

Three-dimensional image reconstruction by means of two-dimensional Radon inversion

Citation for published version (APA):

Eggermont, P. P. B. (1975). *Three-dimensional image reconstruction by means of two-dimensional Radon inversion*. (EUT report. WSK, Dept. of Mathematics and Computing Science; Vol. 75-WSK-04). Eindhoven University of Technology.

Document status and date:

Published: 01/01/1975

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

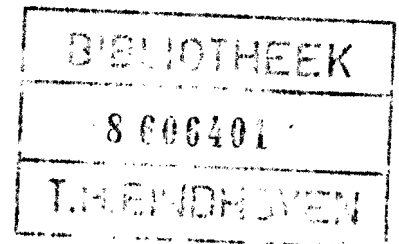
If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

TECHNISCHE HOGESCHOOL EINDHOVEN
NEDERLAND
ONDERAFDELING DER WISKUNDE

TECHNOLOGICAL UNIVERSITY EINDHOVEN
THE NETHERLANDS
DEPARTMENT OF MATHEMATICS



Three-dimensional image reconstruction by means of
two-dimensional Radon inversion

by

P.P.B. Eggermont

T.H.-Report 75-WSK-04

July 1975

<u>Contents</u>	pages
Chapter 1. <u>Introduction</u>	2
PART I.	
Chapter 2. <u>The integral equation</u>	4
2.1. Radon transformations	4
2.1.1. Radons inversion formula	6
2.1.2. Bandlimited functions	9
2.1.3. Again bandlimited function: The solution by means of infinite series	10
Chapter 3. <u>The discrete problem</u>	13
3.1. Sampling of the Radon transform	13
3.2. Fourier methods	13
3.2.1. Interpolation with sincfunctions	13
3.2.2. Interpolation with circlefunctions	14
3.3. Convolution methods	15
Chapter 4. <u>The problem as a set of linear equations</u>	17
4.1. Formulation of the problem as a set of equations	17
4.2. Sincfunctions	18
4.3. Stepfunctions	19
4.4. Series expansions: Orthogonal Radon transforms	23
4.4.1. Continuous case	23
4.4.2. Discrete case	26
Chapter 5. <u>Algebraic reconstruction techniques (ART)</u>	27
5.1. Finite ray width	27
5.2. The set of equations	27
5.3. ART	28
5.4. Limiting behaviour of ART algorithms	30
5.5. Convergence if the stepfunction is refined	34

PART II.

Chapter 6. <u>The problem on the unit disk: Solution of the continuous and discrete problem by means of series expansions</u>	36
6.1. The Radon transformation on the unit disk	36
6.2. The Radon transform of a polynomial on \mathbb{D}	37
6.3. Orthogonal bases of $\mathbb{P}(\mathbb{D})$ and $\hat{\mathbb{P}}(\mathbb{C})$	40
6.4. Radon inversion by means of polynomial expansions	43
6.5. Discrete problem	45
6.6. Stability to noisy data	49
6.7. Smoothing	51
6.8. Parallel- and divergent-ray sampling	54
6.9. Another system of sample points	54
Chapter 7. <u>The calculation of $P^{(M)}$ on \mathbb{D}</u>	59
7.1. Introduction: The polar grid	59
7.2. Calculation of Jacobi series	60
7.3. Reduction to Legendre series	61
7.4. Direct method: A general case	63
7.5. Direct method: Jacobi series	65
7.6. Special case: Legendre series, two algorithms	68
7.7. Numerical stability of Algorithm I	70
7.8. Numerical stability of Algorithm II	71
7.9. Numerical experiments	72
<u>APPENDIX:</u> 7.10. The recurrence relation for Legendre polynomials: general solution	74
Chapter 8. <u>Numerical experiments, conclusions</u>	79
8.1. Implementations	79
8.2. Data and noisy data	79
8.3. Testpatterns	79
8.4. Computation time	80
8.5. Discussion of results	81
8.6. Conclusions	82

Part III.

Chapter 9. <u>The convolution method</u>	83
9.1. The convolution method	83
9.2. A variant of the convolution method	83
9.3. Implementation	85
9.4. Testpatterns	85
9.5. Discussion	85
References	104

Abstract

In this report we consider the problem of three-dimensional reconstruction of structures from their projections, formulated as a multiple two-dimensional problem.

Part I contains a discussion of some reconstruction techniques, that are known in literature, including the analytic backgrounds of some of them.

In Part II a method is discussed that is based on orthogonal polynomial expansion. Notably, attention is payed to the calculation of the resulting finite series of orthogonal polynomials. Finally some numerical experiments concerning the reconstruction of testpatterns are presented.

In part III a variant of the convolution method is derived and some numerical experiments are discussed.

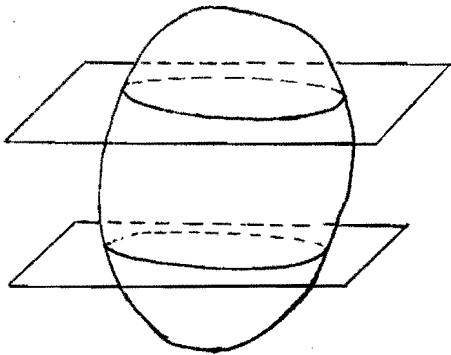
Part I and II were prepared in order to obtain the master's degree (ir.) at the Technological University Eindhoven.

I am greatly indebted to prof. G.W. Veltkamp, for his stimulating criticism during the preparation of this report.

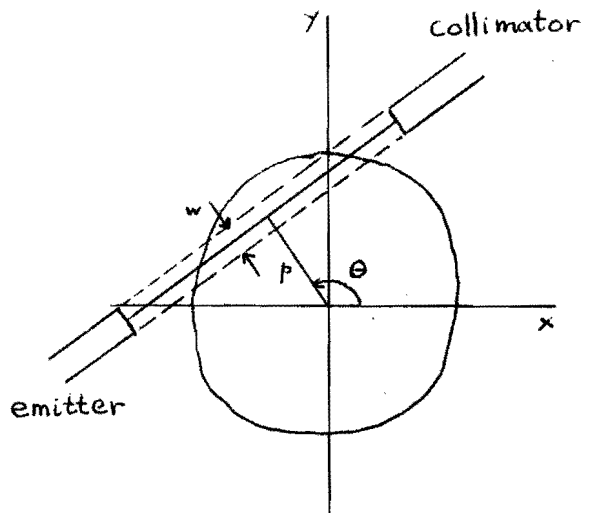
Chapter 1. Introduction

The problem we are dealing with consists of the following: the determination of the structure of an object from X-ray or γ -ray photographs of the object from several directions. This problem arises in nuclear medicine as e.g. localising tumors in the brain ([12]) but also in a different form in radio-astronomy ([2]), aerodynamics ([18]) and electronmicroscopy ([4]).

The problem is essentially three dimensional but it may be seen to be a multiple twodimensional problem: we may consider the object to consist of a number of thin layers, in such a way that the structure does not change much across the thickness of each layer. The problem is then to determine the structure of the layers. So we restrict ourselves to the twodimensional problem.



Three-dimensional



Two-dimensional

An X-ray photograph may be obtained by shifting the object between the emitter-collimator combination (in the direction perpendicular to the line emitter-collimator). The measured quantities depend uniquely on the absorption of the emission by the object, i.e. the integral of the absorption coefficient over the domain \mathbb{B} of the object which is covered by the X-ray. So from X-ray scans we obtain information about the quantities

$$g := \iint_{\mathbb{B}} f \, d\mathbb{B}$$

where f represents the absorption coefficient, for a certain finite set of \mathbb{B} -s and a certain set of angles θ .

So the problem may be posed as:

"Given a sampling of g , determine a good estimate of the absorption coefficient f ".

If the ray width w tends to zero then

$$\frac{1}{w} g \rightarrow \int_L f dL$$

where L is the central line of the ray.

Therefore it is possible to consider the problem as a discrete form of a continuous integral equation. We write the integral equation as $Rf = g$, with the formal solution $f = R^{\text{inv}} g$. It is clear that R is a bounded operator on a space of function which obey certain smoothness conditions, in some prescribed norm, but not so for R^{inv} : a smooth g does not mean that f is smooth. In short: the problem is not well-posed. In practice it means that the discrete problem is poorly conditioned, with the consequence that perturbations of g are amplified in f . This means that smoothing must be applied.

Part I

Chapter 2. The integral equation

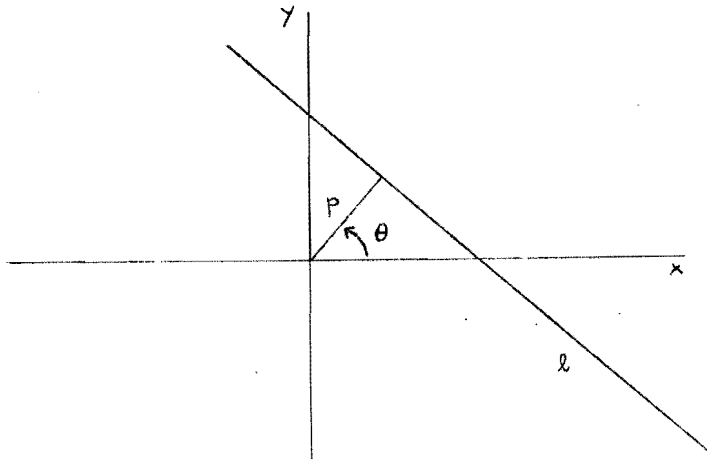
2.1. Radontransformations

Let the space S consist of functions on \mathbb{R}^2 which are infinitely many times differentiable and for which any derivative vanishes at infinity faster than any inverse power of the distance to the origine. Let $f \in S$, then the Radontransform of f exists: for any $(p, \theta) \in \mathbb{R} \times [0, 2\pi]$

$$(2.1.1) \quad g(p, \theta) := \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \delta(p - x \cos \theta - y \sin \theta) dx dy$$

or, in a more convenient form,

$$(2.1.2) \quad g(p, \theta) := \int_{-\infty}^{+\infty} f(p \cos \theta - t \sin \theta, p \sin \theta + t \cos \theta) dt .$$



The equation of the line l is:

$$x \cos \theta + y \sin \theta = p .$$

$g(p, \theta)$ is the integral of f along the line at "distance" p from the origin and of which the normal makes an angle θ with the positive X-as in counter-clockwise direction. (The "distance" p is allowed to be negative.) From (2.1.2) it follows that for all θ and p :

$$(2.1.3) \quad g(p, \theta) = g(-p, \theta + \pi) .$$

Since $f \in S$ also $g \in S$ (on the space $\mathbb{R} \times [0, 2\pi]$). So f , as well as g , has a Fourier transform:

$$(2.1.4) \quad F(X, Y) := \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) e^{2\pi i (xX + yY)} dx dy$$

$$(2.1.5) \quad G(K, \theta) := \int_{-\infty}^{+\infty} g(p, \theta) e^{2\pi i Kp} dp ,$$

and then, with (2.1.3):

$$(2.1.6) \quad G(K, \theta) = G(-K, \theta + \pi)$$

and with (2.1.2):

$$G(K, \theta) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(p \cos \theta - t \sin \theta, p \sin \theta + t \cos \theta) e^{-2\pi i Kp} dp dt .$$

After orthogonal transformation of the coordinate system:

$$p = x \cos \theta + y \sin \theta$$

$$t = -x \sin \theta + y \cos \theta$$

the integral equals

$$G(K, \theta) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) e^{2\pi i K(x \cos \theta + y \sin \theta)} dx dy .$$

So we have established the identity

$$(2.1.7) \quad G(K, \theta) = F(K \cos \theta, K \sin \theta) .$$

Fourier inversion of (2.1.4) results in

$$(2.1.8) \quad f(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(X, Y) e^{-2\pi i (Xx + Yy)} dXdY .$$

After transformation to polar coordinates, we get with (2.1.7) and (2.1.8):

$$(2.1.9) \quad f(x, y) = \int_0^{2\pi} d\theta \int_0^{\infty} KG(K, \theta) e^{-2\pi i K(x \cos \theta + y \sin \theta)} dK$$

and with (2.1.6) we can write the inner integral as a Fourier integral over

the interval $(-\infty, +\infty)$ by

$$(2.1.10) \quad f(x,y) = \int_0^\pi h(x \cos \theta + y \sin \theta, \theta) d\theta ,$$

in which

$$(2.1.11) \quad h(t,\theta) := \int_{-\infty}^{+\infty} |K| F(K,\theta) e^{-2\pi i K t} dK .$$

Two ways to eliminate the integration over K exist, both of which are based on the convolution theorem of Fourier transformations, viz. we may consider $h(t,\theta)$ as the inverse Fourier transform of

$$\text{sign}(K) \times KG(K,\theta)$$

or alternatively of

$$|K| \times G(K,\theta) .$$

2.1.1. Radons Inversion Formula

(The content of this section is the twodimensional case of what is found in Ludwig [16]).

$KG(K,\theta)$ is the Fourier transform of $-\frac{1}{2\pi i} \frac{\partial g(p,\theta)}{\partial p}$ and $\text{sign}(K)$ may be considered as the Fourier transform of $1/(\pi i t)$ since

$$\text{sign}(K) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\sin(2\pi K t)}{t} dt = \frac{1}{\pi i} \int_{-\infty}^{\infty} \frac{e^{2\pi i K t}}{t} dt .$$

In the latter integral the Cauchy principal value should be taken.

If we formally apply the convolution theorem, we obtain

$$(2.1.12) \quad h(t) = \frac{1}{2\pi^2} \int_{-\infty}^{\infty} \frac{\partial g(p,t)}{\partial p} \frac{dp}{t-p}$$

but we will prove this somewhat more rigorously, in the meanwhile obtaining some alternatives for (2.1.12). Since

$$|K| = \pi^{-2} \int_0^{\infty} p^{-2} (1 - \cos 2\pi K p) dp ,$$

we have from (2.1.11)

$$h(t) = \pi^{-2} \int_{-\infty}^{\infty} G(K, \theta) e^{-2\pi i K t} dK \int_0^{\infty} p^{-2} (1 - \cos 2\pi K p) dp .$$

Interchanging of the order of integration is allowed if $g \in S$ for then also $G \in S$, hence G is absolutely integrable and G falls off at least as K^{-2} as $|K| \rightarrow \infty$. Hence

$$\begin{aligned} (2.1.13) \quad h(t) &= \pi^{-2} \int_0^{\infty} p^{-2} dp \int_{-\infty}^{\infty} G(K, \theta) e^{-2\pi i K t} (1 - \cos 2\pi K p) dK = \\ &= (2\pi^2)^{-1} \int_0^{\infty} p^{-2} [2g(t, \theta) - g(t+p, \theta) - g(t-p, \theta)] dp . \end{aligned}$$

Since g is twice differentiable, the integral exists in the ordinary sense. Alternatively, we may write

$$(2.1.14) \quad h(t) = (2\pi^2)^{-1} \int_{-\infty}^{\infty} p^{-2} [g(t, \theta) - g(t+p, \theta)] dp ,$$

or, by partial integration

$$h(t) = -(2\pi^2)^{-1} \int p^{-1} \frac{\partial g(t+p, \theta)}{\partial p} dp = (2\pi^2)^{-1} \int \frac{\partial g(s, \theta)}{\partial p} \frac{ds}{t-s} ,$$

which is (2.1.12). Together with (2.1.10) we obtain the so-called Radon inversion formule (Radon [19])

$$(2.1.15) \quad f(x, y) = (2\pi^2)^{-1} \int_0^{\pi} d\theta \int_{-\infty}^{\infty} (\partial g(p, \theta) \partial p) (x \cos \theta + y \sin \theta - p)^{-1} dp .$$

It is possible to extend (2.1.15) to much larger function classes than S , for instance, to all functions $f \in L^2(\mathbb{D})$ where \mathbb{D} is a compact convex set in \mathbb{R}^2 . The question then arises which condition a function must obey to be the Radon transform of an $f \in L^2(\mathbb{D})$ and whether this f is represented by (2.1.15). This question is answered in section 6.4.

2.2. Bandlimited functions

We return to formula (2.1.11)

$$h(t, \theta) = \int_{-\infty}^{+\infty} |K| G(K, \theta) e^{-2\pi i K t} dK .$$

Suppose that $g(p, \theta)$ is bandlimited, i.e. a constant A exists such that

$$(2.2.1) \quad G(K, \theta) = 0 \quad \text{if } |K| > A/2 .$$

So we may write h as

$$(2.2.2) \quad h(t, \theta) = \int_{-\infty}^{+\infty} H_A(K) G(K, \theta) e^{-2\pi i K t} dK$$

where

$$(2.2.3) \quad H_A(K) = \begin{cases} |K| & \text{if } |K| \leq A/2 \\ 0 & \text{if } |K| > A/2 . \end{cases}$$

Let $H_A(K)$ be the Fourier transform of $h_A(p)$:

$$(2.2.4) \quad h_A(p) = \int_{-\infty}^{+\infty} H_A(K) e^{-2\pi i K p} dK = (A/2) \{ A \operatorname{sinc}(Ap) - (A/2) \operatorname{sinc}^2(Ap/2) \}$$

where

$$(2.2.5) \quad \operatorname{sinc} x := (\sin \pi x) / (\pi x) .$$

Then the convolution theorem gives us

$$(2.2.6) \quad h(t, \theta) = (A/2) \int_{-\infty}^{+\infty} g(p, \theta) \{ A \operatorname{sinc}[A(t - p)] - (A/2) \operatorname{sinc}^2[A(t - p)/2] \} dp .$$

So essentially we have split $H_A(K)$ into a constant function and a finite ramp function inside the region $|K| \leq A/2$:

$$|K| = A/2 - A/2(1 - 2|K|/A)$$

and we have determined the inverse transform of each of these terms. Since g is bandlimited, formula (2.2.6) may be simplified as follows:

$$A \int_{-\infty}^{+\infty} g(p, \theta) \operatorname{sinc}[A(t - p)] dp = \int_{-A/2}^{A/2} G(K, \theta) e^{-2\pi i K t} dK$$

because the Fourier transform of $A \operatorname{sinc}(Ap)$ is unity inside the region $|K| \leq A/2$ and zero outside of it.

The integral on the right hand side also equals

$$\int_{-\infty}^{+\infty} G(K, \theta) e^{-2\pi i K t} dK$$

and this equals $g(t, \theta)$.

Then (2.2.6) reduces to:

$$(2.2.7) \quad h(t, \theta) = A/2 \{ g(t, \theta) - A/2 \int_{-\infty}^{+\infty} g(p, \theta) \operatorname{sinc}^2[A(t-p)/2] dp \} .$$

The solution (2.1.10) then becomes, after a shift of the integration over length p :

$$(2.2.8) \quad f(x, y) = A/2 \int_0^\pi d\theta \{ g(x \cos \theta + y \sin \theta, \theta) - \int_{-\infty}^{+\infty} g(p + x \cos \theta + y \sin \theta, \theta) \times (A/2) \operatorname{sinc}^2(Ap/2) dp \} .$$

This formula was established by Bracewell & Riddle ([2]).

2.3. Again Bandlimited functions: The solution by means of infinite series

Again we suppose that g is bandlimited, i.e.

$$G(K, \theta) = 0 \quad \text{of } |K| > A/2 .$$

Then from the Whittaker-Shannon theorem ([22], [25]) it follows that

$$(2.3.1) \quad g(p, \theta) = \sum_{m=-\infty}^{+\infty} g(m/A, \theta) \operatorname{sinc}[A(p - m/A)] , \quad -\infty < p < \infty ,$$

and

$$(2.3.2) \quad G(K, \theta) = 1/A \sum_{m=-\infty}^{+\infty} g(m/A, \theta) \exp[2\pi i m K/A] , \quad |K| < A/2 .$$

Since $|K|G(K, \theta)$ is the Fourier transform of $h(t, \theta)$, $h(t, \theta)$ is bandlimited too; so

$$(2.3.3) \quad h(t, \theta) = \sum_{m=-\infty}^{+\infty} h(m/A, \theta) \operatorname{sinc}[A(t - m/A)] .$$

Substituting (2.3.2) into (2.1.11) yields

$$(2.3.4) \quad h(n/A, \theta) = 1/A \sum_{m=-\infty}^{+\infty} g(m/A, \theta) \int_{-A/2}^{A/2} |K| e^{2\pi i K(m-n)/A} dK .$$

And

$$(2.3.5) \quad \int_{-A/2}^{A/2} |K| e^{2\pi i K m/A} dK = \begin{cases} A^2/4 & \text{if } m = 0 \\ 0 & \text{if } m \text{ even, } \neq 0 \\ -A^2/(\pi m)^2 & \text{if } m \text{ odd} . \end{cases}$$

So

$$(2.3.6) \quad h(n/A, \theta) = A/4 [g(n/A, \theta) - 4/\pi^2 \sum_{m \text{ odd}} m^{-2} g((m+n)/A, \theta)] .$$

Together with (2.3.3) this results in

$$(2.3.7) \quad h(t, \theta) = A/4 \left[\sum_{n=-\infty}^{+\infty} g(n/A, \theta) \operatorname{sinc}[A(t - n/A)] + \right. \\ \left. - 4/\pi^2 \sum_{n=-\infty}^{+\infty} \left\{ \sum_{m \text{ odd}} m^{-2} g((m+n)/A, \theta) \operatorname{sinc}[A(t - n/A)] \right\} \right] .$$

The first series equals $g(t, \theta)$, according to (2.3.1).

Changing the summation order in the double series yields

$$(2.3.8) \quad \sum_{m \text{ odd}} \left\{ \sum_{n=-\infty}^{+\infty} g((m+n)/A) \operatorname{sinc}[A(t - n/A)] \right\} / m^2 .$$

$g(t, \theta)$ is a bandlimited function, so $g(t + m/A, \theta)$ is bandlimited too, since its Fourier transform equals $\exp(-2\pi i K m/A) G(K, \theta)$. Therefore the inner sum in (2.3.8) is equal to $g(t + m/A, \theta)$.

Finally (2.3.7) becomes

$$(2.3.9) \quad h(t, \theta) = A/4 \left\{ g(t, \theta) - 4/\pi^2 \sum_{m \text{ odd}} g(t + m/A, \theta) / m^2 \right\} .$$

This expression may be interpreted as a discretization of the integral in (2.2.7) with discretization points $p_m = m/A$, m integer. (With $\tilde{p}_m = m/A$, m integer, we loose a factor 2 for the second integral!)

(2.3.9) is the extension of (2.3.6) to all values of t , which formula was deduced by Ramachandran and Lakshminarayanan ([20], [21]).

(2.1.10) gives the solution in the form

$$(2.3.10) \quad f(x,y) = A/4 \int_0^\pi d\theta \{g(x \cos \theta + y \sin \theta, \theta) - 4/\pi^2 \sum_{m \text{ odd}} m^{-2} g(x \cos \theta + y \sin \theta + m/A, \theta)\} .$$

Remark. (2.3.9) may seem to be in disagreement with (2.2.7) since the factors $A/4$ and $A/2$ are quite different. This is an allusion, however. We will show that, actually, both formulae are closely related to (2.1.14). Since

$$\int_{-\infty}^{\infty} (A/2) \text{sinc}^2(Ap/2) dp = 1 ,$$

(2.2.7) is equivalent to

$$(2.3.11) \quad h(t,\theta) = (A/2)^2 \int_{-\infty}^{\infty} [g(t,\theta) - g(t+p,\theta)] \text{sinc}^2(Ap/2) dp = \\ = (1/\pi^2) \int_{-\infty}^{\infty} p^{-2} [g(t,\theta) - g(t+p,\theta)] \sin^2(Ap/2) dp .$$

And since

$$\sum_{m \text{ odd}} m^{-2} = \pi^2/4 ,$$

(2.3.9) is equivalent to

$$(2.3.12) \quad h(t,\theta) = (A/\pi^2) \sum_{m \text{ odd}} m^{-2} [g(t,\theta) - g(t+m/A,\theta)] .$$

We observe that (2.3.11) turns into (2.1.14) if we let $A \rightarrow \infty$, since the mean value of $\sin^2(\pi Ap/2)$ over an interval of length $2/A$ is $1/2$. Further, discretisation of (2.3.11) in the points $p_m = m/A$ (m recurring through the integers) yields (2.3.12). And finally, if (2.1.14) is discretised in the points $p_m = m/A$, with now m running through the odd integers, the result is again (2.3.12). So (2.3.12) or (2.3.9) which hold exactly if $g(p,\theta)$ is bound-limited ($G(K,\theta) = 0$ if $|K| > A/2$) can be regarded as a genuine discretisation of Radon's inversion

formula (2.1.14) with step side $2/A$, the mesh being chosen so that $p = 0$ is just in the middle of two mesh points.

