MATHEMATICAL TECHNIQUES IN BIO-MEDICAL IMAGING : PROBLEMS AND PROSPECTS

D. DUTTA MAJUMDER AND S. BANERJEE

National Centre for Knowledge Based Computing Systems
Indian Statistical Institute, Calcutta 700 035

(Accepted on 24 *June* 1993*)*

Mathematical formalisms underlying some of the principal bio-medical imaging modalities are given. Problems of image reconstruction using these formulations, when limited or noisy data is available are stated and attempts to remedy these problems are discussed. Applications of the mathematical formalisms to X-ray, nuclear magnetic resonance and radionuclide based imaging modalities are presented.

1. INTRODUCTION

Visualization in scientific and engineering research is a rapidly emerging application area aimed at developing approaches and tools to allow researchers and users to "see" and comprehend the systems and components they are studying. It is an application area that requires multidisciplinary endeavour from subjects like computer science, electrical engineering, physics, mathematics and statistics-along with experts from other domains such as medicine biology, astrophysics and environmental sciences. The subject is at the intersection of some of the frontier research area fields such as pattern recognition (PR), imaging processing (IP), computer vision, tomography and computer graphics. Examples of approaches to scientific visualization include the presentation of information in three dimensions, development of dynamic methods to interact with and manipulate multidimensional data, and development of models of visual perception that enhance interpretive and decision making processes. Application areas include, among others, Astronomy, Astrophysics, Environmental Science and Bio-Medicine. In this paper we shall concentrate only in the area of bio-medical computing.

Important examples of applications of visualization in biomedical computing include presentation of anatomy and physiology in 3-D, animated representation of the dynamics of fluid flow, animated representation of bio-medicine and bio-kinetic study for orthopaedic purposes such as gait analysis and graphical rendering of biomolecular structures and their interactions.

We shall try to develop the presentation of the paper in a brief but tutorial fashion indicating the research areas and open problems and then concentrate on the mathematical aspects of 3-D reconstruction from projections in different bio-medical imaging modalities. Researchers may be expected to get acquainted with some of the standard books on pattern recognition[1, 2, 3], image processing[4, 5, 6], computer vision[7],

computer graphics[8] and image reconstruction[9,10], but is not necessary for the professional people. We shall not attempt to present an exhaustive list of references but we shall try to cover the key references on the basis of which our study is made.

The principal imaging modalities to date are X-ray Computed Tomography (CT), Positron Emission Tomography (PET), Single photon Emission Comptued Tomography (SPECT), nuclear Magnetic Resonance Imaging (MRI) and ultrasonography/echocardiography. In all these modalities, the variation of some particular physical property in a 3-D object element or reconstruction space is used in reconstructing the image on pixels. Images obtained by considering this slices of the object are tomographic images which are preferred to planar images owing to their better image contrast and spatial resolution.

There are two broad groups of image reconstruction methods, namely, inverse transform type methods and series expansion type methods. In the former, a "picture" function (i.e. a function that describes the physical property under consideration in Cartesian/polar space) is obtained in terms of its Radon or Fourier transform using a discretized version of the Radon or Fourier inversion formula, respectively[10]. This subject will be elaborated in a later section. On the other hand, in series expansion methods, the problem is discretized at the very beginning. It may be noted here, that in other methods, the physical property used in reconstruction may be modeled either deterministically or stochastically.

In series expansion methods, the idea is to estimate an image vector x such that

$$y = Rx + e$$

for measurement vector y. R is referred to as the projection matrix and e is the error vector. These methods generally start out with the approximation of an uniform gray level matrix and projection are calculated and compared with the measured ones. Then appropriate corrections are incorporated and successive approximations are made until convergence is obtained. Estimation of x is done by requiring that x and e satisfy some kind of optimization criteria. These series expansion methods usually utilize an iterative method of image reconstruction. One such method, known as the Simultaneous Iterative Reconstruction Technique (SIRT)[10] simultaneously corrects all projections by applying a suitable damping factor. In the second method, an individual projection is calculated and corrected. This constitutes the Algebraic Reconstruction Technique (ART)[10]. The disadvantage with these methods is that all projections are not given equal weight with the latter projections carrying more weight. These methods are not generally applied to X-ray CT but are frequently employed in Emission Computed Tomographic (ECT) imaging systems like PET and SPECT, as they allow the incorporation of beam attenuation factors in a fairly straightforward manner[11]. Later on, we shall show how this is done for SPECT.

This paper is organized as follows. Section II contains a formulation of the Convolution Backprojection (CBP) algorithm, an example of an inverse transform method. Extension of the CBP method to 3-D is presented in Section III followed by Section IV describing efforts to reconstruct images using CBP for limited and noisy data. Section V details the Expectation Maximization (EM) formulation of

Maximum Likelihood (ML) estimation (EM-ML) for finite data proposed by Dempster, Laird and Rubin[12]. Iterative vector extrapolation methods to accelerate the convergence of the EM-ML algorithm and their applications to PET imaging are discussed in Section VI. The 3-D version of the EM-ML algorithm used in SPECT imaging is described next (Section VII). Noise in ECT images and efforts for their rectification constitute the subject matter of Section VIII. In the following Section (IX), the Fourier zeugmatogrphic technique in MRI, an inverse transform type method, is elaborated and finally, a few concluding remarks are made.

2. Mathematical and Physical Preliminaries for Projection Reconstruction

The basic idea underlying projection reconstruction is the production of an image of a 2-D distribution of some physical property from estimation of its line integrals along a finite number of lines of known locations[10] (see Fig. 1). The line along which the integral is obtained is called the ray and the integral along this line is

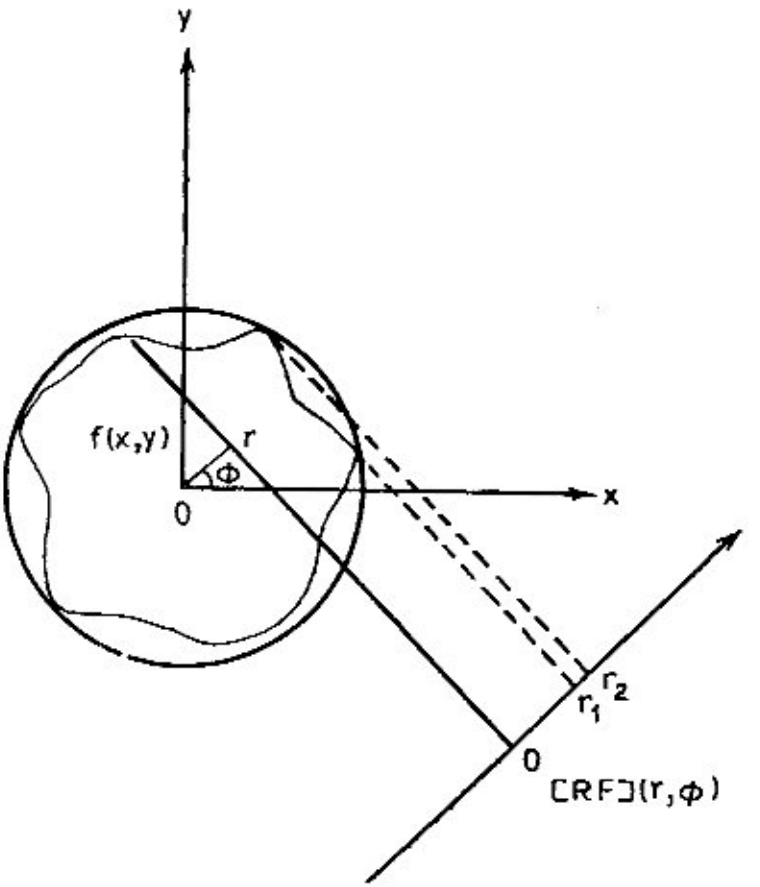

Fig. 1.

called the ray sum. We assume that contributions to the integral are from infinitely thin slices of the object so that distinctions between voxels and pixels is of no consequence. The picture function at point (x, y) of Cartesian space is $f(x, y)$. The ray sum $P(r, \phi)$ in terms of (r, ϕ) variables in projection space is

$$P(r, \phi) = \int_{r,\phi} f(x, y)\, ds \qquad \ldots. (2.1)$$

where $r = x \text{ Cos } \phi + y \text{ Sin } \phi$ and s is the distance along the ray direction. The set of values for a given orientation ϕ is the projection at angle ϕ (see Fig. 2). By chainging ϕ one obtains a set of projections $P(r, \phi)$ which are used to reconstruct the image. The rays considered in Fig. 2 are equally spaced parallel rays.

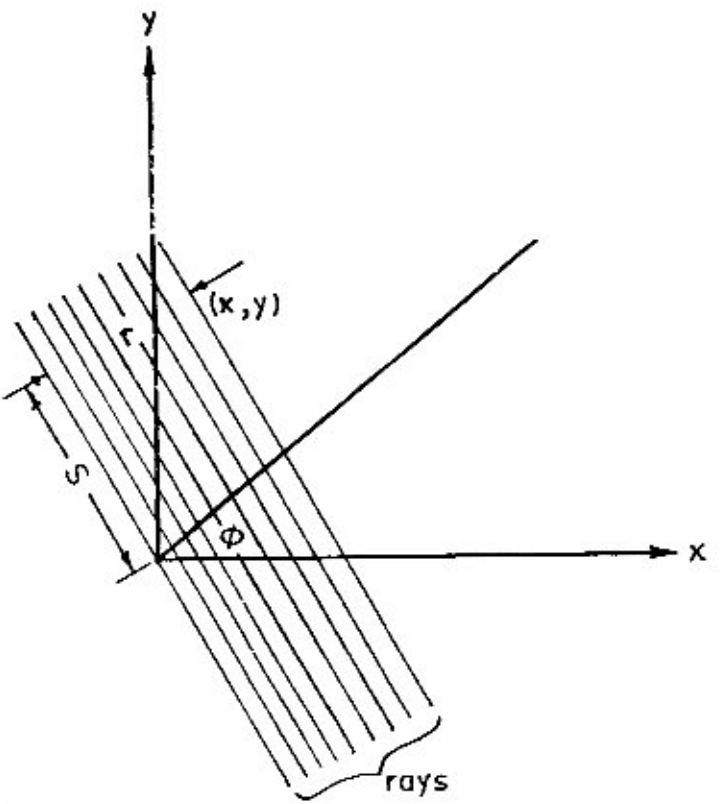

FIG. 2.

Backprojection consists of dividing each ray sum equally among all the pixels which lie on the corresponding ray. This process is repeated for each of the projections and the results are summed[13]. The backprojection is

$$\bar{f}(x, y) = \sum_{j=1}^{m} P(r, \phi_j) \Delta \phi \quad \ldots (2.2)$$

$\bar{f}(x, y)$ is the calculated picture function, m is the total number of orientations and $\Delta \phi$ is the angular increment between adjacent projections. In terms of x and y,

$$\bar{f}(x, y) = \sum_{j=1}^{m} P(x \operatorname{Cos} \phi_j + y \operatorname{Sin} \phi_j, \phi_j) \Delta \phi$$

and in integral form

$$\bar{f}(x\ y) = \int_{0}^{\pi} P(x \operatorname{Cos} \phi + y \operatorname{Sin} \phi, \phi)\, d\phi \quad \ldots (2.3)$$

From Eqns. (2.2) and (2.3) one finds that the projection amplitude needs to be known at arbitrary values of r and the integral over angular increments $\Delta\phi$ is sought after. In practice, only a finite number of equally spaced values are available. These errors result in the appearance of streak artifacts in the image and spatial resolution is decreased. Thus, it is necessary to filter the projections prior to backprojection to make the method analytic. These filters should produce a gradual roll-off of frequency prior to backprojection in order to minimize the effects of aliasing.

The analytic method known as Fourier reconstruction[13] is derived with the help of the "central slice theorem". According to this theorem, the Fourier coefficient of the picture function $f(x, y)$ given by

$$F(k_x, k_y) = \int_{-\infty}^{\infty} P(r, \phi) \exp(-2\pi ikr)\, dr = H(k, \phi) \quad \ldots (2.4)$$

equals the fourier transform of the projection taken in the direction of the Fourier wave defined by the wave numbers k_x, k_y. $H(k, \phi)$ is the Fourier transform of the projection.

Then, on using eqn. (2.4) in (2.3)

$$\bar{f}(x, y) = \int_{0}^{\pi} \int_{-\infty}^{\infty} \frac{H(k, \phi)}{|k|} \exp[2\pi ik(x \cos \phi + y \sin \phi)] |k|\, dk\, d\phi \quad \ldots (2.5)$$

Here the Fourier trasform of $f(x, y)$ is taken and eqn. (2.4) is used, as a consequence of which

$$\frac{H(k, \phi)}{|k|} = \frac{F(k_x, k_y)}{|k|}.$$

So, the actual fourier coefficients $F(k_x, k_y)$ of the object are obtained from those of the backprojected image by multiplying with the wave vector amplitude $|k|$.

