A Comprehensive Statistical Model for Single Photon Emission Tomography

Stuart Geman, Kevin M. Manbeck, and Donald E. McClure
Reports in Pattern Analysis No. 153
Division of Applied Mathematics
Brown University
Providence, RI 02912

January 1991

[1]In *Markov Random Fields: Theory and Applications*, R. Chellappa and A. Jain, editors. Academic Press Boston, 1991.

[2]Supported by Army Research Office Contract DAAL03-86-K-0171 to the Center for Intelligent Control Systems, National Science Foundation grant number DMS-8813699, Office of Naval Research Contract N00014-88-K-0289, and the General Motors Research Laboratories.

Abstract

A Bayesian approach to the reconstruction of SPECT images is presented. The approach requires two separate mathematical models: one model (called the likelihood or data model) describes the process which generates the data; the second model (called the prior model) captures certain prior knowledge about the image under study, via a probability distribution on the ensemble of possible reconstructions. This paper presents a general degradation model for the SPECT problem. In addition, experimental methods for measuring critical parameters in this model are presented. Concerning the prior model, a probability distribution is constructed that enforces a degree of smoothness while accommodating the discontinuities associated with boundaries. Experiments with physical phantoms and patient data demonstrate the effectiveness of the approach. Finally, the isotropic properties of the prior model are examined analytically in a suitable continuum limit.

1 Introduction

The field of tomography is broadly concerned with the application of noninvasive imaging techniques to determine the internal structure of three dimensional objects. Tomographic images must be reconstructed, typically with the aid of computers and computational algorithms, from data which are collected or observed external to the object under study. In the industrial environment, tomographic techniques have found successful application to nondestructive internal inspection. But the primary applications of tomography have been in medicine, where it has been widely used over the past two decades. In particular, computed tomography (CT) scanning, ultrasound, and more recently magnetic resonance imaging (MRI), have revolutionized diagnostic medicine.

Somewhat less successful has been the introduction of emission tomogra-

phy, positron emission tomography (PET) and single photon emission computed tomography (SPECT), to the clinical setting. In these modalities, one seeks to learn the internal distribution of a pharmaceutical. This is accomplished by introducing a radiopharmaceutical—a pharmaceutical which has been chemically combined with a radioactive isotope—into a patient, and then measuring, externally, a distribution of radioactive events. These modalities produce generally lower quality images, with lower spatial resolution, than either CT or MRI.

Despite this relative lack of resolving power, PET and SPECT hold much promise because they can measure *metabolic activity*, which is impossible to measure with CT or, thus far, with MRI. Furthermore, the development of highly discriminating monoclonal antibodies for use in SPECT promises to allow clinicians to accurately locate specific tissue types, both normal and pathological.

The relatively low cost of SPECT systems, the ability of SPECT to measure metabolic activity, and the introduction of new monoclonal antibody imaging agents, combine to make SPECT a potentially important clinical tool. Unfortunately the application of SPECT is currently limited by its poor image quality.

It is popular lore—but incorrect—that a major technological or scientific advance in one tomographic modality (recently MRI) will antiquate and supplant the other modalities. In fact, each modality has a distinct diagnostic or therapeutic application, and it is therefore imperative to develop each imaging method to its fullest potential.

In this paper we will describe mathematical methods which may signifi-

cantly improve SPECT reconstruction techniques. Previous experience with these methods ([5], [16], [4], [29] [25], [9], [10], [13],) has demonstrated a potential utility of Markov random field (MRF) image models, in a Bayesian framework, for various image restoration and reconstruction tasks. There have been, in particular, several studies of MRF-based Bayesian methods ([11], [12], [22], [27], [14], [21], [6]), and related regularization methods ([28], [34]) for emission tomography. In this paper we demonstrate that these methods are well suited to take advantage of the rather detailed available information on the physics of SPECT imaging, and we derive, analytically, some isotropic properties of the particular random fields employed.

In §2 we will briefly review SPECT imaging, and identify the primary physical factors that influence the collection of data in a SPECT imaging session. The Bayesian approach is reviewed in §3, emphasizing the dual requirements for both a likelihood model, describing the distribution of the externally observable data for any fixed internal configuration of isotope intensity, and a prior model, describing *a priori* likely and unlikely configurations. The likelihood (or data) model is studied in detail in §4. The primary physical effects which govern the SPECT imaging modality have been modeled up to machine-specific parameters. Experiments were designed to measure these parameters, and the results were used to build a machine-specific model. Reconstruction experiments were performed with phantoms of known structure and with real patient data. Results are reported in §5. A detailed discussion of the prior probability model is postponed to §6, where the model is presented together with an analytic study of its isotropic properties.

2 Physics of SPECT

In a SPECT imaging session, a patient is injected with a pharmaceutical that is tagged with a radioactive isotope. The pharmaceutical is application dependent and is chosen to concentrate in a region of interest. As this so-called *radiopharmaceutical* undergoes radioactive decay, photons are released from the patient's body and counted by a gamma camera which rotates around the patient and collects data at numerous angles. Once these data are recorded, one is faced with the difficult task of reconstructing a map of the isotope, and thereby the pharmaceutical, concentration.

The physics which govern the release of a photon and the path that a photon travels are well understood:

1. Photon release follows a Poisson distribution with mean proportional to isotope concentration. In a patient study, the concentration of isotope must be kept low in order to avoid the effects of high dose radiation. This adversely affects the signal-to-noise ratio and contributes to low quality reconstructions.

2. After being released, a photon may—at random—experience absorption or Compton scatter. These two effects, together, are known as attenuation. A patient's body contains essentially three types of attenuating material: bone, soft-tissue, and lung. Hence, the attenuation of a photon is intrinsically nonuniform. The amount of scatter is depth dependent: photons that are released deep in an attenuating substance have a higher probability of scattering than those that are released near the surface. The effects of absorption and scatter further contribute to

degraded reconstructions and low resolution. See figure 1, panels a and b.

3. The geometry of the gamma camera introduces a "collimator effect" which amounts to a depth dependent blurring of the reconstructed image. See figure 1, panel c.

The collective effect of these influences on the observed photon counts can be accurately modeled and thereby used to improve reconstruction quality. Since the release of photons, absorption, and Compton scatter are all inherently random events, it is natural to model the SPECT problem in a statistical framework. In this regard, we follow the lead of Shepp and Vardi [33], who were the first to derive reconstruction algorithms for emission tomography from statistical models and basic principles of inference. (Earlier, Rockmore and Makovski [32] recognized the natural role of statistical models in formulating the emission tomography reconstruction problem.)

3 Bayesian Formulation

Commercially available SPECT machines reconstruct isotope concentrations using a version of the filtered backprojection (FB) algorithm. This approach is limited in its ability to accommodate the degrading effects discussed above. In particular, FB techniques ignore the Poisson nature of decay and typically allow for only uniform absorption and depth-*independent* scatter and blur. As a result the commercially available reconstruction methods are characterized by a lack of resolution and other blurring artifacts. Oftentimes it is safe to ignore the random elements in an image restoration task. For exam-

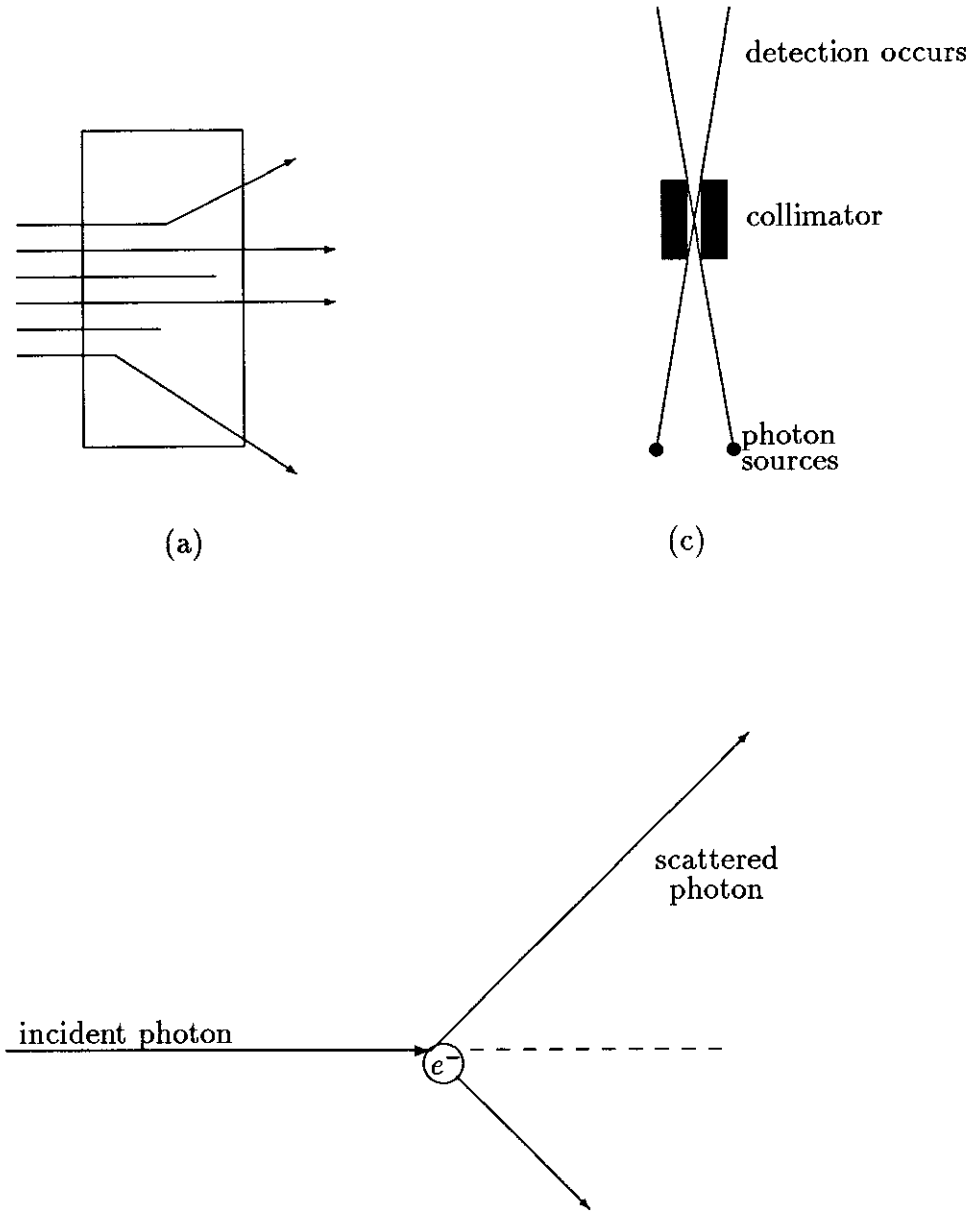


Figure 1: Relevant physical effects. (a) A photon may be attenuated by either being deflected or absorbed. (b) A photon may change direction and energy via Compton scatter with an electron, e^-. (c) Imperfect collimation introduces the solid angle effect: in addition to accepting true projections, a collimator will accept "near" projections.

ple, backprojection is used successfully in CT where the signal-to-noise ratio is favorable, and scattering and collimator effects are negligible. However, the signal-to-noise ratio in SPECT is low, and absorption, scattering, and collimator effects are highly significant.

The model developed here will explicitly account for quantum noise, nonuniform absorption, and depth dependent scatter and blur. The approach is based on realistic models of the stochastic and deterministic components of a SPECT data acquisition system.

The model involves the unknown isotope concentration map, which we denote by X, and the observable photon counts, which we denote by Y. The reconstruction problem is to estimate X from Y. The physics of SPECT imaging determine that Y is distributed according to a Poisson distribution with mean

$$E[Y] = \mathcal{A}X, \tag{1}$$

where the discrete *modified Radon transform* $\mathcal{A}$ describes the process by which an image X is transformed into data Y. $\mathcal{A}$ incorporates the effects of nonuniform absorption and depth dependent scatter and blur. Later (see §4) we will discuss how $\mathcal{A}$ can be *measured* with a simple experimental apparatus. We thus specify a Poisson *likelihood* model $P(Y|X)$ (with mean $\mathcal{A}X$) to evaluate the probability that a given isotope map gives rise to the observed data.

