TOMOGRAPHY

The term *tomography* refers to the general class of devices and procedures for producing two-dimensional (2D) cross-sectional images of a three-dimensional (3D) object. Tomographic systems allow one to view the internal structure of objects in a noninvasive and nondestructive manner. By far the best known application is the computer-assisted tomography (CAT or simply CT) scanner for X-ray imaging of the human body. Other medical devices, including nuclear medicine scanners and magnetic resonance imaging systems, also make use of tomographic principles. Outside the medical realm, tomography is used in applications ranging from microscopy through nondestructive testing and radar imaging to geophysical imaging and radio astronomy. In this article a brief survey of the applications of tomography is presented. We also review the underlying theory of image reconstruction from line integrals and highlight applications that explicitly make use of line-integral measurements. We conclude with a brief discussion of recent and possible future developments.

X-ray Imaging and Motion Tomography

In conventional X-ray radiography, a stationary source and planar detector are used to produce a 2D projection image of the patient (1). This image has an intensity proportional to the amount by which the X rays are attenuated as they pass through the body, that is, the 3D spatial distribution of X-ray attenuation coefficients is projected into a 2D image. The resulting image provides important diagnostic information due to differences in the attenuation coefficients of bone, muscle, fat, and other tissues in the 40 keV to 120 keV range used in clinical radiography. The utility of conventional radiography is limited by the fact that the 3D anatomy is projected into a 2D image, causing certain structures to be obscured. For example, tumors in the lung may be obscured by a more dense rib that projects into the same area in the radiograph.

The earliest examples of tomographic systems were designed to overcome this problem using motion of the X-ray source and planar detector to produce an image of a single 2D section through the patient (2). Motion tomographs work by controlling the source and detector motion so that only points in a single plane of the patient project onto a fixed location in the imaging plane (1). Points that are not in this plane will project onto time-varying locations in the imaging plane. Consequently, the image of out-of-plane structures becomes very blurred due to the motion, and any high-resolution detail in the resulting image must be due to the stationary plane. While improving on conventional radiographs, these systems are limited by the loss in contrast resulting from the superposition, onto the plane of interest, of the blurred images from adjacent planes. Clinical systems of this type, using linear and circular motion, were in use from the 1940s. The development of CT scanners in the 1970s has made these systems virtually obsolete.

X-ray Computerized Tomography

Computerized tomography systems do not form the image directly as in the case of motion tomography. Instead, sampled data are processed by an image-reconstruction algorithm to produce a digital representation of the desired image. A computed tomography system collects measurements of 1D line integrals, or projections, along parallel paths through a 2D slice of a 3D object. By collecting these projections at different angles relative to a fixed coordinate system, we build up the 2D Radon transform from which the image is reconstructed. This function is named after the mathematician J. Radon, who was the first to investigate the properties of the integrals of N-D functions over $(N - 1)$-D hyperplanes (3).

Consider the first clinical X-ray CT system (4) for which the inventor, G. Hounsfield, received the 1979 Nobel prize in medicine (the prize was shared with the mathematician A. Cormack) (5). A collimated X-ray source and detector are translated on either side of the patient as illustrated in Fig. 1(a). The detected X-ray measurements provide a projection through the patient of the 2D distribution of X-ray attenuation coefficients within the plane illuminated by the X-ray source. By rotating the source and detector around the patient other projections are measured. The resulting samples of the Radon transform are then processed using an image-reconstruction algorithm to produce a 2D image of the attenuation coefficients. Since the X-ray beam is confined to the plane of interest, the image does not suffer from the superposition of additional blurred structures as was the case with motion tomographs.

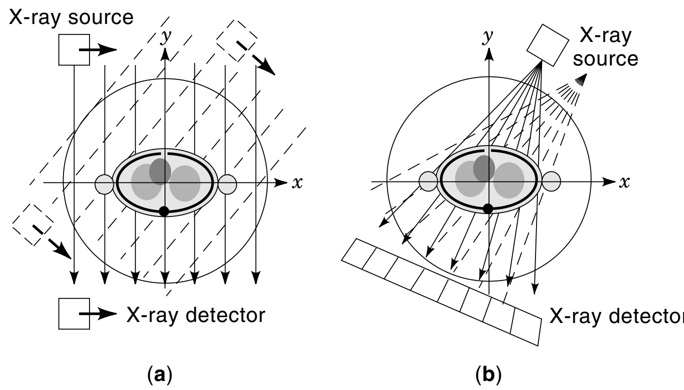

Figure 1. Illustration of the data collection arrangement for parallel and fan-beam tomography. (a) Parallel acquisition requires translation of the source and detector to collect each projection; the process is repeated after rotation to collect additional views. (b) The fan-beam geometry; in this case no translation of the source or detectors is required.

One major limitation of the first generation of CT systems was that the translation and rotation of the detectors were slow and a single scan would take several minutes. X-ray projection data can be collected far more quickly using the fan-beam X-ray source geometry employed in the current generation of CT scanners and illustrated in Fig. 1(b). In this case, an array of detectors is used so that the system can simultaneously collect data for all projection paths that pass through the current location of the X-ray source. In this way, the X-ray source need not be translated, and a complete set of data is obtained through a single rotation of the source around the patient. Using this configuration, modern scanners can scan a single plane in less than 1 s.

Other Applications

Tomographic principles have also been applied in a number of other diagnostic medical imaging instruments. One major application, for which the early developments actually predate X-ray CT, is nuclear medicine imaging (6). Measurements of physiological processes are obtained by tomographic imaging of the spatial distribution of a biochemical tracer or probe that has been tagged with a radioactive isotope. Using these systems with various probes, it is possible to produce 3D images of metabolic activity throughout the body, variations in neurotransmitter and receptor densities in the brain, and the functioning of the heart and blood vessels. In single-photon emission computed tomography (SPECT), γ-ray emitting isotopes are used to tag the probes. A γ camera collects a sequence of planar parallel projections of the 3D tracer distribution as the camera is rotated around the patient. In positron-emission tomography (PET), positron-emitting isotopes are used. Pair of photons produced by the annihilation of a positron, emitted from the tracer, with a nearby electron are detected using scintillation detectors. Using a ring of these detectors it is possible to collect a set of parallel projections of the tracer distribution. Image-reconstruction algorithms similar to those used in X-ray CT are then applied to reconstruct the 3D tracer distribution.

