THE USE OF SIEVES TO STABILIZE IMAGES PRODUCED WITH THE EM ALGORITHM FOR EMISSION TOMOGRAPHY†

Donald L. Snyder* and Michael I. Miller*

ABSTRACT

Images produced in emission tomography with the expectation-maximization (EM) algorithm have been observed to become more 'noisy' as the algorithm converges towards the maximum-likelihood estimate. We argue in this paper that there is an instability which is fundamental to maximum-likelihood estimation as it is usually applied and, therefore, is not a result of using the EM algorithm, which is but one numerical implementation for producing maximum-likelihood estimates. We show how Grenader's method of sieves can be used with the EM algorithm to remove the instability and thereby decrease the 'noise' artifact introduced into the images with little or no increase in computational complexity.

1. INTRODUCTION

A fundamental problem in statistics is the estimation of probability density functions from a finite sample of statistical measurements. We are interested in generating estimates of radioactivity concentrations derived from positron emission tomographs (PET) and single photon tomographs (SPECT) from which information is inferred about physiologic and metabolic function. These concentrations have all of the same properties as probability densities, except they are not normalized to unity. For PET and SPECT imaging,[1-5] compounds containing radioactive isotopes are introduced into the body which result in measurements that are well modeled as a Poisson process with an intensity which we denote by $\lambda(x)$.[6-8] The goal of medical imaging is to estimate, from a set of statistical measurements, the emitter density $\lambda(x)$. We have argued that the measurements $N(du)$ for both PET and SPECT are Poisson distributed[6,8,9] with an intensity function $\Theta(u)$ given by

$$\Theta(u) = \int_E p(u,x)\lambda(x)\,dx, \tag{1}$$

where $p(\cdot)$ is the point spread function describing the PET or SPECT imaging modality. The log-likelihood function $L(\lambda)$ describing the measurements $N(du)$ is

$$L(\lambda) = -\int_M\int_E p(u,x)\lambda(x)\,dx\,du + \int_M \ln\left[\int_E p(u,x)\lambda(x)\,dx\right]N(du). \tag{2}$$

In equation (2), the integrals over the emission space E and the measurement space M as well as the data $N(du)$ are chosen appropriately for the particular imaging modality.[10]

We have used the *expectation-maximization* algorithm of Dempster et al.,[11] first introduced by Shepp and Vardi[7] to derive an algorithm for numerically producing the maximum-likelihood estimate of $\lambda(x)$. The recursive algorithm is as follows:

$$\hat{\lambda}^{k+1}(x) = \hat{\lambda}^{k}(x)\int_M \frac{p(u,x)N(du)}{\int_E p(u,z)\hat{\lambda}^{k}(z)\,dz}. \tag{3}$$

We have found that images of the radioactivity distributions produced using the EM equation of (3) for the iterative implementation of maximum-likelihood estimates of $\lambda(x)$ perform better, as measured by signal-to-noise ratio and resolution metrics, than conventional filtered-back projection and confidence weighted algorithms.[12] However, we have found that PET images based on either time-of-flight or non time-of-flight measurements as well as SPECT images appear to become 'noisy' as iterative estimates generated with the EM algorithm climb the 'likelihood hill' towards the true maximum-likelihood estimate of $\lambda(x)$. This degradation has also been observed by other investigators in positron emission tomography.[13,14] In order to demonstrate this, we have included the results from one simulated time-of-flight image. Shown in Fig. 1 is a radioactivity distribution used for this PET simulation, where the intensity levels and structures were chosen to be in the range of radioactivity levels and structure sizes that have been measured in previous PET tomography studies. We have defined this model previously.[9,15] Parameters were chosen to be consistent with those of the SUPER-PETT I time-of-flight tomograph developed at Washington University.[16] The image shown in Fig. 1 is a 128-by-128 array, obtained by quantizing the two-dimensional image intensity distribution $\lambda(x_1,x_2)$ into square pixels of equal length and width $\Delta = 0.25$cm. The time-of-flight measurements of the simulation were obtained by generating pseudo-random numbers from a Poisson distribution with a mean equal to the convolution of the radioactivity distribution shown in Fig. 1 with the measurement error density describing the SUPER-PETT I measurement system. The point-spread density used for both the generation of the measurements as well as the maximum-likelihood reconstructions was chosen to be an asymmetric Gaussian with a full width at half maximum of 1.0 cm transverse to the line-of-flight and 6.0 cm along the line-of-flight. The maximum-likelihood estimates were generated by implementing equations (2,3).

Plotted in Fig. 2 is the result of generating 50 iterations of the EM equations of (2,3) on the Poisson simulation described above, which contained approximately 100,000 measurement points. This figure illustrates the "noise" artifact that is introduced in the maximum-likelihood reconstruction. Fig. 2 shows that as the iterative algorithm yields estimates which are closer to the maximum-likelihood solution, there is a tendency for large peaks and valleys to be introduced into the image. These spurious peaks and valleys may be viewed as a 'noise' artifact. The purpose of this paper is to show that this 'noise' artifact is *fundamental* to the application of *unconstrained* maximum-likelihood estimation of density functions based on point-process data. We argue that the method of maximum-likelihood estimation as it is usually applied in medical imaging produces inherently unstable estimates of distributions of radioactivity, and that the

Manuscript Received by IEEE: May 23, 1985.

†This work was supported by the National Institutes of Health under Grants RR01379, RR01380 (both from the Division of Research Resources), and Grant HL13851 and by the National Science Foundation under Grant ECS-8215181.

*Institute for Biomedical Computing and Department of Electrical Engineering, Washington University, St. Louis, Missouri 63130.

0018-9499/85/1000-3864$01.00©1985 IEEE

noise artifact is not a result of using the EM algorithm for the attainment of the maximum-likelihood solution, but would occur with any algorithm producing maximum-likelihood estimates. In the remainder of this paper, we show how the elegant *method of sieves*, developed by Grenander,[17] and methods for maximizing the likelihood subject to a penalty function can be used with the EM algorithm to stabilize the estimation procedure thereby resulting in estimates which do not suffer from noise instability as the iterative algorithm converges towards the maximum-likelihood solution. We propose the use of a *convolution-kernel sieve* as it can be directly incorporated into the EM equations of (3) with little increase in computational complexity.