At this stage one could, in principle, solve the inverse problem by finding that X which maximizes $P(Y|X)$. This is, in fact, the *maximum likelihood* approach proposed by Shepp and Vardi [33] for PET imaging (see also Rockmore and Makovski [32]). Unfortunately, maximum likelihood reconstruc-

tions are critically dependent on the chosen pixel resolution, and in general, are badly degraded at clinically interesting pixel resolutions. This difficulty is not entirely unexpected: the reconstruction problem is inherently a nonparametric estimation problem. One seeks to reconstruct a completely general function describing the internal radiopharmaceutical concentration. As is well-known in the context of nonparametric estimation (see e.g. Grenander [15]), as the pixel resolution becomes finer, the variance of the maximum likelihood estimator increases. Maximum likelihood *per se* is generally not consistent for nonparametric estimation. Some sort of regularization is needed.

We regularize using a Bayesian framework in which a "smoothing prior" $P(X)$ is placed on the isotope concentration map X. More specifically, we propose to model, via a prior probability distribution, the expectation that isotope concentration maps are likely to consist of locally smooth regions separated by discontinuities ("boundaries"). (See §6 for details.) Applying Bayes' formula we determine the *posterior distribution*

$$P(X|Y) = \frac{P(Y|X)P(X)}{\sum_X P(Y|X)P(X)}. \tag{2}$$

It is at this stage that we are ready to solve the inverse problem by estimating X from the data Y. In particular, the estimate we will seek is the posterior mean $(\sum_X XP(X|Y))$, or, in practice, an approximation thereof.

Obviously, a central role is played by the particular expectations modeled in the prior. One is tempted to design a highly "informative" prior, perhaps anticipating, for example, known anatomical shapes and their expected locations. While restorations based on problem-specific prior distributions can be effective, we have chosen instead a more conservative and universal ap-

proach by describing a rather general model (the "phi model" - see §6), which has been applied in such diverse areas as tomography [12], movie restoration [13], infrared image enhancement [8], and astronomical image restoration [24]. The phi model is minimal in the sense that it encourages only the most basic regularity properties: local smoothness and the existence of discontinuities associated with boundaries. The idea, of course, is to lessen the chances of introducing artifacts, via the prior probability distribution, into the reconstructions. As we shall see (§6), the model also has, at least approximately, a desirable invariance property with respect to rotation.

4 SPECT Model

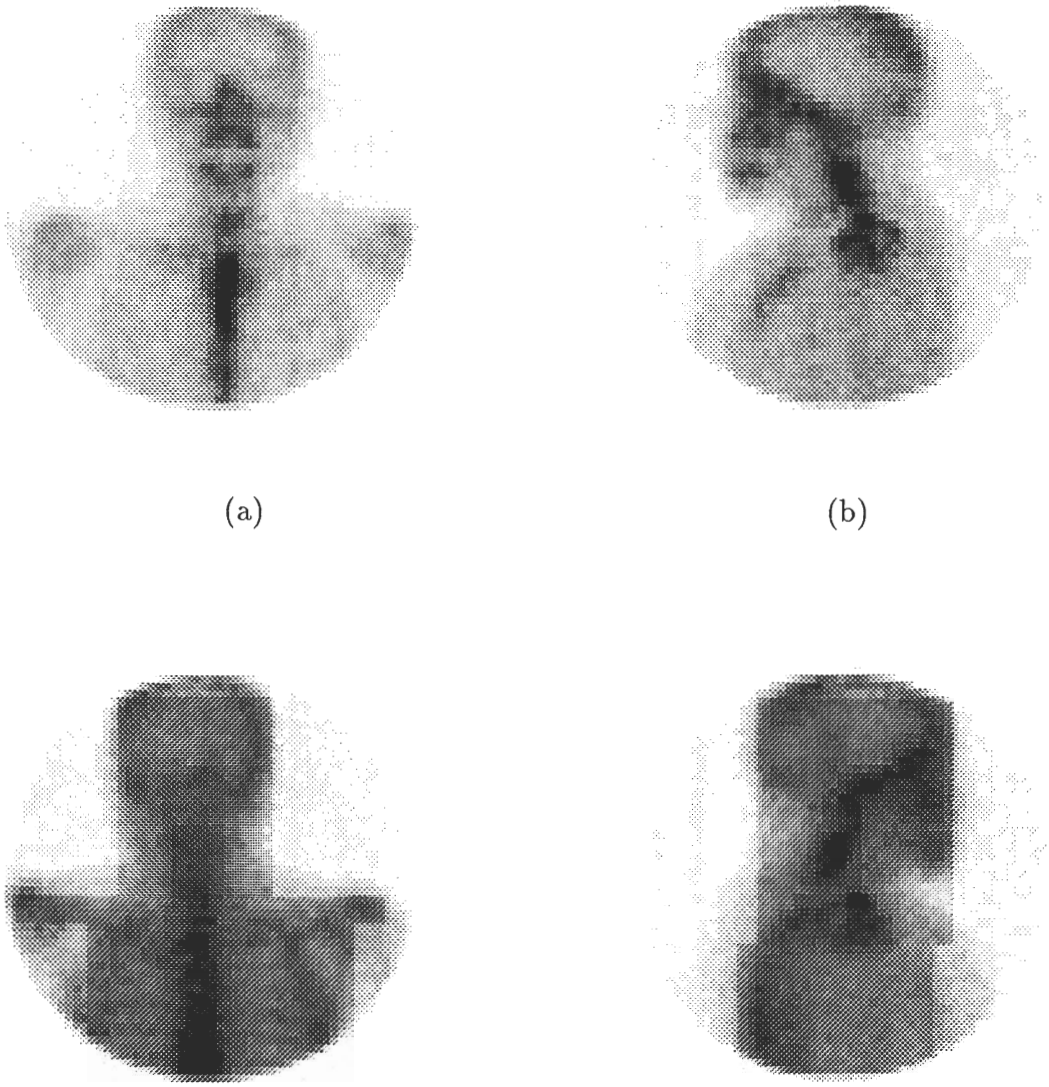
Accurately defining the modified radon transform $\mathcal{A}$, introduced in the previous section, is essential to the proper functioning of the reconstruction process. In order to describe $\mathcal{A}$, it is necessary to be more precise in the formulation of the SPECT model.

In a typical SPECT imaging session, a patient is injected with a radiopharmaceutical and lies face up on a horizontal table. The imaging session begins with the gamma camera directly above and facing the patient. In this position, the camera detects and records photons leaving the patient's body in a certain time period, often 20 seconds. The camera then rotates about an axis parallel to the table, stops after arcing $360/A$ degrees (A is often 64), and collects counts at this new angle. This process is repeated until the gamma camera returns to its original position directly above the patient, at which time the imaging session is over and data have been collected for projections through A different angles.

The gamma camera is discretized into an array of detection bins with R rows and C columns (64×64 is typical). The scintillation crystal is circular but the array of bins is rectangular, so many of the corner bins do not have any counts. See figure 2 for an example of four gamma pictures taken at different angles around a patient's head. Notice that the counts were only collected in a circular region. Also notice the lack of resolution present in the pictures; this is due to the various sources of noise inherent in SPECT (see §2 above).

Once the data have been collected, one is faced with the task of reconstructing a map of isotope intensity within the patient. To save on computations, we have followed convention in that the reconstruction of a single transverse slice, r, was accomplished by using only data from row r in each gamma picture, despite the fact that, due to scatter and collimator effects, data in other camera rows are relevant. Evidently, we have not yet made full use of the available information for reconstructions. From the ensuing discussion, it should be clear to the reader that, computational issues aside, there is no difficulty in accommodating a complete three dimensional SPECT model within the proposed framework.

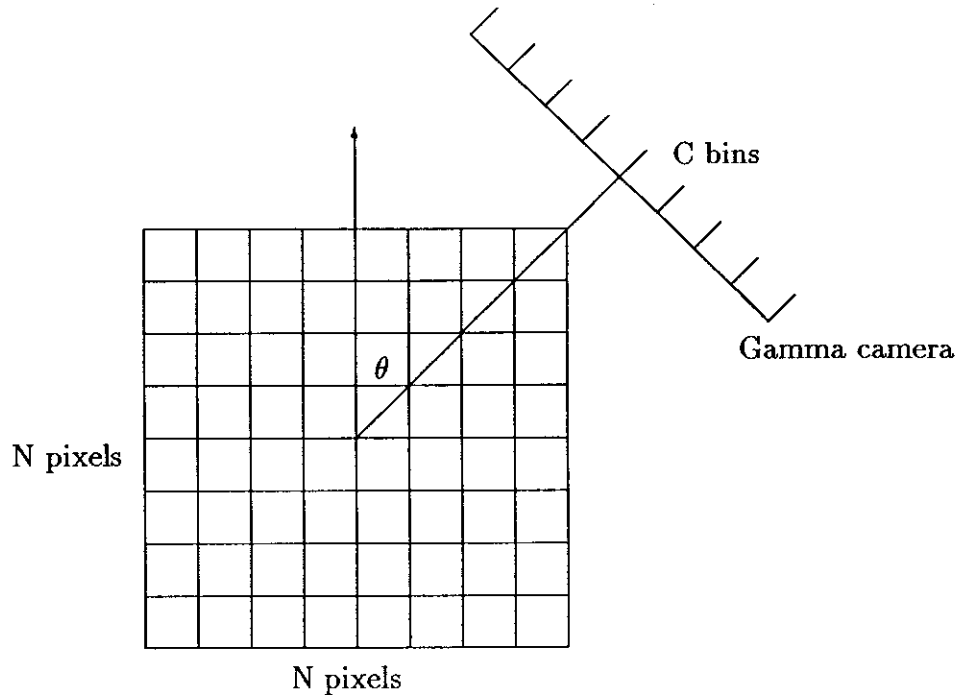
The transverse slice under consideration is discretized into an $N \times N$ array of pixels. A common value for N is 64. See figure 3 for a diagram of the camera and image in one transverse slice. In a given slice, label the image pixels, or image *sites*, S_i for $i = 1$ to N^2 and label the gamma camera bins B_j for $j = 1, C*A$. (Recall that A is the number of discrete angles through which the camera is rotated.) Let X_i be the isotope intensity at site i and let Y_j be the number of photons counted at bin B_j.



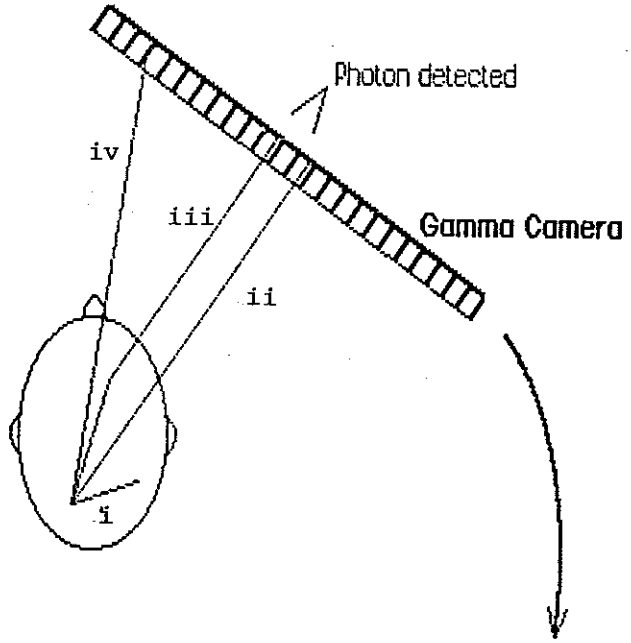
(a) (b)

(c) (d)

Figure 2: Four gamma pictures captured as the camera rotates around a patient's head. Dark areas represent bins with a large number of photon counts. The lack of resolution present in the images is typical of SPECT.



