The Annals of Statistics
1995, Vol. 23, No. 4, 1267–1300

A STUDY OF LEAST SQUARES AND MAXIMUM LIKELIHOOD FOR IMAGE RECONSTRUCTION IN POSITRON EMISSION TOMOGRAPHY[1]

BY FINBARR O'SULLIVAN

University of Washington

Positron emission tomography (PET) is a radiologic tool offering a unique capability for measuring tissue metabolism in vivo. A number of biological and physical factors limit the resolution of PET so often the statistical aspects of image reconstruction have an appreciable effect on the quality of information obtained from a study. To a first approximation, reconstruction involves the solution of a linear inverse problem with a line-integral Radon-type transform. Standard filtered back-projection reconstruction is based on the method of least squares. Although computationally efficient, the method does not enforce positivity constraints leading to undesirable negative artifacts in the results. Maximum likelihood based approaches to reconstruction do not suffer from this problem, but their computational complexity has limited the ability to determine quantitatively the improvements in image quality. In this paper, asymptotic approximations and numerical simulations are used to examine the least squares and maximum likelihood methods in some detail. The studies are carried out for idealized representations of conventional and time-of-flight tomographs. The asymptotic analysis indicates that for a range of Sobolev norms the rates of estimation of least squares and maximum likelihood reconstructions are of the same order. This is borne out by numerical studies. However, in these studies maximum likelihood is found to be more efficient than least squares: on a conventional distance–angle tomograph, the root mean square error is on the order of 10–20% smaller for maximum likelihood reconstructions. The corresponding reduction in the root mean square error on a time-of-flight tomograph is in the 5% range. Similar results are found using more complex region of interest based measures of reconstruction accuracy. In addition it is found that much of the effect of maximum likelihood can apparently be realized by postprocessing least squares solutions in an ad hoc manner to reduce negativity artifacts. Since the postprocessing scheme is computationally fast, this suggests that in PET it may be possible to realize the benefits of maximum likelihood without the substantial computational overhead of the EM algorithm.

1. Introduction. Positron emission tomography uses the decay characteristics of a radioactive isotopes such as F-18 fluoro-deoxyglucose or O-15

Received September 1989; revised December 1994.

[1]This research was supported in part by the National Institutes of Health Grants CA-57903, CA-42593 and CA-42045.

AMS 1991 *subject classifications.* Primary 62G05; secondary 62P10, 41A35, 41A25, 47A53, 45L10, 45M05.

Key words and phrases. Asymptotic approximation, rates of estimation, filtered back-projection, positron emission tomography, time-of-flight, maximum likelihood, positivity, regularization, Sobolev norms.

water [32, 39] to image the biologic function of tissue in vivo. Relative to more conventional radiologic techniques such as X-ray computed tomography [33] and magnetic resonance imaging [24], the spatial resolution of positron emission tomography (PET) is limited by dose constraints of the injected radioisotope. However, because the information provided by PET has the potential to quantify the metabolic state of tissue [32, 39, 58], there has been substantial scientific interest in developing refined approaches to data analysis functions associated with the technology [36, 42]. Motivated by this, the statistical aspects of the reconstruction problem have attracted considerable practical and theoretical interest. (A referee pointed out that other aspects such as variance estimation should be receiving more statistical attention [9, 22].) Mathematically, the reconstruction problem is formulated as the inversion of a linear integral operator based on Poisson-distributed data [5, 40, 29, 54, 59]: the observable data is regarded as a realization of an inhomogeneous Poisson process whose intensity is related to the target source distribution of interest by a linear integral equation. The reconstruction problem is to estimate the source distribution given a realization of the Poisson process. The conventional approach to reconstruction in PET is an algorithm known as filtered back-projection (FBP). This can be viewed as a discrete Fourier approximation to a least squares reconstruction. Maximum likelihood reconstruction was initially suggested by Rockmore and Mackovski [44]; it became feasible when Shepp and Vardi [45] and Vardi, Shepp and Kaufmann [54] showed how the expectation–maximization (EM) algorithm could be used for its computation. Building on this, a rich collection of refinements of the approach have been developed [16, 17, 19, 30, 40, 46, 48]. Unfortunately, the computational burden of these methods has limited the ability to quantify the practical gains in reconstruction accuracy [16, 18, 40]. A simple maximum likelihood reconstruction currently requires at least two orders of magnitude more computing time than FBP. As a result maximum likelihood reconstructions are only rarely used in an operational setting [18]. Comparisons between maximum likelihood and least squares continue to be of some theoretical interest.

In this paper, the least squares and maximum likelihood approaches to PET image reconstruction are studied using asymptotic approximation and numerical simulation tools. The analyses are carried out for idealized representations of conventional and time-of-flight tomographs which have been previously employed in the statistical literature [5, 29, 59]. The measurement models and reconstruction methods are described in Section 2. Section 3 develops upper bounds on rates of convergence in Sobolev norms for regularized versions of least squares and maximum likelihood reconstructions. These calculations indicate that for a fixed target function the rates of convergence of the least squares and maximum likelihood reconstruction methods are the same. This is consistent with min–max error computations of Bickel and Ritov [5], Johnstone and Silverman [29] and Zhang [59]. The analysis of the maximum likelihood case is complicated by the nonlinear nature of the estimator. The results complement the recent work of Eggermont and

LaRiccia [14]. The asymptotic analysis developed in this paper also applies to more general versions of maximum likelihood reconstruction, recently proposed by Vardi and Lee [53] for the solution of linear inverse problems with positivity constraints.

Johnstone and Silverman [29] pointed out the need to demonstrate the relevance of asymptotic analysis, via min–max rates of estimation [5, 29, 59] or even heuristic arguments based on the Cramér–Rao inequality [54], in practice. Motivated by this, numerical studies are carried out to examine the error characteristics of the reconstruction methods in a more practical setting (see Section 4). The results indicate good agreement with the qualitative dose–response (rate of estimation) relationship suggested by the asymptotics. More detailed comparisons between the reconstruction methods are recorded. Over a range of simulated injected doses, maximum likelihood is found to be more efficient than least squares: On a conventional distance–angle tomograph, the root mean square error is typically 10–20% smaller for the maximum likelihood approach. The corresponding reduction in the mean square error on a time-of-flight tomograph is much smaller: around 5%. Similar results are found using more complex region of interest based measures of reconstruction accuracy. The numerical studies also indicate that much of the effect of maximum likelihood can apparently be realized by postprocessing the least squares solution in a rather heuristic manner to reduce negativity artifacts. The computational cost of this postprocessing is roughly equivalent to two or three applications of filtered back-projection. Techniques of this type are of considerable practical interest because currently the computational demands of maximum likelihood reconstructions are too great for routine operational use.

2. Tomograph model and reconstruction methods.

2.1. *Measurement model.* In a PET study a subject is placed in the aperture of the tomograph and injected with a quantity of a positron emitting radioisotope. When the isotope decays the emitted positron is annihilated by a nearby electron resulting in the creation of a pair of 512-keV photons radiating along a straight line from the annihilation site in opposite directions from each other. The orientation of the line of flight of the photons is uniformly distributed in space. Multiple banks of integrated detectors arranged in circular rings describe cross-sectional planes through the tissue section. By appropriate collimation, measurements are made on those emissions for which the line of flight of the photon pair lies in one of these cross-sectional planes. Figure 1 shows a schematic. Conventional tomographs measure the line of flight of the photon pair. This information is encoded in a distance (d) and angle (a) measurement. More elaborate time-of-flight tomographs, such as the SP-3000 UWPET [31], use the difference between the time of detection of the individual photons in the pair to measure the location of the annihilation source along the line of flight. This is referred to as a time (t) measurement (see Figure 1). The system has several sources of noise

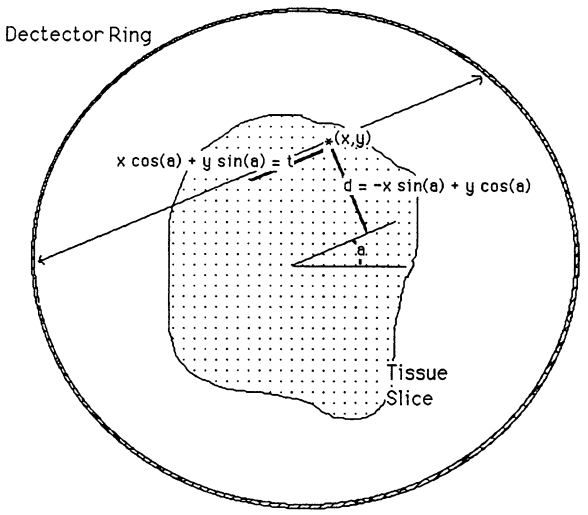


FIG. 1. *Illustration of PET measurements for conventional and time-of-flight tomography: time* (t)*, distance* (d) *and angle* (a) *measurements for a positron emission at the point* (x, y).

which lead to statistical errors in the time, distance and angle measurements. Currently, these errors are modeled as independent Gaussians. The most significant errors occur in the time and distance measurements. The resolution of the tomograph is determined by the standard deviation (σ_b) of the distance measurement. Ultimately this parameter is used to set the size of the picture elements (pixels) in image reconstruction software [26]. The standard deviation of the time-of-flight measurements (σ_t) is substantial, typically 15–20 times larger than σ_b. For example, the UWPET is a time-of-flight tomograph with $\sigma_b = 0.19$ cm and $\sigma_t = 3.82$ cm.

Let Z denote the two- or three-dimensional (with time-of-flight) observation domain. The information gathered on a collection of n detected emissions $\{z_i,\ i = 1, 2, \ldots, n\}$ can be regarded as a random sample from a probability density g with g: $Z \to \mathbb{R}$. The relationship between g and the source distribution of radioisotope (denoted by θ_0) is described by a system operator R with $g = R\theta_0$. Since the distribution of isotope changes over time, due to metabolism, θ_0 actually represents a time-averaged source distribution corresponding to the window of time during which the tomograph measurements are made [32]. In a conventional tomograph, the model in Figure 1 implies the relation

$$g(d, a) = R_a w_b \star \theta_0[d],$$

where $w_b \star \theta_0$ is the convolution of the source distribution with a spherically symmetric Gaussian with componentwise standard deviation σ_b; R_a is the parallel beam line-integral Radon transform at angle a [33, 50]. With time-of-flight data,

$$g(t, d, a) = R_a^T w_b \star \theta[d, t],$$

where the R_a^T convolves the function $w_b \star \theta_0$ in the t-direction (see Figure 1) with a one-dimensional Gaussian whose standard deviation is $\sigma = \sqrt{\sigma_t^2 - \sigma_b^2}$. Note that, as $\sigma_t \to \infty$, $R_a^T w_b \star \theta[d, t] \to R_a w_b \star \theta[d]$ for all t.

Several additional factors must be accounted for in practice: the most important are object attenuation; nonuniformities in detector efficiencies; and random coincidences [26, 29, 40]. Object attenuation and nonuniformities in detector efficiencies modify the probability that an event associated with a particular line of flight will be recorded. The system operator is better approximated by a scaled version of the operator R introduced in the preceding paragraph. Thus the system operator is $K = CR$, where C is a diagonal operator whose action is to multiply a function by a fixed function $c(z)$ for $z \in Z$. In PET, tissue attenuation is routinely measured by a separate transmission scan obtained before emission data is acquired [26]. The random coincidence effect arises due to inaccuracies in detection: a recorded event is defined as a pair of photon detections occurring within a small time interval of one another. The time interval is known as the coincidence time window. Depending on the width of this coincidence window, however, it is possible for two unrelated photons to be detected as a false event. Such events are known as random coincidences. If the size of the coincidence window is made too small, the sensitivity of the tomograph is compromised. In practice, random coincidences account for between 10 and 20% of detected events; thus it is important to account for the effect [21, 40]. With random coincidences, a reasonable model for the probability distribution of recorded events is

$$g = K\theta_0 + g^*, \tag{1}$$

where $K = CR$ and g^* is proportional to the probability distribution of random coincidences. Reliable estimates of C and g^* are available to the reconstruction process, so the only unknown is the probability distribution of the source distribution θ_0. Additional adjustments for the detection of scattered events are also of interest [2, 3]. This is a complex problem. Typically, detector collimation and energy acceptance thresholds are such that scatter can be regarded as a second-order effect, relative to attenuation, detector nonuniformity and random coincidences [54].

2.2. *Reconstruction methods.* Reconstruction is carried out in a discretized framework [26, 40, 46, 50, 54]. Given the resolution limits of the tomograph, the reconstruction goal is focused on the recovery of the source distribution up to convolution with a Gaussian kernel representing the accuracy of the detectors (i.e., the target function is $\theta_b = w_b \star \theta_0$). A similar type of reduction is standard in computerized tomography [33, 47]. Without this reduction the reconstruction problem would be hopelessly difficult from an asymptotic statistical point of view [8]. Let there be I discretization elements in the imaging domain and J discretization elements in the observation domain. In this discretized setting, the data consist of a vector of binned counts y_j, for $j = 1, 2, \ldots, J$, which are modeled as independent Poisson variates with rate parameters given by μ_j, for $j = 1, 2, \ldots, J$. The

source distribution is a vector $\lambda = (\lambda_i;\ i = 1, 2, \ldots, I)$. Formally, λ_i and μ_j's are the integrals of θ_b and g over the appropriate discretization elements in the imaging or observational domains, multiplied by a scale factor associated with the dose of radioisotope injected into the tissue; λ and μ are connected by a linear equation,

$$\mu = K\lambda + \mu^r,$$

where K is a $J \times I$ matrix representing the discretized linear integral operator K in (1) and μ^r are the rates of random coincidences.

Following Vardi, Shepp and Kaufman [54] and Geman, Manbeck and McClure [16], it is convenient to present reconstruction methods in their raw form first and subsequently discuss techniques used to impose spatial smoothing constraints by regularization or by low-pass filtering.

Maximum likelihood and the EM algorithm. The raw maximum likelihood (ML) solution maximizes the likelihood or log-likelihood of the data under the assumed Poisson model. Up to a constant, the negative log-likelihood is given by

$$\ell(y|\lambda) = \sum_{j=1}^{J} \left\{ y_j \log\left(K_j'\lambda + \mu_j^r\right) - K_j'\lambda \right\},$$

where $K_j'\lambda$ is the dot product between the jth row of K and λ. Vardi, Shepp and Kaufman [54] showed that the expectation–maximization (EM) algorithm could be used to compute the ML solution. The extension to a model with random coincidences is developed in Politte and Snyder [40]. The EM iteration involves two steps corresponding to the calculation of a conditional expectation of the log-likelihood and its subsequent maximization. When these two steps are combined the iteration ($\lambda^{\text{old}} \to \lambda^{\text{new}}$) is

$$\lambda_i^{\text{new}} = \lambda_i^{\text{old}} \cdot \sum_j \frac{y_j k_{ji} / k_{\cdot i}}{\sum_s k_{js} \lambda_s^{\text{old}} + \mu_j^r},$$

where k_{ji} is the $(j, i)'$-th element of the matrix K and $k_{\cdot i} = \sum_j k_{ji}$. This iteration is performed to convergence and the raw ML estimate λ^{ml} is set to the final value of λ^{new}. Vardi and Lee [53] pointed out that the EM algorithm could be used even if the Poisson assumptions for the data y did not hold. In this case, the negative log-likelihood $\ell(y|\lambda)$ is merely regarded as a distance criterion. As noted in the next section, for consistency it is important that $\mathbb{E}y = K\lambda^0 + \mu^r$, where λ^0 is the target parameter of interest.