Magnetic resonance (MR) imaging differs from X-ray and emission CT in the sense that the image Fourier transform or "k-space" is measured directly. This is achieved by using a magnetic field gradient to produce a spatial frequency-encoding of the magnetic resonance signal from hydrogen nuclei in the body (7). Using combinations of time-varying magnetic field gradients and radio-frequency pulses, it is possible to obtain k-space measurements with a wide range of sampling patterns. The early MR systems were designed to sample k-space along a set of radial lines (8). One-dimensional Fourier transforms of these samples produce a set of parallel projections. Images can then be reconstructed using the same methods as employed in X-ray CT. Modern MR systems usually collect data on rectangular sample patterns in k-space, and the images are reconstructed directly using discrete Fourier transforms. Fast acquisition can be achieved using more complex k-space sampling patterns such as the spiral scan. In addition to varying the manner in which the Fourier space is sampled, different pulse sequences can be used to alter contrast in the images through varying the impact of spin–spin and spin–lattice relaxation constants on the resonance signal. MR imaging remains a highly active research field with particular interest in the development of methods for dynamic imaging of the beating heart and functional techniques for studying brain activity, blood flow, and other physiological processes. Since the modern MR techniques do not make direct use of line-integral methods we will not consider them further in this article.

Tomographic methods have also proven very powerful in applications other than medical imaging. X-ray CT systems have been widely used for nondestructive testing of manufactured components and materials (9). Tomography has also had an enormous impact in exploring the natural world. Applications range from microscopic imaging using electron micrographs (10) and confocal microscopes (11) to imaging of celestial bodies using radio telescopes (12). Electromagnetic (EM) techniques have been used for resistivity imaging between bore holes in geophysical exploration (13), synthetic aperture radar mapping of the earth and other planets (14), and imaging of ionospherical electron density (15). Similarly, acoustical signals have been used for imaging over a wide range of scales from acoustic microscopy (16) to large-scale mapping of oceanographic temperatures and 3D mapping of the earth's interior using natural seismic data (17).

The following theoretical development is presented from the point of view of X-ray CT and assumes that the radiation follows a straight-line path between the source and detector. For some of the EM and acoustic applications mentioned before, diffractive and other effects are significant so that the radiation is no longer confined to a straight line. Accurate solutions require that these effects be modeled, which can substantially complicate the inverse problem. However, a large number of tomographic problems can be formulated as locally linear inverse problems that admit to iterative solutions of the type described later, if not the closed-form solutions applied in X-ray CT. After presenting the theoretical background to computed tomography, we return briefly to some of these other applications.

THEORY

Conventional Radiography and Motion Tomography

X rays passing through an object experience exponential attenuation proportional to the linear attenuation coefficient of the object. The intensity of a collimated beam of monoenergetic x radiation exiting a uniform block of material with linear attenuation coefficient μ and depth d is given by $I = I_0 e^{-\mu d}$, where I_0 is the intensity of the incident beam. For objects with spatially variant attenuation $\mu(z)$ along the path length z, this relationship generalizes to (1)

$$I = I_0 \exp(-\textstyle\int \mu(z)\,dz) \tag{1}$$

For the case where X rays of different energies are present, an energy dependence should be included in Eq. (1). The theoretical development of CT methods, included that presented here, usually assumes a nonenergetic source. For broadband X-ray sources, the beam becomes "hardened" as it passes through the object, that is, the lower energies are attenuated faster than the higher energies. This effect causes a beam-hardening artifact in CT images that is reduced in practice using a data calibration procedure.

Let $\mu(x, y, z)$ represent the 3D distribution of attenuation coefficients within the human body. Consider the simplified model of a conventional radiography system that has a wide parallel beam of X rays traveling through the patient in the z direction. Assume that a 2D detector array or film in the (x, y) plane has a negative-logarithmic response. The following image would then be formed at an ideal detector:

$$r(x,y) = \textstyle\int \mu(x,y,z)\,dz \tag{2}$$

Since the attenuation coefficients of the body are different for different tissues, the projection image formed according to Eq. (2) can provide useful diagnostic information through exposing internal variations in attenuation coefficients. The limitation of this process is that the image formed projects a 3D distribution $\mu(x, y, z)$ into a 2D image $r(x, y)$, and hence subtle variations in attenuation may be masked in the projection image.

The motion tomography systems described previously attempt to produce an image of a single z plane in the patient through motion of the source and detector. For linear motion restricted to the x direction, the resulting image can be expressed as an integral over time t:

$$f(x,y) = \iint_{-T}^{T} \mu(x - tc_{z_0}(z - z_0), y, z)\,dt\,dz \tag{3}$$

where c_{z_0} is a constant that depends on the distance of the desired plane z_0 from the detector and the rate of translation. Note that only the plane $z = z_0$ projects onto the same position in the imaging plane for the entire imaging period $t \in -[T, T]$. However, structures from other planes are still superimposed, albeit in a blurred form. In contrast, computed tomography forms a sequence of 2D images that represent a reconstruction of a single slice $f(x, y) = \mu(x, y, z_0)$. Images of the patient are formed at different depths z_0 as the patient is translated axially through the CT scanner.

Parallel-Beam Tomography

Consider a 2D image $f(x, y)$. The Radon transform is defined as

$$g(u,\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(u - x\cos\theta + y\sin\theta)\,dx\,dy \tag{4}$$

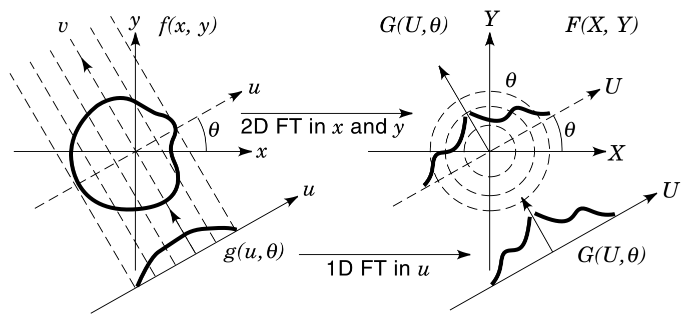

Figure 2. Illustration of the projection slice theorem. The 2D image at left is projected at angle θ to produce the 1D projection $g(u, \theta)$. The 1D Fourier transform, $G(U, \theta)$, of this projection is equal to the 2D image Fourier transform, $F(X, Y)$ along the radial line at angle θ.

where $\delta(u)$ is the Dirac delta function. For any particular value of θ, the Radon transform represents the line integrals of the function along parallel paths ν at angle θ to the y coordinate of the fixed (x, y) coordinate system as illustrated in Fig. 2. The function $g(u, \theta)$ is often referred to as a sinogram since an image consisting of a single point produces a sinusoidal pattern in Radon transform space.