(a)



(b)

Figure 3: Diagram of a transverse slice. (a) The slice is discretized into an $N \times N$ array. In this figure the camera has rotated through angle θ. (b) A transverse slice of a patient's head and some of the physical effects modeled in this paper: i) an absorbed photon, ii) a direct count, iii) a scattered count, iv) a photon rejected by a collimator.

There is a standard mathematical/physical argument (which we will not go into) that leads to the conclusion that the observable photon counts, $Y = \{Y_j\}_{j=1}^{C*A}$, are approximately independent Poisson random variables, with means related linearly to the isotope intensity distribution, $X = \{X_i\}_{i=1}^{N^2}$, as in equation 1. It is thus the modified Radon transform (MRT), $\mathcal{A}$, that determines the relation between the unknown isotope intensities and the observable photon counts. The j, i component of the MRT, $\mathcal{A}_{ji}$, is proportional to the probability that a photon emitted at S_i is recorded at B_j. (As we shall see, the proportionality constant is somewhat arbitrary, and will be used to adjust the scale of the estimated isotope intensity values.) The details of absorption, scattering, and collimator effects are modeled through the MRT. An additional effect that could be accounted for in the MRT is the loss and redistribution of isotope during the imaging session, due to metabolism, diffusion, and radioactive decay. This effect is generally minimal, given the isotopes and pharmaceuticals actually used; we will make the assumption that the distribution and intensity of radiopharmaceutical is unchanged over the imaging session.

$\mathcal{A}_{ji}$ is modeled through a *detection density*, whose integral over the width of the j^{th} detection bin, B_j, is proportional to $\mathcal{A}_{ji}$. Referring to figure 4, the detection density, α, is modeled as a function of: 1) the (perpendicular) distance, d, from location S_i to the camera containing bin B_j; 2) the distance, D, along this same perpendicular, from location S_i to the *boundary* of the object being imaged; and 3) the distance, x, along the camera surface, measured from the perpendicular projection of S_i. Thus $\alpha = \alpha(D, d, x)$ and

$$\mathcal{A}_{ji} = k \int_{B_j} \alpha(D, d, x) dx. \tag{3}$$

The proportionality constant k is independent of i and j. It is a scale parameter, and its value depends on how we choose physical units such as the time unit and the scale for the observed projection counts Y. For convenience, we adjust k in our experiments so that the reconstructions approximately fill the dynamic range [0,255].

A set of experiments was devised in order to measure the dependence of the detection density on the three parameters, D, d, and x. In these experiments, a narrow catheter filled with a radioisotope, called a *line source*, is placed in a cylindrical tank. The line source is inclined at an angle θ relative to the imaging table - see figure 5. A gamma camera positioned directly above the line source records the photons released as the isotope decays. The observed photon counts form a *line spread function* (LSF). Using data gathered in this manner, it is possible to determine the effects of distance on attenuation, scatter, and imperfect collimation, and, from this, to deduce the dependence of α on D, d, and x. By filling the cylindrical tank with different materials, it is possible to measure these physical effects in relation to the surrounding environment.

The angled line source experiment was performed in two environments: air and water. Air was chosen as a medium for the photons to traverse because it has such a small density that a photon will essentially never be scattered or absorbed. Hence, the degradation process for a photon in air is relatively simple, and it is possible to isolate the effects of camera geometry. A photon traveling through denser media, on the other hand, will experience a significant amount of absorption and scatter. Water was chosen as an experimental environment because its attenuating properties approximate

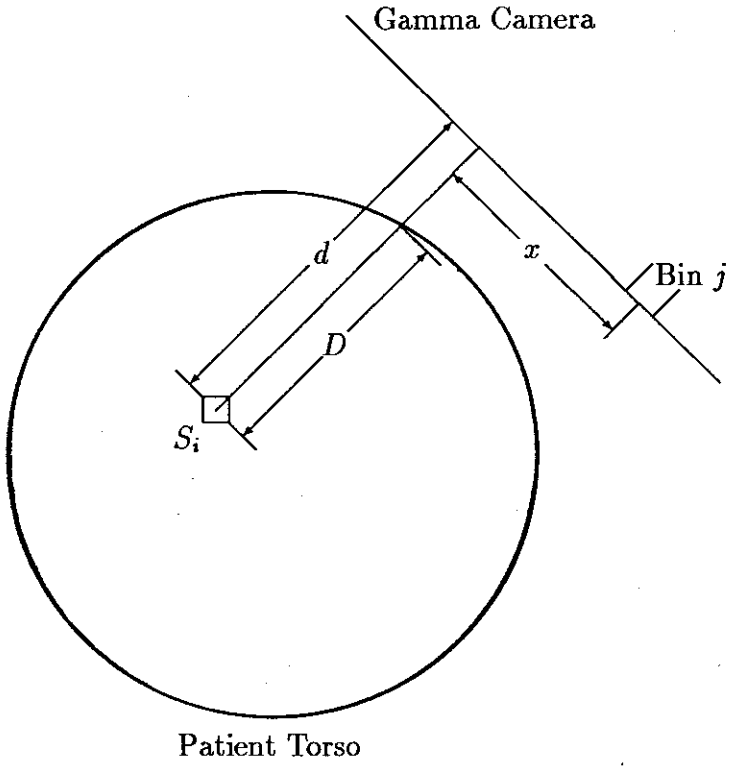
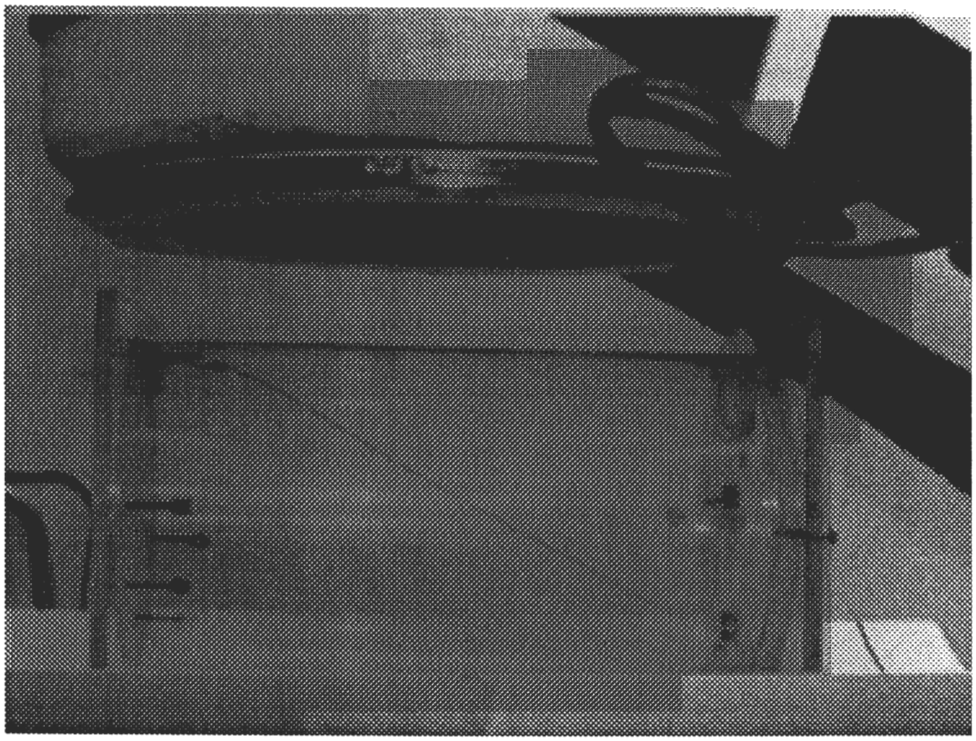
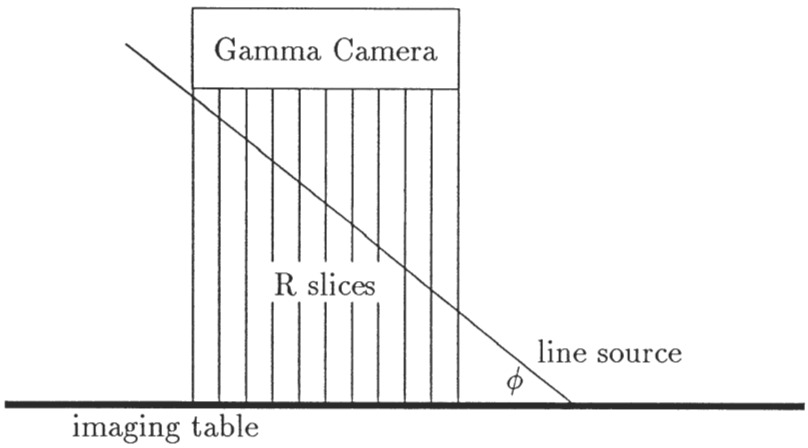


Figure 4: Detection density, due to a source at site S_i, is modeled as a function of: 1) distance d from the source to the camera; 2) distance D from the source to the boundary of the object; 3) distance x from the projection of S_i.



(a)



(b)

Figure 5: Experimental apparatus. (a) Photograph of equipment used in line source imaging experiment. The cylinder is resting on the imaging table, and the gamma camera is directly above the cylinder. (b) Diagram of a side view of the line source imaging experiment. Note that different positions on the gamma camera correspond to photons that originated at different distances from the camera.

those of most human body tissue reasonably well (see, for example, Johns and Cunningham [20]). The two major exceptions to this approximation are bone and lung tissues—the absorption and scatter coefficients for bone are higher than those of water, and the air present in lung tissue dictates that the absorption and scatter coefficients of lung are lower than those of water. In the experiments reported here (see §5), we have used the attenuating characteristics of water to approximate all body tissues. We thus assume uniform attenuation. This approximation is not necessary; the reconstruction method is unchanged by the inclusion of nonuniform attenuation. In this case, however, the MRT must be constructed to reflect the inhomogeneous attenuation function. In any case, even with uniform attenuation, there is an overall depth-dependent scatter and absorption, as we shall now see.

Results of the angled line source experiments, for air and water, are shown in figure 6. The model used here to fit these results is essentially the one presented in Penny et al. [31]. (Beck [2] proposed the same form of decomposition of the line spread function, and strong empirical evidence in support of the model was presented in Floyd et al. [7].) There are two contributions to the detection density: A *direct* contribution from detected photons that have not been scattered, and a *scatter* contribution from photons that have undergone Compton scatter prior to being detected. Thus

$$\alpha(D, d, x) = \alpha_{\text{direct}}(D, d, x) + \alpha_{\text{scatter}}(D, d, x). \tag{4}$$

It is convenient to factor each of the three detection densities into the product of a depth-dependent attenuation term and a residual line spread function, where, by definition, the former depends only on D and d, and the latter

integrates to one with respect to x:

$$\begin{aligned}
\alpha(D,d,x) &= A(D,d)\zeta(x;D,d) \\
\alpha_{\text{direct}}(D,d,x) &= A_{\text{direct}}(D,d)\zeta_{\text{direct}}(x;D,d) \\
\alpha_{\text{scatter}}(D,d,x) &= A_{\text{scatter}}(D,d)\zeta_{\text{scatter}}(x;D,d) \\
\int_x \zeta(x;D,d) &= \int_x \zeta_{\text{direct}}(x;D,d) = \int_x \zeta_{\text{scatter}}(x;D,d) = 1.
\end{aligned}$$

Modeling the MRT amounts to estimating the four functions A_{direct}, A_{scatter}, ζ_{direct}, and ζ_{scatter} from the experimental data.

The contribution from α_{direct} can be deduced from basic physical models for photon attenuation and from the empirical results of the line source experiment in air. Examination of the data (depicted in figure 6, panel a) indicates that there is no apparent loss in *total* counts as a function of distance to the camera; the number of counts in each row is essentially the same. On the other hand, the *shape* of observed counts within a row of detectors is well fit by a normal (Gaussian) probability distribution with standard error increasing linearly as a function of distance to the detector. Since essentially all counts depicted in figure 6, panel a (line source in air), are direct counts:

$$\zeta_{\text{direct}}(x;D,d) \approx \eta_{\text{direct}}(x;\sigma_{\text{direct}}(D,d)),$$

where

$$\begin{aligned}
\eta_{\text{direct}}(x;\sigma) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2} \\
\sigma_{\text{direct}}(D,d) &= \sigma_{\text{direct}}(d) = m_{\text{direct}}d + b_{\text{direct}}.
\end{aligned}$$

(a) (b)

Figure 6: Gamma pictures of (a) an angled line source in air and (b) an angled line source in water. The distance between the camera and the line sources is greatest at the top of the picture.

m_{direct} and b_{direct} were estimated, by simple regression, to be .0315 and .525 respectively.

