COMPUTERIZED TOMOGRAPHY: THE NEW MEDICAL X-RAY TECHNOLOGY

L. A. SHEPP AND J. B. KRUSKAL

Abstract. Computerized X-ray tomography is a completely new way of using X-rays for medical diagnosis. It gives physicians a more accurate way of seeing inside the human body and permits safe, convenient, and quantitative location of tumors, blood clots and other conditions which would be painful, dangerous, or even impossible to locate by other methods. Although each tomography machine costs hundreds of thousands of dollars, hundreds of tomography machines are already in use.

A mathematical algorithm to convert X-ray attenuation measurements into a cross-sectional image plays a central role in tomography. Sophisticated mathematical analysis using Fourier transforms has led to algorithms which are much more accurate and efficient than the algorithm used in the first commercial tomography machines. We show how some of the algorithms in actual use have been developed. We also discuss some related mathematical theorems and open questions.

1. Introduction. In computerized tomography, X-ray transmission measurements are recorded on a computer memory device rather than on film, and a sophisticated mathematical algorithm is applied. This produces a numerical description of tissue density as a function of position within a thin slice through the body. The physician examines this function by use of visual displays.

In the ordinary medical use of X-rays, the picture is something like a shadow; any feature in line with denser bone tissue tends to be blocked out. In other words, if we could make a great many pictures, each of a thin slice perpendicular to the beam of X-rays, the actual X-ray picture is formed by superposition of all these hypothetical pictures, i.e., it is a kind of "multiple exposure." Computerized X-ray tomography provides a picture of a single thin slice through the body, without superposition. The word tomography is related to the Greek word "tomos" meaning cut or slice.

Imagine a thin slice, say through the head, perpendicular to the main body axis. Several hundred parallel X-ray pencil beams are projected through the head in the plane of this slice, and the attenuation of X-rays in each beam is measured separately and recorded. (In earlier machines a single beam has been used by translating it parallel to itself within the plane; some of the later machines discussed in Section 3 use fan rather than parallel arrays of beams.) Another set of parallel beams is used within the same plane but at an angle of perhaps 1° or so with the first set, and measurements are taken again. The process is repeated until measurements have been taken for a grid covering all directions in the plane. An elaborate calculation then permits approximate reconstruction of the X-ray attenuation density as a function of position within the slice.

In appropriate units, tissue density in the head varies roughly between 1.0 and 1.05 with the exception of bone which has a density of about 2. Some features of medical interest are indicated by variations of density as small as .005. Reconstructing tissue density with adequate accuracy at a sufficiently fine grid of points is thus a challenging project.

Mathematically we may describe the problem as follows. Consider a fixed plane through the body. Let $f(x,y)$ denote the density at the point (x,y), and let L be any line in the plane. Suppose we direct a thin beam of X-rays into the body along L, and measure how much the intensity is attenuated by going through the body. It is easy to see that the logarithm of the attenuation factor is given

---

Lawrence Shepp received his Ph.D. in Mathematics in 1961 at Princeton University under W. Feller and has been at Bell Laboratories, Murray Hill, N.J. since 1962, where he is a member of the Mathematics Center. He has made many contributions to probability theory and received a Paul Lévy prize. He became interested in tomography in 1972. He presented this paper as invited lecturer at the MAA and AMS meetings in Toronto, 1976.

Joseph B. Kruskal received his Ph.D. in mathematics at Princeton in 1954. He has been a member of the Bell Laboratories Mathematics Center since 1958, but has taught at Princeton, Madison, Ann Arbor, Yale, Cambridge (England), and Columbia at various times. He has made contributions to combinatorial mathematics, statistics, the mathematics of psychology, and statistical linguistics, and has been president of the Psychometric Society and The Classification Society (NAB).—*Editors*

approximately by the projection or line integral of f along L,

$$P_f(L) = \int_L f(x,y)ds, \tag{1.1}$$

where s indicates length along L. The formula (1.1) is only an approximation because: (1) it assumes the X-ray beam is infinitely thin, and (2) it assumes the beam is monochromatic, or alternatively that the physical attenuation coefficient is independent of the energies of the different X-ray photons, and (3) it ignores the significant statistical fluctuations due to the limited number of photons actually transmitted during each measurement. None of these three assumptions is quite correct, but the error in using (1.1) can in principle be made arbitrarily small.

Although complete mathematical analysis of the above errors is impossible, the effects of the approximations have been studied by various means. Thus consider the first approximation. Suppose we replace the line integrals $P_f(L)$ by integrals $P_f(S)$ over strips S. It is shown in Appendix 1 that the strip integrals of f are the line integrals of the function $k*f$, the convolution of f with a circularly symmetric kernel k, where k is the Abel transform of the strip shape. Although the strip integral is still only an approximation, it suggests that an X-ray beam of nonzero width will merely produce a slight smoothing in the reconstruction. The last approximation can be analyzed with the techniques of signal-in-noise communication theory [**20**]. All three approximations have been studied using simulations [**21**].

The mapping $f \to P_f$ in (1.1) is known as the Radon transform because Radon (1917) studied it extensively. He showed that if f is continuous and has compact support, then f is uniquely determined by the values of $P_f(L)$ for all lines L (i.e., the Radon transform is one-one). Furthermore, Radon gave the fairly simple inversion formula, (2.1) below, for f.

Around 1970, G. N. Hounsfield of EMI, Ltd. invented the first computerized tomography machine to give an image accurate enough to be of value in medical diagnosis [**11**]. Since that time computerized X-ray tomography has assumed great medical importance. It is interesting to note that several years prior to Hounsfield's work, Bracewell and Riddle [**1**] used the tomography principle in radioastronomy, while Cormack [**2a**, **2b**] proposed it for medical use.

Mathematics has played a central role in tomography from the very beginning, because without a mathematical algorithm of some sort, reconstruction of the density f from its projections $P_f(L)$ is impossible. Rigorous deduction plays an important but limited role in these algorithms. Without it, tomography would be far less effective. For example, Radon's work, which appears to have been pure mathematics carried out for its own sake, has played a very important role in the development of tomography. Lebesgue-Stieltjes integration theory provided the foundation on which the present formulation of algorithms is based. Classical theorems by Fourier, Poisson and others play an important role as we shall illustrate. On the other hand, when comparing the value of different algorithms, rigorous deduction hardly enters the picture. For example, consider one important aspect of reconstruction, the accuracy to which the density values are reconstructed. In principle, it might be possible to give rigorous error bounds and these could be used to compare the accuracy of different algorithms. However, this has not yet been possible (and might not even provide a useful comparison). Instead the accuracy is compared by several types of experiments which we shall describe in a moment.

It is interesting to consider the method by which algorithms have been "justified" mathematically in this field. While this method consists of mathematical reasoning in a certain sense, the reasoning is far from rigorous. Approximations are introduced at many steps with only intuition as a guide to the error involved. We do not know of a single instance in which a tomographic algorithm has been justified in a truly rigorous sense. Thus, in contrast to some other workers in this field, we do not feel that one derivation is more rigorous than another, whether it is based on Radon's inversion formula, the Fourier inversion formula or any other foundation.

It is interesting to see how non-rigorous mathematical thinking has played the major role in comparing different algorithms. The most widely used technique for comparing algorithms has been to compare the reconstructions when applied to data taken from real human subjects. Another technique is to use what physicians call "phantoms": this means taking data from a physical object of known structure instead of a human subject. This is useful because we know what the true object is. Errors in the reconstruction, however, may be due to errors in the data or to errors in the algorithm. To separate these, Shepp introduced a "mathematical phantom." As described more fully in [**20**], this involves simulating a body section by a mathematically describable function. It has been convenient to use piecewise constant functions where the pieces are made from circles or ellipses (since for these the projection data is easily obtained). In a mathematical phantom there is no measurement error, so any errors in the reconstruction are due to the algorithm. Furthermore, any desired type of measurement error can be simulated to study its effects, which has been very useful [**21**] in the design of the newer high speed scanning machines.

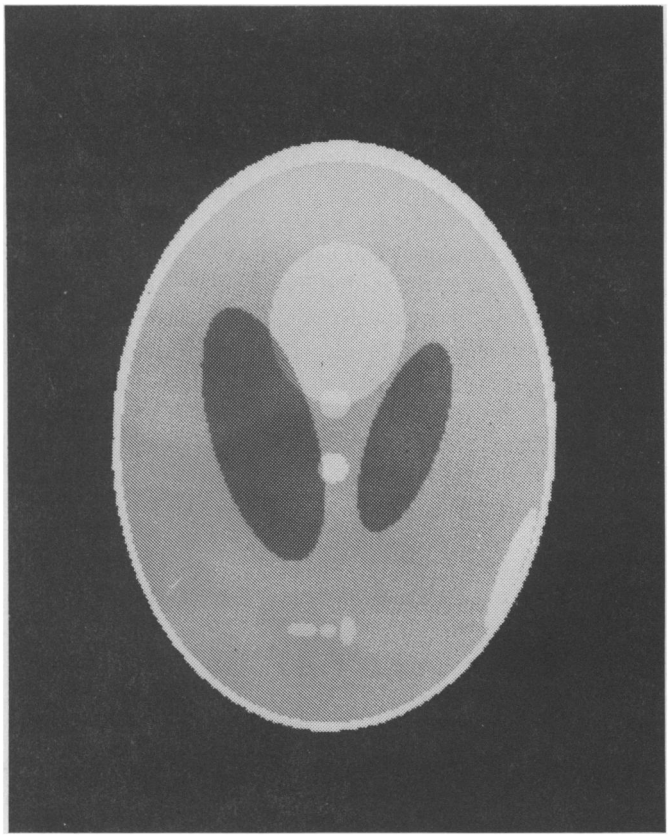

FIG. 1. Simulation of human head using 11 ellipses. The density of the skull is 2.0 and of the ventricles, tumors, etc. is 1.0-1.05 (see [**20**] for more details).

