BULLETIN OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 83, Number 6, November 1977

PRACTICAL AND MATHEMATICAL ASPECTS OF THE PROBLEM OF RECONSTRUCTING OBJECTS FROM RADIOGRAPHS[1]

BY KENNAN T. SMITH, DONALD C. SOLMON, AND SHELDON L. WAGNER

Table of Contents

CHAPTER I. PRACTICAL ASPECT

1. Introduction. Having done pure mathematics for a good many years, I became interested about three years ago in practical problems in medical radiology–specifically in the problem of finding tumors, hemorrhages, and other lesions of the brain with ordinary hospital equipment and without the introduction of contrast material.

This is a significant and unsolved medical problem. The soft tissue density differences within the brain are so small that, masked by the heavy and variable skull, they are invisible on ordinary radiographs. The traditional procedures of radiology involve the injection either of air or of an x-ray opaque dye. Both are painful and dangerous, and the information obtained is often meager. In fact, very sick patients are unable to support the tests. On the other hand, the last few years have seen remarkable advances with the advent of the EMI scanner and its descendants. The scanners take a large number of radiographs around a semicircle and use the computer (in a way that is described in §2) to produce cross sections of the density function. The information obtained in this way is far more accurate and complete than that from the traditional procedures, and it is obtained either without contrast

An invited address delivered at the Far West Sectional Meeting of the American Mathematical Society, Monterey, California, April 19, 1975; received by the editors February 20, 1976.

AMS (MOS) subject classifications (1970). Primary 92A05, 78A55; Secondary 44A15.

[1]This project has been supported by the National Science Foundation under grant no. DCR71-03758 A01.

material or with a relatively noninvasive use of it. The disadvantage of the scanners is that they are so expensive and elaborate as to be out of the range of the ordinary small hospital. Moreover, they are completely inflexible, producing cross sections of the density function whether or not this is the best information for a given problem, and sometimes producing them in far from the best way. The objective of our work has been to produce equivalent results, but with ordinary hospital equipment that is available everywhere, and with the flexibility to produce other kinds of results too.

The brain is a particularly good starting point in the study of medical radiology for another reason than its medical significance. With the small density differences, masked by the skull, it offers the maximum theoretical difficulty, while at the same time, having no moving parts, it offers the minimum technical difficulty. The lessons learned from the brain can be expected to be of good use on other parts of the body, and in nonmedical problems as well. Already these lessons are leading to encouraging experiments also described here in the diagnosis of gall stones, breast cancer, and vascular lesions–all by methods other than cross section reconstruction.

These practical problems lead to interesting mathematical problems that require rather sophisticated theorems for their solution. And the mathematical solutions, in return, although they rarely provide computationally useful formulas, error bounds, and the like, do have a direct and important effect. They set the directions from which the practical problem should be approached.

A vast amount of work is being done these days on cross section reconstructions, and it spans many fields:–medicine, molecular biology, neutron radiography, and radio astronomy, to name a few. This article is not intended as a survey of the field. It is a personal account of our own work, and it makes no mention of any other, apart from occasional comparisons of what we can do (or think we can do) with what the EMI scanner can do (or what we think it can do). These comparisons should be regarded as opinions of the authors. The reader who is interested in surveying the field should consult the extensive bibliography which has been prepared and is being kept up to date by Dr. Richard Gordon of the N.I.H. The reader who is particularly interested in the medical side can refer to the original paper of G. N. Hounsfield [7], the inventor of the EMI scanner; the many reports from the Mayo Clinic; and the book of Drs. P. F. J. New and W. R. Scott [12].

In several aspects our work has been joint work with our colleagues R. B. Guenther of the Oregon State mathematics department, C. Hamaker of the University of Oregon mathematics department, C. W. Kerber of the University of Oregon Medical School, and E. K. Killian of the Good Samaritan Hospital in Corvallis. We have received help and helpful advice from J. P. Kelley of the Oregon State x-ray department and C. Klopfenstein of the University of Oregon chemistry department.

Since this article is an account of my lecture at Monterey, I may take the liberty of offering occasional first person comments–for which my co-authors should not be judged too harshly.

This article is dedicated to my long time teacher, colleague, and friend Nachman Aronszajn on the occasion of his seventieth birthday.

KTS

2. A reconstruction method. For radiographic purposes an object in R^n is determined by its density function f, $f(x)$ being the density at the point x. An x-ray from a direction θ (a point on the sphere S^{n-1}) provides a function $P_\theta f$ on the plane orthogonal to θ whose value at a point z of this plane is the total mass along the line through z in the direction θ:

$$Pf(\theta, z) = P_\theta f(z) = \int_{-\infty}^{\infty} f(z + t\theta)\, dt \quad \text{for } z \in \theta^\perp. \tag{2.1}$$

The reconstruction problem is the problem of recovering the unknown density f from a knowledge of certain of the radiographs[2] $P_{\theta_1} f, \ldots, P_{\theta_M} f$.

Throughout the article it will be assumed that an object is identified with its density function and that the latter is square integrable with compact support, unless the contrary is explicitly stated.

Of course real objects exist in the 3 dimensional space R^3. Note, however, that a solution to the 2 dimensional problem provides immediately a solution to the 3 dimensional problem by recovering all 2 dimensional sections orthogonal to a fixed line. At present almost all 3 dimensional reconstructions are being obtained in this way for reasons of computational simplicity. Thus even the practical situation requires the study of both the 2 and 3 dimensional cases. It turns out that most of the formulas and results remain the same for all dimensions, or at least follow a pattern, so that both cases can be covered at once by allowing the general dimension n. What is more important to a mathematician, however, is that some very interesting distinctions do arise between the dimensions.

Reconstruction methods can easily be devised. The one we are using is a modification of a simple iterative scheme of Kacmarz [3], which is as follows: Suppose that N_j is the null space of P_{θ_j}, that f is the true solution, that P_j is the orthogonal projection (in L^2) on $f + N_j$, and that $P = P_M \cdots P_1$. The method consists in choosing an initial guess g and setting $f_m = P^m g$. The P_j are computable, and it is shown in [1] (by an elegant, and highly recommended argument) that if $N = \bigcap N_j$, then the f_m converge in the L^2 sense to the projection of g on $f + N$. This Kacmarz method is effectively the one used also (at least initially) by the EMI scanner, and it has been rediscovered and used by others too [4]. We have established the following rate of convergence.

THEOREM 2.2. *If each N_j makes an angle $\geq \alpha_j$ with the intersection of the following ones and g_0 is the projection of g on $f + N$, then*

$$\|f_m - g_0\|^2 \leq c^m \|g - g_0\|^2 \quad \text{with} \quad c \leq 1 - \prod_{j=1}^{M-1} \sin^2 \alpha_j \tag{2.3}$$

PROOF. This is a theorem in abstract Hilbert space, the N_j being arbitrary closed subspaces subject to the angle condition. (Recall that the angle between two subspaces U and V of a Hilbert space is $\geq \alpha$ if $|\langle u, v\rangle| \leq \cos \alpha$ for all unit vectors u and v in U and V respectively and orthogonal to $U \cap V$.)

There is no loss of generality in assuming that $f = 0$, for this amounts to making a translation in the Hilbert space, and it is almost obvious that the theorem is invariant under translation. We do this in order to obtain projec-

[2] In the parlance of radiography the x-ray is the photon beam and the radiograph is the picture it produces.

tions that are linear. Since g_0 lies in N and P is the identity on N, the inequality to be proved is

$$\|P^m(g - g_0)\|^2 \leqslant \left(1 - \prod_{j=1}^{M-1} \sin^2\alpha_j\right)^m \|g - g_0\|^2.$$

Setting $v = g - g_0$ and noting that $N^\perp$ is invariant under P, we see that it is sufficient to show that

$$\|Pv\|^2 \leqslant \left(1 - \prod_{j=1}^{M-1} \sin^2 \alpha_j\right)\|v\|^2 \quad \text{for } v \in N^\perp. \tag{2.4}$$

This will be proved by induction on M, the case $M = 1$ being self evident as both sides are 0.

Let $N' = N_M \cap \cdots \cap N_2$ and $P' = P_M \cdots P_2$. For any $v \in N^\perp$ write $v = w + v_1$ with $w \in N_1$ and $v_1 \in N_1^\perp$, so that $Pv = P'w$. Now write $w = w' + w''$ with $w' \in N'$ and w'' in $N'^\perp$. Then $P'w = w' + P'w''$, and the summands are orthogonal, so that

$$\|P'w\|^2 = \|w'\|^2 + \|P'w''\|^2.$$

Induction gives

$$\|P'w''\|^2 \leqslant \left(1 - \prod_{j=2}^{M-1} \sin^2 \alpha_j\right)\|w''\|^2,$$

and these last two formulas give

$$\|P'w\|^2 \leqslant \left(1 - \prod_{j=2}^{M-1} \sin^2 \alpha_j\right)\|w\|^2 + \prod_{j=2}^{M-1} \sin^2 \alpha_j \|w'\|^2.$$

Now, w lies in N_1 and is orthogonal to $N = N_1 \cap N'$, and w' lies in N' and is orthogonal to N. Since the angle between N_1 and N' is at least α_1 it follows that $\|w'\|^2 = \langle w, w'\rangle \leqslant \cos\alpha_1\|w'\|\,\|w\|$ and hence that $\|w'\| \leqslant \cos\alpha_1\|w\|$. Combining this with the last formula and using the facts that $Pv = P'w$ and $\|w\| \leqslant \|v\|$, we get the required formula (2.4).

REMARK. In our reconstructions we have been using 18 x-ray directions ($M = 18$) and 10 iterations ($m = 10$) in which case the running time on the CDC 3300 is about 40 seconds. D. C. Solmon and C. Hamaker [6] have computed the angles α_j and have found that

$$\prod_{j=1}^{17} \sin^2 \alpha_j \approx 2.957 \times 10^{-5}$$

(for a judicious ordering of the projections P_i). However, by applying Theorem (2.2) to a sequence of finite dimensional subspaces invariant under P, they have shown that $c < 2/3$ (where c is the constant in 2.3). Thus the theory states that 12 iterations are sufficient to obtain a relative error of 1%, which is in agreement with the 10 iterations we have been using in practice.

The results of our first significant experiments, one on a pig's head and the other on a brain tumor patient at the University of Oregon Medical School Hospital are shown in Figures 1 and 3 below.

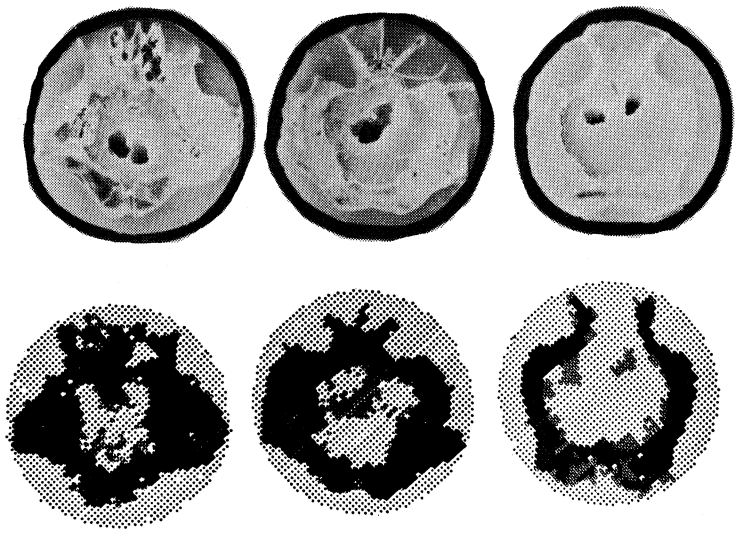


FIGURE 1. Reconstruction of pig's head. The pig's head was a phantom consisting of a real skull about 10 cm in diameter filled with tissue equivalent wax (49% beeswax, 49% paraffin, and 2% resin) in which there were two water filled holes about 0.8 cm in diameter. In the *upper row* are actual photographs of cross sections of the pig's head. In the *lower row* are reconstructions of the same cross sections.

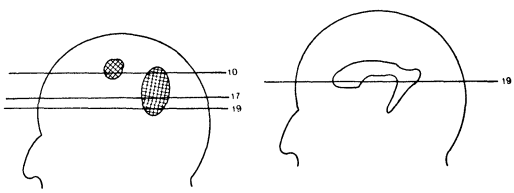


FIGURE 2. Location of brain tumors (as demonstrated by angiography) in a patient. The profile sketches of the patient indicate the levels of the cross sections shown in Figure 3. On the left profile are indicated a small tumor on the left and a large tumor on the right. On the right profile are indicated the ventricles. The numbers correspond to the cross sections shown in Figure 3.

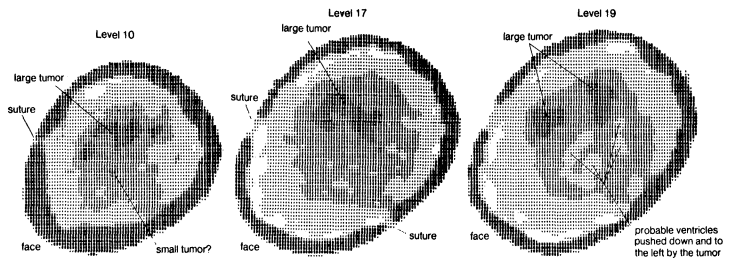


FIGURE 3. Horizontal cross sections of the brain tumors sketched in Figure 2. The metastatic tumors are the dark masses in the interior. Some apparently mysterious features, e.g., the fact that the skull is thicker on the left side, are due to the fact that the head was tipped to the left during the x-rays. Therefore, any horizontal cross section cuts the left side higher than the right. The numbers correspond to the horizontal sections indicated in Figure 2.

The result of a more recent experiment is shown in Figures 4, 5, and 6. This experiment involved a completely known phantom (or model) similar to one being used by the Mayo Clinic to check the uniformity of EMI scanners. It is a plexiglass cylinder about 3″ in diameter wrapped with a teflon tape skull about 1/8″ thick. Inside are two lexan pins of diameters 1/4″ and 1/8″, a water filled hole of diameter about 3/16″ and a *Y* shaped water filled channel of diameter about 3/16″. The lexan pins are smooth and round, but the water filled hole and channel are ragged. In traditional radiography plexiglass, lexan, and water are considered completely equivalent. (According to T. F. Budinger the physical densities are as follows: teflon 2.16, lexan 1.20, plexiglass 1.19, and water 1.00.) A photograph of the phantom is shown in Figure 4, an EMI reconstruction in Figure 5, and our own reconstruction in Figures 6a and 6b. The numbers in Figure 6b are the actual numbers that came straight from the reconstruction program without a posteriori smoothing of any kind with the following inessential modifications: 468 was subtracted from each number in the phantom so that the more interesting ones would stand out as one digit numbers; and any number less than −9 was replaced by −9 so as to stay in the two digit range. (The latter occur only in the teflon skull and on the fringes. Although teflon is denser than the other materials, both the EMI reconstruction and our own showed it to be less dense. Either teflon tape is different from teflon, or else air trapped in the wrapping reduces the density.)