2.4. Back-projection

In the literature a method called back-projection is presented in various forms.

Let $g(p, \theta)$ be defined as in (2.1.1) and let

$$(2.4.1) \quad \tilde{f}(x, y) := \int_0^\pi g(x \cos \theta + y \sin \theta, \theta) d\theta,$$

i.e., $\tilde{f}(x, y)$ is the integral of the values of g over all rays that pass through the point (x, y) . Hence it might be hoped that $\tilde{f}(x, y)$, which is called the back-projection of the observed data $g(p, \theta)$, more over less resembles $f(x, y)$. We will classify the connection (cf. also Zurick and Zeitler, [26]). From the definitions (2.4.1) and (2.1.1) we have

$$\tilde{f}(x, y) = \int_0^\pi d\theta \int_{-\infty}^{\infty} f((x \cos \theta + y \sin \theta) \cos \theta - t \sin \theta, (x \cos \theta + y \sin \theta) \sin \theta + t \cos \theta) dt.$$

Or, substituting $t = -x \cos \theta + y \sin \theta + \rho$,

$$\begin{aligned} \tilde{f}(x, y) &= \int_0^\pi d\theta \int_{-\infty}^{\infty} f(x - \rho \sin \theta, y + \rho \cos \theta) d\rho = \\ &= \int_0^{2\pi} d\theta \int_0^{\infty} \rho^{-1} f(x - \rho \sin \theta, y + \rho \cos \theta) \rho d\rho. \end{aligned}$$

This formula may be interpreted as an integral over \mathbb{R}^2 written in polar coordinates. Passing to rectangular coordinates $\xi = \rho \sin \theta$, $\eta = -\rho \cos \theta$ we find

$$(2.4.2) \quad \tilde{f}(x, y) = \iint_{-\infty}^{\infty} \frac{f(x - \xi, y - \eta)}{\sqrt{\xi^2 + \eta^2}} d\xi d\eta.$$

Hence (Gilbert [7]) $\tilde{f}(x, y)$ is obtained by convolution of $f(x, y)$ with the function $(x^2 + y^2)^{-1}$.

The result (2.4.2) can also be obtained by Fourier theory. From (2.1.5) and (2.1.7) we have

$$g(p, \theta) = \int_{-\infty}^{\infty} F(K \cos \theta, K \sin \theta) e^{-2\pi i K p} dK,$$

hence

$$\begin{aligned} \tilde{f}(x, y) &= \int_0^{\pi} d\theta \int_{-\infty}^{\infty} F(K \cos \theta, K \sin \theta) e^{-2\pi i K(x \cos \theta + y \sin \theta)} dK = \\ &= \int_0^{2\pi} \int_0^{\infty} \dots = \iint_{-\infty}^{\infty} \frac{F(X, Y)}{\sqrt{X^2 + Y^2}} e^{-2\pi i(Xx + Yy)} dXdY. \end{aligned}$$

So the Fourier transform of $\tilde{f}(x, y)$ is

$$(2.4.3) \quad \tilde{F}(X, Y) = (X^2 + Y^2)^{-\frac{1}{2}} F(X, Y),$$

which is in accordance with (2.4.2) since the Fourier transform of $(x^2 + y^2)^{-\frac{1}{2}}$ is

$$\begin{aligned} \iint_{-\infty}^{\infty} (x^2 + y^2)^{-\frac{1}{2}} e^{2\pi i(Xx + Yy)} dx dy &= \int_0^{2\pi} d\theta \int_0^{\infty} e^{2\pi i r(X \cos \theta + Y \sin \theta)} dr = \\ &= 2\pi \int_0^{\infty} J_0(2\pi r(X^2 + Y^2)^{\frac{1}{2}}) dr = (X^2 + Y^2)^{-\frac{1}{2}}. \end{aligned}$$

It follows from (2.4.2) that the resemblance between $\tilde{f}(x, y)$ and $f(x, y)$ is, generally, rather superficial. If $f(x, y) = \delta(x - x_0)\delta(y - y_0)$, i.e., a δ -peak in (x_0, y_0) then

$$\tilde{f}(x, y) = ((x - x_0)^2 + (y - y_0)^2)^{-\frac{1}{2}},$$

which also has its singularity in (x_0, y_0) , but falls off much less rapidly than $f(x, y)$.

The difference between \tilde{f} and f is reflected in the difference between the function g and h occurring in the relations (2.4.1) and (2.1.10), respectively, which is most obvious from the relation between their Fourier transforms which follows from (2.1.11)

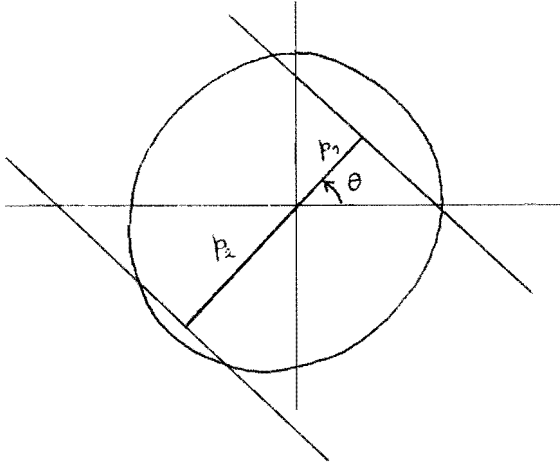
$$H(K,\theta) = |K|G(K,\theta) .$$

We end this section by mentioning that the fact that $\tilde{f}(x,y)$ occurs as a term in (2.2.8) and (2.3.10) (although multiplied with factors $A/2$ and $A/4$, respectively) seems to have given rise to some confusion in the literature.

Chapter 3. The discrete problem

3.1. Sampling of the Radon transform

We describe the sampling of $g(p, \theta)$. For a fixed value of θ , values are obtained for $g(p, \theta)$ for a discrete set of p -values, such that the p cover the interval $[-1, +1]$ as well as possible. The result is called a projection. This procedure is repeated for several angles θ . So the sampling



consists of projections and each projection consist of a number of estimates. The angles θ may cover uniformly the interval $[0, \pi]$, but this is not necessary with all methods to solve the problem. The angles θ are called the projection directions.

3.2. Fourier methods

With Fourier methods we indicate those methods in which approximations of the Fourier transform are obtained explicitly and the inversion of (2.1.9) is carried out in some way. The Fourier transform of $g(p, \theta)$ may be obtained, or, more precisely, approximated, with (2.1.5) but only in the projection directions θ . If these are not distributed uniformly on the interval $[0, \pi]$ a very accurate way of interpolation must be used.

3.2.1. Interpolation with sincfunctions

Since f has a compact support in all practical applications, the Whittaker-Shannon theorem may be applied to the Fourier transform of f :

$$(3.2.1) \quad F(X, Y) = \sum_{m=-\infty}^{+\infty} \sum_{\ell=-\infty}^{+\infty} F(m/A, \ell/B) \text{sinc}[A(X - m/A)] \text{sinc}[B(Y - \ell/B)]$$

in which A and B must satisfy the condition

$$(3.2.2) \quad f(x,y) = 0 \text{ if } |x| > A/2 \text{ or if } |y| > B/2 .$$

In practice we have to replace the infinite series by a finite one. So we put

$$(3.2.3) \quad \tilde{F}(X,Y) := \sum_{m=-M}^{+M} \sum_{\ell=-L}^{+L} \tilde{F}(m/A, \ell/B) \text{sinc}[A(X - m/A)] \text{sinc}[B(Y - \ell/B)]$$

and construct a set of linear equations by assuming:

$$(3.2.4) \quad \tilde{F}(X,Y) = F_s(X,Y)$$

for those values of (X,Y) for which approximations $F_s(X,Y)$ are known for the real $F(X,Y)$.

The solution, in some sense, may be obtained and then the calculated $\tilde{F}(m/A, \ell/B)$ are approximations to $F(m/A, \ell/B)$. The estimate for $f(x,y)$ is then

$$(3.2.5) \quad \tilde{f}(x,y) = 4/AB \sum_{\ell=-L}^{+L} \sum_{m=-M}^{+M} \tilde{F}(m/A, \ell/B) e^{-2\pi i(mx/A + \ell y/B)}$$

or, alternatively, one may use the $\tilde{F}(m/A, \ell/B)$ for a discretization of integral (2.1.9) (see [3] and [23]).

3.2.2. Interpolation with circle functions

Different from the previous method but very much alike is the "Fourier-Bessel" method. Let

$$(3.2.6) \quad F(K \cos \theta, K \sin \theta) = \sum_{n=-\infty}^{+\infty} F_n(K) e^{in(\theta + \pi/2)} .$$

Then it follows that

$$(3.2.7) \quad f(r \cos \psi, r \sin \psi) = \sum_{n=-\infty}^{+\infty} f_n(r) e^{in\psi}$$

in which

$$(3.2.8) \quad f_n(r) = \int_0^{\infty} F_n(K) J_n(2\pi Kr) dK .$$

Again we take a finite series instead of (3.2.6) and construct a set of linear equations

$$(3.2.9) \quad F_s(K_I \cos \theta_J, K_I \sin \theta_J) = \sum_{n=-N}^{+N} \tilde{F}_n(K_I) e^{in(\theta_J + \pi/2)}$$

where the θ_J are the projection directions and $F_s(K_I \cos \theta_J, K_I \sin \theta_J)$ are known approximations for $F(K_I \cos \theta_J, K_I \sin \theta_J)$. From (3.2.9) we obtain estimates $\tilde{F}_n(K_I)$ for the real $F_n(K_I)$ e.g. with least squares. If the θ_J are given by $\theta_J = J/2N$, $J = 1, \dots, 2N$ then the unknown $\tilde{F}_n(K_I)$ are found as finite Fourier series with the $F_s(K \cos \theta_J, K \sin \theta_J)$ as coefficients. With $\tilde{F}_n(K_I)$ the integrals in (3.2.8) are calculated. An estimate for f is then given by

$$\tilde{f}(r \cos \psi, r \sin \psi) = \sum_{n=-N}^{+N} \tilde{f}_n(r) e^{in\psi}$$

for those values of r for which (3.2.9) is calculated (see [3], [4]).

3.3. Convolution methods

The convolution methods find their bases in the solutions given as convolution sums and integrals in the sections 2.1.1, 2.2 and 2.3.

Formula (2.1.15) is deduced by Junginger and Van Haeringen ([14]) and Gilbert ([7]). However, in numerical evaluation of the integral, loss of significant digits will occur because of the numerically unstable behaviour of the integrand for small p . Formula (2.2.8) was derived by Bracewell and Riddle ([2]). We notice that the formula is not well suited for numerical evaluation if A is large. Since

$$(A/2) \int_{-\infty}^{+\infty} \text{sinc}^2(Ap/2) dp = 1$$

loss of significant digits will occur.

Formula (2.3.10), or more precisely (2.3.6), was established by Ramachandran, Lakshminarayanan ([20]). Again for large A formula (2.3.10) is not well fit to numerical evaluation because of

$$4/\pi^2 \sum_{m \text{ odd}} m^{-2} = 1.$$

The most appealing side of these methods is that Fourier integrations are avoided and replaced by integrals over parameters that are relevant to the problem.

We want to make a final note on the stability to noisy data of the methods of Bracewell et.al. and Ramachandran et.al. From the theory of chapter 2 this is easily understood.

If we are solving the problem numerically we must choose an effective bandwidth $A/2$ i.e. we put $G(K,\theta) = 0$ if $|K| > A/2$. If g contains noise the main contribution of the Fourier transform of this noise will lie in the region $|K| > A/2$ and hardly in $|K| < A/2$. So, to some extent, the noise is filtered. Another type of convolution method (or rather, deconvolution method) can be based on the relation (2.4.2) between $f(x,y)$ and the back-projection $\tilde{f}(x,y)$ belonging to $g(p,\theta)$ as defined in (2.4.1). Since $\tilde{f}(x,y)$ is obtained relatively easily from $g(p,\theta)$ (more precisely, with the same effort as with which $f(x,y)$ can be obtained from $h(p,\theta)$). Hence $f(x,y)$ can be obtained by inversion of the convolution operation in (2.4.2) which is most easily performed in Fourier spaces, using (2.4.3). Of course, some filtering would be in order; this can easily be done by multiplication of the Fourier transform of $f(x,y)$ by some bell-shaped function.

Using Fast Fourier algorithms this method seems quite simple and attractive. However, once the decision to use numerical Fourier transforms of sampler data is made, it seems slightly more attractive to use (2.1.5), (2.1.11) (with filtering) and (2.1.10).

Chapter 4. The problem as a set of linear equations

4.1. Formulation of the problem as a set of equations

We describe here a number of methods, all of which are characterized by the following.

Let \mathbb{D} a convex compact set in \mathbb{R}^2 and $L^2(\mathbb{D})$ the space of square Lebesgue integrable functions on \mathbb{D} . Let $f \in L^2(\mathbb{D})$ and g the Radon transform of f , which is sampled in the points (p_j, θ_j) , $j = 1, \dots, N$.

Choose a number of functions $B_m(x, y)$, $m = 1, \dots, M$ in $L^2(\mathbb{D})$ with Radon transforms $C_m(p, \theta)$.

We now want to approximate f by a linear combination of the basis functions $B_m(x, y)$

$$(4.1.1) \quad \tilde{f}(x, y) = \sum_{m=1}^M a_m B_m(x, y) .$$

The coefficients a_m have to be determined such that \tilde{f} fits the samples of g as well as possible. Radon transforming (4.1.1) yields

$$(4.1.2) \quad \tilde{g}(p, \theta) = \sum_{m=1}^M a_m C_m(p, \theta)$$

and therefore, the unknowns a_m are the solution, in some sense, of

$$(4.1.3) \quad \sum_{m=1}^M a_m C_m(p_j, \theta_j) = g(p_j, \theta_j), \quad j = 1, \dots, N .$$

If this set of equations is not too large, standard solution methods may be applied.

The main point of these methods is how to choose the basis functions $B_m(x, y)$ since the quality of the approximation will depend highly on this choice. We notice that we can take an orthogonal set of $B_m(x, y)$ (with a certain weight function). In general this has no effect on the solution of (4.1.3), unless the $C_m(p, \theta)$ are orthogonal too. In that case the a_m can be calculated from (4.1.2) by taking inner products. (See section 4.4 and chapter 6). We now discuss several choices for the $B_m(x, y)$.

4.2. Sincfunctions

The choice of sincfunctions as basis functions is inspired again by the Whittaker-Shannon theorem.

As basis functions we take

$$(4.2.1) \quad B_{m,\ell}(x,y) = \text{sinc}[A(x - m/A)]\text{sinc}[B(y - \ell/B)]$$

and

$$(4.2.2) \quad \tilde{f}(x,y) = \sum_{m=-M}^{+M} \sum_{\ell=-L}^{+L} a_{m,\ell} B_{m,\ell}(x,y) .$$

If we put $a_{m,\ell} = f(m/A, \ell/B)$ this is the truncated form of the series we met in 2.1.3.

We notice that (4.2.2) implies that \tilde{f} is bandlimited and therefore cannot have a compact support. As an approximation to f it might be satisfactory however. The Radon transform of $B_{m,\ell}$ is given by Sweeney & Vest ([22]):

$$(4.2.3) \quad C_{m,\ell}(p,\theta) = \begin{cases} (A|\cos \theta|)^{-1} \text{sinc}[B(p \sec \theta + (m \tan \theta)/A + \ell/B)] & \text{if} \\ \quad 0 \leq |\tan \theta| \leq A/B \\ (B|\sin \theta|)^{-1} \text{sinc}[A(p \csc(\theta) + m/A - \ell/B \cot \theta)] & \text{if} \\ \quad A/B \leq |\tan \theta| < \infty \\ \text{sinc}[B(p + m/A)] & \text{if } \tan \theta = \pm \infty \end{cases}$$

Radon transforming (4.2.2) gives

$$(4.2.4) \quad \tilde{g}(p,\theta) = \sum_{m=-M}^{+M} \sum_{\ell=-L}^{+L} a_{m,\ell} C_{m,\ell}(p,\theta) .$$

And by putting:

$$(4.2.5) \quad \tilde{g}(p_j, \theta_j) = g(p_j, \theta_j), \quad j = 1, \dots, M$$

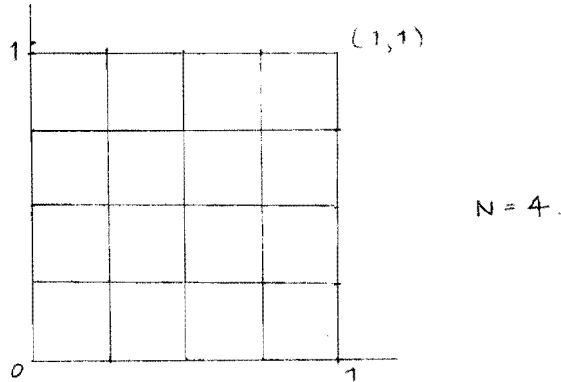
we obtain a set of linear equations in the unknown $a_{m,\ell}$. In general N will exceed $(2M+1)(2L+1)$ i.e. there are more equations than unknowns, so it is possible that no solution of (4.2.5) exists at all, or that infinitely many solutions exist. The method of least squares is then appropriate to obtain a solution, also with respect to smoothing if noisy data are used. The main disadvantage is that it is expensive to construct the coefficients of the equations, if N is large and therefore M and L may be large.

This method is applied to Sweeney & Vest ([23]) but only for small M and L (100 equations in 36 unknowns).

4.3. Stepfunctions

Another choice is the following.

Suppose that \mathbb{D} is the unit square in the first quadrant with one vertex in the origin.



Split the square into N equal squares with sides of length $1/N$. The (m, ℓ) -th square is the square with centre $((m - \frac{1}{2})/N, (\ell - \frac{1}{2})/N)$ or notated as a set

$$(4.3.1) \quad \mathbb{D}_{m, \ell} = \{(x, y) \mid |x - (m - \frac{1}{2})/N| < 1/2N, |y - (\ell - \frac{1}{2})/N| < 1/2N\} .$$

Now let the basisfunctions be defined by

$$(4.3.2) \quad B_{m, \ell}(x, y) = \begin{cases} 1 & \text{if } (x, y) \in \mathbb{D}_{m, \ell} \\ 0 & \text{if } (x, y) \notin \mathbb{D}_{m, \ell} \end{cases}$$

and

$$(4.3.3) \quad \tilde{f}(x, y) = \sum_{m=1}^N \sum_{\ell=1}^N a_{m, \ell} B_{m, \ell}(x, y) .$$

If $a_{m, \ell}$ equals the mean of f over the (m, ℓ) -th square f would be a very good approximation to f (N large).

Analogous to section 4.2 we may construct a set of linear equations. Let $C_{m, \ell}(p, \theta)$ be the Radon transform of $B_{m, \ell}(x, y)$. Then (4.3.3) becomes, after Radon transformation,

$$(4.3.5) \quad \tilde{g}(p, \theta) = \sum_{m=1}^N \sum_{\ell=1}^N a_{m, \ell} C_{m, \ell}(p, \theta)$$

and we obtain a set of linear equations by putting

$$(4.3.6) \quad \tilde{g}(p_j, \theta_j) = g(p_j, \theta_j), \quad j = 1, \dots, M .$$

Since $C_{m,\ell}(p,\theta) = 0$, if the line with parameter representation (p,θ) does not intersect the (m,ℓ) -th square, it is possible that some $a_{m,\ell}$ do not occur in the equations (4.3.6). This then would show that the squares are chosen too small.

Finally we show how the $C_{m,\ell}(p,\theta)$ may be calculated.

(i) First we calculate the Radon transform of the following function: Let

$$S_{\alpha,\beta} := \{(x,y) \mid |x| < \alpha \ \& \ |y| < \beta\}$$

and

$$(4.3.7) \quad h_{\alpha,\beta} = \begin{cases} 1 & \text{if } (x,y) \in S_{\alpha,\beta} \\ 0 & \text{if } (x,y) \notin S_{\alpha,\beta} \end{cases}.$$

The Radon transform $g_{\alpha,\beta}(p,\theta)$ of $h_{\alpha,\beta}$ equals the length of the line segment of the line (p,θ) that lies within the rectangle $S_{\alpha,\beta}$. So the Radon transform may be "calculated" from a diagram but it is difficult to translate this in analytical terms. So we use the theory of 2.1.

The Fourier transform of $h_{\alpha,\beta}$ equals

$$(4.3.8) \quad H(X,Y) = \frac{\sin 2\pi\alpha X}{\pi X} \cdot \frac{\sin 2\pi\beta Y}{\pi Y}$$

and so

$$(4.3.9) \quad g_{\alpha,\beta}(p,\theta) = \int_{-\infty}^{+\infty} \frac{\sin[2\pi\alpha K |\cos \theta|]}{\pi K |\cos \theta|} \cdot \frac{\sin[2\pi\beta K |\sin \theta|]}{\pi K |\sin \theta|} e^{-2\pi i K p} dK$$

so g is the convolution of two functions, the Fourier transforms of which equals the first resp. the second factor of the integrand in (4.3.9).

The following holds if $|\cos \theta| \neq 0$

$$(4.3.10) \quad \int_{-\infty}^{+\infty} \frac{\sin[2\pi\alpha K |\cos \theta|]}{\pi K |\cos \theta|} e^{-2\pi i K p} dK = \frac{1}{|\cos \theta|} H\left(\frac{p}{2\alpha |\cos \theta|}\right)$$

where

$$(4.3.11) \quad H(x) := \begin{cases} 1 & \text{if } |x| \leq \frac{1}{2} \\ 0 & \text{if } |x| > \frac{1}{2} \end{cases}$$

and: if $|\sin \theta| \neq 0$

$$(4.3.12) \quad \int_{-\infty}^{+\infty} \frac{\sin[2\pi\beta K |\sin \theta|]}{\pi K |\sin \theta|} e^{-2\pi i K p} dK = \frac{1}{|\sin \theta|} H\left(\frac{p}{2\beta |\sin \theta|}\right) .$$

So we obtain for $\cos \theta \sin \theta \neq 0$

$$(4.3.13) \quad g_{\alpha, \beta}(p, \theta) = \frac{1}{|\cos \theta \sin \theta|} \int_{-\infty}^{+\infty} H\left(\frac{p-t}{2\alpha |\cos \theta|}\right) H\left(\frac{t}{2\beta |\sin \theta|}\right) dt .$$

This integral is nothing but the length of the intersection of the intervals

$$[p - \alpha |\cos \theta|, p + \alpha |\cos \theta|] \text{ and } [-\beta |\sin \theta|, \beta |\sin \theta|] ,$$

resulting in

$$(4.3.14) \quad g_{\alpha, \beta}(p, \theta) = \begin{cases} 0 & \text{if } p < -\alpha |\cos \theta| - \beta |\sin \theta| \\ \frac{p + \alpha |\cos \theta| + \beta |\sin \theta|}{|\sin \theta| |\cos \theta|} & \text{if } \alpha |\cos \theta| - \beta |\sin \theta| < p \leq \\ & \leq -|\beta |\sin \theta| - \alpha |\cos \theta| \\ \frac{2\alpha}{|\sin \theta|} & \text{if } \alpha |\cos \theta| - \beta |\sin \theta| < p \leq \\ & \leq -\alpha |\cos \theta| + \beta |\sin \theta| \\ \frac{2\alpha}{|\cos \theta|} & \text{if } -\alpha |\cos \theta| + \beta |\sin \theta| < p \leq \\ & \leq \alpha |\cos \theta| - \beta |\sin \theta| \\ \frac{-p + \alpha |\cos \theta| + \beta |\sin \theta|}{|\sin \theta| |\cos \theta|} & \text{if } +|\alpha |\cos \theta| - \beta |\sin \theta| < p \leq \\ & \leq \alpha |\cos \theta| + \beta |\sin \theta| \\ 0 & \text{if } p < \alpha |\cos \theta| + \beta |\sin \theta| . \end{cases}$$

The special cases $\cos \theta = 0$ and $\sin \theta = 0$ need no special treatment anymore.

