.g. $\omega = 0.05$) are used in tomography.

2. Let φ_j be distributed at random in $[0, \pi]$ and $\omega = 1$. Then the convergence is fast on all $\mathcal{C}_m$, $m < p$. The same is true for more sophisticated arrangements of the φ_j, such as for $p = 18$ (Hamaker and Solmon (1978))

$$0, 9, 14, 5, 11, 3, 16, 7, 13, 2, 10, 17, 4, 8, 15, 1, 6, 12$$

or similarly for $p = 30$ in Herman and Meyer (1993).

The practical consequence is that a judicious choice of directions may well speed up the convergence tremendously. Often it suffices to do only 1-3 steps if the right ordering is chosen. This has been observed also for the EM iteration (see below).

Note that the $\mathcal{C}_m$ with $m \geq p$ are irrelevant since they describe those details in f which cannot be recovered from p projections because they are below the resolution limit. This is a result of the resolution analysis in Natterer (1986).

4.2. EM (expectation maximation)

The EM algorithm for solving the linear system $Af = g$ reads

$$f^{(k+1)} = f^{(k)} A^* \frac{g}{Af^{(k)}}, \qquad k = 0, 1, \ldots \ . \tag{4.4}$$

Multiplications and divisions in this formula are understood component wise. It is derived from a statistical model of image reconstruction, see Shepp and Vardi (1982). The purpose is to compute a minimizer of the log likelihood function

$$\ell(f) = \sum_i (g_i \log(Af)_i - (Af)_i) \ . \tag{4.5}$$

The convergence of (4.4) to a minimizer of (4.5) follows from the general EM theory in Dempster et al. (1977). Not that (4.4) preserves positivity if $a_{i\ell} \geq 0$.

The pure EM method (4.4) produces unpleasant images which contain too much detail. Several remedies have been suggested. An obvious one is to stop the iteration early, typically after 12 steps. One also can perform a smoothing step after each iteration (EMS-algorithm of Silverman et al. (1990)). A theoretically more satisfying method is to add a penalty term - $B(f)$ to (4.5), i.e. to minimize

$$\ell(f) - B(f) \tag{4.6}$$

$-B(f)$ may be interpreted either in a Bayesian setting or simply as smoothing term. Typically,

$$B(f) = (f - \overline{f})^T B (f - \overline{f})$$

with a positive definite matrix B and a reference picture $\overline{f}$. Unfortunately, minimizing (4.6) is much more difficult than minimizing (4.5) and cannot be done with a simple iteration such as (4.4). For partial solutions to this problem see Levitan and Herman (1987), Green (1990), Setzepfandt (1992).

As in ART, a judicious arrangement of the equations can speed up the convergence significantly. The directions have to be arranged in "ordered subsets", see Hudson et al. (1992).

4.3. MART (multiplicative algebraic reconstruction technique)

While ART converges in the consistent case to a minimal norm solution of (4.1), MART is designed to converge to a solution of (4.1) which minimizes the entropy

$$E(f) = -\sum_{k=1}^{M} f_k \log f_k \ . \tag{4.7}$$

For this to make sense we assume that (4.1) has a positive solution, and we seek the minimizer of (4.7) among those f which have only positive components. This is reasonable in many tomographic problems.

The step $f^k \to f^{k+1}$ of the MART algorithm for (4.1) is as follows:

$$\begin{aligned} f^{k,0} &= f^k \\ f^{k,i} &= f_\ell^{k,i-1} \left(\frac{g_i}{a_i^T f^{k,i-1}} \right)^{a_{i\ell}} , \qquad i = 1, \ldots, M \ , \\ f^{k+1} &= f^{k,M} \ . \end{aligned}$$

MART is an example of a multiplicative algorithm, see Pierro (1990); another example is the EM algorithm.

5. Circular harmonic algorithms

Circular harmonic algorithms can be derived from the inversion formula of Cormack (1963). For the 2D problem $Rf = g$ it is obtained by Fourier expansions

$$\begin{aligned} f(x) &= \sum_{\ell} f_\ell(r) e^{i\ell\varphi} , \quad x = r(\cos\varphi, \sin\varphi)^T , \\ g(\theta, s) &= \sum_{\ell} g_\ell(s) e^{i\ell\varphi} , \quad \theta = (\cos\varphi, \sin\varphi)^T . \end{aligned}$$

One can show that

$$f_\ell(r) = -\frac{1}{\pi} \int_r^\infty (s^2 - r^2)^{-1/2} T_{|\ell|}(\frac{s}{r}) g'_\ell(s) ds . \tag{5.1}$$

T_ℓ is the Chebysheff polynomial of the first kind of order ℓ. In principle this defines an inversion formula for R: The Fourier coefficients g_ℓ of $g = Rf$ determine the Fourier coefficients f_ℓ of f via (5.1), hence f is determined by g.