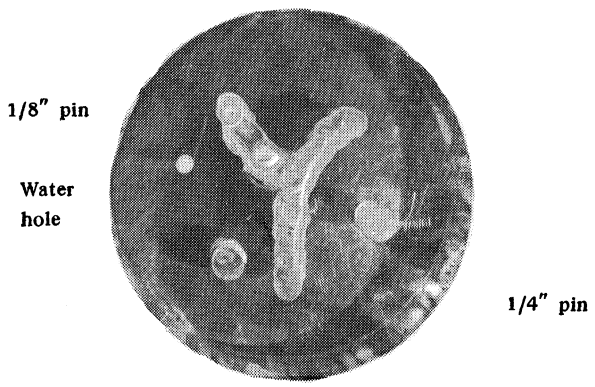


Figure 4. Photo of phantom top view

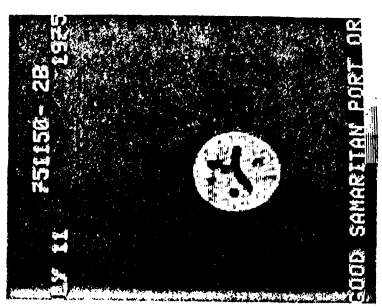


Figure 5. EMI Scan

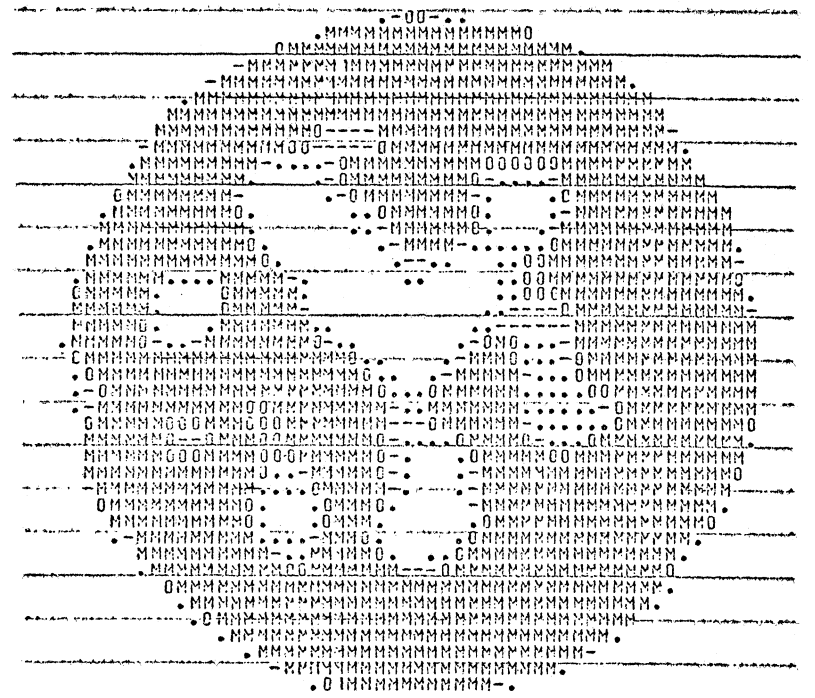


Figure 6a. Reconstruction of Phantom

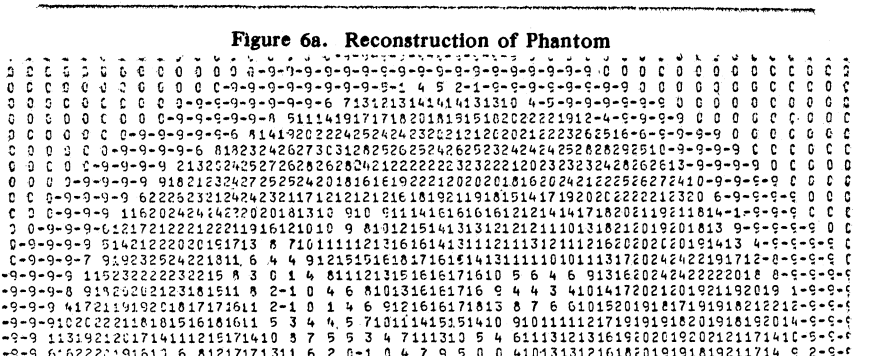


Figure 6b. Reconstruction of Phantom – density numbers

3. Some simple formulas. It is immediately checked that if ρ is a function of one variable, then

$$\int_{\theta^\perp} P_\theta f(x)\rho(\langle x, \xi\rangle)\, dx = \int_{R^n} f(y)\rho(\langle y, \xi\rangle)\, dy \quad \text{for } \xi \in \theta^\perp. \tag{3.1}$$

This formula has two interesting consequences. The first arises on taking $\rho(t) = t^m$, in which case the left side is a homogeneous polynomial in ξ defined on $\theta^\perp$ and depending on both f and θ, while the right side is a homogeneous polynomial in ξ defined on R^n and depending only on f. This means that the functions in the range of the x-ray transform must satisfy certain consistency conditions. D. Ludwig [9] (for $n = 2$) and D. C. Solmon [15] (for $n > 2$) have proved the following theorem. See §13 below for a precise statement and proof of Ludwig's theorem, which is the more difficult of the two.

THEOREM 3.2. *Apart from obvious symmetry, support, and regularity conditions a function $g(\theta, x)$ defined for $\theta \in S^{n-1}$ and $x \in \theta^\perp$ lies in the range of the x-ray transform if and only if for each m the polynomials*

$$p_{m,\theta}(\xi) = \int_{\theta^\perp} g(\theta, x)\langle x, \xi\rangle^m\, dx \quad \text{for } \xi \in \theta^\perp \tag{3.3}$$

fit together to form a polynomial on R^n, i.e. $p_{m,\theta}(\xi) = p_m(\xi)$ for $\xi \in \theta^\perp$.

REMARK 3.4. Just what can be expected in the case of heads I do not know, but I read recently that in the case of chest x-rays it can be expected that 80–90% of the photons reaching the film are noise (scatter)–and of course there is also additional noise coming from the developing process, unevenness in the film emulsion and so on. When one takes a certain number M of x-rays there is little reason to believe that the resulting pictures $g(\theta_1, x), \ldots, g(\theta_M, x)$ satisfy the consistency conditions of the theorem, in which case they are not the true radiographs of any object at all and the iterations described in the last paragraph simply bounce around. The situation is not quite this bad, however, for it can be shown that a finite number of radiographs read at a finite number of points cannot lead to inconsistencies. See §15.

A trivial use of Theorem 3.2, but one that actually has been important in practice is the following. It is clear that the several radiographs must be radiographs of precisely the same object. Under laboratory conditions the object can be held perfectly still, but under hospital conditions the patient cannot. Despite careful precautions the head bobs up and down and around and about to a surprising degree, and the films lie in surprising positions in the cassettes. The problem then is to match the radiographs that result. The theorem shows ($m = 0, 1$) that this can be achieved by matching the centers of gravity.

A second useful consequence of formula (3.1) is the Fourier transform formula that results from taking $\rho(t) = e^{-it}$. Apart from powers of 2π, the left side becomes the Fourier transform of $P_\theta f$ and the right side the Fourier transform of f. When the powers of 2π are accounted for, the result is as follows.

$$(P_\theta f)\hat{}(\xi) = (2\pi)^{1/2}\hat{f}(\xi) \quad \text{for } \xi \in \theta^\perp, \tag{3.5}$$

where

$$\hat{f}(\xi) = (2\pi)^{-n/2}\int e^{-i\langle x,\xi\rangle} f(x)\, dx.$$

REMARK 3.6. Another reconstruction method in widespread current use, usually called the "Fourier Method", is based on formula (3.5). This formula shows that when the radiographs of the object f are known from the directions $\theta_1, \ldots, \theta_M$, then the Fourier transform $\hat{f}$ is known on the lines $\theta_1^\perp, \ldots, \theta_M^\perp$. Then $\hat{f}$ is defined everywhere by an interpolation, and the reconstruction is its inverse Fourier transform.

In the literature, the method described in the last section, which I prefer to call the Kacmarz method, usually is referred to as the ART method, standing for Algebraic Reconstruction Technique.

There is one more method in widespread current use called the "Convolution Method". This is simply a numerical implementation of Radon's original formula for the inversion of the Radon transform. Radon's formula and its proof are given in §§9 and 12.

4. The determination of an object by x-rays–and the lack of it. One of the main points I want to discuss in this article is the extent to which x-rays determine an object. This seems to be widely misunderstood, and the misunderstandings seem to have led to effort in unprofitable directions.

From the Fourier transform formula (3.5) it is clear that an object is determined by its radiographs, for the Fourier transform is determined. Since, however, the objects come from a special class of functions (square integrable with compact support), much more can be said. The Fourier transforms come from a special class of real analytic functions on R^n, and it is well known that real analytic functions cannot vanish on an infinite set of $n-1$ dimensional planes through the origin without vanishing identically. Therefore:

THEOREM 4.1. *An object is determined by any infinite set of radiographs.*

Although this result is reassuring, especially in the problem of making reconstructions from electron micrographs, where the directions must be confined to a relatively narrow cone, the practical question is the extent to which an object is determined by a finite set of radiographs. It has been largely assumed that incorrect objects with many correct radiographs, although they do exist, must be so bizarre as to be unreachable by reconstruction methods, or at least as to be easily rejected by any reasonable interpreter. The fact of the matter is this:

THEOREM 4.2. *A finite set of radiographs tells nothing at all.*

For some reason this theorem provokes merriment. It is so plainly one of those mathematical ideals untainted by any possibility of practical application. The precise statement of the theorem will become clear from the simple proof, and in any event will be given explicitly in Theorem 4.3.

Given a finite number of directions $\theta_1, \ldots, \theta_M$, choose a polynomial q that vanishes on each plane $\theta_j^\perp$, for example

$$q(\xi) = \langle \xi, \theta_1 \rangle \cdots \langle \xi, \theta_M \rangle,$$

and let Q be the corresponding differential operator obtained by replacing ξ_j by $-i\partial/\partial x_j$. For any $g \in C_0^\infty$ (infinitely differentiable with compact support) set $f = Qg$, and consider $P_\theta f$. By formula (3.5) we have

$$(P_\theta f)\hat{}(\xi) = (2\pi)^{1/2} \hat{f}(\xi) = (2\pi)^{1/2} q(\xi) \hat{g}(\xi) \quad \text{for } \xi \in \theta^\perp,$$

thus,

$$P_{\theta_j} f = 0 \quad \text{for } j = 1, \ldots, M.$$

This shows that there are plenty of objects with zero radiographs from the given directions. Indeed, if g_0 is any given function in C_0^∞, the theorem of B. Malgrange [10] on the solvability of linear partial differential equations with constant coefficients can be used to solve the equation $Qg_1 = g_0$. Now, g_1 is not quite suitable as the g above, for it cannot be expected to have compact support. However, if K is any compact set in the interior of the support of g_0, then $\varphi \in C_0^\infty$ can be chosen so as to be 1 on a neighborhood of K and 0 outside a slightly larger neighborhood which is still within the support of g_0. Now take $g = \varphi g_1$, i.e. $f = Q(\varphi g_1)$. Then

(a) $P_{\theta_j} f = 0$ *for* $j = 1, \ldots, M$,

(b) $f = g_0$ *on* K *and outside the support of* g_0.

In other words, f has zero radiographs from the given directions, and yet is equal to the completely arbitrary g_0 except on an arbitrarily small ring just inside the boundary of the support of g_0. A slight modification of this argument gives the following result, which is the precise statement of Theorem 4.2.

THEOREM 4.3. *Suppose given an infinitely differentiable object f_0 and a finite number of directions. Then there is a new infinitely differentiable object f with exactly the same shape, exactly the same radiographs from these directions, and completely arbitrary on any compact set in the interior of the support of f_0.*

Before turning to some illustrations of this theorem and to some of the implications, let us consider first another theorem that seems to point back in the opposite direction. Although a reconstruction method necessarily uses only a finite number of x-rays, it also necessarily produces only a finite dimensional space of possible reconstructions. The question arises as to whether the radiographs serve to distinguish between the objects within this finite dimensional space.

THEOREM 4.4. *For almost any finite dimensional space F the objects in F can be distinguished by a single radiograph from almost any direction.*

The precise statement is as follows.

THEOREM 4.5. *Let F be a finite dimensional space of objects with basis $f_1, \ldots, f_N$, and let V be the set of directions θ such that at least two objects in F have the same radiograph from the direction θ. Then*

(a) *V is an algebraic variety in S^{n-1};*

(b) *$V = S^{n-1}$ if and only if there are polynomials $q_1, \ldots, q_N$ such that*

$$\sum_{j=1}^{N} q_j(\theta) P_\theta f_j(\xi) = 0 \quad \textit{for all } \theta \textit{ and all } \xi \in \theta^\perp. \tag{4.6}$$

Part(a) says that either no single direction serves to distinguish between objects in F, or else almost any direction does that job; while part (b) says that the latter is the usual case. In our reconstructions the f_j are step functions on a fixed grid in R^2. This example is worked out at the end of the proof of Theorem 4.13.

What will be proved is the following slightly more general result that comes from Fourier transforming the theorem.

THEOREM 4.7. *Let $f_1, \ldots, f_N$ be real analytic functions on R^n, and let V be the set of directions θ such that their restrictions to the plane $\theta^\perp$ are linearly dependent. Then*

(a) *V is an algebraic variety in S^{n-1};*

(b) *$V = S^{n-1}$ if and only if there are polynomials $q_1, \ldots, q_N$ such that*

$$\sum_{j=1}^{N} q_j(\theta) f_j(\xi - \langle \xi, \theta \rangle \theta) = 0 \quad \textit{for all } \theta \textit{ and all } \xi. \tag{4.8}$$

PROOF. A fixed point θ lies in V if and only if there are numbers $q_j(\theta)$ satisfying (4.8), hence if and only if each coefficient in the Taylor expansion of (4.8) is 0. Now, if

$$g(\xi) = f(\xi - \langle \xi, \theta \rangle \theta)$$

and

$$A_\theta = D - \langle D, \theta \rangle \theta \quad \textit{with } D = (D_1, \ldots, D_n) = (\partial/\partial x_1, \ldots, \partial/\partial x_n)$$

then $Dg(\xi) = A_\theta f(\xi - \langle \xi, \theta \rangle \theta)$. Thus, θ lies in V if and only if there is a solution $q_1(\theta), \ldots, q_N(\theta)$ to the system of equations

$$\sum_{j=1}^{N} q_j(\theta) A_\theta^\alpha f_j(0) = 0 \quad \textit{for every } \alpha, \tag{4.9}$$

which happens if and only if the coefficient matrix $\{A_\theta^\alpha f_j(0)\}$ has rank less than N. Consequently, V is the set of common zeros of the $N \times N$ minors, and the latter are clearly polynomials in θ, so V is an algebraic variety.

Let r be the maximum rank of the above matrix at a point of V, and rearrange the j's so that $\{A_\theta^\alpha f_j\}, j = 1, \ldots, r$, has rank r at some point of V. Consider the matrix (at 0)

$$\begin{matrix} A_\theta^{\alpha_1} f_1 & A_\theta^{\alpha_1} f_2 & \cdots & A_\theta^{\alpha_1} f_{r+1} \\ \vdots & \vdots & & \vdots \\ A_\theta^{\alpha_r} f_1 & A_\theta^{\alpha_r} f_2 & \cdots & A_\theta^{\alpha_r} f_{r+1} \\ A_\theta^{\alpha} f_1 & A_\theta^{\alpha} f_2 & \cdots & A_\theta^{\alpha} f_{r+1}. \end{matrix}$$

The determinant is 0 on V, and expansion via the last row gives the relation

$$\sum_{j=1}^{N} q_j(\theta) A_\theta^\alpha f_j(0) = 0 \quad \textit{on } V \textit{ for each } \alpha, \tag{4.10}$$

the q_j being 0 for $j > r + 1$, and being the minors in the above matrix, therefore polynomials, for $j \leqslant r + 1$. Hence,

$$\sum_{j=1}^{N} q_j(\theta) f_j(\xi - \langle \xi, \theta \rangle \theta) = 0 \quad \text{for } \theta \text{ on } V \text{ and all } \xi. \tag{4.11}$$

If $V = S^{n-1}$, this is the required relation (4.8). On the other hand, if this relation holds for all $\theta \in S^{n-1}$, then it gives a dependence relation between the f_j for every θ except those in the proper variety W of common zeros of the q_j, and shows that $S^{n-1} = V \cup W$, which is possible only if $S^{n-1} = V$.

While on this order of ideas we give a similar theorem that provides the corresponding information in the case of the Radon transform, which is the subject of much of Chapter II.

THEOREM 4.12. *Let $f_1, \ldots, f_N$ be real analytic functions on R^n, and let V be the set of directions θ such that their restrictions to the line through the origin with direction θ are linearly dependent. Then V is an algebraic variety in S^{n-1}, and $V = S^{n-1}$ if and only if there are homogeneous polynomials $q_1, \ldots, q_N$ of the same degree such that*

$$\sum_{j=1}^{N} q_j f_j = 0. \tag{4.13}$$

The proof is similar to that of Theorem 4.7 except that in place of the Taylor coefficients of the f_j the homogeneous parts of the Taylor expansions are used.

In practical situations the question can often be resolved without recourse to the general theorem. In our case, for example, the situation is as follows. Suppose that all objects to be considered have support in the unit square $|x_1| \leqslant 1$, $|x_2| \leqslant 1$. A number N is chosen (in our case $N = 25$), and arbitrary objects are approximated by step functions on the squares determined by the lines $|x_1| = k/N$ and $|x_2| = l/N$. Thus, a basis of the space F consists of the translates $\chi_j(x) = \chi(x - a_j)$, where χ is the characteristic function of the square $|x_1| \leqslant 1/2N$, $|x_2| \leqslant 1/2N$ and the a_j are the points at the centers of the above squares. The Fourier transforms, the f_j of the last theorem, are therefore the functions

$$f_j(\xi) = e^{-i\langle a_j, \xi \rangle} \hat{\chi}(\xi),$$

and the variety V consists of those θ for which the matrix $\alpha_{jk} = \langle a_j, \theta \rangle^k$ has less than the maximum rank. These are the θ for which the numbers $\langle a_j, \theta \rangle$ are not distinct–hence the θ for which the corresponding line has slope k/l with k and l integers at most $2N$ in absolute value. This particular case is given in R. Mersereau and A. Oppenheim [**11**].

Now, the question is–Which is the practical result, the obviously unrealistic infinite dimensional Theorem 4.3, or the finite dimensional Theorem 4.4? To see what happens in practice consider Figure 7, which shows two reconstructions of the same cross section of the head of another brain tumor patient from the University of Oregon Medical School Hospital. In the first, the tumor (verified by biopsy) is easily visible as the dark mass on the right hand side. In the second, it is merged with the skull (in accordance with Theorem

4.3) in such a way as to make diagnosis very dubious. It is indeed Theroem 4.3 that is in evidence here, not noise, experimental error, or what have you. Medically feasible objects with virtually identical radiographs–yet medically very inequivalent, i.e. one with a tumor and one without–are easy to produce artifically on the computer without noise and without experimental error.