II. THE BASIC PROBLEM

For stating the fundamental problem, we assume that the measurements N(dx) are from a Poisson process with intensity $\lambda(x)$ without any of the additional errors between the underlying sample and measurement points that do occur in emission tomography. The log-likelihood of equation (2) then simplifies to

$$L(\lambda) = -\int_E \lambda(x)\,dx + \int_E \ln[\lambda(x)]\,N(dx). \tag{4}$$

Direct maximization of the log-likelihood functional for the estimation of the density function $\lambda(x)$ yields an estimate which is a set of Dirac delta-functions, centered at the points of the N observations; that is

$$\hat{\lambda}(x) = \sum_{i=1}^{N} \delta(x-x_i), \tag{5}$$

where the N data points occur at $x_1, x_2, \ldots, x_N$. This solution is obviously unacceptable because it is expected that $\lambda(x)$ is bounded, and at the very least piecewise continuous. It is obvious that a direct maximization of equation (4) fails to produce a meaningful estimator. The EM algorithm described in the previous section for the PET and SPECT problems containing measurement errors will suffer in precisely the same way. The best the maximum-likelihood algorithm can do is compensate perfectly for the measurement errors, which would produce an unstable estimate similar to that described by (5).

It is worth noting that a discretization of equation (4), which is required in practice for any digital implementation, does not help matters. Discretizing into pixels of width Δ results in the discrete log-likelihood

$$L_d(\Lambda) = -\sum_{i=1}^{P} \Lambda(i) + \sum_{i=1}^{P} \ln[\Lambda(i)]\,N(i) \tag{6a}$$

where

$$\Lambda(i) = \int_{i\Delta}^{(i+1)\Delta} \lambda(x)\,dx$$

and

$$N(i) = \int_{i\Delta}^{(i+1)\Delta} N(dx). \tag{6b}$$

Then, the maximum-likelihood estimate of Λ becomes a series of discrete pulses with heights proportional to the number of observations in each bin and inversely proportional to the pixel size. A direct maximization of (6a) may result in an extremely rough estimate of the radioactivity distribution Λ; estimates in pixels containing zero measurement points will be set to zero, even if sandwiched between large estimate values. This contradicts our feeling that the function Λ is 'smooth'; i.e., that Λ in adjacent pixels should be close in value. In fact, upon examination of the likelihood equation of (6a), we see that it completely fails to tell us anything about the function Λ in pixels where there are no measurement points. The discrete implementation also has the undesirable property that as the implementation becomes finer, in the sense that the pixel sizes get smaller, we find ourselves in a worse predicament, a situation termed "dimensional instability" by Tapia and Thompson.[18]

II.1 Maximization Using a Penalty Function

As has been suggested by Good and Gaskins,[19] one way to remedy the apparent difficulty is to perform the maximization of the likelihood with a roughness penalty $\Phi(\lambda)$, where we may view the penalty function as describing our prior knowledge about the smoothness of the density λ. Specifically, the approach is to maximize the function $G(\lambda)$ defined by

$$G(\lambda) = L(\lambda) - \Phi(\lambda) \tag{7}$$

where $L(\lambda)$ is the log-likelihood functional given by (4), and $\Phi(\lambda)$ is the penalty which describes our prior knowledge of the density λ. As Good and Gaskins[19] have pointed out, since λ is a positive function, equation (7) must be maximized subject to the constraint that $\lambda \geq 0$. This is equivalent to finding the real function γ, where $\gamma(x) = \sqrt{\lambda(x)}$, that maximizes the functional $G(\gamma)$ given by

$$G(\gamma^2) = L(\gamma^2) - \Phi(\gamma^2). \tag{8}$$

Finding the function γ which maximizes equation (8) is more straightforward than (7) because we have removed the difficult constraint that λ is nonnegative.

All that is left is to choose the penalty which expresses any prior knowledge on the smoothness of λ or expresses some property that the estimate of λ should have. One observation which has motivated us to select a particular penalty function is that there does not appear to be any natural way to introduce bandwidth limitations into the maximum-likelihood estimation procedure as it is presently used for reconstructing activity distributions. This is in sharp contrast to the bandwidth limitations that are readily introduced with the filtered back-projection and confidence-weighted algorithms of conventional and time-of-flight tomography, respectively; in these approaches, the bandwidth of the reconstruction filter is selected so as to be consistent with the sampling plan used during data collection. Such a bandwidth constraint can be introduced into the maximum-likelihood procedure by choosing a penalty function Φ which is a soft bandwidth constraint on γ; that is, choose $\gamma(x)$ which satisfies

$$\int \left| \frac{d\gamma(x)}{dx} \right|^2 dx \leq (2\pi B)^2 \int \gamma^2(x)\,dx, \tag{9a}$$

where B is related to the second central moment of the energy-spectrum of γ. Based on this constraint, the problem becomes one of finding the function $\hat{\gamma}$ which maximizes $G(\gamma^2)$ the penalized log-likelihood functional given by

$$G(\gamma^2) = -\int \gamma^2(x)\,dx + \int \ln[\gamma^2(x)]\,N(dx) - \alpha\left[\int \left| \frac{d\gamma(x)}{dx} \right|^2 dx - (2\pi B)^2 \int \gamma^2(x)\,dx\right] \quad (9b)$$

where the desired estimate is $\hat{\lambda} = \hat{\gamma}^2$. By introducing the bandwidth constraint of (9a) as a penalty function in (9b) with its associated Lagrange multiplier α, we show in Appendix 1 that the function $\hat{\gamma}$ that maximizes $G(\gamma^2)$ is an exponential spline that satisfies the nonlinear equation

$$\hat{\gamma}(x) = \sum_{i=1}^{N} \frac{h(x-x_i)}{\hat{\gamma}(x_i)}, \quad (10a)$$

where N is the total number of points observed, x_i is the location of the i*th* such point, and the function $h(x)$ is the exponential function

$$h(x) = \frac{1}{2\sqrt{\alpha\beta}} \exp\left[-\sqrt{\beta/\alpha}\,|\,x\,|\right]. \quad (10b)$$

and $\beta = 1-\alpha(2\pi B)^2$. The estimate of λ is $\hat{\lambda}(x) = \hat{\gamma}^2(x)$. Additional properties of $\hat{\gamma}$ are discussed in Appendix 1, those being that

$$\int \hat{\lambda}(x)\,dx = N,$$

the number of counts, as well as the property that the root-mean-square bandwidth of γ is a constant B for all choices of the parameter α in the range $0<\alpha(2\pi B)^2<1$. A similar result to that given by equation (10a) with somewhat different constraints has been derived for estimating probability densities by Tapia and Thompson.[18] It is also worth noting, as shown by other investigators,[17,18] if the first derivative constraint of (9a) is chosen to be less than some constant, and not a function of the energy in γ, then a polynomial spline for h occurs in equation (10b).