The formula (5.1) is useless for practical calculations since T_ℓ increases exponentially with ℓ outside $[-1, +1]$. Cormack (1964) derived also a stable version of his inversion formula. It reads

$$\begin{aligned} f_\ell(r) = -\frac{1}{\pi r} \Bigg\{ \int_r^\infty \left(\frac{s^2}{r^2} - 1\right)^{-1/2} \left(\frac{s}{r} + \sqrt{\frac{s^2}{r^2} - 1}\right)^{-|\ell|} & g'_\ell(s) ds \\ - \int_0^r U_{|\ell|-1}(\frac{s}{r}) g'_\ell(s) ds & \Bigg\} \end{aligned}$$

where U_ℓ is the Chebysheff polynomial of the second kind of order ℓ. This formula does not suffer from exponential increase. It is the starting point of the circular harmonic reconstruction algorithms of Hansen (1981), Hawkins and Barrett (1986), Chapman and Cary (1986).

We take a different route and start out from (2.1) again. We consider only the case $n = 2$. Putting $x = x_{ik} = s_i \theta(\varphi_k)$, $\theta(\varphi) = (\cos\varphi, \sin\varphi)^T$, in (2.4) we obtain

$$(V \star f)(x_{ik}) = \frac{2\pi\rho}{pq} \sum_{\ell=-q}^{q} \sum_{j=0}^{p-1} v_\Omega(s_i \cos(\varphi_{j-k}) - s_\ell) g(\theta_j, s_\ell) .$$

The j sum is a convolution. In order to make this convolution cyclic we extend $g(\theta_j, s_\ell)$ by putting $g(\theta_{j+p}, s_\ell) = g(\theta_j, -s_\ell)$, in accordance with the eveness property of the Radon transform. Then,

$$(V \star f)(x_{ik}) = \frac{\pi\rho}{pq} \sum_{j=0}^{2p-1} v_\Omega(s_i \cos(\varphi_{j-k}) - s_\ell) g(\theta_j, s_\ell) \ .$$

Defining

$$\begin{aligned} h_{\ell ik} &= \frac{\pi}{p} \sum_{j=0}^{2p-1} v_\Omega(s_i \cos(\varphi_{j-k}) - s_\ell) g(\theta_j, s_\ell) \ , \\ & \ell = -q, \dots, q \ , \quad i = 0, \dots, q \ , \quad k = 0, \dots, 2p-1 \end{aligned}$$

we have

$$(V_\Omega \star f)(x_{ik}) \sim \frac{\rho}{q} \sum_{\ell=-q}^{q} h_{\ell ik} \ .$$

This defines the circular harmonic algorithm.

Algorithm 4 (Circular harmonic algorithm for standard parallel geometry.)

Data: The values $g_{j,\ell} = g(\theta_j, s_\ell)$, $j = 0, \dots, p-1$, $\ell = -q, \dots, q$, g the 2D Radon transform of f.

Step 1: Precompute the number $v_{\ell,i,j} = v_\Omega(s_i \cos(\varphi_j) - s_\ell)$ and extend $g_{j,\ell}$ to all $j = 0, \dots, 2p-1$ by $g_{j+p,\ell} = g_{j,-\ell}$.

Step 2: For $i = 0, \dots, q$, $\ell = -q, \dots, q$ carry out the discrete cyclic convolutions

$$h_{\ell ik} = \frac{\pi}{p} \sum_{j=0}^{2p-1} v_{\ell,i,j-k} g_{j,\ell} \ , \quad k = -q, \dots, q \ .$$

Step 3: For $i = 0, \dots, q$ and $k = 0, \dots, 2p-1$ compute

$$f_{CH}(x_{ik}) = \frac{\rho}{q} \sum_{\ell=-q}^{q} h_{\ell ik} \ .$$

Result: $f_{CH}(x_{ik})$ is an approximation to $f(x_{ik})$.

The second step of the algorithm has to be done with a fast Fourier transform (FFT) in order to make the algorithm competitive with filtered backprojection. In that case step 2 requires $\mathrm{O}(q^2 p \log p)$ operations. This is slightly more (by the factor $\log p$) than what is needed in the filtered backprojection algorithm. The third step needs $4pq^2$ additions.