The picture function is

$$f(x, y) = \int_0^{\pi} \int_{-\infty}^{\infty} F(k, \phi) \exp(2\pi ikr) \mid k \mid dk\, d\phi$$

$$= \int_0^{\pi} P'(x \operatorname{Cos} \phi + y \operatorname{Sin} \phi, \phi)\, d\phi$$

where the modified projection $P'(r, \phi)$ is given by

$$P'(r, \phi) = \int_{-\infty}^{\infty} \mid k \mid F(k, \phi) \exp(2\pi ikr)\, dk. \qquad \ldots (2.6)$$

This modified backprojection P' is not spatially bounded unlike, P, so Radon filtering[14] is applied. This method incorporates the convolution theorem in eqn. (2.6). According to this theorem, the Fourier transform of the product of two functions $F(k, \phi)$ and $\mid k \mid$ are replaced by the convolution of their respective Fourier transforms, namely $P(r, \phi)$ and $-1/2\pi^2 r^2$.

So eqn. (2.6) becomes

$$P'(r, \phi) = \frac{-1}{2\pi^2} \int_{-\infty}^{\infty} \frac{P'(r', \phi)\, dr'}{(r - r')^2}.$$

The quadratic singularity is removed if k is bounded, i.e. $\mid k \mid \leq k_m$. So

$$\int_{-k_m}^{k_m} \mid k \mid \exp(2\pi ikr)\, dk = \frac{k_m}{\pi r} \operatorname{Sin}(2\pi k_m r) \frac{-\operatorname{Sin}^2(\pi k_m r)}{\pi^2 r^2}$$

Using Eqn. (2.7), the expression for the modified backprojection is

$$P'(r, \phi) = \int_{-\infty}^{\infty} P(r', \phi) \left\{ \frac{k_m \sin[2\pi k_m (r - r')]}{\pi (r - r')} - \frac{\sin^2[\pi k_m (r - r')]}{\pi^2 (r - r')^2} \right\} dr'.$$

This result is valid for equally spaced parallel rays.

In realistic situations data are collected so that they divide into subsets, as in fan beam type of collection mode. Each of these subsets contain ray sums for rays diverging from a single point. This is the divergent mode of data collection shown in Fig. 3. A process called "rebinning" is performed on this data set before applying the parallel beam convolution method described above. Measured ray sums for sets of divergent rays are interpolated in projection space to estimate ray sums for parallel rays (10).

The divergent projection data are $P(r_n, \phi_{n,k})$, for $-N \leq n \leq N$ and $0 \leq \phi_{n,k} \leq \Pi$. From these values, estimates of $p(ud, v\Gamma)$ for $-U \leq u \leq U$ and $0 \leq v \leq V - 1$ are made with $\phi = v\Gamma$ and $(v\Gamma)_{max} = \pi$ and $ud \leq L$ ($\sqrt{2}\, Lx \sqrt{2}\, L$ is the

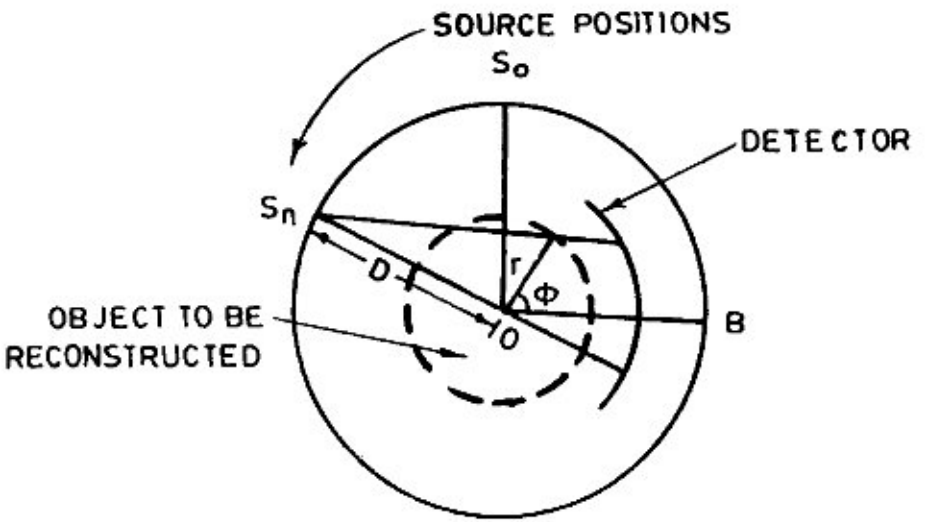

FIG. 3.

size of the 2-D picture space). The interpolation is carried out in two steps.

The first stage, done for each $-N \le n \le N$, consists of interpolation in the second variable. Considering n fixed, estimates of $p(r_n, \phi)$, for $\phi = v\Gamma$, $0 \le v \le V-1$ and $v\Gamma = \pi$, are obtained from values of $p(r_n, \phi_{n,k})$. At the end of this stage, ray sums for sets of parallel but unequally spaced rays are obtained. In the second stage, interpolation in the second variable is performed. Keeping v fixed, estimates of $p(ud, v\Gamma)$ are obtained from $p(r_n, v\Gamma)$ for $-U \le u \le U$. Thus, parallel equally spaced rays are obtained and parallel beam convolution techniques can be applied.

3. CBP METHOD FOR 3-D RECONSTRUCTION

Magnetic Resonance Imaging (MRI) wears the advantage of being a truly multiplanar and 3-D imaging modality. Thus, it can provide images in planes other than the traditional transverse one as in X-ray CT. Three dimensional MRI images can be obtained by Convolution Back Projection. A few words on the extension of the CBP method described in a previous section to three dimension will clarify the concept.

The 3-D equivalent of the picture function is the spin density $\rho(x, y, z)$ for MRI. In terms of its Fourier coefficients $F(k_x, k_y, k_z)$ this is

$$\rho(x, y, z) = \int\int\int F(k_X, k_Y, k_Z) \exp 2\pi i (k_x x + k_y y + k_z z)\, dk_x\, dk_y\, dk_z$$

and, in spherical coordinates (k, θ, ϕ) this becomes

$$\rho(x, y, z) = \int\int\int F(k, \theta, \phi) \exp(2\pi ikr)\, k^2 \operatorname{Sin}^2 d\theta\, d\phi\, dk$$

$$= \int_0^{2\pi} \int_0^{\pi} P'_{\theta\phi}(r) \sin\theta \, d\theta \, d\phi$$

where, the modified projection $P'_{\theta\phi}(r)$ is

$$P'_{\theta\phi}(r) = \int_0^{\infty} k^2 F(k, \theta\phi) \exp(2\pi ikr) \, dk.$$

Using the 3-D extension of the central slice theorem viz.,

$$F(k, \theta, \phi) = \int_{-\infty}^{\infty} P_{\theta\phi}(r) \exp(2\pi ikr) \, dr = H_{\theta\phi}(k).$$

(An additional filtering factor $g(k)$ can be introduced into the above expression) the modified projection becomes

$$P'_{\theta\phi}(r) = \int_0^{\infty} k^2 H_{\theta\phi}(k) \exp(2\pi ikr) \, dk.$$

The 3-D image is reconstructed by backprojecting these modified projections.

In the discrete case

$$\rho(x, y, z) = \sum_{i=1}^{N} P'_{\theta_i \phi_i}(r) \sin\theta_i \Delta\theta_i \Delta\phi_i$$

where N is the number of projections and the solid angle for each projection is $\sin\theta_i \Delta\theta_i \Delta\phi_i$ and this has the same value for all projection angles θ, ϕ in the case of isotropic resolution.

The CBP method has some limitations and Fourier imaging methods in MRI some inherent computational advantages which will be discussed in a later section.

4. CBP Image Reconstruction from Incomplete, Missing and Noisy Data

Improving procedures for obtaining reconstructions from missing/noisy projection data is a challenging area of research, and we present below some of the efforts made to alleviate this difficulty and indicate the need for further research.

Several approaches exist for image reconstruction from limited data. One way is to devise algorithms which do not require completeness of projections. The series expansion methods and Minimum Variance Estimator (MVE) methods[15] fall under this category.Both these methods require a lot of computation time.

Another approach obtains estimates of missing data in projection space by extrapolation and the image is reconstructed by CBP or Fourier inversion (FI)[16].

These extrapolation methods are generally iterative and, therefore, convergence is slow and the computation cumbersome. However, there are two approaches which claim to achieve substantial savings of computation time, and we devote the rest of this section to a brief survey of these approaches.

The first method under consideration estimates missing data by extrapolation, but is non-iterative and utilizes a parametric model of signals. Srinivasa *et al.*[16] have estimated the missing data in projection space using Linear Prediction Theory (LP). They represent the ray sum as a data matrix P whose elements $p(i, j)$ represent line integrals for location vectors i and projection angles j, i.e. i and j are discretized versions of r and ϕ, respectively. For data available over limited angles, only the rows representing projections for $\alpha < j < \beta$ are available in matrix P, and so its columns are not complete. The procedure consists of modelling each column as a 1-D Auto Regressive (AR) process of order N. Thus, a non-recursive digital filter estimates data values $p(i, j)$ from prior observations available in the same column j. The predicted value $\hat{p}(i, j)$ in terms of $p(i, j)$ observed and the AR predictor coefficients $a_{N,j}(k)$ is

$$\hat{p}(i, j) = -\sum_{k=1}^{N} a_{N,j}(k)\, p(i-k, j).$$

The AR predictor coefficients are selected to minimize the mean square error (MSE) between the predicted and observed values. Srinivasa *et al.*[16] considered several methods for obtaining the AR parameters and found that the Levinson recursion method[17] was most suitable.

The data matrix was completed by extrapolating the available column data upwards and downwards using the linear predictor. For column j

$$\hat{p}(\beta, j) = -\sum_{k=1}^{N} a_{N,j}(k)\, p(\beta-k, j)$$

$$\hat{p}(\beta+1, j) = -\sum_{k=2}^{N} a_{N,j}(k)\, p(\beta-k+1, j) - a_{N,j}(1)\, \hat{p}(\beta, j).$$

By running the predictor backwards, missing data for angles less than β are obtained. Apriori information such as non-negativity of the projection data and finite extent of the object have also been taken into consideration. In addition to obtaining images from hollow, limited angle and truncated projections, images from arbitrarily missing/noisy projections have also been obtained[18].

The other approach that we consider is a hierarchical Bayesian approach for image reconstruction proposed by Bresler and Macovski[19] which uses minimum variance filtering. Their algorithm for reconstruction of multiple 3-D objects has been implemented in three steps for hierarchical estimation of the position and shape of the object. The object is considered to be smooth with limited boundary curvature represented by a chain of primitives which are equal length cylindrical sections centered at a point $\bar{c}(t)$ around a trajectory $\{c(t) = [x(t), y(t), z(t)]\}$ in the $(x, y,$

z) space and cross-sectional shape and orientation defined by parameter vector $\overline{\Gamma}$. The curvature and smoothness of objects O_j is obtained from a stochastic difference equation describing the evolution with time, of the vector $X_j(t)$:

$$X_j(t) = [\, c_j'(t),\ \Gamma_j(t)\,]^t$$

and

$$X_j(t+1) = AX_j(t) + BW_j(t) \qquad \ldots (4.1)$$

where A, B are constants and $W_j(t)$ is a zero mean, uncorrelated Gaussian noise vector with covariance Q. Bresler[19] has performed and data analysis for a single cylinder and stated that the application of the algorithm to multicylinder data utilizes a suboptimal approach.

Massive data reduction is obtained in the first step by detecting object primitives in 3-D space and obtaining local maximum likelihood estimates of their parameters. The likelihood function of measurement Y for a single cylinder c is[20].

$$L(c, \Gamma) = -\log p[Y \mid c, \Gamma]$$

$$= \frac{2}{N} \int_{\phi} \int_{-\infty}^{\infty} y(r, \phi)\, g_0\,[r - M(\phi), c, \phi\,;\, \Gamma\,]\, ds\, d\phi - \frac{1}{N}\, \varepsilon\ (\Gamma, \phi\,)$$

where ds is an area differential in projection space and the integral is over all projection angles ϕ.

Here g_0 is the X-Ray Transform (XRT) of the cylindrical distribution in projection space, r is the location vector in projection space, ϕ is an unit vector in the direction of projection view and $M(\phi)$ is a 2×3 matrix determined by projection angles ϕ.