II.2 Maximization Using the Method of Sieves

Grenander[17] presents an alternative approach for preventing the inconsistent estimates described by equation (5) when applying maximum-likelihood estimation to probability densities. He notes that in maximum-likelihood problems similar to those described by equations (4) and (5), the parameter space over which $\hat{\lambda}$ is selected (positive measurable functions of finite area) is too large. He proposes maximizing the likelihood over a constrained subspace S_m and then relaxing the constraint by allowing the subspace to grow with the sample size. This is Grenander's *method of sieves*, which often produces consistent estimators under the condition that the sieve grows sufficiently slowly with the sample size. See Grenander,[17] Geman and Hwang,[20] Karr,[21] and Leadbetter and Wold[22] for various examples. One instance of a sieve, which is particularly pertinent to the bandwidth penalty solution, is the sequence of subsets S^b_m given by

$$S^b_m = \{\gamma : \gamma \text{ differentiable}, \int \left| \frac{d\gamma(x)}{dx} \right|^2 dx \le m \int \gamma^2(x)\,dx\} \quad (11)$$

for m=1,2,... which we shall term a 'bandwidth-penalty sieve'. For a particular value of m, S^b_m is the analogue of the bandwidth penalty of equations (7) and (9a,b). For any choice of the constant B in equation (9a), there results a γ that maximizes $G(\gamma^2)$ in the sieve S^b_m for some m. The introduction of the sieve allows us to increase the bandwidth of the allowable solutions as the sample size grows by increasing the size of the subset S^b_m. For the exponential spline of equations (10a,b) as the bandwidth B increases the Lagrange multiplier α decreases towards zero resulting in the unconstrained maximum-likelihood solution described by equations (4) and (5).

For the tomography problem stated in equations (1), (2) and (3), there is a different and more well suited sieve given by

$$S^k_m = \{\lambda : \lambda(x) = \int K_m(x,y)F(dy)\}, \quad (12)$$

where F is a nonnegative measurable function and $K_m(\cdot)$ a nonnegative kernal with 'window width' inversely proportional to m. $K_m(\cdot)$ is chosen nonnegative so that $\lambda \ge 0$ is automatically satisfied. We shall term S^k_m a 'kernel sieve'. If m grows sufficiently slowly with sample size, the maximum-likelihood estimate $\hat{\lambda} \in S^k_m$ is consistent.[20] As Geman and McClure[23] have shown, it is not in general straightforward to maximize the likelihood of equation (4) subject to λ being in the kernel sieve S^k_m of (12). In fact, they prove that for $K_m(\cdot)$ a Gaussian, homogeneous kernel

$$\frac{m}{\sqrt{2\pi}} \exp\left[-\frac{1}{2} m^2 x^2\right]$$

with standard deviation 1/m, the maximum-likelihood solution $\hat{\lambda} \in S^k_m$ is not the so-called kernel estimate of Rosenblatt[24] and Parzen,[25] which is simply a sum of weighted Gaussians centered over the data points. That is for N data points occurring at distinct points $x_1, x_2, \ldots, x_N$, then

$$\hat{\lambda}(x) = \sum_{i=1}^{N} \frac{m}{\sqrt{2\pi}} \exp\left[-\frac{1}{2} m^2 (x-x_i)^2\right]$$

is not the maximum-likelihood solution in S^k_m. As Geman and Hwang[20] state, a closed-form expression for the function $\hat{\lambda} \in S^k_m$ that maximizes the likelihood is difficult to obtain; this has led Geman and Hwang to suggest the use of kernel-estimators similar to the Rosenblatt-Parzen estimator which for different estimation problems can be shown to yield consistent estimates if the 'window-width' of the kernel is decreased slowly enough as the sample size tends to infinity.[17,20,22]

II.3 EM Algorithm Solution for the Maximum-Likelihood Estimate $\hat{\lambda} \in S^k_m$

We note that the EM algorithm is perfectly suited for maximizing the likelihood of (4) subject to the

sieve constraint of (12); in fact the EM solution which results is identical to the recursion described by equation (3). Maximizing the likelihood $L(\lambda)$ given in (4) subject to $\hat{\lambda} \in S_m^k$ is equivalent to finding the $\hat{F}$ which maximizes $L(F)$ given by

$$L(F) = -\iint K_m(x,\sigma)F(d\sigma)dx + \int \ln\left[\int K_m(x,\sigma)F(d\sigma)\right]N(dx). \tag{13a}$$

The maximum-likelihood estimate $\hat{\lambda} \in S_m^k$ is then

$$\hat{\lambda}(x) = \int K_m(x,\sigma)\hat{F}(d\sigma). \tag{13b}$$

We emphasize the similarity between equations (2) and (13a). Using the EM algorithm to maximize (13a) yields

$$\hat{F}^{k+1}(du) = \hat{F}^k(du)\int \frac{K_m(x,u)N(dx)}{\int K_m(x,\sigma)\hat{F}^k(d\sigma)}. \tag{14a}$$

The estimate $\hat{\lambda} \in S_m^k$ that maximizes the likelihood of (4) becomes

$$\hat{\lambda}(x) = \lim_{k\to\infty} \int K_m(x,u)\hat{F}^k(du). \tag{14b}$$

In particular, when

$$K_m(x,u) = \frac{m}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}m^2(x-u)^2\right]$$

the estimate $\hat{\lambda}$ that maximizes the likelihood of (4) subject to $\hat{\lambda}$ being in the 'Gaussian convolution-kernel sieve' is given by

$$F^{k+1}(du) = \hat{F}^k(du)\int \frac{\exp\left[-\frac{1}{2}m^2(x-u)^2\right]N(dx)}{\int \exp\left[-\frac{1}{2}m^2(x-\sigma)^2\right]\hat{F}^k(d\sigma)} \tag{15a}$$

where $\hat{\lambda}$ is

$$\hat{\lambda}(x) = \lim_{k\to\infty} \int \frac{m}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}m^2(x-u)^2\right]\hat{F}^k(du). \tag{15b}$$