As for the attenuation term, $A_{\text{direct}}(D, d)$, the effect is well known to be exponential in the distance traveled through the uniform attenuating medium. We have already observed that there is essentially no attenuation in air. Therefore, $A_{\text{direct}}(D, d) \approx A_{\text{direct}}(D) = \exp(-\mu D)$. (Because of the proportionality constant k in equation 3, A_{direct} is defined only up to an arbitrary multiplicative constant, which we take to be 1.) Coefficients of attenuation μ can be found in the physics and radiology literature for a wide variety of radiation sources and propagation media; see, for example, Johns and Cunningham [20]. The two sources used in our experiments were isotopes of technetium (Tc) and thallium (Tl). The coefficients of attenuation for these isotopes in water are similar, and approximately equal to that of a 150 keV photon source, for which $\mu = 0.15\text{cm}^{-1}$. This means that approximately 15% of the remaining photons are absorbed, or deflected from their original path, for each centimeter traveled. For comparison, values of μ for a 150 keV source and other propagating media are: Muscle—0.155cm^{-1}; Bone—0.246cm^{-1}; Fat—0.137cm^{-1}; Air—$0.16 \times 10^{-3}\text{cm}^{-1}$; and Lucite—$0.17\text{cm}^{-1}$.

The direct model, α_{direct}, together with the angled line source data in water (figure 6, panel b) can now be used to develop a model for the contribution from scattered photons, α_{scatter}. One must bear in mind that the counts received from the line source placed in water are the *sum* of direct and scattered contributions. Recalling again that attenuation in air is negligible, the attenuation terms A and A_{scatter} can be assumed to depend on D alone. The overall attenuation, $A = A(D)$, can be inferred from the

figure 6, panel b, data by examining the total number of counts received as a function of distance through the attenuating medium. The rows of the camera represented at the top of the figure are further from the line source, and these received fewer photons. The actual data indicates that the total number of counts closely follows an exponential decrease with distance, with the coefficient of attenuation approximately equal to $.12\text{cm}^{-1}$. Hence, $A(D)$ is proportional to $\exp(-.12D)$: $A(D) \approx c\exp(-.12D)$. Since $A_{\text{direct}}(D) = \exp(-.15D)$, c governs the ratio of direct to total counts. At $D = 0$ there is no scatter (all counts are direct), which gives the "boundary condition" $A_{\text{direct}}(0)/A(0) = 1 \Rightarrow c = 1$. We can now integrate equation 4 with respect to x to obtain A_{scatter} : $A_{\text{scatter}}(D) \approx \exp(-.12D) - \exp(-.15D)$.

The remaining term is $\zeta_{\text{scatter}}(x; D, d)$, and this too can be estimated from the data in figure 6, panel b, by first accounting for the known contribution from direct counts, via the model $\alpha_{\text{direct}}(D, d, x)$. When the estimated direct count contribution has been subtracted, and when correction has been made for the attenuation term, $A_{\text{scatter}}(D)$, the remaining counts are well fit by a double exponential whose standard error increases linearly with row number, and therefore also with distance from the line source. Since there is essentially no scatter in air, the standard error was assumed, *a priori*, to depend solely on the depth of the line source within water, i.e. only on D. Therefore:

$$\zeta_{\text{scatter}}(x; D, d) \approx \eta_{\text{scatter}}(x; \sigma_{\text{scatter}}(D, d)),$$

where

$$\begin{aligned} \eta_{\text{scatter}}(x; \sigma) &= \frac{1}{\sqrt{2\sigma^2}} e^{-\sqrt{2}|x|/\sigma} \\ \sigma_{\text{scatter}}(D, d) &= \sigma_{\text{scatter}}(D) = m_{\text{scatter}} D + b_{\text{scatter}}. \end{aligned}$$

The regression estimates for m_{scatter} and b_{scatter} were .212 and 1.615 respectively.

In summary, the following MRT model was estimated from line source data and was used in the experiments reported in §5:

$$\mathcal{A}_{ji} = k \int_{B_j} \alpha(D, d, x) dx,$$

where

$$\begin{aligned}
\alpha(D,d,x) &= e^{-.15D} \frac{1}{\sqrt{2\pi\sigma^2_{\text{direct}}(d)}} e^{-x^2/2\sigma^2_{\text{direct}}(d)} \\
&\quad + \left(e^{-.12D} - e^{-.15D}\right) \frac{1}{\sqrt{2\sigma^2_{\text{scatter}}(D)}} e^{-\sqrt{2}|x|/\sigma_{\text{scatter}}(D)} \\
\sigma_{\text{direct}}(d) &= .0315d + .525 \\
\sigma_{\text{scatter}}(D) &= .212D + 1.615.
\end{aligned}$$

Given $\mathcal{A}$, and given the isotope intensity distribution $X = \{X_i\}_{i=1}^{N^2}$, the Poisson model for Y is then

$$P(Y|X) = \prod_{j=1}^{C*A} \frac{\left(\sum_{i=1}^{N^2} \mathcal{A}_{ji} X_i\right)^{Y_j}}{Y_j!} \exp\left(-\sum_{i=1}^{N^2} \mathcal{A}_{ji} X_i\right).$$

5 Reconstruction Experiments

5.1 Computational Algorithm

Having fixed a transverse slice of interest, we denote by $X = \{X_i\}_{i=1}^{N^2}$ the discretized isotope intensity distribution. The components, X_i, represent concentrations at the internal sites, or pixels, $\{S_i\}_{i=1}^{N^2}$. In all of our experiments

we used the more or less standard discretization $N = 64$. The reconstruction problem is to estimate these isotope concentrations, and thereby the pharmaceutical concentrations, from the observed photon counts $Y = \{Y_j\}_{j=1}^{C*4}$, where, as discussed already in §4, we have restricted ourselves to using only that row of counts, at each angle of rotation, that corresponds to the transverse slice of interest.

Our reconstructions are based on the posterior distribution (equation 2), which, given Y, is proportional to $P(Y|X)P(X)$:

$$P(X|Y) = \gamma P(Y|X)P(X).$$

The data, or likelihood, term $P(Y|X)$ was developed in §4. The prior term $P(X)$ will be laid out, and discussed in detail, in the ensuing section §6. The actual reconstructions are an approximation of the posterior mean, $E[X|Y] = \sum_X XP(X|Y)$, derived by an iterative algorithm known as ICE (for "Iterated Conditional Expectations," see Owen [30]). The posterior mean itself is intractable, since X has $64 \times 64 = 4096$ components, each of which could attain, in our experiments, any of 32 (equally spaced) values.

The ICE algorithm begins with some initial estimate of isotope concentration X^0. We used $X^0 \equiv 0$. A domain of interest is defined (an ellipse known to contain the patient's body or the phantom under study), and X^1 is then computed from X^0 by sequentially modifying each component of X^0 whose corresponding site is within the domain of interest. We followed a "raster scan" ordering of these components, although there is reason to believe that there are better "site visitation" schedules (see Amit and Grenander [1]). Upon visiting a site, S_i, the value X_i^0 is replaced by its conditional mean, given Y and given the current values of the components associated with the

remaining sites S_k, $k \neq i$. The conditional mean is easy to compute from the posterior distribution, despite the fact that the posterior distribution involves an unknown multiplicative (normalizing) constant. We refer the reader to Owen [30] and Manbeck [23] for details. The process is continued, recursively, producing a sequence of images $X^0, X^1, ...$, until there are essentially no further changes. This generally requires about ten or fifteen iterations.

5.2 The Backprojection Algorithm

Most clinical reconstructions are computed by some variant of the backprojection algorithm (cf. [19]). In backprojection, the reconstructed isotope concentration at a given pixel is taken to be proportional to the sum of observed counts over "relevant" bins. A particular bin is relevant to a given pixel if a perpendicular line from the pixel to the camera face falls into the bin. In short, photon counts are back-projected to form the reconstructed image. Attenuation can be approximately corrected for by a depth-dependent weighting of the back-projected counts.

An important modification, that can partially accommodate the effects of scatter and collimator geometry, and, to a degree, mitigate the effects of (Poisson) statistical fluctuations, is the *filtered* backprojection. A suitable filter is first applied to the raw data (observed photon counts), and the result is then back-projected to form the reconstructed image. In general terms, scatter and collimator effects contribute to blurring in the reconstructed image, whereas statistical variation in observed photon counts is manifested in the reconstruction by high frequency noise. It is difficult to design filters that *simultaneously* address both of these undesirable effects; noise removal

tends to blur boundaries, whereas deblurring tends to accentuate noise. We believe that an important advantage of the statistical approach advocated here is its ability to smooth the reconstructed image, while preserving sharp boundaries and, consequently, maintaining good resolution.

5.3 Results

The results of three experiments are presented. Two are with real patient data—a liver scan and a head scan, and one is with a phantom of known structure. The phantom and liver studies used a technetium isotope, and the head scan used a thallium isotope. In each experiment, a reconstruction by filtered backprojection is presented for comparison. The backprojection algorithm used was the one provided with the imaging machine, and, because it is proprietary, we do not know details about the method of attenuation correction or the particular filter used. We should point out that research continues on filtered backprojection for SPECT and other imaging modalities. It is likely that the method can produce better reconstructions than those that we obtained from the package provided with this particular machine.

In addition to the two reconstructions (by filtered backprojection and by Bayesian estimation via ICE), we present for each experiment a gamma camera picture from a single angle, and the *sinogram* associated with the particular transverse slice being reconstructed. The gamma picture simply depicts the 64×64 array of recorded counts at one of the 64 angles of rotation. The actual data used in reconstructing a transverse slice is depicted in the sinogram. The first row of the sinogram is a row of data from the gamma

picture at the first angle of rotation. The second row of the sinogram is from the second angle of rotation, and so on. The data comes from the same level in each gamma picture—corresponding to the particular transverse slice being reconstructed. Notice that the first and last rows of the sinogram are similar, since they come from gamma camera pictures at neighboring angles.

5.3.1 Two Lines in Water

A test for the resolution of the reconstruction was devised. Two catheters were filled with isotope and placed parallel and in close proximity to each other within a tank of water. The tank was situated within the SPECT machine so that the catheters were parallel to the axis of rotation of the gamma camera. Ideally, a transverse slice would contain just two small regions of high intensity. Under the (64×64) digitization used, a single pixel in the image array was larger than the bore of the catheters. Ideally, there would be only one pixel in each intensity peak.

Figure 7 shows the result of one such experiment in which the catheters were separated by 25 mm. Severe degradation from collimator effect, scatter, and Poisson noise is evident in the gamma picture (panel a) and the sinogram (panel b). The filtered backprojection and Bayesian reconstructions are displayed in panels c and d, respectively.[1] Obviously, there is a considerable amount of artifact in the filtered backprojection reconstruction.

[1]Even in the absence of an isotope source, there is a nonzero level of counts recorded by the gamma camera. This is seen as a more-or-less uniform background intensity in figure 7a. These counts produce a systematic artifact in our reconstructions: the boundary of the elliptical domain of reconstruction invariably contains an inappropriately high level of intensity, reflecting the best possible choice for source locations, given the counts in the more extreme columns of the camera. In all of our experiments, this artifact was removed from the reconstructions by setting the intensity values to zero along the bounding ellipse.