Algorithm 4 can be used almost without any changes for the interlaced parallel geometry, i.e. for $g_{j,\ell}$ with $\ell + j$ odd missing (p even). One simply puts $g_{j,\ell} = 0$ for $j + \ell$ odd and doubles $h_{\ell ik}$ in step 2.

Circular harmonic algorithms are also available for standard fan beam data. (2.9) with $x = x_{ik} = t_i\theta(\beta_k)$, $t_i = i\Delta t$, $\Delta t = \rho/q$ reads

$$(V \star f)(x_{ik}) \sim r\Delta\alpha\Delta\beta \sum_{\ell=-q}^{q} \sum_{j=0}^{p-1} v_\Omega(t_i \cos(\beta_{k-j} - \alpha_\ell) - r \sin\alpha_\ell) g(\beta_j, \alpha_\ell) \cos\alpha_\ell$$

where g is now the fan beam data function from (2.8).

Algorithm 5 (Circular harmonic algorithm for standard fan beam geometry.)

Data: The values $g_{j,\ell} = g(\beta_j, \alpha_\ell)$, $j = 0, \dots, p-1$, $\ell = -q, \dots, q$ with g from (2.8).
Step 1: Precompute the numbers $v_{\ell,i,j} = v_\Omega(t_i \cos(\beta_j - \alpha_\ell) - r \sin\alpha_\ell) \cos\alpha_\ell$.
Step 2: For $i = 0, \dots, q$, $\ell = -q, \dots, q$ carry out the discrete cyclic convolutions

$$h_{\ell ik} = \Delta\beta \sum_{j=0}^{p-1} v_{\ell,i,j-k} g_{j,\ell}\ , \quad k = 0, \dots, p-1\ .$$

Step 3: For $i = 0, \dots, q$ and $k = 0, \dots, p-1$ compute the sums

$$f_{CH}(x_{ik}) = r\Delta\alpha \sum_{\ell=-q}^{q} h_{\ell ik}\ .$$

Result: $f_{CH}(x_{ik})$ is an approximation to $f(x_{ik})$.

The complexity of Algorithm 5 is again $\mathrm{O}(q^2 p \log p)$.
A few remarks are in order.

1. Circular harmonic algorithms compute the reconstruction on a grid in polar coordinates. Interpolation to a cartesian grid (for instance for the purpose of display) is not critical and can be done by linear interpolation.
2. The resolution of the circular harmonic algorithms is the same as for the corresponding filtered backprojection algorithms in section 2.1.
3. Even though circular harmonic algorithms are asymptotically a little slower (by a factor of $\log p$) than filtered backprojection they usually run faster due to their simplicity. This is true in particular for fan beam data because in that case the backprojection is quite time consuming.
4. Circular harmonic algorithms tend to be more accurate than filtered backprojection because no additional approximations, such as interpolations or homogeneity approximations (in the fan beam case) are used.
5. The disadvantage of circular harmonic algorithms lies in the fact that they start with angular convolutions. This is considered as unpractical in radiological applications.

6. Fourier reconstruction

We already made use of the relation

$$(Rf)^\wedge(\theta,\sigma) = (2\pi)^{(n-1)/2}\hat f(\sigma\theta) \tag{6.1}$$

for the Radon transform in $\mathbb{R}^n$ in section 2.1, and of the corresponding formula for the nD ray transform

$$(Pf)^\wedge(\theta,\xi) = (2\pi)^{1/2}\hat f(\xi)\ , \quad \xi \in \theta^\perp \tag{6.2}$$

in section 3.4. In this section we use these formulas to derive reconstruction algorithms. To fix ideas we consider the case of the 2D Radon transform, sampled as in the standard parallel geometry. This means that $g = Rf$ is given for $\theta = \theta_j = (\cos\varphi_j, \sin\varphi_j)^T$, $\varphi_j = \pi j/p$, $j = 0,\dots,p-1$ and $s = s_\ell = \ell\rho/q$, $\ell = -q,\dots,q$, as in section 2.1. f is assumed to vanish outside $|x| < \rho$.

The idea of Fourier reconstruction is very simple: Do a 1D Fourier transform on g with respect to the second variable for each θ. According to (6.1) this yields $\hat f$ in all of $\mathbb{R}^2$. Do a 2D inverse Fourier transform to obtain f. Even though this seems fairly obvious, the numerical implementation in a discrete setting is quite intricate. In fact good Fourier algorithms have been found only quite recently.

To begin with we describe the simplest possible implementation. We warn the reader that this algorithm is quite useless since it is not sufficiently accurate.

Algorithm 6 (Standard Fourier reconstruction.)