In X-ray CT scanners, these measurements are collected as the logarithm of the ratio of incident to exiting x-ray intensity. By translating the x-ray source and detector along a linear path at angle θ to the x coordinate, we collect the Radon transform measurements at that angle. By rotating the source and detector relative to the patient, parallel projections for different values of θ can be collected.

The inversion formula for reconstructing a function from its projections was originally derived by Radon in 1917 (3) in the following form:

$$f(r,\phi) = \frac{1}{2\pi^2}\int_0^{\pi}\int_{-\infty}^{\infty}\frac{1}{r\cos(\theta-\phi)-u}\frac{\partial}{\partial u}g(u,\theta)\,du\,d\theta \quad (5)$$

where $f(r, \phi)$ is the image represented in polar coordinates. Direct numerical approximations of this inversion formula are rarely implemented. Instead, the convolution in the variable u and the derivative operation can be combined into a single step, resulting in an inversion formula that is equivalent to the filtered backprojection algorithm described below.

Practical inversion methods can be developed through the relationship between the Radon and Fourier transforms. The basic result that is used in developing these methods is the projection slice theorem (18). This theorem states that the 1D Fourier transform of the parallel projection at angle θ is equal to the 2D image Fourier transform evaluated along the radial line at angle θ with the X axis, that is,

$$G(U,\theta) = \int_{-\infty}^{\infty} g(u,\theta)e^{-juU}du = F(X,Y)|_{X=U\cos\theta, Y=U\sin\theta} \quad (6)$$

where $F(X, Y)$ is the 2D image Fourier transform

$$F(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)e^{-jxX}e^{-jyY}\,dx\,dy \quad (7)$$

This result, which is illustrated in Fig. 2, can be employed in a number of ways. The discrete Fourier transform (DFT) of the samples of each 1D projection can be used to compute approximate values of the image Fourier transform. If the angular projection spacing is $\Delta\theta$, then the DFTs of all projections will produce samples of the 2D image Fourier transform on a polar sampling grid with loci at the intersections of radial lines, spaced by $\Delta\theta$, with circles with radii equal to integer multiples of the DFT frequency sampling interval. Once these samples are computed, the image can be reconstructed by first interpolating these values onto a regular Cartesian grid and then applying an inverse 2D DFT. Design of these Fourier reconstruction methods involves a trade-off between computational complexity and accuracy of the interpolating function (19).

A more elegant solution can be found by reworking Eq. (6) into a spatial domain representation. It is then straightforward to show that the image can be recovered using the following equations (20):

$$f(r,\phi) = \int_0^{\pi} \tilde{g}(r\cos(\theta-\phi),\theta)\,d\theta \quad (8)$$

where

$$\tilde{g}(u,\theta) = \frac{1}{2\pi}\int_{-\infty}^{\infty} G(U,\theta)|U|e^{juU}\,du \quad (9)$$

These two equations form the basis of the widely used filtered backprojection algorithm. Equation (9) is a linear shift-invariant filtering of the projection data with a filter with frequency response $h(U) = |U|$. The gain of this filter increases monotonically with frequency and is therefore unstable. However, by assuming that the data $g(u, \theta)$ and hence the corresponding image are band-limited to a maximum frequency $U = U_{\max}$, we need only consider the finite bandwidth filter with impulse response:

$$h(u) = \int_{-U_{\max}}^{U_{\max}} |U|e^{juU}\,dU \quad (10)$$

The filtered projections $\tilde{g}(u, \theta)$ are found by convolving $g(u, \theta)$ with $h(u)$. The integrand in Eq. (8) can be viewed as an image with constant values along lines at angle θ to the y coordinate that is formed by "backprojecting" the filtered projection at angle θ. Summing (or in the limit, integrating) these backprojected images for all θ produces the reconstructed image.

This method, or the modification described later for the fan-beam geometry, is the basis for image reconstruction in almost all commercially available computed tomography systems. Many practical issues in X-ray CT including sampling requirements, calibration procedures, beam-hardening correction, and the treatment of noise are of great importance in achieving high-quality reconstructions but are beyond the scope of this article; see Ref. 21 for an excellent tutorial that discusses these issues.

Fan-Beam Tomography

X-ray CT data can be collected more rapidly using an array of detectors and a fan-beam X-ray source so that all elements in the array are simultaneously exposed to the X rays. This arrangement gives rise to a natural fan-beam data collection

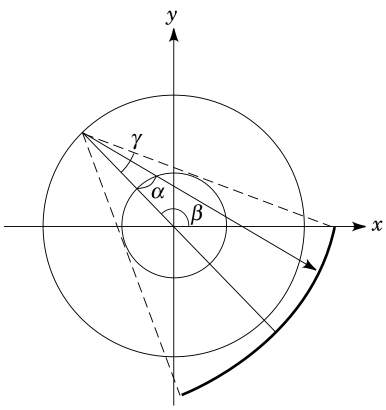

Figure 3. Illustration of the fan-beam geometry. The projection $g(\alpha, \beta)$ corresponds to the line integral along the path from the source to the detector ray at angle α to the line to the center of rotation; the source is rotated by angle β from the x axis. The angle γ represents the largest angular displacement required to collect all line integrals that pass through the object, illustrated here by the inner circle.

geometry as illustrated in Fig. 3. The source and detector array are rotated around the patient and a set of fan beam projections, $g(\alpha, \beta)$, are collected, where β represents the rotation angle of the source and α the angular displacement of the detector in the array relative to the radial line from the source to the center of rotation.