II.4 Simulations Using Penalty and Convolution Sieves

In this section we describe the results from applying the 'bandwidth-penalty sieve' and 'Gaussian convolution-kernel sieve' to the derivation of maximum-likelihood estimates of Poisson data when there are no measurement errors. These preliminary simulations were based on two one-dimensional distributions λ, one being a smooth Gaussian shaped profile and one a rectangularly shaped profile. A Poisson-process was generated with a mean in each pixel of $\Lambda(i)$, where $\Lambda(i)$ is the integral over one pixel of the Gaussian and rectangularly shaped profiles. Fig. 3 shows the maximum-likelihood estimates of the Gaussian (top row) and rectangular profiles (bottom row) based on a Poisson simulation containing an average of 1000 counts in the 512 bin simulation. The histogram shown in Fig. 3 results from a direct maximization of the discrete likelihood of (6); the histogram is, therefore, the unconstrained maximum-likelihood estimate of the function Λ. The one-dimensional histograms of Fig. 3 demonstrate the 'dimensional instability' that the unconstrained maximum-likelihood estimates exhibit. Notice the occurrence of large variations between adjacent pixel estimates of Λ; this effect would get worse if the pixel size or the average number of counts were decreased.

Plotted in Fig's. 4 and 5 are the results of applying the bandwidth-penalty and Gaussian convolution-kernel sieves to the Poisson simulations. The estimates of Fig. 4 were obtained using the exponential spline estimate of (10a,b). Since the nonlinear equation (10a) has no obvious analytical solution, we solved it numerically using an iteration formula motivated by the EM equation (3). If the iterates are denoted by $\hat{\gamma}^1(x), \hat{\gamma}^2(x), \ldots$, then the formula we used for the numerical solution is

$$\hat{\lambda}^{k+1}(x) = \hat{\gamma}^k(x)\sum_{i=1}^{N} \frac{h(x-x_i)}{\hat{\gamma}^k(x_i)}, \quad k=0,1,2,\ldots$$

where $h(x)$ is given in (10b) and

$$\hat{\gamma}^{k+1}(x) = \sqrt{\hat{\lambda}^{k+1}(x)}.$$

For the initial iterate, $\hat{\gamma}^0(x)$, we used the square root of the histogram estimate of $\lambda(x)$. While we have no mathematical proof of the convergence of the iterates to the solution of (10a), we did observe convergence for all the simulation experiments we attempted. Each panel in Fig. 4 shows the estimates derived with a different bandwidth constraint as specified by the value B of equation (9a). For this simulation, we selected $\alpha(2\pi B)^2 = 0.5$ and $B = (BW)(0.5/\Delta x)$, where $\Delta x = 1/512$ is the interval width and BW was either 0.1 or 0.3 as indicated in Fig. 4. As demonstrated by the results of Fig. 4, as the bandwidth B of the maximum-likelihood estimate is decreased, the effect is to smooth the variations between adjacent estimates in the histogram.

The results shown in Fig. 5 demonstrate the use of the EM equation of (15a,b) for the numerical generation of the maximum-likelihood estimate $\hat{\lambda}$, an element of the Gaussian convolution-kernel sieve. The estimates shown in Fig. 5 were obtained by varying the width of the Gaussian kernel. For this simulation, we selected $1/m = \sqrt{2}(2\pi B)$, where B is chosen to be the same as the previous simulation. That is $B = (BW)(0.5/\Delta x)$ and BW has been varied from 0.1 to 0.3, as in Fig. 4.

III. GAUSSIAN CONVOLUTION-KERNEL SIEVE APPLIED TO PET TOMOGRAPH

For the time-of-flight simulation described by Fig's. 1 and 2, an asymmetric Gaussian point-spread density $p(\cdot)$ with FWHM=1.0cm transverse to the line-of-flight and FWHM=6.0cm along the line of flight was used for generating the measurements. For the reconstructions of Fig. 2 the identical point-spread function was used for the EM equations, resulting in the image of Fig. 2. That is, the EM equations of (2) were digitized and implemented with an asymmetric Gaussian point-spread function corresponding to the measurement errors in the simulation. As we have shown via one-dimensional simulations, applying an unconstrained maximum-likelihood estimation procedure such as that described by the EM equations of (3) will result in a noise artifact being introduced. This is in fact what is seen in Fig. 2. In order to prevent this from occurring, we have processed the tomography images

using a Gaussian convolution-kernel sieve.

We have adopted the following procedure. Proceeding as in Section II.3, maximizing the log-likelihood $L(\lambda)$ given in (2), which describes the tomography problem with measurement errors, subject to $\lambda \in S^k_m$ is equivalent to finding the $\hat{F}$ which maximizes $L(F)$ given by

$$L(F) = -\int_M\int_E p(u,x)\ln\int K_m(x,\sigma)F(d\sigma)\,dx\,du \qquad (16a)$$
$$+\int_M \ln\left[\int_E p(u,x)\int K_m(x,\sigma)F(d\sigma)\,dx\right]N(du).$$

The maximum-likelihood estimate $\hat{\lambda} \in S^k_m$ is then

$$\hat{\lambda}(x) = \int K_m(x,\sigma)\hat{F}(d\sigma). \qquad (16b)$$

Choosing the point-spread function $p(\cdot)$ to be the density describing the measurement errors for the tomography system, and the kernel function $K_m(\cdot)$ corresponding to the sieve, results in the following EM implementation equations for maximizing the log-likelihood of equation (16) subject to $\hat{\lambda} \in S^k_m$:

$$\hat{F}^{k+1}(d\sigma) = \hat{F}^k(d\sigma)\int_M \frac{r(u,\sigma)N(du)}{\int r(u,t)\hat{F}^k(dt)} \qquad (17a)$$

where

$$r(u,\sigma) = \int_E p(u,x)K_m(x,\sigma)\,dx.$$

The estimate $\hat{\lambda} \in S^k_m$ that maximizes the likelihood of (2) becomes

$$\hat{\lambda}(x) = \lim_{k\to\infty}\int K_m(x,\sigma)\hat{F}^k(d\sigma). \qquad (17b)$$