Data: The number $g_{j,\ell} = g(\theta_j, s_\ell)$, $j = 0,\dots,p-1$, $\ell = -q,\dots,q$
g the 2D Radon transform of f.

Step 1: For $j = 0,\dots,p-1$ carry out the discrete Fourier transform

$$\hat g_{j,r} = (2\pi)^{-1/2}\frac{\rho}{q}\sum_{\ell=-q}^{q} e^{-i\pi r\ell/q} g_{j,\ell}\ , \quad r = -q,\dots,q\ .$$

Step 2: For each$k \in \mathbb{Z}^2$, $|k| < q$, find j, r such that $r\theta_j$ is as close as possible to k and put

$$\hat f_k = (2\pi)^{-1/2}\hat g_{j,r}\ .$$

Step 3: Do the 2D discrete inverse Fourier transform

$$f_m = \frac{1}{2\pi}\left(\frac{\pi}{\rho}\right)^2 \sum_{\substack{k\in\mathbb{Z}^2\\ |k|<q}} e^{i\pi m\cdot k/q}\hat f_k\ , \quad m \in \mathbb{R}^2\ , \quad |m| < q\ .$$

Result: f_m is an approximation to $f(\frac{\rho}{q}m)$.

The algorithm is designed to reconstruct a function f with support in $|x| < \rho$ which is essentially Ω-band-limited. (2.6), (2.7) have to be satisfied. We stress again that the algorithm as it stands is not to be recommended because of poor accuracy. Better versions are described below.

A few comments are in order. In step 1 we compute an approximation $\hat{g}_{j,r}$ to

$$\hat{g}(\theta_j, \frac{\pi}{\rho}r) = (2\pi)^{-1/2} \int_{\mathbf{R}^1} e^{-is\frac{\pi}{\rho}r} g(\theta_j, s) ds\ , \qquad r = -q, \dots, q\ .$$

Under the assumption (2.6) this approximation is reliable since Fourier transforms are evaluated exactly by the trapezoidal rule if the Nyquist condition, in our case (2.6), is satisfied. According to (6.1), step 1 provides us with the values

$$\begin{aligned} \hat{f}(\frac{\pi}{\rho}r\theta_j) &= (2\pi)^{-1/2}\hat{g}(\theta_j, \frac{\pi}{\rho}r) \\ &\sim (2\pi)^{-1/2}\hat{g}_{j,r}\ . \end{aligned}$$

In step 2 we compute $\hat{f}$ on the cartesian grid $\frac{\pi}{\rho}\mathbb{Z}^2$ by nearest neighbour interpolation:

$$\hat{f}(\frac{\pi}{\rho}k) \sim \hat{f}_k\ , \qquad k \in \mathbb{Z}^2\ , \quad |k| < q\ .$$

Since f vanishes outside $|x| < \rho$, $\hat{f}$ has bandwidth ρ. Thus sampling of $\hat{f}$ on a 2D grid with stepsize $\frac{\pi}{\rho}$ is adequate by the sampling theorem.

Step 3 is the trapezoidal rule for the 2D inverse Fourier transform, properly discretized and complying with the Nyquist condition. Hence f_m is an approximation to $f(\frac{\rho}{q}m)$.

Step 1 and step 3 of Algorithm 6 are justified by the sampling theorem. Thus the failure of the algorithm must be caused by the interpolation in step 2. This is in fact the case. Of course we can replace the interpolation by a more accurate one, such as linear interpolation. However this doesn't help much.

Inspite of its poor accuracy, Fourier reconstruction is attractive because of its favourable complexity which is due to the fast Fourier transform (FFT).

We have used FFT in the circular harmonic algorithm already, but FFT is so essential for Fourier reconstruction that we say a few words about FFT; for a thourough treatment we refer to Nussbaumer (1982). The discrete Fourier transform of length q is defined to be

$$\hat{y}_k = \sum_{\ell=0}^{q-1} e^{-2\pi i k\ell/q} y_\ell\ , \quad k = 0, \dots, q-1\ . \tag{6.3}$$

An algorithm which computes $\hat{y}_0, \dots \hat{y}_{q-1}$ from $y_0, \dots, y_{q-1}$ in less then $\mathrm{O}(q^2)$, typically $\mathrm{O}(q \log q)$, operations is a called an FFT. In the circular harmonic algorithm we have used FFT just for the evaluation of (6.3). In Fourier reconstruction we employ FFT for the evaluation of Fourier integrals

$$\hat{f}(\xi) = (2\pi)^{-1/2} \int e^{-ix\cdot\xi} f(x)dx$$

for the functions f in $\mathbb{R}^1$ with support in $(-\rho, \rho)$. Sampling theory tells us that $\hat{f}$ has to be discretized with stepsize $\frac{\pi}{\rho}$ ($\hat{f}$ has bandwidth ρ). With $h = \rho/q$ the stepsize for f the trapezoidal rule provides the approximation

$$\hat{f}(\frac{\pi}{\rho}k) = (2\pi)^{-1/2}\frac{\rho}{q}\sum_{\ell=-q}^{q-1} e^{-i\pi\ell k/q} f(\frac{\rho}{q}\ell) \ , \quad k = -q, \ldots, q-1 \ . \tag{6.4}$$