The projection data could be resorted into equivalent parallel projections and the preceding reconstruction methods applied. Fortuitously, this re-sorting is unnecessary. It can be shown (22) that reconstruction of the image can be performed using a fan-beam version of the filtered backprojection method. Development of this inverse method involves substitution of the fan-beam data in the Radon inversion formula, Eq. (5), and applying a change of variables with the appropriate Jacobian. After some manipulation, the equations can be reduced to the form

$$f(r, \phi) = \int_0^{2\pi} \frac{\tilde{g}(\alpha, \beta)}{r'^2}\, d\beta \tag{11}$$

where r' is the distance from the point (r, ϕ) to the fan beam source,

$$\tilde{g}(\alpha, \beta) = \frac{-1}{4\pi^2} \int_{-\gamma}^{\gamma} \frac{g(\alpha', \beta)\cos\alpha'}{\sin^2(\alpha - \alpha')}\, d\alpha' \tag{12}$$

and γ is the maximum value of α required to ensure that data are collected for all line-integral paths that pass through the object. As in the parallel-beam case, this reconstruction method involves a two-step procedure: filtering (in this case a weighted filtering) and backprojection. The backprojection for fan-beam data is performed along the paths converging at the location of the X-ray source and includes an inverse square-distance weighting factor.

Numerous variations on these fan-beam formulas exist that deal with issues such as nonuniform angular sampling of the projection data and modifications to deal with noise (18). It is also interesting to note that there is redundancy resulting from 2π angular coverage in β: The image can be reconstructed from data collected over an angular range of $\pi + 2\gamma$. An example of an X-ray CT image reconstructed from fan beam data collected using a modern CT scanner is shown in Fig. 4.

Iterative Approaches

A limitation to the direct or analytic reconstruction approaches is the implicit assumption that the data are exact line integrals of the image. Furthermore, the presence of noise is typically handled by simply modifying the frequency response of the projection filters. In the case of clinical X-ray CT, the spatial sampling rates are very high, as are signal-to-noise ratios, so that the direct methods produce high-quality reconstructions. Performance degrades, however, in applications where sampling is more restricted, where the data do not conform well to the line-integral model, or where noise levels are significant. An entirely different approach to reconstruction from projections that can be used to good effect in these cases is to model the image using a finite-basis-function expansion with unknown coefficients and formulate the inverse problem as one of solving the large set of simultaneous equations relating these coefficients to the sampled data (23, 24).

Let us assume that the image can be represented with sufficient accuracy as a weighted sum of a set of orthogonal basis functions $\phi_j(x, y)$, $j = 1, \ldots, N$, that is,

$$f(x, y) = \sum_{j=1}^{N} f_j \phi_j(x, y) \tag{13}$$

In the following, as in most cases, the basis functions $\phi_j(x, y)$ are chosen as the set of indicator functions on an array of square pixels that collectively tile the region of support Ω of the image. A single index is used to represent each basis function for notational convenience. For a 256×256 pixel image, N will be 256^2; a similar convention is used to index the sampled projections.

The sampled projection data can then be written as

$$y_i = \iint_\Omega h_i(x, y) f(x, y)\, dx\, dy = \sum_{j=1}^{N} H(i, j) f_j \tag{14}$$

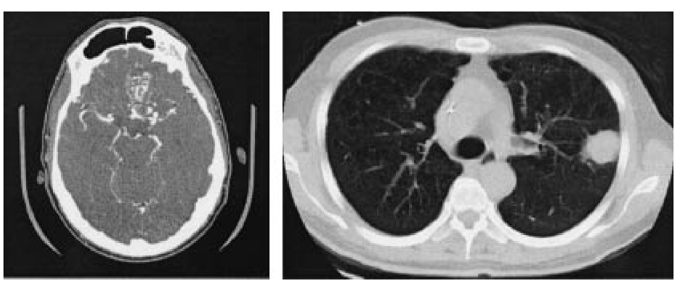

Figure 4. X-ray CT images collected using a GE HiLight spiral scanner (courtesy of Michael McNitt-Gray, M.D., Assistant Professor, Radiology, UCLA). The image on the left shows a cross sectional brain scan of a patient with an aneurism; a contrast agent is used to increase the brightness of the blood vessels in the brain. The image on the right shows a chest scan with a single lesion in the right lung.

where

$$H(i,j) = \iint_{\Omega} h_i(x,y)\phi_j(x,y)\,dx\,dy \tag{15}$$

Here $h_i(x, y)$ represents the line integral kernel from Eq. (4) for the ith sample point. The elements $H(i, j)$ of the projection matrix $\boldsymbol{H}$ contain the integrals of this kernel over the basis functions $\phi_j(x, y)$, which are equal to the length of the intersection of the line along which the integration is performed with the nonzero region of each pixel.

One of the attractions of this formulation is that we are not restricted to the line-integral model in the projection matrix. Most algorithms that are based on the model in Eq. (14) are readily modified to allow the inclusion of physical factors in the projection matrix. Thus, for example, the finite beam width and detector resolution in X-ray CT systems can be modeled by replacing the line-integral model in Eq. (14) with a strip integral in which the image is integrated over the width of the X-ray beam; this case is illustrated in Fig. 5. Further modifications can be included for other applications, such as seismic tomography, where the ray paths are curved (17).

Equation (14) is a huge set of simultaneous linear equations and can, in principle, be solved using standard methods. However, the size of these systems, coupled with the special structure of $\boldsymbol{H}$, has motivated a number of researchers to investigate more efficient specialized numerical procedures. The key property that these methods exploit is that $\boldsymbol{H}$ is highly sparse, that is, most elements in the matrix are zero since the paths along which each integration is performed intersect only a small fraction of the pixels in the image.

One algorithm that makes good use of the sparseness property is the algebraic reconstruction technique (ART) (21). This method finds the solution to the set of equations in an iterative fashion through the successive orthogonal projection of the current image estimate onto the hyperplanes defined by each row of the system of equations. If this procedure converges, the solution will be a point where all of the hyperplanes intersect, that is, a solution to Eq. (14). Let $\boldsymbol{f}^n$ represent the vector of image pixel values at the nth iteration, and $\boldsymbol{h}_i^T$ represent the ith row of $\boldsymbol{H}$. Then the ART method has the following form:

$$\boldsymbol{f}^{n+1} = \boldsymbol{f}^n + \left(\frac{y_i - \boldsymbol{h}_i^T\boldsymbol{f}^n}{\boldsymbol{h}_i^T\boldsymbol{h}_i}\right)\boldsymbol{h}_i \tag{16}$$

ART can also be viewed in terms of the backprojection operator used in the filtered backprojection method: each iteration of Eq. (16) is equivalent to adding to the current image estimate, $\boldsymbol{f}_n$, the weighted backprojection of the error between the ith measured projection sample and that corresponding to $\boldsymbol{f}_n$.