The form of (17a,b) is identical to the EM equations in (3). We emphasize that by choosing the kernel sieve S^k_m, we have not altered the form of the EM equations of (3) in any way. No extra computation is required for generating the maximum-likelihood estimate that is an element of the sieve. We do however, change the density used in the implementation of the EM equations. For generating the unconstrained maximum-likelihood estimate described by (3), the point-spread function was chosen equal to the density describing the measurement errors in the simulation. For the example presented in Fig. 2, the asymmetric Gaussian density used is precisely the density describing the measurement errors in the time-of-flight simulation. However, if S^k_m is chosen to be a Gaussian convolution-kernel sieve, the $K_m(\cdot)$ itself would be a Gaussian density and the core density $r(\cdot)$ used in the EM equations of (17) is made wider; for our tomography example it is given by the convolution of the point-spread function $p(\cdot)$ and the Gaussian kernel $K_m(\cdot)$ describing the sieve. This widening is easily incorporated into the unconstrained EM equations by simply increasing the standard deviation, or full width at half maximum parameter to be the square root of the sum of the squares of the standard deviations of the error density $p(\cdot)$ and the sieve kernel $K(\cdot)$. This is precisely what has been done in generating the image shown in Fig. 6. The image was reconstructed using the EM equations of (17), with the variance m of the circularly symmetric Gaussian density of FWHM=1.8cm transverse to the line-of-flight and FWHM=6.2cm along the line-of-flight. Therefore, in implementing equation (17a), the inner core density $r(\cdot)$ was generated by convolving two Gaussian densities, one corresponding to the measurement errors and the other to the kernel of the sieve. The figure shows the result for fifty iterations of the algorithm. Upon reaching iteration 50, the estimate was then convolved with the Gaussian convolution-kernel $K_m(\cdot)$ as specified by (17b), with the chosen FWHM=1.5cm. As can be seen by comparing Fig's. 2 and 6, the noise artifact has been substantially removed.

IV. CONCLUSIONS

Grenander's method of convolution sieves appears to be particularly well suited for stabilizing images produced with the expectation-maximization algorithm of Dempster, Laird and Rubin for positron-emission tomography. Much work remains to develop this observation fully. Most importantly, effort is needed to quantify the resolution, signal-to-noise ratio, and other performance measures of images produced with this method and to understand how such measures are influenced by sieve parameters. An important issue needing to be addressed is the rationale for selecting the sieve parameter as a function of the number of measured counts in the image. We are optimistic that these issues will be resolved and that improved reconstructions for positron-emission tomography will result.

APPENDIX I

In this appendix, we show that the function $\gamma(x)$ which maximizes the log-likelihood functional $G(\gamma^2)$ of (8) subject to the soft-bandwidth constraint of (9a) is given by equations (10a) and (10b). Using the generalized Kuhn-Tucker theorem,[26] this is equivalent to finding the $\hat{\gamma}$ which maximizes the penalized log-likelihood functional of (8) where the penalty $\Phi(\gamma^2)$ is given by

$$\Phi(\gamma^2) = \alpha\left[\int\left|\frac{d}{dx}(x)\right|^2 dx - (2\pi B)^2\int\gamma^2(x)\,dx\right], \qquad (A1)$$

and the Langrange multiplier α is nonnegative and $\Phi(\hat{\gamma}^2)=0$. Using the calculus of variations, we conclude that $\hat{\gamma}(x)$ is the solution to the nonlinear differential equation

$$\alpha\hat{\gamma}(x)\left[\frac{d^2\hat{\gamma}(x)}{dx^2}\right] - \beta\hat{\gamma}^2(x) = -\sum_{i=1}^{N}\delta(x-x_i), \qquad (A2)$$

where $\beta=1-\alpha(2\pi B)^2$, N is the total number of observed points and x_i is the location of the ith such point. Substituting

$$\hat{\gamma}(x) = \sum_{i=1}^{N}\frac{h(x-x_i)}{\hat{\gamma}(x_i)}$$

directly into the differential equation (A2) where h is given by

$$h(x) = \frac{1}{2\sqrt{\alpha\beta}}\exp\left[-\sqrt{\beta/\alpha}\,|\,x\,|\right],$$

shows that the exponential spline of (10a) and (10b) maximizes the log-likelihood of (8) subject to the bandwidth constraint of (9).

We next demonstrate that $\hat{\lambda}=\hat{\gamma}^2$ satisfies

$$\int \hat{\lambda}(x)\,dx = N, \tag{A3}$$

which is a desirable normalization property. For this purpose, define $\langle a(x),b(x)\rangle$ according to

$$\langle a(x),b(x)\rangle = \alpha\int \frac{da(x)}{dx}\,\frac{db(x)}{dx}\,dx + \beta\int a(x)b(x)\,dx.$$

Using integration-by-parts and (A2), we conclude that $\langle h(x-x_i)\hat{\gamma}(x)\rangle = \hat{\gamma}(x_i)$. Furthermore, this equality and (10a) yields $\langle\hat{\gamma}(x),\hat{\gamma}(x)\rangle=N$. But

$$\langle\hat{\gamma}(x),\hat{\gamma}(x)\rangle = \Phi(\hat{\gamma}^2) + \int\hat{\gamma}^2(x)\,dx,$$

which implies (A3) because $\Phi(\gamma^2) = 0$ for $\gamma=\hat{\gamma}$.

The Fourier transform of the exponential function h(x) of (10b) is given by

$$H(f) = \frac{1}{\alpha\left[(2\pi f)^2-(2\pi B)^2\right] + 1}\,. \tag{A4}$$

As the Lagrange multiplier α tends to zero, H(f) tends to 1, and therefore h(x) tends to a unit impulse $\delta(x)$, resulting in the maximum-likelihood estimate $\hat{\gamma}$ given by

$$\hat{\gamma}(x) = \sum_{i=1}^{N}\frac{\delta(x-x_i)}{\hat{\gamma}(x_i)}\,.$$

This is equivalent to the estimator of (5), which maximizes the unpenalized likelihood of (4). Alternatively, the Lagrange multiplier α can be viewed as a weight determining the contribution of the penalty function $\Phi(\gamma)$ to the maximization of (9b). Since $\alpha=0$ results in the unconstrained solution of (5), we conclude that α must be greater than zero or the bandwidth constraint of (9a) will not be satisfied. So that the kernel h is an exponential and does not oscillate, we also require $\alpha(2\pi B)^2<1$.