The range of k is in agreement with the sampling theorem: The stepsize $h = \frac{\rho}{q}$ corresponds to the bandwidth $\Omega = \frac{\pi}{h} = \frac{\pi}{\rho}q$, hence $|k| \le q$ in (6.4) suffices. Of course (6.4) is a discrete Fourier transform of length $2q$. Sometimes one wants to adjust the stepsizes of f and $\hat{f}$ differently. Then one has to evaluate

$$\hat{y}_k = \sum_{\ell=0}^{q-1} e^{-ic\ell k/q} y_\ell \ , \quad k = 0, \ldots, q-1 \tag{6.5}$$

with an arbitrary real parameter c. This can be done by the chirp-z-algorithm, see Nussbaumer (1982), again with typically $\mathrm{O}(q \log q)$ operations.

Assuming that we have a fast Fourier transform (FFT) algorithm which does a discrete Fourier transform of length q with $\mathrm{O}(q \log q)$ operations, the complexity of Algorithm 6 is as follows. Step 1 does p Fourier transforms of length $2q$, requiring $\mathrm{O}(pq \log q)$ operations. Assuming that the interpolation in step 2 can be done in $\mathrm{O}(1)$ operations per point we get $\mathrm{O}(q^2)$ as the complexity of step 2. The 2D Fourier transform in step 3 can be done with $\mathrm{O}(q^2 \log q)$ operations. Hence the complexity of Algorithm 6 is $\mathrm{O}((pq + q^2) \log q)$. This is much better than the filtered backprojection algorithm (Algorithm 1), which needs $\mathrm{O}(q^2 p)$ operations for a reconstruction on a $q \times q$ grid.

Presently there exist two Fourier methods with satisfactory accuracy and favourable complexity.

1. The linogram algorithm (Edholm and Herman (1987)).

Here, interpolation in Fourier space is avoided altogether by a clever choice of the directions φ_j. The linogram algorithm works on the data

$$\begin{aligned} g^C_{j,\ell} &= g(\theta_j, s_\ell) \quad , \qquad \varphi_j = \operatorname{arccot}\,(j/q) \ , \\ g^T_{j,\ell} &= g(\theta_j, s_\ell) \quad , \qquad \varphi_j = \arctan\,(j/q) \end{aligned}$$

where $j,\ \ell = -q, \ldots, q$. Doing a 1D Fourier transform on $g^C_{j,\ell}$ results in

$$\hat{g}(\theta_j, \sigma) = (2\pi)^{-1/2}\frac{\rho}{q}\sum_{\ell=-q}^{q} e^{-i\sigma\ell\rho/q} g^C_{j,\ell}$$

where $\varphi_j = \text{arccot}\ (j/q)$. For $\sigma = k\pi/\rho\sin\varphi_j$ we get from (6.1)

$$\hat{f}\left(\frac{k\pi}{\rho\sin\varphi_j}\theta_j\right) = (2\pi)^{-1}\frac{\rho}{q}\sum_{\ell=-q}^{q} e^{-i\pi k\ell/(q\sin\varphi_j)} g^C_{j,\ell}\,. \tag{6.6}$$

Note that this can be done efficiently by the chirp-z-algorithm. The key observation is that the points

$$\frac{k\pi}{\rho\sin\varphi_j}\theta_j = \frac{k\pi}{\rho}\begin{pmatrix} 1 \\ \cot\varphi_j \end{pmatrix} = \frac{k\pi}{\rho}\begin{pmatrix} 1 \\ j/q \end{pmatrix}$$

form a grid lying on vertical lines with distance π/ρ, being evenly spaced within each vertical line (though with different stepsizes in different lines). On such a grid we can do a 1D FFT in the horizontal direction in the usual way. In the horizontal direction the stepsize is not what we need for a direct application of the FFT, but the chirp-z-algorithm is still applicable. This takes care of the 2D inverse Fourier transform in $|\xi_2| \geq |\xi_1|$. For $|\xi_1| \leq |\xi_2|$ we proceed analogously with the data $g^T_{j,\ell}$, evaluating $\hat{f}(\sigma\theta_j)$ for $\sigma = k\pi/\rho\cos\varphi_j$. We remark that the linogram data in Edholm and Herman (1987) is a little different from ours, namely $s_\ell = h\ell\sin\varphi_j$, $s_\ell = h\ell\cos\varphi_j$, respectively. The use of these detector positions makes the linogram algorithm even simpler in that the factor $\sin\varphi_j$ in the right hand side of (6.6) dissapears.