Figure 1 shows a mathematical phantom which is a reasonable imitation of a slice through the human head, subject to the limitations mentioned above. The skull has density 2.0 while the diagonal ellipses represent ventricles of the brain which are filled with fluid of density 1.0. The surrounding gray matter has density 1.02; the various tumors, and the blood clot just inside the skull, have densities ranging between 1.03 and 1.05. Thus all interior head tissue has a very narrow range of variation. See [**20**] for further details.

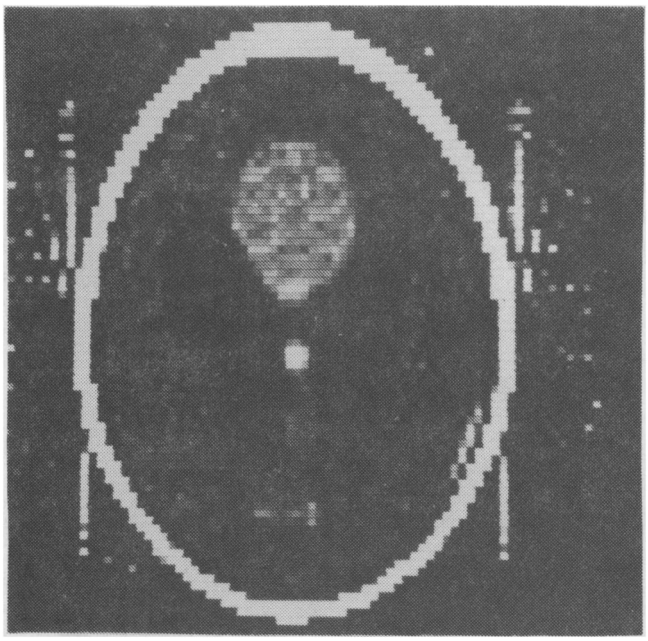

FIG. 2. Reconstruction using the algorithm embodied in the first commercial machine (EMI Ltd.) from 180×160 strip projection data obtained by exact calculation from Fig. 1.

Figures 2,3,4 show reconstructions of this phantom by three different algorithms in chronological order. Figure 2 shows a reconstruction using the algorithm embodied in the first commercial machine, which was constructed and sold by EMI Ltd. We shall call this the Hounsfield iterative algorithm (Hounsfield 1971) since it was based on an iterative relaxation method. Figure 3 shows the reconstruction from the same data by the Fourier-based convolution algorithm due to Shepp (1974). Figure 4 shows the reconstruction by the algorithm now used in EMI machines and due to C. Lemay (1974) of EMI. It is evident that Figs. 3 and 4 are great improvements over Fig. 2 which has artifacts which make it difficult to detect the small tumors. It also has a ring around the inside of the skull, which was also observed in human reconstructions and was believed to be a genuine aspect of human anatomy. Through the use of mathematical phantoms it was first learned that this ring is an artifact due to the algorithm. Figure 4 also shows this artifact, much decreased, and some tomographers argue that such "overshoot" is desirable because it enhances some otherwise less notable features of importance, such as edges of tumors. Other tomographers feel otherwise, pointing out that the overshoot can conceal a feature immediately adjacent to the skull such as a blood clot. We feel that at this stage of tomography the goal should be to make the tomographs as accurate as possible. Enhancement may have value, but it also has dangers which are better deferred until the field has matured further.

The comparison shown in the figures is not quite fair for various reasons, most notably because Fig. 4 is based on a finer grid of simulated X-ray beams than Fig. 2. (However, Fig. 3 is based on the same data as Fig. 2.) Nevertheless, there is little question that the conclusions are correct, because they are supported by a wealth of other experiments, using various algorithms and all three methods of comparison described above.

Some experimental details may be worth mentioning. The reconstructions in Figs. 2 and 4 were made by placing the simulated projection data from the mathematical phantom in the appropriate memory area of an EMI machine (at Columbia-Presbyterian Hospital, NYC) so that the machine could process the data through its algorithm and display the reconstruction in its normal manner. The reconstruction in Fig. 3 was made at Bell Laboratories using a computer program due to Shepp [**20**].

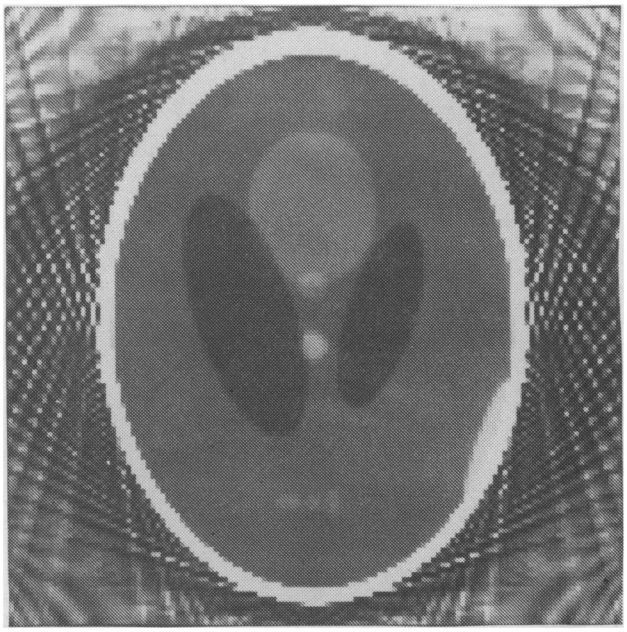

FIG. 3. Reconstruction from the same data using the Fourier based algorithm of Shepp [**20**] (see [**20**] for more details).

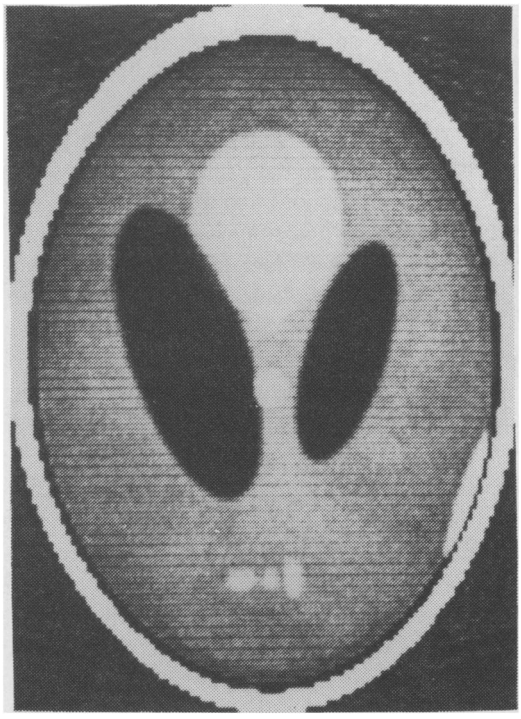

FIG. 4. Reconstruction using the algorithm now embodied in the EMI machine from 180×239 strip projection data obtained by exact calculation from Fig. 1.

In §2, we survey the mathematical foundations and some rigorous models of tomography and state some theorems of mathematical interest which have grown out of tomography. In §3 we briefly describe some of the major tomographic algorithms now in use.

Another survey [22] of the field of tomography has recently appeared and deserves comparison. [22] treats some interesting mathematical questions related to tomography, but has little to do with the way mathematical algorithms are being used in the field.

2. Mathematical foundations. Radon [18] gave a simple formula to invert the transform (1.1). Assume the projections $P_f(L)$ are given for all lines L where f is continuous with compact support. If Q is any point in the plane, denote by $F_Q(q)$ the average value of $P_f(L)$ over all lines L at distance $q>0$ from Q. Then for any Q, f is reconstructed by

$$f(Q)=-\frac{1}{\pi}\int_0^\infty \frac{dF_Q(q)}{q}, \tag{2.1}$$

where the integral converges as a Stieltjes integral in spite of the apparent singularity at $q=0$.

Some pure mathematicians seem to have the mistaken notion that a formula like (2.1) is a complete answer to the tomographic reconstruction problem, and that little more remains to be done. However, it will come as no surprise to most applied mathematicians that the work has just begun. First, an inversion formula is not enough. The stability properties of the inverse mapping are of vital importance since if the inverse mapping is not sufficiently stable then impractically precise measurements of $P_f(L)$ may be required to determine f to sufficient accuracy. Second, in practice it is necessary to use sums rather than integrals and this discretization involves subtle and difficult questions.

To illustrate the latter point first, the natural discrete approximations to (2.1) do not give good reconstructions. This is partly because of the singularity at $q=0$ but mainly because finding a good natural approximation to $F_Q(q)$ is possible only if there are many lines L available at distance exactly q from Q. This occurs only for a few points Q and for a few distances q, no matter how the lines L are chosen. Thus another inversion formula for the mapping $f\to P_f$ is needed which lends itself better to discrete approximations.