Thus the indeterminacy theorem expresses a fundamental fact that must be faced. The consequences can be reduced by the use of a huge number of x-ray directions, but there ought to be better ways. One possibility is described in the next section.

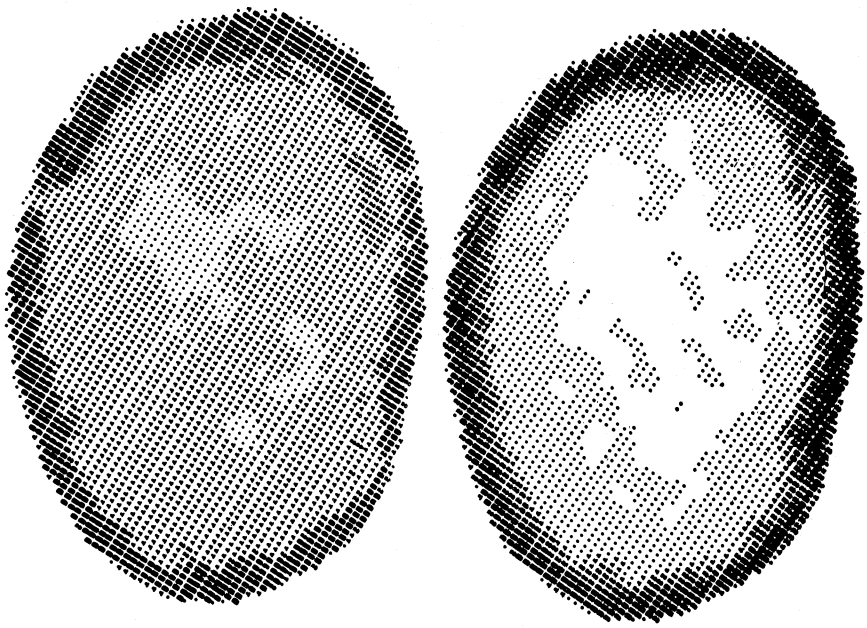


Figure 7

5. Resolution. In this field the term *resolution* refers, in a somewhat imprecise way to the minimum size of details that can be expected to be picked up. Thus, to say that a procedure for diagnosing lesions of the brain has a resolution of 1/8″ would mean that the procedure can be expected to reveal tumors, hemorrhages, etc. which have diameter 1/8″ or better.

There are three main factors governing the resolution of reconstructions from x-rays. The first is the indeterminacy described in Theorem 4.3. The second is the size of the reconstruction matrix relative to the size of the full object being reconstructed. The third is noise. The first two will be discussed in this section and the noise in the next.

It is seen from the actual example in Figure 7 that the indeterminacy is a real problem. In the case of the big scanners there seems to be disagreement as to whether it has been licked or not. Some authorities say that initially it was a problem, but is not one any longer, while others say that there are still difficulties, particularly with lesions just inside the skull.

Any reasonable reconstruction method finds very quickly a solution to the problem of producing an object with the correct, or very nearly correct,

radiographs from the given directions–but it may well find one of the feasible but wrong objects of Theorem 4.3. Ordinarily the wrong objects will be highly oscillatory in the little ring where the compensating errors are made. This is little help, however, in detecting the error, or even that one has been made, for the skull bone itself is highly oscillatory–something like a honeycomb–so that the erroneous oscillations get lost among the ones that belong. What are needed are additional twists to direct the reconstruction procedure toward the right solution instead of one of the wrong ones.

One possibility is to make use of a priori information about the object that does not come from the x-rays. One example of this that is used in many procedures is to set a priori upper and lower bounds for the density of brain tissue and to correct each stage of the iteration so that these bounds are satisfied. Such bounds were used in the wrong reconstruction shown in Figure 7. It appears that work should be done on this question of what kind of a priori information can really be used effectively.

A second possibility, the one that we rely on more and the one that produced the correct reconstruction in Figure 7, is to make use of empirical knowledge of the behavior of the reconstruction method. The idea is simply to determine experimentally what kind of object the procedure handles well, and then to decompose a given problem in stages, each of which involves that kind of object. Experience shows that the kind of object that our procedure handles well is a simple object with relatively good density contrasts in the desired details.

Now consider the head. In regard to what is sought, it certainly is a simple object–consisting of bone, normal tissue, and perhaps a lesion or two. However, the desired details, namely the lesions, certainly do not have good density contrasts. If the lesions are forgotten for a moment, the object becomes still simpler–just bone and tissue–and the density contrast becomes large–bone vs. tissue. It can be expected, therefore, that an initial run of the reconstruction procedure should do a very good job of finding the skull, and indeed it does. In this initial reconstruction the density at each tissue point is replaced by a constant, usually the average density of brain tissue. The computing machine then simulates x-rays of this new object and subtracts them from the original x-rays. If the results were perfect they would be the radiographs of a head without a skull and with zero average density, an object which is simple and in which the density contrasts of the lesions have become relatively large, in short an object which another run of the reconstruction procedure should handle well. Figure 7 shows what is obtained with and without this "removal of the skull". (At the end the skull is replaced.)

Recently we learned from Dr. Paul F. J. New that in the majority of brain scans now being performed in the Massachusetts General Hospital a radiopaque iodine dye is now being injected intravenously prior to the scan. For various reasons blood tends to accumulate at the site of a tumor, and so the dye accumulates there also, providing additional density contrast and thus helping the reconstruction procedure toward the right solution. This intravenous injection is considered safe and relatively noninvasive, as opposed to the arterial injection used in arteriograms which is much less so.

Although such dye injection should help in this way, it can in fact be used

in a much more powerful way akin to the above removal of the skull–provided one is not tied to a fixed machine that can do nothing but produce cross sections. X-rays can be taken before and after injection of the dye, and subtracted. The results are radiographs of nothing but the vascular structure containing the dye–a very simple object with high density contrast which the reconstruction procedure should handle remarkably well. In many cases, however, it will not be necessary to produce cross sections at all, for the regions with dye will stand out sufficiently well on one or two films. This idea is explored further in §8.

Still another problem in which the subtraction idea will be essential is one that we have not tried yet but will be trying shortly. It is the problem of detecting flaws in the fuel assembly of a nuclear reactor. From the point of view of economy it is necessary to replace the fuel pins as seldom as possible, while from the point of view of safety it is necessary to replace them as soon as they become defective. At present there is no way to detect flaws without dismantling the whole thing. Of course an x-ray beam does not penetrate such an object, but a neutron beam does, and this opens the possibility of cross section reconstructions.

The fuel assembly consists of 217 fuel pins in a hexagonal case about 5″ in diameter. Here we have something just about as different as possible from the head. The 217 pins make a very complicated object, while the flaws, consisting of cracks, bulges, bends, etc., provide large density contrasts–essentially fuel or casing vs. air. Such an object has no chance of being reconstructed faithfully by a direct application of our procedure. However, if two radiographs are taken, one of the assembly to be tested and one of a perfect assembly, and the two are subtracted, the result will be a radiograph of nothing but the defects. One would certainly hope that these will be few enough in number to provide the necessary simple object.

Now we turn to the second factor governing the resolution of reconstructions–the size of the reconstruction matrix relative to the size of the object being reconstructed. As mentioned earlier, the reconstructed density function is a step function on a fixed grid, so it can be thought of as a matrix. It is perfectly obvious, but does not always seem to be appreciated, that a detail in the object cannot possibly be revealed properly if it is smaller than the area represented by a single square of the grid. In fact, in practice it turns out that single squares are not reliable, while a group of about four adjacent ones usually is.

To see the practical consequence of this, consider our plexiglass phantom in Figure 4. The diameter is about 3″ and the reconstruction matrix is 50×50, so each square of the matrix represents a square in the phantom of side $3''/50$. Thus, details in the phantom should be picked up if they have diameter at least $2 \times 3''/50 \approx 1/8''$, and indeed the $1/8''$ lexan pin is picked up.

The $1/8''$ lexan pin is also picked up in the EMI scan shown in Figure 5, but in a similar experiment a year or so ago it was not. The reason is as follows. Until very recently it has been part of the EMI procedure to insert the head in a water bath of fixed size. Many reasons were advanced for doing this, but the main one was to produce an effectively round object of fixed size

so that the same x-ray exposure could be used for all x-ray directions and all patients. The big disadvantage is that the object being reconstructed becomes the actual object plus the water bath. Thus the 3″ plexiglass phantom filled only a small portion of the water bath and hence of the EMI reconstruction matrix. At the time that the 1/8″ pin was not found an 80×80 reconstruction matrix was being used, and it simply was not big enough, while at the time of the scan shown in Figure 5 this had been increased to 160×160. Since the diameter of the water bath is about 10.6″, each square in the new matrix represents a square in the object of side $10.6''/160 \approx 1/16''$, and details should be picked up if they have diameter at least 1/8″, as indeed the 1/8″ pin was– in the new matrix.

This is another illustration of the disadvantage of a fixed machine that is completely set in its ways. Our very first experiment, for example, was done on a tooth x-ray in the dentist's office. Since the tooth could be made to fill the full 50×50 reconstruction matrix, it was possible to pick up the tiny root canals. Our solution to the problem of not having round objects is explained in the next section.

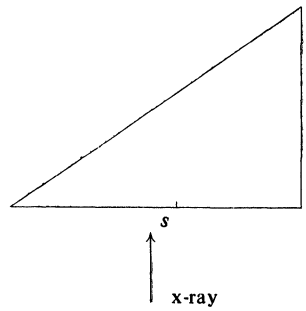
6. Calibration. Up to this point we have been pretending that the result of an x-ray from the direction θ is the total mass function $P_\theta f$ defined on the plane $\theta^\perp$. Of course the direct result is not this function, but an exposed photographic film in which the gray level at a point x depends on the total mass of the object f along the line through x perpendicular to the film. The gray levels on the film are converted to numbers by a *densitometer* which shines a small spot of light on the film (in the case of the plexiglass phantom the spot was a rectangle effectively .1 mm. in width and .6 mm. in height) and records the number

$$D(x) = \log \frac{\text{no. of light photons hitting spot}}{\text{no. of light photons transmitted}}, \tag{6.1}$$

which is called the optical density of the film at the point x.

There is a simple formula relating the number $D(x)$ to the desired number $P_\theta f(x)$, but the simplicity is something of an illusion for the formula contains constants that depend on the kind of film that is being used and on the x-ray exposure. One can always stick to the same kind of film, but hospital equipment is incapable of producing uniform exposures with the accuracy required for this kind of work. What kind of accuracy is needed can be seen from the plexiglass phantom. For x-ray purposes the density difference between plexiglass and lexan is about 1.5%. An average line across the phantom and through the 1/8″ lexan pin has length about 2″, so the pin occupies about 1/16 of it. Therefore, the difference between the total mass along a line through the pin and along a nearby line that misses the pin is about 1/1000. In other words, about 1000 gray levels are needed to do the problem. On the other hand, supposedly identical x-ray exposures of the same object produce densitometer readings that vary by several per cent.

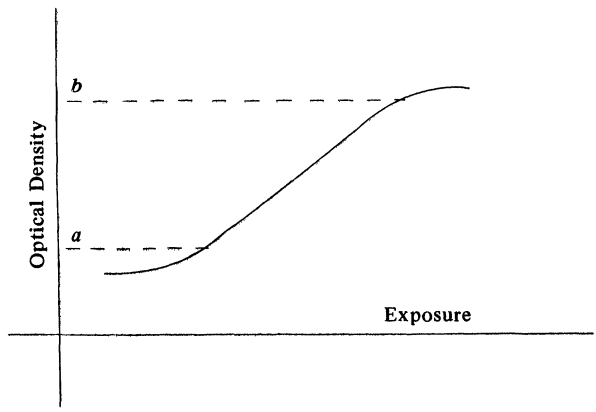
For this reason it was decided at the beginning to pay no attention to the "simple" formula relating $D(x)$ and $P_\theta f(x)$, but rather to calibrate empirically each individual film. An aluminum wedge is included in each x-ray shot.



Let $w(s)$ denote the densitometer reading at the point on the film at distance s from the thin end of the wedge. If the density of the wedge were 1 and the angle 45°, then s would be the true total mass corresponding to the densitometer reading $w(s)$, and hence for any densitometer reading D, $w^{-1}(D)$ would be the corresponding true total mass. A different density and different wedge angle simply introduce an irrelevant proportionality factor. Thus, $w^{-1}(D(x))$ is a function that is proportional to the desired $P_\theta f(x)$, the proportionality factor depending on the wedge density and angle, but not on anything else.

While the calibration of each individual film was introduced initially to counteract the variability in supposedly identical x-ray exposures, it turns out to do much more than that. In particular, it permits intentional variation of the x-ray exposure.

The *characteristic curve* of a photographic film is the graph of the optical density of the film as a function of the exposure received by the film. The characteristic curve of a typical medical film looks something like this.



Within the optical density range from a to b (which depend on the film and intensifying screen in use) the film is sensitive to small changes in exposure, and hence to small changes in total mass in the object. Outside this range the film becomes very insensitive. Consequently it is necessary to choose the exposure so that the relevant areas of the film come out in the sensitive range.

Now consider a head. It is much thicker from front to back than from side to side. An exposure that puts the front view within the sensitive range will make the side view much too dark, and one that puts the side view within the sensitive range will make the front view much too light. The individual

calibration of each film allows the use of different exposures from the different directions so that all films can be kept within the sensitive range.

The wedge also has other functions, such as to cancel out variations in the development process, but the above are the main ones. It will be seen in §8 that in some problems the wedge is not needed, but in reconstructions it is.

REMARK. In the scanners the attenuation of the x-ray beam usually is measured by the effect on scintillation crystals rather than by the effect on photographic film, so the above discussion is not directly applicable. It appears, however, that the crystals share some of the characteristics of the film and that the basic reason for the use of the water bath was to produce a round head for which the x-ray exposures could be kept the same from all directions. The unwanted accidental variations are cancelled by dividing the beam in two parts in each shot–one passing through the head and the other just through the water bath–and using the difference between the two.

7. Noise. Noise is a general term that refers to anything on the film that does not belong there. There are many sources of noise on x-ray films.

(1) SCATTER. When a photon from the x-ray beam passes through the object two things can happen. It can sail through completely undisturbed, or it can interact with atoms of the object. In the case of an interaction the photon may be stopped, which attenuates the x-ray beam. The amount of attenuation at a given point of the film, as measured by the optical density at the point, gives a measure of the number of interactions that took place along the line through the point and perpendicular to the film, hence of the number of object atoms along this line, and hence of the total mass along the line. Unfortunately, however, the interactions are not this simple. Sometimes they involve, not the stopping of the x-ray photon, but a change in its direction. Other times they involve the release of photons from the object atom. And still other things occur too. These new photons and the old ones with the wrong direction are called scatter. They strike and expose the film at points where it should not be exposed.

(2) QUANTUM MOTTLE. A fairly heavy dose of photons is needed to expose the film, unacceptably heavy in the case of human patients. In most medical x-rays the film is not exposed directly by the x-ray beam, but rather by the light from a fluorescent screen called an intensifying screen. A rather sparse x-ray beam causes the intensifying screen to fluoresce sufficiently to expose the film. The result is an exposure to the patient that is a small fraction of what would be needed with direct exposure, but it is also a badly mottled film on which it is more or less the case that the effect of individual x-ray photons has become visible.

(3) OTHERS. There are many other sources of noise, of which here are a few: finger marks of the technicians, streaks left by dirty rollers in the developing machine, incorrect developing temperature, uneven film emulsion.

Figure 8 shows a graph of the noise along a line on one of the films of the plexiglass phantom. The noise is large relative to the total mass differences in these problems, so something must be done to get rid of it.