(ii) By putting $\alpha = \beta = 1/N$ the dimensions of $S_{\alpha, \beta}$ agree with the dimensions of $\mathbb{D}_{m, \ell}$.

(iii) By the translation over the vector $((m - \frac{1}{2})/N, (\ell - \frac{1}{2})/N)$ $S_{1/N, 1/N}$ is transformed into $\mathbb{D}_{m, \ell}$ or equivalently:

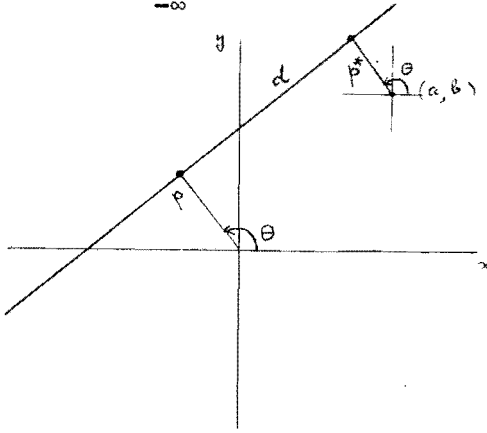
$$(4.3.15) \quad h_{1/N, 1/N}(x - (m - \frac{1}{2})/N, y - (\ell - \frac{1}{2})/N) = B_{m, \ell}(x, y) .$$

(iv) Therefore we want to know the connection of the Radon transforms of a function and the shifted function. The following holds:

Let $g(p, \theta)$ be the Radon transform of $f(x, y)$, then the Radon transform of $f(x - a, y - b)$ equals $g(p - a \cos \theta - b \sin \theta, \theta)$, for all real a, b .

Proof. The Radon transform of $f(x-a, y-b)$ is defined by

$$(4.3.16) \quad g^*(p, \theta) = \int_{-\infty}^{+\infty} f(p \cos \theta - t \sin \theta - a, p \sin \theta + t \cos \theta - b) dt .$$



From a picture it is clear that we have to shift the integration path over distance d :

$$(4.3.17) \quad g^*(p, \theta) = \int_{-\infty}^{+\infty} f(p \cos \theta - t \sin \theta - d \sin \theta - a, p \sin \theta + t \cos \theta + d \cos \theta - b) dt .$$

Now put:

$$(4.3.18) \quad \begin{cases} p^* \cos \theta = p \cos \theta - d \sin \theta - a \\ p^* \sin \theta = p \sin \theta + d \cos \theta - b \end{cases}$$

then, appararently,

$$(4.3.19) \quad g^*(p, \theta) = g(p^*, \theta) .$$

Now (4.3.18) are two equations in two unknowns viz. p^* and d . The unique solution is

$$p^* = p - a \cos \theta - b \sin \theta$$

$$d = -a \sin \theta + b \cos \theta$$

and so (4.3.19) becomes:

$$g^*(p, \theta) = g(p - a \cos \theta - b \sin \theta, \theta) .$$

(v) Combining the preceding results we obtain that the Radon transform $C_{m, \ell}(p, \theta)$ of $B_{m, \ell}(x, y)$ equals

$$(4.3.21) \quad C_{m,\ell}(p,\theta) = g_{1/N,1/N}(p - (m - \frac{1}{2})\cos \theta/N - (\ell - \frac{1}{2})\sin \theta/N, \theta)$$

where g is given by (4.3.14).

This method with stepfunctions is applied by Sweeney & Vest ([23]). The objections are the same as those concerning the method with sincfunctions. However more frequently used is the variant in which the finite ray width is explicitly taken into account. Because of the special way the resulting set of equations is solved we devote a separate chapter to it (chapter 5).

4.4. Series expansions: Orthogonal Radon transforms

Although (4.2.2), (4.2.3) and (4.3.2), (4.3.3) may be interpreted as series expansions for f , we want to connect the name series expansion only to those methods in which $\{B_{m,n}\}$ forms an total orthogonal system in $L^2(\mathbb{D})$ and the corresponding Radon transforms are orthogonal on $\mathbb{R} \times [0, 2\pi]$ in some sense. The further approach looks like that of section 4.1 but because of the orthogonality of the system $\{C_{m,m}\}$ we can avoid solving a set of linear equations. We describe the procedure in the light of the choice Matulka and Collins have made ([18]). In part II we treat more exhaustively the method introduced by Marr ([17]).

4.4.1. Continuous case

Let $\mathbb{D} := \mathbb{R}^2$ and $L^2(\mathbb{D}, W)$ the space of functions for which $f(x,y)\exp((x^2+y^2)/2)$ is square Lebesgue integrable on \mathbb{D} .

We define an inner product on $L^2(\mathbb{D}, W)$:

$$(4.4.1) \quad (f_1, f_2)_{\mathbb{D}} := \iint_{\mathbb{D}} f_1(x,y) \overline{f_2(x,y)} e^{x^2+y^2} dx dy .$$

Let $L^2(\mathbb{R} \times [0, 2\pi])$ be the space of functions g on $\mathbb{R} \times [0, 2\pi)$ for which $g(p,\theta)\exp(p^2/2)$ is square Lebesgue integrable on $\mathbb{R} \times [0, 2\pi)$.

We define an inner product in $L^2(\mathbb{R} \times [0, 2\pi))$

$$(4.4.2) \quad (g_1, g_2) := \int_{-\infty}^{\infty} dp \int_0^{2\pi} d\theta g_1(p,\theta) \overline{g_2(p,\theta)} e^{p^2} .$$

The Radon transform g of an $f \in L^2(\mathbb{D})$ is element of $L^2(\mathbb{R} \times [0, 2\pi))$ as maybe seen from the following.

Let $f \in L^2(\mathbb{D}, W)$ then

$$\|f\|_{\mathbb{D}}^2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |f(x,y)|^2 e^{x^2+y^2} dx dy .$$

After a rotation of the coordinate system we get

$$\|f\|_{\mathbb{D}}^2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |f(p \cos \theta - t \sin \theta, p \sin \theta + t \cos \theta)|^2 e^{p^2+t^2} dp dt .$$

The Cauchy-Schwarz inequality states that

$$(4.4.3) \quad \left| \int_{-\infty}^{+\infty} f(p \cos \theta - t \sin \theta, p \sin \theta + t \cos \theta) dt \right|^2 \leq \int_{-\infty}^{+\infty} e^{-t^2} dt \times \int_{-\infty}^{+\infty} e^{t^2} |f(p \cos \theta - t \sin \theta, p \sin \theta + t \cos \theta)|^2 dt$$

and so

$$\|f\|_{\mathbb{D}}^2 \geq \sqrt{\frac{1}{\pi}} \int_{-\infty}^{+\infty} e^{p^2} \left| \int_{-\infty}^{+\infty} f(p \cos \theta - t \sin \theta, p \sin \theta + t \cos \theta) dt \right|^2 dp$$

or equivalently

$$\|f\|_{\mathbb{D}}^2 \geq \sqrt{\frac{1}{\pi}} \int_{-\infty}^{+\infty} e^{p^2} |g(p, \theta)|^2 dp .$$

After integrating both sides over $\theta \in [0, 2\pi]$ we obtain

$$2\pi \|f\|_{\mathbb{D}}^2 \geq \sqrt{\frac{1}{\pi}} \int_0^{2\pi} d\theta \int_{-\infty}^{+\infty} e^{p^2} |g(p, \theta)|^2 dp = \sqrt{\frac{1}{\pi}} \|g\|^2$$

or:

$$(4.4.4) \quad 2\pi^{3/2} \|f\|_{\mathbb{D}}^2 \geq \|g\|^2$$

so

$$g \in L^2(\mathbb{R} \times [0, 2\pi)) .$$

It also follows that the Radon transformation is bounded, with norm $2\pi^{3/2}$. The weight function $e^{x^2+y^2}$ defines a set of orthogonal functions.

Let, if $-\infty < n < \infty$, $k \geq 0$

$$(4.4.5) \quad F_{n,k}(x,y) = (-1)^k \left(\frac{k!}{\pi(|n|+k)!} \right)^{\frac{1}{2}} (x + iy)^{|n|} L_k^{|n|} (x^2 + y^2) e^{-x^2 - y^2},$$

where $L_k^{|n|}$ are the Laguerre polynomials (see [24]).

In this expression the plus sign must be chosen if n is positive. A different representation is

$$F_{n,k}(r \cos \psi, r \sin \psi) = (-1)^k \left(\frac{k!}{\pi(|n|+k)!} \right)^{\frac{1}{2}} e^{in\psi} r^{|n|} L_k^{|n|} (r^2) e^{-r^2}.$$

The orthogonality relation is

$$(4.4.6) \quad (F_{n,k}, F_{m,\ell})_{\mathbb{D}} = \delta_{nm} \delta_{k\ell}.$$

$F_{n,k}(x,y)e^{x^2+y^2}$ is a polynomial of degree $|n| + 2k$. For every $m \geq 0$ exactly $\frac{1}{2}(m+1)(m+2)$ functions $F_{n,k}$ exist, for which $|n| + 2k \leq m$. Therefore the system $F_{n,k}(x,y)e^{x^2+y^2}$ spans the space of polynomials of degree m . So $F_{n,k}$, $-\infty < n < \infty$, $k \geq 0$ is a total orthogonal system in $L^2(\mathbb{D}, W)$. Thus, for any $f \in L^2(\mathbb{D}, W)$ numbers $a_{n,k}$ exist so that

$$(4.4.7) \quad f(x,y) = \sum_{m=-\infty}^{+\infty} \sum_{k=0}^{\infty} a_{n,k} F_{n,k}(x,y)$$

in the sense that

$$(4.4.8) \quad \lim_{N \rightarrow \infty} \| f - \sum_{n=-N}^{+N} \sum_{k=0}^N a_{n,k} F_{n,k} \|_{\mathbb{D}}^2 = 0.$$

The Radon transform of $F_{n,k}$ is given by

$$(4.4.9) \quad \widehat{F}_{n,k}(p, \theta) = [2^{2|n|+4k} k! (|n|+k)!]^{-\frac{1}{2}} e^{in\theta} H_{|n|+2k}(p) e^{-p^2}$$

where $H_m(p)$ is the Hermite polynomial of degree m (see [24]).

The $\widehat{F}_{n,k}$ form an orthogonal system too:

$$(4.4.10) \quad (\widehat{F}_{n,k}, \widehat{F}_{m,\ell}) = 2\pi^{3/2} \binom{|n|+2k}{k} 2^{-|n|-2k} \delta_{nm} \delta_{k\ell}.$$

From (4.4.4) and (4.4.8) it follows that the Radon transform g of f may be expressed as

$$(4.4.11) \quad g(p, \theta) = \sum_{n=-\infty}^{+\infty} \sum_{k=0}^{\infty} a_{n,k} \hat{F}_{n,k}(p, \theta)$$

in the sense that

$$(4.4.12) \quad \lim_{N \rightarrow \infty} \left\| g - \sum_{n=-N}^{+N} \sum_{k=0}^N a_{n,k} \hat{F}_{n,k} \right\|^2 = 0 .$$

From this it follows, with (4.4.10) that

$$(4.4.13) \quad a_{n,k} = \frac{2^{|n|+2k} (g, \hat{F}_{n,k})}{2^{3/2} \binom{|n|+2k}{k}} .$$

So (4.4.7) and (4.4.8) form the solution of the inversion problem: f is expressed as a "function" of g .

Note. From (4.4.9), (4.4.10) it follows that the Radon transformation is a bounded mapping of $L^2(\mathbb{D}, W)$ into $L^2(\mathbb{R} \times [0, 2\pi])$. It also follows that inverse Radon transformation is unbounded.

4.4.2. Discrete case

Of course we have to replace the infinite series in (4.4.4) by a finite one. The most plausible way is to take a polynomial of some degree M :

$$(4.4.14) \quad \hat{f}(x, y) = \sum_{m=-M}^{+M} \sum_{k=0}^{[(M-|n|)/2]} a_{n,k} F_{n,k}(x, y) .$$

Approximations for the $a_{n,k}$ may be calculated from (4.4.13) with the samples of g .

Since in practice we only meet functions f which have a bounded support it seems to be somewhat innaturally to have a set of basisfunctions with unbounded support. From this point of view the choice of Marr ([17]) is better.

Chapter 5. Algebraic Reconstruction Techniques (ART)

5.1. Finite ray width

In this section we follow the notation of section 4.3. As a consequence of the finite raywidth we obtain estimates of the Radon transform of f integrated over the raywidth, i.e. of the quantities

$$(5.1.1) \quad s(p, \theta) = \int_{p-w/2}^{p+w/2} g(q, \theta) dq .$$

This means that (4.3.3) after transformation to (4.3.5) must be integrated as in (5.1.1) to obtain the right set of equations. Because of the specific form of the $B_{m,\ell}(x,y)$ this means that we must know the area of the intersection of every small square with every ray.

5.2. The set of equations

Now we change the notation. We number the N^2 squares from 1 to N^2 . Let $b_n(x,y)$ denote the characteristic function of the n -th square. The estimate \tilde{f} to f now has the form:

$$(5.2.1) \quad \tilde{f}(x,y) = \sum_{n=1}^{N^2} a_n b_n(x,y) .$$

So we may represent \tilde{f} by a vector $a \in \mathbb{R}^{N^2}$. Also we number the rays, suppose from 1 to M . Let $d_{m,n}$ be the area of the intersection of the m -th ray and the n -th square. Then (5.2.1) becomes after integration over the m -th ray

$$(5.2.2) \quad s_m := \int_{p_m-w_m/2}^{p_m+w_m/2} \tilde{g}(q, \theta_m) dq = \sum_{n=1}^{N^2} a_n d_{m,n}$$

So the set of equations is

$$(5.2.3) \quad \sum_{n=1}^{N^2} d_{m,n} a_n = s_m, \quad m = 1, \dots, M .$$

Or, in matrix-vector notation.

$$(5.2.4) \quad Da = s .$$

It is clear that the structure of D will be complicated. Gordon et.al. advocate that instead of D one may use the matrix E , defined by

$$(5.2.5) \quad E_{m,n} := \begin{cases} 1/N^2 & \text{if } D_{m,n} \geq 1/(2N^2) \\ 0 & \text{if } D_{m,n} < 1/(2N^2) . \end{cases}$$

Then (5.2.4) is replaced by

$$(5.2.6) \quad Ea = s .$$

Geometrically this means that we take the area of the intersection of the m -th ray and n -th square equal to the total area of the n -th square if the centre of the square lies within the ray. Especially near the boundary of the ray this may result in large differences.

In the next section we present an iterative method to solve (5.2.6).

5.3. ART

The solution method which is known by the name of ART originates from Gordon, Bender and Herman ([8]). The idea is the following.

We have M equations in N^2 unknowns,

$$(5.3.1) \quad (d_m, x) = s_m, \quad m = 1, \dots, M ,$$

where d_m is the m -th row of D .

If we choose a vector $x^{(0)} \in \mathbb{R}^{N^2}$, in general this vector will not satisfy equations (5.3.1). We calculate the residu of the first equation: $s_1 - (d_1, x^{(0)})$ and distribute this among the unknowns of the first equation, proportional to their coefficient, to obtain $x^{(1)}$:

$$(5.3.2) \quad x^{(1)} = x^{(0)} + d_1 (s_1 - (d_1, x^{(0)})) / \|d_1\|^2 .$$

So $(d_1, x^{(1)}) = s_1$.

Generally $x^{(1)}$ will not satisfy the other equations, so we may construct an $x^{(2)}$ from $x^{(1)}$, so that $(d_2, x^{(2)}) = s_2$. Then the equations of (5.3.1) except for the second one will not be satisfied etc. The general case is

$$(5.3.3) \quad x^{(k+1)} = x^{(k)} + d_{k+1} (s_{k+1} - (d_{k+1}, x^{(k)})) / \|d_{k+1}\|^2$$

where the numbering of the d and s is "periodically" i.e. if $k > M$ then

$d_k := d_{k \bmod M}$ etc.

This is the original ART algorithm of Gordon et.al. It has previously been discussed by Kaczmarz ([14]). Two variants exist, both of which are based on the physical interpretation of the $x_j^{(k)}$.

(i) Because $x_j^{(k)}$ should represent an absorption coefficient we take only nonnegative values for $x_j^{(k)}$, i.e. we replace $x_j^{(k)}$ by

$$(5.3.5) \quad \max(x_j^{(k)}, 0) .$$

(ii) Moreover, if we know that the absorption coefficient does not exceed 1, we may replace $x_j^{(k)}$ by

$$(5.3.6) \quad \min(1, \max(x_j^{(k)}, 0)) .$$

So there are two ways to iterate in (5.3.4)

algorithm 1 (ART)

$x^{(0)}$ arbitrary

$$(5.3.7) \quad \tilde{x}^{(k+1)} := x^{(k)} + d_{k+1} (s_{k+1} - (d_{k+1}, x^{(k)})) / \|d_{k+1}\|^2$$

and now (i)

$$(5.3.8) \quad x_j^{(k+1)} := \tilde{x}_j^{(k)}, \quad j = 1, \dots, N^2$$

"unconstrained ART", or (ii)

$$(5.3.9) \quad x_j^{(k+1)} = \max(0, \tilde{x}_j^{(k+1)}), \quad j = 1, \dots, N^2$$

"partially constrained ART", or (iii)

$$(5.3.10) \quad x_j^{(k+1)} = \min(1, \max(0, \tilde{x}_j^{(k+1)})), \quad j = 1, \dots, N^2 .$$

algorithm 2 (ART2)

$x^{(0)}$ arbitrary

$$(5.3.11) \quad \tilde{x}^{(k+1)} = \tilde{x}^{(k)} + d_{k+1} (s_{k+1} - (d_{k+1}, \tilde{x}^{(k)})) / \|d_{k+1}\|^2$$

and now again (i)

$$(5.3.12) \quad x_j^{(k+1)} := \max(0, \tilde{x}_j^{(k+1)}), \quad j = 1, \dots, N^2$$

"partially constructed ART2", (ii) or:

$$(5.3.13) \quad x_j^{(k+1)} := \min(1, \max(\tilde{x}_j^{(k+1)}, 0)), \quad j = 1, \dots, N^2 .$$

In the next section we give the convergence properties of algorithm 1. It is remarkable that even in case $Da = s$ has no solution(s) still the sequence $x^{(k)}$ of algorithm 1 does converge in a certain sense. Another alternative, especially suited to noisy data, is to start from a set of inequalities instead of the set of equations in (5.3.1) viz.

$$(5.3.14) \quad s_i - \epsilon_i^{(1)} \leq (d_i, x) \leq s_i + \epsilon_i^{(2)}, \quad 1 \leq i \leq M$$

where the ϵ_i are chosen positive. Because of these inequalities some desirable smooting will occur. Herman ([10]) gives a relaxation method to solve (5.3.14), which method is very much alike to ART. (ART3).

5.4. Limiting behaviour of ART algorithms

In this section we give the theorems of Herman, Lent and Rowland ([11]) with proofs in a slightly altered form.

We have the set of equations

$$(5.4.1) \quad (d_m, x) = s_m, \quad m = 1, 2, \dots, M .$$

Or in matrix-vector notation

$$(5.4.2) \quad Dx = s$$

where d_m^H is the m -th row of D . So

$$(5.4.3) \quad \mathcal{L}\{d_1, \dots, d_m\} = R(D^H) .$$

Let

$$(5.4.4) \quad L_1 := \{z \mid Dz = s\}$$

then the following theorem holds.

Theorem 1. If $L_1 \neq \emptyset$ then the sequence $x^{(k)}$ of algorithm 1(i) converges for every $x^{(0)}$. Moreover, the limit is the element of L_1 with the **smallest** distance to $x^{(0)}$.

Proof. The iteration formula is

$$(5.4.5) \quad x^{(k+1)} = x^{(k)} + d_{k+1} [s - (d_{k+1}, x^{(k)})] / \|d_{k+1}\|^2.$$

Let $z \in L_1$, and let $y^{(k)} := x^{(k)} - z$. Then

$$(5.4.6) \quad y^{(k+1)} = y^{(k)} - d_{k+1} (d_{k+1}, y^{(k)}) / \|d_{k+1}\|^2$$

or

$$(5.4.7) \quad y^{(k+1)} = (I - P_{k+1})y^{(k)}$$

where P_{k+1} is the orthogonal projector on $\mathcal{L}\{d_m\}$, where $m \equiv k+1 \pmod{M}$. Let D^+ be the pseudo inverse of D , defined by

- (i) $D^+ D D^+ = D^+$
- (ii) $D D^+ D = D$
- (iii) $D D^+$ and $D^+ D$ are orthogonal projectors.

Then $D^+ D$ is the orthogonal projector on $R(D^+) = R(D^H)$.

Now we write $y^{(k)}$ as

$$y^{(k)} = D^+ D y^{(k)} + (I - D^+ D)y^{(k)},$$

and consider the effect of operation (5.4.7) to each of these terms. Since P_{k+1} is an orthogonal projector on a subspace of $R(D^H)$ we have $P_{k+1}(I - D^+ D) = 0$, so we obtain

$$(I - P_{k+1})(I - D^+ D)y^{(k)} = (I - D^+ D)y^{(k)}.$$

Consequently, from (5.4.5) we obtain

$$(5.4.8) \quad (I - D^+ D)y^{(k+1)} = (I - D^+ D)y^{(k)}$$

and

$$(5.4.9) \quad D^+ D y^{(k+1)} = (I - P_{k+1})D^+ D y^{(k)}.$$

By induction we obtain

$$(5.4.10) \quad (I - D^+ D)y^{(k)} = (I - D^+ D)y^{(0)} \quad \text{for every } k,$$

$$(5.4.11) \quad D^+ D y^{(k+1)} = (I - P_{k+1}) \dots (I - P_1)D^+ D y^{(0)}.$$

We have for every v

$$(5.4.12) \quad \|(I - P_M) \dots (I - P_1)v\| \leq \|v\|,$$

where the equality sign holds if and only if

$$P_1 v = \dots = P_M v = 0.$$

If $v \in R(D^H)$ this implies $v = 0$. Then it follows that the restriction of $(I - P_M) \dots (I - P_1)$ to $R(D^H)$ has norm $c < 1$. From (5.4.11) it follows, since $D^+ D y^{(0)} \in R(D^H)$ that

$$D^+ D y^{(k+1)} \rightarrow 0 \quad \text{if } k \rightarrow \infty$$

and so

$$y^{(k)} \rightarrow (I - D^+ D)y^{(0)} \quad \text{if } k \rightarrow \infty.$$

For $x^{(k)}$ this means that

$$(5.4.13) \quad \begin{aligned} x^{(k)} &\rightarrow z + (I - D^+ D)(x^{(0)} - z) \\ &\rightarrow D^+ s + (I - D^+ D)x^{(0)} \quad \text{if } k \rightarrow \infty. \end{aligned}$$

This proves the first part of the theorem.

For $z \in L_1$ we may write

$$(5.4.14) \quad z = D^+ s + (I - D^+ D)z.$$

So for the distance of z to $x^{(0)}$ we have

$$\begin{aligned} \|D^+ s + (I - D^+ D)z - x^{(0)}\|^2 &= \\ \|D^+(s - Dx^{(0)}) + (I - D^+ D)(z - x^{(0)})\|^2 &= \\ \|D^+(s - Dx^{(0)})\|^2 + \|(I - D^+ D)(z - x^{(0)})\|^2. \end{aligned}$$

The last equality sign holds because the two vectors are mutually orthogonal. So the element of L_1 with smallest distance to $x^{(0)}$ must satisfy

$$\|(I - D^+ D)(z - x^{(0)})\|^2 = 0$$

i.e. (see (5.4.14))

$$z = D^+ s + (I - D^+ D)x^{(0)},$$

and this is exactly the limit of $x^{(k)}$. This proves the second part of the theorem.