It is also interesting to note that although as α varies the bandwidth of h(x) varies as demonstrated by equation (A4), the root-mean-square bandwidth of $\hat{\gamma}(x)$ is a constant B for all α such that $0<\alpha(2\pi B)^2<1$. This is because the bandwidth penalty $\Phi(\hat{\gamma})$ is identically zero for all values of $\alpha>0$; this somewhat surprising result is a consequence of the nonlinear nature of the exponential spline relating $\hat{\gamma}(x)$ and h(x) in (10a).

ACKNOWLEDGEMENT

We are indebted to Professor Alan F. Karr of the Johns Hopkins University for providing us with a preprint of his most recent paper[21] along with various other references. It was the ideas that grew out of these references that resulted in this paper. We would like to thank Mr. Steve Moore and Mr. Russell Hermes of the Institute for Biomedical Computing at Washington University for implementing the simulations that resulted in the pictures shown in Figs. 1, 2, and 6. We also appreciate the efforts of Mr. John Gorman of the Institute for Biomedical Computing, who was the first at Washington University to demonstrate the instability for non-time-of-flight simulations. We are grateful for the support of Drs. Lewis J. Thomas and Michael M. Ter-Pogossian, directors of the Biomedical Computer Laboratory and the Division of Radiation Science of Washington University, respectively.

REFERENCES

1. T.F. Budinger, S.E. Derenzo, G.T. Gullberg, W.L. Greenberg, and R.H. Heusman, "Emission Computer Assisted Tomography with Single-Photon and Positron Annihilation Photon Emitters," J. Comput. Assisted Tomogr., vol. 1, pp. 131-145, 1977.

2. M.M. Ter-Pogossian, M.E. Raichle, and B.E. Sobel, "Positron Emission Tomography," Scientific Amer., vol. 243, pp. 170-181, October 1980.

3. R.J. Jaszczak, R.E. Coleman, and C.B. Lim, "SPECT: Single Photon Emission Computed Tomography," IEEE Trans. on Nuc. Sci., vol. NS-27, pp. 1137-1153, June 1980.

4. J.W. Keyes, Jr., "Instrumentation," in Computed Emission Tomography, ed. edited by P.J. Ell and B.L. Hofman, Oxford University Press, 1982.

5. D.E. Kuhl and R.Q. Edwards, "Image Separation Radioisotope Scanning," Radiology, vol. 80, pp. 853-662, 1963.

6. D.L. Snyder, L.J. Thomas, Jr., and M.M. Ter-Pogossian, "A Mathematical Model for Positron Emission Tomography Systems Having Time-of-Flight Measurements," IEEE Transactions on Nuclear Science, vol. NS-28, pp. 3575-3583, 1981.

7. L.A. Shepp and Y. Vardi, "Maximum-Likelihood Reconstruction for Emission Tomography," IEEE Transactions on Medical Imaging, vol. MI-1, pp. 113-121, 1982.

8. M.I. Miller, D.L. Snyder, and T. Miller, "Maximum Likelihood Reconstruction for Single Photon Emission Computed Tomography," IEEE Trans. on Nuclear Science, vol. NS-32, no. 1, pp. 769-778, February 1985.

9. D.L. Snyder and D.G. Politte, "Image Reconstruction from List-Mode Data in an Emission Tomography System Having Time-of-Flight Measurements," IEEE Trans. on Nuclear Science, vol. NS-30, pp. 1843-1849, 1983.

10. D.L. Snyder, "Estimating the Intensity Function of a Poisson Process When Each of Its Points is Observed After a Random Translation," Proc. 22nd Annual Allerton Conference on Communication, Control and Computing, p. 343, University of Illinois, Urbana, IL, 1984.

11. A.D. Dempster, N.M. Laird, and D.B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," J. of the Royal Statistical Society, vol. B,39, pp. 1-37, 1977.

12. D. Politte, Reconstruction Algorithms for Time-of-Flight Assisted Positron-Emission Tomographs, M.S. Thesis, Sever Institute of Technology, Washington University, St. Louis, MO, December 1983. Supervised by D.L. Snyder.

13. L.A. Shepp, Y. Vardi, J.B. Ra, S.K. Hilal, and Z.H. Cho, "Maximum-Likelihood with Real Data," IEEE Transactions on Nuclear Science, vol. NS-31, pp. 910-913, 1984.

14. R. Carson, 1985. Personal Communication

15. D.G. Politte and D.L. Snyder, "A Simulation Study of Design Choices in the Implementations of Time of Flight Reconstruction Algorithms," Proc. Workshop on Time of Flight Tomography, vol. IEEE Catalog No. 82CH1791-3, pp. 131-136, Washington University, St. Louis, MO, 1982.

16. M.M. Ter-Pogossian, D.C. Ficke, M. Yamamoto, and J.T. Hood, Sr., "Super PETT I: A Positron Emission Tomograph Utilizing Photon Time-of-Flight Information," IEEE Trans. on Medical Imaging, vol. MI-1, no. 3, pp. 179-187, 1982.

17. U. Grenander, Abstract Inference, John Wiley and Sons, New York, 1981.

18. R.A. Tapia and J.R. Thompson, Nonparametric Probability Density Estimation, John Hopkins University Press, Baltimore, MD, 1978.

19. I.J. Good and R.A. Gaskins, "Nonparametric Roughness Penalties for Probability Densities," Biometrika, vol. 58,2, pp. 255-277, 1971.

20. S. Geman and C-R Hwang, "Nonparametric Maximum Likelihood Estimation by the Method of Sieves," The Annals of Statistics, vol. 10, pp. 401-414, 1982.

21. A.F. Karr, "Maximum Likelihood Estimation of the Multiplicative Intensity Model, via Sieves," The Annals of Statistics, 1985 (in review).