By linearity of the XRT, the noisy projection of N objects with the jth object comprising of T cylinders is

$$y(r, \phi) = \sum_{j=1}^{N} \sum_{t=1}^{T_j} g_0\,[r - M(\phi), c(t), \phi\,;\, \Gamma\,(t)\,] + v(r, \phi)$$

$v(r, \phi)$ being the uncorrelated Gaussian random field of intensity N_0 and

$$\varepsilon\,(r, \phi) = \int_{\phi} \int_{-\infty}^{\infty} g_0^2\ [r, \phi\,;\, \Gamma\,]\, ds\, d\phi$$

being thought of (19) as the energy of the cylinder signal in projection space.

Thresholding and maximization of $L(c, \Gamma)$ w.r.t. c and Γ give the optimal detection of a cylinder's presence and estimation of location and parameters. Thus, at the end of the first step, the algorithm produces a set of N data points

$$Z = [Z_i]_{i=1}^{N}.$$

Assuming the estimation error to be Gaussian distributed with covariance R and uncorrelated between data points

$$Z_i = X_j(t) + V_j(t) \qquad \ldots (4.2)$$

where $V_j(t)$ is additive white noise. The system and measurement models are of the classical form to which Kalman filtering can be applied, in principle, and a minimum variance estimate obtained. However, the uncertainties in this case are not confined to the noises W and v. Also, due to noise and false assumption (false alarm) of the presence of a single cylinder in space, the detected cylinders are not those corresponding to any object and so MVE is not useful at this stage.

In the second stage, a set of feasible objects are constructed by combining individual detected cross-section estimates via minimum variance filtering. The assignment of detections to each feasible object is guided by the sequential hypothesis test.

Bresler[19] thus treat the reconstruction problem as an unsupervised pattern recognition (PR) problem. Number of clusters present (objects) and parameters of individual objects are obtained from the unclassified data set Z. S is partitioned by $\mathcal{T}$ into m disjoint clusters λ^j and a false alarm cluster Φ consisting of data rejected from all λ^j.

ν_j is a vector of indices specifying the order in which n_j data points in the cluster λ^j are to be assembled into corresponding objects. All the m disjoint ν_j comprise an order rule $\nu = \{ \nu_j \}_{j=1}^{m}$

A hypothesis $H = (\mathcal{T}, \nu)$ completely specifies data to object association and S is the set of all valid hypotheses. The choice $H \in S$ is crucial to the solution of the problem and is formulated as a multihypothesis test.

This test is to maximize the posterior probability distribution of data z, i.e. $\max_{H \in S} P(H \mid Z)$ which minimizes the associated Bayes risk provided that the errors are equally weighted and correct decisions are not penalized. According to Bayes' rule

$$P(H \mid Z) = \frac{p(Z \mid H)\, p(H)}{p(z)} \qquad \ldots (4.3)$$

$p(H)$, the apriori distribution, has been assumed constant, and

$$p(Z \mid H) = \prod_{j=1}^{m} p(\lambda^j \mid H,\ \lambda^j \in H)\, p(\phi \mid H)$$

because of independence of the objects, estimation noises V and the constraint

$$\lambda^i \bigcap \lambda^j \quad i \neq j \qquad \ldots (4.4a)$$

which together imply conditional independence of the clusters.

Also

$$\Phi = Z - \bigcup_{i=1}^{m} \lambda^i \qquad \ldots (4.4b)$$

and

$$\tau = \{ m(\lambda^i)_{i=1}^{m} \}. \qquad \ldots (4.4c)$$

Again, because of the equal likelihood of false alarms in the imaged domain of volume V,

$$p(\phi \mid H) = \left[\frac{1}{V} \right]^{n_\Phi} \qquad \ldots (4.5)$$

where n_Φ = number of false alarms = $N - \sum_{j=1}^{m} n_j$.

Taking the log of eqn. (4.3) and dropping H independent terms, one arrives at

$$\min_{H \in S} \sum_{j=1}^{m} [-\log p(\lambda^j \mid H,\ \lambda^j \in H) - n_j \log V]. \qquad \ldots (4.6)$$

It is computationally difficult to solve eqn. (4.6) by direct enumeration of different H in S. Calculation of object likelihood functions for all S hypotheses is unfeasible for all but the smallest problems. Bresler[19] utilizes the conditional independence of objects to decompose eqn. (4.6) into a step of feasible object construction followed by integer optimization.

Thus, in the second step, ordered clusters $O_i = (\lambda^i, \nu_i)$ in Z that are reasonable to incorporate from H are detected. The criterion that O_i be incorporated into feasible set F ($H(F)$, is that it must pass the following hypothesis test :

$$L(O_i) = -\ln p(\lambda^i \mid \nu_i)$$

subject to $X_j(t+1)$ and Z_i of eqns. (4.1) and (4.2) >> a given α_n. Thus a pruned feasible set S^F replace a larger set S and thereby produces computational savings. The Kalman filter is then applied to data λ^i in order specified by ν_i to obtain a minimum variance causal estimate $\{ \hat{X}_j(t \mid t) \}_{\tau=1}^{\nu_\tau}$.

In the final step, a subset $\{ (\lambda^j, \nu_j) \}_{\varphi=1}^{M}$ of F is chosen to maximize eqn. (4.5) subject to the constraints given by eqns. (4.4a), (4.4b) and (4.4c).

This problem has been reformulated as a linear integer optimization problem. A feasible object O_j is completely specified by (λ_j, ν_j) where λ_j is a binary vector of length N, whose ith element is set to 1 if Z is associated with O_j and 0 otherwise. $\bar{\tau}$ is another binary vector of length K, where K is the number of feasible objects in F, iff the ith feasible object in $F \subset H$. So $\bar{\tau}$ decreases a hypothesis H. $\bar{\Lambda} = [\lambda_1, \ldots, \lambda_k]$ is a $N \times K$ matrix with the columns λ_j. The constraints of eqn. (4.4) are reexpressed as

$$\bar{\Lambda}\,\bar{\tau} \leq [1, 1, ..., 1]^t.$$

Also define

$$C_j = L(O_j) - n_j$$

as the cost associated with the choice of O_j. C_js are elements of vector $C = [C_1, ..., C_K]^t$.

Thus, the linear integer optimization problem is

$$\min_{\tau \in S^\tau} C'\tau \text{ subject to } \bar{\Lambda}\,\bar{\tau} \leq [1, 1, ..., 1]^t \qquad ... (4.7)$$

with S^τ = {all binary vectors of length K}.

This is a set packing problem and there are a number of efficient algorithms for solution[21]. The performance is claimed to be improved by preprocessing $\bar{\Lambda}$ and taking advantage of the sparseness of $\bar{\Lambda}$ to decompose eqn. (4.7) into a set of smaller problems.

In both methods discussed above, computation time is reduced. However, no account has been taken of density variations of the object to be reconstructed. In Bresler's[19] method, the extension of the algorithm to reconstruct multi-cylinders is not clear. Extensions to irregular objects with density variations are areas requiring further research.

5. Image Reconstruction from Limited Data in ECT Imaging

In ECT methods, the patient is administered with some physiologically active compound labeled with a radionuclide. These radionuclides decay in the body and emanate radiation which consists of positrons in PET and gamma ray photons in SPECT.

Radionuclides emitting positrons are unstable nuclei of protonrich isotopes of low atomic number element(S).

The emitted position combines with an atomic electron in its immediate vicinity, annihilates and produces two gamma ray photons of energies 511 keV each. These photons fly off in nearly opposite directions along a line with absorption are detected in time coincidence photon pair indicates the presence of an annihilation event in the cylindrical region between the detectors (referred to as tube). The number of coincidence counts observed by all such tubes in a given time interval is the measurement data. Single gamma ray photon emission occurs in unstable nuclie with high atomic numbers. The emitted photon in SPECT travels in a randomly oriented direction and is detected outside the body by a gamma ray camera. The photons are collimated to selectively detect only those photons which fall on the camera surface in specified directions. Photons which lie within an energy window separates unatternuated primary photons from those undergoing interactions, are detected.

Let λ denote the mean intensity function for emission of positrons (PET) or

gamma ray photons (SPECT) resulting from the decay of the radionuclides in the object space. $\lambda(S)$ has to be estimated for all points s within the object. Measurement data is a finite set of photon counts observed by detectors both in PET and SPECT. Thus, one has to estimate the parameter $\lambda(S)$ over the entire, continuous object space based on a finite set of measurement data.

The physical processes involved in ECT imaging are stochastic in nature and are generally described by mathematically appropriate stochastic models. Likewise, the measurement process, which involves observations of photon counts by detectors is also a stochastic process. Hence, image reconstruction in ECT can be treated as a stochastic estimation problem. As mentioned earlier, image reconstruction in ECT generally falls under the iterative series expansion category. An iterative algorithm based on the Expectation Maximization (EM) formulation for the Maximum Likelihood (ML) parameter estimation from incomplete data is often used in ECT imaging. This ML-EM algorithm was introduced independently by Shepp and Vardi (22) and Lange and Carson (23). In this section, we first describe the ML estimation from incomplete data via the EM algorithm developed by Dempster *et al.*[12]. Then we show how this algorithm has been adapted to ECT imaging.

Dempster *et al.*[12] propose the existence of two sample spaces Y and X and a many to one mapping from X to Y. The observed data vector $y \in Y$ and the corresponding $\bar{x}$ in X are observed indirectly through y. A mapping process $\bar{x} \rightarrow \bar{Y}(x)$ from x to y is assumed to exist and x is known to lie in $x(y)$, the subset of x determined by $\bar{Y} = \bar{Y}(x)$, where $\bar{Y}$ is the observed data. $\bar{X}$ is referred as the complete data and can include parameters also.

A family of sampling densities $f(\bar{X} \mid \bar{\lambda})$ depending on parameters $\bar{\lambda}$ is postulated, and a corresponding family of sampling densities $g(\bar{y} \mid \bar{\lambda})$ is derived. The complete data specification $f(\bar{x} \mid \bar{\lambda})$ is related to the incomplete-data specification $g(\bar{y} \mid \bar{\lambda})$ by

$$g(\bar{y} \mid \bar{\lambda}) = \int_{x(y)} f(\bar{x} \mid \bar{\lambda})\, \overline{dx}.$$

The object of the EM algorithm is to find a λ which maximizes $g(\bar{y} \mid \bar{\lambda})$ by making essential use of the associated family $f(\bar{x} \mid \bar{\lambda})$. Each iteration of the EM algorithm involves two steps, the expectation step (E-step) and the maximization step (M-step). We shall elaborate this algorithm in the context of our problem in the following.

The incomplete data likelihood function can be expressed in terms of the complete data likeliho of function through

$$P_y(\bar{y};\bar{\lambda}) = P_x(\bar{x},\bar{\lambda})/P_{x/y}(\bar{x} \mid \bar{y}=\bar{y};\bar{\lambda}) \qquad \ldots (5.1)$$

where $P_{x/y} = (\bar{x} \mid \bar{y}=\bar{y};\bar{\lambda})$ is the conditional density function given $Y = y$.

In our problem, the object is made up of N voxels and the mean intensity function λ_i defined earlier is considered constant in each voxel i.e. $\bar{x} = \{x_i,\ i = 1, \ldots, N\}$ is the emission vector which is a sample from the emission process $\bar{X} = \{X_i,\ i = 1, \ldots, N\}$ and $\bar{Y} = \{Y_j,\ j = 1, \ldots, M\}$ is the observed vector which is a sample from the measurement process $\bar{Y} = \{Y_j,\ j = 1, \ldots, M\}$. Y_j is the number of

photon counts in the jth detector) or detector tube in PET).

E-step of the EM algorithm : The algorithm of eqn. (5.1) is taken and the expectation for given $\lambda = \hat{\lambda}^k$ is obtained.

$$\log P_y(Y, \lambda) = E_x[\log P_x(x, \lambda) \mid Y = y, \hat{\lambda}^k]$$

$$- E_x[\log P_{x/y}(x \mid Y = y, \lambda)\, Y = y; \hat{\lambda}^k] \quad \ldots (5.2a)$$

as

$$E_x[\log P_y(Y, \lambda)\, Y = y; \hat{\lambda}^k] = \log P_y(Y, \lambda).$$

Equating the first term on the r.h.s. of eqn. (5.2a) with $U(\lambda, \hat{\lambda}^k)$ and the second term with $V(\lambda, \hat{\lambda}^k)$

$$\text{Log } P_y(Y; \lambda) = U(\lambda, \hat{\lambda}^k) - V(\lambda, \hat{\lambda}^k). \quad \ldots (5.2b)$$

From Jensen's inequality (24), $V(\lambda, \hat{\lambda}^k) \leq V(\hat{\lambda}^k, \hat{\lambda}^k)$. from eqn. (5.2b) we see that for any $\lambda = \hat{\lambda}^{k+1}$ for which

$$U(\hat{\lambda}^{k+1}, \hat{\lambda}^k) \geq U(\hat{\lambda}^k, \hat{\lambda}^k)$$

$$\log P_y(Y, \hat{\lambda}^{k+1}) \geq \log P_y(Y, \hat{\lambda}^k).$$

The M-step consists of computing the maximum likelihood of λ by the iteration

$$U(\hat{\lambda}^{k+1}, \hat{\lambda}^k) = \max_{\lambda} U(\lambda, \hat{\lambda}^k)$$

where $\hat{\lambda}^{k+1}$ is the new estimate from $\hat{\lambda}^k$.