We mention the analogues for algorithms 1(ii) and 1(iii). However the second part of theorem 1 is not valid in these cases. Let

$$L_2 := \{z \mid Dz = s, z_j \geq 0, 1 \leq j \leq N^2\}$$

$$L_3 := \{z \mid Dz = s, 0 \leq z_j \leq 1, 1 \leq j \leq N^2\}.$$

Theorem 2. If $L_2 \neq \emptyset$ resp. $L_3 \neq \emptyset$ then the sequence $\{x^{(k)}\}$ of algorithm 1(ii) resp. 1(iii) converges to an element of L_2 resp. L_3 . In general the limit is not the element of L_2 resp. L_3 with the smallest norm.

In case $L_1 = \emptyset$ the sequence $\{x^{(k)}\}$ of algorithm 1(i) still shows a limiting behaviour.

Definition. A sequence $\{y^{(k)}\}_{k \geq 1}$ converges cyclically with period M if for every $m, 1 \leq m \leq M$ the subsequences $\{y^{kM+i}\}_{k \geq 0}$ converge.

Theorem 3. If $L_1 = \emptyset$ the sequence $\{x^{(k)}\}$ of algorithm 1(i) converges cyclically.

Proof. The proof shows a great resemblance with the proof of theorem 1. Let

$$(5.4.15) \quad w^{k,m} := x^{kM+m} - x^{(k-1)M+m}.$$

Then from the iteration formula it follows that

$$w^{k,m} = w^{k,m-1} - d_m (d_m, w^{k,m-1}) / \|d_m\|^2.$$

So in the notation as in (5.4.6), (5.4.7), we write

$$(5.4.16) \quad w^{k,m} = (I - P_m) w^{k,m-1}.$$

By induction we obtain, since $w^{k,m-M} = w^{k-1,m}$,

$$(5.4.17) \quad w^{k,m} = A_m B_m w^{k-1,m}$$

where

$$(5.4.18) \quad \begin{aligned} A_m &= (I - P_m)(I - P_{m-1}) \dots (I - P_1) \\ B_m &= (I - P_M)(I - P_{M-1}) \dots (I - P_{m+1}). \end{aligned}$$

Note. If $m = M$ we must take $B_m = I$.

Denoting again by D the matrix with m -th row equal to d_m^H , we see that $w^{0,m} \in R(D^H)$. Then (5.4.17) implies that for every k we have

$$w^{k,m} \in R(D^H)$$

or, equivalently

$$D^+ D w^{k,m} = w^{k,m} .$$

Comparing with (5.4.12) sqq, we see that a positive constant $c < 1$ exists, such that

$$(5.4.19) \quad \| A B_m D^+ D \| = c .$$

From (5.4.17) we conclude that for every k :

$$(5.4.20) \quad \| w^{k,m} \| \leq c \| w^{k-1,m} \| .$$

From the identity

$$x^{kM+m} = \sum_{\ell=1}^k w^{\ell,m} + x^{0,m}$$

we conclude with (5.4.20) that $\lim_{k \rightarrow \infty} x^{kM+m}$ exists.

We want to remark that the theorem holds for any set of equations (5.4.2). We could have taken matrix E (see (5.2.7)) instead of the matrix D .

5.5. Convergence if the stepfunction is refined

We consider the question of convergence of the ART-algorithms if the stepfunction is refined, i.e. if the subsquares are taken smaller. We discuss the conclusion reached by Zwick and Zeitler ([26]) and by Gilbert ([7]).

Intuitively it seems to be clear that ART will provide better approximations to f if the number of basisfunctions increases. (If also the number of equations increases one should expect convergence to the real f .) However, according to Zwick and Zeitler this sequence of estimates converges to the solution of the problem given by the method of "back-projection". The point where they go wrong in their analysis is that in the continuous version of ART they use an unbounded support for f . At the end of this section we will give one correct version of continuous ART.

An example will show that the conclusion of Zwick and Zeitler is not correct viz. in the case that f itself is a stepfunction with a representation as in (5.2.1), let us say for $N = M$. Then the set of m equations in N^2 unknowns (the set of equations (5.2.4)) will have a solution if $N = M, 2M, 3M, \dots$. So if the solutions given by algorithm 1(i) converge as N increases the limit function will satisfy the m equations (5.2.4). It is easy to verify that the solution according to backprojection will not satisfy these equations. Consequently this estimate cannot be the limitfunction.

We proceed by giving a formally correct version of ART. Suppose that f is zero outside the unit disk. (Because f has a compact support this can always be achieved by means of dimensioning.) Then the Radon transform $g(p, \theta)$ of f is equal to zero if $|p| > 1$. Let $g(p, \theta)$ be given for a number of projections $\theta = \theta_m, m = 1, 2, \dots, M$. Then the equations are:

$$(5.5.1) \quad \int_{-\sqrt{1-p^2}}^{\sqrt{1-p^2}} f(p \cos \theta_m - t \sin \theta_m, p \sin \theta_m + t \cos \theta_m) dt = g(p, \theta_m),$$

for every $-1 \leq p \leq 1$ and $m = 1, 2, \dots, M$.

Let $f^{(0)}$ be a given estimate for f . Let $g^{(k)}$ denote the Radon transform of $f^{(k)}$. The iteration equations become:

$$(5.5.2) \quad f^{(k+1)}(p \cos \theta_m - t \sin \theta_m, p \sin \theta_m + t \cos \theta_m) = \\ f^{(k)}(p \cos \theta_m - t \sin \theta_m, p \sin \theta_m + t \cos \theta_m) + \\ \{g(p, \theta_m) - g^{(k)}(p, \theta_m)\} / (2\sqrt{1-p^2}),$$

where $m = (k + 1) \bmod M$.

We notice that after the k -th iteration

$$g^{(k)}(p, \theta_m) = g(p, \theta_m), \quad -1 \leq p \leq 1.$$

Part II

Chapter 6. The problem on the unit disk: solution of the continuous and discrete problem by means of series expansions

6.1. The Radon transformation on the unit disk

Let

$$(6.1.1) \quad \mathbb{D} := \{(x,y) \mid x^2 + y^2 \leq 1\}$$

$$(6.1.2) \quad \mathbb{C} := \{(p,\theta) \mid -1 \leq p \leq 1, 0 \leq \theta \leq 2\pi\} .$$

Let $L^2(\mathbb{D})$ be the space of square Lebesgue integrable functions on \mathbb{D} and $L^2(\mathbb{C},W)$ be the space of square Lebesgue integrable functions on \mathbb{C} with weight-function $W(p,\theta) = (1 - p^2)^{-\frac{1}{2}}$.

Inner products are in $L^2(\mathbb{D})$:

$$(6.1.3) \quad (f_1, f_2)_{\mathbb{D}} := \iint_{\mathbb{D}} f_1(x,y) \overline{f_2(x,y)} dx dy ,$$

In $L^2(\mathbb{C},W)$:

$$(6.1.4) \quad (g_1, g_2)_{\mathbb{C}} := \iint_{\mathbb{C}} g_1(p,\theta) \overline{g_2(p,\theta)} W(p,\theta) dp d\theta .$$

If $f \in L^2(\mathbb{D})$ its Radon transform is defined by

$$(6.1.5) \quad \hat{f}(p,\theta) := \int_{-\sqrt{1-p^2}}^{\sqrt{1-p^2}} f(p \cos \theta - t \sin \theta, p \sin \theta + t \cos \theta) dt .$$

This is a special case of the definition of the Radon transform of a function on \mathbb{R}^2 (see 2.1.2) if we extend $f \in L^2(\mathbb{D})$ to a function on \mathbb{R}^2 by

$$f(x,y) = \begin{cases} f(x,y) & \text{if } (x,y) \in \mathbb{D} \\ 0 & \text{if } (x,y) \notin \mathbb{D} . \end{cases}$$

The following holds: If $f \in L^2(\mathbb{D})$ then $\hat{f} \in L^2(\mathbb{C},W)$. This may be concluded from the inequality (6.1.7) we derive below.

Let $f \in L^2(\mathbb{D})$ then

$$\|f\|_{\mathbb{D}}^2 := \iint_{\mathbb{D}} |f(x,y)|^2 dx dy = \int_{-1}^{+1} dx \int_{-(1-x^2)^{\frac{1}{2}}}^{(1-x^2)^{\frac{1}{2}}} |f(x,y)|^2 dy .$$

After a rotation of the coordinate system this becomes

$$\|f\|_{\mathbb{D}}^2 = \int_{-1}^{+1} dp \int_{-(1-p^2)^{\frac{1}{2}}}^{(1-p^2)^{\frac{1}{2}}} |f(p \cos \theta - t \sin \theta, p \sin \theta + t \cos \theta)|^2 dt.$$

With the Cauchy-Schwarz inequality

$$\left| \int_a^b g(t) dt \right|^2 \leq \int_a^b |g(t)|^2 dt \int_a^b dt$$

we get

$$2\|f\|_{\mathbb{D}}^2 \geq \int_{-1}^{+1} \left| \int_{-(1-p^2)^{\frac{1}{2}}}^{(1-p^2)^{\frac{1}{2}}} f(p \cos \theta - t \sin \theta, p \sin \theta + t \cos \theta) dt \right|^2 (1-p^2)^{-\frac{1}{2}} dp$$

i.e.

$$2\|f\|_{\mathbb{D}}^2 \geq \int_{-1}^{+1} |\hat{f}(p, \theta)|^2 (1-p^2)^{-\frac{1}{2}} dp.$$

After integration of both sides over $\theta \in [0, 2\pi)$ we obtain

$$4\pi\|f\|_{\mathbb{D}}^2 \geq \int_0^{2\pi} d\theta \int_{-1}^{+1} |\hat{f}(p, \theta)|^2 (1-p^2)^{-\frac{1}{2}} dp$$

or

$$(6.1.7) \quad \|\hat{f}\|_{\mathbb{C}}^2 \leq 4\pi\|f\|_{\mathbb{D}}^2.$$

From this it also follows that Radon transforming is a bounded mapping of $L^2(\mathbb{D})$ into $L^2(\mathbb{C}, W)$ with norm $2\pi^{\frac{1}{2}}$, since (6.1.7) is an equality if $f = 1$.

6.2. The Radon transform of a polynomial on \mathbb{D}

In the sections 6.2 - 6.5 we describe somewhat differently a part of the theory of Marr ([17]).

Let $\mathcal{P}(\mathbb{D})$ be the space consisting of all polynomials on \mathbb{D} , and let $\mathcal{P}_M(\mathbb{D})$, $M = 0, 1, \dots$ be the linear subspaces of $\mathcal{P}(\mathbb{D})$ that satisfy the conditions

- (i) $\mathbb{P}_M(\mathbb{D})$ consists of polynomials of exact degree M and the zero polynomial
- (ii) $\mathbb{P}_M(\mathbb{D})$ is orthogonal to $\mathbb{P}_L(\mathbb{D})$ for every $L \leq M - 1$
- (iii) $\mathbb{P}(\mathbb{D})$ is the direct sum of the $\mathbb{P}_M(\mathbb{D})$, $M = 0, 1, \dots$

From this it follows that for every $M \neq L$ and every $P_1 \in \mathbb{P}_M$, $P_2 \in \mathbb{P}_L$ we have

$$(6.2.1) \quad (P_1, P_2)_{\mathbb{D}} = 0 .$$

Since $\mathbb{P}(\mathbb{D})$ is dense in $L^2(\mathbb{D})$, for every $f \in L^2(\mathbb{D})$ polynomials $P_M \in \mathbb{P}_M(\mathbb{D})$ exist such that

$$(6.2.2) \quad f(x, y) = \sum_{M=0}^{\infty} P_M(x, y)$$

and the polynomials are uniquely determined by f . From (6.2.1) it follows that every $P \in \mathbb{P}_M(\mathbb{D})$ is orthogonal to every polynomial of degree $\leq M - 1$. A special case is

$$(6.2.3) \quad \iint_{\mathbb{D}} (x \cos \theta + y \sin \theta)^m P(x, y) dx dy = 0 ,$$

if $m \leq M - 1$. By rotating \mathbb{D} over angle θ we obtain

$$\iint_{\mathbb{D}} p^m P(p \cos \theta - t \sin \theta, p \sin \theta + t \cos \theta) dp dt = 0$$

i.e.

$$\int_{-1}^{+1} p^m dp \int_{-(1-p^2)^{\frac{1}{2}}}^{(1-p^2)^{\frac{1}{2}}} P(p \cos \theta - t \sin \theta, p \sin \theta + t \cos \theta) dt = 0$$

and so

$$(6.2.4) \quad \int_{-1}^{+1} p^m \hat{P}(p, \theta) dp = 0 \text{ if } m \leq M - 1, P \in \mathbb{P}_M(\mathbb{D})$$

with $\hat{P}(p, \theta)$ the Radon transform of P . Then

$$(6.2.5) \quad \hat{P}(p, \theta) = (1-p^2)^{\frac{1}{2}} \int_{-1}^{+1} P(p \cos \theta - t(1-p^2)^{\frac{1}{2}} \cos \theta, p \sin \theta + t(1-p^2)^{\frac{1}{2}} \cos \theta) dt .$$

$P(x,y)$ is a polynomial in x and y with degree M , so the integrand in (6.2.5) is a polynomial in p and $(1 - p^2)^{\frac{1}{2}}t$. Since odd powers of t do not contribute to the integral, the integral is a polynomial in p of degree $\leq M$. Therefore $(1 - p^2)^{-\frac{1}{2}}\hat{P}(p,\theta)$ is a polynomial in p of degree $\leq M$. From (6.2.4) it follows that

$$\int_{-1}^{+1} p^m \{(1 - p^2)^{-\frac{1}{2}}\hat{P}(p,\theta)\} (1 - p^2)^{\frac{1}{2}} dp = 0$$

if $0 \leq m \leq M - 1$.

The only polynomial of degree $\leq M$ satisfying this relation is constructed by orthogonalization of the polynomials $1, p, p^2, \dots, p^M$ with respect to weight-function $(1 - p^2)^{\frac{1}{2}}$. (This polynomial is determined but for a factor.) The resulting polynomial is well known viz. $U_M(p)$, the Chebyshev polynomial of the second kind, defined by (see [24]):

$$(6.2.6) \quad U_M(\cos \psi) = \sin[(M + 1)\psi] / \sin \psi.$$

So we have for any $P \in \mathbb{P}_M(\mathbb{D})$

$$(6.2.7) \quad \hat{P}(p,\theta) = c(\theta) (1 - p^2)^{\frac{1}{2}} U_M(p)$$

where $c(\theta)$ is a function of θ only.

From (6.2.5) we get

$$\lim_{p \rightarrow 1} (1 - p^2)^{-\frac{1}{2}} \hat{P}(p,\theta) = \int_{-1}^{+1} P(\cos \theta, \sin \theta) dt = 2P(\cos \theta, \sin \theta),$$

so finally we obtain

$$(6.2.8) \quad \hat{P}(p,\theta) = 2(M + 1)^{-1} (1 - p^2)^{\frac{1}{2}} U_M(p) P(\cos \theta, \sin \theta).$$

The Radon transforms of polynomials $P_M \in \mathbb{P}_M(\mathbb{D})$, $P_L \in \mathbb{P}_L(\mathbb{D})$ with different degrees (i.e. $M \neq L$) are orthogonal too:

$$(6.2.9) \quad (\hat{P}_M, \hat{P}_L)_{\mathbb{C}} = 4(M + 1)^{-1} (L + 1)^{-1} \int_0^{2\pi} P_M(\cos \theta, \sin \theta) \overline{P_L(\cos \theta, \sin \theta)} d\theta * \\ \int_{-1}^{+1} (1 - p^2)^{\frac{1}{2}} U_M(p) U_L(p) dp.$$

So if $M \neq L$ we obtain

$$(\hat{P}_M, \hat{P}_L)_{\mathbb{C}} = 0 .$$

Let $\hat{P}_M(\mathbb{C})$ be the space consisting of all Radon transforms of polynomials of $P_M(\mathbb{D})$ and $\hat{P}(\mathbb{C})$ be the space consisting of the Radon transforms of all polynomials on \mathbb{D} . (So $\hat{P}(\mathbb{C})$ is the transformed space of $P(\mathbb{D})$.)

Summary.

(i) Let $P \in P_M(\mathbb{D})$. The Radon transform of P is given by

$$(6.2.10) \quad \hat{P}(p, \theta) = 2(M+1)^{-1} \sqrt{1-p^2} U_M(p) P(\cos \theta, \sin \theta) .$$

(ii) $\hat{P}_M(\mathbb{C})$, $M \geq 0$ are mutually orthogonal subspaces of $\hat{P}(\mathbb{C})$ (which is it-

(6.2.11) self a linear subspace of $L^2(\mathbb{C}, W)$).

6.3. Orthogonal bases of $P(\mathbb{D})$ and $\hat{P}(\mathbb{C})$

In searching for orthogonal bases for $P(\mathbb{D})$ and $\hat{P}(\mathbb{C})$ we may restrict ourselves to the bases for the finite dimensional spaces $P_M(\mathbb{D})$ and $\hat{P}_M(\mathbb{C})$, according to (6.2.11). More specifically we would like to have an orthogonal system in $P_M(\mathbb{D})$, the Radon transforms of which form an orthogonal system in $\hat{P}_M(\mathbb{C})$.

From (6.2.9) we see in what direction we may search. If $P_1, P_2 \in P_M(\mathbb{D})$ then we have

$$(6.3.1) \quad (\hat{P}_1, \hat{P}_2)_{\mathbb{C}} = 2\pi(M+1)^{-2} \int_0^{2\pi} P_1(\cos \theta, \sin \theta) \overline{P_2(\cos \theta, \sin \theta)} d\theta .$$

So it seems to be reasonable to decompose the polynomials of $P_M(\mathbb{D})$ into their Fourier series.

Let $P \in P_M$. Then for fixed r , $P(r \cos \psi, r \sin \psi)$ is a trigonometric polynomial with degree $\leq M$ and, if ψ is fixed, a polynomial in r with exact degree M . So

$$(6.3.2) \quad P(r \cos \psi, r \sin \psi) = \sum_{n=-M}^{+M} c_n(r) e^{in\psi}$$

where

$$(6.3.3) \quad c_n(r) = (2\pi)^{-1} \int_0^{2\pi} P(r \cos \psi, r \sin \psi) e^{-in\psi} d\psi$$

and then $c_n(r)$ is a polynomial in r with degree $\leq M$. But we know more about the degree of $c_n(r)$. A term $x^k y^\ell$ of $P(x,y)$ has the contribution to $c_n(r)$:

$$(6.3.4) \quad r^{k+\ell} \int_0^{2\pi} \cos^k \psi \sin^\ell \psi e^{-in\psi} d\psi .$$

Since $\cos^k \psi \sin^\ell \psi$ is a trigonometric polynomial of degree $k + \ell$, the integral is 0 if $k + \ell < |n|$. By writing, instead of (6.3.4)

$$r^{k+\ell} \int_0^\pi \cos^k \psi \sin^\ell \psi e^{-in\psi} d\psi \{1 + (-1)^{k+\ell-|n|}\}$$

we see that the integral in (6.3.4) is zero if $k + \ell - |n|$ is odd. From this observation it follows that

$$(6.3.5) \quad c_n(r) = r^{|n|} d_n(r^2)$$

where $d_n(t)$ is a polynomial of degree $\leq (M - |n|)/2$.

Taking into account (6.3.1) and (6.3.5) we consider polynomials of degree M of the form

$$(6.3.6) \quad P(r \cos \psi, r \sin \psi) = r^{|n|} e^{in\psi} d_n(r^2)$$

where $M - |n|$ is even and the degree of d_n equals $(M - |n|)/2$. If $P \in \mathbb{P}_M(\mathbb{D})$ we must have

$$\iint_{\mathbb{D}} x^k y^\ell P(x,y) dx dy = 0 \quad \text{if } k + \ell \leq M - 1 .$$

In polar coordinates

$$(6.3.7) \quad \int_0^{2\pi} \cos^k \psi \sin^\ell \psi e^{in\psi} d\psi \int_0^1 r^{k+\ell+|n|+1} d_n(r^2) dr = 0$$

if $k, \ell \geq 0, k + \ell \leq M - 1$.

Since the first integral is 0 if $k + \ell < |n|$ or if $k + \ell - |n|$ is odd,

(6.3.7) is certainly valid if

$$\int_0^1 r^{2|n|+2m+1} d_n(r^2) dr = 0 \text{ if } m \geq 0, |n| + 2m \leq M - 1 .$$

Or, equivalently,

$$(6.3.8) \quad \int_0^1 r^{|n|+m} d_n(r) dr = 0 \text{ if } 0 \leq m < (M - |n|)/2 .$$

We remark that the degree of $d_n(r)$ is exactly $(M - |n|)/2$. Polynomials satisfying (6.3.8) are the result of orthogonalization of the polynomials $1, r, r^2, \dots, r^M$ with respect to weightfunction $r^{|n|}$. These polynomials are well known viz. multiples of $P_k^{(0, |n|)}(2r - 1)$, polynomials of Jacobi (see [24]), $k = (M - |n|)/2$.

Also we see that exactly $M + 1$ polynomials of the form (6.3.6) exist, that are element of $\mathbb{P}_M(\mathbb{D})$.

Since the dimension of $\mathbb{P}_M(\mathbb{D})$ is $M + 1$, we have an orthogonal basis for $\mathbb{P}_M(\mathbb{D})$ viz.

$$e^{in\psi} r^{|n|} P_k^{(0, |n|)}(2r^2 - 1), |n| + 2k = M .$$

Their Radon transforms are (see (6.1.7))

$$2(M + 1)^{-1} (1 - p^2)^{\frac{1}{2}} U_M(p) e^{in\theta}$$

and they form an orthogonal basis for $\widehat{\mathbb{P}}_M(\mathbb{C})$.

Summary

Let $F_{n,k}(x,y)$ be defined by

$$(6.3.9) \quad F_{n,k}(r \cos \psi, r \sin \psi) = e^{in\psi} r^{|n|} P_k^{(0, |n|)}(2r^2 - 1)$$

where $n = 0, \pm 1, \pm 2, \dots, k = 0, 1, \dots$

Then $F_{n,k}(x,y)$ is a polynomial of degree $|n| + 2k$. The system of polynomials $F_{n,k}(x,y)$ form an orthogonal basis in $L^2(\mathbb{D})$. Their Radon transforms are given by

$$(6.3.10) \quad \widehat{F}_{n,k}(p, \theta) = 2(|n| + 2k + 1)^{-1} (1 - p^2)^{\frac{1}{2}} U_{|n|+2k}(p) e^{in\theta}$$

and the system $\widehat{F}_{n,k}$ form an orthogonal basis for $\widehat{\mathbb{P}}(\mathbb{C})$. (Subspace of $L^2(\mathbb{C}, W)$.)

We give the norms of $F_{n,k}$ and $\widehat{F}_{n,k}$ (see [24])

$$(6.3.11) \quad (F_{n,k}, F_{m,\ell})_{\mathbb{D}} = \pi(|n| + 2k + 1)^{-1} \delta_{n,m} \delta_{k,\ell}$$

$$(6.3.12) \quad (\widehat{F}_{n,k}, \widehat{F}_{m,\ell})_{\mathbb{C}} = 4\pi^2(|n| + 2k + 1)^{-2} \delta_{n,m} \delta_{k,\ell} .$$

Note. From these relations it follows that the inverse Radon transformation is unbounded.

The polynomials we developed in this section have previously been developed by Zernike. Theory concerning these polynomials may be found also in [1].

6.4. Radon inversion by means of polynomial expansions

Now we are able to solve the inversion problem. Let $f \in L^2(\mathbb{D})$ and let $g \in L^2(\mathbb{C}, W)$ be its Radon transform.

So, numbers $\gamma_{n,k}$ exist, such that

$$(6.4.1) \quad f(x,y) = \sum_{n=-\infty}^{+\infty} \sum_{k=0}^{\infty} \gamma_{n,k} F_{n,k}(x,y)$$

in $L^2(\mathbb{D})$ sense (i.e. $\sum_{n,k} |\gamma_{n,k}|^2 (|n| + 2k + 1)^{-1} < \infty$).
From (6.4.1) and (6.1.6) we have

$$(6.4.2) \quad g(p, \theta) = \sum_{n=-\infty}^{+\infty} \sum_{k=0}^{\infty} \gamma_{n,k} \widehat{F}_{n,k}(p, \theta) .$$

So we must have

$$(6.4.3) \quad \gamma_{n,k} = \frac{(g, \widehat{F}_{n,k})_{\mathbb{C}}}{(\widehat{F}_{n,k}, \widehat{F}_{n,k})_{\mathbb{C}}}$$

or

$$(6.4.4) \quad \gamma_{n,k} = [(|n| + 2k + 1) / 2\pi]^2 \int_0^{2\pi} \int_{-1}^{+1} g(p, \theta) \overline{\widehat{F}_{n,k}(p, \theta)} (1 - p^2)^{-\frac{1}{2}} dp .$$

So (6.4.1) and (6.4.4) yield the solutions of the inversion problem.