Least squares and filtered back-projection. The raw least squares solution minimizes the residual sum of squares between a corrected data vector and its expected value under the model. The corrected data is

$$y^c = C^{-1}(y - \mu^r),$$

where C is the discretized diagonal scaling matrix. With this, the least squares estimate $\lambda^{(\mathrm{ls})}$ minimizes the quadratic

$$l^{(\mathrm{ls})}(y|\lambda) = \tfrac{1}{2}\sum_j (R_j'\lambda)^2 - \sum_j y_j^c R_j'\lambda,$$

where R_j is the jth column of R. The solution is given by

$$\lambda^{\mathrm{ls}} = (R'R)^{-1}R'y^c. \tag{2}$$

In the case that R corresponds to a Radon-type transform (e.g., PET) and the discretization is fine enough, the matrix $R'R$ is approximately Toeplitz (i.e., has spectral decomposition in terms of the two-dimensional discrete Fourier transform). Analytic approximations to the spectrum are available [33, 50]. In conventional tomography the ith eigenvalue is proportional to i^{-1} for $i = 1, 2, \ldots, I$. For time-of-flight tomography, the eigenvalues have a more complicated structure [50]. Note that the eigenvalues of a Toeplitz matrix may be computed numerically in terms of its action on a single vector. If δ is a vector, the application of $R'R$ to δ gives a vector p with $p = R'R\delta$ and to within the Toeplitz approximation,

$$\hat{p} = \hat{\xi}\cdot\hat{\delta},$$

where the caret indicates the discrete Fourier transform. Thus the vector $\hat{\xi} = \hat{p}/\hat{\delta}$ contains the spectrum of $R'R$. Figure 2 shows $\hat{\xi}^{-1}$ for a conventional and time-of-flight tomograph (see Section 4 for details of the discretization). In both cases the spectrum is essentially linear.

The approximate Toeplitz structure is the basis of filtered back-projection (FBP) algorithm. The data are first *backprojected* to create the quantity $R'y^c$; subsequent multiplication by the Fourier approximation to $(R'R)^{-1}$ is the *filtering* step. On a time-of-flight machine FBP is referred to as confidence-weighted filtered back-projection [26]. The connection between least squares and FBP formally depends on the degree of discretization; FBP is the reconstruction algorithm used on almost all operational PET machines. The great advantage of FBP is its computational speed. After back-projection, filtering is carried out using fast Fourier transform techniques. By way of comparison, each iteration of the EM algorithm used to compute the ML reconstruction, involves a projection (multiplication by K) and back-projection (multiplication by K') operation. As a result, a single EM iteration is roughly equivalent to two applications of FBP. The most efficient implementations of the EM algorithm (with appropriate initialization) typically require on the order of 1000 iterations to produce satisfactory estimates [16, 40]; thus EM reconstructions can be as much as three orders of magnitude slower than FBP. The integration of smoothing into the EM algorithm can substantially reduce the number of iterations required (for a fixed amount of smoothness). For example, the approach of Snyder and Miller [48] uses on the order of 100 iterations for practical reconstructions. Still this is two orders of magnitude more computing time than filtered back-projection. As a result maximum likelihood has been only rarely used in operational settings. One could argue

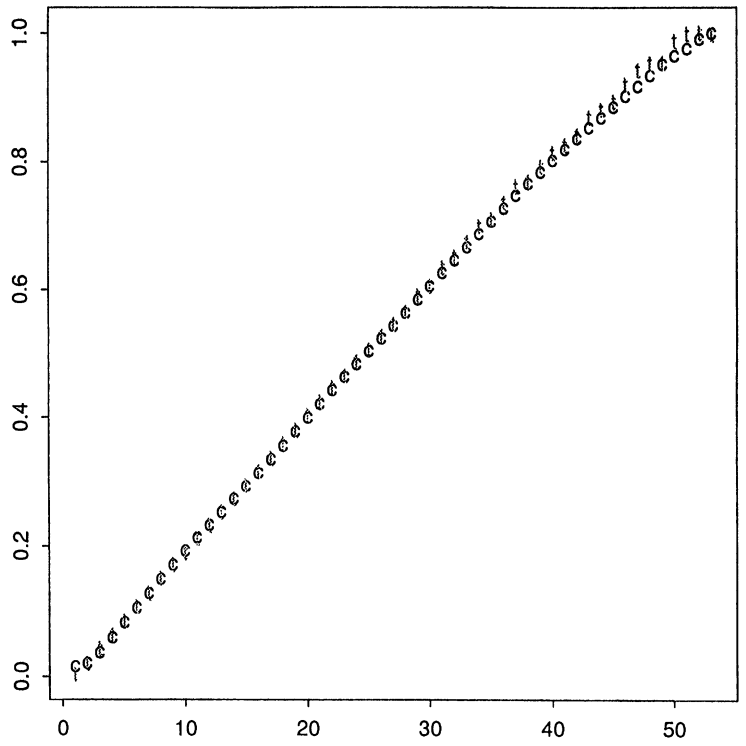


FIG. 2. *Comparison of the reconstruction filters used in conventional (marked "c") and time-of-flight (marked "t") filtered back-projection algorithms: the filters are radially symmetric, only the diagonal of the matrix* ξ^{-1} *is shown up to the folding frequency. The computation is carried out by the approach described in Section 2 using the discretization parameters in Section 4. The full dimension of the image domain is* 26.88 *cm, so the folding frequency occurs at* 13.44 *cm.*

that recent advances in algorithms used to compute ML-type reconstructions [28] and the increasing speed of computers may ultimately change this situation.

Spatial smoothing by filtering and regularization. Reconstructions must be smoothed in some manner in order to obtain acceptable (consistent) results [54]. There are a wide variety of possible approaches, including regularization, the method of sieves and kernel smoothing. The most commonly used approach with filtered back-projection is to smooth the raw FBP solution by convolution with a smoothing kernel [5, 26, 59]. A radially symmetric Gaussian convolution kernel is the typical choice. A similar approach to smoothing ML solutions was suggested by Vardi [52]. An advantage of this is that it allows the examination of a range of solutions corresponding to different amounts of smoothing in a simple manner.

An alternative approach to smoothing is regularization. In regularization (as well as more general Bayesian methods) the objective function (likelihood or least squares criterion) is modified by adding a term which measures the

plausibility (e.g., smoothness), of the solution [15, 16, 30, 46]. Asymptotic properties of continuous analogues of some regularization methods will be examined in the next section. The method of sieves achieves smoothness by restricting the range of optimization in least squares or maximum likelihood [48, 40]. Formally the method of sieves can sometimes be related to regularization because regularization may be regarded as a Lagrangian form for a method of sieves approach.

Further techniques for smoothing, which have been proposed in the context of maximum likelihood reconstructions, include early stopping of the EM iteration [16, 54, 55] and a smoothed EM algorithm introduced by Silverman, Jones, Wilson and Nychka [46]. The latter approach has been extended and developed by Eggermont and LaRiccia [14]. Silverman, Jones, Wilson and Nychka [46] describe a dramatic improvement in computational efficiency for these algorithms relative to the initial approach of Vardi, Shepp and Kaufman [54] similar to the experience of Snyder and co-workers with their method of sieves approaches. Detailed numerical comparisons over a range of count rates between method of sieves and the smoothed EM approach would be interesting.

3. Asymptotic analysis of regularization estimators. Convergence results for regularized versions of the least squares and maximum likelihood type reconstruction methods are developed in this section. In order to cover more general estimators of the type recently introduced by Vardi and Lee [53], the stochastic aspects of the observational model are extended beyond the Poisson structure appropriate for PET. Some statistical background for the asymptotic approximation of regularization estimators is contained in [11], [34] and [56]. Also, Natterer [33] has a brief introduction from a numerical analysis point of view. As in previous work in the field [5, 14, 29, 59], a continuous analogue of the tomograph model will be analyzed. This reduction simplifies the theory; however, the numerical work in the next section indicates that the results from the continuous model provide a reasonable qualitative guide to how least squares and maximum likelihood reconstructions perform in a more practical discretized setting.

Let $\Omega \in \mathbb{R}^2$ be the imaging domain and let L be an mth-order linear differential operator on $L_2(\Omega, \mathbb{R})$. Penalty functionals of the form

$$P(\theta) = \int_\Omega L\theta[x]^2 \, dx = (L\theta, L\theta)$$

will be considered in the definition of regularization functionals (cf. [34] and [56]). Regularization estimators are naturally studied in terms of Sobolev norms (see Natterer [33], section VII.4, for a five-page introduction adequate for the development here). For m a real number (not necessarily an integer), let H^m be the Sobolev space of functions on $\mathbb{R}^d$ with norm

$$\|\theta\|_{H^m}^2 = \int_{\mathbb{R}^d} \left(1 + |\xi|^2\right)^m |\hat{\theta}(\xi)|^2 \, d\xi,$$

where $\hat{\theta}$ is the Fourier transform of θ and $|\xi|$ is the Euclidean norm of a vector in $\mathbb{R}^d$. For a bounded open subset Ω of $\mathbb{R}^d$, $H_0^m(\Omega)$ is the restriction of H^m to functions supported on the closure of Ω; $\bar{H}^m(Z)$ is the Sobolev space of functions defined on the extension of Z to its entire range, that is, the infinite cylinder $[0, 2\pi] \times \mathbb{R}^1$ for conventional PET or $[0, 2\pi] \times \mathbb{R}^2$ with time of flight. (See page 45 of Natterer [33] for further generalizations.) Two sets of assumptions will be made. The first is as follows.

ASSUMPTION 1.

(i) The set Ω is a bounded, open, simply connected, nonempty set with C^∞ boundary. The observational domain Z is a closed bounded subset of $\mathbb{R}^2$ or $\mathbb{R}^3$.

(ii) The operators R: $L_2(Z)$ and L: $L_2(\Omega) \to L_2(\Omega)$ are linear operators, and for some $\beta > 0$ and $m > 0$ there are positive constants M_0 and M_1 such that, for all $\theta \in L_2(\Omega)$,

$$M_0 \|\theta\|_{H_0^{-\beta}(\Omega)} \le \|R\theta\|_{L_2(Z)} \le M_1 \|\theta\|_{H_0^{-\beta}(\Omega)}$$

and, furthermore,

$$M_0 \le \frac{\{\|R\theta\|_{L_2(Z)} + \|L\theta\|_{L_2(\Omega)}\}}{\{\|\theta\|_{H_0^{-\beta}(\Omega)} + \|\theta\|_{H_0^m(\Omega)}\}} \le M_1.$$

The penalty function is $P(\theta) = \|L\theta\|^2_{L_2(\Omega)}$.

Assumption 1(i) is used for the construction of Sobolev spaces and their interpolants on Ω [33]. The first inequality in Assumption 1(ii) specifies smoothness characteristics for the operator R [cf. Natterer [33], equation (IV.1.10)]. The parameter β specifies the difficulty of the reconstruction problem: the larger the value of β, the more difficult the reconstruction (see Theorem 1). The second set of inequalities in Assumption 1(ii) will allow the construction of a natural set of parameter spaces for considering regularization estimators and determining their approximation characteristics [11, 34]. The assumption implies that there are sequences of eigenfunctions $\{\phi_\nu \in H_0^m(\Omega)\colon\ \nu = 1, 2, \ldots\}$ and corresponding eigenvalues $\{\gamma_\nu\colon\ \nu = 1, 2, \ldots\}$ with $0 < \gamma_1 \le \gamma_2 \le \cdots$ satisfying

$$(R\phi_\nu, R\phi_\mu) = \delta_{\nu\mu}, \qquad (L\phi_\nu, L\phi_\mu) = \gamma_\nu \delta_{\nu\mu}, \tag{3}$$

for all pairs ν, μ of positive integers, where $\delta_{\nu\mu}$ is Kronecker's delta. Also with $r = m + \beta$, γ_ν / ν^r is bounded away from zero and infinity as $\nu \to \infty$. For $b \ge 0$, let

$$\|\theta\|_b^2 = \sum_\nu (1 + \gamma_\nu^b)(R\phi_\nu, R\theta)^2,$$

and let Θ_b be the completion of the set $\{\theta \in H_0^m(\Omega);\ \|\theta\|_b < \infty\}$ in the norm $\|\cdot\|_b$; Θ_b is a Hilbert space with inner product

$$\langle \theta, \zeta \rangle_b = \sum_\nu (1 + \gamma_\nu^b)(R\phi_\nu, R\theta)(R\phi_\nu, R\zeta).$$

Regularization will be carried out in the space $\Theta = \Theta_1$. By interpolation theory, Assumption 1(ii) implies that, for $b \in [0,1]$, Θ_b and $H_0^{(m+\beta)b-\beta}(\Omega)$ are equal as sets and have equivalent norms. Throughout this section, the parameters b, s and s_0 are used to refer to interpolation spaces and norms (Θ_b etc.); m and β refer to Sobolev spaces and norms (H^m etc.).

Assumptions for the observational model and target parameter are given next.

ASSUMPTION 2.

(i) The observed data is a random sample $\{(y_i, z_i),\ i = 1,2,\ldots,n\}$, where $z_i \in Z$ and $y_i \in \mathbb{R}$. The conditional mean and variance of y_i given $z_i = z$ are bounded, uniformly in z.

(ii) For some $\delta > d/2$ (where d is the dimension of Z) there is an $s_0 > 0$ and positive constant M such that, for θ: $\Omega \to \mathbb{R}$,

$$\|R\theta\|_{\overline{H}^\delta(Z)} \le M\|\theta\|_{s_0}.$$

(iii) For some $1 \ge s > 2s_0 + 1/r$ with $r = m + \beta$, the target parameter θ_0 is in Θ_s and satisfies

$$g(z) = h_y(z)h(z) = c(z)R\theta_0[z] + g^*,$$

where $h_y(z)$ is the conditional expectation of y given z. The functions c and g^* are uniformly bounded away from zero and infinity; c, g^* and h are in $\overline{H}^\delta(Z)$.