The measurement random variable Y_j belonging to the incomplete data random process, is given by

$$Y_j = \sum_{i=1}^{N} X_{ij}$$

where X_{ij} represents the unobservable random number of photons emitted in voxel i and counted in measurment j.

The emission process in the complete set of voxels is modelled as a spatially independent Poisson process as

$$P(\bar{X} = \bar{x}) = \prod_{i=1}^{N} \frac{e^{-\lambda_i} \lambda_i^{x_i}}{x_i!}, \quad \ldots (5.3)$$

The measurement process is also modelled as an independent Poisson process. Given that the mean of the emission process is $\lambda = \{\lambda_i,\ i = 1, \ldots, N\}$, the conditional probability for observing the measurement data Y is

$$P(Y = y \mid \lambda) = \prod_{j=1}^{M} \frac{e^{-\sum_{i=1}^{N} \lambda_i p_{ij}} \left(\sum_{i=1}^{N} \lambda_i p_{ij} \right)^{y_j}}{Y_j!}. \qquad \ldots (5.4)$$

P_{ij} = probability that emission from voxel i will be detected in tube j. This eqn. (5.4) is also the likelihood function $L(\lambda)$ of the observed data.

Since $\tilde{X}_{ij}$ represents only those photons of the emission process X_i which is detected in measurment j, $\tilde{X}_{ij}$ is obtained by thinning X_i with probability p_{ij}. Since X_i is Poisson, $\tilde{X}_{ij}$ is also a Poisson random variable with mean $p_{ij}\ \lambda_i$. The complete data likelihood function is

$$P_{\tilde{X}}(\tilde{x}\,; \lambda) = \prod_{j=1}^{M} \prod_{i=1}^{N} \frac{e^{-\lambda_i p_{ij}} (\lambda_i p_{ij})^{\tilde{x}_{ij}}}{\tilde{X}_{ij}!}$$

where $\tilde{x} \in \tilde{X}$.

The log-likelihood function is

$$\log P_{\tilde{X}}(\tilde{x}\,; \lambda) = \sum_{j=1}^{M} \sum_{i=1}^{N} \{ -\lambda_i p_{ij} + \tilde{x}_{ij} \log \lambda_i p_{ij} - \log (\tilde{x}_{ij}\ !) \}.$$

E-step of the EM algorithm

$$U(\lambda, \hat{\lambda}^k) = E_{\tilde{x}} [\log P_x(x; \lambda) \mid Y = y, \hat{\lambda}^k)$$

$$= \sum_{j=1}^{M} \sum_{i=1}^{N} \{ -\lambda_i p_{ij} + E_{\tilde{x}} [\tilde{x}_{ij} \mid Y = y, \hat{\lambda}^k] \log \lambda_i p_{ij} \} + C.$$

$\hat{\lambda}^k$ is the current estimate of λ and C contains λ independent terms. Since the observed measurement data $y_j = \sum_i \tilde{x}_{ij}$ for a fixed data y_j, the random variables $\tilde{x}_{ij}, i = 1, \ldots, N$ can be assumed to be multinomially distributed. The expectation $E_{\tilde{x}} [\tilde{x}_{ij} \mid Y = y, \hat{\lambda}^k]$ for these multinomially distributed random variables is

$$E_{\tilde{x}} [\tilde{x}_{ij} \mid Y = y, \hat{\lambda}^k] = \frac{\hat{\lambda}_i^k p_{ij} y_j}{\sum_i \hat{\lambda}_i^k p_{i,j}}.$$

So

$$U(\lambda, \hat{\lambda}^k) = \sum_{j=1}^{M} \sum_{i=1}^{N} \left\{ -\lambda_i p_{ij} + \frac{\hat{\lambda}_i^k p_{ij} y_j}{\sum_i \lambda i' p_{ij'}} \log \lambda_i p_{ij} \right\} + C$$

The M-step is

$$\left[\frac{\partial U(\lambda, \hat{\lambda}^k)}{\partial \lambda_i} \right]_{\lambda = \hat{\lambda}^{k+1}} = \sum_{j=1}^{M} \left\{ -p_{ij} + \frac{\hat{\lambda}_i^k \, p_{ij} \, y_j}{\sum_{i'} \hat{\lambda}_{i'}^k \, p_{i'j}} \frac{1}{\hat{\lambda}_i^{k+1}} \right\} = 0$$

giving the new estimate

$$\hat{\lambda}_i^{k+1} = \frac{\hat{\lambda}_i^k}{\sum_j p_{ij}} \sum_{j=1}^{M} \frac{y_j \, p_{ij}}{\sum_{i'} \hat{\lambda}_{i'}^k \, p_{i'j}} \qquad \text{... (5.5a)}$$

This expression in additive form is

$$\hat{\lambda}_i^{k+1} = \hat{\lambda}_i^k + \frac{\hat{\lambda}_i^k}{\sum_j p_{ij}} \sum_{j=1}^{M} \frac{y_j - \sum_{i'} \lambda_{i'}^k \, p_{i'j}}{\sum_{i'} \hat{\lambda}_{i'}^k \, p_{i'j}} \, p_{ij} \qquad \text{... (5.5b)}$$

and in vector matrix notation is

$$\hat{\lambda}^{k+1} = \hat{\lambda}^k + \Gamma(\hat{\lambda}^k) \, g(\hat{\lambda}_i^k) = \hat{\lambda}^k + \Delta(\hat{\lambda}_i^k). \qquad \text{... (5.6)}$$

The diagonal matrix is given by

$$\Gamma(\hat{\lambda}^k) = \frac{\hat{\lambda}_i^k}{\Sigma_j \, p_{ij}} \qquad i = 1, \ldots, N$$

and the log-likelihood gradient vector $g(\hat{\lambda}_i^k)$ at $\lambda = \hat{\lambda}^k$ has ith component

$$[g(\hat{\lambda})]_i = \sum_j \frac{y_i - \sum_{i'} \hat{\lambda}_{i'}^k \, p_{i'j}}{\sum_{i'} \hat{\lambda}_{i'}^k \, p_{i'j}} \, p_{ij}$$

and $\Delta(\hat{\lambda}^k)$ is the correction vector.

The iterates of the EM algorithm have non-decreasing log-likelihood values, i.e. possess the monotonicity property. The non-negativity constraint is automatically satisfied if the initial estimate has non-negative pixel values, i.e. $\hat{\lambda}_i^k \geq 0$ if $\hat{\lambda}_i^0 \geq 0$. The EM estimate also has the self normalizing property of preserving the total activity of the estimate a constant, i.e.,

$$\sum_{i=1}^{N} \hat{\lambda}_i^{k+1} = \sum_{i=1}^{N} \frac{\hat{\lambda}_i^k}{\sum_j p_{ij}} \sum_{j=1}^{M} \frac{y_j \, p_{ij}}{\sum_{i'} \hat{\lambda}_{i'}^k \, p_{i'j}} = \frac{\sum_j y_j}{\sum_j P_{ij}}.$$

Since the right hand side is independent of k,

$$\sum_{i=1}^{N} \hat{\lambda}_i^{k+1} = \sum_{i=1}^{N} \hat{\lambda}_i^{k}.$$

Therefore,

$$\sum_{i=1}^{N} \Delta\, \hat{\lambda}_i^{k} = 0.$$

Assuming that all the emissions occurring in a voxel i are detected by one of the detectors, $\sum_j p_{ij} = 1$. Then $\sum_i \hat{\lambda}_i^k = \sum_j y_j$. Thus, the reconstructed image will have the same activity as in the measurements.

Although the EM algorithm produces images that have better quantitative accuracy and resolution than those obtained by the CBP method, they are nor generally suitable for use in routine diagnostic and research algorithms due to their high computational requirement at each EM iteration and slow convergence. Hence, a number of schemes, including the EM Search (EMS) algorithm[25] have been suggested for accelerating convergence.

In eqn. (5.6), the correlation vector can be considered as defining a direction to which $\hat{\lambda}^k$ moves to the new estimate. By multiplying $\Delta(\hat{\lambda}^k)$ by a factor μ^k, this movement can be enhanced. This forms the basis of the gradient search algorithms of which the EMS[26] produces the most satisfactory results for ECT imaging.

In EMS, the parameter μ^k is chosen to maximize the log-likelihood $1(\lambda)$ along a line segment $\{\lambda = \hat{\lambda}^k + \mu\, \Delta\, (\hat{\lambda}^k);\ \mu \geq 0\}$. Line search for μ^k is made in the line segment setermined by the vector inequality

$$\hat{\lambda}^k + \mu\, \Delta\, (\hat{\lambda}^k) \geq \in \hat{\lambda}^k$$

where ε is an infinitesimal positive quantity.

Defining $\gamma_j = \sum_{i=1}^{N} p_{ij}\, \hat{\lambda}_i^k$ and

$\delta_j = \sum_{i=1}^{N} p_{ij}\, [\Delta\, (\hat{\lambda}^k)]_i$, the log-likelihood becomes

$$1(\mu) = [-\gamma_j - \mu\, \delta_j + Y_j \log\, (\gamma_j + \mu\, \delta_j)\,] + \text{constant}.$$

The maximum value is obtained as usual. To maintain the positivity of the estimate, the line search is made in the range $1.0 \leq \mu \leq \mu_{max}$. Here

$$\mu_{max} = \min_i\ \mu_i'$$

where $\mu_i' = \infty$ if $[\Delta\, (\hat{\lambda}^k)]_i \geq 0$

$$= \frac{(\varepsilon - 1)\,\hat{\lambda}_i^k}{[\Delta(\hat{\lambda}^k)]_i} \quad \text{if } [\Delta(\hat{\lambda}^k)]_i < 0$$

Optimum μ_k can be obtained using the Newton-Raphson method or simply by checking for maximum $1(\mu)$ for a number of intermediate in the specified range.

Rajeevan[26] has found that the EMS algorithm accelerated the convergence twofold. However, he has developed a class of vector extrapolation based iterative algorithms, coupled with the above mentioned gradient based ML estimation (MLE) algorithms for convergence of base iterations, which are claimed to substantially reduce the number of iterations required for obtaining MLE. A presentation of these two algorithms along with results obtained on application to PET constitute the subject matter of the next section.

6. Vector Extrapolated Fast ML Algorithms

In the nonlinear vector extrapolated fast ML algorithms of Rajeevan[26], each cycle of the algorithm consists of generating a finite number consecutive estimate by a gradient based algorithm.

The maximum likelihood estimate $\hat{\lambda}^{ML}$ to which the iterates generated by a gradient based algorithm converges is

$$\hat{\lambda}^{ML} = \hat{\lambda}^0 + \sum_{i=0}^{\infty} \Delta(\hat{\lambda}^i) = \hat{\lambda}^k + \sum_{i=0}^{\infty} \Delta(\hat{\lambda}^{k+1})$$

for the kth iteration.

The estimates converge asymptotically to $\hat{\lambda}^{ML}$ and

$$|| \Delta(\hat{\lambda}^k) || \to 0 \text{ as } k \to \infty.$$

At each step k, the new estimate is considered to be generated by a nonlinear operator A_k, i.e.

$$\hat{\lambda}^{k+1} = A_k\, \hat{\lambda}^k.$$

The two cyclic algorithms presented below use a local linearization of an operator A. In each cycle, a small number m of consecutive estimates are generated by a fixed linear operator using a gradient based algorithm, i.e.

$$\hat{\lambda}_n^k = B_n\, \hat{\lambda}_n^{k-1} + b_n, \; k = 1, \ldots, m \qquad \ldots (6.1)$$

where B_n is the linear operator, b_n is a constant vector and an initial estimate obtained in a given cycle n of the vector extrapolation technique, which is used as a starting estimate of the next cycle. $\hat{\lambda}_{n,m}$ is computed as a weighted sum of the estimates $\hat{\lambda}_n^k$, $k = 0, \ldots, m$. In the Minimum Polynomial Extrapolation (MPE) method, the weights are computed from the coefficients of the minimum polynomial

of the linear operator B_n. In the reduced rank Extrapolation (RRE) method, $\hat{\lambda}_{n,m}$ is computed by adding a weighted sum of the correction vectors $\Delta(\hat{\lambda}_n^k)$, $k = 0, \ldots, m-1$ to the initial estimate of the cycle, where

$$\Delta(\hat{\lambda}_n^k) = \hat{\lambda}_n^{k+1} - \hat{\lambda}_n^k.$$

The procedure for obtaining the weights in each of these two methods is given below.

