

# Hybrid Poisson/Polynomial Objective Functions for Tomographic Image Reconstruction from Transmission Scans

Jeffrey A. Fessler, *Member, IEEE*

**Abstract**—This paper describes rapidly converging algorithms for computing attenuation maps from Poisson transmission measurements using penalized-likelihood objective functions. We demonstrate that an under-relaxed cyclic coordinate-ascent algorithm converges faster than the convex algorithm of Lange [1], which in turn converges faster than the expectation-maximization (EM) algorithm for transmission tomography [1]. To further reduce computation, one could replace the log-likelihood objective with a quadratic approximation. However, we show with simulations and analysis that the quadratic objective function leads to biased estimates for low-count measurements. Therefore we introduce hybrid Poisson/polynomial objective functions that use the exact Poisson log-likelihood for detector measurements with low counts, but use computationally efficient quadratic or cubic approximations for the high-count detector measurements. We demonstrate that the hybrid objective functions reduce computation time without increasing estimation bias.

## I. INTRODUCTION

MANY important medical problems occur in the human thorax, such as breast cancer, heart disease, and lung cancer. To produce quantitatively accurate images of physiology within the thorax using emission computed tomography, one must correct for the effects of photon absorption or *attenuation* [2]. Conventional *calculated* methods for attenuation correction [2] are inappropriate in the thorax, due to the nonuniform attenuation properties of bone, lungs, and soft tissue. Thus, most positron emission tomography (PET) centers have adopted the *measured* attenuation correction method, in which one precedes the emission scan with a transmission scan that measures the unique attenuation characteristics of each patient over the slices of interest [3]. Many SPECT centers are now using measured transmission scans for cardiac studies as well [4]–[7]. In PET and SPECT the primary medical interest is in physiology rather than anatomy. Thus the transmission scan is somewhat a measurement of “nuisance parameters,” so it is desirable to minimize its duration. Short scans yield noisy measurements, leading to noisy attenuation correction factors that propagate unwanted errors into the reconstructed emission image. This paper describes statistical methods for

reconstructing images of attenuation coefficient distributions (or *attenuation maps*) from noisy Poisson transmission scans.

The conventional method for attenuation correction in PET using measured transmission scans consists of two steps: first compute attenuation correction factors by smoothing the ratio of the *blank scan*<sup>1</sup> to the *transmission scan* measurements; then multiply the emission measurements by the attenuation correction factors in sinogram space. Therefore often no attenuation map is needed. However, in some cases it is desirable to reconstruct attenuation maps: for anatomical localization [8], for fully 3-D PET studies [9], for improved noise performance [10], and for quantitative SPECT [11]. In this paper we compare reconstruction methods in terms of the statistical accuracy of the reconstructed attenuation maps. These attenuation maps will often be reprojected to form attenuation correction factors, in which cases a more appropriate figure of merit would be the statistical accuracy of these correction factors, or of the final reconstructed emission images. Future studies should evaluate the overall emission/transmission process.

The conventional method for reconstructing attenuation maps from transmission measurements is a two-step process. First compute the logarithm of the ratio of the blank scan to the transmission scan, which gives a (noisy) estimate of the line integral of the attenuation distribution along each measurement ray. Then reconstruct the attenuation map by applying the conventional filtered backprojection (FBP) reconstruction method to the logarithmic data. As we describe in Appendix A, for Poisson data this logarithm of a ratio leads to biased estimates when the transmission counts are small (see Fig. 1). To avoid this bias problem, one must use a method that incorporates the measurement statistics [12], so we estimate the attenuation map using a penalized-likelihood approach. We compare several algorithms for maximizing the objective function: a cyclic-coordinate ascent algorithm based on the work of Sauer and Bouman [13]–[15], a convex algorithm and a gradient algorithm described by Lange [1], [16], [17], and the transmission EM algorithm [18], [19]. We demonstrate that the coordinate-ascent method converges significantly faster than the alternatives when implemented on a serial computer. The first part of this paper focuses on this comparison.

For routine clinical use, we would like still faster convergence, or less computation per iteration, even for the coordinate ascent algorithm. Therefore, in the second part of

Manuscript received July 22, 1994; revised November 8, 1994. This work was supported in part by NIH Grant CA-60711 and DOE Grant DE-FG02-87ER60561. The associate editor coordinating the review of this paper and approving it for publication was Prof. Jerry L. Prince.

The author was with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109-0552 USA.

IEEE Log Number 9413839.

<sup>1</sup> A blank scan is a transmission scan without the patient in the scanner that is acquired for the purpose of calibrating the measurements.

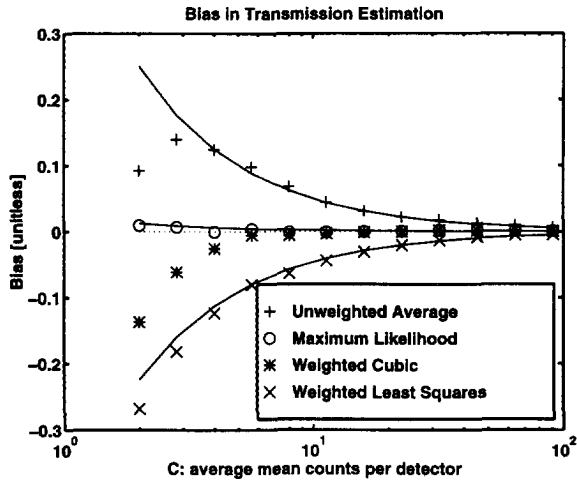


Fig. 1. Plots of estimator bias for the scalar transmission problem discussed in Appendix A. The unweighted average (i.e., FBP) estimator is positively biased and the weighted least-squares estimator is negatively biased, whereas the maximum likelihood estimator is nearly unbiased. The weighted cubic estimator is also nearly unbiased, except for very low count measurements.

this paper we propose hybrid Poisson/polynomial objective functions. The basis for these hybrids is very simple: since the transmission measurements are independent, the Poisson log-likelihood separates into a sum of terms corresponding to the marginal log-likelihood for each measurement. For measurements with only a few counts, we retain the corresponding marginal log-likelihood term. However, for measurements with a large number of counts, we replace the corresponding marginal log-likelihood term with a quadratic or cubic approximation closely related to the expansion of Sauer and Bouman [13]–[15]. These polynomial approximations significantly reduce the computation per iteration relative to the transcendental form of the log-likelihood for Poisson transmission measurements.

To summarize, the contributions of this paper are the following. We analyze the bias properties of transmission estimators (Appendix A). We show empirically that coordinate ascent converges faster than alternative algorithms, and that under-relaxation accelerates convergence (Sections II and III). We introduce the hybrid objective functions which lead to faster computation. Unlike the weighted least-squares objective, the hybrid approach properly treats nonpositive measurements (Section IV). We show that the resolution/noise tradeoff of penalized likelihood estimates outperforms the FBP method, and that the hybrid method has similar accuracy as penalized-likelihood but with much less computation (Section V).

