# Penalized Weighted Least-Squares Image Reconstruction for Positron Emission Tomography

Jeffrey A. Fessler

*Abstract*—This paper presents an image reconstruction method for positron-emission tomography (PET) based on a penalized, weighted least-squares (PWLS) objective. For PET measurements that are precorrected for accidental coincidences, we argue statistically that a least-squares objective function is as appropriate, if not more so, than the popular Poisson likelihood objective. We propose a simple data-based method for determining the weights that accounts for attenuation and detector efficiency. A nonnegative successive over-relaxation (+SOR) algorithm converges rapidly to the global minimum of the PWLS objective. Quantitative simulation results demonstrate that the bias/variance trade-off of the PWLS+SOR method is comparable to the maximum-likelihood expectation-maximization (ML-EM) method (but with fewer iterations), and is improved relative to the conventional filtered backprojection (FBP) method. Qualitative results suggest that the streak artifacts common to the FBP method are nearly eliminated by the PWLS+SOR method, and indicate that the proposed method for weighting the measurements is a significant factor in the improvement over FBP.

## I. INTRODUCTION

**P**ET IMAGING provides noninvasive quantification of human physiology for medical diagnosis and research. The quantitative accuracy of PET is limited by the imperfect system response and by the methods used to reconstruct trans-axial images from projection measurements. The conventional FBP reconstruction method is based on a mathematical idealization of tomographic imaging [1]. The FBP method disregards the spatially-variant system response of PET systems, and statistical noise is treated in a post-hoc manner by spatially-invariant smoothing. Although these approximations may be adequate for some purposes, there is little question that the FBP method is suboptimal for quantitative applications such as brain activation studies [2] and nonlinear functional images [3]. Such studies are particularly challenging since the total numbers of detected photon coincidence events per slice are often fairly low.

Statistical image reconstruction (SIR) methods can account for spatially-variant system responses, and can also incorporate the measurement statistics. This potential has motivated the development of a great many iterative reconstruction algorithms. Ironically, most of the SIR methods reported for PET have been implemented using spatially-*invariant* approximations to the system response, and have been based on an idealized model for the measurement statistics. Since the full capabilities of SIR methods have therefore not been realized in practice, it is perhaps unsurprising that there is ongoing debate as to whether the potential improvements of SIR over FBP are significant enough to justify the additional computation time involved.

The benefits of SIR methods are likely to depend on the task of interest. In this paper, we focus on the specific context of quantifying radiotracer concentrations within small structures. This task is important because small structures are poorly quantified by FBP [4], which in turn degrades the accuracy of parametric images [3]. Recent work by several groups has indicated potential improvements for similar tasks using SIR methods [5]-[7]. Furthermore, the bias and variance within a small point source in a uniform background are directly related to resolution and noise, respectively, so this task provides a somewhat generic measure of reconstruction algorithm performance.

SIR methods require five components:

1) a finite parameterization of the positron-annihilation distribution, e.g., its representation as a discretized image,
2) a system model that relates the unknown image to the expectation of each detector measurement,
3) a statistical model for how the detector measurements vary around their expectations,
4) an objective function that is to be maximized to find the image estimate, and
5) an algorithm, typically iterative, for maximizing the objective function, including specification of the initial estimate and stopping criterion.

In Section II, we review the prevailing choices for the five SIR components, and, where appropriate, contrast them with the approach proposed in this paper. We argue that the measurement statistics are non-Poisson and analytically intractable when accidental coincidence events are precorrected, and therefore propose that a penalized, weighted least-squares objective is an appropriate practical compromise. We apply the +SOR "coordinate-descent" method for fast, globally convergent minimization of that objective, subject to nonnegativity constraints. In Section III, we describe the computer simulations used to compare quantification by FBP, ML-EM, and the PWLS+SOR reconstruction method. Section IV summarizes the results of this comparison, and also qualitatively illustrates the different noise characteristics of the methods on FDG thorax images. Future directions are discussed in Section V.

## II. THEORY

This section briefly reviews the literature for the five components of SIR methods, and describes the method proposed in this paper.

### A. Object Parameterization

The blurring effect of positron range implies that the annihilation distribution for a PET study is band limited. Thus, discretization of the distribution is acceptable. Decomposing the annihilation distribution into rectangular voxels is the *de facto* standard for parameterizing images. Although smoother bases have been suggested [8], [9], the voxel basis has the important practical property that its support is minimal (no overlap), thus the system matrix (described below) is maximally sparse. This sparseness facilitates computations. If $\lambda(\underline{x})$ denotes the spatial distribution of the positron annihilations, then we approximate $\lambda$ by:

$$\lambda(\underline{x}) \approx \sum_j \lambda_j I_j(\underline{x}) \qquad (1)$$

where $\lambda_j$ denotes the mean activity in the $j$th voxel, and $I_j(\underline{x})$ is the indicator function with the $j$th voxel as its support [8].

The dimension of the voxels is an important design issue that is unique to SIR methods. The classical Nyquist sampling theory was developed for noiseless, spatially-invariant systems, and does not directly apply to PET reconstruction. An example of this was illustrated by Mintun *et al.* [10] in their discussion of axial resolution. The FBP method can, in principle, reconstruct images with arbitrarily fine pixel grids, whereas with unregularized SIR methods, voxel sizes that are too small lead to over-parameterization and numerical instability. Conversely, voxel sizes that are too large can produce model mismatch and loss of image features. Although the importance of system sampling is well understood for FBP [11], the effect of image sampling for SIR methods appears to have been addressed only indirectly under an idealized model for X-ray CT [12]. Recently developed Cramér-Rao bound methods [13]–[15] may help address the question of voxel dimension for realistic system models.

### B. System Model

Having discretized the annihilation distribution into a set of voxels, one represents a model for the tomographic system by a "system matrix" $\mathbf{P}$. An element $p_{ij}$ of $\mathbf{P}$ denotes the probability that an annihilation in the $j$th voxel is detected by the $i$th detector pair. Ideally, perhaps after certain corrections, the mean of the $i$th detector pair measurement would be approximately

$$\bar{y}_i = \sum_j p_{ij} \lambda_j \qquad (2)$$

for an annihilation distribution $\lambda(\underline{x})$ represented by (1).

Although the importance of accurate system modeling has been amply illustrated in the SPECT literature, a spatially-*invariant* Gaussian response has been the most popular approximation for PET [5], [16]. The spatially-invariant Gaussian

method was convenient due to its computational simplicity for algorithms that use "run-time" probability calculations. However, the decrease in cost of computer memory has diminished this motivation, and allows precomputing the nonzero elements of $\mathbf{P}$, which significantly reduces the computations per iteration.

Even with precomputed system matrices, there remains a trade-off between accuracy and sparseness. The most accurate analytical approach might be to use an "inverse Monte Carlo" approach analogous to that proposed for SPECT [17], but the resulting system matrix is not sparse. Practical considerations dictate use of sparse approximations, which inevitably introduce some system model inaccuracies. The effects of such model mismatch on reconstruction by SIR methods is not well understood. Presumably one would want to use the most sparse system matrix that adequately describes the system, i.e. the effects of modeling errors are well below the statistical noise.

Although a thorough treatment of system modeling is beyond the scope of this paper, future efforts should consider (1) the difference between cross-slice and direct-slice system responses, (2) the "third dimension" effect described by Silverman *et al.* [18] due to the finite axial width of the detector crystals, and (3) the spatially variant crystal response inherent to PET detector blocks due to inter-crystal mispositioning errors [19].

### C. Statistical Model: Non-Poission

The statistical model describes the distribution of each measurement about its mean, and consequently determines a measure of similarity between the actual measurements and the calculated projections (2). Since the introduction of an ML-EM [20] algorithm for PET a decade ago [21], [22], SIR methods based on a Poisson statistical model [23], [24] have been studied extensively.

The original formulations were based on an idealized PET system, and ignored the effects of accidental coincidence (AC) events. Since accurate quantification of radiotracer activity using PET must include corrections for the effects of AC events [25], [26], several recent papers have attempted to incorporate AC effects into the Poisson framework under the assumption that the AC events are additive Poisson variates with exactly known mean [6], [16], [27]. This assumption is unrealistic for many PET systems.