For the MPE algorithm, the extrapolated estimate is

$$\hat{\lambda}_{n,m} = \sum_{k=0}^{m} w_k \hat{\lambda}_n^k$$

with w_k being the weights.

The minimal polynomial of an operator B w.r.t. a vector v is the unique monic polynomial $P(B)$ of minimal degree which annihilates the vector v. The minimal polynomial $P(B_n)$ for the linear operator B_n w.r.t. the correlation vector $\Delta(\hat{\lambda}_n^0)$ is thus

$$P(B_n)\ \Delta(\hat{\lambda}_n^0) = 0$$

where

$P(B_n) = \sum_{k=0}^{m} \{c_K B_n^k\}$; m being the degree of the minimal polynomial and c_k ($k = 0, \ldots, m$) are the coefficients with $c_m = 1$. The number of estimates used for extrapolation is mathematically equal to the degree of the minimal polynomial m. From the definition of minimal polynomial

$$\sum_{k=0}^{m} c_K B_m^k\ \Delta(\hat{\lambda}_n^0) = 0$$

using the definition of $\Delta(\hat{\lambda}_n^k)$ viz.,

$$B_n\ \Delta(\hat{\lambda}_n^k) = \Delta(\hat{\lambda}_n^{k+1})$$

$$\sum_{k=0}^{m} c_k\ \Delta(\hat{\lambda}_n^k) = 0$$

or, as $P(B_n)$ is a monic polynomial, $c_m = 1$,

$$\sum_{k=0}^{m-1} c_k \Delta(\hat{\lambda}_n^k) = -\Delta(\hat{\lambda}_n^m)$$

or, in vector matrix notation,

$$Dc = -\Delta(\hat{\lambda}_n^m) \qquad \ldots (6.2a)$$

where D is the m dimensional column vector

$$D = [\, \Delta(\hat{\lambda}_n^0)\, \Delta(\hat{\lambda}'_n) \ldots \Delta(\hat{\lambda}_n^{m-1}) \,]$$

and c is the vector of coefficients

$$c = [c \;\; c \;\; \ldots \;\; c_{m-1}]^t.$$

The least square solution to the overdetermined system of eqn. (6.2a) is

$$c = -D^{\dagger}\, \Delta(\hat{\lambda}_n^m). \qquad \ldots (6.2b)$$

$D^{\dagger}$ being the Moore Penrose generalised inverse of D

$$D^{\dagger} = [D^t D]^{-1} D^t$$

Thus the coefficients c_k are computed.

If 1 is not an eigenvalue of B_n, then $\hat{\lambda}_{n,m}$ is unique and satisfies

$$\hat{\lambda}_{n,m} = B_n\, \hat{\lambda}_{n,m} + b_n \qquad \ldots (6.3)$$

Using the definition of $\Delta(\hat{\lambda}_n^k)$ and eqn. (6.1)

$\Delta(\hat{\lambda}_n^0) = (B_n - I)\, \hat{\lambda}_n^0 + b_n$, I being the identity matrix. Substituting for b_n from Eqn. (6.3)

$$\Delta(\hat{\lambda}_n^0) = (B_n - I)\, (\hat{\lambda}_n^0 - \hat{\lambda}_{n,m})$$

Premultiplying both sides of the above equation by $P(B_n)$ and recalling the definition $P(B_n)$

$$P(B_n)\, (B_n - I)\, (\hat{\lambda}_n^0 - \hat{\lambda}_{n,m}) = 0.$$

Since $(B_n - I)$ commutes with any polynomial in B_n and is assumed invertible

$$P(B_n)\, (\lambda - \hat{\lambda}_{n,m}) = 0$$

or

$$\sum_{k=0}^{m} c_k B_n\, (\hat{\lambda}_n^0 - \hat{\lambda}_{n,m}) = 0.$$

Using eqns. (6.1) and (6.3), the above equation becomes

$$\sum_{k=0}^{m} c_k\, (\hat{\lambda}_{n,m}) = \sum_{k=0}^{m} c_k\, \hat{\lambda}_m^k.$$

So,

$$\hat{\lambda}_{n,m} = \sum_{k=0} w_k \hat{\lambda}_m^k$$

with the weights given by

$$w_k = \frac{c_k}{\sum_{j=0}^{m} c_j}$$

In the second vector extrapolation method, RRE, $\hat{\lambda}_{n,m}$ for each cycle n is obtained by adding a weighted sum of correction vectors $\Delta(\hat{\lambda}_n^k)$ to the initial estimate $\hat{\lambda}_n^0$, i.e.

$$\hat{\lambda}_{n,m} = \hat{\lambda}_n^0 + \sum_{k=0}^{m-1} w_k \, \Delta(\hat{\lambda}_n^k). \qquad \ldots (6.4)$$

As in MPEML, we adopt a local linearization scheme using the linear operator B_n. We also assume that 1 is not an eigenvalue of B_n. Thus eqns. (6.1) and (6.3) still remain valid.

Multiplying both sides eqn. (6.4) by B_n

$$B_n \hat{\lambda}_{n,m} = B_n \hat{\lambda}_n^0 + \sum_{k=0}^{m-1} w_k \, B_n \Delta(\hat{\lambda}_n^k)$$

which becomes

$$\hat{\lambda}_{n,m} = \hat{\lambda}_n^1 + \sum_{k=0}^{m-1} w_k \Delta(\hat{\lambda}_n^{k+1}). \qquad \ldots (6.5)$$

Subtracting eqn. (6.4) from (6.5)

$$0 = \Delta(\hat{\lambda}_n^0) + \sum_{k=0}^{m-1} w_k \left\{ \Delta(\hat{\lambda}_n^{k+1}) - \Delta(\hat{\lambda}_n^k) \right\}$$

which can be rewritten as

$$\sum_{k=0}^{m-1} w_k \Delta^2(\hat{\lambda}_n^k) = -\Delta(\hat{\lambda}_n^0)$$

with $\Delta^2(\hat{\lambda}_n^k) = \Delta(\hat{\lambda}^{k+1}) - \Delta(\hat{\lambda}^k)$

or, in matrix notation

$$D^2 w = -\Delta(\hat{\lambda}_n^0).$$

The matrix D^2 is

$$D^2 = [\Delta^2(\hat{\lambda}_n^0) \, \Delta^2(\hat{\lambda}_n^1) \ldots \Delta^2(\hat{\lambda}_n^{m-1})]$$

and the vector of weights

$$w = [w_0\ w_1\ \ldots\ w_{m-1}]^t$$

The least squares solution for obtaining the weights is

$$w = -[D^2]^{\dagger}\,\Delta(\hat{\lambda}_n^0) \qquad \ldots (6.6)$$

where $[D^2]^{\dagger}$, the Moore-Penrose generalized inverse of matrix D^2 is computed using

$$[D^2]^{\dagger} = [\,[D^2]^t D^2\,]^{-1}\,[D^2]^t.$$

The extrapolated matrix is

$$\hat{\lambda}_{n,m} = \hat{\lambda}_n^0 + Dw = \hat{\lambda}_n^0 - D[D^2]^{\dagger}\,\Delta(\hat{\lambda}_n^0)$$

In the next cycle, the starting estimate is

$$\hat{\lambda}_{n+1}^0 = \hat{\lambda}_{n,m}.$$

The cyclic iteration is continued until acceptable convergence is obtained. This is the reduced rank estimate as the matrices D, D^2 each consist of m column vectors having ranks smaller than B_n. If the number of estimates equals the dimensions of solution space, D D^2 are square matrices. The expression for full rank extrapolation is obtained as follows.

$$\hat{\lambda}_{n,m} = -(B_n - I)^{-1}\,b_n \qquad \ldots (6.7)$$

Provided that $(B_n - I)$ is invertible.

From the definitions

$$\Delta(\hat{\lambda}_n^0) = (B_n - I)\,\hat{\lambda}_n^0 + b_n \qquad \ldots (6.8)$$

$$\Delta^2(\hat{\lambda}_n^k) = (B_n - I)\,\Delta(\hat{\lambda}_n^k) \qquad \ldots (6.9)$$

Substituting for b_n from eqn. (6.8) into (6.7)

$$\hat{\lambda}_{n,m} = \hat{\lambda}_n^0 - (B_n - I)^{-1}\,\Delta(\hat{\lambda}_n^0).$$

From eqn. (6.9) the matrix D^2 is

$$D^2 = (B - I)\,D$$

So $\hat{\lambda}_{n,m}$ for full rank extrapolation is

$$\hat{\lambda}_{n,m} = \hat{\lambda}_n^0 - D[D^2]^{-1}\,\Delta(\hat{\lambda}_n^0).$$

This equation for full rank extrapolation is smaller to that of reduced rank extrapolation in which the inverse of D^2 is replaced by its generalized inverse.

As with EM algorithms, the MPEML and RREML algorithms possess the self normalizing property if the gradient based algorithms generating the intermediate estimates possess the self normalizing property. Some of the weights can be negative

as they are obtained through minimum norm solutions. It may be possible that, when image values are small, the extrapolated image becomes negative. These negative pixels are offset to a very small positive value at the onset of the next cycle.

The error introduced in the estimate due to the local linearization assumption is reduced by the process of cyclic iteration. Sidi[27] has shown that, with cyclic iteration, nonlinear vector extrapolation methods generally converge quadratically.

Results Obtained for a Computer Simulated 2-d PET System

Rajeevan[26] has found that PET (the geometry of the simulated 2D PET system is given in Fig. 4) images for both of the vector methods coupled with the EM gradient based algorithms, namely EMMPE and EMRRE, i.e. 2 cycles of EMMPE and EMRRE (i.e. $n = 2$) with 3 EM iterations ($m = 3$), are comparable to those produced by 30 EM iterations alone. Likewise, 2 cycles of EMSMPE (EMSRRE) with 2 EMS iterations each per cycle, produces images that are comparable to those produced by 10 EMS iterations. It has also been shown[26] that 5 or 10 iterations of the EMS algorithm produces images of the same quality as 13 or 24 EM interactions,

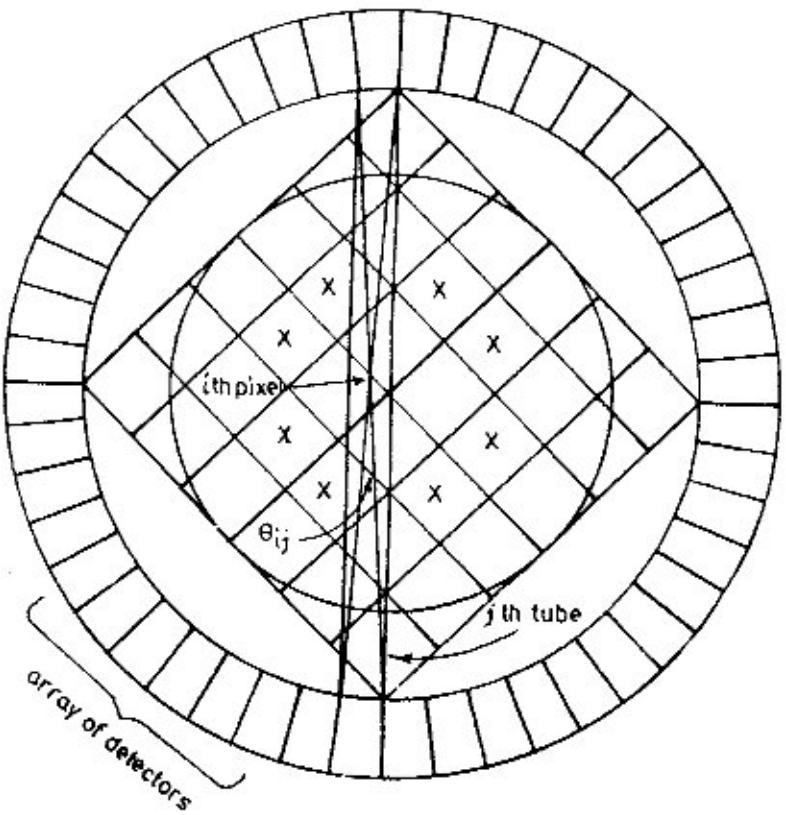

Fig. 4.

respectively. A glance at these numbers indicated that the vector methods produce substantial acceleration of convergence. These images are also found to produce images with desired likelihood values and low residual errors[26].