## PART 1: PENALIZED MAXIMUM LIKELIHOOD

### II. OBJECTIVE AND ALGORITHMS

#### A. The Problem

Let  $\mu = [\mu_1, \dots, \mu_p]'$  denote the vector of unknown linear attenuation coefficients (having units of inverse length). Let  $y = [y_1, \dots, y_N]'$  denote the vector of measured transmission counts. We assume that the  $y_n$ 's are realizations of statistically independent random variables  $\{Y_n\}_{n=1}^N$  having Poisson

distributions with expectations  $\{\bar{y}_n(\mu)\}$ :

$$\bar{y}_n(\mu) = b_n e^{-l_n(\mu)} + r_n \quad (1)$$

where

$$l_n(\mu) = \sum_{j=1}^p a_{nj} \mu_j.$$

The  $b_n > 0$  factors denote the blank scan mean counts. The  $r_n \geq 0$  factors represent additive background events such as random coincidences (PET) [19], scatter (SPECT, PET, X-ray CT) [19], [20], or energy channel cross-talk (SPECT) [21]. The  $a_{nj} \geq 0$  factors have units of length, and describe the tomographic system geometry. Thus  $l_n(\mu)$  is (approximately) the line integral of the attenuation distribution along the  $n$ th ray. We assume that  $\{a_{nj}\}$ ,  $\{b_n\}$ , and  $\{r_n\}$  are known constants, so our goal is to estimate the unknown attenuation map  $\mu$  from the measurements  $y$ . This problem is relatively challenging due to the nonlinear relationship in (1).

Under the Poisson model, the log-likelihood for  $\mu$  is:

$$\begin{aligned} L(\mu) &= \log p(Y = y; \mu) \\ &= \sum_n y_n \log \bar{y}_n(\mu) - \bar{y}_n(\mu). \end{aligned} \quad (2)$$

(We ignore constants independent of  $\mu$  throughout.) Since image reconstruction is ill-conditioned, rather than estimating  $\mu$  by maximizing  $L(\mu)$ , we include a roughness penalty of the form:

$$P(\mu) = \sum_{j=1}^p \frac{1}{2} \sum_{k \in \mathcal{N}_j} w_{jk} \phi(\mu_j - \mu_k) \quad (3)$$

where  $\mathcal{N}_j$  is a neighborhood of the  $j$ th pixel,  $\phi$  is a symmetric, twice-differentiable scalar function, and  $w_{jk} = w_{kj} \geq 0$ . In Appendix B we specify the  $w_{jk}$  factors using a method that leads to uniform image resolution when  $\phi$  is the quadratic function  $\phi(x) = x^2/2$ . If one adopts the conventional choice for  $w_{jk}$ , then the reconstructed image resolution will be nonuniform [22].

Combining the penalty (3) and the log-likelihood (2) yields the penalized-likelihood objective function:

$$\Phi(\mu) = L(\mu) - \beta P(\mu). \quad (4)$$

We describe methods for choosing  $\beta > 0$  to achieve a desired resolution in [22]. Our goal is to estimate  $\mu$  by maximizing  $\Phi(\mu)$  subject to the nonnegativity constraint:

$$\hat{\mu} = \arg \max_{\mu \geq 0} \Phi(\mu). \quad (5)$$

In the remainder of this section we discuss iterative algorithms for performing this maximization.

#### B. Concavity and Convergence

The second partials of the log-likelihood (2) are

$$-\frac{\partial^2}{\partial \mu_j \partial \mu_k} L(\mu) = \sum_n a_{nj} a_{nk} \cdot \left[ 1 - \frac{y_n r_n}{\bar{y}_n(\mu)^2} \right] b_n e^{-l_n(\mu)}. \quad (6)$$

Therefore,  $L(\mu)$  is concave if  $r_n = 0$  for all  $n$  [18]. However, if  $r_n \neq 0$  then  $L(\mu)$  is *not* globally concave, so it is difficult to establish global convergence for any algorithms. Previous global convergence proofs for transmission algorithms have relied heavily on the use of concavity [1], [17], [18], [23]. From (6) one sees that  $L(\mu)$  is locally concave over the set  $\{\mu: \bar{y}_n^2(\mu) \geq y_n r_n \forall n\}$ . Fortunately, in PET and SPECT the  $r_n$  factors are fairly small, and since  $y_n \approx \bar{y}_n(\hat{\mu})$  and  $\bar{y}_n(\mu) > r_n$ , we will usually operate in or near the concave region of  $L(\mu)$ . All of the algorithms given below should have local convergence, which may be adequate since in practice one can initialize the iteration with a FBP image which is usually fairly close to the maximum. Nevertheless, further investigation of the convergence properties is needed for the case  $r_n \neq 0$ .

### C. Cyclic Coordinate Ascent Algorithm

Cyclic coordinate-ascent is a particularly simple approach to maximizing an objective function: simply update the parameters one by one while holding the others fixed, always using the most recent value for each parameter. This method is inherently sequential, and is more suited to conventional serial workstations than to fine-grain or mesh parallel computers. Although this approach may have poor repute in the optimization literature (e.g., [24, p. 310]), Sauer and Bouman have shown that it is remarkably effective for tomography [13]–[15]. Advantages of coordinate-ascent methods include: fast convergence for high spatial frequencies, easily enforced nonnegativity constraints, and decoupled parameter updates even when using smoothing penalties. In [25] we extended the analysis of [13] to show that *under-relaxation* can further improve the convergence rates of coordinate-ascent methods, so here we present the under-relaxed version.

Unfortunately, for transmission tomography there is no analytical closed form for the exact maximizer of  $\Phi$  over  $\mu_j$  even when holding the other parameters fixed, so one must resort to approximate methods. If the penalty is quadratic, we apply one iteration of a 1-D Newton's method with an under-relaxation parameter, i.e., we cyclically update the parameters using the following iteration:

$$\mu_j^{\text{new}} = \left[ \mu_j^{\text{old}} + \omega \frac{\frac{\partial}{\partial \mu_j} \Phi(\mu^{\text{old}})}{-\frac{\partial}{\partial \mu_j^2} \Phi(\mu^{\text{old}})} \right]_+ \quad (7)$$

where  $\omega \in (0, 1]$  is a relaxation parameter, and  $[x]_+ = x$  if  $x > 0$  and is 0 otherwise. One iteration of this algorithm updates all parameters in some order; we alternate between four raster scan orders to improve the convergence rate [13]. The updates are done "in place," always using the most recent estimate of  $\mu_j$ . For nonquadratic penalties, we use the efficient 1-D line-search method of Bouman and Sauer [14], [15].

Since  $\Phi$  is not quadratic, (7) does not guarantee monotonic increases in  $\Phi$ . If  $\mu^*$  satisfies the Karush–Kuhn–Tucker

conditions for (5), i.e.

$$\begin{cases} \frac{\partial}{\partial \mu_j} \Phi(\mu^*) = 0, & \mu_j^* > 0 \\ \frac{\partial}{\partial \mu_j} \Phi(\mu^*) \leq 0, & \mu_j^* = 0 \end{cases}$$

and if  $\Phi$  is strictly concave in a neighborhood of  $\mu^*$ , then  $\mu^*$  is a fixed point of (7), and using continuity one can prove local convergence of the iteration (7) to  $\mu^*$ . For reasons of convergence rate [25]–[27], we usually use  $\omega = 0.6$ . This under-relaxation improves the odds that (7) will yield an increase in  $\Phi$ . With  $\omega = 0.6$  we have never observed a decrease in  $\Phi$  over a full iteration, although we have observed small decreases with  $\omega = 1$ . Fortuitously, using  $\omega < 1$  not only improves the convergence rate, it appears to improve the monotonicity as well.