In routine use, our PET systems[1] use real-time subtraction of delayed-window coincidences [25], [28] to correct for AC events. The system detects coincidence events during two time windows. For events within the first "prompt" window, the corresponding sinogram bin is incremented. These increments should be well approximated by a Poisson process. However, for events within the second "delayed" window, the corresponding sinogram bin is decremented [28]. Although these decrements should also be a Poisson process, the combined effect of the increments and decrements is not Poisson. Even for moderate AC rates (10–20%), this correction produces many negative measurements, clearly violating the Poisson statistical
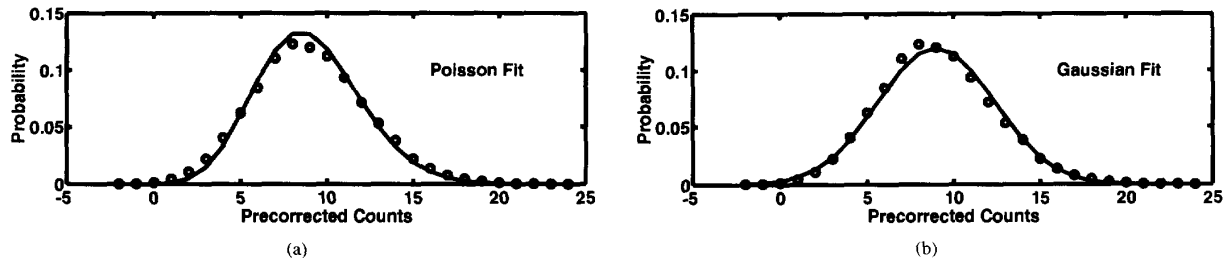
---

[1]CTI ECAT 931 and 921.

Fig. 1. Comparison of Poisson and Gaussian fits (−) to the distribution (o) of PET measurements precorrected for accidental coincidences (see text). The Gaussian fit is more accurate as measured by the $\chi^2$ statistic.

model. Higher percentages of AC events are common for scans acquired shortly after radiotracer injection.

To illustrate the inaccuracy of the Poisson measurement model for AC precorrected measurements, we have performed a small Monte Carlo simulation summarized by Fig. 1. Let $y_i$ be the precorrected measurement for the $i$th coincidence detector pair, then

$$y_i = y_{i,p} - y_{i,d} \tag{3}$$

where $y_{i,p}$ and $y_{i,d}$ are the number of coincidences within the prompt and delayed windows, respectively [28]. If the mean numbers of true coincidence events and AC events during the acquisition are $n_t$ and $n_a$, respectively, then a reasonable model is:

$$y_{i,p} \sim \text{Poisson}\{n_t + n_a\}$$
$$y_{i,d} \sim \text{Poisson}\{n_a\}. \tag{4}$$

The expectation of $y_i$ is $n_t$, so $y_i$ is an unbiased estimate of the number of true coincidences. Since $y_{i,p}$ and $y_{i,d}$ are statistically independent, the variance of $y_i$ is $n_t + 2n_a$, a larger variance than would be expected for a pure Poisson variate.

For concreteness, let $n_t = 9$ and $n_a = 1$. The circles in Fig. 1 show a simulated histogram for $y_i$ generated by a pseudo-random number generator in accordance with the distributions described above ($N = 30,000$). The top figure shows the approximation based on a Poisson distribution with mean nine, the ideal mean. The bottom figure shows the approximation by a Gaussian distribution also with mean nine and with variance 11. As measured by the $\chi^2$ statistic, the Gaussian distribution is the better approximation. Of course for large means, the Poisson distribution is also approximately Gaussian by the Central Limit Theorem [29]. But this example illustrates that even for small true rates and 10% accidental coincidence rates, a Gaussian approximation is as appropriate, if not more so, than the Poisson approximation.

If one could acquire separate sinograms for the prompt and delayed coincidences, then one could consider jointly estimating[2] the AC means and the $\lambda_j$'s from the two sinograms [22]. Alternatively, one could exploit the spatial smoothness of AC events, form an estimate of their means using the delayed-window measurements, and then incorporate those estimates as "known AC means" into the ML-IB method of

Politte and Snyder [16]. In principle such methods would have the advantage that they retain the higher-order moments associated with the skewness of the Poisson distribution, whereas a Gaussian approximation only models the first and second moments. Whether that theoretical advantage produces practical improvements is an open question.

Since the mean AC contributions to the precorrected measurements are unknown[3], the probability distributions of the precorrected measurements are also unknown. Thus, pure likelihood-based methods are inapplicable, and one must resort to approximate similarity measures. In light of Fig. 1, we propose using a weighted least-squares (WLS) similarity measure:

$$\frac{1}{2}(\hat{\mathbf{y}} - \mathbf{P}\lambda)'\mathbf{\Sigma}^{-1}(\hat{\mathbf{y}} - \mathbf{P}\lambda) \tag{5}$$

where $'$ denotes transpose, $\lambda$ is the vector of annihilation rates $\lambda_j$, $\mathbf{P}$ is the system matrix, and the measurement vector $\hat{\mathbf{y}}$ represents an emission sinogram that has been precorrected for the effects of dead-time, attenuation, detector sensitivity, AC events, and possibly scatter. (Thus $\hat{y}_i$ is an estimate of $\bar{y}_i$.) The matrix $\mathbf{\Sigma}$ is diagonal with $i$th entry $\sigma_i^2$, an estimate of the variance of the $i$th precorrected measurement $\hat{y}_i$. This weighting is critical to the method, and our approach to computing $\mathbf{\Sigma}$ is described in the Appendix.

Is an approximate statistical model likely to achieve the goals of SIR methods? One aim is to achieve resolution recovery and uniformity by incorporating the system response. The WLS similarity measure accommodates the system response through the first-order moment. Another benefit of SIR methods is a nonuniform weighting of the measurements, where the weighting reflects the relative information of each measurement. The ML-EM algorithm implicitly incorporates such a weighting by dividing each measurement by its predicted value before backprojecting. This is in complete contrast to the conventional FBP method which treats all measurements equally, despite the large variations in counts and correction factors. The WLS similarity measure also accounts for the relative information of each measurement through the weights. Even if the weights are suboptimal, as the data-weighting discussed in the Appendix may be, it should nevertheless be an improvement over the uniform weights implicit in FBP!

---

[2] We have studied a similar joint estimation method for accounting for statistical uncertainties in transmission scans [30], [31].

[3] The AC contributions are recorded over the entire slice only, not on a ray-by-ray basis.

## D. Objective Function

Objective functions based solely on the measurement statistics, be they Poisson or Gaussian, perform poorly due to the ill-conditioned nature of tomographic reconstruction. Unregularized methods produce increasingly noisy images with iteration [32]. To remedy this problem, several regularization methods have been investigated that impose smoothness constraints on the image estimate.

One approach is the method of sieves [33], [34]. When AC effects are included in the Poisson case, the ML-IB method of Politte and Snyder apparently requires that the resolution and kernel sieves be equal, in which case the method of sieves is equivalent[4] to post-filtering the ML image estimate [16]. Therefore the method of sieves retains the slow convergence of the ML-EM algorithm, for which a few hundred [33], if not several thousand [6], [35] iterations are required.

A more flexible approach is to incorporate a smoothness penalty or "prior" [36]–[39], which is particularly straightforward with the WLS similarity measure. Sauer and Bouman [40] have proposed one approach in the context of X-ray transmission tomography that we have adapted to PET reconstruction. This method is based on the following *penalized, weighted least-squares* objective function:

$$\Phi(\lambda) = \frac{1}{2}(\hat{\mathbf{y}} - \mathbf{P}\lambda)'\Sigma^{-1}(\hat{\mathbf{y}} - \mathbf{P}\lambda) + \beta R(\lambda) \qquad (6)$$

where $R(\lambda)$ is a regularizing penalty term. (Similar objectives have been used for "Bayesian" methods [8], [40]). The goal is to estimate $\lambda$ from $\hat{\mathbf{y}}$:

$$\hat{\lambda} = \arg\,\min_{\lambda \geq 0}\Phi(\lambda).$$

The effect of the penalty term is to discourage disparities between neighboring pixel values, while the effect of the first term in (6) is to encourage agreement with the measured data. These are usually conflicting goals, and the smoothing parameter $\beta$ controls the trade-off between the two, in analogy with the filter window that one must choose for FBP reconstruction.