The open question at this point concerns the stability properties of these two cyclic algorithms. The MPEML uses first order difference vectors for computing weights, whereas the RREML algorithm uses second order difference vectors. In this sense, MPEML appears to be more stable than RREML. The number of intermediate estimates used for vector extrapolation in each cycle of the algorithm also influences the stability. In the MPEML algorithm, the degree of the minimum polynomial of B_n is not known explicitly and so an exact estimate of the intermediate estimates cannot be made. In computer simulations[26], typical values of $m = 2, 3$ are chosen. No criteria has been presented for choosing the value of m in the RREML algorithm, and a typical value of $m = 2$ has been taken. Also, the number of cycles n of the vector extrapolation has been taken to be 2 in both cases. Therefore, in both the algorithms, appropriate criteria need to be established for obtaining optimum values of m and n.

The other problem that arises is that of noise present in the images. We shall address this question in a later section.

An extension of the EM algorithm to 3-D, applicable to real SPECT imaging, along with a discussion of how beam attenuation factors can be incorporated, is considered in the subsequent section.

7. SPECT IMAGING

We have stated in the Introduction that beam attenuation factors can be incorporated in an easy manner when using series expansion methods for ECT imaging. In this section we shall show how this is achieved for SPECT imaging. Results of the previous section will be extended to obtain 3-D Maximum Likelihood images in SPECT, which incorporate corrections to minimize image degradation. In SPECT, images are degarded due to attenuation and scattering effects of photons in addition to collimator divergence. Inadequate number of detected photons in measurements also limits the accuracy of the imaging process. Scattering and septal penetration of photons result in blurred images of point sources. Both of these effects, namely, blurring and attenuation, vary with depth and location of the emission source inside the object. The width of the geometric response varies with the distance of the source from the camera and the width of the scatter response varies with the depth of the source from the medium. So the point source response of a SPECT measurement system has a 2-D distance and depth dependent functional form. Thus, SPECT image reconstruction algorithms need to account for this 2-D nature of the response functions. Compensations for the aforementioned degrading effects can be obtained through point response functions incorporated into iterative MLE algorithms. The vector extrapolated MLE algorithms described in the previous section has been extended to 3-D by Rajeevan *et al.*[28] and applied to SPECT image reconstruction. In these 3-D versions, voxel detector probabilites are computed from 2-D point source

response functions which approximate the effects of attenuation, sacttering and collinear divergence. These point source functions are modelled as 2-D Gaussian functions, assuming that the object has uniform attenuation. A projector-backprojector pair[29] is then developed and used in the implementation of the 3-D MLE algorithm.

The primary and scatter radiations emitted from the object are modelled as a pair of parallel subsystems[30] both of which are in series with the detector system[31]. The point spread function (PSF) of the total imaging system (h_t) in the spatial domain is[29]

$$h_t(d, D) = e^{-\mu D}\ h_g(d) + e^{-\mu D}\, SF(D)\, [h_g(d) ** h_s(D)]. \qquad \ldots (7.1)$$

The geometric detector PSF (h_g) and scatter PSF (h_s) are taken to be rotationally invariant PSFs in 2-D, whose parameters are functions of the distance from the collimator (d) and depth of the medium (D) of the point source and and linear attenuation coefficient of the medium (μ). A 2-D convolution between h_g and h_s is performed. The variance of the geometric PSF(σ_ρ) is assumed to be linear function of d given by (32)

$$\sigma_\rho = a_1 d + a_2$$

and the variance of the frequency domain scatter PSF($\tilde{\sigma}_s$) is taken to be

$$= b_1 e^{-b_2 D} + b_3 e^{-b_4 D}$$

giving σ_s in the spatial domain as

$$= \frac{1}{2\pi\ b_1 e^{-b_2 D} + b_3 e^{-b_4 D}}$$

and the scatter function (SF) is (32)

$$SF = \frac{c_1 e^{-c_2 D} + e^{-\mu D}}{e^{-\mu D}}.$$

The constants $a_1, a_2, b_1, b_2, b_3, b_4, c_1$, and c_2 can be estimated from SPECT acquisitions on a point source at various distance from the collinear and depths in the medium.

The reconstruction space is a cubic region enclosing a cylindrical object of uniform cross-section and is decomposed into $N \times N \times N$ voxels of which only those that are within the object are taken into account. Each projection image is made of $N \times N$ pixels. The cross section of the object is considered to be uniform. Thus, all voxels in a column in the axial direction have equal d and D and the center of the voxels project at the same lateral distance[1] onto a projection image space. These values for voxels in a cross-section for all projection angles are computed and the corresponding widths of the geometric and scatter responses are calculated and stored.

As the imaging system is considered to be a combination of two parallel subsystems and the total PSF is assumed to be the sum of the primary and scatter PSFs, the projector-backprojector and pair compute the primary and scatter component

separately and then add them together.

The projector is implemented by first computing the widths of the geometric and scatter PSFs at projection angle θ, from precomputed values of σ_g, σ_s and 1. The projection of an image is obtained by each voxel column convolved with the 2-D PSFs. The PSFs being rotationally invariant 2-D Gaussian functions are separated into axial and transverse components which are easily computed from a knowledge of width and amplitude and the lateral shift for the transverse component.

Let h_g^a and h_g^t be the $N \times 1$ vectors representing axial and transverse components of h_g respectively, and h_s^a and h_s^t be the corresponding quantities for the scatter PSF h_s, at a given angle. The ith voxel column of the image is the $N \times 1$ vector. The convolution of the total PSF h_t is done in two stages corresponding to the two terms of Eqn. (7.1). In the first stage, the primary component of the PSF is convolved. This again is obtained in two steps as follows :

Step 1 : $\tilde{v}_i = \mathcal{F}^{-1}(\mathcal{F}(h_g^a) * \mathcal{F}(v_i))$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ are the forward and inverse Fourier transforms, compuetd using FFT.

Step 2 : Each element of vector $\tilde{v}_i$ is smeared in the transverse direction with the wights givne by the transverse component of h_g^t multiplied by the attenuation factor $e^{-\mu D}$ before adding to the projection image, i.e.

$$[p_\theta]_{j_1, j_2} \leftarrow [p_\theta]_{j_1, j_2} + e^{-\mu D} [h_g^t]_{j_1} [\tilde{v}_{j_1}]_{j_2}$$

where $[p_\theta]_{j_1, j_2}$ is the (j_1, j_2) the element of the projection angle θ.

In the second stage, the scatter component is projected as

$$\tilde{\tilde{v}}_i = \mathcal{F}^{-1}(\mathcal{F}(h_s^a)\, \mathcal{F}(\tilde{v}_i))$$

and

$$[p_\theta]_{j_1, j_2} \leftarrow [p_\theta]_{j_1, j_2} + SF\, [h_s^t * h_g^t]_{j_1} [\tilde{\tilde{v}}_i]_{j_2}$$

All voxel columns are thus projected in all projection angles.

Back-projection, as the name implies, consists of backprojecting each image onto a voxel column. An $N \times 1$ column vector c is first obtained by taking a weighted sum of each row of projection image with weights given by h_g^t i.e.

$$[c]_{j_2} = \sum_{j_1 = 1}^{N} [h_g^t]_{j_1} [p_\theta]_{j_1, j_2}$$

for $j_2 = 1, \ldots, N$. c is then convolved with the axial component of the geometric PSF (h_g^a) to obtain $\tilde{c}$,

$$\tilde{c} = \mathcal{F}^{-1}(\mathcal{F}(h_g^a) * \mathcal{F}(c)).$$

The primary component of the projection image is backprojected onto vexel column v_i.

$$v_i \leftarrow v_i + e^{-\mu D}\,\tilde{c}.$$

The scatter is likewise obtained in two steps, viz.,

$$\textit{Step 1}: \quad [c]_{j_2} = \sum_{j_1=1}^{N} [h_g^t * h_s^t]_i\,[p_\theta]_{j_1, j_2}$$

$$j_2 = 1, \ldots, N$$

$$\textit{Step 2}: \quad v_i \leftarrow v_i + e^{-\mu D}\, SF\,(c * (h_g^a * h_s^a)).$$

In 3-D SPECT imaging, each detector is indexed by its position (j_1, j_2) and projection angle θ. So, the index of a detector becomes (θ, j_1, j_2) instead of j in eqn. (5.5b). Thus, the 3-D EM algorithm gives the estimate for the image density

$$\hat{\lambda}_i^{k+1} = \hat{\lambda}_i^k + \frac{\hat{\lambda}_i^k \sum_{\theta} \sum_{j_1, j_2} y_{\theta, j_1, j_2} - \hat{y}_{\theta, j_1, j_2}^k}{\sum_{\theta} \sum_{j_1, j_2} p_{i, \theta, j_1, j_2} \quad \hat{y}_{\theta, j_1, j_2}^k} \; p_{i, \theta, j_1, j_2}$$

where y_{θ, j_1, j_2} is the observed measurement data in the (θ, j_1, j_2) the detector p_{i, θ, j_1, j_2} is the probability that a photon emitted in voxel i will be detected in the (θ, j_1, j_2)th detector and

$$\hat{y}_{\theta, j_1, j_2}^k = \sum_i \hat{\lambda}_i^k \; p_{i, \theta, j_1, j_2}$$

is the projection of image $\hat{\lambda}^k$ onto the (j_1, j_2)th pixel at angle θ.

Results for a Real SPECT Imaging System

It has been found by Rajeevan *et al.* (26,29) that 4 cycles of the 3D Expectation Maximization RRE, (EMRRE) with 2 iterations per cycle have the same log-likelihood value and residual image as 26 iterations of the 3D EMMLE image. Transverse and coronal (of Liver and Spleen) views[26,29] indicate that 4 cycles of the 3D EMRRE algorithm with 2 iterations per cycle produce images of comparable quality as 26 iterations of the unaccelerated EMMLE algorithm. Thus vector acceleration techinques also reduce computation time in real SPECT imaging.

As in PET imaging, noise in the image is a problem and we shall treat this subject in the following section.

8. Noise in ECT Images and Their Rectification

Maximum Likelihood estimates of emission densities in ECT images produced

by iterative algorithms such as those described in the last two sections produce images that have better quantitative accuracy and resolution than CBP images. However, one problem associated with MLE algorithms is the slow convergence which has been alleviated by vector extrapolation algorithms discussed before. The second problem is due to the noise generated in the image in successive iterations.

Two approaches have been adopted to minimize this noise. Several authors[33] have suggested that ML iterations should be terminated before the images turn noisier. However, this approach has consistently failed[34] with real PET data. The second approach is to use prior information about the object to be reconstructed.

In cases where some prior knowledge about the object distribution is available, it can be formulated into a probability distribution for the parameters λ. These mean parameters λ_j can be assumed to be random variables. Image reconstruction in these situations is performed in a Bayesian framework.

According to Bayes' theorem, the posterior density function $P(\lambda \mid Y)$ is given by

$$P(\lambda \mid Y) = \frac{P(Y \mid \lambda)\, P(\lambda)}{P(Y)} \qquad \ldots (8.1)$$

where $P(\lambda \mid Y)$ is the likelihood function. $P(\lambda)$ the prior distribution and other terms have their usual significance.

Taking the logarithm of both sides of eqn. (8.1)

$$\log P(\lambda \mid Y) = \log P(Y \mid \lambda) + \log P(\lambda) - \log P(Y). \qquad \ldots (8.2)$$

The maximum a posteriori (MAP) estimate $\hat{\lambda}^{MAP}$ which is the estimate of λ which has the maximum conditional probability for measurement data Y is obtained by maximizing $\log P(\lambda \mid Y)$.

The log-likelihood function is obtained as described earlier and $\log P(Y)$ is independent of λ. Thus, in order to minimize $\log P(\lambda \mid Y)$, information about the prior distribution is required.

Rajeevan[26] has tried three distributions-Gaussian, Gamma and Gibbs for the prior image reconstruction implemented along with MAP algorithms. A penalty of deviating from the mean image, known apriori, is imposed. In realistic situations, such images are rarely known apriori. Ideally, in Bayesian reconstruction, it is desirable to use distributions for which the mean image does not have to be known apriori, or at least, to use image distributions for which a minimal set of parameters need to be known. Of three distributions, the Gibbs distributions best fits into this category. Therefore, the use of Gibbs functions as prior distributions in Bayesian image reconstruction has been explored.

The Gibbs distribution, which is an example of a Markov Random Field (MRF) is a suitable stochastic model for the images produced in ECT and can incorporate image characteristics such as spatial smoothness and edge properties through local correlations. Some relevant properties of MRFs are described below. Other properties can be found in Dubes and Jain[35].