For the objective function (4), one can show that

$$\begin{aligned} \frac{\partial}{\partial \mu_j} \Phi(\mu) &= \dot{L}_j(\mu) - \beta \dot{P}_j(\mu) \\ -\frac{\partial^2}{\partial \mu_j^2} \Phi(\mu) &= -\ddot{L}_j(\mu) + \beta \ddot{P}_j(\mu) \end{aligned}$$

where

$$\begin{aligned} \dot{L}_j(\mu) &= \sum_n a_{nj} \left[ 1 - \frac{y_n}{\bar{y}_n(\mu)} \right] b_n e^{-l_n(\mu)} \quad (8) \\ -\ddot{L}_j(\mu) &= \sum_n a_{nj}^2 \left[ 1 - \frac{y_n r_n}{\bar{y}_n(\mu)^2} \right] b_n e^{-l_n(\mu)} \\ \dot{P}_j(\mu) &= \sum_{k \in \mathcal{N}_j} w_{jk} \dot{\phi}(\mu_j - \mu_k) \\ \ddot{P}_j(\mu) &= \sum_{k \in \mathcal{N}_j} w_{jk} \ddot{\phi}(\mu_j - \mu_k) \quad (9) \end{aligned}$$

and where  $\dot{\phi}(x) = (d/dx)\phi(x)$  and  $\ddot{\phi}(x) = (d^2/dx^2)\phi(x)$ . Substituting these formulas into (7) yields the cyclic coordinate-ascent algorithm. To minimize the computation in calculating (8) and (9) we maintain the current state [13] of the line-integrals during the iterations, i.e., after updating pixel  $j$ :

$$l_n(\mu^{\text{new}}) := l_n(\mu^{\text{old}}) + a_{nj}(\mu_j^{\text{new}} - \mu_j^{\text{old}}), \quad \forall n: a_{nj} \neq 0.$$

### D. Convex Algorithm

De Pierro [28], [29] has analyzed the emission EM algorithm using convexity, and Lange [1], [16] has adapted De Pierro's method to the transmission problem. Here we extend [1] slightly to include nonzero background events ( $r_n > 0$ ).

Since the measurements are independent, the log-likelihood separates into a sum of  $N$  terms:

$$L(\mu) = \sum_n h_n [l_n(\mu)] \quad (10)$$

where from (1) and (2)

$$h_n(l) = y_n \log(b_n e^{-l} + r_n) - (b_n e^{-l} + r_n).$$

If we define

$$l_n^{\max} = \begin{cases} \infty, & r_n \geq \sqrt{y_n r_n} \\ \log \left( \frac{b_n}{\sqrt{y_n r_n} - r_n} \right), & \text{otherwise} \end{cases}$$

then  $h_n(\cdot)$  is strictly concave on the interval  $(-\infty, l_n^{\max})$ . Following [1], if  $\mu_j^i > 0$  then the inequality

$$\begin{aligned} L(\mu) &= \sum_n h_n \left[ \sum_{j=1}^p \frac{a_{nj} \mu_j^i}{l_n(\mu^i)} \frac{\mu_j}{\mu_j^i} l_n(\mu^i) \right] \\ &\geq \sum_n \sum_{j=1}^p \frac{a_{nj} \mu_j^i}{l_n(\mu^i)} h_n \left[ \frac{\mu_j}{\mu_j^i} l_n(\mu^i) \right] \\ &\triangleq Q^*(\mu; \mu^i) \end{aligned}$$

holds for any  $\mu$  in the set<sup>2</sup>:

$$\mathcal{L}_{\mu^i} = \{ \mu \geq 0: \frac{\mu_j}{\mu_j^i} l_n(\mu^i) \leq l_n^{\max}, \quad \forall j, n \}.$$

As shown by De Pierro [28], [29], a similar decomposition applies to  $P(\mu)$ . If  $\phi$  is convex, then

$$\begin{aligned} P(\mu) &= \sum_j \sum_{k \in \mathcal{N}_j} \frac{w_{jk}}{2} \\ &\quad \cdot \phi \left( \frac{2\mu_j - \mu_j^i - \mu_k^i}{2} + \frac{-2\mu_k + \mu_j^i + \mu_k^i}{2} \right) \\ &\leq \sum_j \sum_{k \in \mathcal{N}_j} \frac{w_{jk}}{4} \\ &\quad \cdot [\phi(2\mu_j - \mu_j^i - \mu_k^i) + \phi(-2\mu_k + \mu_j^i + \mu_k^i)] \\ &\triangleq P^*(\mu; \mu^i) \end{aligned}$$

holds for any  $\mu$ . Thus if we define

$$\Phi^*(\mu; \mu^i) = Q^*(\mu; \mu^i) - \beta P^*(\mu; \mu^i) \quad (11)$$

then one can show [1], [28], [29] that for  $\mu \in \mathcal{L}_{\mu^i}$ :

$$\Phi(\mu) - \Phi(\mu^i) \geq \Phi^*(\mu; \mu^i) - \Phi^*(\mu^i; \mu^i). \quad (12)$$

Therefore, if we choose  $\mu^{i+1}$  to maximize  $\Phi^*(\cdot; \mu^i)$ , then if  $\mu^{i+1} \in \mathcal{L}_{\mu^i}$  we ensure monotonic increases in  $\Phi$ . If  $r_n \neq 0$ , then the inequality in (12) does not necessarily hold globally, and monotonicity is not guaranteed intrinsically. Our implementation always checks (12) to ensure monotonicity, and even with as much as 10% background events we have never observed a violation of (12) when the algorithm is initialized with a FBP image. As discussed in Section II-B, apparently, one usually operates in the concave part  $\mathcal{L}_{\mu^i}$  of  $Q^*(\mu; \mu^i)$ .

Unlike  $\Phi(\mu)$ , the function  $\Phi^*(\mu; \mu^i)$  is separable in  $\mu_j$ , so we can find the  $\mu^{i+1}$  that maximizes  $\Phi^*(\cdot; \mu^i)$  easily using  $p$  separate 1-D maximizations, which is ideal for parallel processors. Again there is no analytical form for this maximization,

<sup>2</sup>Note that in the simple case where  $r_n = 0$ , then  $l_n^{\max} = \infty$  and  $\mathcal{L}_{\mu^i}$  is the entire nonnegative orthant.

so as in [1] we apply one step of Newton's method to each parameter. The partials of  $Q^*$  and  $P^*$  are

$$\begin{aligned} \frac{\partial}{\partial \mu_j} Q^*(\mu; \mu^i) \Big|_{\mu=\mu^i} &= \dot{L}_j(\mu^i) \\ -\frac{\partial^2}{\partial \mu_j^2} Q^*(\mu; \mu^i) \Big|_{\mu=\mu^i} &= \frac{1}{\mu_j^i} \sum_n a_{nj} q_n(\mu^i) \\ \frac{\partial}{\partial \mu_j} P^*(\mu; \mu^i) \Big|_{\mu=\mu^i} &= \dot{P}_j(\mu^i) \\ \frac{\partial^2}{\partial \mu_j^2} P^*(\mu; \mu^i) \Big|_{\mu=\mu^i} &= 2\ddot{P}_j(\mu^i) \end{aligned} \quad (13)$$

(see (8)), where we precompute the terms

$$q_n(\mu^i) = l_n(\mu^i) \left[ 1 - \frac{y_n r_n}{y_n^2(\mu^i)} \right] b_n e^{-l_n(\mu^i)}$$

prior to computing the sums in (13). Combine the above partials with (11) to yield the iteration:

$$\mu_j^{i+1} = \max \left\{ \mu_j^i + \omega^i \frac{\frac{\partial}{\partial \mu_j} \Phi^*(\mu^i; \mu^i)}{-\frac{\partial^2}{\partial \mu_j^2} \Phi^*(\mu^i; \mu^i)}, \epsilon^i \right\},$$

$$j = 1, \dots, p \quad (14)$$

where  $\epsilon^i \rightarrow 0$  as  $i \rightarrow \infty$ , and where  $\omega^i$  is chosen to assure monotonicity by starting with  $\omega^i = 1$  and then if necessary repeatedly halving it until  $\Phi(\mu^{i+1}) > \Phi(\mu^i)$ . The key difference between (7) and (14) is that the latter uses a *simultaneous* update, and as such it is more amenable to parallel implementations (although the step of choosing  $\omega^i$  to ensure monotonicity may not parallelize easily).