We want to discuss the conditions a function $g \in L^2(\mathbb{C}, W)$ must satisfy in order that a $f \in L^2(\mathbb{D})$ exists such that g is the Radon transform of f . Or, stated differently, we want to characterize the subspace of $L^2(\mathbb{C}, W)$, consisting of Radon transforms of function in $L^2(\mathbb{D})$.

It is clear that a necessary and sufficient condition is that g has an expansion as in (6.4.2). From this it follows that necessary conditions are

$$(i) \quad g(p, \theta) = g(-p, \theta + \pi)$$

$$(ii) \quad \iint_{\mathbb{C}} e^{in\theta} p^\ell g(p, \theta) dp d\theta = 0 \text{ if } 0 \leq \ell < n .$$

$$(iii) \quad \sum_{n,k} |\gamma_{n,k}|^2 (|n| + 2k + 1)^{-1} < \infty .$$

These conditions are also sufficient as may be seen from the following.

Let $p = \cos \varphi$, $0 \leq \varphi \leq \pi$, then the inner product in $L^2(\mathbb{C}, W)$ can be written as

$$(g_1, g_2)_C = \int_0^{2\pi} d\theta \int_0^\pi g_1(\cos \varphi, \theta) \overline{g_2(\cos \varphi, \theta)} d\varphi .$$

Then we see that, with $\cos \varphi = p$,

$$e^{in\theta} e^{i(|n|+2k+1)\varphi}, \text{ } n \text{ and } k \text{ integer},$$

is an orthogonal basis in $L^2(\mathbb{C}, W)$.

This may be modified to

$$e^{in\theta} \cos[(|n| + 2k + 1)\varphi]$$

$$e^{in\theta} \sin[(|n| + 2k + 1)\varphi], \text{ } n \text{ integer, } k \geq -[(|n| + 1)/2] .$$

Note. If $x \in \mathbb{R}$ then $[x]$ denotes the largest integer that is smaller than or equal to x .

So, a basis for $L^2(\mathbb{C}, W)$ is

$$T_{|n|+2k+1}(p) e^{in\theta}$$

$$(1 - p^2)^{\frac{1}{2}} U_{|n|+2k}(p) e^{in\theta}, \text{ } n \text{ integer, } k \geq -[(|n| + 1)/2] ,$$

where $T_\ell(p)$ is the Chebyshev polynomial of the first kind of degree ℓ (see [24]).

Because $g \in L^2(\mathbb{C}, W)$ numbers $a_{n,k}$ and $b_{n,k}$ exist such that

$$(6.4.5) \quad g(p, \theta) = \sum_{n=-\infty}^{+\infty} \sum_{k=-[(|n|+1)/2]}^{\infty} e^{in\theta} \{ a_{n,k} T_{|n|+2k+1}(p) + b_{n,k} (1-p^2)^{\frac{1}{2}} U_{|n|+2k}(p) \} .$$

Since $T_{|n|+2k+1}(-p) = (-1)^{n+1} T_{|n|+2k+1}(p)$, we obtain, applying the symmetry relation

$$g(p, \theta) = g(-p, \theta + \pi) ,$$

that $a_{n,k} = 0$ for every n, k .

Then we have

$$g(p, \theta) = \sum_{n=-\infty}^{+\infty} \sum_{k=-[(|n|+1)/2]}^{\infty} b_{n,k} (1-p^2)^{\frac{1}{2}} U_{|n|+2k}(p) e^{in\theta} .$$

From condition (ii) it follows that

$$\sum_{k=-[(|n|+1)/2]}^{\infty} b_{n,k} \int_{-1}^{+1} p^{\ell} (1-p^2)^{\frac{1}{2}} U_{|n|+2k}(p) dp = 0 ,$$

if $0 \leq \ell < |n|$. The terms with $k \geq 0$ in this series are 0 so we have

$$\sum_{k=-[(|n|+1)/2]}^{-1} b_{n,k} \int_{-1}^{+1} p^{\ell} U_{|n|+2k}(p) (1-p^2)^{\frac{1}{2}} dp = 0 , \quad 0 \leq \ell < |n| .$$

By combining terms p^{ℓ} in the right way, we obtain

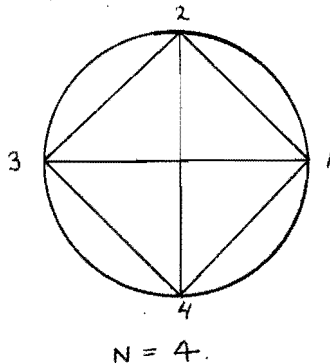
$$\sum_{k=-[(|n|+1)/2]}^{-1} b_{n,k} \int_{-1}^{+1} U_{\ell}(p) U_{|n|+2k}(p) (1-p^2)^{\frac{1}{2}} dp = 0 , \quad 0 \leq \ell < |n| ,$$

from which it follows that $b_{n,k} = 0$, $-[(|n| + 1)/2] \leq k \leq -1$ and thus g has an expansion as in (6.4.2). Condition (iii) implies that the corresponding f is element of $L^2(\mathbb{D})$.

6.5. Discrete problem

Up to now we only developed the analogue of section 4.4 for the unit disk. Solving the discrete problem could be done as described in section 4.4, but Marr [17] has shown that we can do better if we choose the sample points of $g(p, \theta)$ correctly.

Let N be a natural number.



Take N points on the circumference of \mathbb{D} , so that they form a regular polygon. Number the points from 1 to N . We denote the points by (x_I, y_I) , $1 \leq I \leq N$ so that

$$(6.5.1) \quad \begin{cases} x_{I+1} = \cos(2\pi I/N) \\ y_{I+1} = \sin(2\pi I/N) . \end{cases}$$

$N(N-1)/2$ different chords may be drawn between the points (x_I, y_I) . These chords may be represented by $N(N-1)$ parameter combinations $(p, \theta) \in \mathbb{C}$ (if (p, θ) is a representation, then $(-p, \theta + \pi)$ is one too). Let

$$(6.5.2) \quad \begin{aligned} p_J &= \cos(\pi J/N), \quad 1 \leq J \leq N-1 \\ \theta_{IJ} &= (2I + J - 2)\pi/N, \quad 1 \leq I \leq N . \end{aligned}$$

Then (p_J, θ_{IJ}) represents the chord between the points (x_I, y_I) and (x_{I+J}, y_{I+J}) (if $I + J > N$ then $I + J$ should be replaced by $I + J - N$). If g is a function defined on \mathbb{C} we denote

$$(6.5.3) \quad g(p_J, \theta_{IJ}) \quad \text{by} \quad g[I, J] .$$

The symmetry relation $g(p, \theta) = g(-p, \theta + \pi)$ becomes

$$(6.5.4) \quad g[I, J] = g[I + J, N - J] .$$

Because p_J are zeroes of $U_{N-1}(p)$ we obtain

$$(6.5.5) \quad \sum_{J=1}^{N-1} (1 - p_J^2) U_L(p_J) U_M(p_J) = N/2 \delta_{L,M}$$

and, consequently,

$$(6.5.6) \quad \sum_{J=1}^{N-1} (-1)^J (1 - p_J^2) U_L(p_J) U_M(p_J) = -\frac{1}{2} N \delta_{N-L-2, M},$$

if $M \leq N - 2$, $L \leq N - 2$.

Combining this with the identity

$$(6.5.7) \quad \sum_{I=1}^N e^{i(n-m)\theta_{IJ}} = \begin{cases} 0 & \text{if } n \not\equiv m \pmod{N} \\ N & \text{if } n \equiv m \pmod{2N} \\ (-1)^{J_N} & \text{if } n \equiv m + N \pmod{2N} \end{cases}$$

we obtain that the $\widehat{F}_{n,k}$ are orthogonal on the sample points

$$(6.5.8) \quad \sum_{I=1}^N \sum_{J=1}^{N-1} \widehat{F}_{n,k}^{[I,J]} \overline{\widehat{F}_{m,\ell}^{[I,J]}} = 2N^2 (|n| + 2k + 1)^{-2} \delta_{n,m} \delta_{k,\ell}$$

if $|n| + 2k \leq N - 2$ and $|m| + 2\ell \leq N - 2$.

Now we formulate the following theorem.

Theorem. Given the array of numbers

$$g[I, J], \quad 1 \leq I \leq N, \quad 1 \leq J \leq N - 1$$

satisfying the condition

$$(6.5.9) \quad g[I, J] = g[I + J, N - J] \text{ if } I + J \leq N.$$

Let

$$(6.5.10) \quad \gamma_{n,k} := (|n| + 2k + 1)^2 / (2N^2) \sum_{I,J} g[I, J] \overline{\widehat{F}_{n,k}^{[I,J]}} \quad |n| + 2k \leq N - 2.$$

Let $M \leq N - 2$,

$$(6.5.11) \quad P^{(M)}(x, y) = \sum_{n=-M}^{+M} \sum_{k=0}^{[(M-|n|)/2]} \gamma_{n,k} \widehat{F}_{n,k}(x, y).$$

Then $P^{(M)}$ is the polynomial that minimizes the functional

$$L(P) := \sum_{I=1}^N \sum_{J=1}^{N-1} |g[I, J] - \widehat{P}[I, J]|^2$$

considered on the space of all polynomials P of degree $\leq M$. Moreover, $L(P^{(N-2)}) = 0$, i.e. $P^{(N-2)}$ interpolates g in the given points.

Proof. The fact that $P^{(M)}$ minimizes $L(P)$ follows from (6.5.8). $L(P^{(N-2)}) = 0$ follows from the fact that functions g satisfying the symmetry condition (6.5.9) span a linear space S with dimension $N(N-1)/2$ and exactly $N(N-1)/2$ functions $\hat{F}_{n,k}$ exist that satisfy the same symmetry relation and are mutually orthogonal on the discrete points. So they form a basis for S .

By the special choice of (p_J, θ_{IJ}) the interpretation of $P^{(N-2)}$ is twofold. Primo, the coefficients of $P^{(N-2)}$ may be considered to be the solution of the set of equations

$$\hat{P}^{(N-2)}[I, J] = g[I, J], \quad 1 \leq I \leq N, \quad 1 \leq J \leq N-1,$$

(see section 4.1), and $\hat{P}^{(M)}$ is the least squares solutions of

$$\hat{P}^{(M)}[I, J] = g[I, J].$$

Secundo, the right hand side of (6.5.10) may be interpreted as discretizations of the integrals in (6.4.4).

Finally we conclude that $P^{(N-2)}$ is the solution of the discrete problem (however, see section 6.7).

We mention the version of $P^{(M)}$ if all data $g[I, J]$ are real.

Let

$$\gamma_{\pm n, k} = \frac{1}{2} \{ \beta_{n, k} \mp i \alpha_{n, k} \}, \quad n \geq 1$$

$$\gamma_{0, k} = \beta_{0, k}.$$

Then (6.5.10) becomes

$$(6.5.12) \quad \begin{cases} \alpha_{n, k} \\ \beta_{n, k} \end{cases} = 4(n+2k+1)N^{-2} \sum_{J=1}^{N-1} \sin[(n+2k+1)\pi J/N] * \\ \sum_{I=1}^{N-J} \begin{cases} \sin \\ \cos \end{cases} [(2I+J-2)n\pi/N] g[I, J]$$

$n \geq 0, k \geq 0, n+2k \leq N-2.$

And $P^{(M)}$ obtains the form:

$$(6.5.13) \quad P^{(M)}(r \cos \psi, r \sin \psi) = \sum_{n=0}^M \sum_{k=0}^{[(M-n)/2]} \{ \alpha_{n, k} \sin n\psi + \\ + \beta_{n, k} \cos n\psi \} r^{|n|} P_k^{(0, |n|)}(2r^2 - 1).$$

Note. The $\alpha_{0,k}$ need not be calculated, since they do not really appear in (6.5.13).

The calculation of $\alpha_{n,k}$ and $\beta_{n,k}$ is just finite Fourier summation. The calculation of $P^{(M)}$ on a set of points that cover \mathbb{D} more or less uniformly is treated in chapter 7.

6.6. Stability to noisy data

We consider the effect of noise in $g[I,J]$ on the solution $P^{(M)}$. Because of the linearity we may restrict ourselves to reconstruction of a sample consisting of "pure" noise.

Let $\underline{g}[I,J]$, $1 \leq I \leq N$, $1 \leq J \leq N - 1$ be stochastic variables that satisfy the conditions

- (i) $\underline{g}[I + J, N - J] = \underline{g}[I,J]$ if $I + J \leq N$
- (ii) $E\{\underline{g}[I,J]\} = 0$, all I,J
- (iii) $E\{\underline{g}[I,J]\overline{\underline{g}[K,L]}\} = \delta_{IK}\delta_{JL}$, $1 \leq J \leq N - 1$, $1 \leq I \leq N - J$, $1 \leq K \leq N - 1$, $1 \leq L \leq N - K$

So the $\underline{g}[I,J]$ are mutually orthogonal stochastic variables (taking (i) into account). The coefficients $\gamma_{n,k}$ are stochastic variables too.

$$(6.6.1) \quad \gamma_{n,k} = [(|n| + 2k + 1)/N]^2 / 2 \sum_{J=1}^{N-1} \sum_{I=1}^N \underline{g}[I,J] \overline{\hat{F}_{n,k}[I,J]} .$$

Therefore, making use of symmetry condition (i) we obtain

$$(6.6.2) \quad \gamma_{n,k} = [(|n| + 2k + 1)/N]^2 \sum_{J=1}^{N-1} \sum_{I=1}^{N-J} \underline{g}[I,J] \overline{\hat{F}_{n,k}[I,J]} .$$

From this it follows that

$$(6.6.3) \quad E\{\gamma_{n,k}\} = 0$$

and because of condition (iii):

$$E\{\gamma_{n,k} \overline{\gamma_{m,\ell}}\} = [(|n| + 2k + 1) (|m| + 2\ell + 1) / N^2]^2 \sum_{J=1}^N \sum_{I=1}^{N-J} \hat{F}_{m,\ell}[I,J] \overline{\hat{F}_{n,k}[I,J]} .$$

Or, equivalently,

$$E\{\underline{Y}_{n,k} \overline{Y_{m,\ell}}\} = [(|n| + 2k + 1) (|m| + 2\ell + 1) / N^2]^2 / 2 \cdot \\ \sum_{J=1}^{N-1} \sum_{I=1}^N \widehat{F}_{m,\ell}^{[I,J]} \overline{\widehat{F}_{n,k}^{[I,J]}} .$$

With (6.5.8) we finally obtain

$$(6.6.4) \quad E\{\underline{Y}_{n,k} \overline{Y_{m,\ell}}\} = [(|n| + 2k + 1) / N]^2 \delta_{nm} \delta_{k\ell} ,$$

if $|n| + 2k \leq N - 2$ and $|m| + 2\ell \leq N - 2$.

So the $\underline{Y}_{n,k}$, $|n| + 2k \leq N - 2$ are statistically orthogonal. For

$$\underline{P}^{(M)}(x,y) = \sum_{|n|+2k \leq M} \underline{Y}_{n,k} F_{n,k}(x,y)$$

we obtain that

$$(6.6.5) \quad E\{\underline{P}^{(M)}(x,y)\} = 0$$

and

$$(6.6.6) \quad E\{|\underline{P}^{(M)}(x,y)|^2\} = \sum_{|n|+2k \leq M} \sum_{|m|+2\ell \leq M} E\{\underline{Y}_{n,k} \overline{Y_{m,\ell}}\} F_{n,k}(x,y) \overline{F_{m,\ell}(x,y)} .$$

With (6.6.4) we see that

$$(6.6.7) \quad E\{|\underline{P}^{(M)}(x,y)|^2\} = N^{-2} \sum_{|n|+2k \leq M} (|n| + 2k + 1)^2 |F_{n,k}(x,y)|^2 .$$

The following inequality is known

$$(6.6.8) \quad |F_{n,k}(r \cos \psi, r \sin \psi)| = |e^{in\psi} r^{|n|} P_k^{(0,|n|)}(2r^2 - 1)| \leq 1 \text{ if } r \leq 1 .$$

(See [24], theorem 7.2.)

The equality sign holds only if $r = 1$. Then (6.6.7) yields

$$(6.6.9) \quad E\{|\underline{P}^{(M)}(x,y)|^2\} \leq M^4 / (2N)^2 (1 + O(M^{-1})), \quad M \leq N - 2, \quad N \rightarrow \infty, \quad M \rightarrow \infty .$$

However this estimate is too crude. This may be seen after integration of (6.6.7) over \mathbb{D} (apply (6.3.11))

$$(6.6.10) \quad \iint_{\mathbb{D}} E\{|P^{(M)}(x,y)|^2\} dx dy = \pi/N^2 \sum_{|n|+2k \leq M} (|n| + 2k + 1) = \\ = \pi M^3 / (3N^2) (1 + O(M^{-1})), \quad M \leq N - 2, \\ N \rightarrow \infty, M \rightarrow \infty.$$

So (6.6.9) is a good estimate on a set with measure at most $O(M^{-1})$.

Note. We have calculated numerically the right hand side of (6.6.6) for $M = 10, 20, 40, 80$. In these cases the right hand side was bounded by M^3/N^2 on the region $\sqrt{x^2 + y^2} \leq 0.96$, so in these cases (6.6.10) gives a fairly good estimate in this region.

From (6.6.10) we see that the effect of noisy data may be reduced by decreasing the degree of the polynomial.

We have restricted ourselves to noisy data that have variances independent of $[I, J]$. If the variances do depend on the $[I, J]$ the coefficients $\gamma_{n,k}$ will not be mutually orthogonal anymore, so that it becomes difficult to estimate $E\{|P^{(M)}(x,y)|^2\}$. However (6.6.6) remains valid, so integration over \mathbb{D} yields

$$\iint_{\mathbb{D}} E\{|P^{(M)}(x,y)|^2\} dx dy = \sum_{|n|+2k \leq M} E\{|\gamma_{n,k}|^2\} (|n| + 2k + 1)^2 \pi / N^2.$$

We may use the right hand side as estimate to $E\{|P^{(M)}(x,y)|^2\}$ for lack of something better. However a rigorous upper bound may be developed. If m is the maximum of the variances, we put all the variances equal to m . The variance of the resulting polynomial will be larger than the previous one. Since now all variances are equal we may apply (6.6.9).

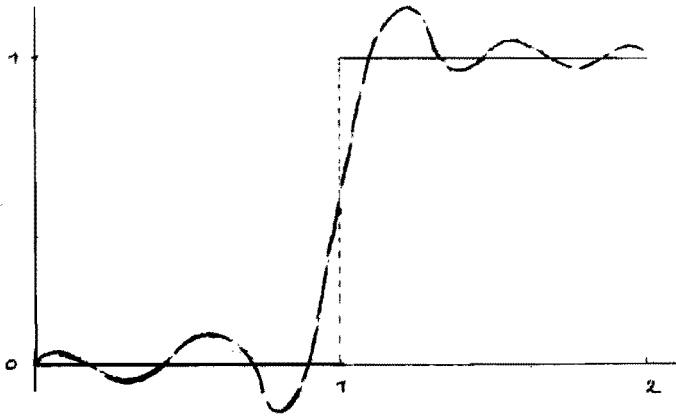
6.7. Smoothing

Two reasons may be mentioned why smoothing should be applied.

The first reason is that inverse Radon transforming is not a bounded mapping, so noise in the data may be largely amplified. However, the suggestion of Marr to take $P^{(M)}$ for some suitable M is not recommendable since in this case we throw away the information that is contained in the coefficients $\gamma_{n,k}$ with

$|n| + 2k > M$, but we could use $P^{(N-2)}$ in which the $\gamma_{n,k}$, $|n| + 2k > M$ are multiplied by a suitable weightfactor.

A second reason is that in general the function f we want to estimate is discontinuous, more specifically piecewise continuous. If we approximate such a function by a finite series expansion we will see a Gibbs phenomenon i.e. in the neighbourhood of the discontinuities we have "overshoot" that cannot be diminished by taking a higher degree approximation. Moreover, the overshoot damps out slowly so that in the output we are troubled by harmonic fluctuations. By means of smoothing we can remove the fluctuations except in neighbourhoods of the discontinuities of f .



Smoothing might be done by averaging in the final result, but it is more attractive to modify the coefficients $\gamma_{n,k}$.

There are two ways to do so:

(i) By smoothing $\hat{P}^{(M)}(p, \theta)$. $\hat{P}^{(M)}(\cos \varphi, \theta)$ is a double finite Fourier series

$$\hat{P}^{(M)}(\cos \varphi, \theta) = \sum_{|n|+2k \leq M} 2\gamma_{n,k} (|n| + 2k + 1)^{-1} \sin[(|n| + 2k + 1)\varphi] e^{in\theta}.$$

Smoothing $\hat{P}^{(M)}$ can be done by calculating $\hat{P}^{(M)}$ on the grid $(\cos \pi J/M, 2\pi I/M)$, $1 \leq J \leq M$, $1 \leq I \leq M$ and averaging as follows

$$\begin{aligned} \hat{P}_{J,I}^* = 1/16 [& (\hat{P}_{J-1,I-1} + 2\hat{P}_{J-1,I} + \hat{P}_{J-1,I+1}) + \\ & 2(\hat{P}_{J,I-1} + 2\hat{P}_{J,I} + \hat{P}_{J,I+1}) + \\ & (\hat{P}_{J+1,I-1} + 2\hat{P}_{J+1,I} + \hat{P}_{J+1,I+1})] \end{aligned}$$

where $\hat{P}_{J,I}$ denotes $\hat{P}^{(M)}(\cos(\pi J/N), 2\pi I/N)$.

Then we have

$$\hat{P}^*(\cos \varphi, \theta) = \sum_{|n|+2k \leq M} 2\gamma_{n,k}^* (|n| + 2k + 1)^{-1} \sin[(|n| + 2k + 1)\varphi] e^{in\theta}$$

where

$$(6.7.1) \quad \gamma_{n,k}^* = \gamma_{n,k} \cos^2[(|n| + 2k + 1)\pi/2M] \cos^2[n\pi/2M] .$$

Occasionally we might modify only those $\gamma_{n,k}$ for which $M_0 \leq |n| + 2k \leq M$ for some M_0 .

(ii) $P^{(M)}$ is the sum of a number of polynomials that are mutually orthogonal, since

$$(6.7.2) \quad P^{(M)}(x,y) = \sum_{\ell=0}^M \left[\sum_{k=0}^{[\ell/2]} \{ \gamma_{\ell-2k,k} e^{i(\ell-2k)\psi} + \gamma_{-\ell+2k,k} e^{-i(\ell-2k)\psi} \} * r^{\ell-2k} P_k^{(0,\ell-2k)}(2r^2 - 1) \right]$$

where $x = r \cos \psi$, $y = r \sin \psi$.

The expression between the square brackets is a polynomial of exact degree ℓ and an element of $\mathbb{P}_\ell(\mathbb{D})$. So we write

$$(6.7.3) \quad P^{(M)}(x,y) = \sum_{\ell=0}^M P_\ell(x,y)$$

where $P_\ell \in \mathbb{P}_\ell(\mathbb{D})$.