ART will converge to a solution of Eq. (14) provided the system of equations is consistent. If the data are inconsistent or the system is ill-conditioned, then problems with convergence or numerical instability may arise. In these cases, additional information should be introduced. The need for additional information is particularly important in "limited data" problems in which, for example, complete projection views are missing, or the presence of an X-ray opaque object in the field of view obscures parts of each projection.

Other constraints can be introduced by extending the idea of projection onto hyperplanes to the more general method of projection onto convex sets (POCS) (25). In this approach, constraints are introduced in the form of a collection of convex constraints sets, C_k, which represent the set of images that satisfy the kth constraint. A solution to the problem is then found by computing the orthogonal projection of the current image estimate onto each constraint set in turn. Under certain restrictions, including the existence of a non empty intersection of all constraint sets, this method will converge to a point in this intersection.

Inconsistency in the data due to noise or modeling errors can be allowed for by relaxing the constraint that the equations are solved exactly. In some versions of ART, this is achieved using the constraint that the forward projection of the solution differs from the measured data by a maximum of $\pm\delta$, where δ is a small constant (24). For the case where noise is Gaussian, it is more appropriate to constrain the average squared error rather than the error in each measurement. In

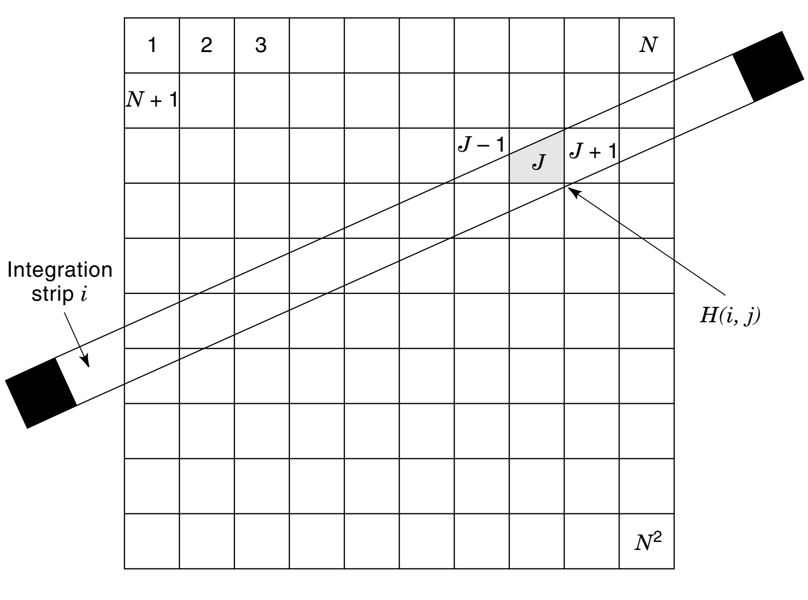

Figure 5. Illustration of the pixel-based finite-dimensional formulation used in the iterative reconstruction methods. The ith measurement is proportional to the areas of intersection of each pixel with the strip that joins the source and detector.

either case, appropriate convex constraints sets are readily defined. Certain properties of the solution are often known independently of the data, for instance, the images are typically non-negative and of finite spatial extent. Again these properties are readily expressed in terms of convex constraints. The general form of the POCS approach makes it attractive for developing general inverse methods for tomographic problems in which the effect of a variety of different constraints can be evaluated.

A limitation of the POCS methods is that if there is more than one solution that satisfies all constraints, then the convergence point is dependent on the initialization of the search. A second problem in both ART and POCS methods is that they can both be sensitive to ill-conditioning in $\boldsymbol{H}$, resulting in unstable solutions that are very sensitive to small changes in the data. Regularization methods can be used to overcome the ill-conditioning problems. The regularizing function resolves ambiguities resulting from ill-conditioning by choosing the solution that minimizes the regularizing function among those that give essentially the same fit to the data. Typically this function is some measure of smoothness or energy in the solution (26).

Consider the case where the regularizing function is chosen as a weighted quadratic norm or semi-norm $\|\boldsymbol{f}\|_W^2 = \boldsymbol{f}^T\boldsymbol{W}^f$ on the solution, and the presence of noise in the data is allowed for by minimizing the squared error in the fit to the data. Then the regularized solution is found by solving the optimization problem:

$$\min_{\boldsymbol{f}} \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{f}\|^2 + \lambda\|\boldsymbol{f}\|_W^2 \tag{17}$$

This problem can be solved using general numerical optimization techniques such as steepest descent or the method of conjugate gradients. The method can also be modified to introduce additional constraints, such as non-negativity, using constrained optimization approaches. Among the common choices of regularizing function are energy ($\boldsymbol{W}$ is the identity matrix) and spatial smoothness functions such as the Laplacian of the image ($\boldsymbol{W} = \boldsymbol{L}^T\boldsymbol{L}$ where $\boldsymbol{L}$ is a finite-difference representation of the spatial Laplacian operator). The most commonly studied nonquadratic regularizing function is probably the entropy measure $E_f = -\sum_j f_j \log f_j$ (26).

Statistical Methods

Tomographic inverse problems can also be formulated in terms of classical estimation techniques using statistical models of the data. We can compute a maximum likelihood (ML) estimate using the finite-dimensional formulation in Eq. (14) where the coefficients $\boldsymbol{f}$ are the unknown parameters. The ML solution is found as the maximum of $p(\boldsymbol{y}|\boldsymbol{f})$, the conditional probability for the data given the image coefficients. A well-known example of an ML method in computed tomography is the EM algorithm for Poisson data developed by Shepp and Vardi (27). This method uses the Poisson distribution to model the photon-limited nature of data acquisition in SPECT and PET instruments for nuclear medicine imaging. The improved modeling of the spatially variant noise process leads to superior performance for the ML method in comparison with direct methods or iterative linear methods such as ART. However, the ML solution in PET and SPECT is usually ill-conditioned resulting in high-variance solutions. In practice this problem can be controlled, to some extent, by early termination of the iterative process.