To illustrate the point about stability of the inversion mapping, let us consider a more general inversion problem. Since the Radon transform uses all lines L, while discrete approximations deal with only a finite number of lines, it is natural to consider how the inversion depends on the set $\mathcal{L}$ of available lines. Furthermore, the density functions $f(x,y)$ must have smoothness properties, so it is natural to specify that f belongs to some linear space $\mathcal{F}$ of functions, and to consider how the reconstruction depends on $\mathcal{F}$.

Now consider three different sets of lines: the set of all lines, the set $\mathcal{L}_D$ of all lines which intersect a disk D and the set $\mathcal{L}'_D$ of all lines which are exterior to D. We shall refer to the complete Radon transform, the interior Radon transform, or the exterior Radon transform, according to whether L ranges over the first, second, or third of these sets. Then Radon's theorem (2.1) shows that the complete Radon transform is a 1-to-1 mapping. That is, P_f determines f everywhere if f belongs to $\mathcal{F}_\infty$, the set of functions which are continuous and have compact support. The complete Radon transform is even 1-to-1 on the larger set $\mathcal{F}_k$, the set of continuous functions f with $f(x,y)=o\left((x^2+y^2)^{-k}\right)$ as $(x,y)\to\infty$, $k>1$, [18].

Inversion of the interior or exterior Radon transform is of medical interest. For example, the fewer measurements required for the interior transform could reduce patient dosage in imaging a specific organ or part of the body within D. Inversion of the exterior transform could solve the problem caused by rapid heart motion in imaging the chest region around the heart. If there were an accurate inverse exterior transform, the measurements through the beating heart could simply be omitted.

Let us first consider inversion of the interior transform. There is an elementary "solution" which is often rediscovered, sometimes called "old tomography" because there are older devices (not very

successful) which used this method. Old tomography merely forms the average $P_f(L)$ over all lines L which go through each point (x,y). This however gives not f itself but a smoothed version of f, namely

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x+u,y+v)\frac{du\,dv}{\sqrt{u^2+v^2}}. \tag{2.2}$$

Old tomography is actually a particular algorithm for reconstructing f at a point Q from its parallel projections in each of n directions by merely averaging the line integrals $P_f(L)$ over the lines L nearest to Q in each direction. Figure 5 was obtained in this way and shows that this algorithm does not give good reconstructions. It is easy to see that the interior Radon transform does not determine f inside D, i.e., there are nonzero functions f with zero interior Radon transform. Nevertheless, practical inversion of the interior transform may be possible for certain more limited classes $\mathfrak{F}$.

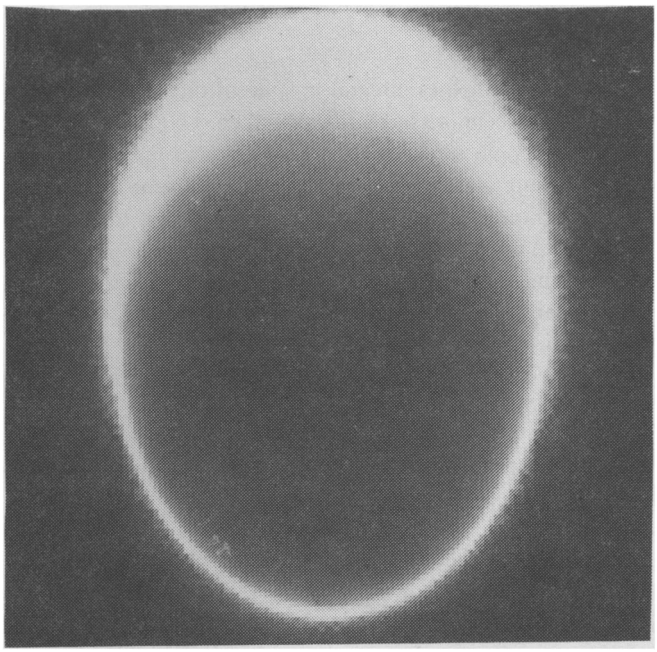

FIG. 5. Reconstruction using the algorithm of "old tomography" which gives for each point Q the average of that line integral in each direction whose line is nearest to Q. There are 128 lines in each of 64 directions and the line integrals were obtained by exact calculation from Fig. 1.

For the exterior Radon transform, theorems of Helgason [8] and Ein-Gal [4] show that the exterior Radon transform is 1-to-1 on $\mathfrak{F}_\infty$ and these authors provide explicit inversion formulas for $f\in\mathfrak{F}_\infty$. However, the inversion mapping for the exterior Radon transform has a property which suggests to us that a practically useful numerical inversion may never be possible. In particular let us extend the Radon transform to the class of functions $\mathfrak{F}_k$ for $k>1$ as above. The complete Radon transform remains 1-to-1 over $\mathfrak{F}_k$. However, an example of D. J. Newman below shows that the exterior Radon transform is not 1-to-1 for any k. In fact the same example shows that if D_1 is a disk containing D, there exist $f\in\mathfrak{F}_k$ with zero exterior transform and which are arbitrarily small outside of D_1 but large in D_1-D. Of course this does not constitute a proof that exterior inversion is unstable in some sense for $\mathfrak{F}_\infty$ itself, but it certainly suggests trouble. What is needed is a study of the stability of the inversion formulas of [8] and [4] in the presence of noise in the measurements.

Newman's examples of $f\in\mathfrak{F}_k$ for which $P_f(L)=0$ for $L\in\mathfrak{L}'_D$ are constructed as follows: Let

$$f(x,y)=\operatorname{Re}\frac{1}{z^n},\quad z=x+iy,\qquad n\geqslant 2. \tag{2.3}$$

For all lines L not passing through $z=0$,

$$\int_L f(x,y)ds = \mathrm{Re}\, c \int_L \frac{1}{z^n} dz = 0, \tag{2.4}$$

because, along L, $ds = cdz$ (ds and dz are proportional) and $z^{-n}dz$ is an exact differential. It is easy to modify f inside a disk D containing $x=y=0$ so as to make f continuous. Note $f \in \mathcal{F}_k$ for $n > k$.

Another example of $\mathcal{L}$ and $\mathcal{F}$ where a mathematical inversion formula exists, but which is believed to be impractical due to instabilities, occurs with $\mathcal{L} = \mathcal{L}_1 =$ the set of all lines which make an angle of say $\leqslant 1°$ with the x-axis. Here again for $f \in \mathcal{F}_\infty$ explicit inversion is possible but the inversion involves a process of analytic continuation. Measuring $P_f(L)$ for only those $L \in \mathcal{L}$, would present significant engineering and cost advantages if a practical realization of the analytic continuation inversion were possible but this seems unlikely, again because of noise instabilities. What is needed but not available is a theory which would assign a measure of stability to the various cases above. We shall return to this point at the end of §2.

For a restricted set of lines $\mathcal{L}$ there are usually many $f \in \mathcal{F}$ with the given projections. However, a unique reconstruction may be obtained by placing an additional criterion of optimality on f. Thus Marr [**16**] has considered $\mathcal{L} = \mathcal{L}_n =$ the (finite) set of $n(n-1)/2$ lines joining all pairs of n equally spaced points on the circumference of the unit disk D, and $\mathcal{F}_m =$ the set of polynomials in x and y of degree m. For $m \leqslant n-2$, Marr gives a formula for the polynomial $g \in \mathcal{F}_m$ which minimizes the sum-of-squares error,

$$\sum_{L \in \mathcal{L}_n} \left(P_f(L) - P_g(L)\right)^2. \tag{2.5}$$

Although Marr's model is realistic in the sense that $\mathcal{L}_n$ is finite, it does not seem desirable in practice to restrict g to be a polynomial. Also if $m > n-2$, g is not unique since (2.5) can be made zero in many ways.

Moving closer to the situation of central interest, let $\mathcal{L}$ consist of all lines at one of n distinct angles $\theta_1, \ldots, \theta_n$. Let $L_{t,\theta}$ be the line with equation

$$x\cos\theta + y\sin\theta = t \tag{2.6}$$

and use $P_f(t,\theta)$ for convenience to denote $P_f(L_{t,\theta})$. We assume $f \equiv 0$ outside a disk D which we take for convenience to be the unit disk. Consequently, $P_f(t,\theta) = 0$ for $|t| \geqslant 1$. We want to keep $\mathcal{F}$ reasonably wide so as not to force an inappropriate structure on f, and hence use $\mathcal{F} = L^2(D)$. Denoting the given projection data by $P_j(t)$, f must satisfy the basic equations

$$P_f(t,\theta_j) = P_j(t), \quad -1 < t < 1, \quad j = 1, \ldots, n. \tag{2.7}$$

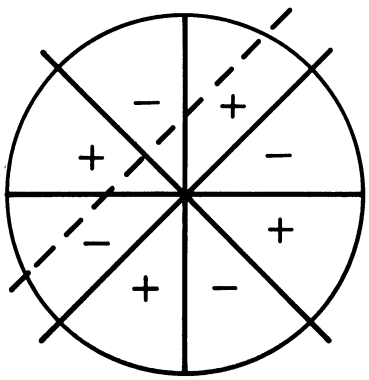

FIG. 6. Example of a function which projects to zero in each of the 4 directions, 0°, 45°, 90°, 135°.