2. The gridding algorithm (O'Sullivan (1985), Kaveh (1987), Schomberg and Timmer (1995).

This algorithm works on the standard parallel data which we used in Algorithm 6. It does the interpolation in step 2 of Algorithm 6 in the following way. Let w be a smooth function in $\mathbb{R}^2$ with $w = 1$ on $|x| < \rho$ which is decaying exponentially at infinity. Put $f_w = wf$. Then,

$$\begin{aligned} \hat{f}_w(\xi) &= (2\pi)^{-1}\int_{\mathbb{R}^2} \hat{w}(\xi-\eta)\hat{f}(\eta)d\eta \\ &= (2\pi)^{-1}\int_0^\infty \sigma\int_{S^1} \hat{w}(\xi-\sigma\theta)\hat{f}(\sigma\theta)d\theta d\sigma \\ &= (2\pi)^{-3/2}\int_0^\infty \sigma\int_{S1} \hat{w}(\xi-\sigma\theta)\hat{g}(\theta,\sigma)d\theta d\sigma \end{aligned}$$

by (6.1). Using a quadrature rule with nodes σ_r, θ_j and weights α_{jr} we obtain the approximation

$$\hat{f}_{w,k} = (2\pi)^{-3/2} \sum_{j=0}^{p-1} \sum_{r=-q}^{q} \alpha_{jr} \sigma_r \hat{w}(\frac{\pi}{\rho}k - \sigma_r \theta_j) \hat{g}_{jr} \tag{6.7}$$

to $\hat{f}_w(\frac{\pi}{\rho}k)$. The method relies on the following assumptions:

1. $\hat{w}$ is decaying at infinity so fast that only a few terms of the r sum in (6.7) have to be retained.
2. The dependence on the angle is not critical, so that it suffices to retain only a few terms in the j sum of (6.7).

If these conditions are met then $\hat{f}_{w,k}$ of (6.3) is a good approximation to $\hat{f}_w(\frac{\pi}{\rho}k)$ which can be evaluated essentially in O(1) operations for each k. This takes care of step 2. Of course when using (6.7) we have to divide f by w after step 3 to make up for the previous multiplication with w.

It is needless to say that our derivation of the gridding algorithm is purely heuristic. It seems that presently there exists no convincing theoretical analysis of the gridding algorithm.

7. Outlook

In the preceding sections we have given the fundamentals of the most widely used algorithms in tomography. In many ways these fundamentals are quite different from traditional numerical analysis, the main difference being the consistent use of sampling theory and Fourier analysis. The developement of algorithms is still very lively, in particular in 3D and in Fourier based algorithms.

We dealt only with the most simple problems and with standard situations. Practical problems deviate in many ways from the simple ones we considered. Often the data is incomplete (see Louis (1980)), leading to non uniqueness and instability. Sometimes the integral equation to be solved are not completely specified, be it that the weight function (as in emission tomography, Welch et al. (1998)) or the directions (as in electron microscopy, Wuschke (1990), Gelfand and Goncharov (1990)) are unknown. In particular in technical applications, the number of data is often so small (see e.g. Sielschott and Derichs (1995)) that full reconstruction is impossible and special algorithms have to be developed, usually taylored to the specific application. Sometimes only certain features of the object, such as boundaries between regions of different densities are searched for (Faridani et al. (1997)), Ramm and Katsevich (1996)), calling for special algorithms.

At present we have an adequate understanding of the fundamentals of tomographic reconstruction algorithm. However, new applications of tomography are coming up almost daily, each one presenting new challenges to the numerical analyst. So I guess that research in this field will go on for ever.

REFERENCES

M. Abramowitz and I.A. Stegun (1970), *Handbook of Mathematical Functions.* Dover.

Y. Censor (1981),'Row-action methods for huge and sparce systems and their applications', *SIAM Review* **23**, 444-466.

Y. Censor, P.B. Eggermont and D. Gordon (1983), 'Strong underrelaxation in Karczmarz's method for inconsistent systems', *Numer. Math.* **41**, 83-92.