The introduction of the additional complication to the measurement, the term y_i, allows the theory to be applied to the generalized versions of maximum likelihood *type* reconstructions discussed by Vardi and Lee [53]. In a discretized setting the observed data could be a realization from a marked Poisson process. In PET, the y_i's would traditionally be assumed to be identically equal to 1. Note that the target parameter lies in a space Θ_s which does not have to be as smooth as the nominal parameter space Θ_1.

Letting the discretization in Section 2.2 get finer yields continuous analogues of the least squares and maximum likelihood criteria. Regularized least squares and maximum likelihood type estimators are minimizers over Θ of the functionals

$$\ell_{n\tau}^{(\mathrm{ls})}(\theta) = \frac{1}{2}\int_Z R\theta[z]^2\,dz + \int_Z \frac{R\theta[z]g^*(z)}{c(z)}\,dz - \frac{1}{n}\sum_i \frac{y_i R\theta[z_i]}{c(z_i)} + \frac{\tau}{2}P(\theta),$$

$$\ell_{n\tau}^{(\mathrm{ml})}(\theta) = \int_Z c(z)R\theta[z]\,dz - \frac{1}{n}\sum_i y_i \log(c(z_i)R\theta[z_i] + g^*(z_i)) + \frac{\tau}{2}P(\theta),$$

respectively. The regularization parameter is $\tau > 0$. Bayesian interpretations of regularization estimators can be found in Wahba [56], for example. In general it should be emphasized that the "likelihood" component of the functional $\ell_{n\tau}^{(\mathrm{ml})}$ is not necessarily interpretable as a likelihood except for the

Poisson sampling model of Vardi, Shepp and Kaufman [54] (where $y_i = 1$). Other examples of regularization functionals of this type are discussed in Cox and O'Sullivan [12]. Note that by Assumption 2(ii) and Sobolev's imbedding theorem, the regularization functionals are continuous in Θ_b for $b \geq s_0$. Also, both functionals are convex and thus amenable to gradient based minimization algorithms. The elements of Θ can be written in a series expansion (convergent in Θ)

$$\theta = \sum_{\nu} \theta_\nu \phi_\nu,$$

where $\theta_\nu = (R\theta, R\phi_\nu)$ for $\nu = 1, 2, \ldots$. An explicit representation for the least squares estimator $\theta_{n\tau}^{(\mathrm{ls})}$ is

$$\left(R\theta_{n\tau}^{(\mathrm{ls})}, R\phi_\nu\right) = \frac{\int_Z R\phi_\nu[z] g^*(z)/c(z)\, dz - (1/n)\Sigma_i y_i R\phi_\nu[z_i]/c(z_i)}{[1 + \tau\gamma_\nu]}.$$

It is easy to check that $\theta_{n\tau}^{(\mathrm{ls})}$ is in Θ. The maximum likelihood estimator does not have an explicit representation; however, by the following theorem its existence in Θ is guaranteed, at least for τ sufficiently small. The following theorem is proved in the Appendix.

THEOREM 1. *Suppose there is an* α *satisfying*

$$s_0 < \alpha < \frac{s - 1/r}{2}, \tag{4}$$

and a sequence $\tau_n \to 0$

$$n^{-1}\tau_n^{-2(\alpha + 1/r)} \to 0. \tag{5}$$

Then under Assumptions 1 *and* 2, *given* $\varepsilon > 0$, *there are a constant* M *and an* n_0 *such that, for all* $n > n_0$,

$$P\Big\{\theta_{n\tau_n}^{(\mathrm{ls})} \text{ and } \theta_{n\tau_n}^{(\mathrm{ml})} \text{ exist in } \Theta \text{ and } \forall\, b \in [0, \alpha], \text{ and satisfy} \\ \|\theta_{n\tau_n}^{(\mathrm{ls}),(\mathrm{ml})} - \theta_0\|_b \leq M\left[\tau_n^{(s-b)/2}\|\theta_0\|_s + n^{-1/2}\tau_n^{-(b+1/r)/2}\right]\Big\} > 1 - \varepsilon. \tag{6}$$

The proof of the theorem is given in the Appendix. Note that the limits on s in Assumption 2(iii) implies there is an α satisfying (4). The first term on the r.h.s. of the inequality in (6) gives an upper bound on the order of the asymptotic bias in the $\|\cdot\|_s$-norm and the second term gives the order of the asymptotic standard deviation. Minimizing the sum with respect to τ gives

$$\tau_n \approx n^{-1/(s+1/r)}, \tag{7}$$

which results in

$$\|\theta_{n\tau_n} - \theta_0\|_b^2 = O_p\left(n^{-(s-b)/(s+1/r)}\right).$$

It is easy to see that we can choose α so that both (4) and (5) hold when τ_n is given by (7). With $b = \beta/(\beta + m)$, with Assumption 1(ii) an upper bound on

the integrated squared error of the regularization estimators is obtained. If $\theta \in H_0^m(\Omega)$ (i.e., $s = 1$), then this bound is of the form

$$\|\theta_{n\tau_n} - \theta_0\|^2_{L_2(\Omega)} = O_p(n^{-m/(m+1+\beta)}).$$

This can be compared with the $O_p(n^{-m/(m+1)})$ bound for two-dimensional nonparametric regression function estimation in similar spaces [10]. Cox [10] actually considers Sobolev spaces without the zero boundary condition; however, the results rely on the asymptotic spectral characteristics of certain elliptic differential operators not affected by replacing the Sobolev spaces of Cox [11] with the $H_0^m(\Omega)$ spaces considered here [51].

For the least squares estimator, the upper bound on the rate of convergence obtained from Theorem 1 is also the upper bound on the expected value of the reconstruction error in the $\|\cdot\|_b$-norm (see Lemma 1). In the case that the y_i's are all unity (i.e., the usual PET model) min–max optimality of this rate can be established by simple adaptation of the arguments in Johnstone and Silverman [29]. The argument relies on the fact that balls in the spaces Θ_b have a rather simple structure so it is straightforward to modify the proofs of Propositions 5.1 and 5.2 of Johnstone and Silverman [29]. Indeed, this is already contained in the comments of paragraph 3, Section 8 of their paper.

Application to PET. Verification of Assumptions 1 and 2 is straightforward in the context of PET. We only consider the case that $\sigma_b = 0$. The structure of R^*R is described as a ramp filter in the Fourier domain [33]. Here the asterisk denotes the $L_2(Z)$ adjoint. Figure 2 is consistent with this. For $\theta \in H_0^m(\Omega)$,

$$\widehat{R^*R\theta} = \frac{m_0 \hat{\theta}(\xi)}{|\xi|},$$

where the caret denotes the Fourier transform and m_0 is a normalization factor which ensures that the integral of $R^*R\theta$ and θ are the same. If L is an mth-order differential operator, L^*L is an elliptic differential operator of order $2m$ in $\mathbb{R}^2$ [51]. Using the Sobolev space estimates for the Radon transform (e.g., Natterer [33], Theorem II.5.1), Assumption 1(ii) follows with $\beta = \frac{1}{2}$ for conventional PET. Also, Theorem II.5.3 gives Assumption 2(ii) with $s_0 = \delta/r$ and $s_0 > 1/r$.

With time-of-flight data [50, 59] the Fourier domain representation of $R^{T*}R^T$ becomes

$$\begin{aligned}\widehat{R^{T*}R^T\theta} &= m_0 \hat{\theta}(\xi)\left[\int_{-\pi/2}^{\pi/2} \sigma_t \exp\left\{-|\xi|^2\sigma_t^2 \cos(a)^2\right\} da\right] \\ &= 2m_0 \hat{\theta}(\xi)\left[\int_0^{\pi/2} \sigma_t \exp\left\{-|\xi|^2\sigma_t^2 \cos(a)^2\right\} da\right],\end{aligned}$$

where m_0 is another normalization factor. Here, since $\cos(a)/(\pi/2 - a)$ is bounded away from zero and infinity on $[0, \pi/2]$, by change of variable the

integral can be related to a Gaussian form and from this it follows that for $|\xi| > \eta > 0$ there are finite positive constants M_1 and M_2 for which

$$\frac{M_1}{|\xi|} < \left[\int_{-\pi/2}^{\pi/2} \sigma_t \exp\left\{-|\xi|^2 \sigma_t^2 \cos(a)^2\right\} da\right] < \frac{M_2}{|\xi|}.$$

Hence, using the Cauchy–Schwarz inequality, the operators R and R^T are equivalent in the sense that

$$M_1 \|R\theta\|_{L_2(Z)} \leq \|R^T\theta\|_{L_2(Z^t)} \leq M_2 \|R\theta\|_{L_2(Z)},$$

for M_1 and M_2 finite strictly positive constants and Z^t denoting the observational domain for time-of-flight PET measurements. Using interpolation (see [33], Theorem II.5.2), the equivalence is easily extended to more general Sobolev norms on Z or Z^t to obtain

$$M_1 \|R\theta\|_{\bar{H}^{2\delta/3}(Z)} \leq \|R^T\theta\|_{\bar{H}^{\delta}(Z^t)} \leq M_2 \|R\theta\|_{\bar{H}^{2\delta/3}(Z)},$$

for $\delta > 0$. Thus for time-of-flight PET we can take $\delta > \frac{3}{2}$ and again let $s_0 > 1/r$ to obtain Assumption 2(ii).

Relative to estimation of a density based on direct measurements, the indirect measurements of PET is seen to slow the asymptotic convergence by introducing an extra 0.5 in the denominator of the rate; that is, the integrated squared error is $O_p(n^{-m/(m+1.5)})$ as opposed to $O_p(n^{-m/(m+1)})$ for regular two-dimensional density estimation with a target function having m orders of smoothness. This result has a very similar flavor to those obtained by Bickel and Ritov [5] and Zhang [59] in Hölder spaces. Also note the calculations indicate that the rates of convergence for conventional and time-of-flight PET are the same, which compares well with the work of Zhang [59].

4. Numerical studies. Practical versions of least squares and maximum likelihood reconstruction were investigated by some numerical simulation. The focus was to develop an empirical understanding of the relationship between the count rate in the PET study (n) and the reconstruction error and to compare this qualitatively with the behavior suggested by the asymptotic analysis. It was also of interest to study the manner in which time-of-flight information improves reconstruction.

4.1. *Simulation model and reconstruction methods.* Simple computer representations for conventional and time-of-flight tomographs were used [31, 50, 54]. For a conventional tomograph, the specifications were set in accordance with Hoffman, Cutler, Digby and Mazziotta [25]; an imaging domain with $I = 128 \times 128$ pixels of dimension 2.1 mm, an observation domain with $J = 128 \times 160$ distance–angle bins of size 2.1 mm $\times$ $\pi/160$ rad and a transverse detector resolution of $\sigma_b = 1.9$ mm. For the time-of-flight tomograph, a time measurement was added. The standard deviation of the Gaussian time-of-flight measurement error was 3.82 cm and the discretization in the

time direction was 32 bins each of length 2.3 mm [31]. Attenuation, detector efficiencies and random coincidences were not included in the simulation, thus $c(z) = 1$ and $g^*(z) = 0$. The observational model was Poisson, $y \sim P(\mu)$, with

$$\mu = K\lambda = R\lambda_b,$$

where R is a simple trapezoidal approximation to the line integral (or weighted line integral) Radon transforms; λ_b is the discrete convolution between the underlying source distribution λ and a spherically symmetric Gaussian distribution whose standard deviation is σ_b. Note that this implies that the full-width-half-maximum of the Gaussian is 4.5 mm (a scale which is used more frequently in the PET engineering literature).

Three phantom source distributions (λ's) were considered in the simulations (see Figure 4a–c in Section 4.3). The first phantom was slice number 8 of the brain phantom set of Hoffman, Cutler, Digby and Mazziotta [25]. There are three grey levels, corresponding to background, white matter and grey matter. The second phantom has a greater range of grey scales; it comes from Vardi, Shepp and Kaufman [54]. The third phantom is a digitized version of the Jaszczak phantom commonly used for quality monitoring in operational tomographs [31]. Data sets were generated from each of these phantoms using pseudo-Poisson random deviates [43]. The phantom was scaled so that the expected total counts $N = \sum_j \mu_j$ ranged (linearly on a $\log_2$ scale) over nine discrete values between $N = 10^4$ and $N = 10^6$ counts. This range of counts is comparable to what is typically seen on individual scans in a dynamic F-18 deoxyglucose brain study [39]. Reconstructions were obtained using FBP and maximum likelihood computed by the EM algorithm discussed in Section 2. The maximum number of EM iterations used in the ML reconstructions was 3000 with a uniform initial guess. Raw reconstructions were smoothed by convolution with a spherically symmetric Gaussian kernel. The standard deviation of the Gaussian was adapted to minimize the root mean square (RMS) error between the reconstruction ($\hat{\lambda}$) and the true phantom (λ),

$$\text{RMS(error)} = \sqrt{\frac{\sum_i \left(\lambda_i - \hat{\lambda}_i\right)^2}{I}}.$$

(While this can be done in a simulation setting, a data-dependent method is needed for application in practice. Such a method is proposed in Pawitan and O'Sullivan [38].) In the figures below, the RMS error is scaled by the standard deviation of the phantom [s.d. $= \sqrt{\left(\sum_i \left(\lambda_i - \bar{\lambda}\right)^2\right)/I}$ with $\bar{\lambda}$ the mean of λ]. The adaptation of the amount of smoothness was found to be critical as different reconstruction algorithms require different amounts of smoothing [37].

Postprocessing of FBP. Included in the simulations studies was a postprocessing algorithm proposed by O'Sullivan, Pawitan and Haynor [37] for correcting negativity artifacts in FBP solutions. The approach is quite ad hoc.

The motivation is that negative values in FBP reconstructions are typically surrounded by a number of values where the solution is relatively large and positive. Theoretically, this might be explained by the fact that the ramp-type filter used in FBP (see Figure 2) accentuates high frequencies and consequently tends to induce negative spatial autocorrelation [7]. The postprocessing algorithm iteratively applies a local cancellation rule to distribute the negative artifact at a given pixel over the surrounding neighborhood, while preserving the total number of counts in the reconstructed image. The algorithm is as follows (λ^{ls} is the raw FBP reconstruction and λ^{ls^+} is the postprocessed result):

Initialize:

$\lambda \leftarrow \lambda^{ls}$; $crit \leftarrow \infty$; $tol = .0001$; $maxit = 100$; $iter = 0$; $w \leftarrow 2$

While ($crit > tol$ *and* $iter \leq maxit$){

$iter \leftarrow iter + 1$

Scan the image, λ_i for $i = 1, 2, \ldots I$

if ($\lambda_i < 0$){

a. Find the most positive value (λ_{i^+}) of λ in a w-neighborhood of pixel s

b. Adjustments: $\lambda_i \leftarrow \min(\lambda_i + \lambda_{i^+}, 0)$; $\lambda_{i^+} \leftarrow \max(\lambda_i + \lambda_{i^+}, 0)$}

$$crit = \frac{\max_{i:\lambda_i<0}|\lambda_i|}{\text{mean}_{i:\lambda_i>0}(\lambda_i)}\}$$

$\lambda^{ls^+} \leftarrow \lambda$

End

The neighborhood in step (a) is defined as the set of all pixels which are within a distance w of s. In the case that no positive values are found, the neighborhood may be expanded by a factor of 2 until the condition is met. Typically, we find the magnitude of the maximum negative artifact FBP solution is reduced by at least two orders of magnitude by the algorithm. The computational cost appears to be roughly comparable to an additional one or two back-projection steps. Thus the computation of the postprocessed FBP solution, which we will denote FBP-p, retains much of the efficiency of FBP. Raw FBP-p reconstructions were smoothed using the same approach as that used for the FBP and ML reconstructions. Previously it was found that FBP-p was quite competitive with maximum likelihood on idealized distance–angle tomographs [37]. The simulations here examine how the method might be performed in a time-of-flight setting.