We see from Fig. 6 that if there is one solution f to this equation there will be many solutions. (For a stronger assertion of nonuniqueness of f, see [22].) Therefore it is necessary to impose an optimality condition on f in order to obtain a unique inversion formula. One natural condition which leads to interesting results is to select among the solutions of (2.7) that solution g which has minimum possible L^2-oscillation, i.e.,

$$\int_D\int (g(x,y)-\bar{g})^2\,dx\,dy = \text{minimum} \tag{2.8}$$

where $\bar{g}$ is the average of g over D. Note that $\bar{g}$ is determined from the known projections $P_j(t)$ since

$$\bar{g}=\frac{1}{\pi}\int_D\int g\,dx\,dy=\frac{1}{\pi}\int_{-1}^{1}P_j(t)\,dt, \qquad j=1,\ldots,n. \tag{2.9}$$

From (2.9) it is easy to see that (2.8) is equivalent to the simpler condition

$$\int_D\int g^2\,dx\,dy = \text{minimum}. \tag{2.10}$$

This formulation of the problem has the advantage that the optimal g is computable explicitly in terms of P_j, at least for the case of equally spaced angles θ_j. The solution was found by Logan and Shepp [15] and the same problem was posed and solved independently by O. Tretiak [23]. To describe the solution we introduce the useful notion of a ridge function due to B. F. Logan, which is simply a function constant along lines perpendicular to the direction θ,

$$\rho(x,y)=\rho(x\cos\theta+y\sin\theta), \qquad (x,y)\in D. \tag{2.11}$$

Let $\mathcal{R}$ be the subspace of $L^2(D)$ formed by sums of ridge functions in the directions $\theta_1,\ldots,\theta_n$. It is proved in [15] for the case of equally spaced angles θ_j that the solution to (2.7) and (2.10) has the form

$$g(x,y)=\sum_{j=1}^{n}\rho_j(x,y), \tag{2.12}$$

where ρ_j is a ridge function in direction θ_j. In fact, g is the orthogonal projection of any f satisfying (2.7) onto $\mathcal{R}$. Further, in the case of equally spaced angles, it is possible to obtain an explicit expression for ρ_j. For this purpose introduce the change of variables,

$$h_j(\tau)=\rho_j(\cos\tau)\sin\tau, \qquad 0<\tau<\pi \tag{2.13}$$

and expand h_j and P_j into sine series

$$h_j(\tau)=\sum_{\omega=1}^{\infty}\hat{h}_j(\omega)\sin\omega\tau, \tag{2.14}$$

$$P_j(\cos\tau)=\sum_{\omega=1}^{\infty}\hat{\pi}_j(\omega)\sin\omega\tau. \tag{2.15}$$

This allows a complete diagonalization and an explicit solution for $\hat{h}_j(\omega)$ in terms of $\hat{\pi}_k(\omega)$, $k=1,\ldots,n$ for each $\omega=1,2,\ldots$, i.e., $\hat{h}_j(\omega)$ depends only on $\hat{\pi}_k(\omega)$, $k=1,\ldots,n$ for the same value of ω. This leads to an expression for each ridge function ρ_j of g in terms of convolutions with two fixed kernels operating on the projections. Unfortunately each ridge function involves all projections, which makes the formula apparently impractical (except perhaps for circularly symmetric functions).

For any $f\in L^2(D)$, there is of course a unique polynomial $R(x,y)$ of degree $n-1$ which best approximates f in the L^2 sense on D. Curiously every function f satisfying (2.7) has the *same* best fitting polynomial R. In fact this polynomial R is merely the first n terms (in ω) of the expansion of g generated by (2.13)-(2.15). This gives a fairly explicit algorithm for the optimal polynomial $R(x,y)$ of degree $n-1$. Unfortunately, for functions f of medical interest, R does not give sufficiently good approximations to be of value.

It was asserted in [15] without proof that the space $\mathcal{R}$ of sums of ridge functions is a *closed* subspace of $L^2(D)$. This follows from the explicit inversion formula of [15] for the case of equally

spaced angles, but as several readers of [15] pointed out, is not proved in general in [15]. Fortunately the general result has recently been proved by Hamaker and Solmon [7].

B. F. Logan [14] proved a very interesting theorem whose practical implications for tomography are not yet entirely clear, but which deserve further study. Very roughly the theorem states that knowledge of the full projections in each of n directions is sufficient to reconstruct f up to but not beyond bandwidth n. A little more precisely, Logan shows that a function in $L^2(D)$ of essential bandwidth $n(1-\varepsilon)$ can be essentially reconstructed from any n views. On the other hand, he proves that there exist functions of essential bandwidth $n(1+\varepsilon)$ which project to zero in any n given directions. Still more precisely, suppose that n distinct directions $\theta_1,\dots,\theta_n$ are given and that $\hat{f}$ denotes the Fourier transform of $f\in L^2(D)$. The concentration of the energy of f in the frequency band of radius ρ is defined as the ratio

$$\lambda(f;\rho)=\int_{u^2+v^2\leq\rho^2}\int|\hat{f}\ (u,v)|^2\,du\,dv/\int_{u^2+v^2<\infty}\int|\hat{f}\ (u,v)|^2\,du\,dv. \tag{2.16}$$

To measure how n projections limit the concentration, let

$$\lambda_n(\rho)=\lambda_n(\rho;\theta_1,\dots,\theta_n)=\sup_g\lambda(g;\rho), \tag{2.17}$$

where the sup is taken over all $g\in L^2(D)$ which project to zero in the given directions θ_j, i.e., for which $P_g(t,\theta_j)\equiv 0$. Logan proves that $\lambda_n(\rho)$ is independent of $\theta_1,\dots,\theta_n$. He also proves that $\lambda_n(\rho)$ rapidly increases around $\rho=n$,

$$\lim_{n\to\infty}\lambda_n(n(1+\varepsilon))=\begin{cases}1 & \varepsilon>0\\ 0 & \varepsilon<0\end{cases}. \tag{2.18}$$

Actually much more precise estimates of $\lambda_n(\rho)$ are obtained in [14] but (2.18) will suffice for our purposes. Logan shows further that if $\lambda(f;\rho)$ is near one and $\lambda_n(\rho)$ is near zero, then n projections of f suffice to make a good L^2 estimate of f. On the other hand, by definition, if $\lambda_n(\rho)$ is near one then there exist $f\in L^2(D)$ which are well-concentrated in frequency to a circle of radius ρ, but which project to zero in the given directions and hence cannot be reconstructed.

Thus, roughly speaking, to reconstruct functions f of bandwidth ρ one needs $n=\rho$ projections of f. Unfortunately it has not yet been established what value of ρ is needed for functions f of medical interest.

It seems very surprising that the bandwidth does not depend on the angles θ_j but only on how many different angles there are. This suggests the possibility that one could use n angles within a very narrow range, which would limit the motion of the X-ray tube and allow much faster data acquisition. This would have many benefits, including perhaps the possibility of imaging the beating heart. However, it seems quite likely that any algorithm for this purpose (i.e., which would give the same resolution as in the case of equally-spaced angles) would require unrealistically precise estimates of $P_f(L)$. But, as long as such limitations have not been firmly established the narrow angle approach remains a tantalizing possibility, deserving further mathematical quantification.

3. Algorithms for practical use. In practice, measurements $P_l=P_f(L_l)$ are made for a finite set $\mathfrak{L}$ of lines, L_l, and it is desired to find an approximation $\bar{f}_q$ to $f(Q_q)$ for a grid of points Q_q, generally taken to be a square grid. It is not surprising that the properties of the set $\mathfrak{L}$ play an important role. For so-called "parallel-mode" tomography machines, which include all the earlier machines, $\mathfrak{L}$ consists of many equally-spaced *parallel* lines at each of many equally-spaced angles. Most of our discussion will be devoted to algorithms suitable for parallel-mode machines.

Some of the newer machines have adopted a different mode of scanning, in order to reduce the scanning time from minutes to seconds. In such "fan-beam machines, $\mathfrak{L}$ consists of several "fans" of lines. Each fan consists of many lines through a single focal point. The focal points lie on a circle concentric to the disk on which f is supported, which has a radius R typically about 3 times that of D.

The parallel-mode set of lines can be regarded as a limiting ($R\to\infty$) special case of the fan-beam mode set, but we shall only touch lightly on algorithms for fan-beam machines.

There are two major types of algorithms for reconstruction, iterative and convolutional. The first uses an iterative procedure [11], [6], [20] which updates the current estimate of the density using each projection measurement in turn and which converges to the desired solution after about 5 or 10 complete cycles through every measurement. The first successful tomographic reconstruction algorithm (Hounsfield, 1972) used an algorithm of this type. A convolution algorithm (in the parallel mode) forms the density estimate by applying to the set of measurements a linear mapping which has a certain property, namely, that the coefficient which weights the measurement $P_f(L_l)$ in estimating the density $f(Q_q)$ is a function ϕ only of the distance from Q_q to L_l. Many different weight functions, or filters, ϕ have been suggested [1], [11], [20], [2]. One important advantage of the convolutional algorithm over the iterative algorithm is in speed of computation, since experience shows that an iterative algorithm usually takes about i times longer, where i is the number of complete iteration cycles. At one time the iterative procedures were asserted by many people [11], [5], [6], [17] to have advantages in accuracy. It was also commonly believed that there was little to be gained by improvements in the algorithm, and the iterative procedures were considered attractive because they are inherently "digital" or "discrete," or because they can in principle be used interactively for as many iterations as desired (although they are seldom used this way). It is now generally agreed, however, that the convolutional algorithms are not only faster but give reconstructions with much better accuracy and spatial resolution than the earlier iterative procedures.