Y. Censor and S.A. Zenios (1997), *Parallel Optimization.* Oxford University Press.

L.T. Chang and G.T. Herman (1980), 'A scientific study of filter selection for a fan-beam convolution algorithm', *SIAM J. Appl. Math.* **39**, 83-105.

C.H. Chapman and P.W. Cary (1986), 'The circular harmonic Radon transform', *Inverse Problems* **2**, 23-49.

J.G. Colsher (1980), 'Fully three-diemnsional emission tomography', *Phys. Med. Biol.* **25**, 103-115.

A.M. Cormack (1963), 'Representation of a function by its line integrals, with some radiological applications I', *J. Appl. Physics* **34**, 2722-2727.

A.M. Cormack (1964), 'Representation of a function by its line integrals, with some radiological applications II', *J. Appl. Physics* **35**, 195-207.

S.R. Deans (1983), *The Radon Transform and some of its Applications.* Wiley.

M. Defrise, D.W. Townsend and R. Clack (1989), 'Three-dimensional image reconstruction from complete projections', *Phys. Med. Biol.* **34**, 573-587.

M. Defrise and R. Clack (1995), 'A Cone-Beam Reconstruction Algorithm Using Shift-Variant Filtering and Cone-Beam Backprojection', *IEEE Transactions on Medical Imaging* **13**, 186-195.

A.P. Dempster, N.M. Laird and D.B. Rubin (1977), 'Maximum likelihood from incomplete data via the EM algorithm', *J.R. Statist. Soc. B* **39**, 1-38.

P. Edholm and G.T. Herman (1987), 'Linograms in image reconstruction from projections', *IEEE Trans. Med. Imag.* **6**, 301-307.

A. Faridani, D.V. Finch, E.L. Ritman and K.T. Smith (1997), 'Local tomography II', *SIAM J. Appl. Math.* **57**, 1095-1127.

J. Frank (ed.) (1992), *Electron Tomography.* Plenum Press.

I.M. Gelfand and A.B. Goncharov (1987), 'Recovery of a compactly supported function starting from its integrals over lines intersecting a given set of points in space', *Doklady* **290** (1986), English Translation in Soviet Math. Dokl. **34**, 373-376.

I.M. Gelfand and A.B. Goncharov (1990), 'Spatial rotational alignment of identical particles given their projections: Theory and practice', *Translation of Mathematical Monographs* **81**, 97-122.

R. Gordon, R. Bender, G.T. Herman (1970), 'Algebra reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography', *J. Theor. Biol.* **29**, 471-481.

P. Grangeat (1991), 'Mathematical framework of cone-beam reconstruction via the first derivative of the Radon transform', in: *Herman, G.T. - Louis, A.K. - Natterer, F. (eds.): Lecture Notes in Mathematics* **1497**, 66-97.

P.J. Green (1990), 'Baysian reconstruction from emission tomography data using a modified EM algorithm', *IEEE Trans. Med. Imag.* **9**, 84-93.

C. Hamaker and D.C. Solmon (1978), 'The angles between the null spaces of X-rays', *J. Math. Anal. Appl.* **62**, 1-23.

E. W. Hansen (1981), 'Circular harmonic image reconstruction', *Applied Optics* **20**, 2266-2274.

W.G. Hawkins and H.H. Barrett (1986), 'A numerically stable circular harmonic reconstruction algorithm', *SIAM J. Numer. Anal.* **23**, 873-890.

G.T. Herman (1980), *Image Reconstruction from Projection. The Fundamentals of Computerized Tomography.* Academic Press.

G.T. Herman and L. Meyer (1993), 'Algebraic reconstruction techniques can be made computationally efficient', *IEEE Trans. Med. Imag.* **12**, 600-609.

H.M. Hudson, B.F. Hutton and R. Larkin (1992), 'Accelerated EM reconstruction using ordered subsets', *J. Nucl. Med.* **33**, 960-968.

A.J. Jerry (1977), 'The Shannon sampling theorem-its various extensions and applications: a tutorial review', *Proc. IEEE* **65**, 1565-1596.

S. Kaczmarz (1937), 'Angenäherte Auflösung von Systemen linearer Gleichungen', *Bulletin de l'Académie Polonaise des Sciences et des Lettres* **A35**, 355-357.

A.C. Kak and M. Slaney (1987), *Principles of Computerized Tomography Imaging.* IEEE Press, New York.

M. Kaveh - Soumekh, M. (1987), *Computer assisted diffraction tomography, in: Image Recovery: Theory and Application, H. Stark (ed.)*, Academic Press.

H. Kruse (1989), 'Resolution of reconstruction methods in computerized tomography', *SIAM J. Sci. Stat. Comput.* **10**, 447-474.