### E. Other Algorithms

The EM algorithm [1], [18], [19], [30], and the gradient algorithm of Lange [1], [17], [23] are two alternative algorithms for penalized maximum likelihood. The straightforward extensions of these algorithms to include nonzero background events are described in [31].

## III. CONVERGENCE RATE SIMULATIONS

This section describes some representative simulations that demonstrate that the coordinate ascent algorithm converges faster than the convex algorithm, as well as the alternative algorithms described in [1]. For  $\mu$  we used the synthetic attenuation map shown in Fig. 2, which represents a human thorax with linear attenuation coefficients 0.0165/mm, 0.0096/mm, and 0.0025/mm, for bone, soft tissue, and lungs, respectively. The image was a 128 by 64 array of 4.5 mm pixels. We simulated a PET transmission scan with 192 radial bins and 256 angles uniformly spaced over 180°. The  $a_{nj}$  factors correspond to 6 mm wide strip integrals on 3 mm center-to-center spacing. (This is an approximation to the ideal line integral that accounts for finite detector width.) The  $b_n$  factors were generated using pseudo-random log-normal variates with a standard deviation of 0.3 to account for detector efficiency variations, and scaled so that  $\sum_n \bar{y}_n$  was one million counts. The  $r_n$  factors corresponded to a uniform

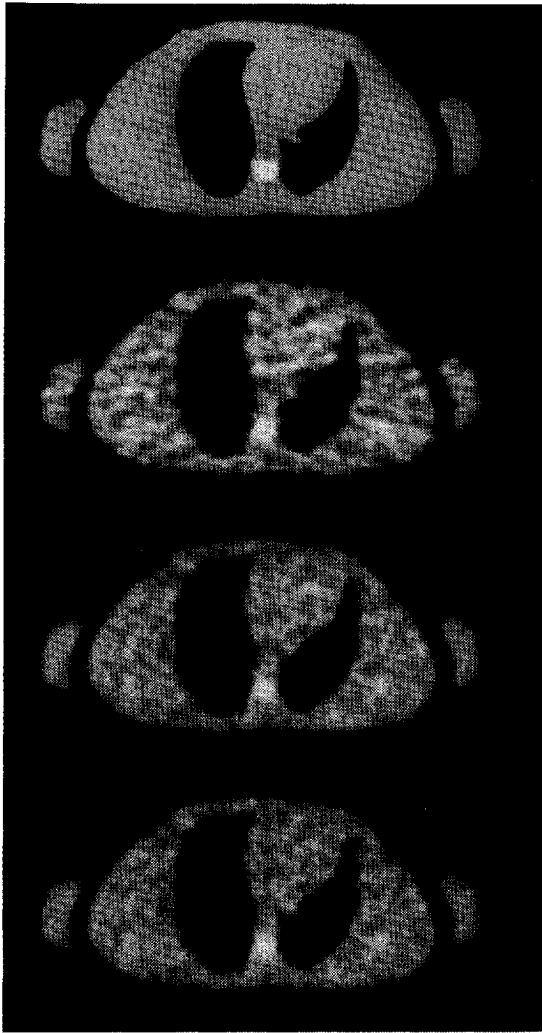


Fig. 2. From top to bottom: digital thorax phantom, typical filtered back-projection reconstructed image, result of 10 coordinate ascent iterations for the penalized likelihood objective, result of 10 coordinate ascent iterations for the hybrid objective with  $\gamma_a = 5$  and  $\gamma_b = 50$ . The bottom two images are visually indistinguishable because the hybrid objective is an accurate approximation to the log-likelihood.

field of 10% random coincidences. Pseudo-random Poisson transmission measurements were generated according to (1). The image shown in Fig. 2 was reconstructed using FBP with a second order Butterworth filter at a resolution of 2.5 pixels or 1.125 cm full-width half-maximum (FWHM). This FBP image was used to initialize the iterative algorithms, after setting all negative pixels to zero. For the iterative methods, we used the function  $\phi(x) = x^2/2$  for the penalty in (3).

Fig. 3 compares the rates of convergence for the coordinate ascent algorithm with and without under-relaxation. These typical results confirm that under-relaxation accelerates convergence for the transmission problem just as it does in the emission case.

Fig. 4 shows a plot of the objective increase  $\Phi(\mu^i) - \Phi(\mu^0)$  versus iteration  $i$  for the penalized maximum likelihood algorithms. The computation time per iteration varies between algorithms, so a more objective comparison is shown in Fig. 5, in which the abscissa is CPU time as measured on

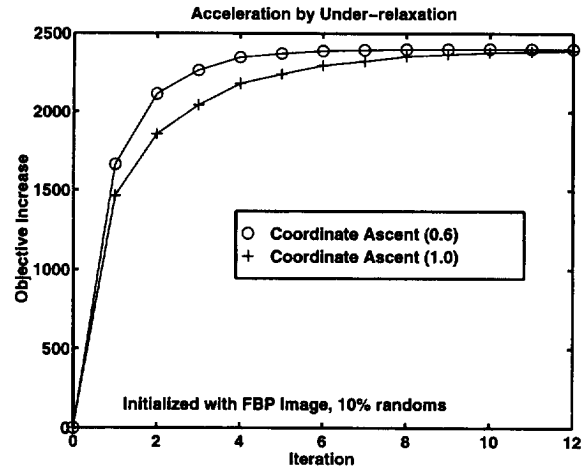


Fig. 3. Plots of the objective  $\Phi(\mu_i) - \Phi(\mu^0)$  versus iteration for the coordinate ascent algorithm with  $\omega = 0.6$  and  $\omega = 1.0$ . Underrelaxation accelerates the convergence rate of coordinate ascent.

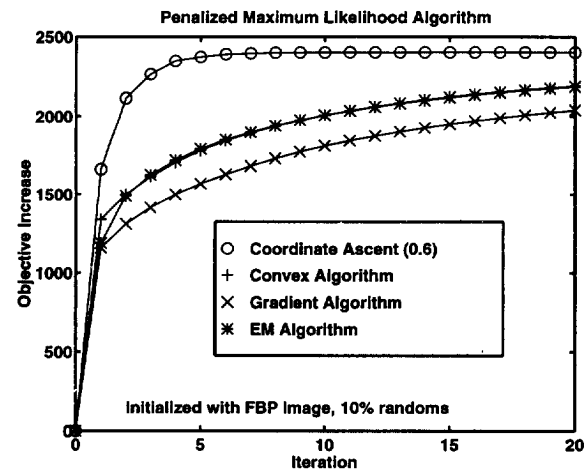


Fig. 4. Plots of the objective  $\Phi(\mu_i) - \Phi(\mu^0)$  versus iteration for the penalized likelihood algorithms described in the text. The coordinate ascent algorithm converges fastest.

a DEC 3000/800. The coordinate ascent algorithm effectively converges (i.e.,  $\Phi(\mu^i)$  reaches 99.9% of its peak value) in less than 10 iterations, whereas the other algorithms require many more iterations. From Fig. 5, one sees that it may be advantageous to run two or three iterations of the convex algorithm and then switch to coordinate ascent.