4.2. *Evaluation of performance.* The overall reconstruction was measured using the RMS error as well as with a more elaborate region of interest (ROI) based measurements. For the ROI error, regions S_j for $j = 1, 2, \ldots, J$, were defined on each phantom as follows: white and grey matter for the Hoffman phantom ($J = 2$); the six homogeneous elliptical regions as well as the complement of those structures over the head region for the Shepp–Vardi phantom ($J = 7$); and the six cold spheres in the Jaszczak phantom ($J = 6$).

Note that none of the regions of interest considered include background. The ROI error was defined as an average relative error over the regions in the phantom. Formally, let $\lambda^{(j)}$ and $\hat{\lambda}^{(j)}$ be the average value over region S_j in the true phantom and reconstructed image, respectively. Then we use

$$\text{ROI(error)} = \sqrt{\sum_j \left(\frac{\hat{\lambda}^{(j)} - \lambda^{(j)}}{\lambda^{(j)} + 0.1/I} \right)^2}.$$

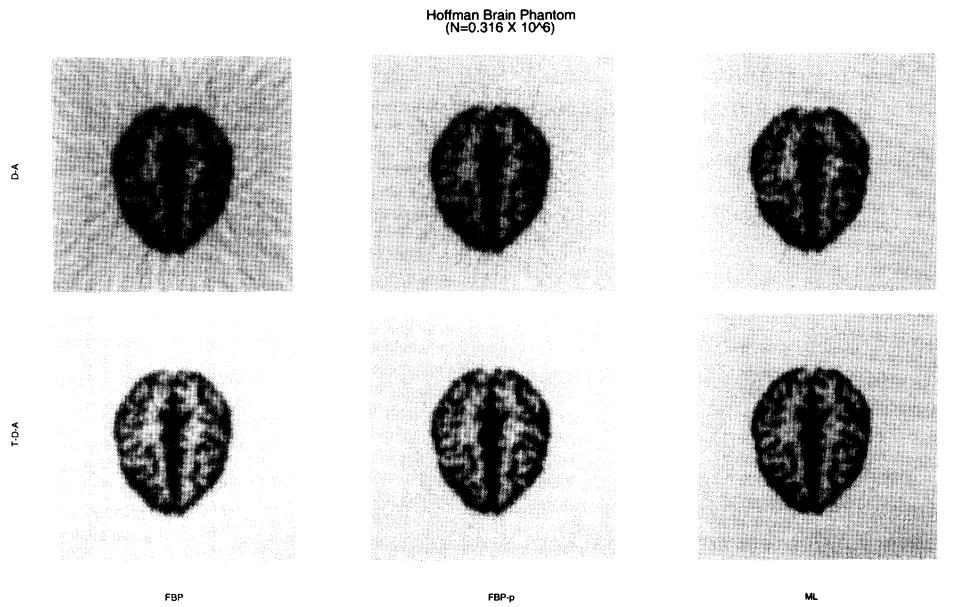
The scaling of the difference between $\lambda^{(j)}$ and $\hat{\lambda}^{(j)}$ is to ensure that each region receives equal weight; the factor $0.1/I$ is particularly important for the Jaszczak because the average intensity of each of the regions of interest is zero there.

4.3. *Results.* Figure 3a–c shows sample reconstructions for each of the phantoms on conventional (distance–angle) and time-of-flight (time–distance–angle) tomographs. Background negative artifact is only apparent with the FBP reconstructions on a conventional tomograph. The effect is quite diminished with postprocessing. Time-of-flight reconstructions are much improved. Here differences between the reconstruction methods are quite difficult to discern, qualitatively.

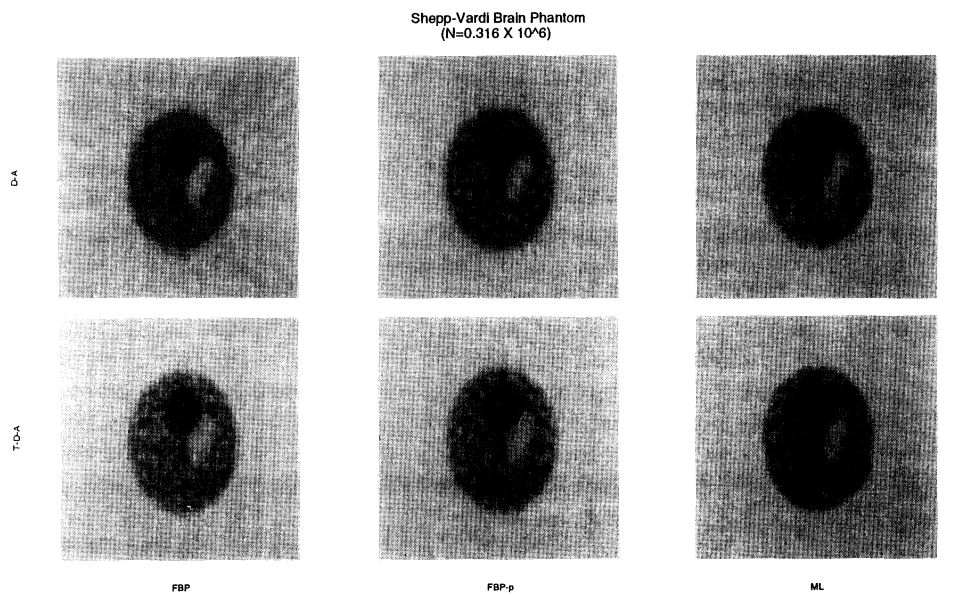
The behavior of the RMS and ROI errors as a function of total expected counts is shown in Figure 4a–c. The RMS errors are more stable than the ROI error. Not surprisingly the ROI errors for the Jaszczak phantom are particularly variable. The differences between FBP and maximum likelihood reconstructions are consistently greater for the RMS error. The reason is that the RMS error includes background while the ROI error does not. For both error assessments, the postprocessed FBP is seen to improve on FBP; indeed, for the Shepp–Vardi and Jaszczak phantoms postprocessed FBP is very close to maximum likelihood at least for moderate counts. The greatest discrepancy between the postprocessed FBP and maximum likelihood occurs at high counts with the Hoffman phantom; here, in the extreme, the ROI error for the postprocessed FBP is 42% greater than the maximum likelihood error. Note that FBP is only an approximation to least squares, and this may be what drives the disparity at high counts. Table 1 summarizes relative increases in the RMS and ROI errors averaged across count rates N. For the RMS error the FBP is 13–20% greater than ML, while the postprocessed FBP is 2 and 5% greater than ML. The corresponding relative increases for the ROI error range from 13 to 26% for FBP and from 3 to 15% for the postprocessed FBP reconstruction.

Apparent rates of estimation for the different reconstruction methods were computed from the data sets plotted in Figure 4a–c. This calculation was based on fitting the model

$$\log(\text{Error}(N)) = \alpha_0 - \alpha_1 \log N + \text{noise},$$

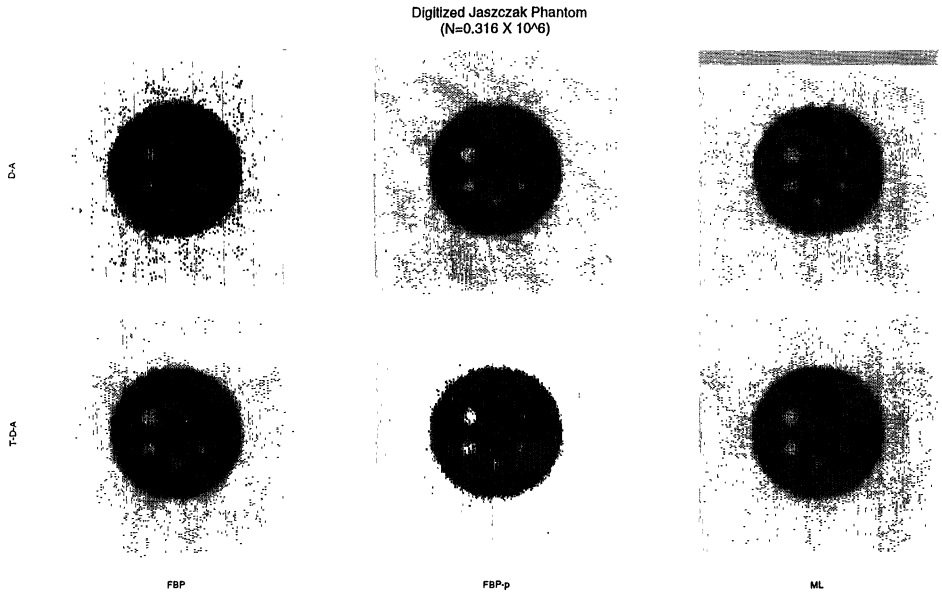


(a)



(b)

FIG. 3. *Sample two-dimensional reconstructions for conventional and time-of-flight tomography with expected counts* $N = 0.316 \times 10^6$; *there are three sets of plots* (a, b, c) *with six images on each*: (a) *Hoffman brain phantom*; (b) *Shepp–Vardi brain phantom*; (c) *Digitized Jaszczak phantom. Note that the color scales on each image range from the maximum to the minimum value for that image. Maximum likelihood* (*ML*) *and modified filtered back-projection* (*FBP-p*) *achieve a dramatic reduction in the background artifact seen in filtered back-projection* (*FBP*).



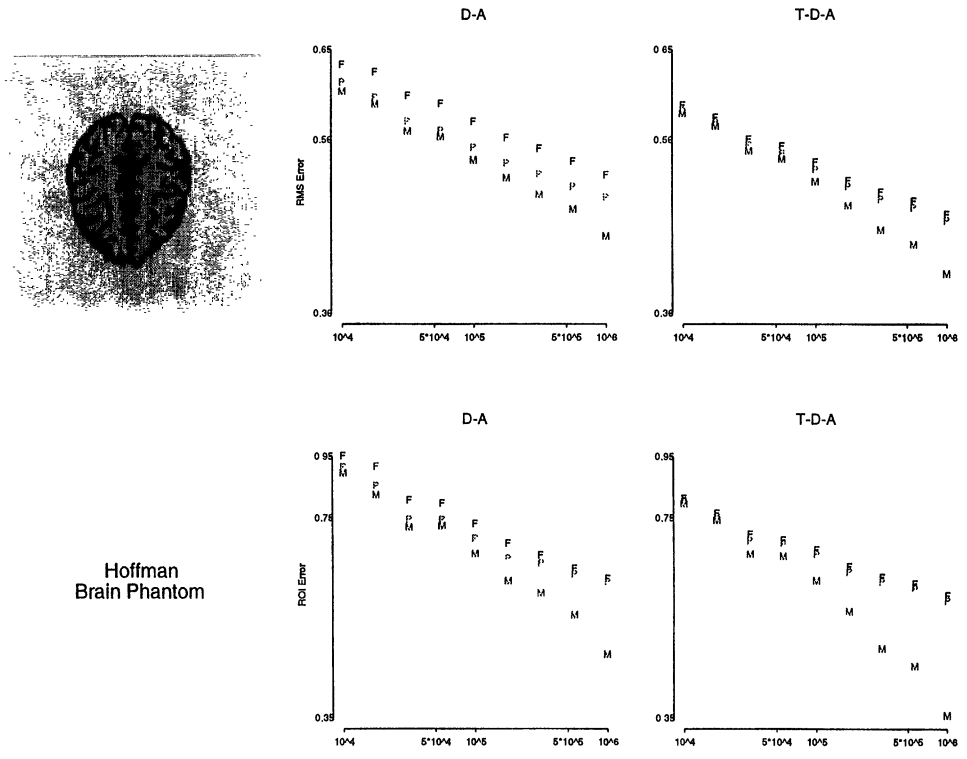
(c)

FIG. 3. *continued.*

where the "Error" is either the RMS or ROI error. For a given phantom, the slope α_1 was constrained to be the same across the different reconstruction methods; the intercepts were not. Since the model implies

$$\text{Error}(N) \approx A_0 N^{-\alpha_1},$$

the estimate of α_1 is the apparent rate of estimation. The results are shown in Table 2 (estimated standard errors are in parentheses). Overall the rates of estimation for the ROI error are seen to be slower than the RMS error. With the RMS error, the rate of estimation for the Hoffman phantom is seen to be slower than for the other phantoms. The theoretical rate of estimation suggested by Theorem 1 as well as by similar results in Zhang [59] would be $\tilde{\alpha}_1 = m[2(m + 1.5)]$, where m is a measure of the smoothness of the target function. It would be satisfying if the theoretical smoothness of the target function were predictive of the empirical rate of estimation observed over the count rate considered. For example, the Shepp–Vardi phantom is piecewise continuous, so one would expect it to lie in a Sobolev space H^m with m arbitrarily close to 0.5 (see, e.g., [33]). Based on this, one would predict a rate of $\tilde{\alpha}_1 = 0.125$ rather than the value of 0.10 which was observed. Some related work on simpler nonparametric regression found the same effect [6]. The reason for the mismatch may have something to do with the fact that the experimental setup is discretized and so the sampled target function could