Recently still further light has been shed on the comparison between iterative and convolutional algorithms. Experiments by Shepp, which were confirmed by others [9], have pinpointed the chief reason for inaccuracy in the early iterative algorithms [11], [6], and he has given an iterative algorithm [20] whose accuracy is roughly comparable with that of the better convolutional algorithms. To understand the source of the earlier inaccuracy, note that the iterative algorithms attempt to find a function f which is piecewise constant on the squares in a grid, and which has the given projections. The X-ray beam is represented by a strip rather than a line (see Appendix) and each strip gives an equation in the unknown values of f on the grid squares. The equations are solved by Gauss-Seidel relaxation. It turns out that in setting up the equation for each strip it is important to use the actual area of intersection of the strip with each grid square rather than an average value approximation to this area as described in [11] or the even cruder approximation used in [6], where weight one or zero is assigned to each square, according as its center is or is not inside the strip. That the exact coefficients are important did not become clear until the experiments by Shepp with mathematical phantoms [20].

Another limitation of the iterative procedure has to do with spatial resolution. To avoid excessive computation time, iterative procedures have been used only with grids up to 80×80, while the more efficient convolutional algorithms typically use 160×160 or larger. This is a real increase in spatial resolution and has great practical significance. We suspect that even if computation time did not limit the iterative procedures in this way, the necessity of having more equations than unknowns in iterative methods might prevent them from achieving the same spatial resolution as the convolutional algorithms. However, the question is now moot and iterative algorithms are no longer used in commercial machines. Of course it must be acknowledged that for certain oscillatory functions f and noisy measurements of $P_f(L_l)$, the iterative method produces a reconstruction which is closer in some sense to f than the convolution reconstruction.

The convolutional algorithms are fast because advantage can be taken of the special structure of the linear mapping: Direct computation of

$$\bar{f}_q = \sum_{l=1}^{N_L} c_{ql} P_f(L_l), \qquad q = 1,\ldots,N_Q, \tag{3.1}$$

for a general set of weighting constants c_{ql} would involve $N_Q N_L$ multiplications and additions, where

N_Q and N_L are the numbers of grid points and measurement lines. Since typical values are $N_Q = 160 \times 160 = 25{,}600$ and $N_L = n \cdot m =$ (number of directions) $\times$ (numbers of lines or strips per direction) $= 180 \times 160 = 28{,}800$, the direct computation via (3.1) would be rather time consuming. The convolutional algorithm, as is clear from the explicit Fortran program [**20**], involves fewer than $4m^2n + 4N_Q n$ multiplications and additions. In the convolution method c_{ql} is a function ϕ

$$c_{ql} = \phi(d_{ql}) \tag{3.2}$$

of the distance, d_{ql}, from Q_q to L_l. The linear mapping (3.1) is then carried out in 2 stages. The first stage is to perform convolutions. Each convolution is over the set of parallel projections in one of the directions with the function ϕ. The convolution is stored and the second stage is to back-project each convolution to a ridge function as will be described below. The back-projections are summed to obtain the reconstruction. (In the fan-beam case a convolution is carried out for each fan, but the back-projection is not a ridge function but depends also on the distance of Q_q to the focal point of the fan. The calculation and use of this distance makes the fan-beam computation somewhat more difficult than the parallel-beam case.)

There have been two major approaches to choosing the weight function ϕ in the convolutional algorithms. One of them, which is based on the Fourier inversion formula, we refer to as the Fourier approach. The other approach seeks to compensate for the smoothing introduced by old tomography. Because this approach results in seeking to give good approximate reconstructions of a δ-like function, we call this approach the δ-function approach. The Fourier approach has a long history, going back at least to [**3**]. Further contributions were made in [**1**], [**19**], [**20**]. In [**20**] the earlier works were unified and generalized by focusing attention on the role of ϕ ([**1**] and [**19**] had proposed specific but different ϕ's) which can be thought of as a filter function, in analogy with problems in communication theory and antenna design. The δ-function approach is due to Cho [**2**] and was later independently rediscovered and extended [**13**]. While both approaches lead to more or less equivalent algorithms, we feel the Fourier approach has provided a framework in which ϕ may be varied conveniently. One wants to vary ϕ in order to: (a) diminish artifacts (see the discussion in Chapter 1 of the artifact inside the skull in Fig. 4); (b) trade-off between density and spatial resolution [**20**]; (c) sharpen the smoothing effect (Appendix A) resulting from the X-ray beam having nonzero width (the beam width cannot be decreased without losing X-ray flux). The Fourier approach allows one to limit the search for a good ϕ to those with an appropriate Fourier transform, i.e., which satisfy (3.13), below.

In estimating the density at Q_q, should the weight given to the measurement $P_f(L_l)$ depend only on the distance from L_l to Q_q? This seems natural enough intuitively; why should two measurements at the same distance from Q be given different weights? It is also suggested by Radon's formula (2.1) which can be interpreted roughly speaking to say that $f(Q)$ is the integral of $P_f(L)$ over every L with a weight function which depends only on the distance q between Q and L. This intuition led Shepp to the idea that it would be no loss to restrict attention to algorithms which have this property in discrete form. This means that the weight c_{ql} in (3.1) is a function only of the distance d_{ql} from Q_q to L_l as in (3.2). As we shall see below, this restriction has provided helpful guidance in choosing an algorithm for tomographic machines operating in the parallel mode. However, the emergence of "fan-beam" machines, and the important advance in fan-beam algorithms by Lakshminarayanan [**12**] has shown that this restriction excludes valuable algorithms.

The general theory of parallel-mode algorithms having "function-of-the-distance" form due to Shepp and Logan [**20**], has proved very useful. Both this theory and the Lakshminarayanan fan-beam algorithm are obtained by suitable (nonrigorous) approximation from the Fourier reconstruction formula. This rigorous formula for reconstructing a function from all its projections is an alternative to Radon's formula (2.1), and is almost as old ([**24**] seems to be the earliest reference, but see also [**3**]). To obtain this formula define the two-dimensional Fourier transform $\hat{f}$ of f in polar coordinates (ω, θ) by

$$\hat{f}(\omega,\theta)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)e^{-i\omega(x\cos\theta+y\sin\theta)}dx\,dy \tag{3.3}$$

and write the Fourier inversion formula,

$$f(x,y)=\left(\frac{1}{2}\pi\right)^2\int_0^{\pi}\int_{-\infty}^{\infty}\hat{f}(\omega,\theta)e^{i\omega(x\cos\theta+y\sin\theta)}|\omega|d\omega\,d\theta. \tag{3.4}$$

This formula can be obtained from the ordinary Fourier inversion formula by changing from cartesian to polar coordinates (ω,θ). The factor $|\omega|$, which will be very important, comes from the Jacobian of the transformation. Next we express $\hat{f}(\omega,\theta)$ in terms of P_f by changing variables in the definition, and simplifying. Set in (3.3)

$$\begin{aligned} t&=x\cos\theta+y\sin\theta\\ s&=-x\sin\theta+y\cos\theta \end{aligned} \tag{3.5}$$

and note the Jacobian of this rotation is 1, so

$$\begin{aligned}\hat{f}(\omega,\theta)&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)e^{-i\omega t}ds\,dt\\ &=\int_{-\infty}^{\infty}P_f(t,\theta)e^{-i\omega t}dt=\hat{P}_f(\omega,\theta),\end{aligned} \tag{3.6}$$

where $\hat{P}_f(\omega,\theta)$ is the one-dimensional Fourier transform of $P_f(t,\theta)$ with respect to t. Substituting this in the Fourier inversion formula, we obtain

$$f(x,y)=\left(\frac{1}{2\pi}\right)^2\int_0^{\pi}\int_{-\infty}^{\infty}\hat{P}_f(\omega,\theta)e^{i\omega(x\cos\theta+y\sin\theta)}|\omega|d\omega\,d\theta. \tag{3.7}$$

If there were a function ϕ whose Fourier transform $\hat{\phi}$ had the form

$$\hat{\phi}(\omega)=|\omega|, \tag{3.8}$$

we could use Parseval's identity,

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{P}(\omega)\hat{\phi}(\omega)e^{i\omega\tau}d\omega=\int_{-\infty}^{\infty}P(t)\phi(\tau-t)dt \tag{3.9}$$

with $\tau=x\cos\theta+y\sin\theta$ in (3.7) to obtain

$$f(x,y)=\frac{1}{2\pi}\int_0^{\pi}\int_{-\infty}^{\infty}P_f(t,\theta)\phi(x\cos\theta+y\sin\theta-t)dt\,d\theta. \tag{3.10}$$

Note this has the function-of-the-distance form since $x\cos\theta+y\sin\theta-t$ is the distance from (x,y) to $L(t,\theta)$.