## PART 2: HYBRID OBJECTIVES

### IV. POLYNOMIAL APPROXIMATIONS

The results given above demonstrate that the under-relaxed coordinate-ascent method for transmission tomography converges faster than the alternative algorithms, in terms of both number of iterations and CPU time. Since the coordinate-ascent algorithm usually converges in seven to ten iterations, it will be challenging to further reduce the number of iterations. However, it would be desirable to further reduce the computation *per iteration*, in light of the complexity of (8)–(9). One could reduce computation by replacing the

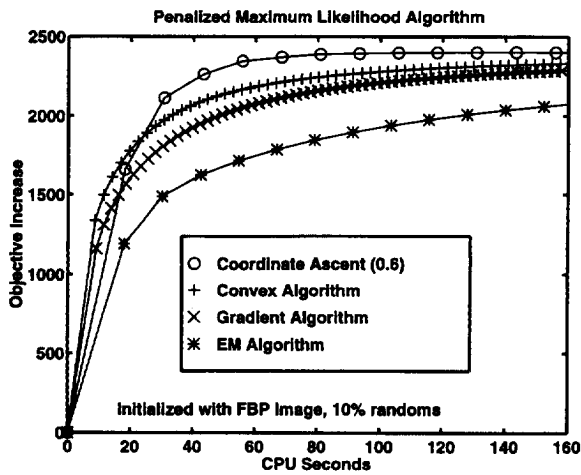


Fig. 5. Plots of the objective  $\Phi(\mu_i) - \Phi(\mu^0)$  versus CPU seconds (DEC 3000/800) for the penalized likelihood algorithms described in the text. The coordinate ascent algorithm converges fastest despite using more CPU time per iteration.

log-likelihood with its quadratic approximation [13]–[15], which simplifies the calculations corresponding to (8)–(9). As shown in Appendix A, the quadratic approximation leads to systematic negative bias. However, the bias decreases as the number of counts increase. Therefore, in this section we propose a hybrid Poisson/polynomial objective function in which we replace only the marginal log-likelihood terms for the high-count measurements with polynomial approximations, whereas for the low-count measurements we retain the exact log-likelihood. As we show in Section V, this approximation significantly reduces computation with negligible bias.

The quadratic approximation of Sauer and Bouman [13]–[15] disregarded the effects of nonzero background events, so we begin by extending their expansion slightly to accommodate  $r_n > 0$ . Since the Poisson measurements are independent, the log-likelihood separates into a sum of  $N$  terms:

$$L(\mu) = \sum_n h_n[l_n(\mu)] \quad (15)$$

where

$$\begin{aligned} h_n(l) &= y_n \log g_n(l) - g_n(l) \\ g_n(l) &= b_n e^{-l} + r_n. \end{aligned}$$

Applying Taylor's expansion to  $h_n(l)$  about some value  $\hat{l}_n$

$$\begin{aligned} h_n(l) &\approx h_n(\hat{l}_n) + h_n^{(1)}(\hat{l}_n)(l - \hat{l}_n) \\ &\quad + \frac{h_n^{(2)}(\hat{l}_n)}{2}(l - \hat{l}_n)^2 + \frac{h_n^{(3)}(\hat{l}_n)}{6}(l - \hat{l}_n)^3 \end{aligned} \quad (16)$$

where  $h_n^{(i)}(l) = (\partial^i / \partial l^i) h_n(l)$  and

$$h_n^{(1)}(l) = \left[ 1 - \frac{y_n}{g_n(l)} \right] b_n e^{-l} \quad (17)$$

$$-h_n^{(2)}(l) = \left[ 1 - \frac{y_n r_n}{g_n^2(l)} \right] b_n e^{-l}$$

$$h_n^{(3)}(l) = \left[ 1 + r_n y_n \frac{b_n e^{-l} - r_n}{g_n^3(l)} \right] b_n e^{-l}. \quad (18)$$

Now consider a measurement for which  $y_n > r_n$ , and define the following method-of-moments estimate of the line integral of attenuation  $\hat{l}_n$ :

$$\hat{l}_n = \log \left( \frac{b_n}{y_n - r_n} \right). \quad (19)$$

This estimate satisfies the equality  $g_n(\hat{l}_n) = y_n$ . Substituting into (17)–(18) yields

$$\begin{aligned} h_n^{(1)}(\hat{l}_n) &= 0 \\ -h_n^{(2)}(\hat{l}_n) &= \frac{(y_n - r_n)^2}{y_n} \triangleq u_n \\ h_n^{(3)}(\hat{l}_n) &= y_n + \frac{r_n^2}{y_n^2} (2r_n - 3y_n) \triangleq t_n. \end{aligned} \quad (20)$$

Substituting into (16) thus shows that the  $n$ th term in (15) is approximately:

$$h_n(l) \approx (y_n \log y_n - y_n) - \frac{u_n}{2} (l - \hat{l}_n)^2 + \frac{t_n}{6} (l - \hat{l}_n)^3. \quad (21)$$

#### A. Purely Quadratic Approximation

Since the first term in (21) is independent of  $l$ , we can disregard it, and then substituting into (15) and dropping the cubic term yields the following objective function, which is based on a purely quadratic approximation to the log-likelihood:

$$\Phi_q(\mu) = \sum_{n: y_n > r_n} -\frac{u_n}{2} [l_n(\mu) - \hat{l}_n]^2 - \beta P(\mu). \quad (22)$$

We refer to the attenuation map obtained by maximizing  $\Phi_q(\mu)$  over  $\mu \geq 0$  as the *penalized weighted least-squares (PWLS) estimate*. As shown in Appendix A, this type of estimator is biased.

#### B. Polynomial Approximation

To remedy this bias problem, consider the following hybrid objective function. First pick two count thresholds  $\gamma_b \geq \gamma_a \geq 0$  and split the measurements into three groups:

$$\begin{aligned} N_L &= \{n: y_n - r_n \leq \gamma_a\} \\ N_M &= \{n: \gamma_a < y_n - r_n < \gamma_b\} \\ N_H &= \{n: y_n - r_n \geq \gamma_b\}. \end{aligned}$$

Note that  $N_L \cup N_M \cup N_H = \{1, 2, \dots, N\}$ . For  $n \in N_H$  (high count measurements) we use the quadratic approximation to  $h_n(l)$ . For  $n \in N_M$  (medium count measurements) we use the cubic approximation to  $h_n(l)$ . For  $n \in N_L$  (low count measurements) we retain the original log-likelihood term  $h_n(l)$ . This leads to the following *hybrid Poisson/polynomial objective function*:

$$\Phi_h(\mu) = L_L(\mu) + L_M(\mu) + L_H(\mu) - \beta P(\mu) \quad (23)$$

where

$$L_L(\mu) = \sum_{n \in N_L} y_n \log \bar{y}_n(\mu) - \bar{y}_n(\mu)$$

$$L_M(\mu) = \sum_{n \in N_M} -\frac{u_n}{2} [l_n(\mu) - \hat{l}_n]^2 + \frac{t_n}{6} [l_n(\mu) - \hat{l}_n]^3$$

$$L_H(\mu) = \sum_{n \in N_H} -\frac{u_n}{2} [l_n(\mu) - \hat{l}_n]^2.$$

This hybrid objective function includes the penalized likelihood objective as a special case by choosing  $\gamma_a = \gamma_b = \infty$ . If one chooses  $\gamma_a = \gamma_b = 0$ , then one obtains the penalized weighted least-squares objective (22), unless there are any  $n$  such that  $y_n \leq r_n$ . In that case, unlike the purely quadratic approximation given by (22), which disregards any measurements such that  $y_n \leq r_n$ , the hybrid objective function retains the log-likelihood corresponding to those terms, since  $\gamma_a \geq 0$ . By varying  $\gamma_a$  and  $\gamma_b$ , one can compromise between computation time and strict faithfulness to the log-likelihood.