Inspection of Figure 8 shows that the nature of the noise at first poses a problem and then provides its solution. Ordinarily one expects to eliminate noise by averaging of some kind. In this case the huge isolated peaks of noise

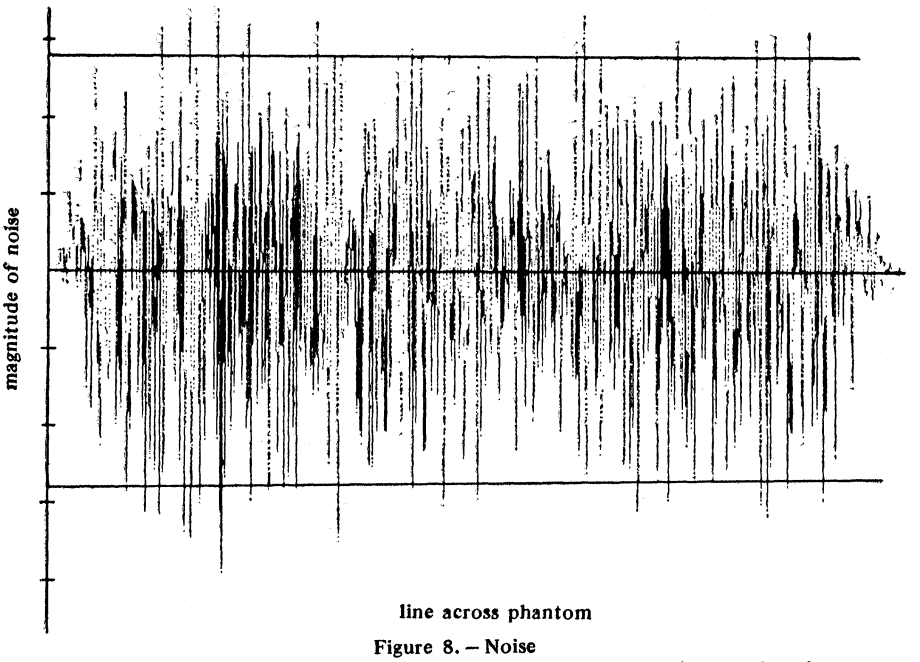


Figure 8. – Noise

provide the problem: if averages are taken over small intervals, then any interval containing a huge peak will contain only one or two and the average will be so distorted that things will end up worse than they began; while if averages are taken over large enough intervals to assure cancellation of the peaks, then all possibility of preserving fine details will be lost.

The source of the problem, the isolation of the largest peaks, the somewhat reduced isolation of the next largest, etc., provides also the answer: to smooth in stages, at each stage correcting only the very worst peaks, and taking advantage of the fact that in these medical problems it is known a priori that there are not huge local oscillations in the total mass.

The first stage of the noise elimination proceeds as follows. Suppose that $D(x)$ is the densitometer reading at the point x along a line on the film. Fix a number L, fairly large at the first stage, and let $A(x)$ be the average of $D(y)$ over the interval of length L centered at x. The noise at this stage is considered to be the difference

$$N(x) = D(x) - A(x)$$

(which is reasonably valid because there are no large local oscillations). Next, fix a number OK. At this stage the point x is considered *ok* and no correction is made if $|N(x)| \leqslant OK$. Finally, fix a number OK'. At the points x that are not *ok*, $D(x)$ is more or less replaced by $A(x)$, but not quite, for the very bad points contribute to $A(x)$. Therefore, define $A'(x)$ to be the average of $D(y)$ over those points y in the interval of length L centered at x which satisfy $|N(y)| \leqslant OK'$ and at these points correct $D(x)$ to be $A'(x)$.

At each stage the procedure is the same, except that $D(x)$ is the outcome of the previous stage instead of the original densitometer readings.

Now, the question is, what are the principles governing the choices of the numbers L, OK, and OK'. First of all, the number L should be chosen to

provide a fairly reliable average–not unduly distorted by the large peaks. Second, the number OK should be chosen so that corrections are made at relatively few points–the number of such points being dependent on L. At the first stage, for example, L will be quite large because of the great height and severe isolation of the worst peaks. The replacement of $D(x)$ by an average over a large interval of length L will not cause the loss of detail if it is done only at very few points, where it would cause the loss of detail if done at many. At later stages L will not be so large, for the worst peaks will be smaller and less isolated. At such stages, when averaging is done over smaller intervals, it can be done at more points without the loss of detail. The basis for choosing OK' remains rather vague, except that it should not be too different from OK. One should not insist that things be too close to a possibly bad average.

In the experiments performed so far five stages have been used. With densitometer readings taken every .1 mm. on the film the values of L were 18, 18, 14, 10, 8. The values of OK varied from one experiment to another (but not from one line to another or one film to another within the experiment). They were chosen so that corrections were made at about 2% of the points in the first two stages, about 4% of the points in the second two, and at all points in the fifth. The values of OK' were chosen so that each interval could be expected to have perhaps 80% of its points OK'. The point to insist on here is not to take these numbers too small, not to try to do too much at any given stage. The procedure is one of "disaster elimination", and, until the end, only the disasters at a given stage should be touched. The results of the noise elimination are shown in Figure 9 for the case of the plexiglass phantom. It is plain from the graphs that most of the noise has been eliminated, and from the quality of the reconstruction (Figure 6) that the essential details have been preserved. Another and quite different example is presented in §8, Figure 10. The mathematical basis of this noise elimination has not been analyzed.

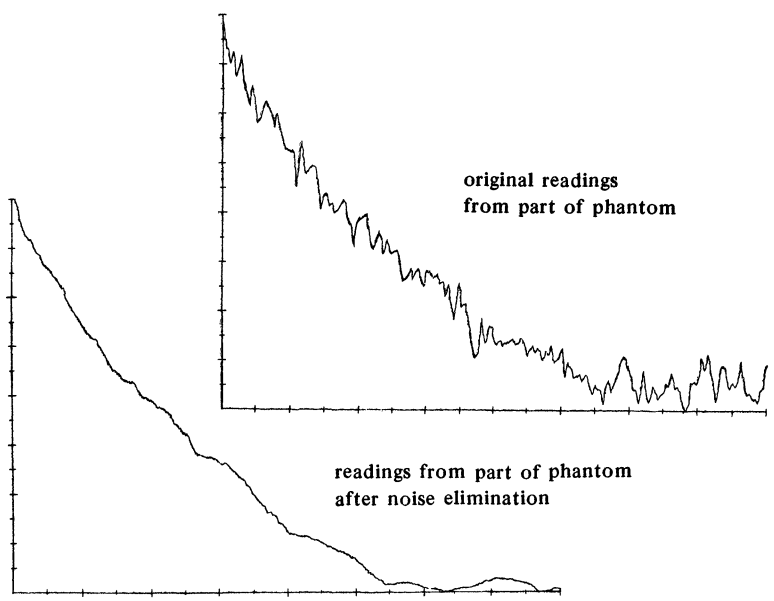


Figure 9

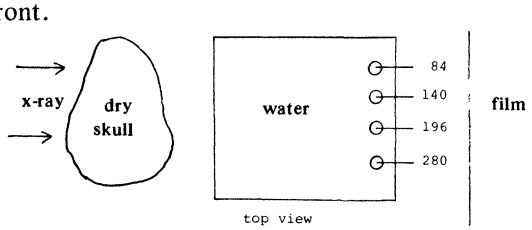
8. Other problems. It is to be expected that work on a very delicate problem such as the revealing of soft tissue distinctions hidden behind a heavy and variable skull ought to furnish along the way the means to do other and simpler problems. In this section I shall discuss a few on which work has already begun.

(1) NONINVASIVE ANGIOGRAMS. Quite often it is important to be able to visualize some part of the vascular structure of the body. One of the traditional brain examinations, the angiogram, has the purpose of visualizing the arteries in the head, as well as certain tumors. Iodine contrast dye is injected into the carotid artery in the neck. The dye goes directly to the head, and then on its way through the body. While the full amount is still in the head an x-ray is taken. At this point the dye is present in sufficient concentration to be visible on the film, though it is not visible later on after spreading through the body. Angiograms of other arteries follow the same principle–direct injections of contrast dye into an artery and an x-ray taken while the dye remains in the neighborhood. This direct arterial injection is painful and dangerous, always calling for hospitalization, and often (perhaps 1% of the time) producing grave complications.

On the other hand, a slow intravenous injection of the dye is considered safe and relatively painless and harmless, though occasional allergic reactions do occur. Hospitalization is rarely necessary. The intravenous injection has been useless in visualizing the arterial structure because there is never a time at which the dye is present in sufficient concentration, but it is commonly used in kidney examinations. The kidneys extract the dye from the blood, and hold it there in high concentrations. Intravenous injection is also used to visualize veins in the way the arterial injection is used to visualize arteries, but, depending upon the vein, this can be a tricky business.

We learned recently from Dr. Paul F. J. New, however, that the stablized intravenous concentrations (200–300 milligrams of iodine per 100 milliliters of saline solution) are sufficient to be detectable in EMI scans–and this opened the exciting possibility of "noninvasive" angiograms. If these concentrations are detectable in cross section reconstructions, then there must be some evidence of their presence on each individual film, even if it cannot be seen. The reason it cannot be seen is simply that it is so meager and is hidden in the skull variations and noise. Why not, therefore, take two films , one before and one after the dye injection, run them through the noise elimination to get rid of the noise and then subtract them to cancel the skull variations?

To test this idea four balloons made from surgical gloves, the first three of diameter 8 mm. and the fourth of diameter 7 mm. (just accident), were filled with iodine contrast dye in concentrations of 84, 140, 196, and 280 milligrams of iodine per 100 milliliters of saline solution. The balloons were placed in a 6″ square plastic milk bottled filled with water, and a real dry skull was placed out in front.



One x-ray was taken, then the balloons were removed and another was taken–the two corresponding to x-rays of a patient taken before and after the injection of dye. Figure 10 shows the graphs of the densitometer readings before and after noise elimination along a line through the balloons on the first film, the corresponding line on the second film, and the difference.

The balloons show up clearly, except for the first one with the very low iodine concentration (which does in fact show up too but at a level not readily distinguishable from the remaining background noise). The different heights reflect the different iodine concentrations. The height of the last one seems at first less than might be expected, but this is due to the smaller diameter which is also apparent on the graph. The fact that the balloon with 84 mg. iodine/100 ml. saline appears slightly, if indecisively, above the background, while the one with 140 mg. iodine/100 ml. saline appears so far above the background, indicates that an 8 mm. balloon with 100 mg. iodine/100 ml. saline should surely be visible, hence also a 4 mm. balloon with 200 mg. iodine/100 ml. saline, etc. Since the latter concentration is still at the bottom of the concentrations found in the blood after intravenous dye injection, it should be possible to visualize blood vessels or tumors of

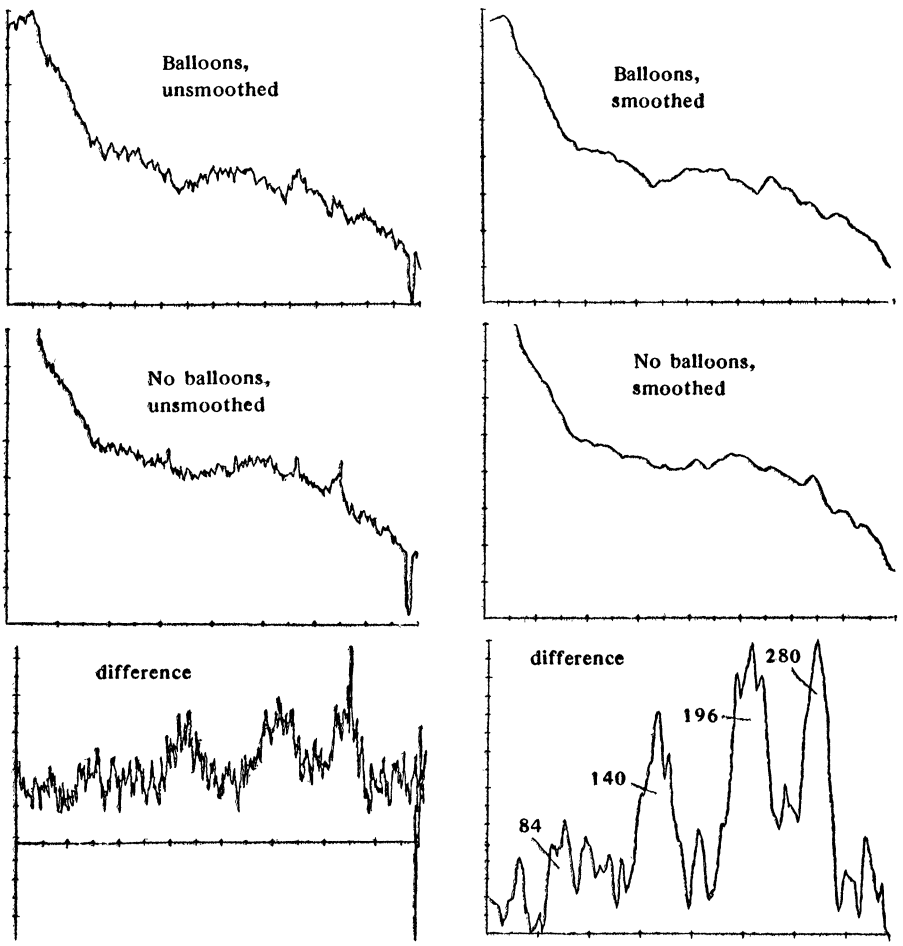


Figure 10

diameter about 4 mm. When many lines are read and a picture is printed back by means of a reverse densitometer (film writer) it should be possible to do much better. For example, film defects that are large enough to be preserved by the noise elimination do not cancel in the subtraction, and they show up glaringly under the magnification and isolation coming from looking at a single line of readings. Note also that the large oscillations in the dry skull make the question of matching the two films accurately much more critical than it will be with other parts of the body.

In this procedure the calibration of §6 is unnecessary because arithmetic is not being done with the numbers in any essential way–though probably it would help a little. Note that in Figure 10 the graph without balloons is perceptibly higher than the one with balloons–which illustrates the lack of uniformity in supposedly identical exposures.

As we see it now the procedure will be used in two ways:

(a) Individual lines can be inspected, as was done with the balloons. The presence of a bump of sufficient size on such a line will indicate an accumulation of dye and hence the presence of a lesion and its location along the line. Thus films from two directions will locate the lesion precisely. The magnitude of the bump will indicate the size of the lesion, as is shown by the difference in the widths of the bumps corresponding to the 8 mm. balloons and the 7 mm. balloon. If lines are read at, say, 2 mm. intervals, the procedure should reveal the presence of a 4 mm. lesion with just two x-rays and localize it completely with four.

If additional information is needed, cross sections can be reconstructed at the level where the lesion is now known to exist.

(b) A subtraction angiogram type film can be made by preparing a large number of lines and running them through a reverse densitometer (film writer). Suppose that a first x-ray is taken before dye injection, a second shortly after when the dye has passed from the anticubital vein through the heart and out into the arteries, but has not yet reached the veins, and a third when the dye has spread through the body. Subtraction of number 1 from number 2 will reveal the arteries, and the film produced by the film writer will be a subtraction angiogram. Subtraction of number 1 from number 3 will reveal both the arteries and the veins in the same way. Subtraction of number 2 from number 3 will reveal the veins alone. While this article was in press this procedure was used with apparent success to visualize the vascular structure in the neck of a patient at the Good Samaritan Hospital in Corvallis.

(2) DISTINGUISHING BETWEEN CANCER AND FIBROCYSTIC DISEASE IN THE BREAST. Even when a mass is known to be present in the breast (e. g. palpable by the physician) it remains a difficult problem to determine without a biopsy whether the mass is cancer or is a fibrocystic lesion. Since the fibrocystic lesion is more dense than the fatty breast tissue, while the cancer is normally still more dense, and since the situation is not obscured by a highly oscillatory skull, it appeared that something ought to be possible simply by examining the densitometer readings along a line through the lesion.

If the lesion has density d and is round with radius r, then the total mass function along the line is

$$y = M(x) = 2d\sqrt{r^2 - x^2}\,, \quad \text{or } x^2 + \frac{y^2}{4d^2} = r^2.$$

Thus, the graph is an ellipse in which the ratio of the axes is $2d$–and it might be expected that if the units on the coordinate axes are chosen judiciously, then cancers will have vertical major axes and fibrocystic lesions horizontal ones, so that the two will be distinguishable at a glance. Figure 11 shows the graphs of six cancers (scirrhous adenocarcinoma, the most common of breast cancers), and Figure 12 shows the graphs of five isolated fibrocystic lesions. These came from breast films (mammograms) in the radiology files at the local hospital, and all were cases where the diagnosis could not be made without a biopsy.