Many penalty functions $R(\lambda)$ have been proposed for image reconstruction [36], [37], [41]–[46], some of which aim to smooth "uniform" regions while maintaining edge sharpness. Since we are interested in low-count scans where edge preservation is probably unrealizable, in this paper we use a simple quadratic smoothness penalty:

$$R(\lambda) = \frac{1}{2}\lambda'\mathbf{R}\lambda = \frac{1}{2}\sum_{j}\sum_{k \in N_j} w_{jk}\frac{1}{2}(\lambda_j - \lambda_k)^2 \qquad (7)$$

where $N_j$ is the set of eight neighbors of the $j$th pixel. The weights $w_{jk}$ equal 1 for horizontal and vertical neighbors, and $1/\sqrt{2}$ for diagonal neighbors. The following theorem shows that this penalty leads to a strictly convex objective function $\Phi$.

*Theorem 1: If* $\Sigma^{-1/2}\mathbf{P}\mathbf{1} \neq \mathbf{0}$ *where* $\mathbf{1}$ *is a column vector of ones, i.e., the projection of a uniform source is nonzero, then* $\Phi$ *is strictly convex for* $\beta > 0$, *i.e. its Hessian*

$$\mathbf{H} = \nabla^2\Phi = \mathbf{P}\Sigma^{-1}\mathbf{P} + \beta\mathbf{R} \qquad (8)$$

*is positive definite.*

[4] Under the often disregarded assumption that the smoothing operator and the projection operator commute [33, equation (12)].

*Proof:* It suffices to show that $\mathbf{x}'\mathbf{H}\mathbf{x} \neq 0\ \forall\mathbf{x} \neq \mathbf{0}$. From (7) it is clear that $\mathbf{x}'\mathbf{R}\mathbf{x} = 0$ only when $\mathbf{x} = \mathbf{0}$ or $\mathbf{x} = c\mathbf{1}$ for some $c \neq 0$. But $c\mathbf{1}'\mathbf{H}c\mathbf{1} = c^2\|\Sigma^{-1/2}\mathbf{P}\mathbf{1}\|^2 \neq 0$ by assumption. $\qquad\square$

## E. Iterative Algorithm

Ideally the objective function alone would determine the statistical properties of an estimator. In practice, the convergence characteristics of the algorithm that maximizes the objective may also influence those properties. For example, if the algorithm only finds local extrema of $\Phi$, then the estimator is inefficient. The ML-EM algorithm for the unpenalized Poisson objective is converges to a global maximum [21], [22]. However, when one regularizes the Poisson objective with a smoothness penalty, the maximization step of the EM algorithm becomes cumbersome, and the corresponding iterative algorithms converge slowly to possibly local extrema [41], [43].

The classical methods for minimizing quadratic objectives, such as steepest descent or conjugate gradient, do not easily accommodate the physical nonnegativity of $\lambda$. However, minimizing a quadratic objective subject to a nonnegativity constraint is a type of "linear complementarity problem" [47], [48], for which the (projected) successive overrelaxation (+SOR) method is a natural algorithm since the nonnegativity constraint applies independently to each parameter. A special case of the +SOR method is the Gauss-Siedel algorithm [49], [50], which has been applied to transmission tomography by Sauer and Bouman [40]. In the Bayesian literature it is known as ICM [51].

The +SOR algorithm updates each image parameter individually by minimizing the objective function (6) over that parameter while holding the other parameters fixed. Since our objective is quadratic, the minimization is computed analytically (no line searches are required). One iteration consists of updating every pixel value in some sequence.

A detailed discussion of +SOR is given in [47], [40], so we only summarize the algorithm here. Let $\hat{\lambda}$ denote the current estimate of $\lambda$, and let $\mathbf{p}_j$ denote the $j$th column of $\mathbf{P}$. The PWLS+SOR procedure is as follows.

*Initialization:*

$$\hat{\lambda} = \mathrm{FBP}\{\mathbf{y}\}$$
$$\hat{\mathbf{r}} = \hat{\mathbf{y}} - \mathbf{P}\hat{\lambda}$$
$$s_j = \mathbf{p}_j'\Sigma^{-1}\mathbf{p}_j, \forall j,$$
$$d_j = s_j + \beta\sum_{k \in N_j} w_{jk}. \qquad (10)$$

*For each j:*

$$\hat{\lambda}_j^{\mathrm{old}} := \hat{\lambda}_j$$
$$\hat{\lambda}_j^{\mathrm{new}} := \frac{\mathbf{p}_j'\Sigma^{-1}\hat{\mathbf{r}} + s_j\hat{\lambda}_j^{\mathrm{old}} + \beta\sum_{k \in N_j} w_{jk}\hat{\lambda}_k}{d_j}$$
$$\hat{\lambda}_j := \max\{0, (1 - \omega)\hat{\lambda}_j^{\mathrm{old}} + \omega\hat{\lambda}_j^{\mathrm{new}}\}$$
$$\hat{\mathbf{r}} := \hat{\mathbf{r}} + \mathbf{p}_j(\hat{\lambda}_j^{\mathrm{old}} - \hat{\lambda}_j)$$

Note that the updates to $\hat{\lambda}$ are done sequentially in place, in contrast to most reconstruction algorithms that simultaneously update all pixels. Although successive algorithms are difficult to parallelize in general, parallel methods for +SOR are available [48].

### F. Convergence Properties

Since $\Phi$ is strictly convex by Theorem 1, it follows from [47, p. 465] that there is a unique $\hat{\lambda} \geq 0$ that minimizes $\Phi$ (i.e., satisfies the Karush-Kuhn-Tucker conditions [47, p. 560]), and that the +SOR sequence converges from any initial estimate to that unique minimum for $\omega \in (0, 2)$ [47, p. 372]). Furthermore, if $\omega \in (0, 1]$, then the sequence of estimates monotonically decreases $\Phi$.

The convergence *rate* of the SOR algorithm depends on $\omega$. Sauer and Bouman analyzed the convergence properties of Gauss-Siedel ($\omega = 1$) [40], and in the remainder of this section we apply their analysis method to SOR. First, decompose the Hessian (8) by:

$$H = L + D + L'$$

where $D$ is the diagonal of $H$, and $L$ is a strictly lower triangular matrix. Then without the nonnegativity constraint, the SOR method can be compactly written [47]:

$$\lambda^{i+1} = (1 - \omega)\lambda^i + \omega D^{-1}(A'\Sigma^{-1}y - L'\lambda^i - L\lambda^{i+1})$$

or

$$\lambda^{i+1} = (D + \omega L)^{-1}[-((\omega - 1)D + \omega L')\lambda^i + \omega A'\Sigma^{-1}y].$$

This sequence converges geometrically, and its convergence rate is governed by the eigenvalues of

$$G_\omega = -(D + \omega L)^{-1}((\omega - 1)D + \omega L') \qquad (11)$$

see [40, equation (24)] for $\omega = 1$.

To analyze the eigenvalues of $G_\omega$ as a function of $\omega$, we adopt simplifications similar to those in [40], i.e.: $\Sigma = \sigma^2 I$, and the matrices $P'P$, $R$, and $H$ are circulant-block-circulant. The latter assumption implies that multiplication by any of these matrices is equivalent to periodic convolution of the image by a spatially-invariant 2D kernel. Since the discrete Fourier transform diagonalizes circulant matrices, we can use 2D-FFT of the 2D kernels to study the eigenvalues of $G_\omega$. To determine the kernel of the matrix $P'P$, Sauer and Bouman projected and then backprojected a point source. Here, we use the following analytical approximation:

$$f(r) = \begin{cases} \pi - 2r, & r \in [0, 1] \\ 2(\arcsin(1/r) + (r - \sqrt{r^2 - 1})) & r > 1 \end{cases}$$

which is shown in Fig. 2 [40, Fig. 11] and [52, Fig. 1]). This function has the expected $1/r$ asymptotic form, but is well behaved near zero—as it must be for a discrete system. Ignoring edge effects, the kernel of the regularization matrix $R$ described by (7) is

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} + \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 0 & -1 \\ 0 & 4 & 0 \\ -1 & 0 & -1 \end{bmatrix}. \qquad (12)$$
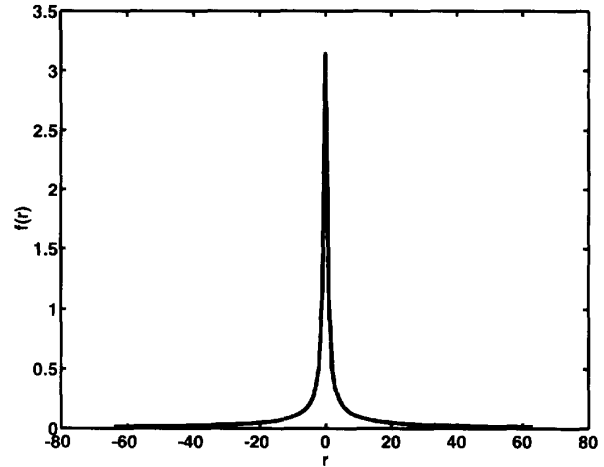


Fig. 2. Plot of $f(r)$, the cross-section of the kernel of the circulant-block-circulant approximation to the projection/backprojection operator $P'P$. The tails decrease like the expected $1/r$ response.