One could imagine a variety of algorithms for maximizing  $\Phi_h(\mu)$  subject to the constraint  $\mu \geq 0$ . Based on the results above and in [13], [25], we again use the under-relaxed cyclic coordinate-ascent algorithm given by (7), except here applied to  $\Phi_h(\mu)$ . From (7) one sees that we need expressions for the partial derivatives of  $\Phi_h(\mu)$ . The first and second partials of  $L_L(\mu)$  are given by the corresponding terms of the sums in (8)–(9). The partials of  $L_M(\mu)$  are

$$\begin{aligned} \frac{\partial}{\partial \mu_j} L_M(\mu) &= \sum_{n \in N_M} a_{nj} \left( u_n [\hat{l}_n - l_n(\mu)] + \frac{t_n}{2} [\hat{l}_n - l_n(\mu)]^2 \right) \\ -\frac{\partial^2}{\partial \mu_j^2} L_M(\mu) &= \sum_{n \in N_M} a_{nj}^2 \{ u_n + t_n [\hat{l}_n - l_n(\mu)] \} \end{aligned} \quad (24)$$

and the partials of  $L_H(\mu)$  are

$$\begin{aligned} \frac{\partial}{\partial \mu_j} L_H(\mu) &= \sum_{n \in N_H} a_{nj} u_n [\hat{l}_n - l_n(\mu)] \\ -\frac{\partial^2}{\partial \mu_j^2} L_H(\mu) &= \sum_{n \in N_H} a_{nj}^2 u_n. \end{aligned} \quad (25)$$

Prior to iterating, we precompute the  $u_n$  and  $t_n$  terms, as well as the  $\sum a_{nj}^2 u_n$  terms, all of which are independent of  $\mu$ . The computational savings result both from this precomputation as well as from the fact that the above derivatives require no exponentials.

Combining (23) with the above expressions for the partial derivatives yields all of the terms necessary to perform the coordinate ascent update given by (7). The complete mathematical expression is more daunting than the actual software implementation, which is fairly straightforward if one has already implemented coordinate ascent iterations for either the pure penalized likelihood or penalized least-squares objectives.

Convergence of the coordinate ascent algorithm has the same caveats described in Section II. However, if  $\gamma_a = \gamma_b = 0$  and the potential function  $\phi$  is quadratic, then the entire objective function is quadratic if  $y_n > r_n \forall n$ . In this special case the coordinate ascent algorithm is guaranteed to converge monotonically to the unique global maximum [25].

TABLE I

PERCENT BIAS AND PERCENT ERROR STANDARD DEVIATION FOR RECTANGULAR REGIONS OF INTEREST TAKEN FROM THE PHANTOM SHOWN IN FIG. 2. THE CPU TIMES ARE IN SECONDS FOR 10 ITERATIONS (FOR THE ITERATIVE METHODS) ON A DEC 3000/800. THE CPU-2 TIMES ARE FOR USING THE QUADRATIC APPROXIMATION IN THE DENOMINATOR OF (7), WHEREAS THE CPU-1 TIMES ARE FOR USING THE SECOND PARTIALS OF THE OBJECTIVE (SEE SECTION IV-C)

	FBP	PML	Hybrid ( $\gamma_a, \gamma_b$ )			PWLS
			(5, 50)	(5, 20)	(5, 5)	
% Std. Dev.	12.0	6.85	6.87	7.14		7.32
% Soft Tissue Bias	4.5	-0.2	0.2	-0.3	-3.2	-5.8
% Lung Tissue Bias	3.0	0.5	0.7	0.9	0.3	-4.0
% Bone Tissue Bias	2.6	-0.7	0.0	-0.3	-4.6	-8.3
CPU-1 Seconds	1.6	136	50	47		23
CPU-2 Seconds		119	46	44	38	
% $N_L$		100	5	5	5	0
% $N_M$		0	50	25		0
% $N_H$		0	45	70	95	100

### C. Denominator Approximation

The denominator of the coordinate ascent algorithm (7) using Newton's method is usually taken to be the second partial derivative of the objective function  $(\partial^2 / \partial \mu_j^2) \Phi(\mu^i)$ . There is no guarantee that Newton's method will ensure monotone increases in  $\Phi$ , but again, underrelaxation seems to help both convergence rate and monotonicity. Therefore it is natural to ask: is it really worth computing the second partials given by (9) and (24), or would the quadratic approximation (25) (summed over all  $n$ ) suffice?

We have implemented a modified hybrid coordinate ascent algorithm in which we use the proper first partial derivatives in the numerator of (7), but we replace the denominator with the quadratic approximation (25). The resulting estimates after 10 iterations differed by less than 0.1%, which is negligible relative to the noise.

The CPU times for 10 iterations of the various algorithms are given in Table I. Using the precomputed quadratic approximation in the denominator of (7) reduces CPU time by 5–15% for the case described in Section III.

## V. PERFORMANCE SIMULATIONS

To study the bias properties of the estimators described above, we performed additional simulations using the thorax phantom and PET system described in Section III. In this case we generated 50 independent realizations of the transmission measurements. For each measurement realization, we reconstructed an estimate of the attenuation map using 10 iterations of the coordinate ascent algorithms applied to the following objective functions: penalized likelihood (4), penalized weighted least squares (22), and the hybrid objective (23) with  $\gamma_a = 5$  and  $\gamma_b = 50$ . (We chose these values based on the results in Fig. 1, which suggest that the cubic approximation is reasonable above 5 counts, and the quadratic approximation is reasonable above 50 counts.) We also reconstructed each realization using the FBP method with the filter described in Section III, and we used those FBP reconstructed images to initialize the iterative algorithms in all cases.

Fig. 6 displays the sample mean and sample standard deviation images of the four methods. The sample means are visually very similar, since we have chosen  $\beta$  to match resolution [22]. The standard deviation of the FBP images