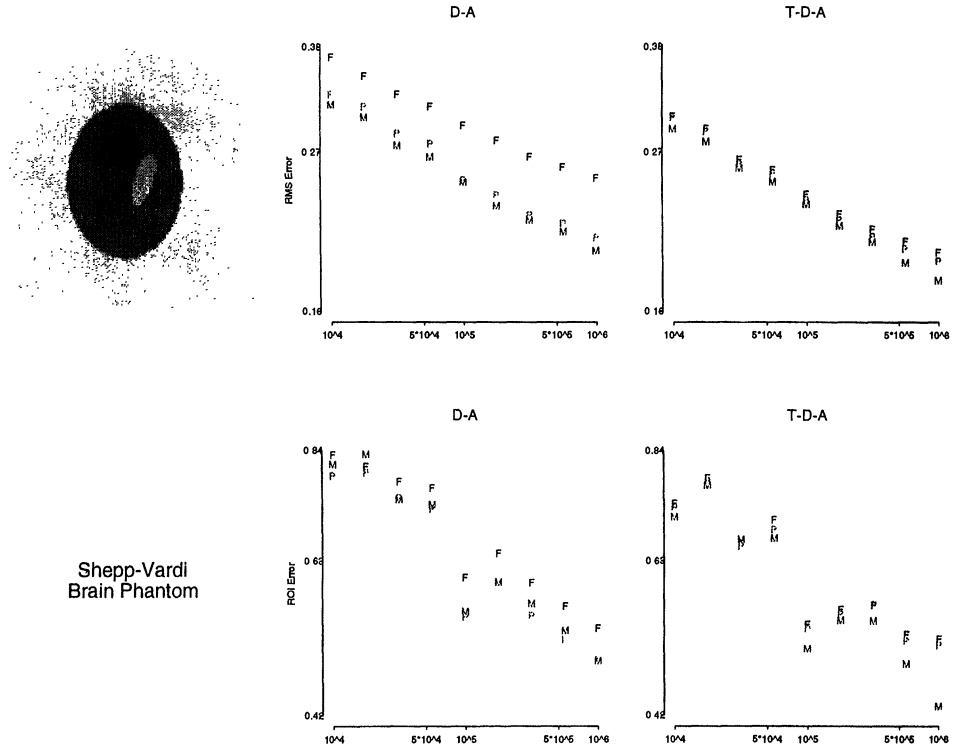


(a)

FIG. 4. *Reconstruction error as a function of expected counts (plotted on a log–log scale): there are three parts to the figure, corresponding to* (a) *Hoffman brain phantom,* (b) *Shepp–Vardi phantom and* (c) *digitized Jaszczak phantom. The top panel on each part shows the RMS error; the bottom panel, the ROI error (see Section* 4). *Results for conventional (DA) and time-of-flight (TDA) appears in the second and third columns, respectively.*

equally well correspond to an analytic or perhaps not even an L_2 function. If the function were analytic, theory would predict an essentially parametric rate of 0.5; on the other extreme, if the function is just L_2, the rate would be arbitrarily close to zero. Thus it would seem very difficult to predict the empirical estimation characteristic in a discretized setting using asymptotic theory. This result may help to stimulate some thought about the practical applicability of asymptotic rate of estimation formulae. Perhaps there would also be an implication for how one might refine the implementation of modern bandwidth selection procedures for density estimation that are derived from such formulae [20].

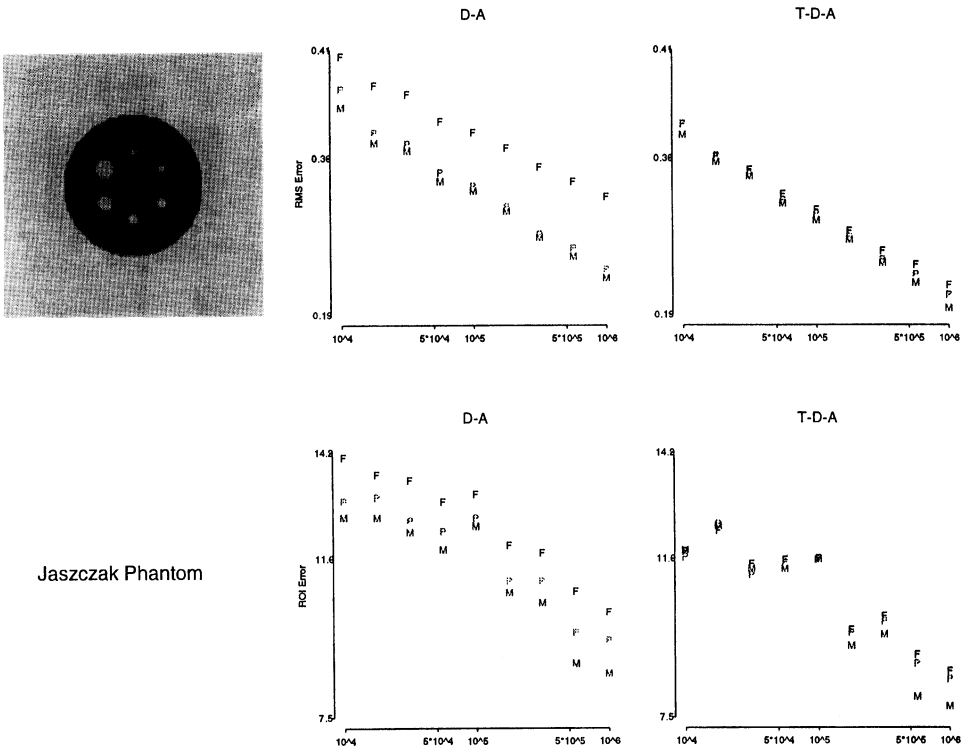
Comparisons between conventional and time-of-flight reconstructions are shown in Figure 5. The ratios of the reconstruction error with time-of-flight data to the corresponding error with conventional distance–angle data are



(b)

FIG. 4. *continued.*

represented in boxplots. Each boxplot has nine values corresponding to the reconstructions done at different counts. For the RMS error, time-of-flight is seen to yield a major improvement with FBP reconstructions. The differences are still appreciable although less dramatic with other reconstruction methods and the ROI error. An asymptotic analysis of least squares estimators by Zhang ([59], see Theorem 5) suggests that in the case that the target source distribution lies in a disc of radius c and the rate of estimation for the mean square error is $2\alpha_1$, the ratio of the mean square reconstruction error for time-of-flight PET to the mean square for conventional PET might be bounded by the constant $[\sigma_t \pi^{3/2}/(2c)]^{2\alpha_1}$, with equality for source distributions putting all their mass on the edge of the disc. In the present circumstance, the upper bound is plotted using $\sigma_t = 3.82$, $c = 13.44$ cm (admittedly liberal) and α_1 taken from Table 2. While there is no reason why asymptotics should be predictive of small-sample performance, it is remarkable that the bound of Zhang [59] has the right order of magnitude. Technically the bound is appropriate for the FBP reconstructions with the RMS error. There is obviously some discrepancy. Part of the problem may be that the bound is based



(c)

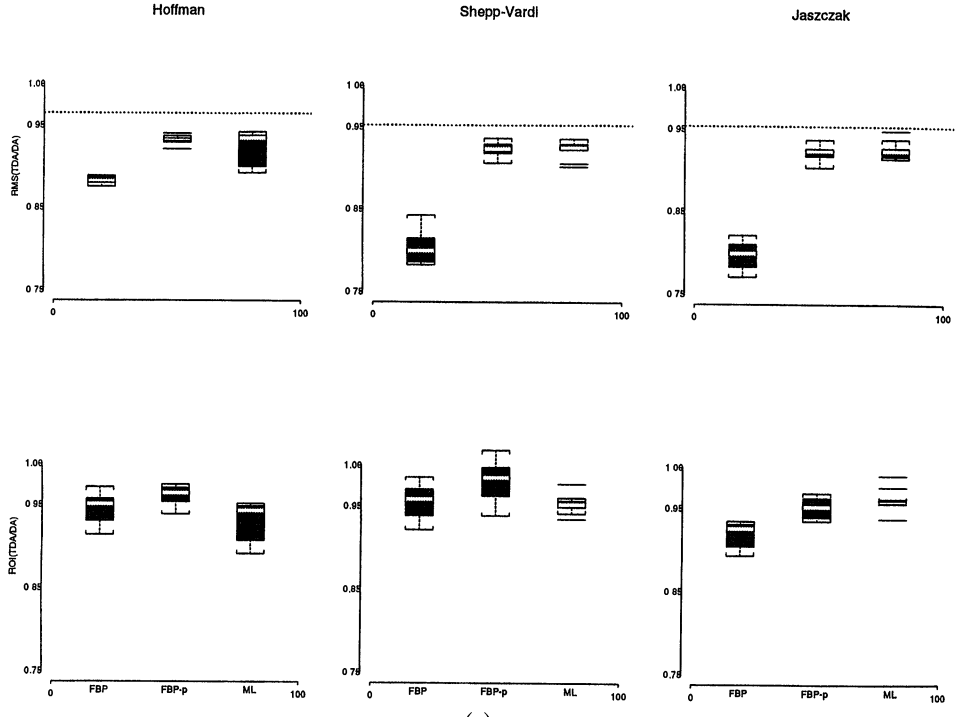
FIG. 4. *continued.*

TABLE 1
Summary of simulation results

Phantom	**Hoffman**	**Shepp–Vardi**	**Jaszczak**
Relative improvement in RMS error: mean (s.d.)			
100 × (FBP-ML)/ML [DA]	12.5 (3.5)	20.4 (4.1)	20.2 (3.5)
100 × (FBP-ML)/ML [TDA]	7.8 (5.9)	4.8 (2.2)	3.4 (1.7)
100 × (FBP-p-ML)/ML [DA]	4.7 (3.2)	3.1 (1.3)	2.4 (1.4)
100 × (FBP-p-ML)/ML [TDA]	5.9 (5.7)	3.1 (1.8)	1.8 (1.1)
Relative improvement in ROI error: mean (s.d.)			
100 × (FBP-ML)/ML [DA]	15.0 (7.8)	5.2 (3.8)	13.2 (3.2)
100 × (FBP-ML)/ML [TDA]	19.5 (18.7)	5.6 (5.9)	3.3 (4.1)
100 × (FBP-p-ML)/ML [DA]	9.7 (9.7)	−1.7 (1.7)	4.6 (2.1)
100 × (FBP-p-ML)/ML [TDA]	17.5 (18.1)	4.3 (5.6)	2.2 (3.5)

TABLE 2
Estimated rates of estimation

Phantom	Hoffman	Shepp–Vardi	Jaszczak
RMS error	0.075 (0.002)	0.103 (0.001)	0.101 (0.001)
ROI error	0.051 (0.002)	0.054 (0.002)	0.041 (0.002)



(a)

FIG. 5. *Comparisons between conventional (DA) and time-of-flight (TDA) reconstructions: the top panel shows the ratio of the RMS error for TDA reconstruction to the RMS error for the DA reconstruction; the bottom panel shows the corresponding ratios for the ROI errors. Columns correspond to the different phantoms. The asymptotic bound suggested by Zhang for the RMS error is shown by the dotted line.*

on consideration of a worst-case source distribution. A further explanation, less likely in my view, is that Zhang's calculation uses Epanechnikov rather than Gaussian smoothing kernels.

Replication of the above experiment did not change the conclusions. It is perhaps noteworthy that the within-count variability in the RMS error was found to be remarkably small, presumably the result of it being a rather smooth functional.

5. Discussion. The reconstruction problem of PET has been examined from a theoretical point of view using analytic and computational tools. The asymptotic analysis indicates that the rate of convergence of maximum likelihood and least squares reconstruction methods are the same. This is largely supported by numerical simulations evaluating the mean square reconstruction errors as a function of count rate. Similar results are found for more complex region of interest based measures of reconstruction accuracy. Time-of-flight significantly reduces reconstruction errors. In practice, however, this result needs to be balanced against the additional complications associated with obtaining time-of-flight data [31]. Maximum likelihood reconstructions were always found to be more accurate than least squares implemented by filtered back-projection. The differences were typically quite modest in the time-of-flight setting. An ad hoc postprocessing of FBP reconstruction to reduce negativity artifacts resulted in reconstructions which were competitive with maximum likelihood. This may be of some practical interest because the computation of maximum likelihood by the EM algorithm requires orders of magnitude more computing time than the postprocessed filtered back-projection reconstruction. Thus some further investigation of the approach might be of interest. In this context alternate methods of postprocessing of FBP, perhaps using statistically based image restoration algorithms [4], could prove useful.

The analysis fails to account for some physical effects, photon scatter, for example, and as a result the relationship between counts in the simulation model and an equivalent count rate for an actual tomograph requires careful validation. In practice, detector efficiencies (and even attenuation) can be highly variable and sometimes close to zero leading to severe streak artifacts in FBP reconstructions [15]. Qualitatively the reconstructions shown in Figure 3 are substantially more resolved than what would be produced on an operational tomograph with the same number of counts. Thus the results here would benefit from some calibration with actual tomograph phantom studies.

The extent to which the results might carry over to other situations in which positivity constrained reconstruction is of interest [53] is not clear. For a simple parametrized family of one-dimensional Poisson deconvolution problems, numerical experiments in O'Sullivan, Pawitan and Haynor [37] found the differences between least squares and maximum likelihood reconstruction to be remarkably small, after correcting for the different amounts of smoothing required by least squares and maximum likelihood reconstructions. I doubt that this is true in general, although it does seem to be true for PET.

APPENDIX

PROOF OF THEOREM 1. The proof uses analysis tools developed by Cox and O'Sullivan [12]. The approach is based on Taylor series expansions for the

regularization functional, very much in the style of the Cramér approach to the analysis of M-estimators developed by Huber [27]. Theorem 1 follows from linearization results in Theorems 2 and 3. To apply the theory of [12], a set of assumptions ([12], A.1–A.6) must be verified. The first four assumptions develop properties of some Hilbert spaces related to the regularization functionals. The finite positive constant M is used generically in the following. Note that successive appearances in a proof typically correspond to larger values.

Assumptions A.1–A.4 *of* [12]. Let U and W be operators defined on Θ by

$$U\theta = \sum_\nu \theta_\nu \left[\frac{1}{1+\gamma_\nu}\right]\phi_\nu, \qquad W\theta = \sum_\nu \theta_\nu \left[\frac{\gamma_\nu}{1+\gamma_\nu}\right]\phi_\nu.$$

These extend to bounded linear operators on Θ_b for $b \geq 0$. In fact U is a bounded linear operator from Θ_b into Θ_{b+2}. Also $\langle \theta, U\theta \rangle = (R\theta, R\theta)$ and $P(\theta) = \langle \theta, W\theta \rangle$ and Assumptions A.1 and A.2 of [12] follow. Note the sequences $\{\phi_\nu\colon\ \nu = 1,2,\ldots\}$ of eigenfunctions and corresponding eigenvalues $\{\gamma_\nu\colon\ \nu = 1,2,\ldots\}$ may equivalently be expressed as the spectral decomposition of W relative to U, that is,

$$\langle \phi_\nu, U\phi_\mu \rangle = \delta_{\nu\mu}, \qquad \langle \phi_\nu, W\phi_\mu \rangle = \gamma_\nu \delta_{\nu\mu}. \tag{8}$$

Thus, as in [12] the norm and inner product on Θ_b can be expressed as

$$\|\theta\|_b = \langle \theta, \theta \rangle^{1/2}, \qquad \langle \theta, \zeta \rangle_b = \sum_\nu \left[1 + \gamma_\nu^b\right] \langle \theta, U\phi_\nu \rangle \langle \zeta, U\phi_\nu \rangle.$$

Useful properties of the norms $\|\cdot\|_b$ are given in [12, Lemma 2.2].