Unfortunately, if ϕ is an "honest" function, its Fourier transform $\hat{\phi}(\omega)$ tends to zero as $\omega\to\infty$, so (3.8) is impossible. However, by using a function ϕ whose Fourier transform $\hat{\phi}$ approximates $|\omega|$ in a suitable sense we may be able to use (3.10) as an approximate formula. We want to insure that the error in (3.10), namely

$$\left(\frac{1}{2\pi}\right)^2\int_0^{\pi}\int_{\infty}^{\infty}\hat{P}_f(\omega,\theta)e^{i\omega(x\cos\theta+y\sin\theta)}\left(\hat{\phi}(\omega)-|\omega|\right)d\omega\,d\theta \tag{3.11}$$

is small. Now it is plausible to assume that

$$\hat{f}(\omega,\theta)\text{ is small for large }\omega\text{ (say }|\omega|>\Omega) \tag{3.12}$$

since this is a smoothness assumption on f, i.e., it means that f contains little high frequency energy. In the region where $\hat{f}(\omega,\theta)\equiv\hat{P}_f(\omega,\theta)$ is small, we can afford a large discrepancy between $\phi(\omega)$ and $|\omega|$. Thus if we weaken (3.8) to

$$\hat{\phi}(\omega)\approx|\omega|\quad\text{for small }\omega\text{ (say }|\omega|<\Omega), \tag{3.13}$$

then (3.10) may be a good approximation. Of course the same Ω must be used in (3.12) and (3.13).

The earlier tomography machines all operate in the "parallel mode." Without going into the details

this means that the measurements $P_f(t,\theta)$ are available for equally-spaced parallel lines at each angle θ_j. If a is the spacing between adjacent parallel lines then t takes on the values ka for $k=0,\pm 1,\pm 2,\ldots,\pm 1/a$, for each angle. In practice, equally-spaced angles are always used, so $\theta_j=(j-1)\pi/n, j=1,2,\ldots,n$. Then the natural discrete approximation to (3.10) is

$$f_\phi(x,y)=\frac{a}{2n}\sum_{j=1}^{n}\sum_{k=-\infty}^{\infty}P_f(ka,\theta_j)\phi(x\cos\theta_j+y\sin\theta_j-ka). \tag{3.14}$$

Of course there are only finitely many terms in the inner sum since $P_f(ka,\theta_j)=0$ for $ka>1$ because f is supported on the unit disk. (3.13) gives a simple algorithm for reconstructing f which depends upon the choice of ϕ.

How ϕ should be chosen is not completely understood except for guiding principles. Intuitively, it seems likely that if the spacing between the parallel lines is not fine relative to the high frequencies in $f(x,y)$, it will not be possible to reconstruct f very well. Since the spacing between the grid lines is a (a itself is chosen to achieve a desired resolution), we may assume that f has little energy above the Nyquist frequency of π/a. Thus we may plausibly take $\Omega\approx\pi/a$. This still leaves a lot of room for choice of ϕ. One possibility, suggested by Bracewell and Riddle (1956), who were the first to use the Fourier method in an algorithmic context, is

$$\hat{\phi}(\omega)=\begin{cases}|\omega|, & |\omega|<\Omega\\ 0, & |\omega|>\Omega\end{cases}. \tag{3.15}$$

However, this choice does not work well, apparently because the corresponding $\phi(t)$,

$$\phi(t)=\Omega\frac{\sin\Omega t}{\pi t}-\frac{1-\cos\Omega t}{\pi t^2} \tag{3.16}$$

only decays like $1/t$ for large t so that the weight given by (3.14) to a measurement line L far away from (x,y) is still too large. Intuitively it seems undesirable to permit such long range effects. If $\hat{\phi}(\omega)$ is absolutely continuous, then ϕ will decay at least as fast as $1/t^2$ for large t. On the other hand, if $\hat{\phi}'(\omega)$ has the same discontinuity at $\omega=0$ as $|\omega|$ (this is expected if (3.13) holds) then it can be expected that $\phi(t)$ will decay no faster than $1/t^2$.

In order that (3.14) should be a good approximation to f, the inner sum in (3.14) should be a good approximation to the inner integral in (3.10), i.e., we should have

$$\int_{-\infty}^{\infty}P_f(t,\theta)\phi(u-t)dt\approx\sum_{k=-\infty}^{\infty}P_f(ka,\theta)\phi(u-ka)a. \tag{3.17}$$

However, a function ϕ satisfying (3.13) for $\Omega\approx\pi/a$ has the shape shown in Fig. 7. Note that there are only one or two sampling points in the main lobe of ϕ. Under such sparse sampling one ordinarily would not expect the Riemann sum in (3.17) to be a good approximation to the integral. Fortunately, however, the sampling is regular and the Poisson summation formula (identity)

$$\sum_{k=-\infty}^{\infty}\hat{\psi}\left(\frac{2\pi k}{a}\right)=a\sum_{k=-\infty}^{\infty}\psi(ka) \tag{3.18}$$

may be applied. In (3.18) we use

$$\psi(t)=\psi_u(t)=P_f(t,\theta)\phi(u-t) \tag{3.19}$$

and note that the right side of (3.18) is the right side of (3.17) whereas the left side of (3.17) is the $k=0$ term of (3.18). Therefore the error in (3.17) consists of the $k\neq 0$ terms on the left of (3.18). Using (3.6),

$$\begin{aligned}\hat{\psi}_u\left(\frac{2\pi k}{a}\right)&=\frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{\phi}\left(\omega-\frac{2\pi k}{a}\right)\hat{P}_f(\omega,\theta)e^{i\omega u}d\omega\\ &=\frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{\phi}\left(\omega-\frac{2\pi k}{a}\right)\hat{f}(\omega,\theta)d\omega\end{aligned} \tag{3.20}$$

since the transform of a product is the convolution of the transforms. We want (3.20) to be small for $k \neq 0$.

By (3.12) $\hat{f}(\omega,\theta)$ may be large for $\omega < \Omega$. Indeed, since the X-ray density typically stays fairly constant within many organs we may expect that $\hat{f}(0,\theta)$ is large. To make (3.20) small, we want $\hat{\phi}(\omega-(2\pi k/a))$ to be small whenever $|\omega| < \Omega$ and $k \neq 0$ and particularly small whenever $\omega = 2\pi k/a$.

These arguments may seem crude to those accustomed to pure mathematics, but they are typical of much applied mathematics. The guidance they provide in choosing ϕ has been very helpful. Reconstructions of f with different ϕ's are compared by the method discussed in Chapter 1.

As we shall see later on, in the case of fan-beam projection data, the values of $P_f(t,\theta)$ are only available for unequally-spaced values of t. In this situation the Poisson sum argument used to obtain (3.17) breaks down, and it is not surprising that there seems to be no choice of ϕ which yields good reconstructions. However, by interpolating in t to obtain approximate values of $P_f(t,\theta)$ at regularly spaced values of t, it is possible to obtain good reconstructions. Unfortunately this interpolation leads to significant errors in case the measurements of $P_f(t,\theta)$ are corrupted by drift and gain variation in X-ray detectors. This is a crucial point which will be referred to later in discussing the adaptation of the parallel-mode algorithm to fan-beam machines.

Many convolution algorithms can be thought of in the framework we have described above, e.g., old tomography corresponds to use of the special ϕ where $\phi(t)=1$ for $|t| \leqslant a$ and 0 otherwise. Bracewell and Riddle [1], who gave the first Fourier based convolution algorithm, used (3.16). Ramachandran and Lakshminarayanan [19] used a function ϕ which is the same as (3.16) for $\Omega = \pi/a$ at the points $t = ka$, $k = 0, \pm 1, \ldots$ and is piecewise linear in between. They introduced this function to achieve substantial savings in computation time (since it avoids a great many sine and cosine evaluations required by (3.16)) but assumed that the modification would degrade the image somewhat. Within our framework, however, there is no reason to expect the piecewise linear modification to do worse than the original (see [20] for more details) and experimental results bear this out.

Shepp and Logan [20] were the first to look at the choice of ϕ in a systematic way. They found the principles above to provide good guidance. They also expressed the trade-off relation between spatial and density resolution in terms of the choice of ϕ. The choice (3.16) and the modification both have the property that $\phi(t)$ continues to oscillate even for large t.

To avoid these oscillations, and to achieve the computational efficiency of a piecewise linear filter, a new filter was designed by the following approach. We wish to emphasize the approach, rather than the particular filter which was chosen, since many other choices might be quite as effective. As a source of piecewise linear filters, we may take even functions $\hat{\phi}(\omega)$ which are approximately equal to (3.15) in a suitable sense, calculate their inverse transforms, and then take piecewise linear approximations. To evaluate the piecewise linear filters, we may examine their transforms, the image reconstructions which they yield, and various other considerations. The filter used in [20] was chosen after examining several filters which were generated in this way. It is a piecewise linear approximation to the inverse transform of

$$\hat{\phi}(\omega) = \begin{cases} 2|\sin(\omega a/2)|/a & \text{for } |\omega| \leqslant 2\pi/a, \\ 0 & \text{elsewhere,} \end{cases}$$

a function which was suggested by B. F. Logan. This filter is given by:

$$\begin{cases} \phi(ka) = -\dfrac{4}{\pi^2 a}\dfrac{1}{4k^2-1}, & k = 0, \pm 1, \ldots \\ \phi \text{ piecewise linear in between these values.} \end{cases} \tag{3.21}$$

The piecewise linearity saves substantial computation time (as was first observed in [19] with no apparent loss in image quality). Using any such function ϕ, the calculation of (3.14) may be arranged to be very efficient (see [20] for more details and a simple 60-line Fortran program).