There is one minor point here. The densitometer readings do not provide the total mass function of the lesion alone, but of the lesion plus the breast, in which the characteristic shapes are completely lost. Indeed, because of the variation in the thickness of the breast, it is plain that the same lesion will look very different in different locations. The total mass function $M(x)$ must be decomposed in a normal and an abnormal part, $M(x) = N(x) + A(x)$, where $N(x)$ is the total mass function that would occur if the lesion were not there and the difference $A(x) = M(x) - N(x)$ is what we shall call the abnormality function. It is the abnormality function that presents the characteristic shape. To obtain the abnormality function we combine the

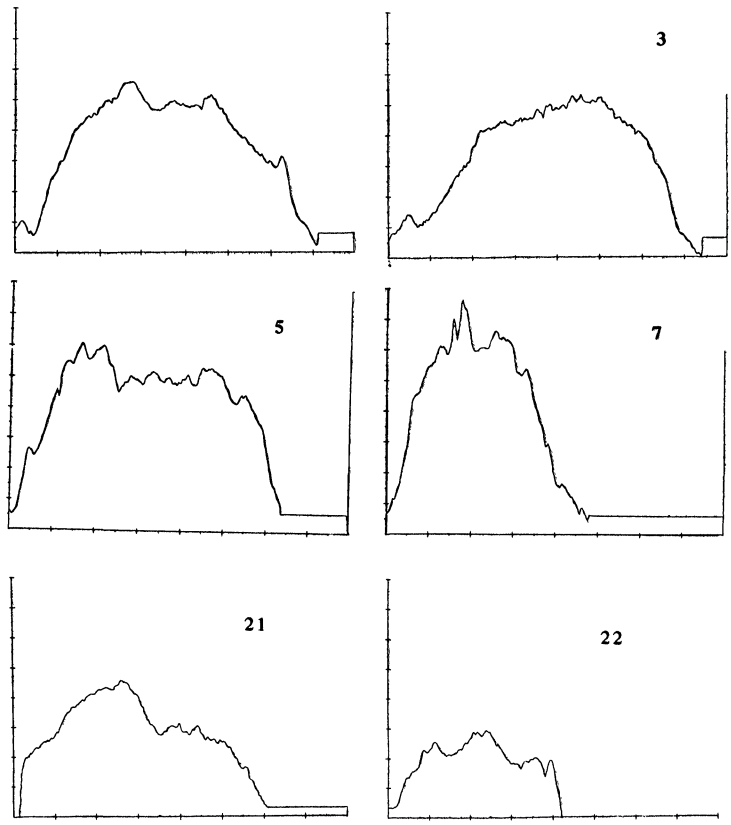


Figure 11. Cancer

available clinical information with inspection of the x-ray films and inspection of the graph of the densitometer readings. Usually it is possible to locate the edges of the lesion fairly easily. Within the lesion, where the function $N(x)$ is not known, it is approximated by a straight line. This approximation is what produced the graphs in Figures 11 and 12.

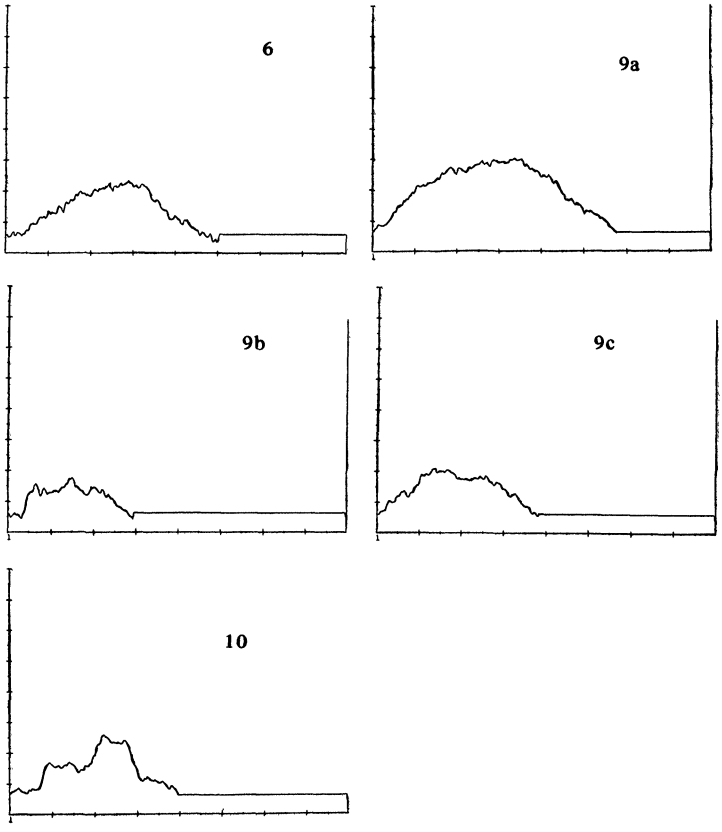


Figure 12. Isolated fibrocystic lesions

By a stroke of good luck calibration is unnecessary in this problem. If D is the densitometer reading corresponding to a total mass M, it can be verified experimentally that, within the range of masses and exposures occurring in mammograms, D is a linear function of M: $D = pM + q$ with numbers p and q that depend on the film and the exposure. Thus

$$D(x) = p(N(x) + A(x)) + q = pN(x) + q + pA(x) = D_n(x) + D_a(x).$$

The term $D_n(x) = pN(x) + q$ is simply the densitometer reading that would occur if the breast were normal, while the interesting term $D_a(x) = pA(x)$, apart from the irrelevant additive constant q, is the densitometer reading that would be produced by the abnormality alone. The stroke of luck that obviates the need for calibration is that the number p is independent of the x-ray exposure, so that the function D_a, which is produced without calibration, is equivalent to the desired function A.

Figure 13 shows six cases of fibrocystic disease in which the primary lesion was not isolated but surrounded by secondary lesions. In such cases the edges of the lesion are naturally more difficult to locate and the characteristic shape

is partially lost. Nevertheless, only two of these, numbers 24 and 25, present the massive, steep sided, rock like appearance of the cancers. These two were almost entirely fibrous, rather than fibrocystic, in nature. For the time being nonisolated and almost entirely fibrous lesions appear to be difficult to distinguish from cancer.

There are other kinds of breast cancer than scirrhous adenocarcinoma and other kinds of benign lesions, but these that we have begun to study are by far the most common.

The noise elimination procedure was not needed in this experiment. The reason is that mammogram films do not make use of the intensifying screen. This produces better results (no quantum mottle) but at the expense of a heavy x-ray exposure: 175 mas. in the top or caudad views, where the breast is flattened out considerably, and 350 mas. in the lateral views. When the noise elimination is brought into play it should be possible to make use of the intensifying screens and to reduce the exposure to an insignificant 10–15 mas.

While this article was in press it was discovered that the ratio between the area under the graph of $y = D_a(x)$ and the area of the rectangle circumscribing this graph is a significant parameter in discriminating between scirrhous adenocarcinoma and fibrocystic disease. Additional mammograms were obtained from the University of Oregon Medical School Hospital and from the Roswell Park Memorial Institute, Buffalo, New York, to bring the total to 33. The above ratio discriminated correctly in every case. This fact throws both light and shadow upon a number of things. See [14].

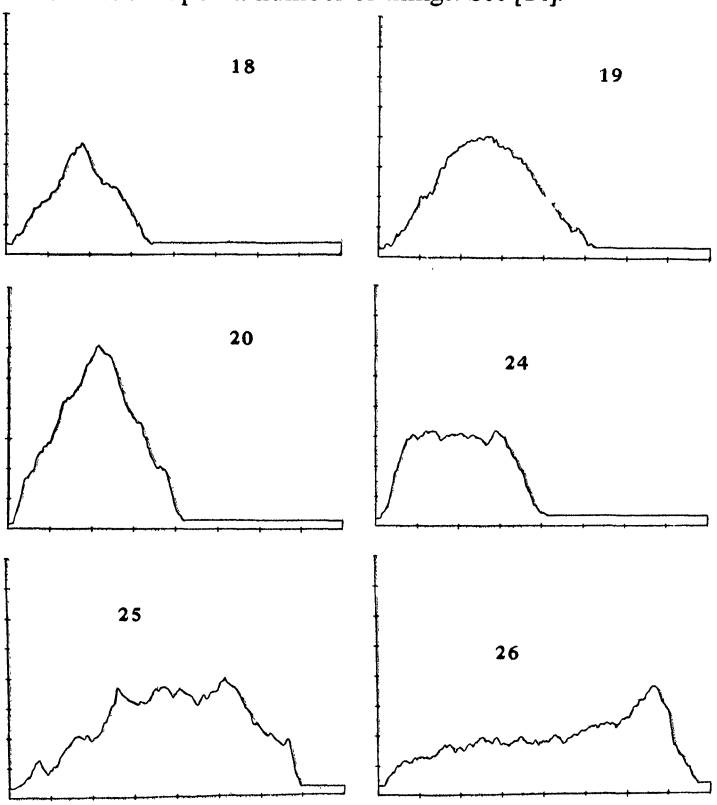


Figure 13. Non-Isolated fibrocystic lesions

(3) DISTINGUISHING BETWEEN AIR BUBBLES AND GALL STONES. Following the removal of stones from the gall bladder the surgeon must determine whether any stones have escaped into the bile duct. Since stones are invisible on a radiograph, he injects iodine contrast dye into the duct, which then appears as a white path surrounded by dark normal tissue. If a stone is present it appears as a dark shadow in the white duct, as in Figure 14.

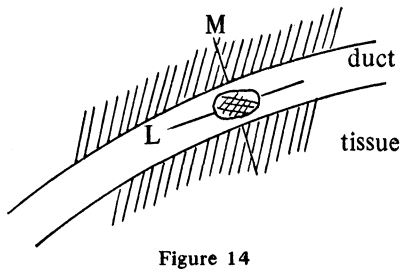


Figure 14

Once in a while an air bubble is injected along with the dye, and it too appears as a dark shadow very difficult to distinguish from the shadow of a stone. In the case of a stone or an indeterminate diagnosis much more extensive surgery must be performed.

It is to be expected that densitometer readings along a line L in the duct should produce results something like those of the last section–that the denser stones should have an appearance something like that of the cancers and the less dense air bubbles an appearance something like that of the fibrocystic lesions. Figure 15 illustrates this with two stones and two air bubbles. Figure 16 shows an irregular stone in which the diagnosis on the basis of shape is hampered by the irregularity. In this case a second line M is read crosswise to the duct so that the shadow can be compared with normal tissue rather than contrast dye which is denser than either bubble or stone. Now the bumps become dips, indicating that the shadow is denser than the normal tissue and that the shadow must be a stone.[3]

When the bubble or stone does not fill the duct it can be tricky to find adjacent normal tissue to use in comparison. Therefore, a better procedure would be to take an x-ray before as well as after the injection of the dye, using the latter only to locate the exact position of the bubble or stone. In this way the dye would not interfere at all and the bubbles would present bumps in sharp contrast to the dips coming from stone.

In this problem the subtraction used to produce the abnormality function is not needed, as the body changes thickness very little in crossing from one side of the shadow to the other, nor is the calibration needed. On the other hand, the noise elimination is needed.

[3] In most of our figures larger densities or total masses are represented by larger numbers, but straight densitometer readings produce the reverse, for larger total masses permit less exposure which results in lighter film. When the films are calibrated the wedge automatically straightens this out, and when subtractions are made, as in the cases of the subtraction angiograms and the mammograms, they are done in an order that straightens it out. In Figures 15 and 16 the densitometer readings are graphed directly, so larger densities are represented by smaller numbers.

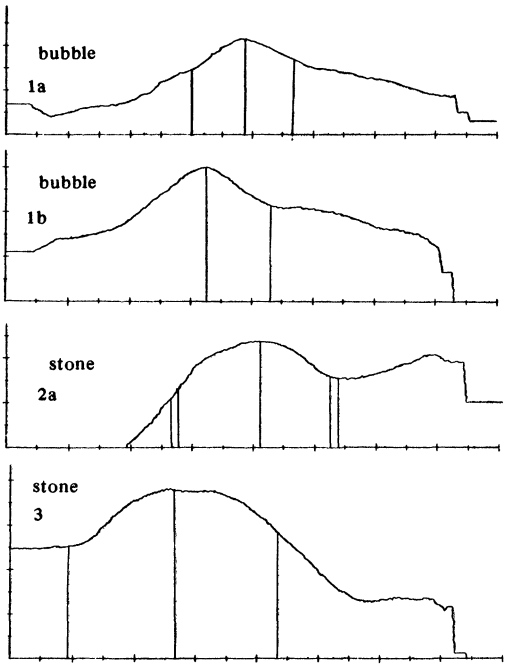


Figure 15

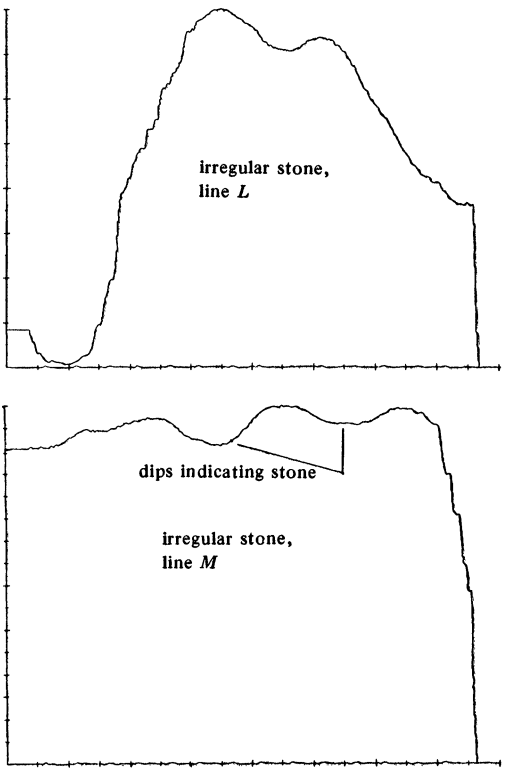


Figure 16

CHAPTER II. MATHEMATICAL ASPECTS

9. The x-ray transform as an operator on L^2. The functions $g(\theta, x)$ in the range of the x-ray transform (see (2.1)) are functions on the space

$$T = \{(\theta, x): \theta \in S^{n-1}, x \in \theta^{\perp}\},$$

which is the tangent bundle to S^{n-1} for those who want to think of it that way. T has a natural measure μ such that

$$\int_T g d\mu = \int_{S^{n-1}} \int_{\theta^{\perp}} g(\theta, x) dx\, d\theta.$$

If $f \in L_0^2(R^n)$ (L^2 with compact support) has support in a compact set K with diameter $\delta(K)$, then the Cauchy-Schwarz inequality gives

$$\int_{\theta} |P_\theta f|^2\, dx \leqslant \delta(K) \|f\|^2_{L^2(R^n)}$$

for each θ, and hence

$$\|Pf\|^2_{L^2(T)} \leqslant \delta(K) |S^{n-1}|\, \|f\|^2_{L^2(R^n)}.$$

This section contains some results about P as an unbounded operator from $L^2(R^n)$ to $L^2(T)$ with domain $L_0^2(R^n)$.

The adjoint of P is easy to compute. If $f \in L_0^2(R^n)$ and $g \in L^2(T)$ (both nonnegative at first to justify the interchanging and combining of integrals) and E_θ is the orthogonal projection on $\theta^{\perp}$, then

$$\begin{aligned} \langle Pf, g \rangle &= \int_{S^{n-1}} \int_{\theta^{\perp}} \int_{-\infty}^{\infty} f(x + t\theta)\, \overline{g(\theta, x)}\, dt dx d\theta \\ &= \int_{S^{n-1}} \int_{R^n} f(y)\, \overline{g(\theta, E_\theta y)}\, dy d\theta = \langle f, P^{\#} g \rangle, \end{aligned}$$

with

$$P^{\#} g(y) = \int_{S^{n-1}} g(\theta, E_\theta y)\, d\theta. \tag{9.1}$$

This gives the following result.

THEOREM 9.2. *For every $g \in L^2(T)$, $P^{\#} g$ is defined almost everywhere and is locally square integrable. Moreover, g is in the domain of P^* if and only if $P^{\#} g$ is globally square integrable, in which case $P^* g = P^{\#} g$.*

For $n = 2$ the square integrability of $P^{\#} g$ depends entirely on cancellation in (9.1). It is immediately seen that if $0 \neq g \in L^2(T)$ is nonnegative, then $P^{\#} g$ cannot be square integrable, while if g is odd as a function of θ, then $P^{\#} g = 0$. Thus the domain of P^* is somewhat peculiar.

A simple integration formula will be needed below.

$$\int_{S^{n-1}} \int_{\theta^{\perp}} h(\eta) |\eta|\, d\eta\, d\theta = |S^{n-2}| \int_{R^n} h(\xi) d\xi. \tag{9.2'}$$

Consider first functions $\tilde{h}$ on S^{n-1}. For these the formula is

$$\int_{S^{n-1}}\int_{S^{n-1}\cap\theta^{\perp}} \tilde{h}(\varphi)\,d\varphi\,d\theta = |S^{n-2}|\int_{S^{n-1}} \tilde{h}(\theta)\,d\theta$$

which is true because both sides define rotation invariant measures on the sphere and they agree for $\tilde{h} = 1$. Now (9.2′) results from expressing the inner integral on the left and the integral on the right in polar coordinates.

Define the operator $\bigwedge^s$ by

$$(\bigwedge{}^s f)^{\wedge}(\xi) = |\xi|^s \hat{f}(\xi) \tag{9.3}$$

for

$$f \in \mathcal{D}_s = \{f \in L^2: |\xi|^s \hat{f} \in L^2\}. \tag{9.4}$$

These operators act on functions in $L^2(R^n)$ and also in $L^2(T)$. In the latter case they act with respect to the variable $x \in \theta^{\perp}$ alone. It is well known and easy to verify that each $\bigwedge^s$ is selfadjoint on $L^2(R^n)$ and that for $s > -n/2$ it is the closure of its restriction to $L_0^2(R^n)$.