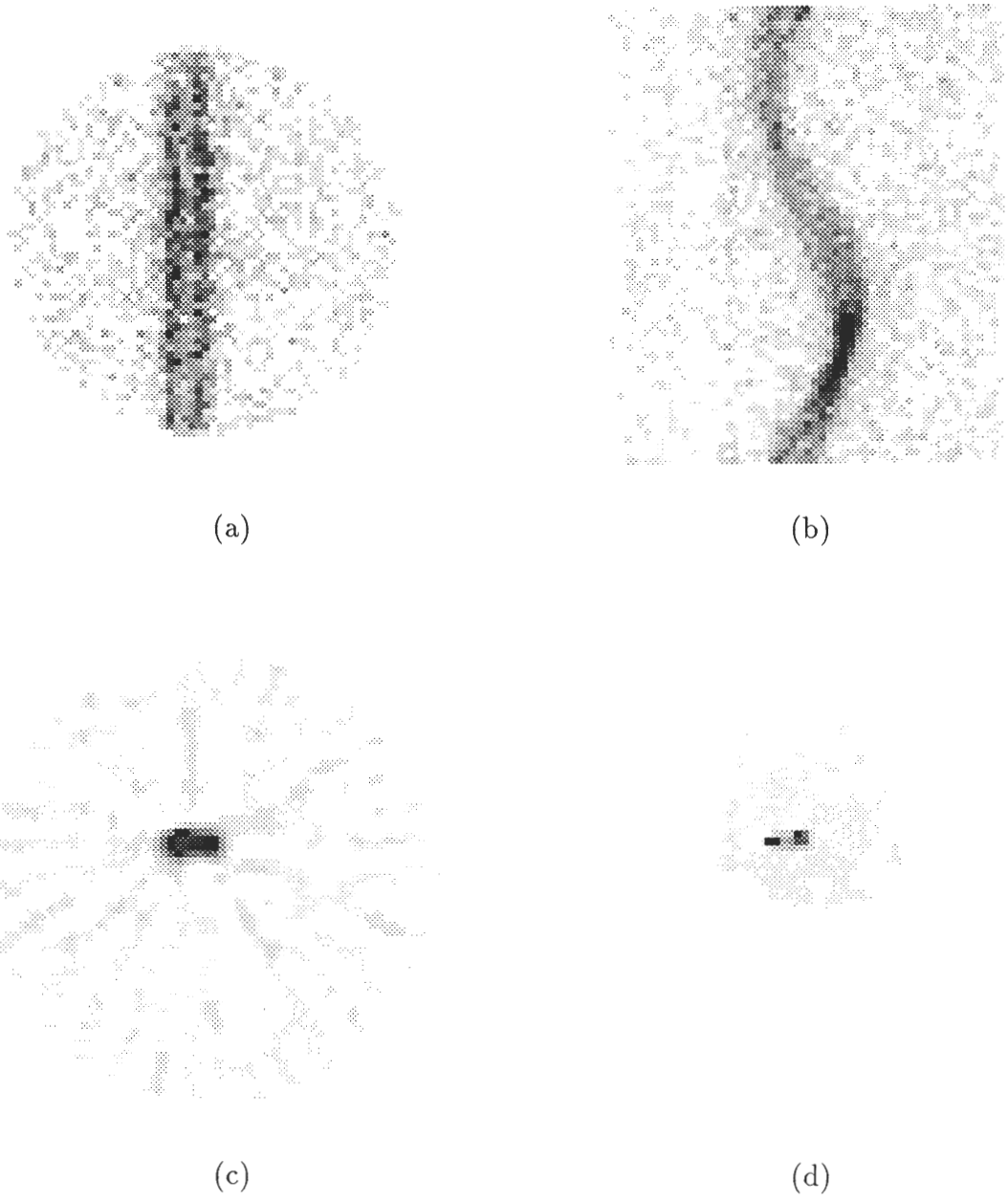


Figure 7: Transverse reconstruction of two line sources in water. (a) Gamma picture. (b) Sinogram. (c) Reconstruction by filtered backprojection. (d) Bayesian reconstruction using ICE.

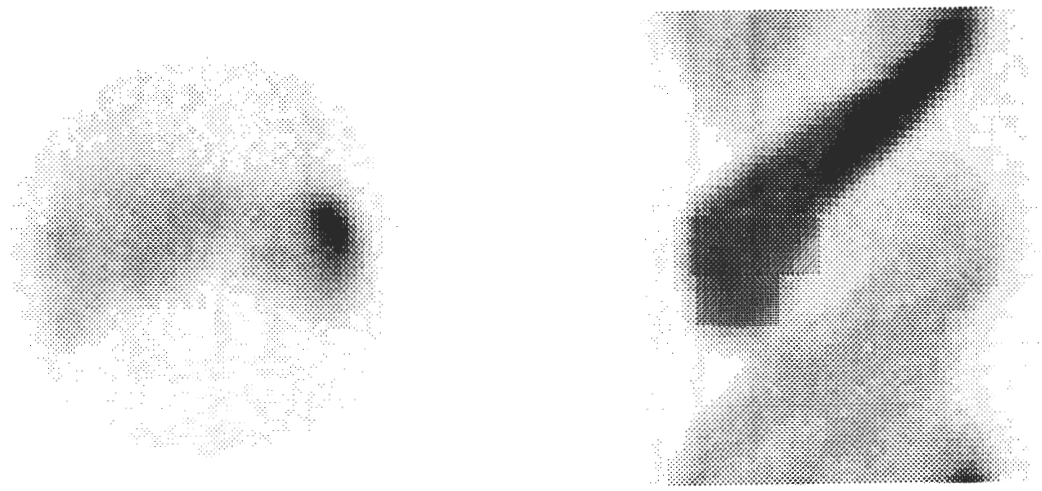
5.3.2 Liver Scan

The results are shown in figure 8. The actual distribution of isotope intensity is, of course, unknown. The larger and lighter structure is the liver, the smaller and darker structure is the spleen. Poor uptake of radiopharmaceutical by the liver indicates disease. The intense ring of radiopharmaceutical seen on the spleen in the filtered backprojection reconstruction (panel c) is probably artifact.

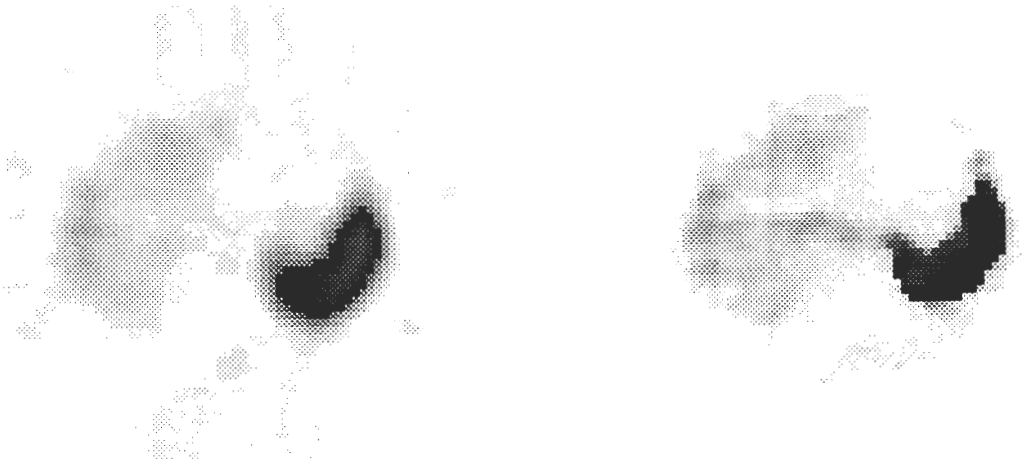
5.3.3 Head Scan

The particular pharmaceutical used is retained mostly by bone, which was the object of interest in this scan. Reconstructions are shown in figure 9. Again, there is no "true image" to go by, and in fact the two reconstructions are rather different. Given the apparent tendency of this filtered backprojection algorithm to blur the isotope distribution (see figure 7), we are inclined to believe that the true distribution is less homogeneous than indicated in figure 9c, and perhaps more along the lines of the reconstruction in figure 9d. In any case, this experiment points out the importance of using phantoms, where structure is known *a priori*, in comparing and assessing reconstruction strategies.

In an experiment such as this one, involving both bone and soft tissue, it would be highly desirable to use a binary-valued attenuation map, thereby taking full advantage of the known gross anatomy of the two tissue types. The distribution of bone and soft tissue can be conveniently encoded into the MRT ($\mathcal{A}$), and there is little additional computational cost for using the resulting inhomogeneous attenuation.



(a) (b)



(c) (d)

Figure 8: Liver scan. (a) Gamma picture. (b) Sinogram. (c) Reconstruction by filtered backprojection. (d) Bayesian reconstruction using ICE. The darker region is the spleen, the lighter and larger region is the liver. The low concentration of radiopharmaceutical in the liver indicates disease.

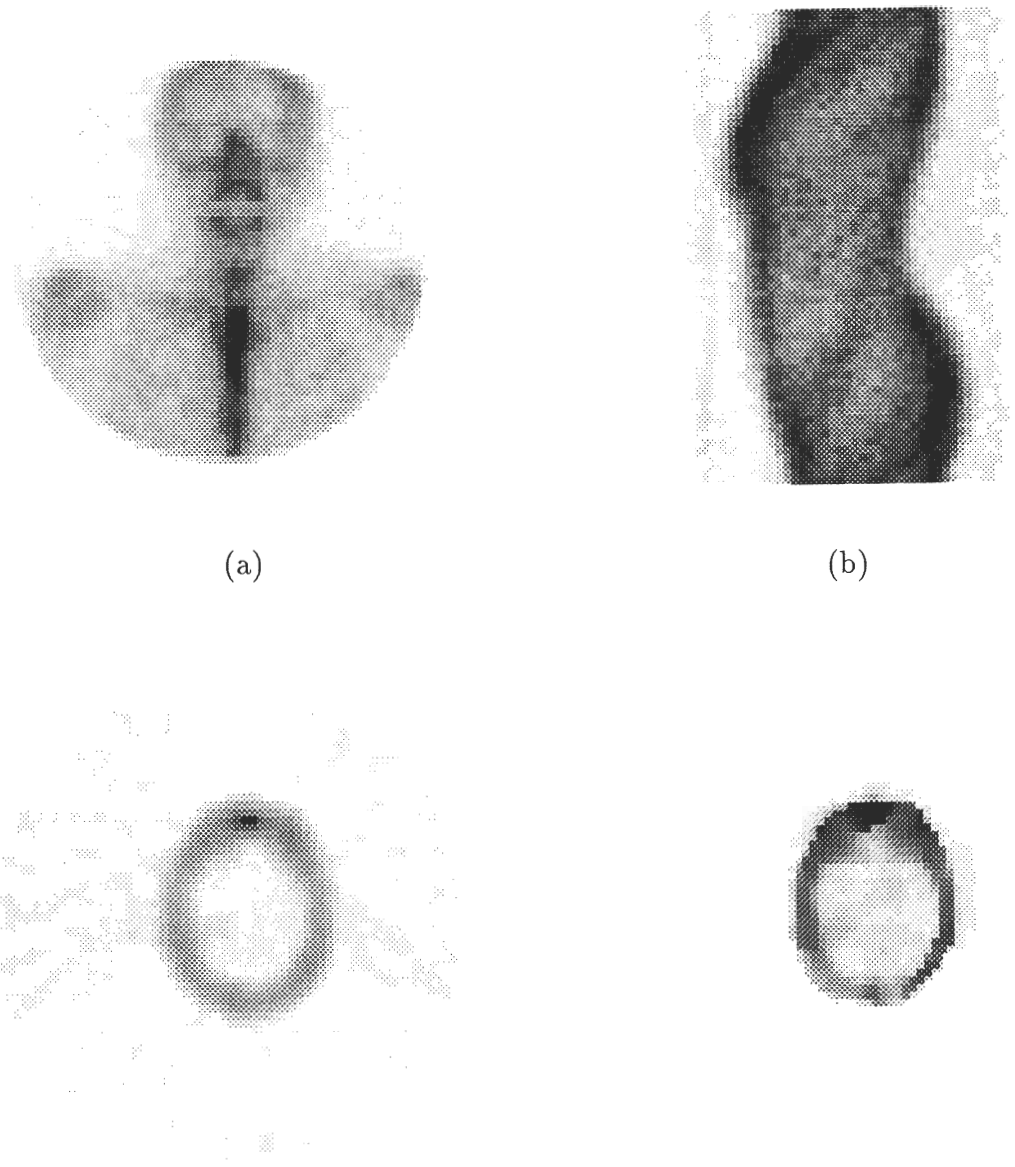


Figure 9: Bone scan of the head. (a) Gamma picture. (b) Sinogram. (c) Reconstruction by filtered backprojection. (d) Bayesian reconstruction using ICE.