E. Levitan and G.T. Herman (1987), 'A maximum a posteriori probability expection maximation algorithm for image reconstruction in emission tomography', *IEEE Trans. Med. Imag.* **6**, 185-192.

R.M. Lewitt (1992), 'Alternatives to voxels for image representations in iterative reconstruction algorithms', *Phys. Med. Biol.* **37**, 705-716.

A.K. Louis (1980), 'Picture reconstruction from projections in restricted range', *Math. Meth. Appl. Sci.* **2**, 109-220.

R. Marabini, G.T. Herman and J.M. Carazo (1998), 'Fully Three-dimensional reconstruction in electron microscopy', in: Borgers, C. - Natterer, F. (eds.): *Computational Radiology and Imaging: Therapy and Diagnostics*, IMA Volumes in Mathematics and its Applications **110**, Springer 1998.

R.B. Marr, C.N. Chen and P.C. Lauterbur (1981), 'On two approaches to 3D reconstruction in NMR zeugmatography', in: Herman, G.T. and Natterer, F. (eds.): *Mathematical Aspects of Computerized Tomography*, Proceedings, Oberwolfach 1980, Springer.

F. Natterer (1986), *The Mathematics of Computerized Tomography.* John Wiley & Sons and B.G. Teubner.

F. Natterer (1993), 'Sampling in fan beam tomography', *SIAM J. Appl. Math.* **53**, 358-380.

F. Natterer and A. Faridani (1990), 'Basic algorithms in tomography', *in: Grünbaum, F.A. et al. (eds.): Signal Processing Part II: Control Theory and Applications*, 321-334. Springer.

S. Nilsson (1997), 'Application of fast backprojection techniques for some inverse problems of integral geometry', *Linköping Studies in Science and Technology. Dissertation, No. 499. Department of Mathematics, Linköping University, Linköping Sweden.*

H.J. Nussbaumer (1982), *Fast Fourier Transform and Convolution Algorithm.* Springer.

S.S. Orlov (1976), 'Theory of three dimensional reconstruction II. The recovery operator', *Sov. Phys. Crystallogr.* **20**, 429-433.

J.D. O'Sullivan (1985), 'A fast sinc function gridding algorithm for Fourier inversion in computer tomography', *IEEE Trans. Med. Imag.* **4**, 200-207.

A.R. De Pierro (1990), 'Multiplicative interative methods in computed tomography', in: Herman, G.T. - Louis, A.K. - Natterer, F. (eds.), *Mathematical Methods in Tomography*, 167-186, Springer.

J. Radon (1917), 'Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten', *Berichte Sächsische Akademie der Wissenschaften, Math.-Phys. Kl.*, **69**, 262-267, Leipzig.

A. Ramm and A. Katsevich (1996), *The Radon Transform and Local Tomography.* CRC Press.

H. Schomberg and J. Timmer (1995), 'The gridding method for image reconstruction bx Fourier transformation', *IEEE Trans. Med. Imag.* **14**, (596-607).

L.A. Shepp and B.F. Logan (1974), 'The Fourier reconstruction of a head section', *IEEE Trans. Trans. Nucl. Sci.* **NS-21**, 21-43.

L.A. Shepp and Y. Vardi (1982), 'Maximum likelihood reconstruction for emission tomography', *IEEE Trans. Med. Imag.* **1**, 113-122.

B. Setzepfand (1992), 'ESNM: Ein rauschunterdrückendes EM-Verfahren für die Emissionstomographie'. Dissertation, Fachbereich Mathematik, Universität Münster.

H. Sielschott and W. Derichs (1995), 'Use of collocation methods under inclusion of a priori information in acoustic pyrometry', *Proc. European Concerted Action on Process Tomography, Bergen, Norway*, 110-117.

B.W. Silverman, M.C. Jones, D.W. Nychka and J.D. Wilson (1990), 'A smoothed EM approach to indirect estimation problems, with particular reference to stereology and emission tomography', *J. R. Statist. Soc. B* **52**, 271-324.

H.K. Tuy (1983), 'An inversion formula for cone-beam reconstruction', *SIAM J. Appl. Math.* **43**, 546-552.

A. Welch, R. Clack, F. Natterer and G.T. Gullberg (1997), 'Towards accurate attenuation correction in SPECT', *IEEE Trans. Med. Imag.* **16**, 532-541.

K. Wuschke (1990), 'Die Rekonstruktion von Orientierungen aus Projektionen'. Diplomarbeit, Institut für Numerische und instrumentelle Mathematik, Universität Münster.