Define $K_P$ to be the $128 \times 128$ matrix with $(i, j)$th element equal to $f(\sqrt{(i - 65)^2 + (j - 65)^2})$. Define $K_R$ to be the $128 \times 128$ matrix of zeros except let the $3 \times 3$ block centered at $(65, 65)$ equal the kernel of $R$ (12). Defining $K = \sigma^{-2}K_P + \beta K_R$, then $K$ is the kernel of the circulant-block-circulant approximation to $H$ see (8)). Let $K_L$ be the "causal" part of $K$ with respect to the conventional left-right/top-down ordering, i.e., $K_L$ is identical to $K$ for the first 64 rows and for the first 63 elements of the 65th row, and zero elsewhere. Let $K_D$ be element $(65, 65)$ of $K$, and let $l(f_x, f_y)$ be the 2D FFT of $K_L$. Then the eigenvalues of $G_\omega$ are given by

$$g_\omega(f_x, f_y) = -\frac{(\omega - 1)K_D + \omega\, l^\star(f_x, f_y)}{K_D + \omega\, l(f_x, f_y)}$$

where $\star$ denotes complex conjugate (see (11) and [40, equation (25)]).

Figs. 3 and 4 show plots of

$$\max_{f_x} |g_\omega(f_x, \cdot)| \quad \text{and} \quad \max_{f_y} |g_\omega(\cdot, f_y)|$$

for $\beta = 1$ and a few values of $\omega$. One sees that using $\omega > 1$ would increase all of the eigenvalues, and thus reduce the convergence rate. On the other hand, using $\omega < 1$ will increase the convergence rate of the low-frequency components, at the expense of slower convergence for the high-frequency components. We have found that this trade-off is useful for improving the overall convergence rate. We usually initialize the iteration with a smooth FBP image, for which the low spatial-frequency components of the initial estimate are nearly correct. A few iterations with $\omega < 1$ will quickly fine-tune the low frequencies, followed by a few more iterations with $\omega = 1$ to converge the high frequencies. To counteract the directional effect illustrated in Figs. 3 and 4, we update the image pixels in four different raster scan orderings.

Note that for any $\omega$, the high frequencies will converge faster with SOR (smaller eigenvalues) than the low frequencies. This characteristic of successive algorithms is the
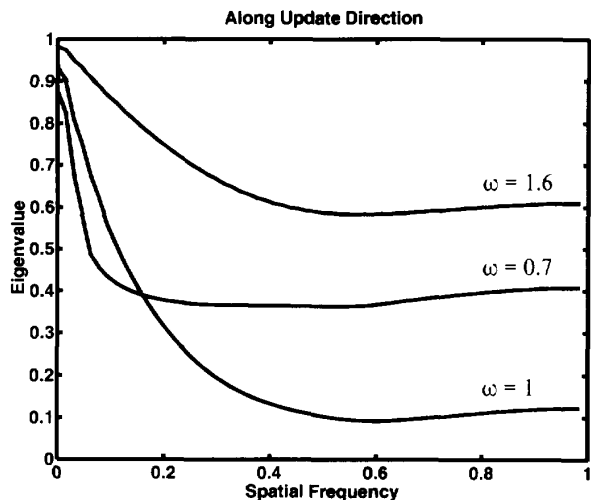
Fig. 3. Maximum eigenvalues of SOR algorithm along update direction, for $\omega = 1, 0.7,$ and 1.6.



Fig. 5. Simulated annihilation distribution. The bias and variance of the reconstructed values within the small hot and cold pixels serve as measures of resolution and noise.
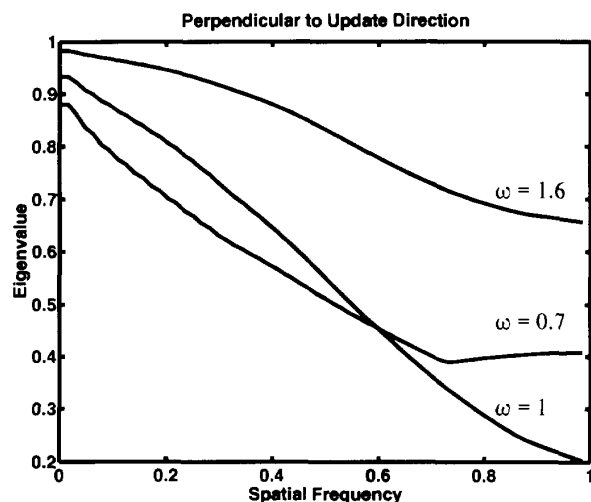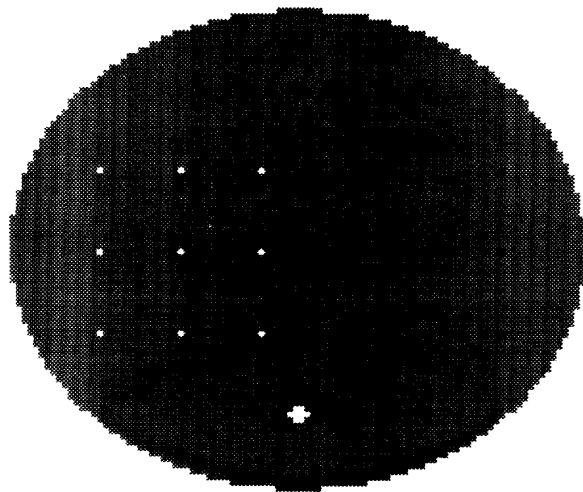


Fig. 4. Maximum eigenvalues of SOR algorithm perpendicular to update direction, for $\omega = 1, 0.7,$ and 1.6.

opposite of the usual simultaneous algorithms (ML-EM, conjugate gradient, etc.) for which the low frequencies converge fastest. Since FBP provides a reasonable initial estimate of the low frequencies, fast convergence of the high frequencies (with suitable regularization) is desirable. Typically the pixel estimates change very little after about 20 iterations. In contrast, ML-EM pixel values continue to change substantially after dozens of iterations.

## III. SIMULATION

Every reconstruction method has a parameter that affects the trade-off between bias and variance. For FBP it is the window type and the cutoff frequency $\alpha$, for ML-EM it is the number of iterations, and for PWLS+SOR it is the parameter $\beta$. Our aim was to address the question: for various levels

of bias (i.e. various resolutions), how do the variances of the image estimates compare between algorithms? To address this question, we performed a simulation using the software phantom shown in Fig. 5, consisting of a uniform background with intensity 1, several hot pixels with intensity 2, and several cold pixels with intensity 0. (Several pixels were used so that we could average among them and obtain statistically significant comparisons with a moderate number of noise realizations.) The pixel grid is $128 \times 128$, with 3 mm pixels. The ellipse radii were 125 mm and 150 mm (approximate abdomen dimensions).