As we mentioned earlier, there is another approach to convolution algorithms which may be

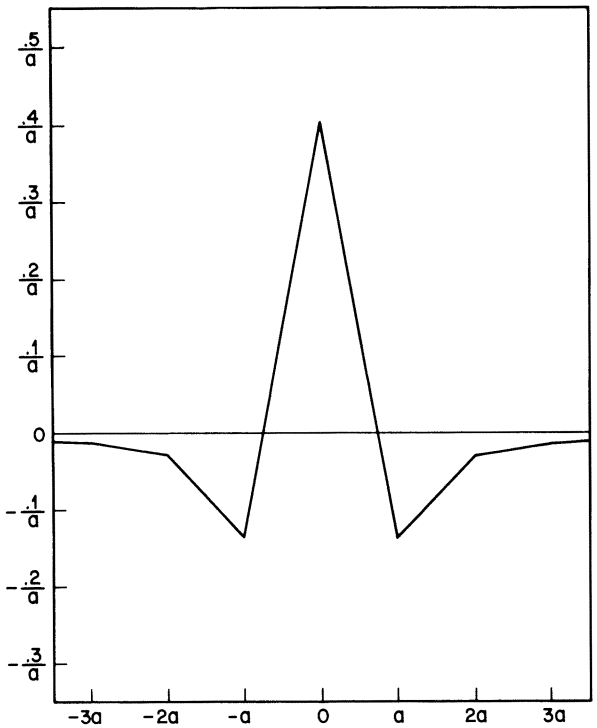

FIG. 7. The reconstruction filter ϕ of (3.21). The algorithm assigns weight $\phi(d)$ to a line L at a distance d from the point of reconstruction. Note that a is the distance between adjacent centerlines of ray measurements. This is typical of reconstruction filters, having positive main lobe and negative side lobes.

discussed in the same framework. This interesting and natural approach, which is due to Z. H. Cho [2] but was also arrived at by LeMay [13], avoids Fourier transforms altogether. Roughly speaking, the idea is to start with "old tomography" and correct its inadequacies. Old tomography forms a ridge function from $P_f(t,\theta_j)$ in the direction orthogonal to θ_j. The ridge function, which is called the "back projection" of $P_f(\cdot,\theta_j)$, has $P_f(\cdot,\theta_j)$ as a cross-section. Old tomography then sums the back projections. The result, as we saw in (2.2), is a smeared version of the desired function f.

To correct this smearing, Cho decided to convolve $P_f(t,\theta)$ with some function $\phi(t)$ before back projecting. By this means he arrived at (3.14) using a much more direct route than the one we described. To achieve the desired effect, one can see that ϕ should have a central peak surrounded by negative "sidelobes."

Roughly speaking, any function $f(x,y)$ can be formed as a linear combination of functions which are sharply peaked at one location and zero elsewhere, in other words, approximate δ-functions. Therefore a linear algorithm which correctly reconstructs an approximate δ-function will correctly reconstruct any density function. Cho used a cylindrical spike S of radius a and height one as an approximate δ-function and chose ϕ so that the average of the back projections

$$\frac{1}{n}\sum_{j=1}^{n}\phi(x\cos\theta_j+y\sin\theta_j)\approx\frac{1}{\pi}\int_0^{\pi}\phi(x\cos\theta+y\sin\theta)\,d\theta \tag{3.22}$$

would approximate S well.

Cho's algorithm [2] also involves several other features of the back projection which are quite outside the simple framework we have discussed. Experiments by Shepp seem to show, however, that these other features do not play an important role in his algorithm, and that the algorithm (3.14) using Cho's function ϕ (modified to be piecewise linear) gives results quite like Cho's. LeMay's algorithm [13] has much in common with Cho's.

The algorithms of [2] and [13] give good results, as illustrated in Fig. 4. However, as we discussed in the introduction, they give a characteristic artifact just inside the skull, which could conceal medically important conditions. Also the functions ϕ which Cho and LeMay obtained are defined in terms of a sequence of numerical values. This type of ϕ is more complicated to use in noise analyses [21] than is the analytically explicit and especially simple ϕ in (3.21). Furthermore, this approach does not provide as much guidance in modifying ϕ to achieve the desirable trade-off of density vs. spatial resolution.

An important virtue of any theoretical framework in which to obtain ideas or algorithms is flexibility in adapting to new situations. Some of the new tomography machines do not operate in the "parallel-mode" mentioned earlier, but in the "fan-beam mode," in order to reduce the X-ray scanning time from minutes to seconds. In this mode $P_f(L)$ is measured for n "fans" of lines L, where the lines in each fan have a common focal point (at an X-ray detector). The n focal points lie on a circle, concentric with the unit circle on which f is nonzero, which has a radius R typically about 3 times as large. (The limiting case as $R\to\infty$ is the parallel mode.)

It is interesting to see how the Fourier approach to convolutional algorithms can be adapted to this situation. While it is possible to regroup the lines into sets of (almost) parallel lines at each of several angles, the parallel lines are not equally-spaced. One approach would be to maintain the restriction to linear algorithms satisfying (3.2), which states that the reconstruction weight for each line L depends only on the distance from L to the point Q of reconstruction. However, as discussed earlier, the Poisson-sum argument breaks down because the parallel lines are not equally spaced. No function ϕ has been found which gives good reconstructions in this situation.

Another approach is to interpolate P_f between the irregularly spaced and not quite parallel lines to obtain pseudomeasurements at an equally spaced set of exactly parallel lines. Although this interpolation is neither simple nor elegant, this method can be made to work reasonably well, subject to one important proviso which makes it quite useless in practice.

The proviso is that if there are any systematic differences between the fans, due to differences between the X-ray detectors with which they are associated, the image is seriously marred by streaks. Unfortunately, it seems difficult to control the drift and gain problems in detector behavior to a sufficient degree of uniformity to avoid this problem.

Fortunately, an elegant effective algorithm has been found by a young biophysicist, A. V. Lakshminarayanan [12], by an approach which uses the fan-beam geometry rather than fighting it. Starting with the Fourier inversion formula (3.7) he made changes of variables to adapt these formulas to the fan-beam coordinates. He then succeeded in making approximations similar to those described above, and obtained an algorithm [12] which is now widely used in fan-beam scanners. Subsequently another derivation for this algorithm, based on Radon's inversion formula and using different approximations, has been found by Herman and Naparstek [10].

Appendix. We study the relationship between the line integral of f

$$P_f(t,\theta)=\int_{-\infty}^{\infty} f(t\cos\theta+s\sin\theta,t\sin\theta-s\cos\theta)\,ds \tag{A.1}$$

along the line $L(t,\theta)$ in (2.6) and the strip integral of f

$$Q_f(t,\theta)=(1/2\delta)\int_{-\delta}^{\delta} P_f(t-u,\theta)\,du \tag{A.2}$$

obtained by integrating f over a strip of width 2δ about $L(t,\theta)$ (and dividing by the width of the strip). We show that (A.2) is actually the line integral of a smoothed version, $f^k(x,y)$, of f obtained by convolving f with a centrally symmetric two-dimensional kernel k,

$$f^k(x,y)=\int_0^{2\pi}\int_0^{\infty} f(x-\rho\cos\alpha,y-\rho\sin\alpha)k(\rho)\rho\,d\rho\,d\alpha. \tag{A.3}$$

In the more general case, where the strip has weight function W, denote the W-weighted strip integral by

$$Q_f^W(t,\theta)=\int_{-\infty}^{\infty} W(u)P_f(t-u,\theta)du. \tag{A.4}$$

Note (A.2) is the special case

$$W(u)=\begin{cases} 1/2\delta; & |u|\leqslant\delta \\ 0; & |u|>\delta \end{cases}. \tag{A.5}$$

We show that the line integrals of f^k are the strip integrals of f, i.e.,

$$P_{f^k}(t,\theta)=Q_f^W(t,\theta) \tag{A.6}$$

where k is the kernel given by the Abel transform of W, i.e.,

$$W(u)=\int_{-\infty}^{\infty} k(\sqrt{u^2+v^2})dv; \qquad k(\rho)=-\frac{1}{\pi\rho}\frac{d}{d\rho}\int_{\rho}^{\infty}\frac{W(u)u}{\sqrt{u^2-\rho^2}}du. \tag{A.7}$$

This shows that regarding actual measurements as line measurements, when of course the beam has a nonzero width, is not serious, in the sense that one is merely reconstructing f^k rather than f, which for thin beams is a reasonable approximation to f. However, we hasten to point out that in actual X-ray measurements neither $P_f(t,\theta)$ nor $Q_f^w(t,\theta)$ is measured (neglecting statistical errors, the nonzero height of the beam, and polychromatic effects) but rather a nonlinear weighting is obtained. Thus if a parallel X-ray beam with profile $W(u)$ is passed through an object of attenuation f, the log of the ratio input/output intensities is given by

$$\log I_{\text{in}}/I_{\text{out}}=-\log\int_{-\infty}^{\infty} W(u)\exp^{-P_f(t-u,\theta)}du. \tag{A.8}$$

If W is δ-function like, i.e., if the beam is very narrow and if $P(t,\theta)$ is smooth in t, then the linearized approximation to (A.8) is (assuming W integrates to unity)

$$\log I_{\text{in}}/I_{\text{out}}\doteq\int_{\infty}^{\infty} W(u)P(t-u,\theta)du=Q_f^W(t,\theta) \tag{A.9}$$

which is a good approximation. This approximation is the spatial analogue of the linearization in energy with a polychromatic beam.