Formula (3.5) shows that

$$\left|(\bigwedge{}^{s+1/2}Pf)^{\wedge}(\xi)\right|^2 = 2\pi|(\bigwedge{}^s f)^{\wedge}(\xi)|^2|\xi| \quad \text{for } \xi \in \theta^{\perp}. \tag{9.5}$$

Integration over $\theta^{\perp}$ and S^{n-1} and the integration formula (9.2) give

$$\|\bigwedge{}^{s+1/2}Pf\|^2_{L^2(T)} = 2\pi|S^{n-2}|\,\|\bigwedge{}^s f\|^2_{L^2(R^n)}. \tag{9.6}$$

It follows immediately from the case $s = -1/2$ that P admits a closure $\bar{P}$ whose domain is the same as that of $\bigwedge^{-1/2}$, i.e.

$$\mathcal{D}_{\bar{P}} = \mathcal{D}_{-1/2}.$$

On this domain formula (9.6) and the Fourier transform formula (3.5) continue to hold. It is shown in [13], [14] that if $n > 2$, then every $f \in L^2$ is integrable over almost all lines in almost all directions and that $f \in \mathcal{D}_{\bar{P}}$ if and only if the line integrals produce a function in $L^2(T)$ and that in this case $\bar{P}f$ is given by the line integrals. If $n = 2$, however, there are $f \in \mathcal{D}_{\bar{P}}$ that are not integrable over any line.

For $s \geqslant 0$, $\mathcal{D}_s$ is simply the Sobolev space $\mathcal{H}^s$, but for $s < 0$ this is not so. In general $\mathcal{D}_s$ is a strictly decreasing function of $|s|$. We are not much interested in $\mathcal{D}_s$ for $s < 0$ except in the case $s = -1/2$ when $\mathcal{D}_s$ is the domain of $\bar{P}$ and in the case $s = (1-n)/2$ when, as is shown in §12, $\mathcal{D}_s$ is the domain of the closure of the Radon transform.

Formula (9.6) with $s = 0$ shows that the operator

$$V_0 = (2\pi)^{-1/2}|S^{n-2}|^{-1/2}\bigwedge{}^{1/2}\bar{P} \tag{9.7}$$

extends to an isometry V of $L^2(R^n)$ into $L^2(T)$. Therefore, $V^*V = 1$, and we have the inversion formula for the x-ray transform:

$$\text{If } \bar{P}f = g, \text{ then } f = (2\pi)^{-1}|S^{n-2}|^{-1}\bigwedge{}^{1/2}P^{\#}\bigwedge{}^{1/2}g, \tag{9.8}$$

with $P^{\#}$ given by formula (9.1).

REMARK. Although V is an isometry, its component parts $\wedge^{1/2}$ and P are unbounded, so the above derivation is somewhat formal. In particular, although $\wedge^{1/2}$ and P commute formally, they do not commute as unbounded operators. Nevertheless, formula (9.8) is correct in the sense of products of unbounded operators, as can be seen in [15]. In the case $n = 2$, where the x-ray and Radon transforms coincide (see the next section), formula (9.8) becomes the classical inversion formula of Radon, except that in the latter the powers of $\wedge$ are collected on the right. Hence the classical formula is

$$\text{If } \bar{P}f = g, \text{ then } f = (4\pi)^{-1}P^{\#} \wedge g, \tag{9.9}$$

with $P^{\#}$ given by (9.1). It is shown in §12 that this formula is justified if and only if g is even and lies in $\mathfrak{D}_1$, the domain of $\wedge$. (When $g = \bar{P}f$, what is automatically true is that g is even and lies in $\mathfrak{D}_{1/2}$.)

10. The general k-plane transform. The x-ray transform (involving integration over lines) and the classical Radon transform (integration over $(n-1)$-planes) are the two extreme cases of the general k-plane transform. If π is a subspace of R^n of dimension k, write $x = (x', x'')$ with $x' \in \pi$ and $x'' \in \pi^{\perp}$ and define

$$P_\pi f(x'') = \int_\pi f(x', x'')dx' \quad \text{for } x'' \in \pi^{\perp}. \tag{10.1}$$

The x-ray transform is the case where $k = 1$ and π is the line through the origin with direction θ, while the Radon transform is the case where $k = n - 1$ and π is the subspace orthogonal to θ. Apart from notation the two are the same in the crucial dimension 2, and this was the origin of our interest in the Radon transform.

Many interesting similarities, distinctions, and subtleties arise as k and n vary. An example is the following result, which becomes false at $k = n/2$.

THEOREM 10.2. *If f is square integrable on R^n, then for almost every k-space π, f is actually integrable over almost all k-planes parallel to π, provided $k < n/2$. Moreover,*

$$\int_{G_{n,k}} \|P_\pi f\|^2_{L^q(\pi^\perp)}\, d\mu \leqslant c^2 \|f\|^2_{L^2(R^n)}, \qquad q = 2(n-k)/(n-2k). \tag{10.3}$$

Here $G_{n,k}$ is the Grassmann manifold of k-spaces in R^n and μ is the finite measure on $G_{n,k}$ (unique to within a constant factor) that is invariant under orthogonal transformations. The proof can be found in [13]. In the remainder of the article, in order to avoid having to deal with the Grassmann manifolds, I shall concentrate on the Radon transform and a few of its applications. The ideas needed to handle the general k-plane transform are contained in the two extreme cases of the x-ray transform and the Radon transform, so nothing much will be lost.

11. The Radon transform. The Radon transform of an integrable function f is the function

$$Rf(\theta, t) = R_\theta f(t) = \int_{\langle x,\theta\rangle = t} f(x)d\mu_{n-1} \tag{11.1}$$

where μ_{n-1} is the Lebesgue measure on the plane $\langle x, \theta \rangle = t$. For fixed $\theta \in S^{n-1}$, $R_\theta f$ is defined almost everywhere in t and satisfies

$$\int |R_\theta f(t)|\, dt \leqslant \int |f(x)|\, dx.$$

Unless it is stated otherwise or obvious from the context the function f is assumed to be square integrable with compact support.

If ρ is a function of one variable, the analogue of (3.1) is

$$\int R_\theta f(t)\rho(t)dt = \int f(x)\rho(\langle x, \theta \rangle)\, dx. \tag{11.2}$$

Taking $\rho(t) = t^m$ we get the consistency condition that

$$p_m(f; \theta) = \int R_\theta f(t) t^m\, dt = \int f(x)\langle x, \theta \rangle^m\, dx \tag{11.3}$$

must be a homogeneous polynomial of degree m. Taking $\rho(t) = e^{-i\pi t}$ we get the Fourier transform formula

$$(R_\theta f)\hat{}(\tau) = (2\pi)^{(n-1)/2} \hat{f}\,(\tau\theta). \tag{11.4}$$

It is apparent from formula (11.3) that

$$p_m(f; \theta) = (2\pi)^{n/2}\big(\langle x, \theta \rangle^m f(x)\big)\hat{}_{\xi=0} = (2\pi)^{n/2}\langle iD, \theta \rangle^m \hat{f}\,(0)$$

and consequently that the p_m are effectively the homogeneous parts of the Taylor expansion of $\hat{f}$:

$$\hat{f}\,(\zeta) = (2\pi)^{-n/2} \sum_{m=0}^{\infty} i^{-m} p_m(f; \zeta)/m!. \tag{11.5}$$

The information about the Radon transform needed to determine the function can be described as follows.

THEOREM 11.6. *For any square integrable f with compact support the set of directions θ with $R_\theta f = 0$ is a homogeneous algebraic variety in S^{n-1}. Conversely, for any given proper homogeneous variety V and any given function $g_0 \in C_0^\infty$ there is an $f \in C_0^\infty$ such that $f = g_0$ except on an arbitrary ring just inside the support of g_0 and $R_\theta f = 0$ for all $\theta \in V$.*

PROOF. From formula (11.3) it follows that $R_\theta f = 0$ if and only if all $p_m(f; \theta) = 0$ and hence that the set of such θ is a homogeneous algebraic variety. For the converse, choose a homogeneous polynomial q that vanishes on V, form the corresponding differential operator Q, and repeat the proof of Theorem 4.3. (Any variety in S^{n-1} is homogeneous, i.e. the intersection of S^{n-1} with a homogeneous variety, so the term homogeneous can be dropped if desired.)

The question of when the functions in a given finite dimensional space can be distinguished by the Radon transform from a single direction is answered by Theorem 4.12.

It is useful to know which sequences of polynomials come via (11.3) from some function f. A simple initial characterization is the following.

THEOREM 11.7. *For each m, let p_m be a homogeneous polynomial of degree m. There is a square integrable f with support in the unit ball such that $p_m(\theta) = p_m(f;\theta)$ for all m if and only if*

(a) $|p_m(\zeta)| \leqslant c|\zeta|^m$ *for all m and all $\zeta \in C^n$.*

(b) *The sum $\sum i^{-m} p_m(\xi)/m!$ is square integrable on R^n.*

PROOF. The necessity of the two conditions is obvious (Cauchy-Schwarz in (11.3)). For the sufficiency, define $\hat{f}$ by (11.5) with $p_m(\zeta)$ in place of $p_m(\hat{f};\zeta)$. According to condition (a) the sum converges on C^n and $\hat{f}$ satisfies

$$|\hat{f}(\zeta)| \leqslant (2\pi)^{-n/2} c e^{|\zeta|} \leqslant c_1 e^{|\zeta_1| + \cdots + |\zeta_n|}.$$

By the Paley-Wiener theorem (see §14) the inverse Fourier transform f has support in the cube $Q = \{x: |x_i| \leqslant 1\}$. By a rotation Q becomes any cube that circumscribes the unit ball, so f has support in the ball.

While Theorem 11.7 is a useful initial step, it is undesirable to require condition (a) on the complex space. When f is known there is control over the p_m on the complex space by means of the second integral in (11.3). When only Rf is known, however, there is control only on the real space. Somewhat surprisingly, condition (a) on the real space (along with condition (b), of course) turns out to be enough, as is shown in §15.

Before going on we record the simple commutation relation between the Radon transform and differentiation. If q is a polynomial, let $q_\theta(\tau) = q(\tau\theta)$, and let Q and Q_θ be the corresponding differential operators. Then

$$R_\theta Q f = Q_\theta R_\theta f, \tag{11.8}$$

as is immediately checked by Fourier transform.

12. The Radon transform as an operator on $L^2(R^n)$. If $f \in L_0^2(R^n)$ (square integrable with compact support) has support in the compact set K, then the Cauchy-Schwarz inequality gives

$$\|R_\theta f\|^2_{L^2(R^1)} \leqslant c\delta(K)^{n-1} \|f\|^2_{L^2(R^n)}, \quad \delta = \textit{diameter}. \tag{12.1}$$

Thus it is natural to consider R as an unbounded operator from $L^2(R^n)$ to $L^2(S^{n-1} \times R^1)$ with domain $L_0^2(R^n)$. The adjoint of R is easy to compute. If $f \in L_0^2(R^n)$ and $g \in L^2(S^{n-1} \times R^1)$, then

$$\begin{aligned}
\langle Rf, g\rangle &= \int_{S^{n-1}} \int_{-\infty}^{\infty} \int_{\langle x,\theta\rangle = t} f(x)\, d\mu_{n-1}\, \overline{g(\theta, t)}\, dt\, d\theta \\
&= \int_{S^{n-1}} \int_{R^n} f(x)\, \overline{g(\theta, \langle x, \theta\rangle)}\, dx\, d\theta \\
&= \int_{R^n} f(x) \int_{S^{n-1}} \overline{g(\theta, \langle x, \theta\rangle)}\, d\theta\, dx = \langle f, R^{\#} g\rangle
\end{aligned}$$

with

$$R^{\#} g(x) = \int_{S^{n-1}} g(\theta, \langle x, \theta\rangle)\, d\theta. \tag{12.2}$$

This computation should be done first for $f \geqslant 0$ and $g \geqslant 0$, in which case all the combinations and exchanges of integration are automatically justified,

and then a posteriori they are justified for all

$$f \in L_0^2(R^n) \quad \text{and} \quad g \in L^2(S^{n-1} \times R^1).$$

This gives the following result.

THEOREM 12.3. *For every $g \in L^2(S^{n-1} \times R^1)$, $R^\# g$ as defined by* (12.2) *is defined almost everywhere and is locally square integrable. Moreover, g is in the domain of R^* if and only if $R^\# g$ is globally square integrable, in which case $R^* g = R^\# g$.*

The domain of R^* is rather peculiar. It is easily verified, for example, that no nonnegative $0 \neq g \in L^2(S^{n-1} \times R^1)$ lies in the domain, while every odd g most certainly does. The condition of lying in the domain of R^* is one of cancellation in (12.2).

Formula (11.4) shows that

$$\left|\left(\bigwedge^{(n-1)/2+s} Rf\right)^{\wedge}(\tau)\right|^2 = (2\pi)^{n-1}|\tau|^{2s}\left|\hat{f}\,(\tau\theta)\right|^2|\tau|^{n-1}. \tag{12.4}$$

Integration with respect to τ and θ gives

$$\left\|\bigwedge^{(n-1)/2+s} Rf\right\|^2_{L^2(S^{n-1}\times R^1)} = 2(2\pi)^{n-1}\|\bigwedge^s f\|^2_{L^2(R^n)}. \tag{12.5}$$

It is easily seen from the case $s = (1-n)/2$ that the operator R admits a closure $\bar{R}$ whose domain is the same as the domain of $\bigwedge^{(1-n)/2}$, i.e.

$$\mathfrak{D}_{\bar{R}} = \mathfrak{D}_{(1-n)/2}.$$

On this domain formulas (12.4) and (12.5) continue to hold with R replaced by $\bar{R}$, but it is not to be expected that $\bar{R}$ is definable by integrals over $(n-1)$-planes. There are functions $f \in \mathfrak{D}_{\bar{R}}$ not integrable over any $(n-1)$-plane. See [13]–[15]. In general functions belong to $\mathfrak{D}_{\bar{R}}$ because of cancellations in the integrals rather than because of absolute integrability. About the best that can be said on the latter score comes from the Hausdorff-Young theorem and [13] which show that

$$L^2 \cap L^p \subset \mathfrak{D}_{\bar{R}} \quad \text{if } p \leqslant 2n/(2n-1).$$

It can be shown that for the functions in this class the required integrals do exist and do give $\bar{R}$.

Formula (12.5) with $s = 0$ shows that

$$V = 2^{-1/2}(2\pi)^{(1-n)/2}\bigwedge^{(n-1)/2}\bar{R}$$

is an isometry, so that $V^* V = 1$ and hence that formally at least

$$\bar{R}^{-1} g = 2^{-1}(2\pi)^{1-n} R^* \bigwedge^{n-1} g = 2^{-1}(2\pi)^{1-n}\int_{S^{n-1}} \bigwedge^{n-1} g(\theta, \langle x, \theta\rangle)\, d\theta,$$

which is the classical inversion formula for the Radon transform. While the powers of $\bigwedge$ commute formally with R and R^*, as unbounded operators they do not.

THEOREM 12.6. *The domain of $\bar{R}$ is $\mathfrak{D}_{(1-n)/2}$, and the range consists of the even functions in $\mathfrak{D}_{(n-1)/2}$. More generally, the range of $\bar{R}$ on $\mathfrak{D}_s \cap \mathfrak{D}_{\bar{R}}$ consists of the even functions in $\mathfrak{D}_{(n-1)/2+s} \cap \mathfrak{D}_{(n-1)/2}$, and on this domain*

(12.7) $\bar{R}^{-1} = 2^{-1}(2\pi)^{1-n}\bigwedge^{(n-1)/2-\sigma} R^* \bigwedge^{(n-1)/2+\sigma}$ *if* $0 \leqslant \sigma \leqslant s$.

PROOF. Formula (12.5) shows that the range is contained in

$$\mathfrak{D}_{(n-1)/2+s} \cap \mathfrak{D}_{(n-1)/2+\sigma}$$

and clearly any function in the range is even. If g is any even function in this space, then $\hat{f}(\tau\theta) = (2\pi)^{(1-n)/2}\hat{g}(\theta, \tau)$ is well defined, and for all α

$$\int_{R^n} |\hat{f}(\xi)|^2 |\xi|^{2\alpha} d\xi = 2^{-1}(2\pi)^{1-n} \int_{S^{n-1}} \int_{-\infty}^{\infty} |g(\theta, \tau)|^2 |\tau|^{n-1+2\alpha} d\tau\, d\theta.$$

Taking $\alpha = 0$ we see that $\hat{f}$ lies in L^2, so the inverse Fourier transform f exists and lies in L^2. Taking $\alpha = (1-n)/2$ (and recalling that all the domains $\mathfrak{D}_\beta$ are contained in L^2) we see that f lies in $\mathfrak{D}_{\bar{R}}$, and that $\bar{R}f = g$. Taking $\alpha = s$ we see that f lies in $\mathfrak{D}_s$.