6 A Smoothing Prior and its Isotropic Properties

One can adopt two points of view towards the role of the prior probability distribution $P(X)$ on possible reconstructions X. On one hand, we can regard $P(X)$ from a decidedly Bayesian perspective as an *a priori* assignment of probabilities to possible states of nature for X. This perspective is useful when we design P, that is, when we prescribe its exact functional form. On the other hand, we can simply regard $P(X)$ as a methodological tool introduced for the purpose of regularizing the otherwise highly variable reconstructions obtained, for example, by the method of maximum likelihood. From this perspective, $P(X)$ is analogous to a *penalty function* introduced as part of the method of penalized maximum likelihood. The two viewpoints complement each other and both are worthwhile to keep in mind in the implementation of the Bayesian approach.

The prior distribution quantifies likelihoods of characteristics of the images X. We adopt a rather general approach which seeks to quantify likelihoods of local characteristics of X, rather than to model highly detailed structural information, such as the likely shapes or locations of structures in X. In particular, we design $P(X)$ to embody two local regularity properties of radiopharmaceutical concentrations:

1. Isotope concentrations tend to be fairly constant within small regions of common tissue type and common metabolic activity. Neighboring pixels are more likely than not to have similar isotope concentrations.

2. Still, sharp boundaries will occur. For example, at the interface be-

tween two tissue types, or between two organs, there may be a sharp gradient in the intensity values of neighboring pixels.

Local characteristics such as these are conveniently modeled in terms of the Gibbs representation of P. Let

$$P(X) = \frac{1}{Z} e^{-U(X)},$$

where U is an *energy function* and Z is a normalizing constant; X will have discrete finite range and $Z = \sum_X \exp\{-U(X)\}$. Under P, the likely states are the low energy states. We therefore construct U so that low energy states are consistent with our expectations about isotope concentrations X. One could, equivalently, work directly with P, but this turns out to be much more difficult.

The two characteristics we wish to quantify suggest defining U in terms of differences $X_s - X_t$ between neighboring pixel values. Let Λ_N be the $N \times N$ square lattice of pixel sites S_i; s and t will refer to generic points in Λ_N. Then define

$$U(X) = \beta \left[\sum_{[s,t]} \phi(X_s - X_t) + c \sum_{<s,t>} \phi(X_s - X_t) \right]; \tag{5}$$

here $[s,t]$ indicates that s and t are nearest horizontal or vertical neighbors in the finite lattice Λ_N, commonly referred to as the *first-order neighbors*, and $< s,t >$ indicates nearest diagonal neighbors, commonly referred to as *second-order neighbors*. The constants β and c are positive. c determines the relative contributions of the first-order and second-order neighbors to U as a whole. We shall return to β's interpretation below. The function $\phi(\xi)$ is even and minimized at $\xi = 0$. Thus, U is minimized by images X of constant

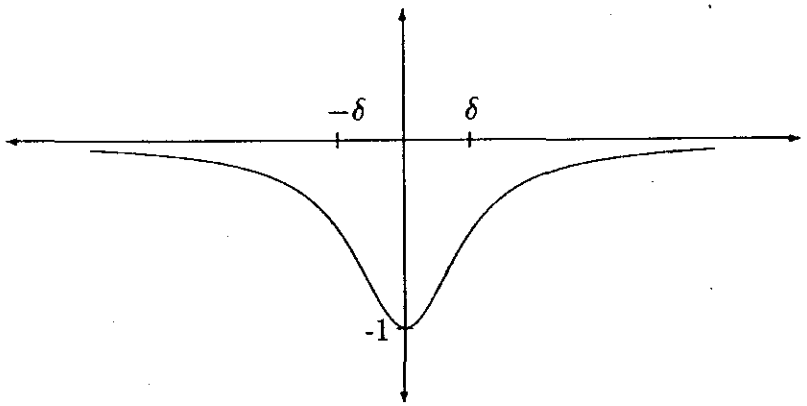


Figure 10: The phi function which is used to specify an energy function U.

intensity. Images composed of homogeneous subregions are more likely than ones that have a high degree of local variability.

This definition of U induces a graph on Λ_N in which each pixel site s is linked to its eight nearest neighbors in the square lattice. The distribution $P(X)$ then determines a Markov random field with this neighborhood structure.

The qualitative behavior of ϕ is crucial for balancing the competing demands of (i) local smoothness and (ii) permitting sharp boundaries. In a variety of applications, we have used the function ϕ defined by

$$\phi(\xi) = \frac{-1}{1 + (\xi/\delta)^2}, \tag{6}$$

where δ, like β, is a constant. ϕ is depicted in figure 10.

The shape of ϕ, especially the fact that it is bounded above, is important for accommodating the second of the two structural properties of isotope concentrations—that sharp gradients may occur across boundaries between regions of different tissue type or metabolic activity. If two pixels s and t

are on opposite sides of a boundary, then ϕ, and hence U, does not associate substantially greater penalties with larger values of the difference $X_s - X_t$. Once a boundary is introduced, the function ϕ is rather indifferent to the size of the jump across the boundary. Relatively high prior probabilities are assigned to images with constant subregions separated by sharp boundaries.

Variations of the energy model of equations 5 and 6 have been explored in the context of tomography, as well as other applications. Green [14] studies the effect of a different phi-function, $\log(\cosh(\xi/\delta))$, which is not bounded, but which is asymptotically linear and more amenable to certain optimization algorithms. We have recently reported on variations of the energy in equation 5 for which low energy (high probability) images are piecewise planar or piecewise quadratic, rather than piecewise constant [13].

The energy function U depends on three parameters: δ is easily interpreted as a scale parameter on the range of values of X_s; β controls the "strength" of the interactions between a pixel and its neighbors, and in *maximum a posteriori* (MAP) estimation it corresponds to a weighting factor balancing the contributions of a data term and a penalty term, just as in penalized maximum likelihood; as noted above, c controls the relative contributions of first-order and second-order neighbors to the total energy.

We have always found reconstructions to be relatively insensitive to the choice of δ, provided only that δ is large enough so that the shape of ϕ near the origin is not too singular. As a rule-of-thumb, we set δ to be approximately 20% of the dynamic range of X. In the experiments reported in §5, δ was set to 50.

The reconstructions are more sensitive to the choice of β, but again as

long as β is within a reasonably conservative range of values, the precise value of β is not too critical. If β is too large, the reconstructions will be oversmoothed and if β is too small, they will be undersmoothed. In [12], [27], and [23], experiments are reported that show the effect of different choices of β. The model parameter β can actually be estimated from the observed data Y using the principle of maximum likelihood and the EM algorithm; again we refer to [12], [27], and [23]. In the experiments reported in §5, for convenience, we chose the scale parameter k in equation 3 so that the reconstructions always nearly fill the dynamic range $[0, 255]$. Then β was fixed once and for all to the "moderate" value $\beta = 1$. The same values of β and δ were used for all three experiments shown in §5.

Less obvious is the effect of different choices for c. One invariance property that we would like the energy function U to possess is isotropy—invariance to rotations of the underlying coordinate system. At first it is natural to expect that c might influence isotropy of U, since c controls how interactions between sites vary with their relative orientations. Individual horizontal and vertical interactions have weight one in equation 5 and individual diagonal interactions have weight c. Further, diagonal neighbor pairs are separated by a greater distance than are horizontal/vertical neighbor pairs. Based on (i) the goal of isotropy and (ii) information about the relative orientations and separations of the two types of neighbors, a number of heuristic arguments have been espoused in support of certain specific choices of c. In [12], in fact, we suggested that the "natural" choice for c is $1/\sqrt{2}$. But as we shall show here, a quadratic approximation of U is isotropic in a continuum limit for any choice of c. The value of c does not affect rotational invariance.

For the experiments reported in §5, c was set to one, largely as a matter of computational convenience and because there is no solid analytical support for any other choice.

6.1 A Two-Dimensional Continuum Limit

The results on isotropy of U presented here are developed in Manbeck [23], and are patterned after similar arguments developed in the context of mathematical physics [35]. See also Besag [3] for a discussion of isotropic properties in the *absence* of a diagonal contribution: $c = 0$. The issue of isotropy is a natural one to question concerning the fidelity between our discrete lattice-based mathematical model on one hand and the continuous physical phenomenon it models (radiopharmaceutical concentration) on the other hand.

The plausibility of rotational invariance, regardless of c, is supported by a simple local asymptotic expansion of the two sums in U. Most neighboring pixel pairs s and t will have similar values X_s and X_t. Thus it is reasonable to approximate $\phi(X_s - X_t)$ by low-order terms of the Taylor series for ϕ. In particular,

$$\phi(X_s - X_t) \sim -1 + \frac{(X_s - X_t)^2}{\delta^2}.$$

Assuming that, in the continuous domain, horizontal neighbors have the same separation as vertical neighbors (call the spacing Δ), and imagining the values of X_s to lie on a smooth surface, then

$$\sum_{[s,t]} \phi(X_s - X_t) \sim \text{Constant} + \frac{\Delta^2}{\delta^2} \sum_{\Lambda_N} \|\nabla X_s\|^2$$

and

$$\sum_{<s,t>} \phi(X_s - X_t) \sim \text{Constant} + 2c\frac{\Delta^2}{\delta^2} \sum_{\Lambda_N} \|\nabla X_s\|^2.$$

Each of the sums is similar to an approximate integral of the squared gradient of X. The gradient expression by itself is rotationally invariant, and thus each of the sums in U is (approximately) isotropic. Of course, this is not a rigorous argument; there is actually no smooth surface which we can legitimately imagine the values of X_s to be sampled from.

A more formal argument can be developed using a Gaussian approximation of U. From the asymptotic expansion of ϕ above and assuming the differences $X_s - X_t$ are small, it follows that up to an additive constant, the function $\delta^2 U(X)/\beta$ is approximately equal to the quadratic form

$$V(X) = \sum_{[s,t]} (X_s - X_t)^2 + c \sum_{<s,t>} (X_s - X_t)^2, \tag{7}$$

for s and t in Λ_N. Let $\mathcal{X}$ denote the vector $\mathcal{X} = \{X_s\}_{s \in \Lambda_N}$ and let Q_N be the $N^2 \times N^2$ array of coefficients of the quadratic form $2 \times V(X)$;

$$V(X) = \frac{1}{2} \mathcal{X}' Q_N \mathcal{X}.$$

The factor $1/2$ simply makes it easier to relate the expression for $V(X)$ to the standard representation of a multivariate Gaussian distribution. Strictly speaking, since $s \in \Lambda_N$ is a multi-index, $\mathcal{X}$ is not a vector and Q_N is not a matrix in the traditional sense. But there is no inconsistency or harm in the simplified notation.