This pixelated software phantom was forward projected (2) using a precomputed system model corresponding to an idealized PET system with 128 angular samples over $180°$, and 110 radial samples with 3 mm spacing. Each $p_{ij}$ was calculated as the area of intersection between the $j$th pixel and a strip of width 6 mm. (Since the strip width of 6 mm is wider than the detector spacing of 3 mm, the strips overlap.) The detector response of this system is thus 6 mm wide rectangular function. Since this system model is spatially-invariant, this is a "best-case" situation for the FBP algorithm. The same system model was used for calculating the projections and for the ML-EM and PWLS+SOR algorithms, so they also represent best-case performance. A more rigorous comparison between FBP and iterative methods would use a nonpixelated (or finely binned) phantom. Since this paper emphasizes the comparison between Poisson likelihood and weighted least-squares similarity measures, we used a pixelated phantom to eliminate possible confounding effects due to system model mismatch. The effects of system model mismatch needs further investigation for all statistical reconstruction methods.

The projections were multiplied by nonuniform attenuation factors corresponding to an ellipse with radii 125 mm and 150 mm and attenuation coefficient of 0.01/mm. Nonuniform detector sensitivities were applied by using pseudo-random log-normal variates with standard deviation 0.4, (based on

empirical fits to the logarithm of measured efficiency nor-
malization factors). After globally scaling the sinogram to
a mean sum of 700,000 true events, $N = 100$ realizations
of AC precorrected sinogram measurements were generated
using pseudo-random Poisson variates according to (13) in
the Appendix. The mean AC contribution to each bin was
about 9%. We chose a low number of counts and a small AC
percentage because one expects the WLS approximation to be
the poorest at low event rates. If the AC rates were increased
the ML-EM algorithm would be increasingly positively biased
since the negative measurements must be set to 0, whereas
the WLS would become increasingly more accurate since
the measurements will approach a Gaussian distribution by
the central limit theorem. Thus, a low AC percentage puts
PWLS+SOR at the *least* advantage relative to ML-EM.

### A. ML-EM

For the ML-EM algorithm, the noisy measurements were
forced nonnegative, and the (known) effects of attenuation,
detector efficiency, and global scaling were incorporated into
P. Each of the 100 measurement realizations were recon-
structed starting from a uniform ellipse with intensity 1. The
estimates from iterations 10, 20, 30, 40, 50, 100, ..., 400
were archived for subsequent statistical analysis. Each ML-
EM iteration required approximately 1.5 seconds on a DEC
3000/400.

### B. PWLS+SOR

For PWLS+SOR the noisy measurements were precorrected
for the (known) effects of attenuation, detector efficiency,
and global scaling, and the variance weights were estimated
using the smoothing method described in the Appendix. The
resulting precorrected measurements $\hat{y}$ were reconstructed
using 20 iterations of PWLS+SOR, again initialized with a
uniform ellipse, for $\beta = 2^{-7}, 2^{-6}, \ldots, 2^{-1}$. Each iteration
required approximately 2.0 seconds. To put this in perspective,
all 47 slices of a CTI 921 EXACT could be reconstructed in
about 30 minutes.

For both ML-EM and PWLS+SOR, only pixels within
a support ellipse with radii 150 mm and 159 mm were
updated. Using this support, there were 8,104 unknown pixels
and 13,394 relevant sinogram measurements; such marginal
sampling makes regularization essential.

### C. FBP

The measurements were precorrected as for PWLS+SOR.
One filter used for radial smoothing was a third-order Butter-
worth filter:

$$\frac{1}{1 + \left(\frac{f}{\alpha f_N}\right)^6}$$

where $f_N$ corresponds to 0.5 cycles per radial bin, for $\alpha = 0.3$,
0.4, 0.6, 0.8, and 0.9. Or, to "restore" some of the high
frequencies attenuated by the rectangular system response, the

following Wiener filter was substituted:

$$\frac{\mathrm{sinc}(f/f_N)}{\mathrm{sinc}^2(f/f_N) + \left(\frac{f}{\alpha f_N}\right)^{10}}$$

for $\alpha = 0.4, 0.6, 0.8$, and 1.0.

### D. Statistics

Let $\hat{\lambda}_j^n$ be the estimate of the $j$th pixel from the $n$th noise
realization, $n = 1, \ldots, N = 100$. We define the within-image
average of the hot pixels to be:

$$\hat{\theta}_{\mathrm{hot}}^n = \frac{1}{9} \sum_{j:\mathrm{hot}} \hat{\lambda}_j^n$$

where the summation is over the 9 small hot pixels, and
similarly define the within-image averages of the cold pixels.
Let $\hat{\theta}_{\mathrm{hot}}$ and $\hat{\theta}_{\mathrm{cold}}$ be the ideal values for $\hat{\theta}_{\mathrm{hot}}^n$ and $\hat{\theta}_{\mathrm{cold}}^n$,
respectively, i.e., $\hat{\theta}_{\mathrm{hot}} = 2$ and $\hat{\theta}_{\mathrm{cold}} = 0$. Then by standard
definitions:

$$\mathrm{bias}_{\mathrm{hot}} = \bar{\theta}_{\mathrm{hot}} - \hat{\theta}_{\mathrm{hot}} = \frac{1}{N} \sum_{n=1}^{N} \hat{\theta}_{\mathrm{hot}}^n - \hat{\theta}_{\mathrm{hot}}$$

and

$$\mathrm{variance}_{\mathrm{hot}} = \frac{1}{N-1} \sum_{n=1}^{N} (\hat{\theta}_{\mathrm{hot}}^n - \bar{\theta}_{\mathrm{hot}})^2$$

with similar definitions for the cold pixels. Since the contrast
is 1 for both hold and cold pixels, the percent bias is simply
$100 \cdot \mathrm{bias}$. Likewise for the percent standard deviation.

## IV. RESULTS

### A. Quantitative

Figs. 6 and 7 show the trade-off between bias and variance
for the estimated activity in the cold and hot pixels respec-
tively. Because the point sources are in a uniform background,
there is an inverse monotonic relationship between bias (more
smoothing) and variance for all methods. The following con-
clusions can be drawn from Figs. 6 and 7:

- Although FBP with a Wiener filter did have the desired
  effect of reducing bias relative to FBP with a Butterworth
  filter, it did so at a price of increased variance; at any
  given bias level the Wiener filter had no advantage.
- For both the hot and cold pixels, the ML-EM algorithm
  and the PWLS+SOR methods had comparable bias-
  variance curves, although clearly with fewer iterations
  for PWLS+SOR. In our opinion this is unsurprising since
  both methods are based on reasonable approximations to
  the measurement statistics.
- For the cold pixels both SIR methods both showed
  significantly reduced variability relative to FBP for any
  level of bias. For the hot pixels the SIR methods offered
  only a slight improvement. This is consistent with studies
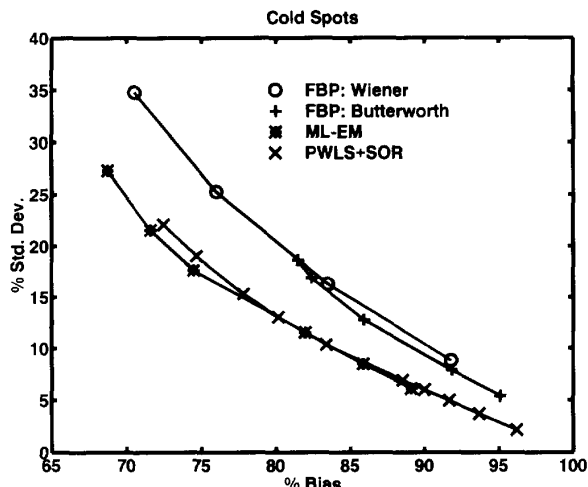  by other investigators.

Fig. 6. Trade-off between bias and variance in the cold pixels as reconstructed by FBP, ML-EM, and PWLS+SOR. For a given bias (i.e. resolution), the standard deviations (i.e. noise) of ML-EM and PWLS+SOR are about 40% smaller than FBP.
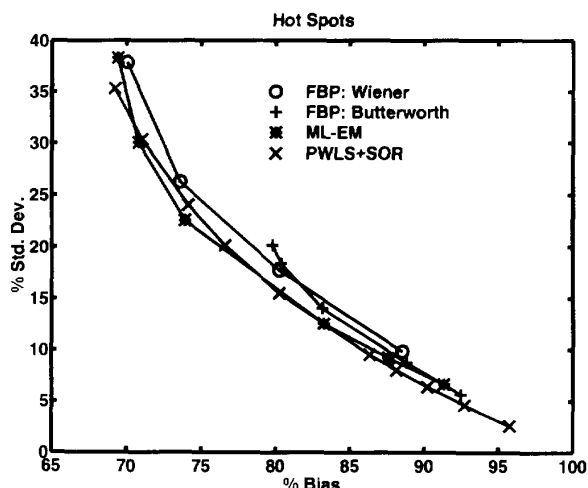


Fig. 8. FDG thorax scan of breast cancer patient via FBP with third-order Butterworth filter with $\alpha = 0.4$ (approximately 8 mm FWHM).