Now, for any $u \in \mathfrak{D}_{\bar{R}}$ and $0 \leqslant \sigma \leqslant s$ we have

$$\left\langle \bar{R}u, \bigwedge^{(n-1)/2+\sigma} g \right\rangle = 2(2\pi)^{n-1} \int_{S^{n-1}} \int_0^{\infty} \hat{u}(\tau\theta) |\tau|^{(1-n)/2+\sigma} \overline{\hat{f}(\tau\theta)} |\tau|^{n-1} d\tau\, d\theta$$

$$= 2(2\pi)^{n-1} \left\langle u, \bigwedge^{(1-n)/2+\sigma} f \right\rangle.$$

Since $(1-n)/2 \leqslant (1-n)/2 + \sigma \leqslant s$ it follows that $\bigwedge^{(1-n)/2+\sigma} f \in L^2$, hence that $\bigwedge^{(n-1)/2+\sigma} g$ is in the domain of R^* and that

$$R^* \bigwedge^{(n-1)/2+\sigma} g = 2(2\pi)^{n-1} \bigwedge^{(1-n)/2+\sigma} f.$$

REMARK. Closer inspection of the proof shows that formula (12.7) holds if $(n-1)/2 + s \leqslant \sigma \leqslant 0$.

13. Ludwig's theorem on supports. This section contains a proof of a very nice theorem of D. Ludwig [8], [9] that characterizes the functions $g \in L^2(S^{n-1} \times R^1)$ such that $g = Rf$ where $f \in L^2(R^n)$ has support in a given compact convex set K. This proof may be somewhat simpler than the existing ones in that it makes no use of spherical harmonics or properties of the wave equation. We first consider the case where $g(\theta, t)$ is independent of θ.

LEMMA 13.1. *If* $\rho \in C_0^\infty([-1, 1])$ *is even, then* $\rho = R\psi$ *with* $\psi \in C_0^\infty(R^n)$ *and* $\psi(x) = 0$ *for* $|x| > 1$.

PROOF. We suppose at first that n is odd so that $\bigwedge^{n-1} = (-id/dt)^{n-1}$. The inversion formula of the last section shows that if

$$\psi(x) = 2^{-1}(2\pi i)^{1-n} \int_{S^{n-1}} \rho^{(n-1)}(\langle x, \theta \rangle)\, d\theta$$

$$= 2^{-1}(2\pi i)^{1-n} |S^{n-2}| \int_0^{\pi} \rho^{(n-1)}(|x| \cos\varphi) \sin^{n-2}\varphi\, d\varphi$$

$$= 2^{-1}(2\pi i)^{1-n} |S^{n-2}| \frac{1}{|x|} \int_{-|x|}^{|x|} \rho^{(n-1)}(s) \left(1 - \frac{s^2}{|x|^2}\right)^{(n-3)/2} ds$$

then $\psi \in C^\infty \cap \mathfrak{D}_{\bar{R}}$ and $\bar{R}\psi = \rho$. When $|x| > 1$, $n-2$ integrations by parts gives 0, for $(1 - s^2/|x|^2)^{(n-3)/2}$ is a polynomial of degree $n-3$.

If n is even, we form the function $\psi_0(x_0, x)$ for R^{n+1} as above and put

$$\psi(x) = \int \psi_0(x_0, x)\, dx_0.$$

Since ψ_0 is a function of the absolute value alone, so is ψ, and their Radon transforms obviously coincide from the direction of the x_n-axis.

LEMMA 13.2. *If $\varphi \in C_0^\infty$ is 0 for $|x| < 1$, then*

$$\int_{S^{n-1}} \bigwedge^{n-1} R\varphi \, d\theta = 0 \text{ for } |t| > 1.$$

PROOF. Since $\bigwedge^{n-1} R\varphi$ is even, the integral is an even function of t, and it is enough to prove that it is orthogonal to any even $\rho \in C_0^\infty([-1, 1])$. By Lemma 13.1, $\rho = R\psi$, and by the isometry of the last section the inner product of ρ with the integral is

$$\langle R\psi, \bigwedge^{n-1} R\varphi \rangle = 2(2\pi)^{n-1} \langle \psi, \varphi \rangle = 0,$$

since $\psi = 0$ for $|x| > 1$ while $\varphi = 0$ for $|x| < 1$.

The next lemma is given in [**8**], but proved in a different way.

LEMMA 13.3. *If $\varphi \in C_0^\infty$ is 0 for $|x| < 1$ and q is a polynomial of degree m, then*

$$\int_{S^{n-1}} q(\theta) \bigwedge^{n-1} R\varphi \, d\theta$$

is a polynomial of degree $< m$ on $|t| < 1$.

PROOF. It is no loss of generality to assume that q is homogeneous. In this case, if Q is the corresponding differential operator, then formula (11.8) shows that

$$RQ\varphi = \frac{\partial^m}{\partial t^m} q(\theta) R\varphi.$$

Consequently. Lemma 13.2 (with $Q\varphi$ in place of φ) shows that the mth derivative of the integral in question is 0 on $|t| < 1$, hence that the integral itself is a polynomial of degree $< m$.

In what follows K is an n-dimensional compact convex set with support function

$$\|\xi\|_K = \sup_{x \in K} \langle x, \xi \rangle.$$

The theorem of Ludwig [**9**] is as follows.

THEOREM 13.4. *Let $g \in L^2(S^{n-1} \times R^1)$. The necessary and sufficient conditions that $g = Rf$ for some $f \in \mathcal{D}_s$ with support in K are:*

(a) *$g \in \mathcal{D}_{(n-1)/2+s}$ and is even.*

(b) *$g(\theta, t) = 0$ for $t > \|\theta\|_K$.*

(c) *$p_m(\theta) = \int g(\theta, t) t^m \, dt$ is a homogeneous polynomial of degree m.*

PROOF. The necessity of (a) is given in Theorem 12.6 and that of (c) in (11.3). The necessity of (b) is evident as the inequality $t > \|\theta\|_K$ means that the plane $\langle x, \theta \rangle = t$ does not touch K.

Moreover, Theorem 12.6 shows that conditions (a) and (b) provide a function $f \in \mathcal{D}_s \cap \mathcal{D}_{\bar{R}}$ with $\bar{R}f = g$. The point at issue is to show that this function f has support in K, and this will be done by showing that f is orthogonal to every $\varphi \in C_0^\infty$ that is 0 on K. For any $\varphi \in C_0^\infty$ the isometry of the last section gives

$$\langle f, \varphi \rangle = c\langle g, \wedge^{n-1} R\varphi \rangle = c\int_{S^{n-1}} \langle g_\theta, \wedge^{n-1} R_\theta \varphi \rangle \, d\theta. \tag{13.5}$$

Now we assume that K is the unit ball $B(0; 1)$ and establish the theorem in this case. (Afterward the general case will follow easily.) For fixed θ we expand g in a Legendre series:

$$g(\theta, t) = \sum g_m(\theta) L_m(t)$$

with

$$g_m(\theta) = \langle g_\theta, L_m \rangle = \int_{-1}^{1} g(\theta, t) L_m(t) \, dt. \tag{13.6}$$

All that needs to be known about the Legendre polynomials L_m is that they are obtained by applying Gram-Schmidt to the powers of t in the space $L^2([-1, 1])$. Thus, they form an orthonormal basis of this space, L_m is a polynomial of degree m, and L_m is orthogonal to all polynomials of degree $< m$. We have

$$\langle g_\theta, \wedge^{n-1} R_\theta \varphi \rangle = \sum g_m(\theta) \langle L_m, \wedge^{n-1} R_\theta \varphi \rangle.$$

This series can be integrated term by term over S^{n-1}, since the sum of the absolute values is at most $\| g_\theta \|_{L^2} \| \wedge^{n-1} R_\theta \varphi \|_{L^2}$, and formula (13.5) gives

$$\langle f, \varphi \rangle = c \sum \int_{-1}^{1} L_m(t) \int_{S^{n-1}} g_m(\theta) \wedge^{n-1} R\varphi \, d\theta \, dt. \tag{13.7}$$

Since L_m is a polynomial of degree m, it follows from (13.6) and condition (c) that g_m is a polynomial of degree m, then from Lemma 13.3 that the inner integral in (13.7) is a polynomial of degree $< m$–orthogonal therefore to L_m. Thus, each term in the series is 0.

This completes the proof for the ball $B(0; 1)$. Of course, the proof for $B(0; r)$ is the same, and the proof for $B(a; r)$ follows from this and obvious identities for translations (e.g. $\|\xi\|_{K+a} = \|\xi\|_K + \langle a, \xi \rangle$). Finally, if K is any compact set and B is any ball containing it, then $\|\xi\|_K \leqslant \|\xi\|_B$, so what has been done shows that f has support in B, and hence in K, since K is the intersection of such balls.

REMARK. Given an arbitrarily large number N, it is not hard to find a function f with the following properties.

(a) $f \in \mathcal{D}_s$ *for all* $s > -N$, *so in particular* $f \in C^\infty$ *and all derivatives are square integrable.*

(b) $(1 + |x|)^N f$ *is integrable.*

(c) Rf *has compact support, while* f *does not.*

Simply start with a function $E \in C^\infty$ which is equal to $|x|^{2-n}$ (or $\log |x|$ if $n = 2$) on $|x| \geqslant 1$, and take f to be any sufficiently high derivative of E. Formula (11.8) shows that $(\bar{R}_\theta f)'' = 0$ for $|t| > 1$, and hence that $\bar{R}_\theta f(t) = 0$ for $|t| > 1$. It is of course the polynomial condition (c) that is violated in

Ludwig's theorem, with $g = \bar{R}f$. From the proof of (11.3) it follows that the p_m are indeed polynomials for $m \leqslant N$ but not necessarily otherwise. This means that if $f \in L^2$ is rapidly decreasing (i.e. satisfies (b)) for all N, then $\bar{R}f$ has compact support if and only if f does.

It is interesting that the polynomial consistency condition (c) in Ludwig's theorem disappears completely when the compact support is dropped (cf. Theorem 12.6). It is not a continuous condition, and cannot be preserved, but some remnant might be expected to remain.

14. Paley-Wiener theorems. The purpose of the Paley-Wiener theorems is to give the conditions under which a function $\hat{f}$ is the Fourier transform of a function or distribution f with support in a given compact n dimensional convex set K. There are three classical versions.

PALEY-WIENER I. *Let $\hat{f} \in L^2$. Then f has support in K if and only if $\hat{f}$ has an analytic extension to C^n satisfying*

$$|\hat{f}(\zeta)| \leqslant ce^{\|\operatorname{Im}\zeta\|_K} \quad \text{for } \zeta \in C^n. \tag{14.1}$$

When K is the rectangle,

$$K = \{x: |x_j| \leqslant a_j, j = 1, \ldots, n\},$$

the support function has the natural extension to C^n

$$\|\zeta\|_K = \sum a_j|\zeta_j|.$$

PALEY-WIENER II. *Let $\hat{f} \in L^2$ and let K be a rectangle. Then f has support in K if and only if $\hat{f}$ has an analytic extension to C^n satisfying*

$$\limsup_{|\zeta|\to\infty} \frac{\log|\hat{f}(\zeta)|}{\|\zeta\|_K} \leqslant 1 \quad \text{for } \zeta \in C^n. \tag{14.2}$$

PALEY-WIENER-SCHWARTZ. *Let $\hat{f}$ be a tempered distribution and let K be a rectangle. Then f has support in K if and only if $\hat{f}$ has an analytic extension to C^n satisfying* (14.2).

Paley-Wiener I is simple and straightforward to prove. Paley-Wiener II, which is much stronger because of the replacement of Im ζ_j by $|\zeta_j|$, is rather slippery, and in fact has been the subject of erroneous proofs. Paley-Wiener-Schwartz is not difficult to prove on the basis of Paley-Wiener II and simple facts about tempered distributions.

In working with x-rays, one acquires the habit of looking at the Fourier transforms along individual lines through the origin (corresponding to individual x-rays). This habit and the theorem of Ludwig lead to some improvements in the Paley-Wiener theorems.

THEOREM 14.3. *Let $\hat{f} \in L^2$ and let K be symmetric about 0. Then f has support in K if and only if $\hat{f}$ is C^∞ and for almost every real direction θ the function $\hat{f}_\theta(\tau) = \hat{f}(\tau\theta)$ has an analytic extension to C^1 satisfying*

$$\limsup_{|\tau|\to\infty} \frac{\log|\hat{f}_\theta(\tau)|}{|\tau|} \leqslant \|\theta\|_K \quad \text{for } \tau \in C^1. \tag{14.4}$$

This improves Paley-Wiener II in two ways. It requires the analytic extensions only along individual lines through the origin, which form a manifold in C^n of real dimension $(n-1)+2 = n+1$, and moreover requires no matching of the extensions. Also it allows any symmetric K, not just rectangles–showing, for instance, that Paley-Wiener II is true with any symmetric K and any complex norm extending the support function. In all of the Paley-Wiener theorems the necessity is obvious and we prove only the sufficiency.

PROOF. Since $\hat{f}$ is square integrable and locally bounded, it follows that for almost all θ, $\hat{f}_\theta$ is square integrable, and we have a square integrable g_θ such that $\hat{g}_\theta = (2\pi)^{(n-1)/2}\hat{f}_\theta$. Set $g(\theta, t) = g_\theta(t)$. From the fact that $\hat{f}$ is square integrable and locally bounded it follows (polar coordinates) that $g \in \mathcal{D}_{(n-1)/2}$, and it is obvious that g is even. From Paley-Wiener II in the case of 1 variable it follows that $g_\theta(t) = 0$ for $t > \|\theta\|_K$. Thus, conditions (a) and (b) in Ludwig's theorem ($s = 0$) hold, and it remains to check (c). To establish (c) we compute the Fourier transform of $(2\pi)^{(1-n)/2}g_\theta$ by expanding $e^{-i\tau t}$ in its Taylor series and integrating term by term. On the one hand the result is $\hat{f}_\theta(\tau) = \hat{f}(\tau\theta)$, and on the other it is a power series in τ with coefficients $i^{-m}p_m(g;\theta)/m!$. Consequently, $p_m(g;\theta) = i^m\langle\theta, D\rangle^m\hat{f}(0)$ is a homogeneous polynomial of degree m. Now, Ludwig's theorem gives that g is the Radon transform of an f with support in K, and it is plain that the Fourier transform of f is $\hat{f}$.

The Paley-Wiener-Schwartz theorem can be improved in various ways. One of them is as follows.

THEOREM 14.5. *Let $\hat{f}$ be a tempered distribution and let K be symmetric. Then f has support in K if and only if $\hat{f}$ is C^∞, $(1+|x|)^{-N}\hat{f} \in L^2$ for some N, and for almost every real direction θ, $\hat{f}_\theta(\tau)$ has an analytic extension to C^1 satisfying* (14.4).

PROOF. Fix $\varepsilon > 0$ and let $\varphi \in C_0^\infty$ have support in the ball $B(0;\varepsilon)$. Then $\hat{f}\hat{\varphi}$ is square integrable and

$$\limsup_{|\tau|\to\infty} \frac{\log|\hat{f}_\theta(\tau)\hat{\varphi}_\theta(\tau)|}{|\tau|} \leqslant \|\theta\|_K + \varepsilon|\theta|.$$

Since the term on the right is the support function for $K + B(0;\varepsilon)$, it follows from Theorem 14.3 that $f * \varphi$ has support in $K + B(0;\varepsilon)$. Since this is true for every φ, it follows that f itself has support in $K + B(0;\varepsilon)$, and then that f has support in K.

REMARK. The polynomial growth condition, $(1+|x|)^{-N}\hat{f} \in L^2$, is of course necessary, but we do not know to what extent it is actually needed in the proof of sufficiency. The Paley-Wiener-Schwartz theorem shows that if a tempered distribution on R^n extends to an analytic function of exponential type on C^n, then it automatically has polynomial growth on R^n. We do not know what happens in this regard when the extendability is only assumed along lines through the origin with real directions. Some information, but probably not the best possible, is given in the next theorem.

Before stating the theorem we note that from the definition of a tempered distribution and the elementary inequality

$$\int |t|^{2\alpha}|\varphi|^2\,dt \leqslant \frac{4}{(2\alpha+1)^2}\int |t|^{2\alpha+2}|\varphi'|^2\,dt, \tag{14.6}$$

which holds for $\alpha \geqslant 0$ and $\varphi \in C_0^\infty(R^1)$, it follows that every tempered distribution g on R^1 has a bound of the form

$$|\langle g, \varphi\rangle|^2 \leqslant c^2\int (1+t^2)^N |D^q\varphi|^2\,dt = c^2\|\varphi\|_{N,q}^2 \quad \text{for } \varphi \in C_0^\infty. \tag{14.7}$$

THEOREM 14.8. *Let $\hat{f}$ be a C^∞ function on R^n such that:*
(a) *Each $\hat{f}_\theta$ has an analytic extension to C^1 satisfying* (14.4).
(b) *Each $\hat{f}_\theta$ is tempered on R^1 with a bound of the type* (14.7) *where N and q are fixed and $c = c(\theta)$ satisfies $\int_{S^{n-1}} c^2\,d\theta < \infty$. Then f has support in K.*

PROOF. It is readily seen that when $q = 0$ we are back in the case of Theorem 14.5, so we shall prove the theorem by induction. For any function $\hat{g}$ on R^1 we set $S\hat{g}(t) = \int_0^t \hat{g}(s)(t-s)\,ds$. It is plain that $(S\hat{g})'' = \hat{g}$; that if $\hat{g}$ is C^∞, so is $S\hat{g}$; that if $\hat{g}$ has an analytic extension of exponential type, so does $S\hat{g}$; and that if $\hat{g}$ has compact support, so does $S\hat{g}$ provided

$$\int \hat{g}(t)\,dt = 0 \quad \text{and} \quad \int t\hat{g}(t)\,dt = 0. \tag{14.9}$$

Fix an even function $e \in C_0^\infty([-1, 1])$ with integral 1. If $\varphi \in C_0^\infty$ and

$$\alpha = \int \varphi(t)\,dt \quad \text{and} \quad \beta = \int t\varphi(t)dt,$$

then $\varphi = \alpha e - \beta e' + \psi$ where $\psi \in C_0^\infty$ satisfies (14.9) and

$$|\alpha| \leqslant \sqrt{\pi}\,\|\varphi\|_{1,0} \quad \text{and} \quad |\beta| \leqslant \sqrt{\pi}\,\|\varphi\|_{2,0}. \tag{14.10}$$

Consequently, if $N \geqslant q$ and $q \geqslant 2$, then, by (14.6),

$$\|\psi\|_{N,q-2} \leqslant k\|\varphi\|_{N,q-2} \tag{14.11}$$

where the constant k depends only on e and q.