Therefore it seems to be reasonable to smooth as follows:

$$P^*(x,y) = \sum_{\ell=0}^M c_\ell P_\ell(x,y)$$

where e.g.

$$c_\ell = \cos^2(\ell\pi/2M) .$$

For the coefficients $\gamma_{n,k}$ this implies

$$(6.7.4) \quad \gamma_{n,k}^* = \gamma_{n,k} \cos^2[(|n| + 2k + 1)\pi/2M] .$$

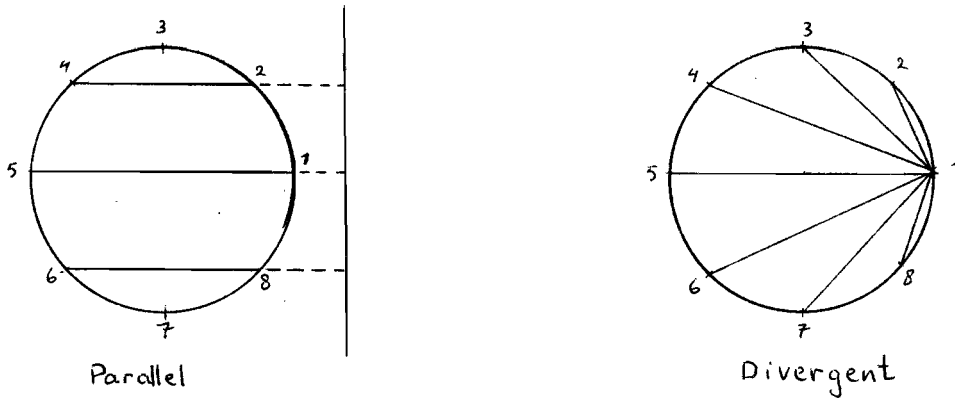
The difference with (6.7.1) is the factor $\cos^2(n\pi/2M)$.

Again we may apply (6.7.4) only to those $\gamma_{n,k}$ for which $M_0 \leq |n| + 2k \leq M$ for some M_0 .

Note. If we are dealing with the $a_{n,k}$ and $b_{n,k}$ of (6.5.12) then for fixed n,k we must treat $a_{n,k}$ and $b_{n,k}$ in exactly the same way.

6.8. Parallel- and divergent-ray sampling

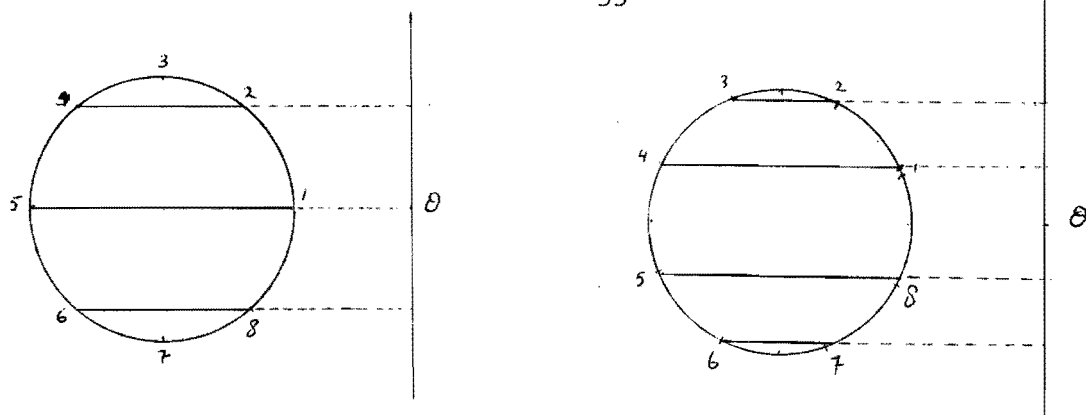
We want to make a remark about the way the samples $g[I,J]$ may be obtained. The routine described in the introduction has been called "parallel ray sampling". The object, or emitter-collimator combination, must be shifted $[N/2]$ resp. $[(N - 1)/2]$ times to sample one projection. Apart from that, the shifts are not constant. Because this must be done for every projection this procedure may be very time consuming. One way to reduce the time acquired for sampling is by means of "divergent ray sampling".



The idea is to put a point emitter on the circumference of \mathbb{D} and in the other points of the regular polygon collimators. Then only N rotations over a constant angle are required.

6.9. Another system of sample points

If we consider the way of sampling of g by means of the $g[I,J]$ then it is striking that not all projections are treated in the same way if N is even. Alternately $N/2$ and $(N - 2)/2$ sample points per projection are taken. Also the sample points on the projections are different as may be seen in the diagram.



One may wonder why this particular system of sample points is chosen, viz. an obvious system seems to be

$$(6.9.1) \quad (\cos(\pi J/N), \pi I/N), \quad 1 \leq J < N, \quad 1 \leq I \leq 2N .$$

Then each projection is sampled in exactly $N-1$ points.

We proceed by developing a total orthogonal system of functions on the sample points (6.9.1).

This can be done by means of the $F_{n,k}$ we met in section 6.5.

Instead of (6.9.1) essentially the following system is important:

$$(6.9.2) \quad P := \{(\cos(\pi J/N), \exp(i\pi I/N) \mid 1 \leq J \leq N, 1 \leq I \leq N) \} .$$

We may partition P as follows. Let

$$(6.9.3) \quad P_1 := \{(\cos(\pi J/N), \exp[i\pi(2L+J-2)/N]) \mid 1 \leq J < N, 1 \leq L \leq N\} ,$$

$$(6.9.4) \quad P_2 := \{(\cos(\pi J/N), \exp[i\pi(2L+J-1)/N]) \mid 1 \leq J < N, 1 \leq L \leq N\} .$$

By writing

$$P_1 = \{(\cos(\pi J/N), \exp(i\pi I/N)) \mid 1 \leq J \leq N, 1 \leq I < N, I-J \text{ even}\}$$

and

$$P_2 = \{(\cos(\pi J/N), \exp(i\pi I/N)) \mid 1 \leq J \leq N, 1 \leq I < N, I-J \text{ odd}\} ,$$

we see that

$$(6.9.5) \quad P_1 \cap P_2 = \emptyset$$

and

$$(6.9.6) \quad P = P_1 \cup P_2 .$$

In section 6.5 we have met functions that are mutually orthogonal on P_1 resp. P_2 . Let

$$(6.9.7) \quad \hat{F}_{n,k}^{(1)}(J,I) := \begin{cases} \sin[(|n| + 2k + 1)\pi J/N] \exp(in\pi I/N) & \text{if } I - J \text{ even} \\ 0 & \text{if } I - J \text{ odd} \end{cases}$$

and

$$(6.9.8) \quad \hat{F}_{n,k}^{(2)}(J,I) := \begin{cases} 0 & \text{if } I - J \text{ even} \\ \sin[(|n| + 2k + 1)\pi J/N] \exp(in\pi I/N) & \text{if } I - J \text{ odd} . \end{cases}$$

Then, from (6.5.8), we have

$$(6.9.9) \quad \sum_{I=1}^{2N} \sum_{J=1}^{N-1} \hat{F}_{n,k}^{(p)}(J,I) \overline{\hat{F}_{m,\ell}^{(q)}(J,I)} = N^2 / 2 \delta_{pq} \delta_{nm} \delta_{k\ell}$$

if $p, q = 0$ or 1 , $|n| + 2k \leq N - 2$, $|m| + 2\ell \leq N - 2$, because the sum is apparently zero if $p \neq q$ and if $p = q$ we have a sum as in (6.5.8).

So we have $N(N - 1)$ orthogonal functions (or rather vectors). Let $g(p, \theta)$ be a function on \mathbb{C} that satisfies the condition $g(p, \theta) = g(-p, \theta + \pi)$. The samples of g are denoted as

$$(6.9.10) \quad g(J,I) := g(\cos(\pi J/N), \pi I/N)$$

and the symmetry relation is

$$(6.9.11) \quad g(J,I) = g(N - J, I + N) ,$$

if $I \leq N$.

So all possible samples $g(J,I)$, $1 \leq J \leq N$, $1 \leq I \leq N$ that satisfy (6.9.11) span a linear space S of dimension $N(N - 1)$.

Therefore we may conclude that the $\hat{F}_{n,k}^{(1)}, \hat{F}_{n,k}^{(2)}$, $|n| + 2k \leq N - 2$ form an orthogonal basis on S .

From (6.9.9) it follows that

$$(6.9.12) \quad \hat{F}_{n,k}^{(1)} + \hat{F}_{n,k}^{(2)}, \hat{F}_{n,k}^{(1)} - \hat{F}_{n,k}^{(2)}, |n| + 2k \leq N - 2$$

form an orthogonal basis of S too.

We have

$$(6.9.13) \quad \hat{F}_{n,k}^{(1)}(J,I) + \hat{F}_{n,k}^{(2)}(J,I) = \sin[(|n| + 2k + 1)\pi J/N] \exp[in\pi I/N] ,$$

$$(6.9.14) \quad \hat{F}_{n,k}^{(1)}(J,I) - \hat{F}_{n,k}^{(2)}(J,I) = (-1)^{I-J} \sin[(|n| + 2k + 1)\pi J/N] \exp[in\pi I/N] .$$

In (6.9.13) we have discrete values of the function

$$(6.9.15) \quad \widehat{G}_{n,k}(p, \theta) = (1-p^2)^{\frac{1}{2}} U_{|n|+2k}(p) e^{in\theta}, \quad |n| + 2k \leq N - 2,$$

for $(p, \theta) = (\cos(\pi J/N), \pi I/N)$.

We like to write (6.9.14) as a polynomial of the form (6.9.15) by modifying n and k .

Let $m_n := n + s(n)N$ where

$$(6.9.16) \quad s(n) = \begin{cases} 1 & \text{if } n \geq 0 \\ -1 & \text{if } n < 0. \end{cases}$$

Then we have $|m_n| = N + |n|$. Instead of (6.9.14) we may write

$$(6.9.17) \quad \widehat{F}_{n,k}^{(1)}(J, I) - \widehat{F}_{n,k}^{(2)}(J, I) = \sin[(|m_n| + 2k + 1) \pi J/N] \exp[i m_n \pi I/N].$$

These are discrete values of the function

$$(6.9.18) \quad \widehat{G}_{m_n,k}(p, \theta) = (1-p^2)^{\frac{1}{2}} U_{|m_n|+2k}(p) e^{i m_n \theta}, \quad |n| + 2k \leq N - 2$$

for $(p, \theta) = (\cos(\pi J/N), \pi I/N)$.

From (6.9.12) we conclude that for every g numbers $a_{n,k}$ and $b_{n,k}$ exist such that for every (J, I) we have

$$(6.9.19) \quad g(J, I) = \sum_{|n|+2k \leq N-2} \{ a_{n,k} \widehat{G}_{n,k}(J, I) + b_{n,k} \widehat{G}_{m_n,k}(p, \theta) \}.$$

Therefore a solution of the discrete problem is obtained by inversion of the continuous version of (6.9.19):

$$(6.9.20) \quad \widehat{P}^{(N-2)}(x, y) = \sum_{|n|+2k \leq N-2} \{ a_{n,k} G_{n,k}(x, y) + b_{n,k} G_{m_n,k}(x, y) \}.$$

The analogue of the theorem in section 6.5 is easily formulated.

We consider formula (6.9.18).

From (6.9.15) and (6.9.16) we obtain

$$\widehat{G}_{n,k}(p, \theta) = (|n| + 2k + 1)/2 \widehat{F}_{n,k}(p, \theta)$$

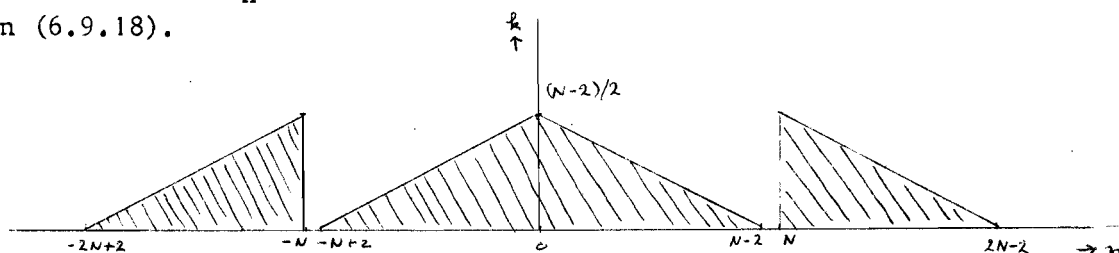
so we must have

$$(6.9.21) \quad G_{n,k}(x, y) = (|n| + 2k + 1)/2 F_{n,k}(x, y)$$

and also

$$(6.9.22) \quad G_{m_n, k}(x, y) = (|m_n| + 2k + 1)/2 \quad F_{m_n, k}(x, y) .$$

Recalling that $m_n = n + s(n)N$ we may plot the indices (n, k) that are allowed in (6.9.18).



The allowed indices will lie in the shaded regions. Now we see the difference with section 6.5. Then all the allowed indices will lie in the central triangles while the two side bands are missing.

We see that with this new system of sample points we are still very well capable of solving the discrete problem.

In practice we will remove the two side bands (and not even calculate these coefficients). This may be interpreted as smoothing. To calculate the coefficients of the central triangle is twice as expensive as in the system of section 6.5. The question is whether this is worth while.

We liked to make a final note on the reasoning of (6.9.14) sqq. To remove the factor $(-1)^{I-J}$ we must modify n and $|n| + 2k + 1$ by adding or subtracting odd multiples of N . For instance $n \rightarrow n + s(n)N$ and k unchanged does the trick. It is easy to see that all other possibilities yields indices (n, k) for which $|n| + 2k > 2N - 2$. So (6.9.17) is the polynomial with the smallest degree.

Chapter 7. The calculation of $P^{(M)}$ on \mathbb{D}

7.1. Introduction: The polar grid

We must calculate

$$(7.1.1) \quad P^{(M)}(r \cos \psi, r \sin \psi) = \sum_{n=0}^M \sum_{k=0}^{[(M-n)/2]} \{ \alpha_{n,k} \sin n\psi + \beta_{n,k} \cos n\psi \} r^n P_k^{(0,n)}(2r^2 - 1)$$

on a finite set of points $(r \cos \psi, r \sin \psi)$ that cover \mathbb{D} more or less uniformly.

We rewrite (7.1.1) as

$$(7.1.2) \quad P^{(M)}(r \cos \psi, r \sin \psi) = \sum_{n=0}^M \{ A_{n,k}(r) \sin n\psi + B_{n,k}(r) \cos n\psi \}$$

where

$$(7.1.3) \quad A_{n,k}(r) = r^n \sum_{k=0}^{[(M-n)/2]} \alpha_{n,k} P_k^{(0,n)}(2r^2 - 1)$$

$$(7.1.4) \quad B_{n,k}(r) = r^n \sum_{k=0}^{[(M-n)/2]} \beta_{n,k} P_k^{(0,n)}(2r^2 - 1) .$$

So it is obvious that we choose the points where we calculate $P^{(M)}$, as much as possible on concentric circles. Then the following "polar" grid comes into consideration.

For natural N we have

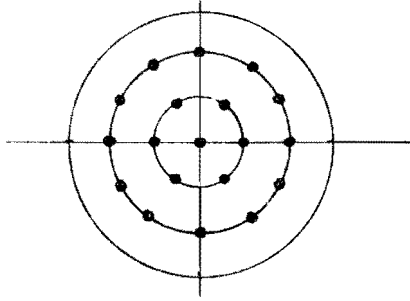
$$(r_0, \psi_0) = (0, 0)$$

$$(r_i, \psi_{ij}), \quad i = 1, 2, \dots, N$$

where

$$r_i = i/N \quad \text{and} \quad \psi_{ij} = \pi j / (3i), \quad j = 0, 1, \dots, 6i - 1 .$$

Then the distance in radial direction is $1/N$ and in tangential direction ca. $\pi / (3N)$.



Gridpoints for $r_i = 0, 1/N$ and $2/N$. The arc between two consecutive points has length $\pi/(3N)$.

Because for fixed r , $P^{(M)}$ is a trigonometric polynomial of degree $\leq M$ it is reasonable to add the restriction that if $3i > M$ we put

$$\psi_{ij} = \pi j/M, \quad j = 0, 1, \dots, 2M-1 .$$

If (7.1.3) and (7.1.4) are calculated then calculating $P^{(M)}$ is just a Fourier synthesis.

Therefore the problem is to calculate series of the form

$$(7.1.5) \quad r^n \sum_{k=0}^m a_k P_k^{(0,n)}(2r^2 - 1)$$

to which we shall refer as Jacobi series.

7.2. Calculation of Jacobi series

Let

$$(7.2.1) \quad Q_{n,k}(x) := P_k^{(0,n)}(2x - 1) .$$

We have the following formulas at our disposal (see [17], [24])

$$(7.2.2) \quad (n + 2k + 2)Q_{n-1,k+1} = (k + 1)Q_{n,k} + (n + k + 1)Q_{n,k+1}, \quad n \geq 0, \quad k \geq 0 ,$$

$$(7.2.3) \quad (n + 2k + 1)xQ_{n,k} = (n + k)Q_{n-1,k} + (k + 1)Q_{n-1,k+1} ,$$

where we have dropped the arguments x of the $Q_{n,k}$.

The "starting" values are

$$(7.2.4) \quad Q_{n,0}(x) = 1, \quad Q_{0,k}(x) = P_k(2x - 1) .$$

The $P_k(2x - 1)$ are shifted polynomials of Legendre. These formulas give rise to two different methods.

(i) With (7.2.2) we are able to reduce the series

$$\sum_{k=0}^m a_k Q_{n,k}(x)$$

to

$$\sum_{k=0}^m a_k^* P_k(2x - 1).$$

This reduction is independent of r (see section 7.3).

Calculation of (7.2.5) may be done by a summing method for Legendre series (see section 7.6).

(ii) With a direct method, i.e. a method that is comparable to the Reinsch-Goertzel method for Fourier series (see section 7.5).

7.3. Reduction to Legendre series

With (7.2.2) we may reduce $\sum_{k=0}^m a_k Q_{n,k}(x)$ to $\sum_{k=0}^m a_k^* P_k(2x - 1)$. The a_k^* will depend on n . For we have

$$Q_{n,k+1} = (n + 2k + 2)/(n + k + 1)Q_{n-1,k+1} - (k + 1)/(n + k + 1)Q_{n,k}$$

so given the coefficients $a_k^{(j)}$ we may find coefficients $a_k^{(j-1)}$ so that

$$(7.3.1) \quad \sum_{k=0}^j a_k^{(j)} Q_{n,k} + \sum_{k=j+1}^m a_k^{(j)} Q_{n-1,k} = \sum_{k=0}^{j-1} a_k^{(j-1)} Q_{n,k} + \sum_{k=j}^m a_k^{(j-1)} Q_{n-1,k}$$

viz.

$$(7.3.2) \quad a_k^{(j-1)} = a_k^{(j)}, \quad 0 \leq k \leq j - 2$$

$$(7.3.3) \quad a_{j-1}^{(j-1)} = a_{j-1}^{(j)} - j a_j^{(j)} / (n + j)$$

$$(7.3.4) \quad a_j^{(j-1)} = (n + 2j) a_j^{(j)} / (n + j)$$

$$(7.3.5) \quad a_k^{(j-1)} = a_k^{(j)}, \quad j + 1 \leq k \leq m.$$

Because (7.3.2) and (7.3.5) yield no change in $a_k^{(j)}$ we must take into account only (7.3.3) and (7.3.4), and therefore we may store the new coefficients

$a_{j-1}^{(j-1)}$ and $a_j^{(j-1)}$ at the memory locations occupied by $a_{j-1}^{(j)}$, $a_j^{(j)}$. By executing this procedure for $j = m, m-1, \dots, 1$ we reduce

$$\sum_{k=0}^m a_k^{(m)} Q_{n,k} \quad \text{to} \quad \sum_{k=0}^m a_k^{(0)} Q_{n-1,k} .$$

By executing this algorithm consecutively for $n, n-1, n-2, \dots, 1$ we reduce the Jacobi series to a Legendre series. In pseudo algol

```

for l := n step -1 until 1 do
for j := m step -1 until 1 do
begin a_{j-1} := a_{j-1} - j a_j / (l + j);
      a_j := (l + 2j) a_j / (l + j)
end;

```

After execution the coefficients a_k contain the a_k^* from formula (7.2.5).

However this method is not very useful because the coefficients a_k^* may be extremely large compared to the a_k . Therefore a loss of significant digits will occur in calculating $\sum a_k^* P_k$. We may see as follows that the a_k^* are large. We have ([24]):

$$P_k(1) = 1, Q_{n,k}(1) = 1.$$

$$P_k(-1) = (-1)^k, Q_{n,k}(0) = (-1)^k \binom{n+k}{k}$$

and therefore obtain

$$(7.3.6) \quad \sum_{k=0}^m a_k^* = \sum_{k=0}^m a_k$$

and

$$(7.3.7) \quad \sum_{k=0}^m (-1)^k a_k^* = \sum_{k=0}^m (-1)^k a_k \binom{n+k}{k} .$$

Since $\binom{n+k}{k}$ rapidly increases as n increases for positive k , the a_k^* may be very large indeed compared to a_k . From (7.3.6) it follows also that the a_k^* will show an alternating behaviour. In fact this is confirmed by numerical experiments since for small arguments x we have no loss of significant digits at all (see section 7.9(i)).

7.4. Direct method: A general case

We indicate how we could proceed in a general case. Given a set of orthogonal functions $F_k(x)$, $k \geq 0$ on the interval $[-1,+1]$, that satisfies

$$(7.4.1) \quad F_{k+1}(x) = (a_k x + b_k)F_k(x) - c_k F_{k-1}(x), \quad k \geq 1$$

and normed such that $F_k(1) = 1$.

The problem is to calculate

$$\sum_{k=0}^m d_k F_k(x)$$

for x close to $+1$, given the numbers d_k .

The summing of this series is essentially generating recursively the sequence $F_m(x), F_{m-1}(x), \dots, F_1(x), F_0(x)$. If we have a method to calculate this sequence we may derive from this method a routine to sum the series. This routine will be stable if and only if the method for the sequence is stable.

From (7.4.1) we may calculate the sequence $F_2(x), F_3(x), \dots$, given x , $F_0(x)$ and $F_1(x)$. In discussing the stability of this method we assume that we start from $\bar{F}_0(x) = F_0(x) + \delta_0$, $\bar{F}_1(x) = F_1(x) + \delta_1$ i.e. noisy $F_0(x)$, $F_1(x)$. Then we are generating the sequence $\bar{F}_k(x) = F_k(x) + \delta_k$, if we assume perfect arithmetic, for which holds the following

$$(7.4.3) \quad \delta_{k+1} = (a_k x + b_k)\delta_k - c_k \delta_{k-1}$$

and if x is close to $+1$ we have

$$\delta_{k+1} \approx (a_k + b_k)\delta_k - c_k \delta_{k-1}.$$

From (7.4.1) and (7.4.2) it follows that

$$(7.4.4) \quad a_k + b_k - c_k = 1.$$

If we have the auxiliary conditions

$$(7.4.5) \quad a_k \geq 0, \quad b_k \geq 0, \quad c_k \leq 0$$

then we must have from (7.4.3)

$$|\delta_{k+1}| \leq \max(|\delta_k|, |\delta_{k-1}|),$$

so this method probably is stable. However in general we have instead of (7.4.5) that

$$(7.4.6) \quad a_k \geq 0, b_k \geq 0, c_k \geq 0.$$

(E.g. if F_k is the polynomial of Chebyshev, Legendre, then $c_k = 1$ resp. $\frac{k}{k+1}$). Then this method will most certainly be unstable. Therefore we proceed in a different way. We try to calculate recursively the numbers $F_k(x)$ and $F_{k+1}(x) - F_k(x)$. With

$$(7.4.7) \quad V_{k+1}(x) := F_{k+1}(x) - F_k(x)$$

we obtain

$$(7.4.8) \quad F_k(x) = F_{k-1}(x) + V_k(x)$$

$$(7.4.9) \quad V_{k+1}(x) = a_k(x-1)F_k(x) + c_k V_k.$$

We see that in (7.4.9) the coefficient of F_k is small if x is close to $+1$. Since $V_k(x)$ is small compared with $F_k(x)$ and $F_{k+1}(x)$, if x is close to $+1$ it is not necessary for $V_k(x)$ to have high relative precision in order to obtain $F_k(x)$ with a high relative precision. Thus we expect that such a method will be numerically stable in a neighbourhood of $x = +1$. Then by means of (7.4.8),

(7.4.9) we may efficiently sum a series $\sum_{k=0}^m d_k F_k(x)$ by the Reinsch-Goertzel method. Assume

$$(7.4.10) \quad \sum_{k=\ell}^m d_k F_k(x) = e_\ell F_\ell(x) + f_\ell V_\ell(x)$$

then with (7.4.8) and (7.4.9) we may express e_ℓ and f_ℓ into $e_{\ell+1}$ and $f_{\ell+1}$, so that we may calculate e_1 and f_1 .