An alternative way to avoid the instability of the ML method is to reformulate the problem in a Bayesian framework. Spatial random field priors $p(\boldsymbol{f})$ can be used to characterize the statistical properties of the images. The posterior probability for the image conditioned on the data is then given by Bayes theorem:

$$p(\boldsymbol{f}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{f})p(\boldsymbol{f})}{p(\boldsymbol{y})} \tag{18}$$

The most widely used class of priors in Bayesian tomography are the Markov random field (MRF) models (28). Their densities are conveniently expressed in terms of a Gibbs energy function, which is a sum over a set of potential functions, each of which is usually taken to be a function of only a few pixels. These potential functions can be chosen to reflect the locally smooth property of many images. The existence of sharp intensity changes, corresponding to the edges of objects in the image, can also be modeled using more complex MRF models. The Bayesian formulation also offers the potential for combining data from multiple modalities. For example, high-resolution anatomical X-ray CT or MR images can be used to improve the reconstructions of functional images from low-resolution PET or SPECT data (29).

Bayesian estimators in tomography are usually of the maximum *a posteriori* (MAP) type (30). The MAP solution is given by maximizing the posterior probability $p(\boldsymbol{f}|\boldsymbol{y})$ with respect to $\boldsymbol{f}$. For each data set, the denominator of the right-hand side of Eq. (18) is a constant, so that the MAP solution can be equivalently found by maximizing the log of the numerator, that is,

$$\max_{\boldsymbol{f}} \ln p(\boldsymbol{y}|\boldsymbol{f}) + \ln p(\boldsymbol{f}) \tag{19}$$

An example of the improvement that can be achieved in the quality of a reconstructed PET image using a Bayesian approach, rather than direct reconstruction, is shown in Fig. 6.

Comparing Eq. (19) to Eq. (17), we see a strong parallel between regularized methods and MAP estimators. The conditional probability term, $\ln p(\boldsymbol{y}|\boldsymbol{f})$, plays the role of the (first) data matching term in Eq. (17); the prior fulfills the role of

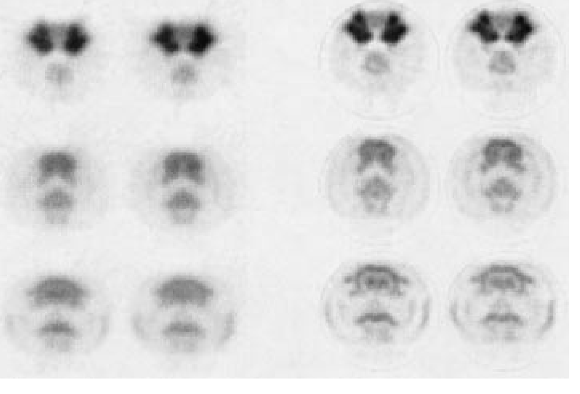

Figure 6. Comparison of coronal PET images of glucose metabolism in a rat brain reconstructed using filtered backprojection with maximum resolution (left) and Bayesian estimation (right). Data were collected using the microPET small animal scanner (courtesy of S. Cherry, Ph.D., Associate Professor, Pharmacology, UCLA). These images demonstrate the ability of the Bayesian estimation process to produce higher resolution images than FBP at similar signal to noise ratios.

the (second) regularizing term. For the case where the noise and image are mutually independent Gaussian processes, the two methods become identical (26). However, in quantum limited imaging systems such as X-ray CT, PET and SPECT the noise is not Gaussian so that the Bayesian approach will lead to different algorithms and solutions than the regularization methods described above.

OTHER APPLICATIONS

Nuclear Medicine

As mentioned in the introduction, PET and SPECT are medical imaging modalities that produce 3D images of the spatial distribution of biochemical tracers within the human body. To enable imaging of these tracers, they are tagged with γ-ray emitting (SPECT) or positron-emitting (PET) isotopes.

SPECT images are reconstructed from data formed through the detection of γ rays by a collimated planar Anger camera that produces an image of the number of photons detected at each pixel element (1). The collimator allows detection of only those photons that are traveling normally to the plane of the camera. The total number of photons collected at a single pixel element in the camera is proportional to the total number of radioactive nuclei, and hence tracer density, integrated along the straight path through the object and normal to the camera (this is a simplified model, an accurate model must account for attenuation and scatter of the γ rays as they pass through the body). By rotating the camera around the patient and collecting images at multiple angles of view, a set of parallel projections of the tracer distribution are collected for multiple 2D slices (1).

PET systems also collect parallel projections of the tracer distribution. The physical basis for PET imaging lies in the fact that a positron produced by a radioactive nucleus annihilates with an electron to form a pair of high-energy (511 keV) photons after traveling a very short distance (31). The pair of photons travel in opposite directions along a straight-line path. Detection of the positions at which the photon pair intersects a ring of detectors allows these systems to approximately determine the straight line path of the photons along which the position was initially emitted. The total number of such events measured by a pair of detectors will be proportional to the total number of such emissions along the line joining the detectors. By surrounding the patient with a ring of 511 keV photon detectors and associated photon-coincidence electronics, a PET system can simultaneously measure line integrals of the tracer distribution along the many paths between all pairs of detectors. Using multiple detector rings allows simultaneous acquisition of multiple 2D slices (32).

Since PET and SPECT data are approximate line integrals of the image, they can be reconstructed using analytical inversion formulas (32). These direct methods, however, do not allow accurate modeling either of the detector system or of the inherent statistical fluctuations in the data. These instruments detect relatively small numbers of individual photons, in some cases averaging less than 10 photons per projection sample. Consequently, the photon-limited statistical variations, which are well modeled using Poisson processes, play an important role in limiting reconstructed image quality. By using the maximum likelihood (27) or Bayesian (28) methods for Poisson data, substantial improvements in image quality can be realized in comparison to direct reconstruction methods. A second factor that limits performance in both modalities is that the line-integral model does not account well for the finite and spatially variant resolution of the detector and other physical factors. Using iterative or statistical approaches, these factors can also be readily included in the model (33).