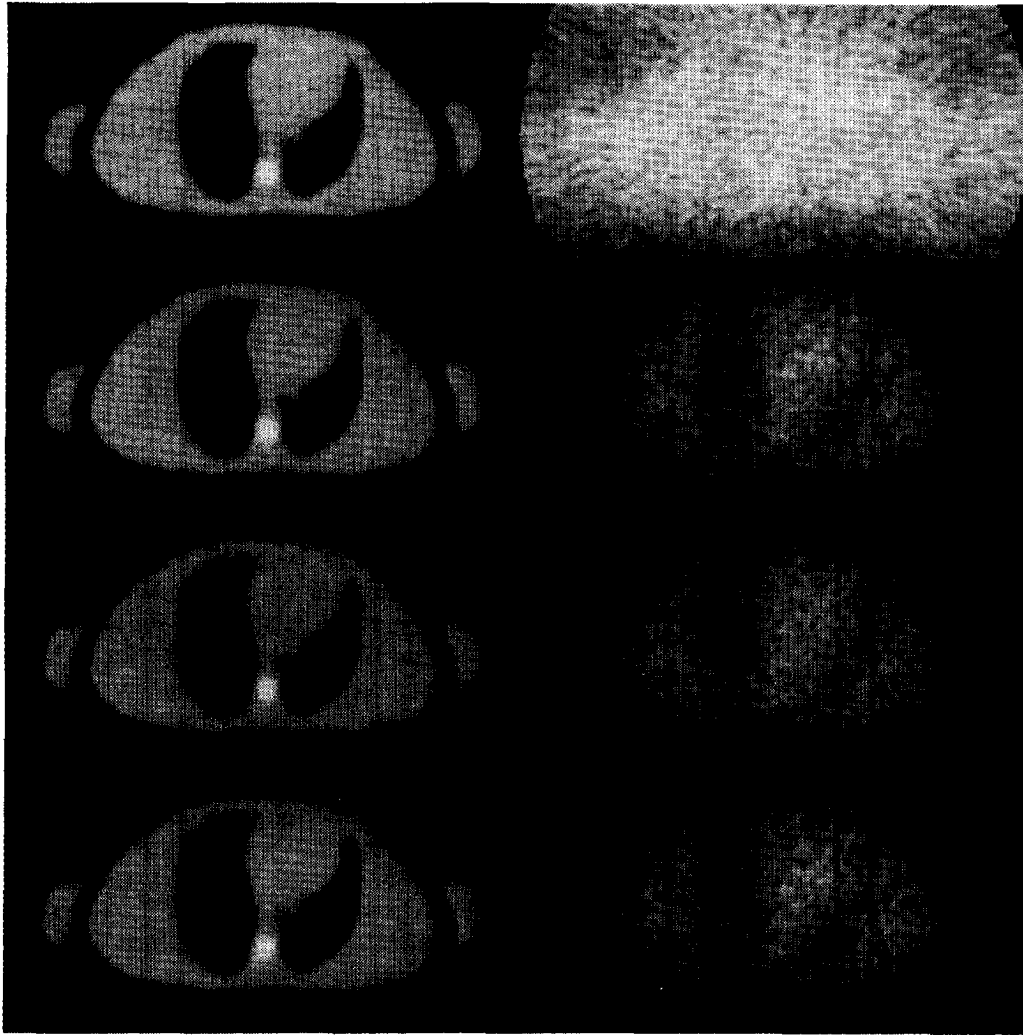


Fig. 6. Mean (left) and standard deviation (right) images from reconstructions computed from 50 realizations. From top to bottom: FBP method, penalized likelihood, penalized weighted least squares, and hybrid objective with  $\gamma_A = 5$  and  $\gamma_B = 50$ .

is fairly uniform, whereas those of the iterative methods look like a blurred version of the object in Fig. 2, i.e., the noise is spatially variant.

The gray scale in Fig. 6 does not reveal the bias properties, so Fig. 7 shows part of a horizontal profile through the sample mean images. This profile shows that FBP is positively biased and PWLS is negatively biased *even in a large uniform region*. Table I shows average percent bias in several rectangular regions. The bias of penalized likelihood and the hybrid method are negligible, whereas the bias of FBP and PWLS are significant.

In addition to reduced bias, the other benefit of statistical methods is an improved tradeoff between resolution and noise. To measure resolution, we created another object by adding several small point sources to the object in Fig. 2. We repeated the process described above for generating Poisson measurements, and reconstructed 50 realizations of this new object for each of the algorithms. Subtracting the sample means of the two cases (with and without the point sources) yields an image consisting only of the blurred point sources.

We averaged those point sources together to further reduce noise, and then computed the FWHM spread of the point sources using linear interpolation. This process was repeated for four values of  $\beta$  for the iterative methods and for four cutoff frequencies for the FBP method. Fig. 8 shows a plot of resolution versus noise standard deviation for the four methods. At any resolution over the range studied, the statistical methods show significantly smaller error standard deviations than the FBP method. Furthermore, the penalized likelihood method and the hybrid method were again indistinguishable, whereas the PWLS method showed a slightly higher noise variance for finer resolution images.

As shown in Table I, we also reconstructed images using the hybrid algorithm with smaller values for  $\gamma_a$  and  $\gamma_b$ . These choices yielded quantitative differences between the hybrid reconstructed images and the penalized likelihood images, with the former having either slightly higher variances or higher biases. Therefore for this data we prefer the higher value  $\gamma_b = 50$  since the additional computation (over  $\gamma_b = 20$ ) is small. Of course one will want to adjust this tradeoff



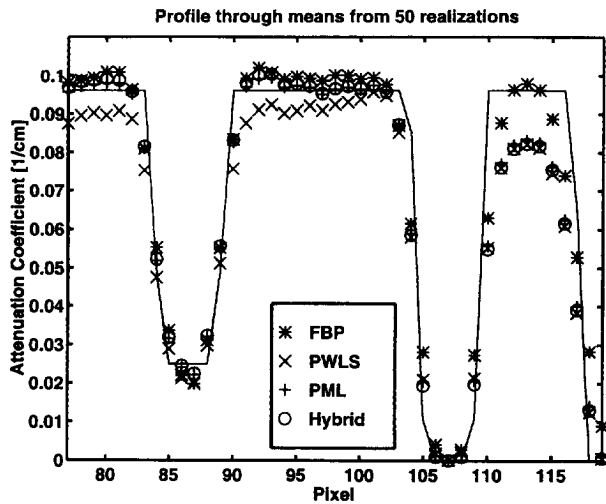


Fig. 7. Horizontal profile through the sample mean images shown in Fig. 6. The FBP method has a systematic positive bias and the penalized weighted least squares (PWLS) objective has a systematic negative bias. The penalized maximum likelihood (PML) and hybrid estimators perform comparably with negligible bias.

between computation and bias and variance for different types of transmission studies (based on total counts, etc.).

## VI. DISCUSSION

In Part 1 of this paper, we compared several algorithms for maximizing penalized-likelihood objective functions for transmission tomography. The coordinate ascent algorithm converged faster than three alternatives, which is consistent with previous work in emission tomography. Another alternative algorithm is the preconditioned conjugate gradient method developed by Mumcuoglu *et al.* [32], [33]. We have not implemented this method because it is more complicated due to the barrier functions that are needed to enforce the nonnegativity constraint. Sauer and Bouman found coordinate ascent to converge faster than conjugate gradient for emission tomography [13], although without using preconditioning. Except for the inconvenience of the nonnegativity constraint, conjugate gradients should be a natural algorithm to apply to the hybrid objective function since part of that objective is quadratic. For a meaningful comparison of coordinate ascent and preconditioned conjugate gradient for transmission tomography, the methods should be compared on the same data with the same convergence criteria.

In the second part of the paper we abandoned the pure log-likelihood in favor of partial approximations to it that reduce computation. We demonstrated that the purely quadratic approximation leads to systematic negative bias for low count PET transmission scans. By using a hybrid Poisson/polynomial objective we obtained reconstruction accuracy indistinguishable from penalized maximum likelihood, but with a factor of two less computing time. A statistical comparison of the resolution/noise tradeoffs demonstrated that all of the statistical methods improve signal to noise ratio relative to FBP.