To prove (A.6) and (A.7), note that from (A.1) and (A.3)

$$P_{f^k}(t,\theta)=\int_{-\infty}^{\infty}\int_0^{2\pi}\int_0^{\infty} f(t\cos\theta+s\sin\theta-\rho\cos\alpha,\, t\sin\theta-s\cos\theta-\rho\sin\alpha)k(\rho)\rho d\rho d\theta ds. \tag{A.10}$$

Make the change of variables

$$\begin{aligned} \rho\cos\alpha&=u\cos\theta+v\sin\theta \\ \rho\sin\alpha&=u\sin\theta-v\cos\theta \\ s&=s-v \end{aligned} \tag{A.11}$$

and use the definition of W to obtain (A.6) after a short calculation. The second part of (A.7) is the well-known Abel inversion formula which is easily proved.

From (A.7), for a square beam as in (A.5),

$$k(\rho)=\begin{cases} (1/2\pi\rho)(\delta^2-\rho^2)^{-1/2} & 0\leqslant\rho<\delta \\ 0 & \delta\leqslant\rho. \end{cases} \tag{A.12}$$

For a circular beam of radius δ, from (A.7),

$$W(u)=\begin{cases} (1/\pi\delta^2)(\delta^2-u^2)^{1/2} & |u|\leqslant\delta \\ 0 & u>\delta \end{cases} \tag{A.13}$$

$$k(\rho)=\begin{cases}(1/\pi\delta^2), & \rho<\delta\\ 0, & \rho>\delta\end{cases} \tag{A.14}$$

which is a uniform weight.

Acknowledgments. This project involved close collaboration over several years with Sadek K. Hilal of The Neurological Institute in N.Y.C. to whom we are grateful for access to the EMI machine involved in the experiments described, and with Jay A. Stein of American Science and Engineering, Inc. Many others at Bell Laboratories also deserve explicit thanks but are too numerous to mention. Finally, we are grateful to R. A. Schulz for writing the programs for several of the experiments described.

This work is based on invited talks at the AMS and MAA meetings in Toronto, Canada, August 24, 28, 1976.

References

1. R. N. Bracewell and A. C. Riddle, Inversion of fan-beam scans in radio astronomy, Astro Phys. J., 150 (1967) 427–434.

2. Z. H. Cho, et al, Computerized image reconstruction methods with multiple photon/X-ray transmission scanning, Phys. Med. Biol., 19 (1974) 511–522.

2a. A. M. Cormack, Representation of a function by its line integrals, with some radiological applications, J. Applied Physics, 34 (1963) 2722–2727.

2b. ———, Representation of a function by its line integrals, with some radiological applications, II, J. Applied Physics, 35 (1964) 2908–2912.

3. H. Cramer and H. Wold, Some theorems on distribution functions, J. London Math. Soc., 11 (1936) 209–294.

4. M. Ein-Gal, The shadow transform: An approach to cross-sectional imaging, Stanford Univ. Tech., Report No. 6851-1 (1974).

5. R. Gordon, A tutorial on ART (Algebraic Reconstruction Techniques), IEEE Trans. Nucl. Sci., NS-21 (1974) 78–93.

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

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

8. S. Helgason, The Radon transform in Euclidean spaces, compact two point homogeneous spaces and Grassman manifolds, Acta Math., 113 (1965) 153–180.

9. G. T. Herman, A. Lent, and P. H. Lutz, Relaxation methods for image reconstruction, Comm. Asso. Comp. Mach. (to appear).

10. G. T. Herman and A. Naparstek, Fast image reconstruction based on a Radon inversion formula appropriate for rapidly collected data, SIAM J. Appl. Math., 33 (1977) 511–533.

11. G. N. Hounsfield, A method of and apparatus for examination of a body by radiation such as X or gamma radiation, The Patent Office, London, (1972) Patent Specification 1283915.

12. A. V. Lakshminarayanan, Reconstruction from divergent ray data, SUNY Technical Report Number 92 (1975) Computer Science Department, Buffalo, N.Y.

13. C. A. G. LeMay, A method of and apparatus for constructing a representation of a planar's slice of body exposed to penetrating radiation, U.S. Patent 3 (1974) 924,129.

14. B. F. Logan, The uncertainty principle in reconstructing functions from projections, Duke Math J., (1975) 661–706.

15. B. F. Logan and L. A. Shepp, Optimal reconstruction of a function from its projections, Duke Math. J., (1975) 645–659.

16. R. B. Marr, On the reconstruction of a function on a circular domain from a sampling of its line integrals, J. Math. Anal. Appl., 45 (1974) 357–374.

17. B. E. Oppenheim, More accurate algorithms for iterative 3-dimensional reconstruction, IEEE Trans. Nucl. Sci., NS-21 (1974) 72–77.

18. J. Radon, Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten, Berichte Saechsische Akademie der Wissenschaften, 69 (1917) 262–277.

19. G. N. Ramachandran and A. V. Lakshminarayanan, Three dimensional reconstruction from radiographs and electron micrographs: application of convolutions instead of Fourier transforms, Proc. Natl. Acad. Sci. U.S.A. 68 (1971) 2236–2240.

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

21. L. A. Shepp and J. Stein, Simulated artifacts in computerized tomography. Chapter in book, Reconstructive Tomography in Diagnostic Radiology and Nuclear Medicine, edited by M. Ter-Pogossian, 1974, 33–48.

22. K. T. Smith, D. C. Solmon, and S. L. Wagner, Practical and mathematical aspects of the problem of reconstructing objects from radiographs, Bull. Amer. Math. Soc., 83 (1977) 1227–1270.

23. O. Tretiak, Talk at Brookhaven symposium on computerized tomography, (1974).

24. A. Rényi, On projections of probability distributions, Acta Math Acad. Sci. Budapest, 3 (1952) 131–141.

BELL LABORATORIES, MURRAY HILL, N.J. 07974

THE PLANE SYMMETRY GROUPS: THEIR RECOGNITION AND NOTATION

DORIS SCHATTSCHNEIDER

Introduction. Groups of transformations which leave invariant a specified item are familiar objects of study for students and researchers alike. Finite groups of plane isometries which leave invariant a regular polygon are elementary examples: C_n, the cyclic group of order n, can be realized as the group of rotations leaving invariant a regular n-gon, and D_n, the dihedral group of order $2n$, can be realized as the group of all isometries (rotations and reflections) leaving invariant the same polygon. A very interesting collection of discrete groups of plane isometries which are natural extensions of these examples exists, but is lacking in most introductory algebra texts. These are the groups of plane isometries which leave invariant a design or pattern in the plane. If the pattern is finite, such a group is necessarily a subgroup of some dihedral group. If the pattern is repeated regularly in one or in two directions, translations and glide-reflections are additional possible isometries of the pattern, and so the group leaving such a design invariant will be an infinite discrete group. Designs which are invariant under all multiples of just one translation are frieze, or border ornaments, and their associated groups are commonly called "frieze groups." Patterns which are invariant under linear combinations of two linearly independent translations repeat at regular intervals in two directions, and hence their groups are often termed "wallpaper groups."

The interweaving of elementary aspects of Euclidean transformation geometry and group theory makes these groups excellent ones for study—but there are several non-mathematical bonuses which make their study especially appealing. To analyze a repeating design to see what makes it "work," and to create original designs using the power of the mathematical "laws" which govern these designs, is a strong non-mathematical motive for studying these groups. (Suddenly, the word "symmetry" has true dual meaning; both its artistic and mathematical connotations are seen as inseparable.) Rudiments of elementary crystallography are part of the theory as well—another bonus.

A very specific incentive to learn about these groups is the opportunity to study examples of the imaginative interlocking patterns by the Dutch artist M. C. Escher (1898–1972). His work is perhaps the most concrete testament to the power gained in understanding these groups. He struggled for several years to produce animate interlocking designs, with very primitive results. When he became aware that these types of designs were governed by groups of isometries, he studied the mathematical literature available. In examining Escher's notebooks, this author discovered that he copied in full the

---

Doris Schattschneider received her Ph.D. from Yale University in 1966 (in the area of algebraic groups). After teaching at Northwestern University and the University of Illinois at Chicago Circle, she came to Moravian College in 1968. Here, her interest in art led her to create a January Term course, "Tessellations, the Mathematical Art," in which much of the information for this MONTHLY article was developed. A deepened interest in the work of M. C. Escher was a natural outgrowth of this course. Recently, she has collaborated with a graphic artist to produce a book and a collection of unique geometric models, *M. C. Escher Kaleidocycles*, published by Ballantine Books.—*Editors*.