We add some remarks of the stability of this method.

We write (7.4.1) and (7.4.8), (7.4.9) in matrix-vector notation:

$$(7.4.11) \quad \begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix} = \begin{bmatrix} a_k x + b_k & -c_k \\ 1 & 0 \end{bmatrix} \begin{bmatrix} F_k \\ F_{k-1} \end{bmatrix}$$

resp.

$$(7.4.12) \quad \begin{bmatrix} V_{k+1} \\ F_{k+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} V_{k+1} \\ F_k \end{bmatrix}$$

and

$$(7.4.13) \quad \begin{bmatrix} V_{k+1} \\ F_k \end{bmatrix} = \begin{bmatrix} c_k & -a_k(1-x) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} V_k \\ F_k \end{bmatrix} .$$

Combining (7.4.12) and (7.4.13) yields

$$(7.4.14) \quad \begin{bmatrix} V_{k+1} \\ F_{k+1} \end{bmatrix} = \begin{bmatrix} c_k & -a_k(1-x) \\ c_k & 1 - a_k(1-x) \end{bmatrix} \begin{bmatrix} V_k \\ F_k \end{bmatrix} .$$

Let $W_k := V_k / \sqrt{1-x}$ then this becomes

$$(7.4.15) \quad \begin{bmatrix} W_{k+1} \\ F_{k+1} \end{bmatrix} = \begin{bmatrix} c_k & -a_k \sqrt{1-x} \\ c_k \sqrt{1-x} & 1 - a_k(1-x) \end{bmatrix} \begin{bmatrix} W_k \\ F_k \end{bmatrix}$$

(so we have scaled the matrix in (7.4.14)).

Now compare the spectral properties of the matrices in (7.4.11) and (7.4.15).

If $x = 1$ the former matrix has only one eigenvector while the latter matrix has two mutually orthogonal eigenvectors.

Next, if we assume a perturbation in $\begin{pmatrix} F_1 \\ F_0 \end{pmatrix}$ we see that this perturbation will be increased in $\begin{pmatrix} F_{k+1} \\ F_k \end{pmatrix}$ if x is close to $+1$, since the coefficients of the decomposition of $\begin{pmatrix} F_1 \\ F_0 \end{pmatrix}$ into the two eigenvectors of the matrix (7.4.11) will be very large ($x \neq 1$). In fact, such a composition does not exist if $x = 1$, and $c_k = 1$.

A perturbation of $\begin{bmatrix} V_1 \\ F_1 \end{bmatrix}$ will not be amplified that much.

7.5. Direct method: Jacobi series

We work out the method described in section 7.4 for Jacobi series.

We distinguish between the cases $x \geq \frac{1}{2}$ and $x < \frac{1}{2}$.

7.5.1. The case $x > \frac{1}{2}$

Define

$$d_k := Q_{n-1,k+1} - Q_{n,k}$$

$$v_k := Q_{n,k} - Q_{n-1,k}$$

$$q_k := Q_{n,k} .$$

Then from (7.2.2), (7.2.3) we have

$$(7.5.1) \quad (k+1)d_k = (n+k+1)v_{k+1}$$

$$(7.5.2) \quad (n+2k+1)(x-1)q_k = (k+1)d_k - (n+k)v_k$$

and

$$(7.5.3) \quad q_{k+1} - q_k = v_{k+1} + d_k .$$

Let $w_k := (n+k)v_k$ then we may describe the scheme

$$(7.5.4) \quad q_{k+1} = q_k + (n+2k+2)w_{k+1}/((n+k+1)(k+1))$$

$$(7.5.5) \quad w_{k+1} = w_k + (n+2k+1)(x-1)q_k .$$

These are the formulas (7.4.8), (7.4.9) for the $Q_{n,k}$. So, in calculating

$$S := \sum_{k=0}^m a_k Q_{n,k}(x)$$

we put

$$\sum_{j=k}^m a_j q_j = \beta_k q_k + \gamma_k w_k, \quad 0 \leq k \leq m$$

then:

$$\beta_k q_k + \gamma_k w_k = \beta_{k+1} q_{k+1} + \gamma_{k+1} w_{k+1} + a_k q_k .$$

From (7.5.4), (7.5.5) we obtain

$$(7.5.6) \quad \gamma_k = \gamma_{k+1} + (n+2k+2)\beta_{k+1}/((n+k+1)(k+1)) ,$$

$$(7.5.7) \quad \beta_k = \beta_{k+1} + (n+2k+1)(x-1)\gamma_k + a_k$$

with starting conditions

$$(7.5.8) \quad \beta_m = a_m, \quad \gamma_m = 0 .$$

Then finally:

$$(7.5.9) \quad S = \beta_0 q_0 + \gamma_0 w_0 = \beta_0 .$$

We write this algorithm in pseudo algol

```

β := am; γ := 0
for k := m - 1 step -1 until 0 do
  begin comment at this moment β contains βk+1 and γ contains γk+1;
    γ := γ + (n + 2k + 2)β / ((n + k + 1)(n + k));
    β := β + (n + 2k + 1)(x - 1)γ + ak
  end;
S := β;

```

7.5.2. The case $x < \frac{1}{2}$

Because we have

$$Q_{n,k}(0) = (-1)^k \binom{n+k}{k}$$

we must accept dimensionizing factors in d_k, v_k (compare to the beginning of 7.5.1). Let

$$\begin{aligned} d_k &= (k+1)Q_{n-1,k+1} + nQ_{n,k} \\ v_k &= nQ_{n,k} - (n+k)Q_{n-1,k} \\ q_k &= Q_{n,k} \end{aligned}$$

From (7.2.2) and (7.2.3) it follows that

$$(7.5.10) \quad d_k = -(n+k+1)v_{k+1}/(k+1),$$

$$(7.5.11) \quad (n+2k+1)xq_k = d_k - v_k,$$

and

$$(7.5.12) \quad (k+1)q_{k+1} = -(n+k+1)q_k - (n+2k+2)v_{k+1}/(k+1)$$

from which we derive the scheme

$$(7.5.13) \quad v_{k+1} = [-(k+1)v_k - (n+2k+1)(k+1)xq_k]/(n+k+1)$$

$$(7.5.14) \quad q_{k+1} = -(n+k+1)q_k/(k+1) - (n+2k+2)v_{k+1}/(k+1)^2.$$

To calculate $S = \sum_{k=0}^m a_k Q_{n,k}(x)$ we assume

$$\sum_{j=k}^m a_j q_j = \beta_k q_k + \gamma_k v_k, \quad 0 \leq k \leq m,$$

and obtain, with (7.5.13), (7.5.14):

$$(7.5.15) \quad \gamma_k = [-(k+1)\gamma_{k+1} + (n+2k+2)\beta_{k+1}/(k+1)]/(n+k+1)$$

$$(7.5.16) \quad \beta_k = -(n+k+1)\beta_{k+1}/(k+1) + (n+2k+1)x\gamma_k + a_k$$

with starting conditions

$$(7.5.17) \quad \beta_m = a_m; \quad \gamma_m = 0.$$

Then we have $S = \beta_0 q_0 + \gamma_0 v_0 = \beta_0$.

In pseudo algol we describe the algorithm:

```

β := am, γ := 0
for k := m - 1 step -1 until 0 do
begin comment at this moment β contains βk+1 and γ contains γk+1;
  γ := [-(k+1)γ + (n+2k+2)β/(k+1)]/(n+k+1);
  β := -(n+k+1)β/(k+1) + (n+2k+1)xγ + ak
end;
S := β;

```

7.6. Special case: Legendre series; two algorithms

If $n = 0$ the problem is to calculate a series of Legendre polynomials with argument $2x - 1$.

The recurrence relation for Legendre polynomials is

$$(7.6.1) \quad \begin{cases} P_{k+1}(t) = (2k+1)t P_k(t)/(k+1) - kP_{k-1}(t)/(k+1) \\ P_0(t) = 1, P_1(t) = t. \end{cases}$$

By putting

$$\sum_{j=k}^m a_j P_j(t) = b_k P_k(t) + d_k P_{k-1}(t)$$

we may derive the following algorithm.

Algorithm I.

$$\begin{cases} b_k = (2k+1)t b_{k+1}/(k+1) - (k+1)b_{k+2}/(k+2), & 0 \leq k \leq m \\ b_{m+1} = 0, & b_{m+2} = 0. \end{cases}$$

Then

$$\sum_{k=0}^m a_k P_k(t) = b_0.$$

Let $c_k = b_k - kb_{k+1}/(k+1)$ then this algorithm becomes

Algorithm II^A.

$$\begin{cases} c_k = (2k+1)(t-1)b_{k+1}/(k+1) + c_{k+1} + a_k \\ b_k = c_k + kb_{k+1}/(k+1) \\ c_{m+1} = 0, & b_{m+1} = 0. \end{cases}$$

By the substitution $f_k := 2b_{k+1}/(k+1)$ we see that this algorithm is equivalent with (7.5.6), (7.5.7), if $n = 0$.

Let $c_k = b_k + kb_{k+1}/(k+1)$ then algorithm I becomes

Algorithm II^B.

$$\begin{cases} c_k = (2k+1)(t+1)b_{k+1} - c_{k+1} + a_k \\ b_k = c_k - kb_{k+1}/(k+1) \\ c_{m+1} = 0, & b_{m+1} = 0. \end{cases}$$

With the substitution $f_k = 2b_{k+1}/(k+1)$ it is easy to show that this algorithm is equivalent to (7.5.15), (7.5.16), if $n = 0$.

Both in algorithm II^A, II^B we have

$$\sum_{k=0}^m a_k P_k(t) = b_0.$$

7.7. Numerical stability of algorithm I

We discuss the numerical stability of the algorithms I and II^A, II^B (see section 7.8).

Let $b_k = E_k(r)$ be the solution fo algorithm I if $a_r = 1$, $a_k = 0$ if $0 \leq k \leq r - 1$.

Then we have

$$(7.7.1) \quad \begin{cases} E_k(r) = (2k+1)t E_{k+1}(r) - (k+1)E_{k+2}(r)/(k+2), & 0 \leq k \leq r \\ E_{r+1}(r) = 0, E_r(r) = 1. \end{cases}$$

But also, see section 7.10.2,

$$(7.7.2) \quad \begin{cases} E_0(r+1) = (2r+1)t E_0(r) - rE_0(r-1)/(r+1), & r \geq 0 \\ E_0(0) = 1, E_0(-1) = 0, \end{cases}$$

from which we conclude that $E_0(r) = P_r(t)$.

We may interprete $E_k(r)$ as the error in b_k as a result of error 1 in b_r and error 0 in b_{r+1} . Then, in the general case that errors ε_k occur in every b_k , we may write the final error in b_0 as

$$(7.7.3) \quad b_0 - \bar{b}_0 = \sum_{k=0}^m \varepsilon_k E_0(k) = \sum_{k=0}^m \varepsilon_k P_k(t)$$

where $\varepsilon_k \leq |b_k| 2^{-t}$. (2^{-t} is the arithmetic precision.)

So the error in b_0 depends on the magnitude of b_k , $0 < k \leq m$. Let

$$(7.7.4) \quad \tilde{P}_n(t) := \sum_{k=1}^n P_{k-1}(t)P_{n-k}(t)/k, \quad n \geq 0,$$

then we have, see section 7.10.2,

$$(7.7.5) \quad \begin{aligned} b_k &= k \sum_{r=k}^m a_r \{P_{k-1}(t)\tilde{P}_r(t) - \tilde{P}_{k-1}(t)P_r(t)\} \quad \text{if } k \geq 1 \\ b_0 &= \sum_{r=0}^m a_r P_r(t). \end{aligned}$$

So if $t = 1$ we have

$$b_k = k \sum_{r=k}^m a_r \sum_{\ell=k}^r 1/\ell, \quad k \geq 1$$

resulting in

$$|b_0 - \bar{b}_0| \leq 2^{-t} \left\{ \sum_{k=1}^m \sum_{r=k}^m |a_r| \sum_{\ell=k}^r k/\ell + \sum_{r=0}^m |a_r| \right\},$$

or equivalently

$$|b_0 - \bar{b}_0| \leq 2^{-t} \left\{ \sum_{r=1}^m \sum_{k=1}^r |a_r| \sum_{\ell=k}^r k/\ell + \sum_{r=0}^m |a_r| \right\}.$$

By means of the identity

$$\sum_{k=1}^r \sum_{\ell=k}^r k/\ell = \sum_{\ell=1}^r \sum_{k=1}^{\ell} k/\ell = \frac{1}{2} \sum_{\ell=1}^r \ell(\ell+1)/\ell = r(r+2)/4$$

we may reduce this to

$$|b_0 - \bar{b}_0| \leq 2^{-t} \left\{ \frac{1}{4} \sum_{r=1}^m r(r+2) |a_r| + \sum_{r=0}^m |a_r| \right\}.$$

The right hand side of this inequality rapidly increases as m increases (if a_r does not decrease faster than e.g. r^{-3} if $r \rightarrow \infty$). Of course this is worst case analysis, so as a result of statistical effects the error will always be smaller. However it appears that algorithms II^{A,B} are better behaved.

7.8. Numerical stability of algorithm II^{A,B}

We consider algorithm II^A if $t \geq 0$ (the case of algorithm II^B if $t \leq 0$ is analogous). The numbers b_k, c_k generated by algorithm II^A satisfy the relations (see section 7.7.5):

$$(7.8.1) \quad b_k = k \sum_{r=k}^m a_r \{P_{k-1}(t) \tilde{P}_r(t) - \tilde{P}_{k-1}(t) P_r(t)\}$$

$$(7.8.2) \quad c_k = k \{P_{k-1}(t) - P_k(t)\} \sum_{r=k}^m a_r \tilde{P}_r(t) + k \{\tilde{P}_{k-1}(t) - \tilde{P}_k(t)\} \sum_{r=k}^m a_r P_r(t).$$

Let $\cos \theta = t$ then we discuss $P_r(\cos \theta)$ and $\tilde{P}_r(\cos \theta)$. From [5] we have the formulas

$$(7.8.3) \quad P_n(\cos \theta) = (\theta/\sin \theta)^{\frac{1}{2}} J_0[(n + \frac{1}{2})\theta] + O(n^{-3/2}), \quad n \rightarrow \infty$$

$$(7.8.4) \quad |(\sin \theta)^{\frac{1}{2}} P_n(\cos \theta)| \leq (4/(\pi n))^{\frac{1}{2}}.$$

Both relations hold uniformly in $\theta \in [0, \pi/2]$. (This corresponds to $t \in [0, 1]$.) From (7.8.3) it follows by means of the mean value theorem that

$$\begin{aligned} P_{n+1}(\cos \theta) - P_n(\cos \theta) &= (\theta/\sin \theta)^{\frac{1}{2}} \theta J_0'[(n + \frac{1}{2} + \xi)\theta] + O(n^{-3/2}) \\ &= -(\theta/\sin \theta)^{\frac{1}{2}} \theta J_1'[(n + \frac{1}{2} + \xi)\theta] + O(n^{-3/2}), \end{aligned}$$

if $n \rightarrow \infty$, uniformly in $\theta \in [0, \pi/2]$, where ξ is some number, dependent on θ and n , in the interval $[0, 1]$.

Therefore constants K_1, K_2 exist such that

$$(7.8.5) \quad |P_{n+1}(\cos \theta) - P_n(\cos \theta)| \leq K_1(\theta/n)^{\frac{1}{2}} + K_2 n^{-3/2}, \quad n \geq 1$$

if $\theta \in [0, \pi/2]$.

From the definition of \tilde{P}_n and (7.8.4) it follows with (7.8.4) that a constant K_3 exists such that

$$(7.8.6) \quad |(\sin \theta)^{\frac{1}{2}} \tilde{P}_n(\cos \theta)| \leq K_3 \text{ if } n \geq 0, \theta \in [0, \pi/2].$$

Then we have from (7.8.1) resp. (7.8.2) that

$$(7.8.7) \quad |(\sin \theta)^{3/2} b_k| \leq c^{(1)} \sum_{r=k}^m |a_r|$$

$$(7.8.8) \quad |c_k| \leq c^{(2)} b^{1/2} \sum_{r=k}^m |a_r|.$$

Considering algorithm II^A and recalling that $\cos \theta = t$ and therefore $t - 1 = -2 \sin^2(\theta/2)$, we see that the coefficients c_k and b_k are calculated with good relative precision. Because c_k is small compared with b_{k+1} we see that the b_k are calculated accurately.

7.9. Numerical experiments

(i) Reduction to Legendre series. We have experimented with the algorithm of section 7.3. To test the numerical stability we added a relative error of at most $.5_{10}^{-6}$ to the elements a_j and a_{j-1} at the moment they are calculated. These experiments have been carried out on the P9200 with machine precision of 10^{-8} . If the algorithm is executed a number of times (e.g. 10 or 20) we find an interval for the b_k . The (relative) magnitude of these intervals is an indication of the stability of this algorithm to round off errors. As an example we have chosen

$$S := \sum_{k=0}^m a_{n,k} Q_{n,k}$$

and reduce this to

$$S = \sum_{k=0}^m b_{n,k} Q_{0,k}$$

with $n = 20$ and $a_{n,k} = (n+k+1)^{-2}$.

We give the intervals we mentioned above for some b_k (between brackets the number of times that the algorithm was executed).

$$m = 10 \quad b_0 \quad (.134 \ 231 \ 2 \pm .000 \ 000 \ 3) \quad 10^{+4} \quad (21)$$

$$b_1 \quad (-.360 \ 306 \ 0 \pm .000 \ 000 \ 1) \quad 10^{+4} \quad (3)$$

$$b_5 \quad (-.250 \ 039 \ 5 \pm .000 \ 000 \ 5) \quad 10^{+4} \quad (21)$$

$$b_{10} \quad (.477 \ 419 \ \pm .000 \ 001 \) \quad 10^{+4} \quad (21)$$

$$m = 10 \quad b_0 \quad (+.532 \ 381 \ \pm .000 \ 001 \) 10^{+10} \quad (21)$$

$$b_1 \quad (-.157 \ 307 \ 23 \pm .000 \ 000 \ 01) 10^{+10} \quad (3)$$

$$b_{20} \quad (+.651 \ 858 \ \pm .000 \ 001 \) 10^{+10} \quad (21)$$

$$b_{40} \quad (+.343 \ 626 \ \pm .000 \ 001 \) 10^{+2} \quad (21)$$

We see that the coefficients b_k are determined with high relative precision (since we introduced relative errors of $.5 \cdot 10^{-6}$). So the stability of the b_k to round off errors is satisfactory, but the absolute errors are very large. If x is close to $+1$ we have loss of significant digits so the large absolute errors show in the results.

t	$t^{n/2} \sum_{k=0}^m b_k Q_{0,k}(t), n = 20, m = 40$	number of executions
0	$(+.666 \ 087 \ \pm .000 \ 001) 10^{+12} \ ^*)$	11
.1	$(+.19 \ \pm .03 \) 10^{-4}$	11
.5	$(0 \ \pm .3 \) 10^{+2}$	11
.75	$(0 \ \pm .16 \) 10^{+4}$	11

We see that if x is close to 0, we have no loss of significant digits.

*) In this case we have calculated the series

$$\sum_{k=0}^m b_k Q_{0,k}(0) .$$

(ii) Direct method.

We have tested the algorithms of section 7.5 on the P9200 computer at the T.H.E. To test the stability to round off error the newly calculated β and γ are perturbed with relative error of at most $.5 \cdot 10^{-6}$. Again the algorithm was executed a number of times to obtain an impression of the numerical stability.

As a testcase we used again

$$\sum_{k=0}^m a_{n,k} Q_{n,k}(t)$$

where $a_{n,k} = (n+k+1)^{-2}$.

The results are tabulated in the table below.

x	$t^{n/2} \sum_{k=0}^m a_{n,k} Q_{n,k}(t); n = 20; m = 40$	number of executions
0	$(+.666\ 085 \pm .000\ 004)_{10^{+12}}$ *)	11
.1	$(+.198\ 244 \pm .000\ 001)_{10^{-4}}$	25
.5	$(+.196\ 180 \pm .000\ 002)_{10^{-4}}$	11
.75	$(+.291\ 820 \pm .000\ 004)_{10^{-4}}$	11

7.10. APPENDIX: The recurrence relation for Legendre polynomials: General solution

7.10.1. The solution $\tilde{P}_n(t)$ that is independent of $P_n(t)$

We recall the recurrence relation for the Legendre polynomials

$$(7.10.1) \quad p_{k+1} = (2k+1)t p_k / (k+1) - k p_{k-1} / (k+1), \quad k \geq 1.$$

We know one solution viz. $p_k = P_k(t)$, the k-th Legendre polynomial. We may find the other solution by means of generating functions. We have from

(7.10.1) that

$$(k+1)p_{k+1} = 2k t p_k - (k-1)p_{k-1} + t p_k - p_{k-1}, \quad k \geq 1.$$

*) In this case we calculated

$$\sum_{k=0}^m a_{n,k} Q_{n,k}(0).$$

Multiplying both sides with t^k and summation over $k \geq 1$ yield, on the assumption that the infinite series converge,

$$\sum_{k=1}^{\infty} (k+1)p_{k+1}x^k = 2tx \sum_{k=1}^{\infty} kp_k x^{k-1} - x^2 \sum_{k=1}^{\infty} (k-1)p_{k-1}x^{k-2} ,$$

$$+ t \sum_{k=1}^{\infty} p_k x^k - x \sum_{k=1}^{\infty} p_{k-1}x^{k-1} .$$

Let

$$(7.10.2) \quad p(x) = \sum_{k=0}^{\infty} p_k x^k$$

then we have

$$p'(x)(1 - 2xt + x^2) + p(x)(-t + x) = p_1 - tp_0 ,$$

or

$$(7.10.3) \quad \frac{d}{dx}\{p(x)(1 - 2xt + x^2)^{\frac{1}{2}}\} = (p_1 - tp_0)(1 - 2xt + x^2)^{-\frac{1}{2}} .$$

We still have the freedom to choose p_0 and p_1 .

(i) $p_0 = 1, p_1 = t$ then $p(0) = 1$ and from (7.10.3) it follows then that

$$p(x) = (1 - 2xt + x^2)^{-\frac{1}{2}} ,$$

which is the generating function of the Legendre polynomials, i.e.

$$(7.10.4) \quad \sum_{k=0}^{\infty} (P_k(t)x^k = (1 - 2xt + t^2)^{-\frac{1}{2}} .$$

In this case the assumption that the infinite series converge holds if $|t| \leq 1, |x| < 1$.

(ii) $p_0 = 0, p_1 = 1$. Comparing with (i) we see that this will result in a solution p_n that is independent of P_n . Now we have $p(0) = 0$ and from (7.10.3), (7.10.4) we have

$$p(x)(1 - 2xt + x^2)^{\frac{1}{2}} = \sum_{k=0}^{\infty} (k+1)^{-1} P_k(t)x^{k+1}$$

and again with (7.10.4) we obtain that

$$p(x) = \left\{ \sum_{k=0}^{\infty} (k+1)^{-1} P_k(t) x^{k+2} \right\} \left\{ \sum_{k=0}^{\infty} P_k(t) x^k \right\} .$$

Therefore we have

$$(7.10.5) \quad p(x) = \sum_{n=0}^{\infty} \left\{ \sum_{k=0}^n (k+1)^{-1} P_k(t) P_{n-k}(t) \right\} x^{n+1} .$$

So, looking back to the definition of $p_n(x)$ we obtain

$$p_n(x) = \sum_{k=0}^{n-1} (k+1) P_k(t) P_{n-k-1}(t), \quad n \geq 1 .$$

We denote this solution as

$$(7.10.6) \quad \tilde{P}_n(t) = \sum_{k=1}^n P_{k-1}(t) P_{n-k}(t) / k, \quad n \geq 1 .$$

The independence of this solution from $P_n(t)$ follows from the relation

$$(7.10.7) \quad P_n(t) \tilde{P}_{n+1}(t) - \tilde{P}_n(t) P_{n+1}(t) = (n+1)^{-1} ,$$

which is proved as follows.