Fig. 7. Trade-off between bias and variance in the hot pixels as reconstructed by FBP, ML-EM, and PWLS +SOR.

The agreement between the performance of ML-EM and PWLS+SOR suggests that the smoothing method for estimating the variances described in Appendix A is an adequate approximation. To further demonstrate this, we applied the PWLS+SOR method using the "ideal" variances $\sigma_i^2 = \mathrm{Var}\{\hat{y}_i\}$, which one can only do in a simulation. The results were indistinguishable both visually and in terms of the statistical analyses describes above. Apparently either the image estimates are somewhat insensitive to the weights, or at the count rates simulated in this study the accuracy of the data-based variance estimate is adequate.

B. Qualitative

The noise properties of reconstruction methods are also of considerable interest because noise structure affects the
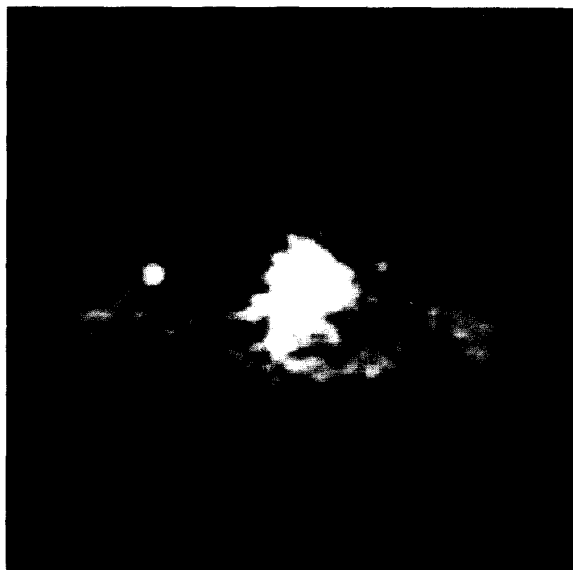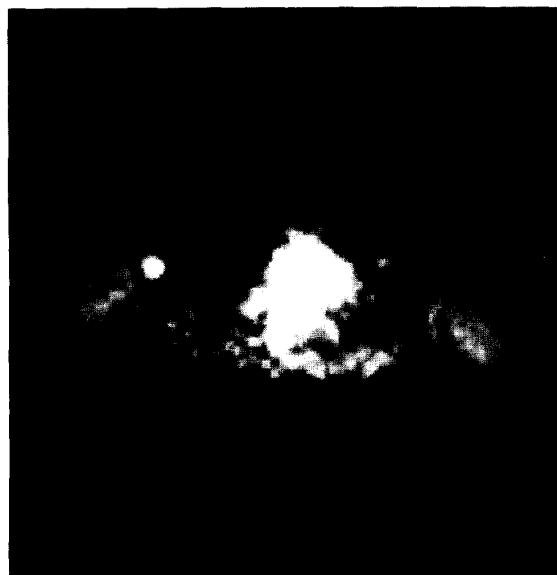


Fig. 9. FDG thorax scan reconstructed with 100 ML-EM iterations post-filtered to approximately 8mm FWHM resolution.

detectability of small lesions. Figs. 8–10 compare the FBP, ML-EM, and PWLS+SOR reconstructions of an FDG thorax image of a patient with breast cancer. There were about 750K prompt coincidences and 20 K delayed coincidences for the slice shown. The noise structure is strikingly different. The reduction in streak artifacts may lead to improved detection of lower contrast lesions. It may also improve the detection of brain activation foci by statistical criteria [2].

Readers who are accustomed to simulated ML-EM studies without accidental coincidences may find the grey background
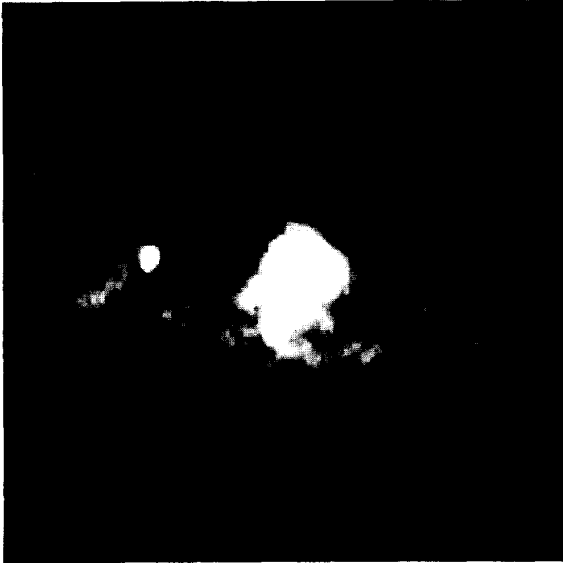
Fig. 10. FDG thorax scan reconstructed with 20 PWLS+SOR iterations. The streak artifacts of the FBP method are nearly eliminated by the statistical ML-EM and PWLS+SOR methods, which may lead to improved detection of lesions with lower contrast than the one shown.
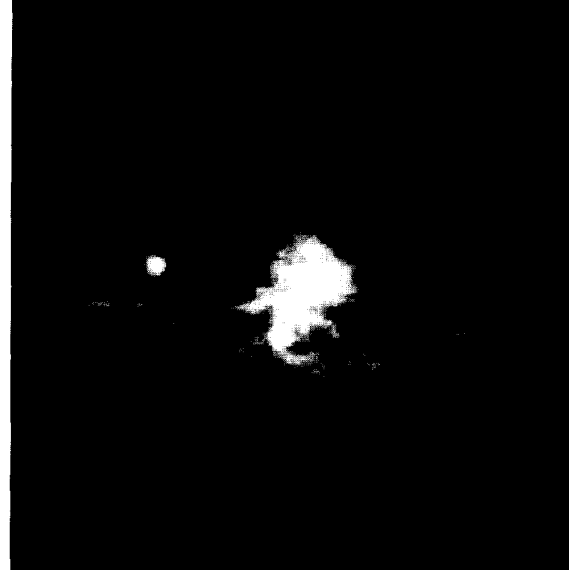


Fig. 11. An *unweighted* penalized least-squares reconstruction of the FDG scan. The reappearance of the streak artifacts strongly suggests that the variance weighting is essential to the PWLS+SOR method, and plays a significant role in the improvement over FBP.

in Fig. 9 to be unexpected. This positive bias is apparently due to the unmodeled accidental coincidence events, and continues to persist after hundreds of iterations. In the absence of AC events and scatter, the sinogram measurements outside of the object would be zero, and the ML-EM algorithm would quickly converge those pixels toward zero.

We conjecture that the reduced streak artifacts in Fig. 10 are due to the variance weighting of PWLS+SOR. Since the attenuation correction factors for the thorax can be very large, even small measurement errors can be amplified by the attenuation correction. The FBP method ignores such statistical differences between different projection elements, whereas the PWLS+SOR method explicitly accounts for them. To substantiate this conjecture, Fig. 11 displays a penalized least-squares reconstruction using *uniform* variance weights. The reappearance of the streak artifacts strongly suggests that the variance weighting is essential, and it plays a crucial role in the improved noise structure of PWLS+SOR.

## V. DISCUSSION

We have considered the measurement statistics for PET systems that precorrect for AC events, and have argued that a PWLS objective is appropriate for such measurements. We summarized the +SOR algorithm for minimizing that objective, and demonstrated that it has fast convergence. Quantitative comparisons to FBP on a simple phantom with small hot and cold pixels demonstrated significant reductions in variance for any level of bias. Qualitative comparisons suggest that the variance weighting of PWLS+SOR significantly improves the noise structure. Although the PWLS objective and +SOR algorithm are not necessarily optimal for PET, the method appears to have some

quantitative and qualitative advantages over FBP. The required computation time is nearing the realm of being practical for routine use.