If $\hat{g}$ has the bound (14.7), then with ψ as above we have

$$|\langle S\hat{g}, \psi\rangle| = |\langle \hat{g}, S\psi\rangle| \leqslant c\|\psi\|_{N,q-2} \leqslant ck\|\varphi\|_{N,q-2}.$$

When this is combined with (14.10) we find that if $\hat{g}$ has the bound (14.7) with $N \geqslant q$ and $q \geqslant 2$, then

$$|\langle S\hat{g}, \psi\rangle| \leqslant \max_{|t|\leqslant 1}|\hat{g}|ck\|\varphi\|_{N,q-2}, \tag{14.12}$$

where the constant k depends only on e and q.

This effectively completes the proof by induction. We can always assume that $N \geqslant q$, and by (14.6) we can assume that $q \geqslant 2$. For fixed θ, we consider $\hat{g}_\theta = S\hat{f}_\theta$. It is plain that $\hat{g}_\theta(t)$ is an even function of (θ, t), so it determines a function $\hat{g}$ on R^n with $\hat{g}(t\theta) = \hat{g}_\theta(t)$. What we have done shows that $\hat{g}$ satisfies the conditions of the theorem with q replaced by $q-2$. It follows by induction that g has compact support, hence that $\hat{g}$ has polynomial growth, and hence that

$$\hat{f}(\xi) = \left\langle \frac{\xi}{|\xi|}, D\right\rangle^2 \hat{g}(\xi)$$

does also, so we are back in the case of Theorem 14.5.

The rest of the section contains some results that will be of use in the next and also have some interest of their own. It is assumed throughout that K is symmetric as well as compact, convex, and n dimensional. We define $E_N(K)$ to be the space of entire functions $\hat{f}$ on C^n satisfying:

$$\|\hat{f}\|_N^2 = \int_{R^n} \left(1 + |\xi|^2\right)^{-N} |\hat{f}|^2 \, d\xi < \infty, \tag{14.13}$$

and

$$|\hat{f}(\tau\theta)| \leqslant c\left(1 + |\tau|^2\right)^p e^{|\tau| \, \|\theta\| K} \tag{14.14}$$

with constants c and p which may depend on $\hat{f}$. According to Theorem 14.5, $E_N(K)$ consists of the Fourier transforms of the distributions which have support in K and have order N in the sense of (14.13). The number N is assumed to be a nonnegative even integer.

LEMMA 14.15. *There is a constant M depending only on N and K such that if $\hat{f} \in E_N(K)$, then*

$$|\hat{f}(\zeta)| \leqslant M\left(1 + |\zeta|^2\right)^{N/2} e^{\|\operatorname{Im}\zeta\| K} \|\hat{f}\|_N \quad \text{for } \zeta \in C^n. \tag{14.16}$$

PROOF. The proof is patterned on that of the theorem on p. 211 of [2]. Set $\hat{g} = (1 + |\xi|^2)^{-N/2}\hat{f}$, so that

$$\|\hat{g}\|_{L^2} = \|\hat{f}\|_N \quad \text{and} \quad f = (1 - \Delta)^{N/2} g$$

If $\psi \in C_0^\infty(R^n)$ is $e^{i\langle x,\zeta\rangle}$ on a neighborhood of K, then, as f has support in K,

$$\hat{f}(\zeta) = (2\pi)^{-n/2}\langle f, \psi\rangle = (2\pi)^{-n/2}\left\langle g, (1 - \Delta)^{N/2}\psi\right\rangle,$$

and hence

$$|\hat{f}(\zeta)| \leqslant (2\pi)^{-n/2}\|\hat{f}\|_N \|(1 - \Delta)^{N/2}\psi\|_{L^2}.$$

Thus, it suffices to show that for each fixed ζ, ψ can be chosen to be $e^{i\langle x,\zeta\rangle}$ on a neighborhood of K, to have support in a compact set independent of ζ, and to satisfy

$$|D^N\psi(x)| \leqslant M(1 + |\zeta|)^N e^{\|\operatorname{Im}\zeta\| K} \tag{14.17}$$

with a constant M independent of ζ, D^N denoting an arbitrary derivative of order at most N.

To this end we fix a function $\varphi \in C_0^\infty(R^n)$ which is 1 on a neighborhood of K and a function $\rho \in C_0^\infty(R^1)$ which is 1 for $t \geqslant -1$ and is 0 for $t \leqslant -2$, and set

$$\psi(x) = \varphi(x)\rho(\langle x, \eta\rangle + \|\eta\|_K) e^{i\langle x,\zeta\rangle}, \quad \text{where } \eta = \operatorname{Im}\zeta.$$

On K we have $-\langle x, \eta\rangle \leqslant \|\eta\|_K < \|\eta\|_K + 1$, so on a neighborhood of K we have $\langle x, \eta\rangle + \|\eta\|_K > -1$, and therefore $\rho(\langle x, \eta\rangle + \|\eta\|_K) = 1$. Thus ψ is $e^{i\langle x,\zeta\rangle}$ on a neighborhood of K, and because of φ it has support in a compact set independent of ζ. It is plain that the derivatives of ψ satisfy (14.17) with the exponential replaced by $e^{-\langle x,\eta\rangle}$. However, ψ is identically 0 unless $\langle x, \eta\rangle + \|\eta\|_K \geqslant -3$, hence unless $-\langle x, \eta\rangle \leqslant \|\eta\|_K + 3$, in which case

$$e^{-\langle x,\eta\rangle} \leqslant e^3 e^{\|\eta\|_K}.$$

THEOREM 14.18. *$E_N(K)$ is complete and $\hat{C}_0^\infty(K)$ is a dense subspace.*

PROOF. The completeness is obvious from Lemma 14.15. To see that $\hat{C}_0^\infty(K)$ (the space of Fourier transforms of functions in $C_0^\infty(K)$) is dense, note first that if $\hat{f}_\rho(\zeta) = \hat{f}(\rho\zeta)$, $\rho < 1$, then $\hat{f}_\rho \to \hat{f}$ in $E_N(K)$ as $\rho \to 1$. Therefore it is enough to approximate functions $\hat{f}$ which satisfy

$$|\hat{f}(\zeta)| \leqslant M\left(1 + |\zeta|^2\right)^{N/2} e^{\rho\|\operatorname{Im}\zeta\|_K}$$

for an arbitrary but fixed $\rho < 1$. In this case the inverse Fourier transform has support in ρK, so it can be regularized without getting outside of K.

LEMMA 14.19. *If K has diameter $\leqslant 2\delta$ and $\hat{f} \in E_N(K)$, then*

$$\left|\frac{D^k\hat{f}(0)}{k!}\right| \leqslant M\delta^{|k|}\left(1 + \frac{|k|^2}{\delta^2}\right)^{N/2}\left(\frac{k}{e}\right)^{-k}\|\hat{f}\|_N \tag{14.20}$$

where M is the constant in (14.16).

PROOF. Evaluate the derivative by Cauchy's formula, integrating over the circles $|\zeta_j| = k_j/\delta$ and using the fact that

$$\|\operatorname{Im}\zeta\|_K \leqslant \delta|\zeta| \leqslant \delta\sum|\zeta_j|.$$

15. More on the polynomials p_m. In this final section we go back to the polynomials p_m that can arise via formula (11.3) from some square integrable function f with support in a given n dimensional compact convex symmetric set K. The promised improvement of Theorem 11.7 is now obvious.

THEOREM 15.1. *For each m, let p_m be a homogeneous polynomial of degree m. There is a square integrable f with support in the unit ball such that $p_m(\theta) = p_m(f;\theta)$ if and only if*

(a) *$|p_m(\xi)| \leqslant c|\xi|^m$ for $\xi \in R^n$.*

(b) *The sum $\sum i^{-m}p_m(\xi)/m!$ is C^∞ and square integrable on R^n.*

PROOF. The sum in (b) obviously satisfies the conditions of Theorem 14.3.

THEROEM 15.2. *Let $p_0, \ldots, p_M$ be any finite sequence of polynomials with p_m homogeneous of degree m. Then there is a function $f \in C_0^\infty(K)$ such that $p_m(f;\theta) = p_m(\theta)$.*

PROOF. Let P be the sum of the p_m, and choose N large enough so that $P \in E_N(K)$. Let F be the subspace of $E_N(K)$ with zero Taylor coefficients through order N. By Lemma 14.19, F is closed, and of course it has finite codimension. By Theorem 14.18, $\hat{C}_0^\infty(K) + F$ is both closed and dense, so it is equal to $E_N(K)$. Consequently, we can find $\hat{f} \in \hat{C}_0^\infty(K)$ and $\hat{f}_0$ in F so that $\hat{f} + \hat{f}_0 = P$. This means that the Taylor series of $\hat{f}$ begins with P, which is the assertion of the theorem.

Another way of stating Theorem 15.2 is as follows.

COROLLARY 15.3. *If U is any neighborhood of 0 and P is any polynomial, then there is an $f \in C_0^\infty(U)$ such that the Taylor series for f begins with P.*

To end the article, let us go back to "practical" matters. Suppose that a certain number M of x-rays are taken with the resulting radiographs $g_{\theta_1}, \ldots, g_{\theta_M}$. Because of the noise involved the question arises as to whether the g_{θ_j} are in fact the radiographs of any object at all. Formula (11.3) provides the necessary condition that for each m the numbers

$$\alpha_{jm} = \int g_{\theta_j}(t) t^m \, dt \tag{15.4}$$

must be the values of a homogeneous polynomial p_m at the points θ_j. Now, the values of a homogeneous polynomial of degree m can be prescribed arbitrarily at any $m + 1$ points of the sphere so this condition is vacuous for $m \geqslant M - 1$, and the set of these conditions becomes finite. The question is whether this finite set of conditions is sufficient. (It is assumed, of course, that $\theta_i \neq \pm \theta_j$.)

THEOREM 15.5. *Let $g_{\theta_1}, \ldots, g_{\theta_M}$ be square integrable and vanish for $|t| \geqslant 1$, and suppose that for $m \leqslant M - 2$ the numbers α_{jm} in (15.4) are the values of a homogeneous polynomial p_m of degree m. Then for each $r > 1$ there is a square integrable f with support in $B(0; r)$ such that $R_{\theta_j} f = g_{\theta_j}$.*

PROOF. By Theorem 15.2 it can be assumed that $\alpha_{jm} = 0$ for $m \leqslant M - 2$. Fix one of the θ_j's, say θ_M, and write $x = (x', t)$ with $x' \perp \theta_M$. Since $\alpha_{nm} = 0$ for $m \leqslant M - 2$, it follows that $\hat{g}_{\theta_M}(\tau) = \tau^{M-1} \hat{h}(\tau)$ where h is square integrable and vanishes for $|t| \geqslant 1$.

Let q be a homogeneous polynomial of degree $M - 1$ such that $q(\theta_j) = 0$ for $j \leqslant M - 1$ and $q(\theta_M) = 1$, and choose $\psi \in C_0^\infty(R^{n-1})$ with integral 1 and vanishing for $|t| \geqslant \varepsilon$. Now set

$$(2\pi)^{(n-1)/2} \hat{f}(\zeta', \tau) = q(\zeta', \tau) \hat{h}(\tau) \hat{\psi}(\zeta').$$

Then $\hat{f}(0, \tau) = \hat{g}_{\theta_M}(\tau)$ and

$$\left| \hat{f}(\zeta', \tau) \right| \leqslant c' e^{|\tau| + \varepsilon |\zeta'|}$$

so that f has support in $\{(x', t): |x'| \leqslant \varepsilon, |t| \leqslant 1\}$, which is contained in any desired ball of radius > 1. Moreover, $\hat{f}(\tau\theta_j) = 0$ because $q(\theta_j) = 0$. Thus we have produced an f with support in the ball such that

$$R_{\theta_M} f = g_{\theta_M} \quad \text{and} \quad R_{\theta_j} f = 0 \quad \text{for } j < M.$$

Doing this for each of the θ_j and taking the sum, we get the required f.

COROLLARY 15.6. *A finite number of noisy radiographs read at a finite number of points cannot be inconsistent.*

PROOF. Suppose that radiographs from directions $\theta_1, \ldots, \theta_M$ are read at points $t_1, \ldots, t_N$, producing readings β_{jk}, $j = 1, \ldots, M$, $k = 1, \ldots, N$. Choose any homogeneous polynomials $p_0, \ldots, p_{M-2}$. It is obvious that for each j we can find a function $g_{\theta_j} \in L^2$ such that

$$g_{\theta_j}(t_k) = \beta_{jk} \quad \text{and} \quad \int g_{\theta_j}(t) t^m \, dt = p_m(\theta_j), \qquad m = 0, \ldots, M - 2.$$

The theorem now provides an f with $R_{\theta_j} f(t_k) = \beta_{jk}$.

The g_{θ_j} can be chosen from any specified subspace of L^2 of dimension at

least $(M - 1)N$, and hence the f from any specified reconstruction subspace F of L^2 such that $F - \cap N_{\theta_j}$ has dimension at least $(M - 1)N$. Here again, however, theory and practice come into conflict. It seems to be useful to choose the number N of readings quite large so that there is redundancy (and with it the possibility of inconsistency) in the equations for the reconstruction. In our reconstructions we have been taking F of dimension 2500 (50×50 matrices), $M = 18$, and $N = 200$.

References

1. I. Amemiya and T. Ando, *Convergence of random products of contractions in Hilbert space*, Acta. Sci. Math. (Szeged) **26** (1965), 239–244. MR **32** #4570.

2. W. F. Donoghue, *Distributions and Fourier transforms*, Academic Press, New York and London, 1969.

3. E. Durand, *Solutions numériques des équations algébriques*. Tome II, Masson, Paris, 1960, p. 120. MR **24** #B1754.

4. R. Gordon, R. Bender and G. T. Herman, *Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography*, J. Theoret. Biol. **29** (1970), 471–481.

5. R. B. Guenther, 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. U.S.A. **71** (1974), 4884–4886. MR **50** #6547.

6. C. Hamaker and D. C. Solmon, *The angles between the null spaces of x-rays*, J. Math. Anal. Appl. (to appear).

7. G. N. Hounsfield, *Computerized transverse axial scanning (tomography)*. I: *Description of system*, Brit. J. Radiol. **46** (1973), 1016–1022.

8. P. D. Lax and R. S. Phillips, *The Paley-Wiener theorem for the Radon transform*, Comm. Pure Appl. Math. **23** (1970), 409–424. MR **42** #8189.

9. Donald Ludwig, *The Radon transform on euclidean space*, Comm. Pure Appl. Math. **69** (1966), 49–81. MR **32** #8064.

10. B. Malgrange, *Existence et approximation des solutions des equations ∂ux derivees partielles et des equations de convolution*, Ann. Inst. Fourier (Grenoble), **6** (1955–1956), 271–355. MR **19**, 280.

11. R. M. Mersereau and A. V. Oppenheim, *Digital reconstruction of multidimensional signals from their projections*, Proc. IEEE **62** (1974), 1319–1338.

12. P. F. J. New and W. R. Scott, *Computed tomography of the brain and orbit*, Williams and Wilkins, Baltimore, Maryland, 1975.

13. K. T. Smith and D. C. Solmon, *Lower dimensional integrability of L^2 functions*, J. Math. Anal. Appl. **51** (1975), 539–549.

14. K. T. Smith, S. L. Wagner, R. B. Guenther and D. C. Solmon, *The diagnosis of breast cancer in mammograms by the evaluation of density patterns*, Radiology (to appear).

15. D. C. Solmon, *The x-ray transform*, J. Math. Anal. Appl. **56** (1976), 61–83.

Department of Mathematics, Oregon State University, Corvallis, Oregon 97331