We have

$$\tilde{P}_{n+1}(t) = (2n+1)t\tilde{P}_n(t)/(n+1) - n\tilde{P}_{n-1}(t)/(n+1), \quad n \geq 1$$

resp.

$$P_{n+1}(t) = (2n+1)tP_n(t)/(n+1) - nP_{n-1}(t)/(n+1), \quad n \geq 0 .$$

Multiplying the first equation with $P_n(t)$ and the second with $\tilde{P}_n(t)$ and subtraction of this two equations results in

$$P_m(t) \tilde{P}_{m+1}(t) - \tilde{P}_m(t) P_{m+1}(t) = m \{ P_{m-1}(t) \tilde{P}_m(t) - \tilde{P}_{m-1}(t) P_m(t) \} / (m+1)$$

if $m \geq 1$.

By induction we obtain:

$$P_m(t) \tilde{P}_{m+1}(t) - \tilde{P}_m(t) P_{m+1}(t) = \{ P_0(t) \tilde{P}_1(t) - \tilde{P}_0(t) P_1(t) \} / (m+1)$$

and (7.10.7) follows.

7.10.2. The numbers generated by algorithm I

We reconsider algorithm I of section 7.6. We would like to have an expression for the b_k . Analogous to (7.7.3) we may derive that

$$(7.10.8) \quad b_k = \sum_{r=k}^m a_r E_k(r)$$

where the $E_k(r)$ satisfy (7.7.1).

These recurrence relations may be written in matrix-vector notation as

$$\begin{bmatrix} E_k(r) \\ E_{k+1}(r) \end{bmatrix} = \begin{bmatrix} (2k+1)t/(k+1) & -(k+1)/(k+2) \\ 1 & 0 \end{bmatrix} \begin{bmatrix} E_{k+1}(r) \\ E_{k+2}(r) \end{bmatrix} .$$

Denote the matrix by A_k .

By induction we have

$$\begin{bmatrix} E_k(r) \\ E_{k+1}(r) \end{bmatrix} = A_k A_{k+1} \cdots A_{r-1} \begin{bmatrix} E_r(r) \\ E_{r+1}(r) \end{bmatrix}, \quad k \leq r-1 .$$

Since $E_r(r) = 1$; $E_{r+1}(r) = 0$ it is not difficult to derive that

$$\begin{bmatrix} E_k(r+1) \\ E_{k+1}(r+1) \end{bmatrix} - (2r+1)t/(r+1) \begin{bmatrix} E_k(r) \\ E_{k+1}(r) \end{bmatrix} + r/(r+1) \begin{bmatrix} E_k(r-1) \\ E_{k+1}(r-1) \end{bmatrix} = 0, \\ k \leq r-1 .$$

Therefore

$$(7.10.9) \quad E_k(r+1) = (2r+1)tE_k(t)/(r+1) - rE_k(r-1)/(r+1)$$

with the starting conditions

$$(7.10.10) \quad E_k(k) = 1, \quad E_k(k-1) = 0 .$$

Since $P_k(t)$ and $\tilde{P}_k(t)$ are the two independent solutions of recurrence relation (7.10.9) we may write

$$(7.10.11) \quad E_k(r) = A_k P_r(t) + B_k \tilde{P}_r(t) ,$$

where the A_k and B_k do not depend on r .

Fitting (7.10.11) to (7.10.10) we obtain

$$E_k(r) = \{P_{k-1}(t)\tilde{P}_r(t) - \tilde{P}_{k-1}(t)P_r(t)\} / \{P_{k-1}(t)\tilde{P}_k(t) - \tilde{P}_{k-1}(t)P_k(t)\},$$

$k \geq 0$

$$E_0(r) = P_r(t) .$$

With (7.10.7) we obtain

$$(7.10.12) E_k(r) = \begin{cases} k\{P_{k-1}(t)\tilde{P}_r(t) - \tilde{P}_{k-1}(t)P_r(t)\} & \text{if } k \geq 1 \\ P_r(t) & \text{if } k = 0 . \end{cases}$$

(7.10.8) finally yields

$$(7.10.13) b_k = k \sum_{r=k}^m a_r [P_{k-1}(t)\tilde{P}_r(t) - \tilde{P}_{k-1}(t)P_r(t)], \quad k \geq 1$$

$$b_0 = \sum_{r=0}^m a_r P_r(t) .$$

Chapter 8. Numerical experiments conclusions

8.1. Implementations

We have implemented the method described in chapter 6 in the following way.

- i) Since calculating the coefficients $\alpha_{n,k}$ and $\beta_{n,k}$ (see section 6.5) as well as calculating $P^{(M)}$ involves summing of finite Fourier series, a procedure has been written that sums finite Fourier series, based on the Reinsch-Goertzel algorithm.
- ii) A procedure that calculate the coefficients $\alpha_{n,k}$ and $\beta_{n,k}$, given the array $g[I,J]$, $1 \leq J \leq N - 1$, $1 \leq I \leq N - J$, has been written, according to the formulas (6.5.11). This procedure makes use of the procedure mentioned under i).
- iii) A smoothing procedure that modifies the coefficients $\alpha_{n,k}$ and $\beta_{n,k}$ according to section 6.7i), has been written. We remark that smoothing might also take place as soon as a coefficient $\alpha_{n,k}$, $\beta_{n,k}$ has been calculated.
- iv) A procedure that calculates the polynomial $P^{(M)}$ (see section (6.5.12)) on the polar grid of section 7.1, has been written. A procedure to sum Jacobi series, based on the algorithms of section 7.5, and the procedure of (i) to sum Fourier series have been incorporated.
- v) A procedure that approximates $P^{(M)}$ on a square grid, by means of bilinear interpolation in the polar grid, has been written.

8.2. Data and noisy data

For data we use the Radon transform of the testpattern at hand. With "data with d% noise" we indicate that each value of the Radon transform is perturbed with Gaussian noise with zero mean and variance equal to $[d/100 * \text{maximum of Radon transform}]^2$.

8.3. Testpatterns

We mainly experimented with two testpatterns. The first one consisted of a two dimensional Gaussian manifold. The Radon transform may be expressed as an error function, that can be approximated by simple functions (see Abramowitz and Stegun, Handbook of Math. functions). The reconstructions are

with extremely high precision for a limited number of projections. The second testpattern is a simplification of a braintumor (see [12]).

We have the following model. Let E_1 be the region inside the ellips

$$[(6.7/3.9)x]^2 + y^2 = 0.8 .$$

Let E_2 be the region inside the ellips

$$[(6.7/3.9)x]^2 + y^2 = 0.8 * 6.2/6.7 .$$

Let E_3 be the region inside the circle

$$(x - 0.78/6.2)^2 + (y - 0.2)^2 = 0.1 .$$

Define the absorption coefficient field

$$f_1(x,y) = \begin{cases} 0 & \text{if } (x,y) \in \mathbb{D}/E_1 \\ 100 & \text{if } (x,y) \in E_1/E_2 \\ 50 & \text{if } (x,y) \in E_2/E_3 \\ 55 & \text{if } (x,y) \in E_3 \end{cases}$$

(braintumor 1).

The Radon transform is easily calculated as a combination of Radon transforms of characteristic functions of ellipses inside \mathbb{D} . To overcome the large difference between \mathbb{D}/E_1 and E_1 we also experimented with

$$f_2(x,y) = \begin{cases} 53 & \text{if } (x,y) \in \mathbb{D}/E_1 \\ f_1(x,y) & \text{if } (x,y) \in E_1 \end{cases}$$

(braintumor 2).

E_3 is the actual braintumor and E_1/E_2 is the skull.

8.4. Computation time

We give the required computation time to calculate all the coefficients $\alpha_{n,k}$ and $\beta_{n,k}$ (analysis). N is the number of projections, t is the computation time in seconds on the B6700 (we give maximum values because t shows a large variance, induced by the machine situation. To give an impression, we mention that for $N = 90$ we also found once $t = 100$ sec.)

N	t (sec)
16	1
32	6
90	150

N: number of projections, t computation time (analysis).

We also give the time to calculate $P^{(M)}$ on the polar grid; M: degree of polynomial; t computation time (synthesis)

M	t (sec)
88	187
78	159
68	118
58	91

The conversion of polar grid-cartesian grid took about 3 sec. (polar grid: 50 radii, cartesian grid: 60×60 grid points).

8.5. Discussion of results

We show a number of reconstructions from the braintumors 1, 2. In all reconstructions we used 90 projections. The reconstructions are calculated from $P^{(M)}$ in the unsmoothed case (i.e. in $P^{(N-2)}$ we have thrown away the coefficients $\alpha_{n,k}$ and $\beta_{n,k}$ with $|n| + 2k > M$. In case smoothing is applied it has been done according to section 6.7i) with $M_0 = [M/2]$.

Plate I shows the reconstruction of braintumor 1 without noise. The polynomial with maximum degree has been chosen i.e. $P^{(88)}$. The braintumor is well visible, but it is difficult to say whether it is the only irregularity. Also the skull is clearly visible (plus overshoot). However outside the skull we see large fluctuations. Plate II shows the reconstruction of braintumor 2 where in the output we have subtracted the mass outside the skull (in \mathbb{D}/E_1). In Plate II^A we have the polynomial $P^{(78)}$. In Plate II^B we have the smoothed polynomial $P^{(78)}$. We see that the fluctuations in II^A are removed for a great part in II^B. We also see that the contours of the braintumor are somewhat less sharp, as was to be expected.

The effects of smoothing on the sharpness in the contours, such as between skull and brain, and brain and braintumor, is visible in plates III^{A,B,C,D}. Again we have braintumor 2 with the smoothed polynomials $P^{(M)}$ of degree resp. 88, 78, 68, 58. We see that the fluctuations outside the skull disappear practically completely.

In Plate IV we show reconstructions of braintumor 2 with 0.2% noise. The polynomials are the smoothed versions of $P^{(M)}$ where M is resp. 88 (IV^A), 78 (IV^B), 68 (IV^C) and 58 (IV^D). We see that noise is removed effectively by smoothing, but we pay for this by the diminished sharpness of the "picture".

8.6. Conclusions

We see that the method of Marr (chapter 6) is a good method in the sense that it gives good reconstructions, if the number of projections is large enough, and that it shows good stability to noisy data (and round off errors). However, the computation time is rather long, although we believe that it can be reduced considerably, e.g. by applying Fast Fourier procedures instead of the Fourier procedure based on the Reinsch-Goertzel algorithm. Especially for large N this will be effective.

Part III

Chapter 9. The convolution method

9.1. The convolution method

We have experimented also with the convolution method of Ramachandran and Lakshininarayanan, described in section 2.3 of this report.

The computations are based on the formulae

$$(9.1.1) \quad f(x,y) = \int_0^\pi h(x \cos \theta + y \sin \theta, \theta) d\theta ,$$

where h is approximated by

$$(9.1.2) \quad h(t,\theta) = \int_{-A/2}^{+A/2} |K|G(K,\theta)e^{-2\pi iKt} dt ,$$

from which we obtained

$$(9.1.3) \quad h(n/A,\theta) = A/4[g(n/A,\theta) - 4/\pi^2 \sum_{m \text{ odd}} m^{-2}g((n+m)/A,\theta)] .$$

9.2. A variant of the convolution method

Since the function

$$H_A(K) = \begin{cases} |K| & \text{if } |K| < A/2 \\ 0 & \text{if } |K| > A/2 , \end{cases}$$

that appears in the integrand in (9.1.1) varies abruptly if $|K| = A/2$, we might replace this function by

$$(9.2.1) \quad \hat{H}_D(K) = \begin{cases} |K| & \text{if } |K| < D/2 \\ D-|K| & \text{if } D/2 < |K| < D \\ 0 & \text{if } |K| > D , \end{cases}$$

where $D = A/2$.

Then (A2) is replaced by

$$(9.2.2) \quad h(t,\theta) = \int_{-D}^{+D} \hat{H}_D(K)G(K,\theta)e^{-2\pi iKt} dt .$$

Assuming again that $G(K, \theta) = 0$ if $|K| > D$, the Whittaker-Shannon theorem can be applied once more.

We have

$$(9.2.3) \quad g(p, \theta) = \sum_{m=-\infty}^{+\infty} g(m/2D, \theta) \text{sinc}[2D(p - m/2D)]$$

and

$$(9.2.4) \quad G(K, \theta) = 1/2D \sum_{m=-\infty}^{+\infty} g(m/2D, \theta) \exp(imK/D)$$

and also

$$(9.2.5) \quad h(p, \theta) = \sum_{m=-\infty}^{+\infty} h(m/2D, \theta) \text{sinc}[2D(p - m/2D)] .$$

Substituting (9.2.4) in (9.2.5) yields

$$(9.2.6) \quad h(n/2D, \theta) = \sum_{m=-\infty}^{+\infty} g(m/2D, \theta) \frac{1}{2D} \int_{-\infty}^{+\infty} \hat{H}_D(K) e^{i\pi(m-n)K/D} dK .$$

For the integral we have

$$(9.2.7) \quad \int_{-\infty}^{+\infty} \hat{H}_D(K) e^{-2\pi i K t} dK = \frac{1}{2} D^2 \cos(\pi D t) \text{sinc}^2(\frac{1}{2} D t) .$$

So

$$(9.2.8) \quad \int_{-\infty}^{+\infty} \hat{H}_D(K) e^{-\pi i K m/D} dK = \begin{cases} 0 & \text{if } m \text{ is odd} \\ \frac{1}{2} D^2 & \text{if } m = 0 \\ 0 & \text{if } m \text{ even multiple of } 2 \\ -8D^2/\pi^2 m^2 & \text{if } m \text{ odd multiple of } 2 . \end{cases}$$

Then we obtain

$$(9.2.9) \quad h(n/2D, \theta) = D/4 [g(n/2D, \theta) - 4/\pi^2 \sum_{m \text{ odd}} m^{-2} g((n + 2m)/2D, \theta)] .$$

Comparing to (9.1.3), with $D = A/2$, we see that the structures of these formulae are the same. In this case, the steps in the array $g(n/A, \theta)$ are twice as large as in (9.1.3).

Using (9.2.9) instead of (9.1.3), we obviously expect smoother reconstructions.

9.3. Implementation

Let A and B be positive integers.

Given the array of values $g(n/A, \pi\ell/B)$, $-A < n < A$, $0 \leq \ell < B - 1$ we can calculate the array $h(n/A, \pi\ell/B)$ exactly ((9.1.3), (9.2.9)). We used the trapezoidal rule with discretization points $\theta_\ell = \pi\ell/B$, $0 \leq \ell < B - 1$ to discretize the integral (9.1.1). Approximations to the integrands $h(x \cos \theta_\ell + y \sin \theta_\ell, \theta_\ell)$ were obtained by linear interpolation in the array $h(n/A, \theta_\ell)$.

9.4. Testpatterns

As testpattern we used braintumor 1, described on page 80 of this report.

Reconstructions are shown in the plates $V^{A,B}$. In plate V^A , we have $A = 20$, $B = 24$, in all the other ones $A = 40$, $B = 48$. In plates $V^{D,E}$ 0,2% noise was added to the data.

In plates $V^{A,B,D}$ we used formula (A3) to calculate the array $h(n/A, \pi\ell/B)$, in plates $V^{C,E}$ formula (A12) with $D = A/2$.

9.5. Discussion: Comparison to Marr's method

Comparing plate V^B to the plates II^B or III^C that were constructed by Marr's method (chapter 7 of this report), it strikes that inside the skull the convolution method (first variant) gives for better results, whereas Marr's method is better outside the skull. One might think that the convolution method (second variant) should overcome this disadvantage of the convolution method. This is partly true. If we compare plate V^B to V^C we see that the large fluctuations outside the skull have been reduced considerably. However, inside the skull, but close to the boundary, we get some fluctuations. So it is difficult to decide which variant is best. Looking at the plates $V^{D,E}$, where noisy data has been used, we notice that inside the skull the two variants result in about the same quality of reconstruction. Moreover outside the skull, the second variant results in smaller fluctuations than the first one.

Finally, we compare the convolution method to Marr's method.

In case of noiseless data, the convolution method gives better results inside the skull than Marr's method, but outside the skull Marr's method is better. (Compare plates V^B and III^C). In case of noisy data, the two methods yield about the same reconstructions inside the skull, but outside of it Marr's method is better. However, a (perhaps theoretical) advantage of Marr's method is the flexibility towards noisy data. Independent of the number of

data, one can decide to reduce the degree of the reconstructing polynomial and so smooth harder, whereas in case of the convolution methods the smoothing is inherent to the sampling.

Considering the computation times (for the plates V^{B-E} in the order of 60-80 seconds on the B6700) of the two methods, it is clear that the convolution methods are more practicle. Marr's method is only cheaply applicable in case reconstructions of smooth functions (inside the unit circle) have to be made.

Note. In the plates we only show the second and third quadrant. The second quadrant shows about the same behaviour as the first and the fourth one. We mention that the cartesian grid has uniform mesh width, i.e. the distance between two consecutive points in the horizontal direction and the distance between two consecutive points in the vertical direction are equal.

In the plates the boundaries of the skull and tumor have been drawn in order to facilitate the judging of the quality of the reconstructions.

References

- [1] Born and Wolfe, Principles of Optics, Pergamon Press, New York, 1964.
- [2] Bracewell and Riddle, Inversion of fan-beam scans in radioastronomy, *Astrophys. J.* 150 (1967), 427-434.
- [3] Crowther, DeRosier and Klug, The reconstruction of a three-dimensional structure from projections and its applications to electron microscopy. *Proc. Roy. Soc. A* 317 (1970), 319-340.
- [4] DeRosier and Moore, Reconstruction of three-dimensional images from electronmicrographs of structures with helical symmetry. *J. Mol. Biol.* 52 (1970), 355-369.
- [5] Erdelyi et al., Higher transcendental functions, vol. 2, McGraw-Hill, New York, 1953.
- [6] Fox and Parker, Chebyshev polynomials in numerical analysis, Oxford University Press, London, 1968.
- [7] Gilbert, The reconstruction of a three-dimensional structure from projections and its application to electronmicroscopy II, Direct methods. *Proc. Roy. Soc. B* 182 (1972), 89-102.
- [8] Gordon, Bender and Herman, Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography. *J. Theor. Biol.* 29 (1970), 471-481.
- [9] Gordon and Herman, Three-dimensional reconstructions from projections: A review of algorithms, *Int. Rev. Cyt.* (1974), 111-151.
- [10] Herman. A relaxation method for reconstructing objects from noisy X-rays, State university of New York, Dept. of Comp. Science, Buffalo N.Y. Techn. Report 63 (1973).
- [11] Herman, Lent and Rowland, ART: Mathematics and applications, *J. Theor. Biol.* 42 (1973), 1-32.
- [12] Herman and Rowland, Three methods for reconstructing objects from X-rays. *Computer graphics and image processing* 2 (1973), 151-178.
- [13] Hounsfield, A method of and apparatus for examination of a body by radiation. Application 40317/68, The Patent Office, London, Pat. Spec. 1283915.
- [14] Junginger and Van Haringen, Calculation of three-dimensional refractive index fields using phase integrals. *Opt. Comm* 5 (1972), 1-14.

- [15] Kaczmarz, Angenäherte Auflösung von Systemen linearer Gleichungen. Bull. Acad. Pol. Sci. Lett. Cl. Sci. Math. Natur. A (1937), 355-357.
- [16] Ludwig, The Radon transformation on Euclidean space, Comm. Pure Appl. Math. 19 (1966), 49-81.
- [17] Marr, On the reconstruction of a function on a circular domain from a sampling of its line integrals. J. Math. Anal. Appl. 45 (1974), 357-374.
- [18] Matulka and Collins, Determination of three-dimensional density fields from holographic interferograms. J. Appl. Phys. 42 (1971), 1109-1119.
- [19] Radon, Über die Bestimmung von Funktionen durch ihre Integralwerte langs gewisser Männigfaltigkeiten. Ber. Saechs. Akad. Wiss. Leipzig, Math. Phys. Kl. 69 (1917), 262-277.
- [20] Ramachandran and Lakshminarayanan. Three-dimensional reconstruction from radiographs and electron micrographs. Part III-description and application of the convolution method. Ind. J. Pure Appl. Phys. 3 (1971), 997-1003.
- [21] -----, Three-dimensional reconstruction from radiographs and electron micrographs: Application of convolutions instead of Fourier transforms, Proc. Nat. Acad. Sci. USA 68 (1971), 2236-2240.
- [22] Shannon, Proc. Inst. Radio Eng. N.Y., 37 (1949).
- [23] Sweeney and Vest, Reconstruction of three-dimensional refractive index fields from multidimensional interferometric data. Appl. Opt. 12 (1973), 2649-2664.
- [24] Szegő, Orthogonal polynomials, AMS Colloquium Publications. Vol. XXIII, New York (1959).
- [25] Whittaker, Proc. Roy. Soc. Edinburgh, 35 (1915).
- [26] Zwick and Zeitler, Image reconstruction from projections, Optik. 38 (1973), 550-565.

A more extensive list of references may be found in [9].

Additional list of references.

- [1] Berry, M.V. and D.F. Gibbs, The interpretation of optical projections, Proc. Soc. Lond. A 314 (1970), 143-152.
- [2] Buddinger, T.F. and G.T. Gullberg, Three-dimensional reconstruction in nuclear medicine emission imaging, IEEE Trans. Nucl. Sci. NS-21 (1974), 2-20.
- [3] Cho, Z.H., I. Ahn, C. Bohm and G. Huth, Computerized image reconstruction methods with multiple photon/X-ray transmission scanning, Phys. Med. Biol. 19, 4 (1974), 511-522.
- [4] Cormack, A.M., Representation of a function by its line integrals, with some radiological applications, J. Appl. Phys. 34, 9 (1963), 2722-2727.
- [5] Cormack, A.M., Representation of a function by its line integrals, with some radiological applications II, J. Appl. Phys. 35, 10 (1964), 2908-2913.
- [6] Cormack, A.M., Reconstruction of densities from their projections, with applications in radiological physics, Phys. Med. Biol. 18, 2 (1973), 195-207.
- [7] Crowther, R.A., L.A. Amos, J.T. Finch, D.J. DeRosier and A. Klug, Three-dimensional reconstructions of spherical viruses by Fourier synthesis from electron micrographs, Nature 226, (1970), 421-425.
- [8] Crowther, R.A. and A. Klug, ART and Science or conditions for three-dimensional reconstruction from electron microscope images, J. Theor. Biol. 32 (1971), 199-203.
- [9] DeRosier, D.J. and A. Klug, Reconstruction of three-dimensional structures from electron micrographs, Nature 217 (1968), 130-134.

- [10] Ein-Gal, M., The shadow-transform: an approach to cross-sectional imaging, Stanford University, Information Systems Laboratory, Technical Report No. 6851-1 (1974).
- [11] Gilbert, P., Iterative methods for the three-dimensional reconstruction of an object from projections, *J. Theor. Biol.* 36 (1972), 105-117.
- [12] Gordon, R. and G.T. Herman, Reconstruction of pictures from their projections, *CACM* 14, 12 (1971), 759-768.
- [13] Gordon, R., A tutorial on ART, *IEEE Trans. Nucl. Sci.* NS-21 (1974), 78-93.
- [14] Herman, G.T., A.V. Lakshminarayanan and S.W. Rowland, The reconstruction of objects from shadowgraphs with high contrasts, State University of New York, Dept. of Comp. Science, Buffalo N.Y., Techn. Report 84 (1974).
- [15] Logan, B.F. and L.A. Shepp, Optimal reconstruction of a function from its projections, *Duke Math. J.* 42, 4 (1975), 645-659.
- [16] Mersereau, R.M., Recovering multidimensional signals from their projections, *Computer graphics and image processing* 1 (1973), 179-195.
- [17] Shepp, L.A. and B.F. Logan, The Fourier reconstruction of a head section, *IEEE Trans. Nucl. Sci.* NS-21 (1974), 21-37.
- [18] Tewarson, R.P., Solution of linear equations in remote sensing and picture reconstruction, *Computing* 10 (1972), 221-230.
- [19] Guenther, R.B., C.W. Kerber, E.K. Killian, K.T. Smith and S.L. Wagner, Reconstruction of objects from radiographs and the location of brain tumors, *Proc. Nat. Acad. Sci. USA* 71 (1974), 4884-4886.