For the generic quantification task studied here, the bias/variance trade-off of PWLS+SOR and ML-EM were comparable. This does not exclude the possibility of other scenarios where the Poisson likelihood has measurable advantages over weighted least squares. However, the PWLS objective proved viable even in our test case which was deliberately chosen with low counts and low AC events to "stress" the Gaussian approximation.

How one chooses to trade-off bias and variance is clearly task dependent. For certain kinetic estimation tasks, uptake bias leads to inaccuracies in functional parameters [3]. On the other hand, some increase in variance may be tolerable for such tasks since one is generally fitting a low-order parametric model to multiple images. For PWLS+SOR, the parameter $\beta$ controls this trade-off. We are currently investigating the relationship between $\beta$, $\Sigma$, and reconstructed resolution using methods similar to that in Section II-F. The result of this study should be a method for specifying $\beta$ in terms of the desired "average" reconstructed resolution as a function of the measurement noise. The PWLS objective is easier to analyze in this context than a penalized Poisson likelihood, since without the nonnegativity constraint the image estimate is linear in the measurements after the weights are specified.

Although it was high AC fraction studies that initially motivated our considering the PWLS alternative to the Poisson criterion, the method also appears to work well for low AC fraction scans. The FDG scan shown in Fig. 8 was acquired about an hour after injection, and there were less than 3% AC coincidences.

There remain several questions pertaining to the PWLS method that may be worth pursuing. These include: (1) What is the optimal voxel size? (2) How should the different system response for direct and cross planes be incorporated? (3) Would a method such as iteratively reweighted least-squares [53] for variance estimation [53] improve performance enough to offset its considerable computational cost? (The results of our comparison using ideal variances suggest not.) and (4) Should the nonnegativity constraint be enforced in all situations? If the nonnegativity constraint is unneeded or undesirable for some tasks, then there may be even faster alternatives than SOR for minimizing the objective [52], [54].

For simplicity, we have adopted a quadratic penalty function, which permits an analytical minimization of the objective function with respect to each pixel value. There may be non-quadratic penalty functions that result in an even more favorable bias-variance trade-off. It remains to be seen whether or not the benefits of such penalty functions are significant enough to outweigh the increased computational requirements for a non-quadratic objective. As observed by Herman [8, p. 107] long before the advent of fast workstations: "It is unlikely that an efficacious reconstruction algorithm would for long remain unused solely because of computational reasons."

## VI. APPENDIX: VARIANCE

Unlike the Poisson objective, for which the variance equals the mean, the Gaussian objective requires a separate estimate of the variances or weights $\sigma_i^2$. This appendix describes the data-based variance estimate used in the simulations above. This estimate is based on the *a priori* expectation that an object's projections are smooth.

If $\mathbf{y}$ denotes the raw sinogram measurement, then the ideal Poisson-difference model is:

$$y_i \sim \text{Poisson}\left\{n_i^{-1}\left(a_i^{-1}\bar{y}_i + r_i\right)\right\} - \text{Poisson}\left\{n_i^{-1}r_i\right\} \quad (13)$$

where $n_i$ is the $i$th detector efficiency normalization factor, $a_i$ is the attenuation correction factor for the $i$th detector pair, $r_i$ is the mean AC contribution to the $i$th detector pair, and $\bar{y}_i$ is defined by (2). The precorrected measurement is then:

$$\hat{y}_i = n_i a_i y_i$$

which is an unbiased estimator of $\bar{y}_i$, as desired. The variance of this precorrected measurement is:

$$\text{Var}\{\hat{y}_i\} = n_i^2 a_i^2 \text{Var}\{y_i\} = n_i a_i^2 (a_i^{-1}\bar{y}_i + 2r_i).$$

The factor "2" reflects the fact that independent AC events are being added and subtracted from $\mathbf{y}$, so their variances add. We seek an estimate of the variance of $\hat{y}_i$.

If $\hat{y}_i$ is an estimate of $\bar{y}_i$, then a natural choice for the variance estimate is:

$$\sigma_i^2 = n_i a_i^2 (a_i^{-1}\hat{y}_i + 2\hat{r}_i)$$

where $\hat{r}_i$ is an estimate of the mean AC event rate for the $i$th detector pair, typically the total delayed-window events for the slice divided by the total number of sinogram bins. For the simulations above, we have used the following estimate of $\bar{y}_i$:

$$\hat{y}_i = \max\{\text{Smooth}\{\hat{y}_i\}, 7\} \quad (14)$$

where the smoothing was performed with a 1 pixel FWHM Gaussian kernel in the radial direction only. The threshold of seven ensures that the method is not overly sensitive to bins with only a few counts.

## REFERENCES

[1] A. C. Kak and M. Slaney. *Principles of Computerized Tomographic Imaging.* New York: IEEE Press, 1988.
[2] P. T. Fox, M. A. Mintun, E. M. Reiman, and M. E. Raichle. "Enhanced detection of focal brain responses using intersubject averaging and change-distribution analysis of subtracted PET images," *J. Cerebral Blood Flow and Metabolism,* vol. 8, no. 5, pp. 642–653, 1988.
[3] S. C. Huang, D. K. Mahoney, and M. E. Phelps, "Quantitation in positron emission computed tomography: 8. Effects of nonlinear parameter estimation on functional images," *J. Comput. Assisted Tomography,* vol. 11, no. 2, pp. 314–325, 1987.
[4] E. J. Hoffman, S. C. Huang, and M. E. Phelps, "Quantitation in positron emission computed tomography: 1. Effect of object size," *J. Comput. Assisted Tomography,* vol. 3, no. 3, pp. 299–308, 1970.
[5] J. S. Liow and S. C. Strother. "Practical tradeoffs between noise, quantitation and number of iterations for maximum likelihood-based reconstructions," *IEEE Trans. Medical Imaging,* vol. 10, no. 4, pp. 563–571, Dec. 1991.
[6] J. S. Liow and S. C. Strother. "The quantitative performance of maximum likelihood based reconstruction with image wide resolution convergence," *J. Nucl. Medicine.* vol. 33, no. 5, pp. 871, May 1992.
[7] J. A. Fessler, W. L. Rogers, N. H. Clinthorne, G. D. Hutchins, and R. A. Koeppe, "Quantification of the human basal ganglia via iterative reconstruction," *J. Nucl. Medicine.* vol. 33, no. 5, pp. 878, May 1992.
[8] G. T. Herman. *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography.* New York: Academic Press, 1980.
[9] R. M. Lewitt. "Multidimensional digital image representations using generalized Kaiser-Bessel window functions," *J. Optical Soc. Am.,* Ser. A, vol. 7, no. 10, pp. 1834–1846, Oct. 1990.
[10] M. A. Mintun, P. T. Fox, and M. E. Raichle. "A highly accurate method of localizing regions of neuronal activity in the human brain with positron emission tomography," *J. Cerebral Blood Flow and Metabolism,* vol. 9, no. 1, pp. 96–103, 1989.
[11] S. C. Huang, E. J. Hoffman, M. E. Phelps, and D. E. Kuhl, "Quantitation in positron emission computed tomography: 3 Effect of sampling," *J. Comput. Assisted Tomography.* vol. 4, no. 6, pp. 819–826, 1980.
[12] R. H. Huesman. "The effects of a finite number of projection angles and finite lateral sampling of projections on the propagation of statistical errors in transverse section reconstruction," *Phys. Med. Biol.,* vol. 22, no. 3, pp. 511–521, 1977.
[13] A. O. Hero, J. A. Fessler, and W. L. Rogers. "A fast recursive algorithm for computing CR-type bounds for image reconstruction problems," in *Conf. Rec. IEEE Nuclear Science Symp. and Medical Imaging Conf.,* vol. 2, pp. 1188–1190, 1992.
[14] A. O. Hero and J. A. Fessler. "A recursive algorithm for computing CR-type bounds on estimator covariance," *IEEE Trans. Informat. Theory,* in press.
[15] J. A. Fessler and A. O. Hero. "Cramer-Rao bounds for biased estimators in image restoration," in *Proc. Midwest Symp. on Circ. and Syst.,* 1993.
[16] D. G. Politte and D. L. Snyder. "Corrections for accidental coincidences and attenuation in maximum-likelihood image reconstruction for positron-emission tomography," *IEEE Trans. on Medical Imaging,* vol. 10, no. 1, pp. 82–89, Mar. 1991.