Q_N is analogous to the inverse of the covariance operator of a Gaussian random vector $\mathcal{X}$. However, Q_N is a singular operator since the energy functions U and V are defined in terms of first differences of X. Q_N has rank $N^2 - 1$. To remove the singularity and analyze a proper, nondegenerate Gaussian distribution, we approximate Q_N by adding a small homogeneous

perturbation. Let $\epsilon > 0$, and define

$$\begin{aligned} V_\epsilon(X) &= \sum_{[s,t]}(X_s - X_t)^2 + c\sum_{<s,t>}(X_s - X_t)^2 + \epsilon\sum_s X_s^2 \\ &= \frac{1}{2}\mathcal{X}' Q_{N,c}\mathcal{X}. \end{aligned} \tag{8}$$

The operator $Q_{N,\epsilon}$ is a finite $N^2 \times N^2$ section of an infinite Toeplitz form [17]. Let Q_ϵ denote the doubly infinite Toeplitz form. By simply expanding the squares in equation 8, it is easy to see that $(Q_\epsilon)_{ss} = 8 + 8c + 2\epsilon$, $(Q_\epsilon)_{st} = -2$ for first-order neighbors, and $(Q_\epsilon)_{st} = -2c$ for second-order neighbors. As a homogeneous, positive-definite quadratic form, Q_ϵ admits a spectral representation [17] [36]. Let $g(\xi, \eta)$ be the spectral density function

$$g(\xi, \eta) = 8 + 8c + 2\epsilon - 4\cos\xi - 4\cos\eta - 8c\cos\xi\cos\eta.$$

Then

$$(Q_\epsilon)_{st} = \frac{1}{4\pi^2}\int\int_{[-\pi,\pi]^2} g(\xi,\eta)e^{-i(s-t)\cdot(\xi,\eta)}d\xi d\eta.$$

This particular integral representation can also be verified trivially by substituting the simple expression for g into the integral.

The spectral representation for Q_ϵ is extremely useful for computing and analyzing the inverse of Q_ϵ. Note that $g(\xi, \eta)$ is strictly positive, and bounded below by $2\epsilon > 0$. The reciprocal of g is a bounded, continuous, strictly positive function on $[-\pi, \pi]^2$. Define $\mathcal{R}_\epsilon$, which we shall relate to the inverse of Q_ϵ, by

$$(\mathcal{R}_\epsilon)_{st} = \frac{1}{4\pi^2}\int\int_{[-\pi,\pi]^2} \frac{1}{g(\xi,\eta)}e^{-i(s-t)\cdot(\xi,\eta)}d\xi d\eta. \tag{9}$$

$\mathcal{R}_\epsilon$ is a doubly infinite, symmetric Toeplitz form. All of its finite sections are positive definite, since its spectral density is positive and integrable. Thus

$\mathcal{R}_\epsilon$ is uniquely identified with a zero-mean stationary random field $\tilde{X}_s$, for $s \in \Lambda_\infty$, the doubly infinite integer square lattice. $\mathcal{R}_\epsilon$ is the covariance operator of $\tilde{X}_s$;

$$\mathrm{Cov}(\tilde{X}_s, \tilde{X}_t) = (\mathcal{R}_\epsilon)_{st}.$$

The process $\tilde{X}_s$ and its covariance $\mathcal{R}_\epsilon$ have two important properties. First, when N is large, the section of $\mathcal{R}_\epsilon$ identified with the finite square lattice Λ_N approximates the covariance of our original process X_s with energy function $V(X)$. Second, in the sense of an appropriate continuum limit, $\mathcal{R}_\epsilon$ is rotationally invariant.

The actual covariance matrix of the Gaussian random vector $\mathcal{X}$ is $(Q_{N,\epsilon})^{-1}$. For large N, $(Q_{N,\epsilon})^{-1}$ is approximately equal to the finite $N^2 \times N^2$ section $(\mathcal{R}_\epsilon)_N$, that is, the finite section of $\mathcal{R}_\epsilon$ which restricts the indices s and t to Λ_N. By using the techniques of Grenander and Szegö [17], §7.4, it is straightforward to show that the so-called "trace norm" of $(Q_{N,\epsilon})^{-1} - \mathcal{R}_\epsilon$ goes to zero as $N \to \infty$. This is a form of l_2 convergence of the covariance operators:

$$\lim_{N\to\infty} \frac{1}{N^2} \sum_{s\in\Lambda_N} \sum_{t\in\Lambda_N} \left\{ \left[(Q_{N,\epsilon})^{-1} \right]_{st} - (\mathcal{R}_\epsilon)_{st} \right\}^2 = 0.$$

The methods of Grenander and Szegö are used in [26], for example, to compute approximate inverses of covariance operators for random fields indexed on a two-dimensional lattice, exactly the type of approximation needed here for $Q_{N,\epsilon}$.

In this sense, the covariance structure of $\mathcal{X}$ is approximated by $\mathcal{R}_\epsilon$.

We have used the spectral representation of $\mathcal{R}_\epsilon$ to compute values of $(\mathcal{R}_\epsilon)_{st}$ for various choices of the parameters ϵ and c, in order to examine the

of the ray from the origin to the point (h, v). The numerical evidence of isotropy is compelling. When $\epsilon = 0.01$ and $c = 2$, a comparison of two points five units from the origin shows $r(0,5) = 0.026 = r(3,4)$. Also, the values of r at points on the horizontal axis were compared to values of r on the 45° diagonal. First the values of r on the integer lattice were interpolated to form an extension of r that could be evaluated for noninteger arguments. Simple linear interpolation was done on the diagonal, so a value $r(h/\sqrt{2}, h/\sqrt{2})$ at distance h from the origin is fit by the linear interpolant between the values of r at the two integer-lattice points on the diagonal nearest to $(h/\sqrt{2}, h/\sqrt{2})$. Then the ratios $\rho(h) = r(h,0)/r(h/\sqrt{2}, h/\sqrt{2})$ were calculated for $h = 2, 3, \ldots, 20$. When $\epsilon = 0.01$ and $c = 2$, the ratios decreased from $\rho(2) = 1.06$ to $\rho(20) = 1.001$. Similar results were obtained when $\epsilon = 0.01$ and $c = \sqrt{2}$, $c = 1$, and $c = 0$. For further numeric evidence of isotropy in the latter $(c = 0)$ case, see Besag [3].

The continuum limit will involve shrinking the lattice spacing, or, equivalently, examining $r(h, v)$ at progressively larger distances $\sqrt{h^2 + v^2}$. The numerical results, however, strongly indicate that the distribution is already nearly isotropic well before the continuum is approached.

To prove a limiting form of rotation invariance for $\mathcal{R}_\epsilon$, we need to identify lattice points with points in the continuous plane. We shall define a sequence of lattices, indexed by n, which become progressively finer discretizations of the continuum as $n \to \infty$. Assume that the lattice spacing between first-order neighbors is $\Delta = 1/n$, the same in both the horizontal and vertical directions. Further, we shall allow the parameter ϵ to depend on n $(\epsilon \to \epsilon_n)$ so that we can obtain a nondegenerate limit for the associated spectral density

functions.

Consider a fixed point (x, y) in the plane, and assume for now that nx and ny are integers. The Cartesian coordinates (x, y) correspond to the lattice coordinates (nx, ny), since the spacing of the refined lattice points is assumed to be $1/n$. The covariance of the lattice-based process $\tilde{X}_s$ between the two sites (nx, ny) and $(0, 0)$ is given by $(\mathcal{R}_{\epsilon_n})_{(0,0),(nx,ny)}$. We will show that there is a rotationally invariant (in (x, y)) limit function (as $n \to \infty$).

Rather than worry about whether nx and ny are integers, we will *define* a function on all of R^2 by extending the formula for $\mathcal{R}_\epsilon$ (see equation 9) to the continuum. Formally, $r_n(x, y) = (\mathcal{R}_{\epsilon_n})_{(0,0),(nx,ny)}$, or, explicitly, for any $(x, y) \in R$

$$r_n(x, y) = \frac{1}{4\pi^2} \int \int_{[-\pi,\pi]^2} \frac{1}{g_n(\xi, \eta)} e^{-i(nx,ny)\cdot(\xi,\eta)} d\xi d\eta, \tag{10}$$

where

$$g_n(\xi, \eta) = 8 - 4\cos\xi - 4\cos\eta + 8c - 8c\cos\xi\cos\eta + 2\epsilon_n. \tag{11}$$

The change-of-variable $n\xi \to \xi$ and $n\eta \to \eta$ in equation 10 gives

$$r_n(x, y) = \frac{1}{4\pi^2} \int \int_{[-n\pi,n\pi]^2} \frac{1}{n^2 g_n(\xi/n, \eta/n)} e^{-i(x,y)\cdot(\xi,\eta)} d\xi d\eta. \tag{12}$$

Let $F_n(\xi, \eta)$ denote the function whose Fourier transform is performed in equation 12:

$$F_n(\xi, \eta) = \chi_{[-n\pi,n\pi]^2} \cdot \frac{1}{n^2 g_n(\xi/n, \eta/n)}.$$

The limiting properties of F_n are readily seen by substituting the first two terms of the Taylor series for the cosine in equation 11. If $\epsilon_n = \epsilon/n^2$, then

$$\lim_{n\to\infty} F_n(\xi, \eta) = \frac{1}{(2 + 4c)(\xi^2 + \eta^2) + 2\epsilon}; \tag{13}$$

call this limiting function $F_\infty(\xi, \eta)$.

The pointwise convergence of F_n does not imply the pointwise convergence of r_n; this step needs to be analyzed carefully since F_n is not integrable on R^2.

We can, however, show that F_n converges to F_∞ in $L_2(R^2)$, which guarantees convergence of r_n, also in $L_2(R^2)$. We start by observing the following properties of $\{F_n\}$ and the limit F_∞.

Proposition 6.1 *Let F_∞ be the pointwise limit of the sequence F_n as defined by equation 13.*

1. *$F_\infty \in L_2(R^2)$;*
2. *There exists a function $F \in L_2(R^2)$ such that for all n, $F_n(\xi, \eta) \leq F(\xi, \eta)$;*
3. *The sequence $\{F_n\}_{n=1}^\infty$ converges to F_∞ in $L_2(R^2)$.*

Proof 1. Obvious. In fact, $F_\infty \in L_p$ for all $p > 1$.
2. The inequality $(1 - \cos x) \geq \alpha x^2$ for all $x \in [-\pi, \pi]$ and $\alpha \leq 2/\pi^2$ is proved in Simon [35]. Fix any $\alpha \leq 2/\pi^2$.

$$\begin{aligned} F_n(\xi, \eta) &= \frac{\chi_{[-n\pi, n\pi]^2}}{n^2 \left[4 - 4\cos\frac{\xi}{n} + 4 - 4\cos\frac{\eta}{n} + c(8 - 8\cos\frac{\xi}{n}\cos\frac{\eta}{n}) + 2\epsilon_n\right]} \\ &\leq \frac{1}{4n^2\left[1 - \cos\frac{\xi}{n} + 1 - \cos\frac{\eta}{n}\right] + 2\epsilon_n n^2} \\ &\leq \frac{1}{4\alpha\left[\xi^2 + \eta^2\right] + 2\epsilon}. \end{aligned}$$

The first inequality follows from dropping positive terms from the denominator on the left, and the second inequality follows from Simon's lower bound

on $1 - \cos x$. The final upper bound on F_n is in $L_2(R^2)$, as in the first part of the proposition.

3. The second part of the proposition allows us to apply the dominated convergence theorem to $|F_n(\xi,\eta) - F_\infty(\xi,\eta)|^2$ to conclude L_2-convergence of F_n to F_∞.

Since the limit F_∞ is *not* in L_1, but only in L_2, it is not a spectral density function itself. F_∞ is *not* identified with the covariance function of any continuous parameter Gaussian random field. Nonetheless, the covariance functions r_n do converge in an L_2 sense to the L_2 Fourier transform, i.e. the Plancherel transform, of F_∞.

To see this, define, for any function F in L_2,

$$\hat{F}(x,y) = \frac{1}{2\pi} \lim_{A\to\infty} \int_A^A \int_A^A F(\xi,\eta) e^{-i(x,y)\cdot(\xi,\eta)} d\xi d\eta, \tag{14}$$

where the limit is taken in L_2. See, for example, Helmberg [18]. This is the Plancherel transform, and it is unitary on L_2. From equation 12, we have

$$r_n(x,y) = \frac{1}{2\pi} \hat{F}_n(x,y).$$

Since $F_n \to F_\infty$ in L_2, and since the Plancherel transform is an isometry on L_2,

$$\lim_{n\to\infty} \hat{F}_n = \hat{F}_\infty$$

in L_2. Even though $\hat{F}_\infty(x,y)$ is not a covariance function, the limiting function is still well-defined as an element of L_2. Thus r_n has a limit in L_2 as $n \to \infty$.