Next it is necessary to consider derivatives (score functions) of the regularization functionals and their limiting forms. These are defined, using Assumption 2(i) and (iii), as the expected value of regularization functionals, that is,

$$l_\tau(\theta) = \begin{cases} \dfrac{1}{2}\displaystyle\int_Z R\theta[z]^2\, dz - \int_Z R\theta[z] R\theta_0[z] + \frac{\tau}{2}P(\theta), & \text{LS}, \\ \displaystyle\int_Z c(z) R\theta[z]\, dz & \\ \quad - \displaystyle\int_Z \log(c(z) R\theta[z] + g^*(z)) g(z)\, dz + \frac{\tau}{2}P(\theta), & \text{ML}, \end{cases}$$

where, from Assumption 2(iii), $g(z) = c(z)R\theta_0(z) + g^*(z)$. The symbols LS and ML refer to least squares and maximum likelihood, respectively. Let $l(\theta) = l_0(\theta)$ and $l_n(\theta) = l_{n0}(\theta)$. By Assumption 2(iii) and the definition of U, it is clear that in the least squares situation θ_0 is the unique global

minimizer of $l(\theta)$ in Θ_α. For the maximum likelihood case, let $N_{\theta_0} = \{\theta \in \Theta_\alpha: \|\theta - \theta_0\|_\alpha \le \varepsilon\}$, where ε is chosen so that $c(z)R\theta(z) + g^*(z)$ is bounded away from zero (uniformly in z). This is possible by Assumption 2(ii) and (iii) using Sobolev embedding. It follows that θ_0 is the unique minimizer of $l(\theta)$ in N_{θ_0}.

From Assumption 2(ii), the third order Frechet derivatives [41] of l and l_n exist and are continuous. Letting D denote differentiation, the derivatives of $l_n(\theta)$ are given by the following:

$$Dl_n(\theta)u = \begin{cases} \int_Z R\theta[z]Ru[z]\,dz - \dfrac{1}{n}\sum_i \dfrac{y_i Ru[z_i]}{c(z_i)} \\ \quad + \int_Z \dfrac{Ru[z]g^*(z)}{c(z)}\,dz, \\ \int_Z c(z)Ru[z]\,dz - \dfrac{1}{n}\sum_i \dfrac{y_i c(z_i)Ru[z_i]}{c(z_i)R\theta[z_i] + g^*(z_i)}; \end{cases}$$

(9)

$$D^2 l_n(\theta)uv = \begin{cases} \int_Z Rv[z]Ru[z]\,dz, \\ \dfrac{1}{n}\sum_i \dfrac{y_i c(z_i)^2 Ru[z_i]Rv[z_i]}{(c(z_i)R\theta[z_i] + g^*(z_i))^2}; \end{cases}$$

$$D^3 l_n(\theta)uvw = \begin{cases} 0, \\ -\dfrac{2}{n}\sum_i \dfrac{y_i c(z_i)^3 Ru[z_i]Rv[z_i]Rw[z_i]}{(c(z_i)R\theta[z_i] + g^*(z_i))^3}. \end{cases}$$

Here the first row in each expression corresponds to least squares and the second to maximum likelihood. Corresponding expressions for the derivatives of $l(\theta)$ are readily obtained. Assumption A.3 of [12] follows; $D^2 l(\theta)$ defines a bounded linear operator $U(\theta)$ on Θ by $D^2 l(\theta)uv = \langle u, U(\theta), v\rangle$. In the least squares case $U(\theta) = U$. For maximum likelihood,

$$\langle \phi, U(\theta)\phi\rangle = \int_Z R\phi[z]^2 \frac{(c(z)R\theta_0[z] + g^*(z))}{(c(z)R\theta[z] + g^*(z))} c(z)^2\,dz.$$

From Assumption 2(iii), the definition of N_{θ_0} and Sobolev's embedding, for $\theta \in N_{\theta_0}$ there are finite strictly positive constants M_1 and M_2 for which

$$M_1\langle \phi, U\phi\rangle \le \langle \phi, U(\theta)\phi\rangle \le M_2\langle \phi, U\phi\rangle.$$

This proves Assumption A.4 of [12]. The spectral decomposition of W relative $U(\theta^*)$ for $\theta^* \in N_{\theta_0}$ gives rise to sequences of eigenfunctions $\{\phi_{*\nu}: \nu = 1, 2, \ldots\}$

and eigenvalues $\{\gamma_{*\nu}: \nu = 1, 2, \ldots\}$. These may be used to define norms $\|\cdot\|_{*b}$ and associated Hilbert spaces Θ_{*b} by analogy with equation (8); $\Theta_{*b} = \Theta_b$ as sets and they have equivalent norms (see [12], Proposition 2.1).

Linear approximations. Approximations to the systematic (bias) and stochastic (variance) components of the estimation error are developed in terms of first-order Taylor series expansions for the limiting and sample score vectors $Dl_\tau(\theta) = Z_\tau(\theta)$ and $Dl_{n\tau}(\theta) = Z_{n\tau}(\theta)$, respectively. The existence of these score vectors as elements of Θ_b for $b \geq \alpha$ is developed in [12]. For $\theta \in N_{\theta_0}$, let

$$G_\tau(\theta) = U(\theta) + \tau W;$$

$G_\tau(\theta)$ is a bounded linear operator on Θ_b for $b \in [0, 1]$ (see [12], Section 2).

The *bias* is approximated in terms of $\bar{\theta}_\tau - \theta_0$, where

$$\bar{\theta}_\tau - \theta_0 = -G_\tau(\theta_0)^{-1} Z_\tau(\theta_0).$$

The *variance* is approximated using $\bar{\theta}_{n\tau} - \theta_\tau$, where

$$\bar{\theta}_{n\tau} - \theta_\tau = -G_\tau(\theta_\tau)^{-1} Z_{n\tau}(\theta_\tau)$$

and $\theta_\tau \in N_{\theta_0}$. Actually it is required that $Z_\tau(\theta_\tau) = 0$, but this will be guaranteed by Theorem 2, at least for all τ sufficiently small. The asymptotic behavior of the linearizations is as follows.

LEMMA 1. *There exists a $\tau_0 > 0$ such that, for $0 \leq b \leq \alpha$ with α satisfying* (4) *and some constant M, the following hold for all $\tau \in (0, \tau_0]$:*

(i) $d(\tau, b)^2 = \|\bar{\theta}_\tau - \theta_0\|_b^2 \leq M\tau^{s-b}\|\theta_0\|_s^2$;

(ii) $d_n(\tau, b)^2 = \|\bar{\theta}_{n\tau} - \theta_\tau\|_b^2 = O_p(n^{-1}\tau^{-(b+1/r)})$.

PROOF. For part (i), let $\theta^* = \theta_0$,

$$\begin{aligned}
\|\bar{\theta}_\tau - \theta_0\|_b^2 &= \|G_\tau(\theta_0)^{-1} Z_\tau(\theta_0)\|_b^2 \\
&= \|G_\tau(\theta_0)^{-1}(\tau W\theta_0)\|_b^2 \\
&\leq M\|G_\tau(\theta_0)^{-1}(\tau W\theta_0)\|^2{}_{*b} \\
&= M\sum_\nu (1 + \gamma^b{}_{*\nu})\langle G_\tau(\theta_0)^{-1}\tau W\theta_0, U(\theta_0)\phi_{0\nu}\rangle^2 \\
&= M\sum_\nu (1 + \gamma^b{}_{*\nu})(1 + \tau\gamma_{*\nu})^{-2}\langle \tau W\theta_0, \phi_{0\nu}\rangle^2 \\
&\leq M\sum_\nu (1 + \nu^r)^b \left(\frac{\tau\nu^r}{1 + \tau\nu^r}\right)^2 \langle \theta_0, U(\theta_0)\phi_{*\nu}\rangle^2.
\end{aligned}$$

Here the equivalence of the $\|\cdot\|_{*b}$- and the $\|\cdot\|_b$-norms is used in the third inequality, the series expansion of $\|\cdot\|_{*b}$ for the fourth and Lemma 2.2 of [12] for the fifth. The fact that $\theta_0 = \sum_\nu \langle \theta_0, U(\theta_0)\phi_{*\nu}\rangle \phi_{*\nu}$, the defining relations for the $\phi_{*\nu}$'s and the growth rate of $\gamma_{*\nu}$ is used in the sixth inequality. Now the claim follows from the argument of Theorem 2.3 of Cox [11]; see also Lemma 5.4 of Cox [10].

Part (ii) follows by Markov's inequality. For $\theta^* = \theta_\tau$, a reduction similar to that used in part (i) gives

$$\begin{aligned}\|\bar{\theta}_{n\tau} - \theta_\tau\|_b^2 &\le M\sum_\nu (1 + \gamma^b{}_{*\nu})(1 + \tau\gamma_{*\nu})^{-2} \\ &\quad \times \{Dl_n(\theta_\tau)\phi_{*\nu} - Dl(\theta_\tau\phi_{*\nu})\}^2.\end{aligned} \tag{10}$$

Using Assumption 2(i) and (iii),

$$\begin{aligned}\mathbb{E}\{Dl_n(\theta_\tau)\phi - Dl(\theta_\tau)\phi\}^2 &\le \begin{cases} Mn^{-1}\operatorname{Var}\left\{\dfrac{yR\phi[z]}{c(z)}\right\}, & \text{LS} \\ Mn^{-1}\operatorname{Var}\left\{\dfrac{yc(z)R\phi[z]}{c(z)R\theta_\tau[z] + g^*(z)}\right\}, & \text{ML} \end{cases} \\ &\le Mn^{-1}\mathbb{E}\{yR\phi(z)\}^2 \\ &\le Mn^{-1}\cdot\int_Z (R\phi[z])^2\,dz \le Mn^{-1}\|\phi\|^2{}_{*0}.\end{aligned}$$

Since $\|\phi_{*\nu}\|^2{}_{*0} = 1$, substituting into equation (10) and using Lemma 2.2 of [12] gives

$$\mathbb{E}\left[d_n(\tau, b)^2\right] \le Mn^{-1}\tau^{-(b+1/r)}. \qquad \square$$

Bounds on derivatives. In order to show that the linearizations can be used to approximate the behavior of the regularization estimators, some derivative operators related to the regularization functional are analyzed (cf. the analysis of M-estimators in Huber [27]). For $0 \le b \le \alpha$, $\tau > 0$, $\theta_1, \theta_2 \in N_{\theta_0}$ and $u, v \in \Theta_\alpha$ with $\|u\|_\alpha = \|v\|_\alpha = 1$, set

$$\begin{aligned}K_{2n}(\tau, b) &= \sup_{\theta_1,\theta_2}\sup_u \left\|G_\tau(\theta_1)^{-1}\left[D^2 l_n(\theta_2)u - D^2 l(\theta_2)u\right]\right\|_b, \\ K_3(\tau, b) &= \sup_{\theta_1,\theta_2}\sup_{u,v} \left\|G_\tau(\theta_1)^{-1}\left[D^3 l(\theta_2)uv\right]\right\|_b, \\ K_{3n}(\tau, b) &= \sup_{\theta_1,\theta_2}\sup_{u,v} \left\|G_\tau(\theta_1)^{-1}\left[D^3 l_n(\theta_2)uv\right]\right\|_b.\end{aligned}$$

LEMMA 2. *For α satisfying* (4) *there is a constant $0 < M < \infty$ and a random variable $A_n = O_p(1)$ such that, for all $\theta \in N_{\theta_0}$ and all $u, v, w \in \Theta_\alpha$, the following hold*:

(i) $(D^3 l(\theta)uvw)^2 \le M\|u\|_\alpha^2 \cdot \|v\|_\alpha^2 \cdot \|w\|_0^2$;
(ii) $(D^2 l_n(\theta)uv - D^2 l(\theta)uv)^2 \le A_n M n^{-1}\|u\|_\alpha^2 \cdot \|v\|_\alpha^2$;
(iii) $(D^3 l_n(\theta)uvw)^2 \le A_n M\|u\|_\alpha^2 \cdot \|v\|_\alpha^2 \cdot (\|w\|_0^2 + n^{-1}\|w\|_\alpha^2)$.

PROOF. In the least squares case, each of the terms on the left-hand side of the inequalities is identically zero. For the maximum likelihood case, part (i) follows by application of the Cauchy–Schwarz inequality together with Assumption 2(ii) and (iii), noting $s_0 < \alpha$.

The analysis for part (ii) is more elaborate:

$$\{D^2 l_n(\theta)uv - D^2 l(\theta)uv\}^2 = \left\{\int_Z \frac{c(z)^2 Ru[z]Rv[z]}{(c(z)R\theta[z] + g^*(z))^2}[P_z^{(n)} - P_z](dz)\right\}^2,$$

where $P_z^{(n)}$ is the y-weighted empirical measure of the z_i's, $dP_z^{(n)} = (1/n)\Sigma_i y_i \delta_{z_i}$, with δ_{z_i} Dirac's delta function; P_z is the limiting form of $P_z^{(n)}$; from Assumption 2(i) and (iii), $dP_z = [c(z)R\theta_0(z) + g^*(z)]\,dz$.

Let h be a function in $\overline{H}^\delta(Z)$. Using the Fourier transform $\hat{h}$ of h and the Cauchy–Schwarz inequality,

$$\begin{aligned}
&\left\{\int_Z h(z)[P_z^{(n)} - P_z](dz)\right\}^2 \\
&\quad = \left\{\int_Z\int_{\mathbb{R}^3} \hat{h}(\nu)\exp(\iota\nu\cdot z)[P_z^{(n)} - P_z](dz)\,d\nu\right\}^2 \\
&\quad \le \left\{\int_{\mathbb{R}^3}[1 + |\nu|^2]^\delta |\hat{h}(\nu)|^2\,d\nu\right\} \\
&\qquad \times\left\{\int_{\mathbb{R}^3}[1 + |\nu|^2]^{-\delta}\left|\int_Z \exp(\iota\nu\cdot z)[P_z^{(n)} - P_z](dz)\right|^2\right\} \\
&\quad = \|h\|_{\overline{H}^\delta(Z)}^2\left\{\int_{\mathbb{R}^3}[1 + |\nu|^2]^{-\delta}\left|\int_Z \exp(\iota\nu\cdot z)[P_z^{(n)} - P_z](dz)\right|^2\right\},
\end{aligned} \tag{11}$$

where $\iota = \sqrt{-1}$. However, using the argument in Lemma 1(ii),

$$\begin{aligned}
\mathbb{E}\left|\int_Z \exp(\iota\nu\cdot z)[P_z^{(n)} - P_z](dz)\right|^2 &\le Mn^{-1}\mathbb{E}|y\exp(\iota\nu\cdot z)|^2 \\
&\le Mn^{-1}\mathbb{E}|y^2| \le Mn^{-1}.
\end{aligned}$$

However, $\int_{\mathbb{R}^3}[1+|\nu|^2]^{-\delta}\,d\nu$ is finite for $\delta > d/2$; thus

$$\mathbb{E}\left\{\int_{\mathbb{R}^3}\left[1+|\nu|^2\right]^{-\delta}\left|\int_Z \exp(\iota\nu\cdot z)\left[P_z^{(n)}-P_z\right](dz)\right|^2\right\} \le Mn^{-1}.$$

Substituting into (11) gives

$$\left\{\int_Z h(z)\left[P_z^{(n)}-P_z\right](dz)\right\}^2 \le Mn^{-1}A_n\|h\|^2_{\bar{H}^\delta(Z)},$$

with $A_n = O_p(1)$, independent of h. Since $\delta > d/2$ from Lemma 4 of Cox and O'Sullivan [13], with $h(z) = c(z)^2 Ru[z]Rv[z]/(c(z)R\theta[z] + g^*(z))^2$,

$$\begin{aligned}\|h\|^2_{\bar{H}^\delta(Z)} &\le M\|c\|^2_{\bar{H}^\delta(Z)}\|g^*\|^2_{\bar{H}^\delta(Z)}\|R\theta\|^2_{\bar{H}^\delta(Z)}\|Ru\|^2_{\bar{H}^\delta(Z)}\|Rv\|^2_{\bar{H}^\delta(Z)}\\ &\le M\|\theta\|^2_{s_0}\|u\|^2_{s_0}\|v\|^2_{s_0}.\end{aligned} \tag{12}$$

The result follows since $\alpha > s_0$ and $\theta \in \Theta_\alpha$.