22. M.R. Leadbetter and D. Wold, "On Estimation of Point Process Intensities," in CONTRIBUTIONS TO STATISTICS: Essays in Honour of Norman L. Johnson, ed. P.K. Sen, pp. 299-312, North-Holland Publishing Company, 1983.

23. S. Geman, "Sieves for Nonparametric Estimation of Densities and Regressions," Repts. in Pattern Analysis, vol. 99, D.A.M. Brown University, 1981.

24. M. Rosenblatt, "Remarks on Some Nonparametric Estimates of a Density Function," Ann. Math. Stat., vol. 27, pp. 832-837, 1956b.

25. E. Parzen, "On the Estimation of a Probability Density and Mode," Ann. Math. Stat., vol. 33, pp. 1065-1976, 1962.

26. D.G. Luenberger, Optimization by Vector Space Methods, p. 249, John Wiley and Sons, New York, 1969.

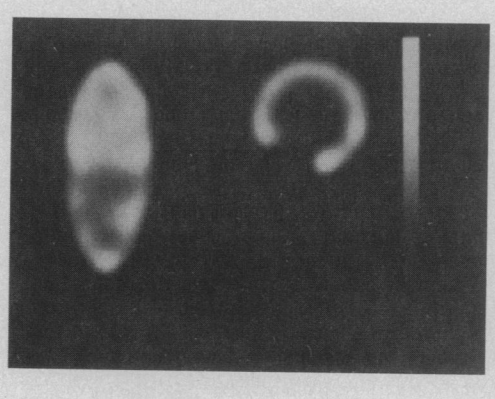

Figure 1. Model radioactivity distribution used for the SUPER PETT-I tomography simulation. The image shown is a 128-by-128 array, obtained by quantizing the heart-liver phantom[15] into square pixels of length and width Δ=0.25cm.

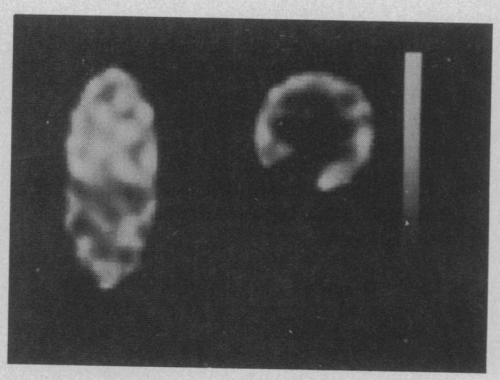

Figure 2. Results of applying the EM equations of (2) and (3) to the time-of-flight simulated measurement data, generated from a Poisson distribution with a mean equal to the convolution of the radioactivity distribution shown in Fig. 1 with a measurement-error density (see text for details). The maximum-likelihood estimates were derived using the EM equations of (3). Iteration 50 of the EM algorithm is shown. Approximately 100,000 measurement points were simulated.

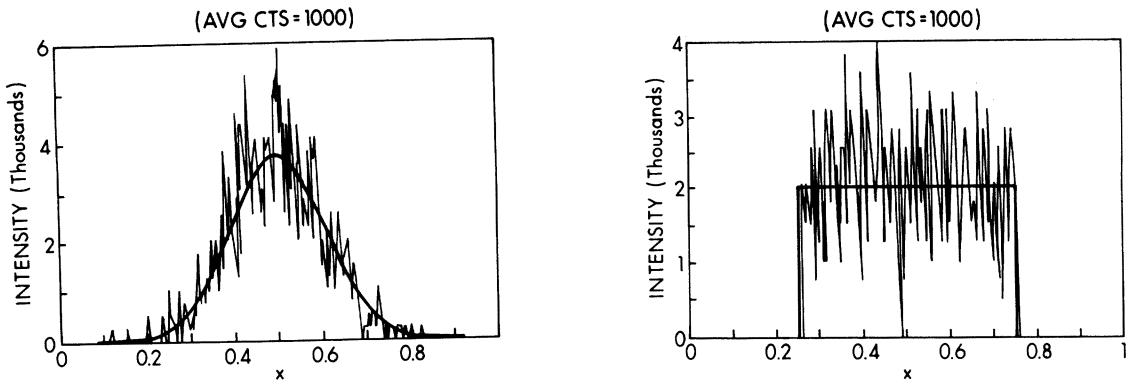

Figure 3. The results of applying maximum-likelihood estimation to a one-dimensional Poisson process, generated with Gaussian (left graph) and rectangular (right graph) mean rate of discharge. An average of 1000 measurement points were simulated in the 512 bit histogram.