Finally, we observe that the limit $\hat{F}_\infty$ is rotationally invariant.

Proposition 6.2 $\hat{F}_\infty(x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta) = \hat{F}_\infty(x,y)$ *a.e. for all* θ *in* $[0, 2\pi)$.

Proof This follows from a simple change of variable in the asymptotic analysis above. The sequence $F_n(x\cos\theta + y\sin\theta, -x\sin\theta + y\cos\theta)$ has the same pointwise and L_2 limit as $F_n(x,y)$, and hence the respective Plancherel transforms have the same L_2 limit, $\hat{F}_\infty$.

Consequently, the L_2 limit $\lim_{n\to\infty} r_n$ is isotropic, for any choice of the constant c.

6.2 Extensions

The three dimensional (3-D) analog of the model for the *a priori* distribution is of interest when analyzing three-dimensional data. In the case of tomography, this 3-D analogue is required for performing true 3-D reconstructions, as opposed to the ordinary 2-D slice reconstructions. The 3-D model adds interactions *between* slices to the prior information *within* slices embodied in equation 5, and the two-dimensional picture elements (pixels) become three-dimensional volume elements (voxels). Since the thickness of slices depends on camera geometry, the voxels will not, in general, be cubes, and there are again the issues of scaling and isotropy. The mathematical analysis of the 3-D prior follows precisely the pattern used above to deduce approximate isotropy for the 2-D phi model.

Consider a prior distribution $P(X)$ on 3-D images $X = \{X_s\}_{s\in\Lambda}$, where Λ is a finite 3-D lattice of size $N_x \times N_y \times N_z$. In analogy with equation 7, we consider Gaussian priors with a local neighborhood structure. The lattice neighbors of a site s may include any lattice point in the $3 \times 3 \times 3$ discrete cube centered at s.

When we associate points of the lattice Λ with points in continuous 3-

space, we specifically allow for different spacing in the x, y and z dimensions. In other words, in the context of tomography, we do not require that pixels within a 2-D cross-section be square nor that the thickness of a slice be constrained by values used to quantize the plane of the slice. Let a, b and c denote the Euclidean distance between neighboring lattice sites in the x, y and z directions, respectively.

The neighborhood of a site s admits a natural decomposition into seven subsets, depending on the difference $(s-t)$. Sites s and t are said to be neighbors of "type i"— denoted $<s,t>_i$—for $i = 1, 2, \ldots, 7$ according to the following scheme:

$$\begin{aligned}
i &= 1 \Longleftrightarrow (s-t) = (\pm 1, 0, 0) \\
i &= 2 \Longleftrightarrow (s-t) = (0, \pm 1, 0) \\
i &= 3 \Longleftrightarrow (s-t) = (0, 0, \pm 1) \\
i &= 4 \Longleftrightarrow (s-t) = (\pm 1, \pm 1, 0) \\
i &= 5 \Longleftrightarrow (s-t) = (\pm 1, 0, \pm 1) \\
i &= 6 \Longleftrightarrow (s-t) = (0, \pm 1, \pm 1) \\
i &= 7 \Longleftrightarrow (s-t) = (\pm 1, \pm 1, \pm 1)
\end{aligned}$$

Then, in analogy with equation 8, we consider

$$V_c(X) = \sum_{i=1}^{7} c_i \sum_{<s,t>_i} (X_s - X_t)^2 + \epsilon \sum_s X_s^2, \tag{15}$$

with $s, t \in \Lambda$. Our goal is to specify values for c_i, $i = 1, \ldots 7$, so that the associated distribution possesses rotational invariance in a suitable continuum limit. The quadratic form in equation 15 has a spectral representation, identified with a strictly positive, continuous spectral density function g on

$[-\pi, \pi]^3$. The reciprocal of g is, accordingly, the spectral density of a well-defined Gaussian process $\tilde{X}_s$, for s in the infinite lattice Z^3. Just as in the 2-D case, finite sections of the covariance operator $\mathcal{R}_\epsilon$ of $\tilde{X}_s$ approximate the covariance operator of the original X_s-process in the trace norm, as $\Lambda \to Z^3$.

To carry out the asymptotic analysis of rotation invariance of $\mathcal{R}_\epsilon$, we assume that the lattice spacing becomes successively finer in the continuum. In particular, we consider a sequence of lattices, indexed by n, with spacings a/n, b/n and c/n in the x, y and z directions, respectively. To obtain a nondegenerate limit for the 3-D analog of equation 10, it is necessary to scale the coefficients in equation 15 to depend on n. Let

$$c_i = \frac{C_i}{n}$$

and

$$\epsilon = \frac{\mathcal{E}}{n^3}$$

for fixed constant values of C_i and $\mathcal{E}$. Here, we require that $\mathcal{E} > 0$ and that the C_i are nonnegative. The asymptotic analysis then proceeds step-by-step as in the 2-D case. Only minor variations occur. For example, the 3-D analogue of F_∞ (equation 13) is in L_p, for $p > 3/2$ (rather than for $p > 1$, as in the 2-D case).

The analysis yields conditions on the constants C_i which assure rotation invariance of the Plancherel transform of F_∞, and hence, approximate isotropy for the process with energy function V_ϵ. Simply stated, the conditions reduce to a "balance condition" on the constants:

$$\begin{aligned} & (C_1 + 2C_4 + 2C_5 + 4C_7)\frac{a}{bc} \\ = \; & (C_2 + 2C_4 + 2C_6 + 4C_7)\frac{b}{ac} \end{aligned}$$

$$= (C_3 + 2C_5 + 2C_6 + 4C_7)\frac{c}{ab}.$$

In addition, we require that the common value in this equation be positive.

The balance condition describes explicitly how the parameters of the energy function are affected by the different scales a, b and c in the three dimensions. Of course, there is a 2-D analogue of the balance condition for non-square pixels.

References

[1] Y. Amit and U. Grenander. *Comparing sweep strategies for stochastic relaxation.* Technical Report, Division of Applied Mathematics, Brown University, 1989.

[2] J.W. Beck. *Analysis of a Camera Based Single Photon Emission Computed Tomography (SPECT) System.* PhD thesis, Duke University, 1982.

[3] J. Besag. On a system of two-dimensional recurrence equations. *J. Royal Statist. Soc., Series B*, 43:302–309, 1981.

[4] J. Besag. On the statistical analysis of dirty pictures (with discussion). *J. Royal Statist. Soc., Series B*, 48:259–302, 1986.

[5] J. Besag. Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. Royal Stat. Soc. Ser. B*, 36:192–236, 1974.

[6] J.-M. Dinten. *Tomographic reconstruction with a limited number of projections: regularization using a Markov model.* Technical Report, Laboratoire de Statistique Appliquée, Université Paris-Sud, Orsay, 1988.

[7] E.C. Floyd, R.J. Jaszczak, C.C. Harris, and R.E. Coleman. Energy and spatial distribution of multiple order Compton scatter in spect. *Phys. Med. Biol.*, 29:1217–1230, 1984.

[8] D. Geman and S. Geman. Bayesian image analysis. In E. Bienenstock, F. Fogelman, and G. Weisbuch, editors, *Disordered Systems and Biological Organization*, Springer–Verlag, Berlin, 1986.

[9] D. Geman, S. Geman, C. Graffigne, and P. Dong. Boundary detection by constrained optimization. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12:609–628, 1990.

[10] D. Geman and G. Reynolds. *Constrained restoration and the recovery of discontinuities.* Technical Report, Department of Mathematics and Statistics, University of Massachusetts at Amherst, 1990.

[11] S. Geman and D.E. McClure. Bayesian image analysis: an application to single photon emission tomography. In *1985 Proceedings of the Statistical Computing Section, American Statistical Association*, pages 12–18, 1985.

[12] S. Geman and D.E. McClure. Statistical methods for tomographic image reconstruction. *Bulletin of the International Statistical Institute*, 52:5–21, 1987.

[13] S. Geman, D.E. McClure, and D. Geman. *A nonlinear filter for film restoration and other problems in image processing.* Technical Report, Division of Applied Mathematics, Brown University, 1990.

[14] P.J. Green. Bayesian reconstruction from emission tomography data using a modified EM algorithm. *IEEE Trans. Med. Imaging*, 9:84–93, 1990.

[15] U. Grenander. *Abstract Inference.* John Wiley & Sons, New York, 1978.

[16] U. Grenander. *Tutorial in Pattern Theory.* Technical Report, Division of Applied Mathematics, Brown University, 1983.

[17] U. Grenander and G. Szego. *Toeplitz Forms and Their Applications.* Chelsea Publishing Company, New York, second edition, 1984.

[18] Gilbert Helmberg. *Introduction to Spectral Theory in Hilbert Space.* North-Holland Publishing Company, Amsterdam, 1969.

[19] G.T. Herman. *Image Reconstruction from Projections, the Fundamentals of Computerized Tomography. Computer science and applied mathematics*, Academic Press, New York, 1980.

[20] H.E. Johns and J.R. Cunningham. *The Physics of Radiology.* Charles C. Thomas, Springfield, IL, fourth edition, 1983.

[21] V.E. Johnson, W.H. Wong, X. Hu, and C.-T. Chen. *Statistical aspects of image restoration.* Technical Report, Department of Statistics, University of Chicago, 1988.

[22] E. Levitan and G.T. Herman. A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography. *IEEE Trans. on Medical Imaging*, 6:185–192, 1987.

[23] K.M. Manbeck. *Bayesian Statistical Methods Applied to Physical Phantom and Patient Data.* PhD thesis, Division of Applied Mathematics, Brown University, 1990.

[24] K.M. Manbeck. *Hubble telescope image restoration by statistical methods.* Technical Report, Division of Applied Mathematics, Brown University, 1990.

[25] J.L. Marroquin, S. Mitter, and T. Poggio. Probabilistic solution of ill-posed problems in computational vision. *J. Amer. Stat. Assoc.*, 82:76–89, 1987.

[26] D.E. McClure and S.C. Shwartz. *A method of image representation based on bivariate splines.* Technical Report, Division of Applied Mathematics, Brown University, 1989.

[27] J.A. Mertus. *Self Calibrating Methods for Image Reconstruction in Emission Computed Tomography.* PhD thesis, Division of Applied Mathematics, Brown University, 1988.

[28] M.I. Miller, D.L. Snyder, and T.R. Miller. Maximum-likelihood reconstruction for single-photon emission tomography. *IEEE Trans. on Nuclear Science*, 32:769–778, 1985.

[29] D.W. Murray and B.B. Buxton. Scene segmentation from visual motion using global optimization. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 9:220–228, 1987.

[30] A.B. Owen. Discussion of: statistics, images, and pattern recognition, by B.D. Ripley. *Can. J. Statist.*, 14:106–110, 1986.

[31] B.C. Penny, M.A. King, and K. Knesaurek. A projector, back-projector pair which accounts for the two-dimensional depth and distance dependent blurring in SPECT. *IEEE Trans. Nuclear Science*, 37:681–686, 1990.

[32] A. Rockmore and A. Makovski. A maximum likelihood approach to emission image reconstruction from projections. *IEEE Trans. on Nuclear Science*, 23:1428–1432, 1976.

[33] L.A. Shepp and Y. Vardi. Maximum-likelihood reconstruction for emission tomography. *IEEE Trans. on Medical Imaging*, 1:113–121, 1982.

[34] B.W. Silverman, M.C. Jones, J.D. Wilson, and D.W. Nychka. A smoothed EM approach to indirect estimation problems, with particular reference to stereology and emission tomography. *J. Royal Stat. Soc. Ser. B*, 52:271–324, 1990.

[35] B. Simon. *The $P(\phi)_2$ Euclidean (Quantum) Field Theory.* Princeton University Press, Princeton, NJ, 1974.

[36] A.M. Yaglom. *An Introduction to the Theory of Stationary Random Functions.* Prentice-Hall, Englewood Cliffs, NJ, 1962.