Part (iii) is proved by considering

$$\left|D^3 l_n(\theta)uvw\right| \le \left|D^3 l(\theta)uvw\right| + \left|D^3 l_n(\theta)uvw - D^3 l(\theta)uvw\right|. \tag{13}$$

Part (i) is used to bound the first term, and the second term is handled as in part (ii). □

From the above the following estimates are obtained.

LEMMA 3. *There is a constant* $0 < M < \infty$ *and a random variable* $A_n = O_p(1)$ *such that the following hold for* $b \le 2 - \alpha - 1/r$, *where* α *satisfies* (4):

(i) $K_3(\tau, b)^2 \le M\tau^{-(b+1/r)}$;
(ii) $K_{2n}(\tau, b)^2 \le A_n Mn^{-1}\tau^{-(\alpha+b+1/r)}$;
(iii) $K_{3n}(\tau, b)^2 \le A_n M\tau^{-(b+1/r)}\{1 + n^{-1}\tau^{-\alpha}\}$.

Both M *and* A_n *are independent of* τ *and* b.

PROOF. The result is established by an analysis almost identical to that used in Lemma 2.5 of O'Sullivan [35]. The details are straightforward. □

Linearization theorems. The main theorems follow after verifying Assumptions A.5 and A.6 in [12].

THEOREM 2. *Let* α *satisfy equation* (4). *There exists* $\tau_0 > 0$ *such that, for all* $\tau \in [0, \tau_0]$, *there is a unique* $\theta_\tau \in N_{\theta_0}$ *with* $Z_\tau(\theta_\tau) = 0$ *satisfying* $\|\theta_\tau - \theta_0\|_\alpha \le 2d(\tau, \alpha)$. *Furthermore, for* $0 \le b \le \alpha$ *and some constant* M, $\|\theta_\tau - \theta_0\|_b \le M\tau^{(s-b)/2}\|\theta_0\|_s$.

PROOF. We check Assumption A.5 of [12], which requires $d(\tau, \alpha) \to 0$ and $r(\tau, \alpha) \to 0$, where

$$r(\tau, b) = K_3(\tau, b)d(\tau, \alpha).$$

However, from Lemmas 2(i) and 3(i), for some constant M, $d(\tau, \alpha)^2 \leq M\tau^{s-\alpha}\|\theta_0\|_s^2$ and $r(\tau, b)^2 \leq M\tau^{-(b+1/r)}\tau^{s-\alpha} \leq M\tau^{s-1/r-2\alpha}$. Since $2\alpha < s - 1/r$, Assumption A.5 of [12] holds, and the result follows from Theorem 3.1 of [12]. □

THEOREM 3. *Suppose* $\{\tau_n\} \subset (0, \tau_0]$ (*with* τ_0 *given in Theorem* 2) *is a sequence such that, for some* α *satisfying equation* (4),

$$n^{-1}\tau_n^{-2(\alpha+1/r)} \to 0. \tag{14}$$

Then there is a constant M *such that the following hold with probability arbitrarily close to* 1 *for all* n *sufficiently large*:

(i) *there is a unique root* $\theta_{n\tau_n} \in \Theta_\alpha$ *of* $Z_{n\tau_n}(\theta) = 0$ *with* $\|\theta_{n\tau_n} - \theta_{\tau_n}\|_\alpha \leq 2d_n(\tau_n, \alpha)$;

(ii) *for* $b \in [0, \alpha]$,

$$\|\theta_{n\tau_n} - \bar{\theta}_{n\tau_n}\|_b \leq Mn^{-1}\tau_n^{-(2\alpha+b+3/r)/2}$$

and

$$\|\theta_{n\tau_n} - \theta_{\tau_n}\|_b \leq Mn^{-1/2}\tau_n^{-(b+1/r)/2}.$$

PROOF. Assumption A.6 of [12] requires $d_n(\tau_n, \alpha) \to_P 0$ and $r_n(\tau_n, \alpha) \to_P 0$, where

$$r_n(\tau_n, b) = K_{2n}(\tau_n, b) + d_n(\tau_n, \alpha)K_{3n}(\tau_n, b).$$

From Lemmas 2(ii) and 3(ii),

$$\begin{aligned} r_n(\tau_n, b) &\leq O_p\big(n^{-1/2}\tau_n^{-(b+\alpha+1/r)/2} + n^{-1/2}\tau_n^{-(b+\alpha+2/r)/2} \\ &\qquad + n^{-1}\tau_n^{-(b+\alpha+\alpha+2/r)/2}\big) \\ &\leq O_p\big(n^{-1/2}\tau_n^{-(2\alpha+1/r)/2} + n^{-1/2}\tau_n^{-(2\alpha+2/r)/2} + n^{-1}\tau_n^{-(3\alpha+2/r)/2}\big). \end{aligned}$$

Thus if τ_n is a sequence such that equation (14) holds then $r_n(\tau_n, \alpha) \to_P 0$. From here, Theorem 3.2 of [12] gives the desired result. □

REFERENCES

[1] ADAMS, R. (1975). *Sobolev Spaces*. Academic Press, New York.

[2] BENDRIEM, B., SOUSSALINE, R., CAMPAGNOLO, R., VERREY, B., WAJNBERG, P. and SYROTA, A. (1986). A technique for correction of scattered radiation in a PET system using TOF information. *Journal of Computer Assisted Tomography* **10** 287–295.

[3] BERGSTROM, M., ERIKSSON, L., BOHM, C., BLOMQVIST, G. and LITTON, J. (1983). Correction for scattered radiation in a ring detector positron camera by integral transformation of the projections. *Journal of Computer Assisted Tomography* **7** 42–50.

[4] BESAG, J. (1986). On the statistical analysis of dirty pictures (with discussion). *J. Roy. Statist. Soc. Ser. B* **48** 259–302.

[5] BICKEL, P. J. and RITOV, Y. (1995). Estimating linear functionals of a PET image. *IEEE Transactions on Medical Imaging* **14** 81–87.

[6] BOSE, S. and O'SULLIVAN, F. (1994). A region based image segmentation method for multi-channel data. Technical Report 280, Dept. Statistics, Univ. Washington (available as the file segment.ps by anonymous ftp from stat.washington.edu).

[7] BRILLINGER, D. (1981). *Time Series: Data Analysis and Theory*. Holden Day, San Francisco.

[8] CARROLL, R. J. and HALL, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83** 1184–1186.

[9] CARSON, R. E., YAN, Y., DAUBE-WITHERSPOON, M. E., FREEDMAN, N., BACHARACH, S. L. and HERSCOVITCH, P. (1993). An approximation formula for PET region-of-interest values. *IEEE Transactions on Medical Imaging* **12** 240–251.

[10] COX, D. D. (1984). Multivariate smoothing spline functions. *SIAM J. Numer. Anal.* **21** 789–813.

[11] COX, D. D. (1988). Approximation of method of regularization estimators. *Ann. Statist.* **16** 694–712.

[12] COX, D. D. and O'SULLIVAN, F. (1990). Asymptotic analysis of penalized likelihood and related estimators. *Ann. Statist.* **18** 1676–1695.

[13] COX, D. D. and O'SULLIVAN, F. (1993). Penalized likelihood-type estimators for generalized nonparametric regression. Unpublished manuscript.

[14] EGGERMONT, P. P. B. and LARICCIA, V. N. (1995). Maximum smoothed likelihood density estimation for inverse problems. *Ann. Statist.* **23** 199–220.

[15] FESSLER, J. A., CLINTHORNE, N. H. and ROGERS, W. L. (1992). Regularized emission image reconstruction using imperfect side information. *IEEE Transactions on Nuclear Science* **39** 1464–1471.

[16] GEMAN, S., MANBECK, K. M. and MCCLURE, D. E. (1991). Design and implementation of a statistical method for emission computed tomography. Reports in Pattern Analysis 154, Div. Applied Mathematics, Brown Univ.

[17] GEMAN, S. and MCCLURE, D. E. (1985). Bayesian image analysis: An application to single photon emission tomography. In *Proceedings of the Statistical Computing Section* 12–18. Amer. Statist. Assoc., Alexandria, VA.

[18] GILLAND, D. R., TSUI, B. M., METZ, C. E., JASZCZAK, R. J. and PERRY, J. R. (1992). An evaluation of maximum likelihood-expectation maximization reconstruction for SPECT by ROC analysis. *Journal of Nuclear Medicine* **33** 451–457.

[19] GREEN, P. J. (1990). Bayesian reconstruction from emission tomography data using a modified EM algorithm. *IEEE Transactions on Medical Imaging* **9** 84–93.

[20] HALL, P., SHEATHER, S. J., JONES, M. C. and MARRON, J. S. (1991). On optimal data-based bandwidth selection in kernel density estimation. *Biometrika* **78** 263–269.

[21] HAYNOR, D. R., HARRISON, R. L. and LEWELLEN, T. K. (1988). A scheme for accidental coincidence correction in time-of-flight positron tomography: theory and implementation. *IEEE Transactions on Nuclear Science* **NS-35** 753–756.

[22] HAYNOR, D. R. and WOODS, S. D. (1989). Resampling estimates of precision in emission tomography. *IEEE Transactions on Medical Imaging* **MI-8** 337–343.

[23] HERTLE, A. (1983). Continuity of the Radon transform and its inverse on Euclidean space. *Math. Z.* **184** 165–192.

[24] HINSHAW, W. S. and LENT, A. H. (1983). An introduction to NMR imaging: form the Block equation to the imaging equation. *Proc. IEEE* **71** 338–350.

[25] HOFFMAN, E. J., CUTLER, P. D., DIGBY, W. M. and MAZZIOTTA, J. C. (1990). 3-D phantom to simulate cerebral blood flow and metabolic images for PET. *IEEE Transactions on Nuclear Science* **NS-37** 616–620.

[26] HOFFMAN, B. D., FICKE, D. C., HOLMES, T. J., POLITTE, D. and TER-POGOSSIAN, M. M. (1986). Image reconstruction of data from SUPER PETT 1: a first generation time-of-flight positron emission tomograph. *IEEE Transactions on Nuclear Science* **33** 428–434.

[27] HUBER, P. (1981). *Robust Statistics*. Wiley, New York.

[28] HUDSON, H. M., MA, J. and GREEN, P. J. (1994). Fisher's method of scoring in statistical image reconstruction: comparison of Jacobi and Gauss–Seidel iterative schemes. *Statistical Methods in Medical Research* **3** 87–101.

[29] JOHNSTONE, I. and SILVERMAN, B. W. (1990). Speed of estimation of positron emission tomography. *Ann. Statist.* **18** 616–629.

[30] LEVITAN, E. and HERMAN, G. T. (1987). A maximum a-posteriori probability expectation maximization algorithm for image reconstruction in emission tomography. *IEEE Transactions on Medical Imaging* **6** 185–192.

[31] LEWELLEN, T. K., BICE, A. B., HARRISON, R. L., PENCKE, M. D. and LINK, J. M. (1988).

Performance measurements of the SP3000/UW time-of-flight positron emission tomograph. *IEEE Transactions on Nuclear Science* **NS-35** 665–669.

[32] MAZZIOTTA, J. C., PHELPS, M. E. and SCHELBERT, H. R. (1986). *Positron Emission Tomography and Autoradiography*. Raven, New York.

[33] NATTERER, F. (1986). *Mathematics of Computerized Tomography*. Wiley, New York.

[34] NYCHKA, D. and COX, D. D. (1989). Convergence rates for regularized solutions of integral equations from discrete noisy data. *Ann. Statist.* **17** 556–572.

[35] O'SULLIVAN, F. (1993). Nonparametric estimation in the Cox model. *Ann. Statist.* **21** 124–145.

[36] O'SULLIVAN, F. (1993). Imaging radiotracer model parameters in PET: A mixture analysis approach. *IEEE Transactions on Medical Imaging* **12** 399–412.

[37] O'SULLIVAN, F., PAWITAN, Y. and HAYNOR, D. R. (1993). Reducing negative artifacts in emission tomography reconstructions: post-processing filtered backprojection solutions. *IEEE Transactions on Medical Imaging* **12** 653–663.

[38] PAWITAN, Y. and O'SULLIVAN, F. (1993). Data dependent bandwidth selection for emission computed tomography reconstruction. *IEEE Transactions on Medical Imaging* **12** 167–172.

[39] PHELPS, M. E., HUANG, S. C., HOFFMAN, E. J., SELIN, C., SOKOLOFF, L. and KUHL, D. E. (1979). Tomographic measurement of local cerebral glucose metabolic rate in humans with [F-18]2-fluoro-2-deoxy-D-glucose: validation of method. *Annals of Neurology* **6** 371–388.

[40] POLITTE, D. and SNYDER, D. (1991). Corrections for accidental coincidences and attenuation in ML image reconstruction. *IEEE Transactions on Medical Imaging* **10** 82.

[41] RALL, L. B. (1969). *Computational Solution of Nonlinear Operator Equations*. Wiley, New York.

[42] RAPOPORT, S. I. (1991). Discussion of PET workshop reports, including recommendations of PET data analysis working group. *Journal of Cerebral Blood Flow and Metabolism* **11** A140–146.

[43] RIPLEY, B. (1987). *Stochastic Simulation*. Wiley, New York.