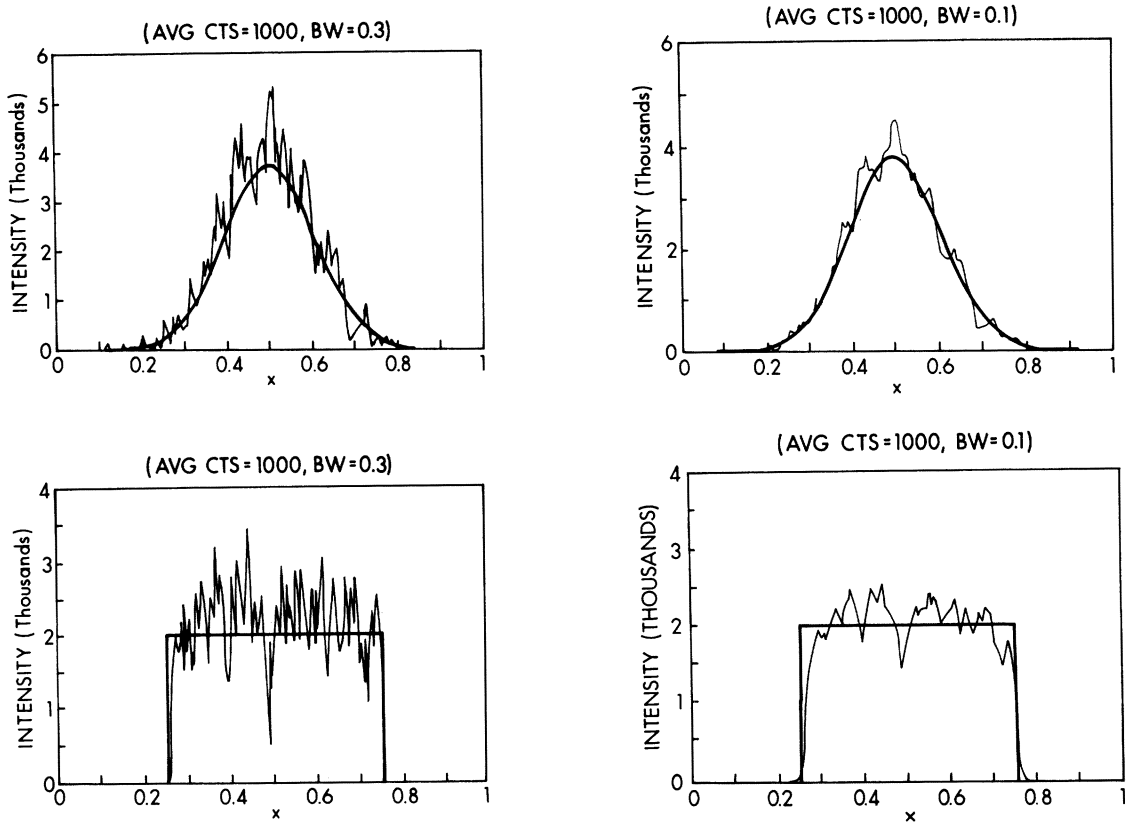

Figure 4. Plots of the maximum-likelihood estimates generated with the bandwidth penalty of equations (9a) and (9b), for different values of the bandwidth BW (see text for definition of BW). Both rows show the estimates derived from the simulation data of Fig. 3, generated with bandwidths BW=0.3 (left column) and BW=0.1 (right column).

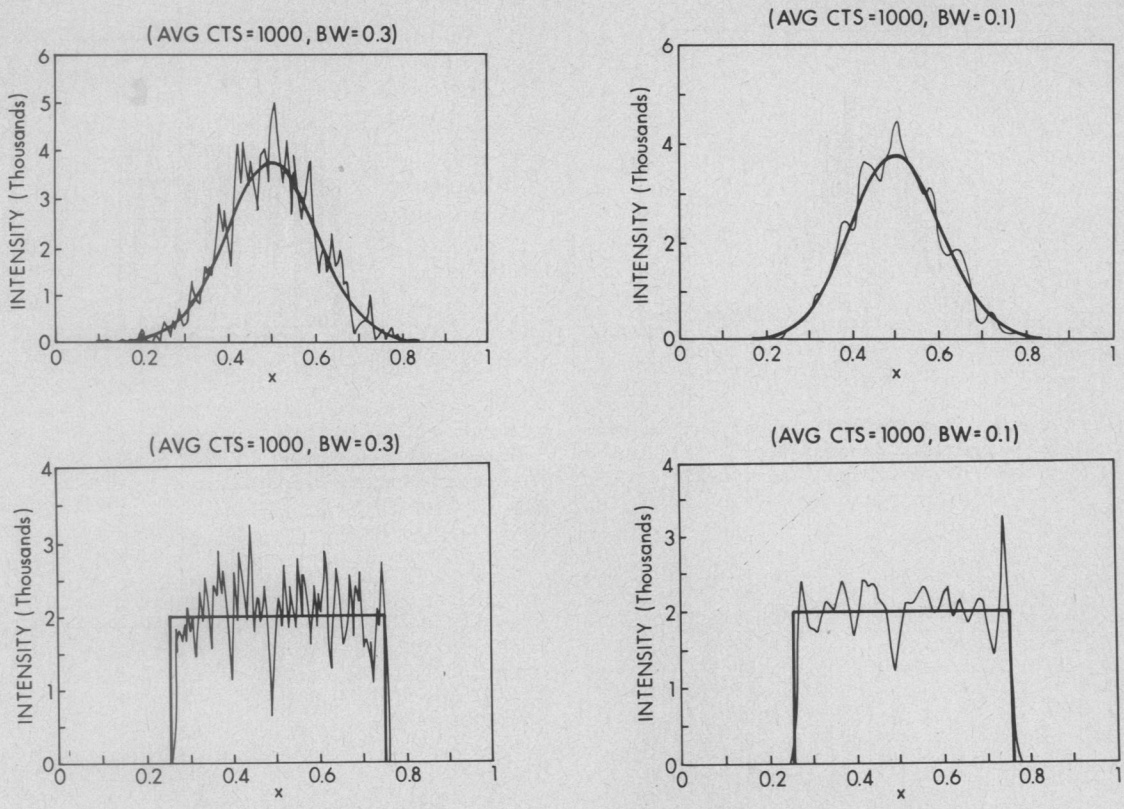

Figure 5. Plots of the maximum-likelihood estimates generated with the *kernel sieve* of equation (12), based on the implementation equations of (13a) and (13b). A Gaussian kernel $K_m(x,\sigma)=(m/\sqrt{2\pi})\exp(-m^2(x-\sigma)^2/2)$ was used, for different values of the kernel width m, where $m=\sqrt{2}(\pi BW/\Delta x)$. Both rows show the kernel-sieve estimates derived from the simulation data shown in Fig. 3, where the kernel width was varied by adjusting BW; that is BW=0.3 (left column) and BW=0.1 (right column).

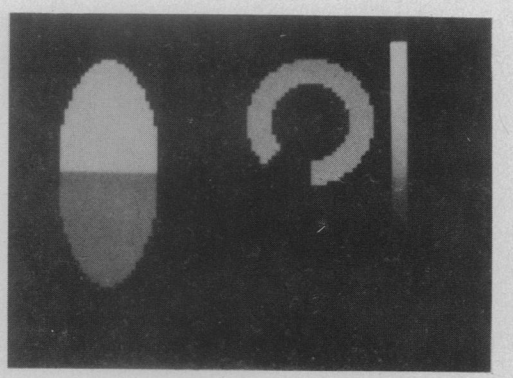

Figure 6. The maximum-likelihood estimate of the heart-liver radioactivity distribution shown in Fig. 1, generated with the convolution-kernel sieve of equation (12), and computed with the EM equations of (17a) and (17b). The estimate derived from the EM algorithm is shown for the fiftieth iteration. The simulated data are identical to the data used for the generation of the unconstrained maximum-likelihood estimates of Fig. 2. The Gaussian convolution kernel used for the sieve was chosen to have a FWHM=1.5cm.