Radio Astronomy

Radio telescopes are used to study the natural radio emission from celestial objects. The resolution of a single radial telescope is related to the size of its reflector. To achieve high resolution would require reflectors with diameters of hundreds of meters or more, depending on the wavelength of interest. Tomographic techniques provide a means of effectively improving the resolution of radio telescopes.

One technique for producing tomographic images is the method of lunar occultation (12). A radio telescope is trained on a limb of the moon. As the moon moves relative to the sky background, the portion of the background that is occluded by the moon will change. The change in received power as the moon moves, revealing an additional strip of background, is proportional to the energy produced from that strip. In this way a set of parallel strip integrals of the region occluded by the moon can be collected. To allow reconstruction, several views are needed, and hence data should be collected simultaneously from several sites. Even in this case, irregular sampling and incomplete data present serious reconstruction challenges.

Arrays of radio telescopes are more commonly used for tomographic imaging. A linear array of receivers can be configured to emulate a single telescope with very high resolution along the length of the array, but poor resolution in the orthogonal direction. Thus the array will produce measurements of the emission from a thin strip, or approximate line integral, of the sky. Motion of the earth allows acquisition of multiple views of the sky. These views can be combined in a tomographic manner to produce high-resolution images. These techniques were pioneered by Bracewell (12) in the 1950s using very similar reconstruction approaches to those developed in 1970s for X-ray CT.

More recent developments in radio astronomy employ interferometry in which an array of antennas, distributed over a large area, are pointed at the same region in the sky. Correlating the signals between pairs of antennas produces measurements of the 2D Fourier transforms of the brightness of the sky. As the earth rotates, these sample points trace out elliptical paths through Fourier space and Fourier interpolation methods can be used to recover the brightness distribution.

Geophysical Tomography

Tomographic methods have been widely used in geophysical exploration. The electromagnetic properties of a region between pairs of bore holes can be mapped by measuring the changes in EM signals as they propagate between the bore holes. Methods for producing images of resistivity and propagation delay are described by Dines and Lytle (13) using the simplifying assumption that propagation occurs along straight-line paths from transmitters, located at several depths within one bore hole, to receivers located at several depths within the second bore hole. A linear algebraic formulation was developed and the problem solved using an ART-like algorithm. Extensions to this approach, which include improvements on the straight-line path assumption, have since been developed, as reviewed by Spies (34).

For mapping the earth on a global scale, scientists have made use of records of natural seismic events. The Bulletin of the International Seismological Center records the arrival time at over 1000 seismic stations of different body wave phases, including the P-wave that passes through the mantle and other waves that reflect from, or travel through, the core. Given the epicenter of each event, these arrival times can then be used to produce 3D maps of anomalies in wave travel times. These maps, in turn, provide important insight into geodynamic questions such as the driving mechanisms of plate motion (17).

A simplified model for the anomalies in arrival times of the body waves is given by assuming propagation of the wave along a known curved ray path. By also assuming that the changes in the delay times are locally linear with respect to changes in the wave speed, the problem reduces to a linear tomographic problem of the form of Eq. (14). The data are the deviations in the arrival time from that expected for a spherically symmetric earth model and the unknown image represents the deviations in wave speed. The kernel matrix $\boldsymbol{H}$ is zero except in those pixels through which the ray is expected to pass and the problem can be solved iteratively using an ART-like method.

While a standard pixel-based decomposition of the earth can be used in these studies, the near-spherical shape of the earth makes a spherical harmonic expansion for the lateral variations, coupled with a Legendre polynomial expansion for radial variations, a more natural representation of the wave speed anomalies. The vector $\boldsymbol{f}$ in Eq. (14) are then the coefficients of this expansion and the elements of the kernel matrix $\boldsymbol{H}$ must be changed to account for this change in basis functions. A review of recent developments in seismic tomography can be found in Ref. 35.

RECENT AND POTENTIAL FUTURE DEVELOPMENTS

Medical tomographic imaging systems and the corresponding image-reconstruction techniques continue to develop at a rapid rate. Current state-of-the-art fan-beam X-ray CT systems have single-plane acquisition times of less than 1 and in-plane resolutions better than 1 mm. Recent advances include spiral scanning systems in which continuous rotation of the source and detectors allows fast volume acquisition while the patient is translated through the scanner. Truly 3D scanners are also now available, for example, the Ultrafast Cine-CT system, which uses multiple rings of detectors for simultaneous imaging of several planes with acquisition times of 50 ms per scan (6). Fast scanning is achieved by replacing the conventional rotating X-ray tube with a steered electron beam that produces X rays as it strikes a cylindrical target in the scanner gantry.

Nuclear medicine imaging has always been viewed as a 3D imaging technique, but recent advances in SPECT collimator design and advances in PET detectors and data acquisition systems has led to the development of fully 3D systems that produce data that cannot be decoupled into independent sets of 2D data. These 3D data sets present both mathematical and computational challenges in developing fast and accurate reconstruction methods and are currently a rich area of research (36).

In addition to the modalities of X-ray CT, PET, SPECT, and MR discussed here, there are an increasing number of experimental medical imaging modalities that use measurements of line integrals, or other mappings between image and data, together with computed inverses, to form images. Ultrasonic tomography of soft tissue presents a challenging inverse problem in which refraction and diffraction effects produce data that are no longer simple integrals along straight lines (16). Electromagnetic methods in the near-DC range are used for dynamic imaging of electrophysiological sources in heart, muscle, and brain. Mapping the electromagnetic properties of transmission through the body using a wide range of the EM spectrum is also of great interest in the search for methods that will provide new diagnostic capabilities (6). These new modalities present special challenges in signal processing because the data are not simple line integrals, and the inverse problems are often highly ill-posed due to combinations of limited data, poor signal-to-noise ratios, and ambiguities inherent in the underlying EM equations.

The emphasis in the article has been primarily in medical imaging, which is the main field of interest of the author. However, the brief discussion of other applications is provided as a starting point for exploring the exciting and growing range of applications of tomography.