[17] C. E. Floyd, R. J. Jaszczak, and R. E. Coleman, "Inverse Monte Carlo: A unified reconstruction algorithm for SPECT," *IEEE Trans. Nuclear Sci.*, vol. 32, no. 1, pp. 779–785, Feb. 1985.

[18] B. W. Silverman, M. C. Jones, J. D. Wilson, and D. W. Nychka, "A smoothed EM approach to indirect estimation problems, with particular reference to stereology and emission tomography," *J. Roy. Statist. Soc. Series B*, vol. 52, pp. (2), pp. 271–324, 1990.

[19] G. Germano and E. J. Hoffman, "An investigation of methods for pileup rejection for 2-D array detectors employed in high resolution PET," *IEEE Trans. Medical Imaging*, vol. 10, no. 2, pp. 223–227, June 1991.

[20] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc. Series B*, vol. 39, no. 1, pp. 1–38, 1977.

[21] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Medical Imaging*, vol. 1, no. 2, pp. 113–122, Oct. 1982.

[22] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comput. Ass. Tomography*, vol. 8, no. 2, pp. 306–316, Apr. 1984.

[23] A. J. Rockmore and A. Macovski, "A maximum likelihood approach to emission image reconstruction from projections," *IEEE Trans. Nucl. Sci.*, vol. 23, pp. 1428–1432, 1976.

[24] Y. Vardi, L. A. Shepp, and L. Kaufman, "A statistical model for positron emission tomography," *J. Amer. Statist. Assoc.*, vol. 80, no. 389, pp. 8–37, Mar. 1985.

[25] E. J. Hoffman, S. C. Huang, M. E. Phelps, and D. E. Kuhl, "Quantitation in positron emission computed tomography: 4. Effect of accidental coincidences." *J. Comput. Ass. Tomography.* vol. 5, no. 3, 391–400, 1981.

[26] M. E. Casey and E. J. Hoffman, "Quantitation in positron emission computed tomography: 7. A technique to reduce noise in accidental coincidence measurements and coincidence efficiency calibration," *J. Comput. Assist. Tomography*, vol. 10, no. 5, pp. 845–850, 1986.

[27] A. Gelman, "Topics in image reconstruction for emission tomography," Ph.D. *dissertation*, Harvard University, Cambridge, MA. , April 1990.

[28] T. J. Spinks, T. Jones, M. C. Gilardi, and J. D. Heather, "Physical performance of the latest generation of commerical positron scanner." *IEEE Trans. Nucl. Sci.*, vol. 35, no. 1, pp. 721–725, Feb. 1988.

[29] R. Billingsley, *Probability and Measure.* New York: Wiley, 1986.

[30] N. H. Clinthorne, J. A. Fessler, G. D. Hutchins, and W. L. Rogers, "Joint maximum likelihood estimation of emission and attenuation denstities in PET," in *Conf. Record IEEE Nuclear Science Symp. and Medical Imaging Conf.*, vol. 3, pp. 1927–1932, 1991.

[31] J. A. Fessler, "On complete data spaces for PET reconstruction algorithms," *IEEE Trans. Nucl. Sci.*, vol. 40, no. 4, pp. 1055–1061, Aug. 1993.

[32] D. L. Snyder and M. I. Miller, "The use of sieves to stabilize images produced with the EM algorithm for emission tomography," *IEEE Trans. Nucl. Sci.*, vol. 32, no. 5, pp. 3864–3871, Oct. 1985.

[33] D. L. Synder, M. I. Miller, L. J. Thomas, and D. G. Polite, "Noise and edge artifacts in maximum-likelihood reconstructions for emission tomography," *IEEE Trans. Medical Imaging*, vol. 6, pp. 3, pp. 228–238, Sept. 1987.

[34] S. Geman and C. R. Hwang, "Nonparametric maximum likelihood estimation by the method of sieves." *The Annals of Statist.*, vol. 10, no. 2, pp. 401–414, 1982.

[35] T. R. Miller, J. W. Wallis, C. S. Butler, M. I. Miller, and D. L. Snyder, "Improved brain SPECT by maximum-likelihood reconstruction." *J. Nucl. Medicine*, vol. 33, no. 5, p. 964, May 1992.

[36] S. Geman and D. E. McClure, "Bayesian image analysis: An application to single photon emission tomography," in *Proc. Stat. Comp. Sec. of Amer. Stat. Assoc.*, pp. 12–18, 1985.

[37] S. Geman and D. E. McClure, "Statistical methods for tomographic image reconstruction." *Proc. 46 Sect. ISI, Bull. ISA*, vol. 52, pp. 5–21, 1987.

[38] Z. Liang, R. Jaszczak, and K. Greer, "On Bayesian image reconstruction from projections: Uniform and nonuniform a priori source information," *IEEE Trans. Medical Imaging*, vol. 8, no. 3, pp. 227–235, Sept. 1989.

[39] J. A. Fessler, N. H. Clinthorne, and W. L. Rogers, "Regularized emission image reconstruction using imperfect side information," *IEEE Trans. on Nucl. Sci.*, vol. 39, no. 5, pp. 1464–1471, Oct. 1992.

[40] K. Sauer and C. Bouman. "A local update strategy for iterative reconstruction from projections." *IEEE Trans. on Signal Processing*, vol. 41, no. 2, pp. 534–548, Feb. 1993.

[41] T. Hebert and R. Leahy. "A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors," *IEEE Trans. Medical Imaging*, vol. 8, no. 2, pp. 194–202, June 1989.

[42] K. Lange, "Convergence of EM image reconstruction algorithms with Gibbs smoothing." *IEEE Trans. Medical Imaging*, vol. 9, no. 4, pp. 439–446, Dec. 1990. (Corrections: June 1991.)

[43] P. J. Green. "Bayesian reconstructions from emission tomography data using a modified EM algorithm." *IEEE Trans. Medical Imaging*, vol. 9, no. 1, pp. 84–93, Mar. 1990.

[44] V. E. Johnson, W. H. Wong, X. Hu, and C. T. Chen, "Image restoration using Gibbs priors: Boundary modeling, treatment of blurring, and selection of hyperparameter." *IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. 13, no. 5, pp. 413–425, May 1991.

[45] I. G. Zubal, M. Lee. A. Rangarajan, C. R. Harrell, and G. Gindi, "Bayesian reconstruction of SPECT images using registered anatomical images as priors," *J. Nucl. Medicine*, vol. 33, no. 5, p. 963, May 1992.

[46] D. S. Lalush and B.M.W. Tsui, "A generalized Gibbs prior for maximum a posteriori-EM SPECT reconstruction with fast and stable convergnece properties." *J. Nucl. Medicine*, vol. 33, no. 5, p. 832, May 1992.

[47] K. G. Murthy, *Linear Complementarity, Linear and Nonlinear Programming.* Berlin: Helderman Verlag, 1988.

[48] W. Niethammer, "A note on the implementation of the successive overrelaxation method for linear complementarity problems," *Numerical Algorithms*, vol. 4, no. 1, pp. 197–200, Jan. 1993.

[49] D. M. Young, "Iterative Solution of Large Linear Systems. New York: Academic Press, 1971.

[50] G. Gulberg et al., "Maximum entropy reconstruction with constraints: Iterative algorithms for solving the primal and dual programs," in *Proc. Tenth Int. Conf. on Information Processing in Medical Imaging*, C. N. de Graaf and M. A. Viergever, Eds., New York: Plenum Press, 1987, pp. 181–200. .

[51] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Statist. Soc. Series B*, vol. 48, no. (3), pp. 259–302, 1986.

[52] N. H. Clinthorne, T. S. Pan, P. C. Chiao, W. L. Rogers, and J. A. Stamos, "Preconditioning methods for improved convergence rates in iterative reconstructions," *IEEE Trans. Medical Imaging*, vol. 12, no. 1, pp. 78–83, Mar. 1993.

[53] P. J. Green, "Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives." *J. Roy. Statist. Soc. Series B*, vol. 46, no. 2, pp. 149–192, 1984.

[54] D. S. Lalush and B.M.W. Tsui, "A fast and stable weighted-least squares MAP conjugate-gradient algorithm for SPECT," *J. Nucl. Medicine*, vol. 34, no. 5, p. 27, May 1993.