[44] ROCKMORE, A. and MAKOVSKI, A. (1976). A maximum likelihood approach to emission image reconstruction from projections. *IEEE Transactions on Nuclear Science* **23** 1428–1432.

[45] SHEPP, L. A. and VARDI, Y. (1982). Maximum likelihood reconstruction in positron emission tomography. *IEEE Transactions on Medical Imaging* **1** 113–122.

[46] SILVERMAN, B. W., JONES, M. C., WILSON, J. D. and NYCHKA, D. (1990). A smoothed EM approach to indirect estimation problems, with particular reference to stereology and emission tomography. *J. Roy. Statist. Soc. Ser. B* **52** 271–324.

[47] SMITH, K. T., SOLMON, D. C. and WAGNER, S. L. (1977). Practical and mathematical aspects of the problem of reconstructing a function from radiographs. *Bull. Amer. Math. Soc.* **83** 1227–1270.

[48] SNYDER, D. L. and MILLER, M. I. (1985). The use of sieves to stabilize images produced with the EM algorithm for emission tomography. *IEEE Transactions on Nuclear Science* **32** 3864–3872.

[49] SNYDER, D. L., MILLER, M. I., LEWIS, J. T. and POLITTE, D. G. (1987). Noise and edge artifacts in maximum-likelihood reconstructions for emission tomography. *IEEE Transactions on Medical Imaging* **6** 228–238.

[50] SNYDER, D. L., THOMAS, L. J. and TER-POGOSSIAN, M. M. (1981). A mathematical model for positron emission tomography systems having time of flight measurement. *IEEE Transactions on Nuclear Science* **28** 3575–3581.

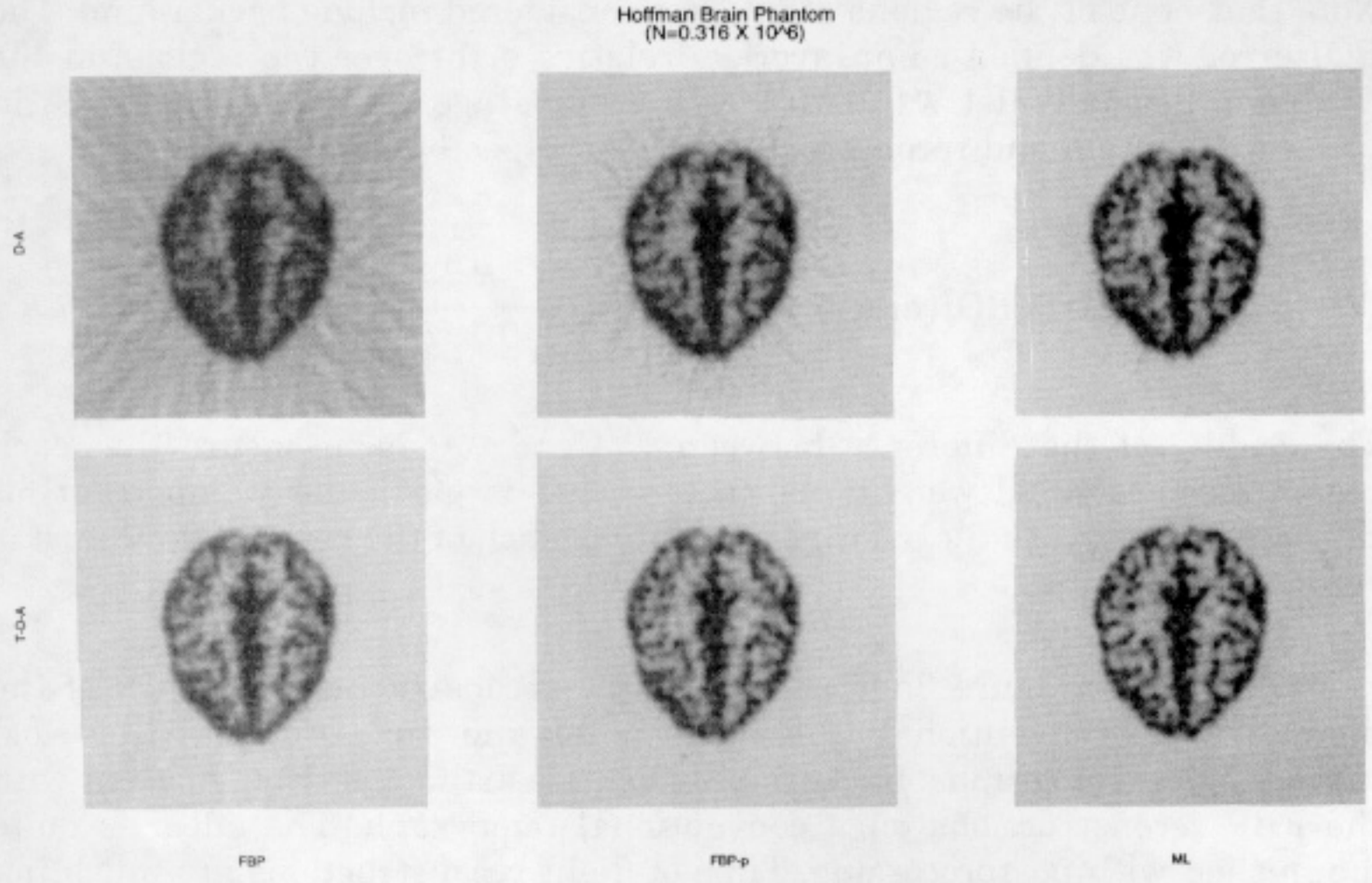
[51] TAYLOR, M. E. (1981). *Pseudodifferential Operators*. Princeton Univ. Press.

[52] VARDI, Y. (1990). Comments on "A smoothed EM approach to indirect estimation problems, with particular reference to stereology and emission tomography," by B. W. Silverman, M. C. Jones, J. D. Wilson and D. Nychka. *J. Roy. Statist. Soc. Ser. B* **52** 319–320.

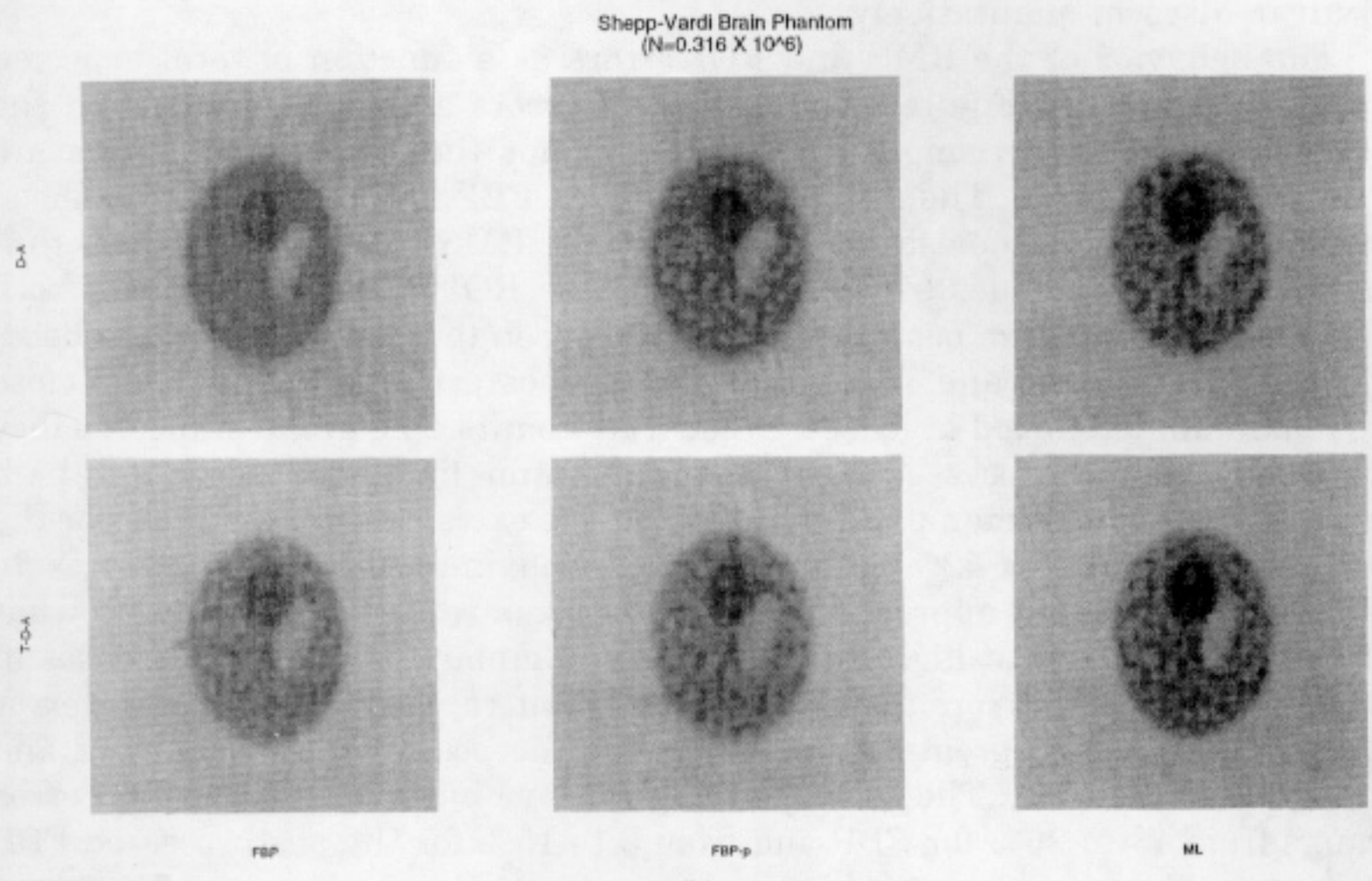
[53] VARDI, Y. and LEE, D. (1993). From image deblurring to optimal investments: Maximum likelihood solutions for positive linear inverse problems (with discussion). *J. Roy. Statist. Soc. Ser. B* **55** 569–612.

[54] VARDI. Y., SHEPP, L. A. and KAUFMAN, L. (1985). A statistical model for positron emission tomography. *J. Amer. Statist. Assoc.* **80** 8–37.
[55] VEKLEROV, E. and LLACER, J. (1990). The feasibility of images reconstructed with the method of sieves. *IEEE Transactions on Nuclear Science* **37** 835–841.
[56] WAHBA, G. (1990). *Spline Models in Statistics*. SIAM, Philadelphia.
[57] WEINBERGER, H. F. (1974). *Variational Methods for Eigenvalue Approximation*. SIAM, Philadelphia.
[58] WILLSON, C. B. J. H. (1992). PET Scanning in oncology. *European Journal of Cancer* **28** 508–510.
[59] ZHANG, P. (1992). On deconvolution using time of flight information in positron emission tomography. *Statist. Sinica* **2** 553–575.

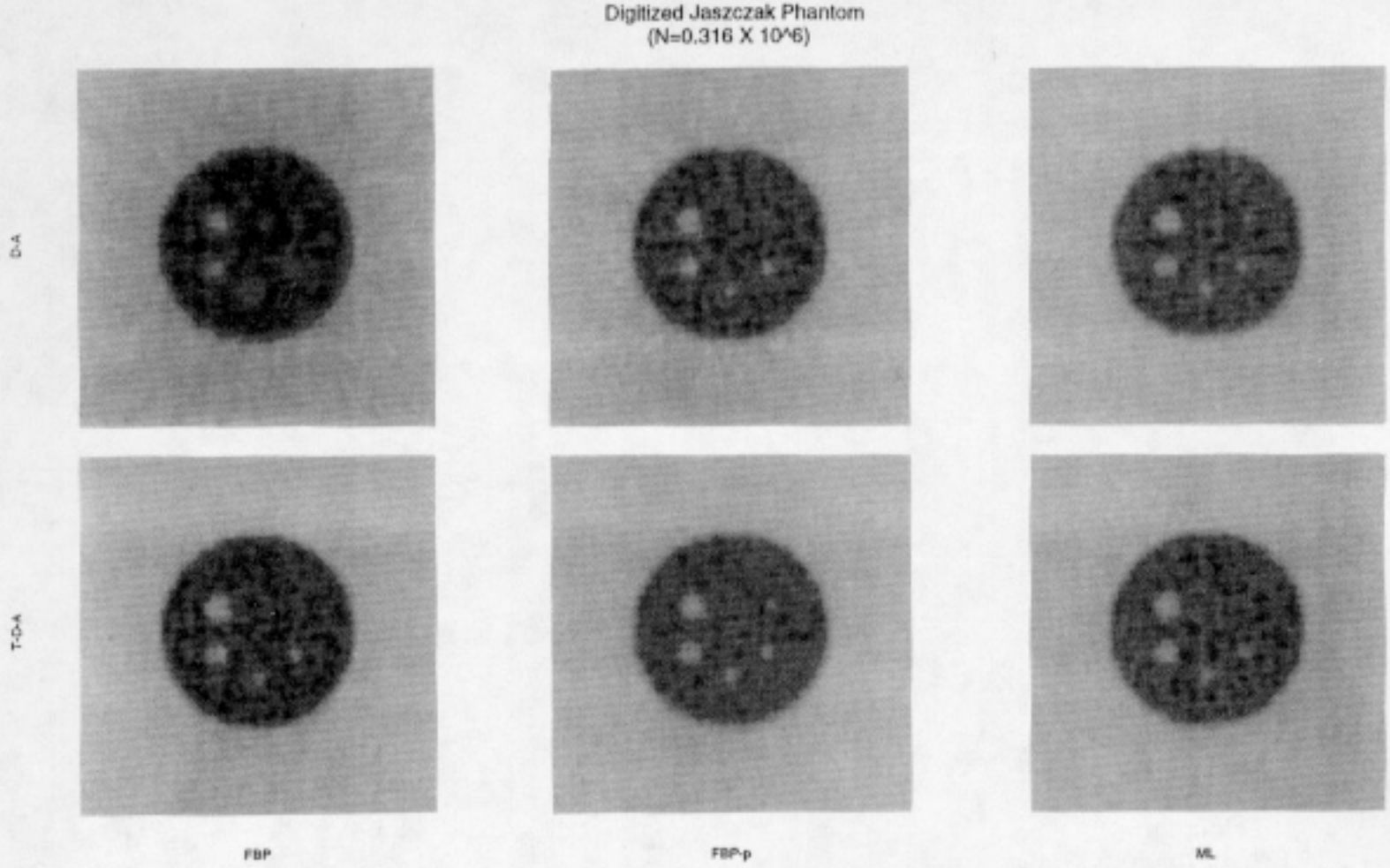
DEPARTMENT OF STATISTICS
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195


Hoffman Brain Phantom
(N=0.316 X 10^6)
D-A
T-D-A
FBP
FBP-p
ML

(a)


Shepp-Vardi Brain Phantom
(N=0.316 X 10^6)
D-A
T-D-A
FBP
FBP-p
ML

(b)



(c)