S is defined as the set of indices for the voxels in an image. Then the image $\lambda = \{\lambda = \lambda_i,\ i = 1, \ldots, N\}$ can be considered to be a collection of N random variables defining a random field on the index set S. The image λ is an MRF if the conditional probability for an image voxel, given the values for all other voxels in the image has the property

$$P(\lambda_i/\lambda_{S/i}) = P(\lambda_i/\lambda_{N_i}) \quad \forall_i \in S$$

where S/i denotes the set of all indices in S excluding the index i. $N_i \in S$ denotes a set of indices forming a neighbourhood for voxel i and $\lambda_{S/i}$ and λ_{N_i} are, respectively, random variables defined on these sets. For 2-D image defined on a rectangular grid, the 1st order neighbourhood consists of all the 4-neighbourhoods of a given pixel and a 2nd order neighbourhood consists of all the 8-nieghbours. For a given neighbourhood type, a set of sites is called a 'clique' if each site in the set is a nighbour of every other site in the set. Following[26], the Gibbs function used is one for which the potentials are nonzero only on cliques of pairs of sites.

The Gibbs distribution for the prior function of eqn. (8.1) is thus

$$P(\lambda) = 1/z \exp\left\{-\frac{1}{\beta} \sum_{(i,\, i_1)} V(\lambda_i, \lambda_{i_1})\right\}$$

where $(i,\ i_1)$ are pairs of neighbouring sites. Using this expression, the log-posterior function is

$$\log\ P(\lambda \mid Y) = \sum_{j=1}^{M} \left\{ - \sum_{i=1}^{N} \lambda_i p_{ij} + y_i \log \left(\sum_{i=1}^{N} \lambda_i p_{ij} \right) \right\}$$

$$- \frac{1}{\beta} \sum_{(i,\, i_1)} V(\lambda_i, \lambda_{i_1}) + \text{terms independent of } \lambda.$$

The expectation step of the EM-MAP algorithm with this Gibbs prior is

$$U(\lambda, \hat{\lambda}^k) = E_x [\log P(\lambda/x) \mid Y, \hat{\lambda}_k]$$

$$= E_x [\log P(x \mid \lambda) \mid Y, \hat{\lambda}^k] + \log P(\lambda) + C$$

$$= \sum^{M} \sum^{N} \left\{ - \lambda_i P_{ij} + \frac{\hat{\lambda}_i^k y_i p_{ij}}{\sum_{i'} \hat{\lambda}_{i'}^k p_{i'j}} \log \lambda_i p_{ij} \right\} - \frac{1}{\beta} \sum_{i,\, i_1} V(\lambda_i, \lambda_{i_1})$$

$$+ \text{terms independent of } \lambda.$$

In the maximization step, $U(\lambda, \hat{\lambda}^k)$ is maximized w.r.t. λ_i, i.e.

$$\frac{\partial U}{\partial \lambda_i} (\lambda, \hat{\lambda}^k) = \sum_{j=1}^{M} \left\{ - P_{ij} + \frac{\hat{\lambda}_i^k y_i p_{ij}}{\sum_{i'} \hat{\lambda}_{i'}^k p_{i'j}} \cdot \frac{1}{\lambda_i} \right\} - \frac{1}{\beta} \sum_{i_1 \in N_i} \frac{\partial V}{\partial \lambda_i} (\lambda_i, \lambda_{i_1}) = 0.$$

N_i being a set of neighbours for site i. The complicated form of

$$\sum_{i_1} \frac{\partial V(\lambda_i, \lambda_{i_1})}{\partial \lambda_i}$$

renders it difficult to solve the coupled equations. In the One Step late (OSL) algorithm of Green[36], $V(\lambda_i, \lambda_{i_1})$ os replaced by $V(\hat{\lambda}_i^k, \hat{\lambda}_{i_1}^k)$ in the potential and enery of the new estimate $\hat{\lambda}^k$. Thus, with the OSL algorithm, the EM-MAP estimate is

$$\hat{\lambda}_i^{k+1} = \frac{\hat{\lambda}_i^k}{\sum_j p_{ij} + \frac{1}{\beta} \sum_{i_1 \in N_i} \left[\frac{\partial V}{\partial \lambda_i} (\lambda_i, \lambda_{i_1}) \right]_{\lambda = \hat{\lambda}^k}} \sum_{j=1}^{M} \frac{y_i p_{ij}}{\sum_l \lambda_l^k p_{lj}}$$

when $\beta \to \infty$, the OSL algorithm reduces to the EM-MLE algorithm.

In order to eliminate the non-smoothness in the ML reconstructed images, the potential function chosen should have a form that penalizes non-smooth images. The following choices have been used (26 and references therein)

$$V_1(\lambda_i, \lambda_{i_1}) = \left(\frac{\lambda_i - \lambda_{i_1}}{\mu} \right)^2$$

$$V_2(\lambda_i, \lambda_{i_1}) = \frac{(\lambda_i - \lambda_{i_1})^2}{\mu^2 + (\lambda_i - \lambda_{i_1})^2}$$

$$V_3(\lambda_i, \lambda_{i_1}) = \log \left\{ 1 + \left(\frac{\lambda_i - \lambda_{i_1}}{\mu} \right)^2 \right\}$$

$$V_4(\lambda_i, \lambda_{i_1}) = \log \left\{ \cosh \left(\frac{\lambda_i - \lambda_{i_1}}{\mu} \right) \right\}$$

V_1 is found to penalize deviations between neighbouring pixels excessively, resulting in an oversmooth image. On the other hand, V_2 incorporates a cut off value of 1.0 and so large differences in neighbouring pixels due to an edge are insufficiently penalized. V_3 and V_4 are compromised between the two extremes.

In the use of Gibbs priors, the two parameters μ and β need to be specified. These have been chosen, for instance in[26], by trial and error. The estimation of these parameters is an open area of research. As a matter of fact, parameter specification of MArkov random Fields is one of the difficult problems that limit the usage of these models for image restoration, edge detection and segmentation[37]. These parameters are often determined by trial and error[38] as in the case of ECT imaging discussed above.

9. Magnetic Resonance Imaging (MRI)

The principles underlying both the commonly used techniques of MRI are basically inverse transform type methods. The physical properties used in this imaging

modality are usually modelled deterministically. In order to provide a clearer understanding of the mathematical formulation of these techniques, a brief overview of the physical processes is presented first.

Nuclear magnetic moments of an object placed in a magnetic field align themselves parallel to the direction of the field, which we consider to be the z direction. If an oscillating magnetic field is applied along a perpendicular direction (e.g. the or the y direction), the magnetic polarization vector is deflected from the z-direction when the z field approaches the reasonance value[13, 39]. The resonance condition is

$$\gamma B = \omega$$

where B is the amplitude of the applied static field, γ the gyromagnetic ratio and ω, the Larmor frequency. The rotation of the polarization vector in a plane perpendicular to the z axis induces an e.m.g. in the detector coil which constitutes the nuclear magnetic resonance (n.m.r.) signal. In practice, nuclear spins are excited with short r.f. pulses and the response of the spin system to such pulses is the free induction decay (f.i.d.) signal.

When magnetic field gradients are superimposed onto the main magnetic field, the resonance frequency becomes a function of the spatial origin of the signal. Due to this gradient field, the frequency domain signal is the equivalent of a projection of the object onto a gradient axis. Lauterbur[40], first proposed the idea of *nmr* zeugmatography, which consists of generating spatial maps of such distributions. In the *CBP* method discussed earlier the gradient is rotated in small angular increments and a series of projections obtained. Image reconstruction is performed using 3-D backprojection.

In the second technique known as Fourier zeugmatography, proposed by Kumar *et al.*[41], a spatial encoding process is involved, which, in effect Fourier transforms the spin density (i.e. the picture function) with appropriately scaled variables to obtain the f.i.d. signal. This f.i.d. signal is then inverse transformed to reconstruct the image. Due to the nature of the physical processes involved, the reconstruction data are automatically discretized.

Consider the spins of spin density $\rho(x, y)$ to be excited by a r.f. pulse due to the magnetic field gradient G_y (along y) called the phase encoding gradient. The spins resonate with relative frequency.

$$\omega(y) = \gamma G_y y.$$

If this gradient for time interval t_y, the phase at the end of this signal is

$$\phi_y = \gamma G_y y t.$$

After interval t_y, G_y is turned off and an orthogonal gradient G_x is applied for a duration t_x when the f.i.d. signal is collected as shown in Fig. 5. During this detection period, the spins process at a frequency

$$\omega(x) = \gamma G_x x.$$

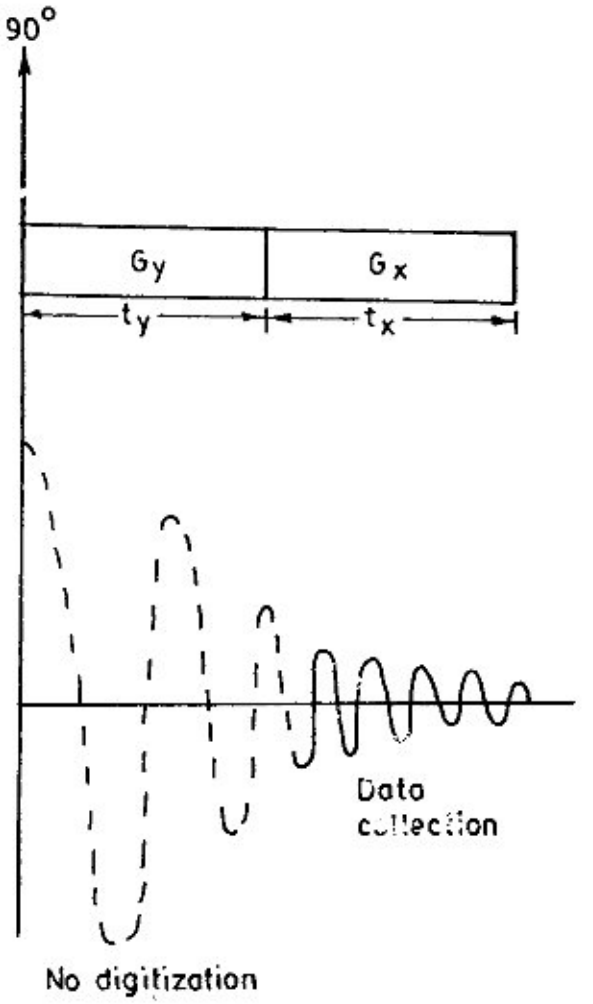

FIG. 5.

Thus, spatial information is encoded into both the phase and frequency of the nmr signal whose amplitude is proportional to $\rho(x, y)$.

An image can be regarded as a 2-D array of signal amplitudes versus spatial coordinates (x, y). The signal corresponding to two subsequent $\pi/2$ pulses of magnitudes G_y and G_x applied during intervals t_y and t_x, respectively, is given by

$$S(t_x, t_y) = \int_{-\infty}^{\infty}\int \rho(x, y) \exp(i\omega_x t_x + i\phi_y)\, dxdy$$

$$= \int_{-\infty}^{\infty}\int \rho(x, y) \exp\{i\gamma(G_x x t_x + G_y y t_y)\}\, dxdy$$

The values of $S(t_x, t_y)$ when appropriately scaled, represent 2-D Fourier coefficients of spin density, the fourier conjugate variables being $\omega_x = \gamma G_x t_x x$ and $\omega_y = \gamma G_y y t_y$.

Spin-spin relaxation (the return of the spin system to equilibrium) effects can easily be incorporated by damping the f.i.d. with the exponential decay term $\exp(-t/T_2)$ in eqn. (9.1), i.e.,

$$S(t_x, t_y) = \int_{-\infty}^{\infty} \int \rho(x, y) \exp \{i\gamma (G_x x t_x + G_y y t_y) - (t_x + t_y)/T_2\} \, dx dy. \quad \ldots (9.2)$$

During the interval t_x, the *nmr* signal is sampled at regular intervals T, so that in a single f.i.d., a set of Fourier coefficients $S(t_0, \tau)$; $S(t_0, 2\tau)$, ..., $S(t_0, N_x \tau)$ is obtained. τ is chosen so that it conforms to the sampling theorem. If τ is too large, aliasing effects occur, and if τ is too small, the object will fill only a small part of the image. The optimal value for a sample of extent L_x along x, is obtained as $\gamma G_x L_x \tau = \pi$,

The total number of sample points N_x determines the number of pixels and hence the spatial resolution in the x- direction. Raising the amplitude of G_y incrementally during each excitation and read cycle, as in spin warp imaging[42] yields an array $N_x x N_y$ of raw data which is double Fourier transformed in order to reconstruct the image which consists of $N_x x N_y$ pels. The process of incrementally increasing the amplitude of the gradient to encode spatial information into the phase of precessing spins is termed as phase encoding.