BIBLIOGRAPHY

1. H. H. Barrett and W. Swindell, *Radiological Imaging,* New York: Academic Press, 1981, Vols. I and II.
2. S. Webb, *From the Watching of Shadows: the Origins of Radiological Tomography,* Institute of Physics Publishing, 1990.
3. J. Radon, On the determination of functions from their integral values along certain manifolds, *IEEE Trans. Med. Imag.,* **MI-4**: 170–176, 1986 (first published 1917 in German).
4. G. N. Hounsfield, *A method of and apparatus for examination of a body by radiation such as x-ray or gamma radiation,* U.K. Patent Specification 1283915, The U.K. Patent Office, England, 1972.
5. B. Holtzmann Kevles, *Naked to the Bone: Medical Imaging in the Twentieth Century,* Rutgers Univ. Press, 1996.
6. S. Webb (ed.), *The Physics of Medical Imaging,* London: Institute of Physics, 1988.
7. P. Lauterbur, Image formation by induced local interactions: Examples employing nuclear magnetic resonance, *Nature,* **242** (5394): 190–191, 1973.
8. A. Kumar, D. Welti, and R. Ernst, NMR Fourier Zuegmatography, *J. Mag. Reson.,* **18**: 69–83, 1975.
9. Z. Zhou et al., Calculation of 3-D internal displacement fields from 3-D x-ray computer tomographic images, *Proc. R. Soc. London Sec. A,* **449**: 537–554, 1995.
10. R. A. Crowther, D. J. DeRosier, and A. Klug, The reconstruction of three-dimensional structure from projections and its applications to electron microscopy, *Proc. R. Soc. London Sec. A,* **317**: 319–340, 1970.
11. T. R. Corle and G. S. Kino, *Confocal Scanning Optical Microscopy and Related Imaging Systems,* New York: Academic Press, 1996.
12. R. N. Bracewell, Image reconstruction in radio astronomy, in G. T. Herman (ed.), *Image Reconstruction from Projections: Implementation and Applications,* New York: Springer, 1979.

13. K. Dines and R. Lytle, Computerized geophysical tomography, *Proc. IEEE,* **67**: 1065–1073, 1979.
14. D. C. Munson, Jr., J. D. O'Brien, and W. K. Jenkins, A tomographic formulation of spotlight-mode synthetic aperture radar, *Proc. IEEE,* **71**: 917–925, 1983.
15. T. D. Raymund et al., Application of computerized tomography to the investigation of ionospheric structures, *Radio Sci.,* **25**: 771–789, 1990.
16. H. W. Jones, Recent activity in ultrasonic tomography, *Ultrasonics,* **31** (4): 353–360, 1993.
17. A. Dziewonski and J. Woodhouse, Global images of the earth's interior, *Science,* **236** (4797): 37–48, 1987.
18. A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging,* New York: IEEE Press, 1988.
19. R. Merserau and A. Oppenheim, Digital reconstruction of multidimensional signals from their projections, *Proc. IEEE,* **62**: 1319–1338, 1974.
20. L. A. Shepp and B. F. Logan, The Fourier reconstruction of a head section, *IEEE Trans. Nucl. Sci.,* **NS-21**: 21–33, 1974.
21. H. P. Hiriyannaiah, X-ray computed tomography, *IEEE Signal Process. Mag.,* **14** (2): 42–59, 1997.
22. B. K. Horn, Fan-beam reconstruction methods, *Proc. IEEE,* **67**: 1616–1623, 1979.
23. G. T. Herman, *Image Reconstruction From Projections: The Fundamentals of Computerized Tomography,* New York: Academic Press, 1980.
24. Y. Censor, Finite series-expansion reconstruction methods, *Proc. IEEE,* **71**: 409–419, 1983.
25. M. I. Sezan and H. Stark, Tomographic image reconstruction from incomplete view data by convex projections and direct Fourier inversion, *IEEE Trans. Med. Imag.,* **MI-3**: 91–98, 1984.
26. D. M. Titterington, General structure of regularization procedures in image reconstruction, *Astron. Astrophys.,* **144**: 381–387, 1985.
27. L. A. Shepp and Y. Vardi, Maximum likelihood reconstruction for emission tomography, *IEEE Trans. Med. Imag.,* **MI-2**: 113–122, 1982.
28. S. Geman and D. E. McClure, Bayesian image analysis: An application to single photon emission tomography, *Proc. Stat. Comput. Sec. Am. Stat. Assoc.,* 1985, pp. 12–18.
29. G. Gindi et al., Bayesian reconstruction of functional images using anatomical information as priors, *IEEE Trans. Med. Imag.,* **MI-12**: 670–680, 1993.
30. E. U. Mumcuoglu et al., Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images, *IEEE Trans. Med. Imag.,* **MI-13**: 687–701, 1994.
31. S. R. Cherry and M. E. Phelps, Imaging brain function with positron emission tomography, in A. W. Toga and J. C. Mazziotta (eds.), *Brain Mapping: The Methods,* New York: Academic Press, 1996.
32. J. Ollinger and J. A. Fessler, Positron emission tomography, *IEEE Sign. Process. Mag.,* **14** (1): 43–55, 1997.
33. J. Qi et al., Fully 3D Bayesian image reconstruction for the ECAT EXACT HR+, *IEEE Trans. Nucl. Sci.,* **45**: 1096–1103, 1998.
34. B. R. Spies, Electrical and electromagnetic borehole measurements: A review, *Surveys Geophys.,* **17**: 517–556, 1996.
35. J. P. Montagner, Can seismology tell us anything about convection in the mantle?, *Rev. Geophys.,* **32** (2): 115–137, 1994.
36. D. Townsend and P. Kinahan (eds.), *Physics in Medicine and Biology (Special Issue on Fully 3D Image Reconstruction),* **43** (4): 1998.

RICHARD M. LEAHY
University of Southern California

TOMOGRAPHY, COMPUTERIZED. See COMPUTERIZED TOMOGRAPHY.

TOMOGRAPHY, EMISSION. See RADIONUCLIDE IMAGING.

TOMOGRAPHY, MEDICAL IMAGING. See BIOMEDICAL NMR; ELECTRIC IMPEDANCE IMAGING.

TOOLKITS. See USER INTERFACE MANAGEMENT SYSTEMS.

TOPOLOGICAL ANALYSIS. See SYMBOLIC CIRCUIT ANALYSIS.