The above procedure can be extended to 3-D by introducing a variable period of evaluation t_z, during which a third mutually orthogonal field G_z is applied. Similar analyses are carried out but with two phase encoding stages instead of one.

Echoplanar imaging, first introduced by Mansfield (43) is a modification of MRI applicable for imaging physiological motion. In this scheme, many lines of data are collected from a single excitation, each being encoded separately. The entire data space is mapped during a single read cycle with a sufficiently fast sample rate.

Spatial resolution is an issue which needs improvement. However, since the limitations are physical, a discussion of this problem is beyond the scope of this article. As with other inverse transform applications, reconstruction of images from limited data remains a problem in this modality also.

10. Concluding Remarks

In summary, we have presented a tutorial introduction to the mathematical formulation of several biomedical imaging modalities which fall under two broad groups of image reconstruction techniques, namely inverse transform and series expansion methods. Reconstruction in X-ray CT and sometimes MRI is performed

using the CBP method while an iterative seris expansion algorithm based on the EM formulation of the ML estimate is used in ECT imaging modalities like SPECT and PET. Physical models in X-Ray CT and MRI are deterministic whereas a stochastic description is adhered to in ET. Fourier zeugmatography, an inverse transform type reconstruction technique widely used in MRI, is also discussed.

It is found that in all the formulations described, reconstructions of continuous density distributions of the objects are carried out from limited, finite or noisy data. Several approaches have been attempted to increase the accuracy of reconstruction from available data but these approahces tend to take too much computation time due to slow convergence, or are simplistic in nature. Therefore, vsualization in biomedical computing is limited in accuracy due to reconstructing continuous objects from limited data. It is hoped that, with the advent of new technologies and sophisticated mathematical tools, it will be possible to obtain more accurate images which will considerably simplify the task of diagnostic radiologists.

REFERENCES

1. K. Fukunaga, Introduction to Statistical Pattern Recognition, Academic Press, New York, (1972).
2. K. S. Fu, Syntactic Pattern Recognition and Applications, Prentice-Hall, New Jersey, (1982).
3. S. K. Pal and D. Dutta Majumdar, Fuzzy Mathematical Approach to Pattern Recognition, John Wiley, New York, (1986).
4. R. C. Gonzalez and P. Wintz, Digital Image Processing, Addison-Wesley, Reading, Masachussetts, (1977).
5. W. K. Pratt, Digital Image Processing, John Wiley & Sons, New York, (1978).
6. B. B. Chowdhury and D. Dutta Majumdar, Two-Tone Image Processing and Recognition, Wiley-Eastern, New Delhi, (1992).
7. D. H. Ballard and C. M. Brown, Computer Vision, Prentice Hall, New Jersey, (1982).
8. R. E. Barnhill and R. F. Reisenfeld, Computer Aided Geometric Design, Academic Press, New York, (1974).
9. K. S. Ray and D. Dutta Majumdar, *J. IETE*, 37 (1991) 443.
10. G. T. Herman, Image Reconstruction From Projections, Academic Press, New York (1980).
11. T. F. Budinger and G. T. Gullberg, *Phys. Med. Biol.* 19 (1974) 387.
12. A. Dempster, N. M. Laird and D. B. Rubin, *J. Roy. Statist.* Soc. : Series B, **39** (1977) 1.
13. P. G. Morris, Nuclear Magnetic Resonance Imaging In Medicine and Biology, Clarendon Press, Oxford (1986).
14. J. Radon, *Ber. Verb. Saechs. Akad. Wiss., Leipzig, Math. Phys. Kl.* **69**, (1917) 262.
15. S. L. Wood, A. Macovski and M. Morf in Computer Aided Tomography and Ultrasonics in Medicine, J. Raviv *et al.* (eds.) North Holland, Amsterdam (1979) 219.
16. N. Srinivasa, V. Krishnan and K. R. Ramkrishnan, Proc. IEEE EMBS, 8th Ann. Conf. Forth Worth, Texas, USA (1986) and references therein.
17. T. J. Ulrych et. al., *Rev. Geophysics*, **13** (1975) 193.
18. N. Srinivasa, K. R. Ramkrishnan and K. Rajgopal, *Proc. ISSPA*, Brisbane, Australia (1987).
19. Y. Bresler and A. Macovski, Proc. ICPR, Montreal, Canada (1984) and references therein.
20. A. P. Sage and J. L. Melsa, Estimation Theory with Applications to Communication and Control, McGraw Hill, USA (1971).
21. H. A. Taha, Integer Programming: Theory, Applications and Computations, Academic Press, New York (1975).
22. L. A. Shepp and Y. Vardi, *IEEE Trans. Med. Imaging, MI-1* (1982) 113.

23. K. Lange and R. Carson, J. Comput. Assist. Tomogr., **8** (1984) 306.
24. C. R. Rao, Linear Statistical Interference and Its Applications, John Wiley & Sons, New York, (1973).
25. L. Kaufman, *IEEE Trans. Med. Imaging, MI-6(1)* (1987) 37.

K. Lange, M. Bahn and R. Little, *IEEE Trans. Med. Imaging, MI- 6(2)*, (1987) 106.

26. N. Rajeevan, Ph. D. thesis, Indian Institute of Science, Bangalore, India, (1991), and references therein.
27. A. Sidi, *SIAM J. Numer. Annl.* **23** (1986) 197.
28. N. Rajeevan, B. C. Penney and C. Byrne, *IEEE Trans. Nucl. Sci.* in press (1991).
29. B. C. Penney, M. A. King and K. Knesaurek, *IEEE Trans. Nucl. Sci.* **37** (1990) 681.
30. C. E. Metz and K. Doi, *Phys. Med. Biol.* **24** (1979) 1079).
31. O. Ying-Lie, Ph.D. thesis, Deft University, Deft (1984).

F. B. Atkins, Ph. D. thesis, University of Chicago, (1978).

32. K. Knesaurek, Med. Phys. submitted for publication (1991).
33. E. Veklerov and J. Llacer, *IEEE Trans. Med. Imaging, MI-6* (1987) 313.

T. J. Hebert, R. Leahy and M. Singh, *IEEE Trans. Nucl.* Sci. (1988) 615.

34. J. Llacer and E. Veleroy, *IEEE Trans. Med. Imaging MI-8(2)* 1989) 186.
35. R. C. Dubes and A. K. Jain, *J. Appl. Statist., 16(2)*, (1989) 131.
36. P. J. Green, IEEE Trans. Med. Imaging, *MI-9(1)*, (1990) 84.
37. S. G. Nadabar and A. K. Jain, *IEEE Trans. Pattern Anal.* Machine Intell. *PAMI-12* (1992) 528.
38. A. K. Jain, and S. G. Nadabar in Data Analysis in Astronomy IV, V. DiGesu et al. (eds.), Plenum Press, New York (1992) 39.
39. F. W. Wehrli, *Phys. Tod. 45(6)* (1992) 34.
40. P. C. Lauterbur, *Nature 243* (1973) 190.
41. A. Kumar, D. Welti and R. Ernst, *J. Magn. Res. 18* (1975) 69; *Naturwissenschaften, 62* (1975) 34.
42. W. A. Edelstein, J. M. S. Hutchinson, G. Johnson and T. Redpath, *Phys. Med. Biol. 25* (1980) 751.
43. P. Mansfield, *J. Phys. C10* (1977) 55.

ADDENDUM

Herman and Odhner[1] have undertaken a study on the evaluation of several reconstruction algorithms, namely two iterative type Expectation Maximization (EM) methods, which are Expectation Maximization-Maximum Likelihood (EM-ML) and Expectation Maximization-Maximum Apriori Probability (EM-MAP) methods, as well as two variants of transform methods (convolution backprojection). This evaluation was carried out in a manner which satisfied statistical hypothesis testing. The evaluation criteria were chosen so as to judge the suitability of a particular algorithm based on three different tasks that they are required to perform. These three tasks consisted of the detection of the relatively higher uptake on the left side of the brain compared to its right side, the estimation of the total uptake in various neurological structures and the uptake at individual points within the neurological structures. Accordingly, three Figures of Merit (FOM) were defined for these tasks. These FOM are, respectively, the hit-ratio, the structural accuracy and pointwise accuracy.

Subjects used for the study consisted of mathematical phantoms containing neuroanatomical structures which are represented as ellipses and rectangles at

appropriate locations symmetrically situated with respect to the midline. The background activity in the brain was taken to be 1.00 while that in the neuroanatomical symmetric structures was taken as 1.95 or 2.00. A random number generator chooses exactly one of a pair to have an activity of 2.00 for each pair of symmetrical structures.

The abnormality index is the average of the reconstructed pixel values for pixels whose centers are within that structure in the phantom. A hit is considered to be made if the abnormality index is higher for that structure in a pair of symmetric structures for which the activity is higher in the phantom. The hit-ratio is the ratio of hits to the total number of pairs in the data set. If the hit-ratio of one algorithm is found to be higher than the other algorithm, the next step was to determine whether this was statistically significant. The sign test[2] was chosen as the test to reject the null hypothesis that both methods are equally good to be hypothesis that the one with the greater hit-ratio is better. In this test, pairs of structures which have been classified as different by two reconstruction methods is considered. Let C be the total number of such pairs and C be the number of pairs correctly classified by the reconstruction with total number of items C and equal probabilities assigned to the two classes. From this binomial distribution, the probability of selecting an element with value C or higher is noted. This probability gives the level of significance for ejecting the null hypothesis.

Inaccuracy is the absolute value of the difference between the abnormality indices of that structure in the reconstruction and in the phantom. Structural accuracy is the negative of the average of inaccuracy over all structures in the phantom. The level of statistical significance for the structural accuracy is obtained as follows. Let β_b and δ_b be the inaccuracies of the bth of a total of B structures reconstructed by two methods. Null hypothesis implies that $(\beta_b - \delta_b)$ is a sample of a zero-mean random variable. For a large enough B

$$\sum_{b=1}^{B} (\beta_b - \delta_b)$$

is a sample of a normally distributed zero-mean random variable. The variance is

$$\sum_{b=1}^{B} (\beta_b - \delta_b)^2.$$

The significance is calculated using the normal distribution.

The phantom and reconstruction are clipped by setting pixel values which are less than 1.65 to 1.65 and values which are greater than 2.05 to 2.05. The normalized root-mean-square distance between clipped reconstruction and clipped phantom is calculated. Pointwise accuracy is the average of the negative clipped normalized distance over all the phantoms. Statistical significance is calculated in the same way as for structural accuracy with the clipped normalized distances for whole phantoms being compared instead of inaccuracies for individual phantoms.

In the iterative algorithms studied, the commonly used practice of preconvergence termination was adopted. After plotting average values of hit-ratios versus number of iterations, the iteration number 2 was considered noteworthy as the highest hit-ratio was obtained at this iteration. The other two iteration numbers were 28 and 80. A plot of the average value of the clipped normalized distance versus the number of iterations for EM-ML and EM-MAP indicated that the average of EM-ML is minimum at iteration number 28. The average monotonically decreases for EM-MAP and does not change significantly after iteration 80.

Results showed that EM-MAP had proise if pointwise accuracy is the important FOM whereas EM-ML was best for detecting relatively higher uptake (hit-ratio) and estimating the average uptake within the structures. In conclusion, it may be remarked that the relative merit of two construction algorithms can be extremely dependent on the medical tasks.

Another class of iterative algorithms known as Multiplicative Algebraic Reconstruction Techniques (MART) have also been described, which have not been extensively used to date[3]. These methods are generated using a multiplicative updating in such a way that non-negative constraints are automatically satisfied and no parameters' choice has to be done, i.e.,

$$x_j^{k+1} = x_j^k \, c_j^k$$

where c_j^k is a positive real number. These algorithms are generally computationally expensive and the convergence rate of the algorithms as well as methods for accelerating them are open areas of research.

ACKNOWLEDGEMENT

One of the authors (SB) would like to acknowledge a grant from CSIR, India which enabled this work on Mathematical techniques in biomedical imaging.

REFERENCES

1. G. T. Herman and D. Odhner, Mathematical Methods in Tomography, Proceedings, Oberwolfach 1990, Springer-Verlag, Berlin, G. Herman, A. Louis and F. Natterer, eds, p. 215-228.
2. R. Mould, Introduction to Medical Statistics, Adam-Hilger, Bristol, England (2nd. ed.) 1989.
3. A. R. De Pierro, Mathematical Methods in Tomography, Proceedings, Oberwolfach 1990, Springer-Verlag, Berlin, G. Herman, A. Louis and F. Natterer, eds, p. 167-186.