Based on our experience, for serial workstations we recommend using the coordinate ascent algorithm for transmission

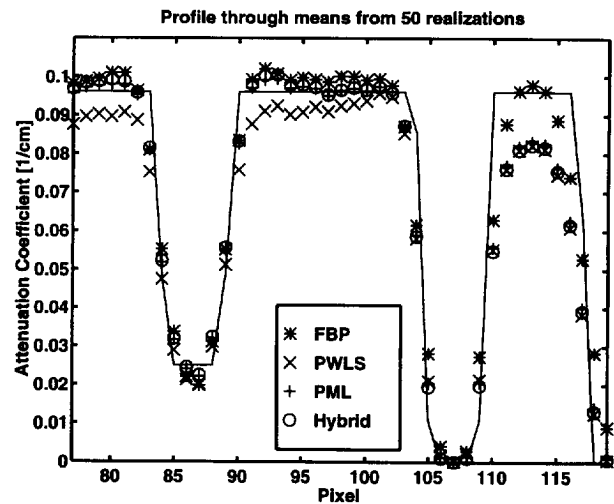


Fig. 8. Resolution versus noise for the various reconstruction algorithms. The FBP method has significantly greater noise than the statistical methods. For finer resolutions, the PWLS estimates are slightly noisier than the PML estimates. The performance of the hybrid objective with  $\gamma_A = 5$  and  $\gamma_B = 50$  is indistinguishable from penalized likelihood.

tomography with the hybrid objective function. For parallel processors the convex algorithm or conjugate gradients are probably preferable, and could easily be generalized to apply to the hybrid objective function.

## APPENDIX A BIAS

In this appendix, we analyze the bias and variance of several nonlinear estimators in the context of a scalar simplification of the transmission tomography measurement model. This analysis gives insight into the origin of the negative bias of the weighted least-squares estimator, the positive bias of the FBP method, and the near-unbiasedness of maximum likelihood or its cubic approximation.

Assume we have independent Poisson measurements  $Y = [Y_1, \dots, Y_N]$ :

$$Y_n \sim \text{Poisson} \{b_n e^{-\theta}\} \quad (26)$$

where  $\theta$  is an unknown parameter and  $b_n$  are known constants. We would like to estimate  $\theta$  from  $\{Y_n\}_{i=1}^N$ . Since there is only one unknown parameter  $\theta$ , this represents a scalar simplification of the transmission tomography problem. Nevertheless, this problem retains the nonlinearity associated with the exponential operation.

The log-likelihood for this problem is

$$L(\theta) = \sum_n Y_n \log(b_n e^{-\theta}) - b_n e^{-\theta}.$$

By zeroing the derivative of  $L(\cdot)$ , one can easily show that the maximum likelihood estimate of  $\theta$  is

$$\hat{\theta}_{\text{ML}} = g_{\text{ML}}(Y) = \log \left( \frac{\sum_n b_n}{\sum_n Y_n} \right). \quad (27)$$

The quadratic approximation to  $L(\cdot)$  (cf. (22)) is

$$\sum_{\{n: Y_n > 0\}} -\frac{Y_n}{2} \cdot \left[ \theta - \log \left( \frac{b_n}{Y_n} \right) \right]^2.$$

By zeroing the derivative of this approximation, one obtains the weighted least squares (LS) estimate of  $\theta$ :

$$\hat{\theta}_{\text{LS}} = g_{\text{LS}}(Y) = \frac{\sum_{\{n: Y_n > 0\}} Y_n \log \left( \frac{b_n}{Y_n} \right)}{\sum_{\{n: Y_n > 0\}} Y_n}. \quad (28)$$

The cubic approximation to  $L(\cdot)$  (cf. (21)) is

$$\sum_{\{n: Y_n > 0\}} -\frac{Y_n}{2} \cdot \left[ \theta - \log \left( \frac{b_n}{Y_n} \right) \right]^2 + \frac{Y_n}{6} \left[ \theta - \log \left( \frac{b_n}{Y_n} \right) \right]^3.$$

By zeroing the derivative of this approximation, one finds that the weighted cubic (WC) estimate of  $\theta$  satisfies the following quadratic expression:

$$\begin{aligned} \hat{\theta}_{\text{WC}}^2 \left( \sum_{\{n: Y_n > 0\}} \frac{Y_n}{2} \right) - \hat{\theta}_{\text{WC}} \sum_{\{n: Y_n > 0\}} Y_n \left[ 1 + \log \left( \frac{b_n}{Y_n} \right) \right] \\ + \sum_{\{n: Y_n > 0\}} Y_n \log \left( \frac{b_n}{Y_n} \right) \left[ 1 + \frac{1}{2} \log \left( \frac{b_n}{Y_n} \right) \right] = 0 \end{aligned} \quad (29)$$

from which one can find an expression of the form  $\hat{\theta}_{\text{WC}} = g_{\text{WC}}(Y)$  by using the quadratic formula.

For comparison purposes, we also analyze the following *unweighted average* estimator:

$$\hat{\theta}_{\text{UA}} = g_{\text{UA}}(Y) = \frac{\sum_{\{n: Y_n > 0\}} \log \left( \frac{b_n}{Y_n} \right)}{\sum_{\{n: Y_n > 0\}} 1}. \quad (30)$$

This estimator is analogous to FBP since all nonzero measurements are treated equally (no weighting).

Since the above estimators are nonlinear, it would be difficult to exactly compute their bias or variance. Therefore we use a second-order Taylor expansion. If  $g(Y) = g(Y_1, \dots, Y_N)$  is an arbitrary functional, and  $\bar{Y}_n = b_n e^{-\theta}$  denotes the mean of  $Y_n$ , then

$$\begin{aligned} g(Y) \approx g(\bar{Y}) + \sum_n \frac{\partial}{\partial Y_n} g(\bar{Y})(Y_n - \bar{Y}_n) \\ + \frac{1}{2} \sum_n \sum_m \frac{\partial^2}{\partial Y_n \partial Y_m} g(\bar{Y})(Y_n - \bar{Y}_n)(Y_m - \bar{Y}_m). \end{aligned}$$

Under the assumption that  $Y_1, \dots, Y_N$  are independent, one can use the Taylor expansion to show

$$E\{g(Y)\} \approx g(\bar{Y}) + \frac{1}{2} \sum_n \frac{\partial^2}{\partial Y_n^2} g(\bar{Y}) \text{Var}\{Y_n\} \quad (31)$$

$$\text{Var}\{g(Y)\} \approx \sum_n \left[ \frac{\partial}{\partial Y_n} g(\bar{Y}) \right]^2 \text{Var}\{Y_n\}. \quad (32)$$

Let  $C = (1/N) \sum_n \bar{Y}_n = (1/N) \sum_n b_n e^{-\theta}$  denote the mean number of counts per measurement.

For the ML estimate, from (27)  $g_{\text{ML}}(\bar{Y}) = \theta$  and

$$\frac{\partial}{\partial Y_n} g_{\text{ML}}(\bar{Y}) = -\frac{1}{(NC)}$$

$$\frac{\partial^2}{\partial Y_n^2} g_{\text{ML}}(\bar{Y}) = \frac{1}{(NC)^2}.$$

Substituting into (31) and (32), we obtain

$$E\{\hat{\theta}_{\text{ML}}\} \approx \theta + \frac{1}{2NC} \quad (33)$$

$$\text{Var}\{\hat{\theta}_{\text{ML}}\} \approx \frac{1}{NC}. \quad (34)$$

For the LS estimator, from (28),  $g_{\text{LS}}(\bar{Y}) = \theta$ , and

$$\frac{\partial}{\partial Y_n} g_{\text{LS}}(\bar{Y}) = -\frac{1}{(NC)}$$

$$\frac{\partial^2}{\partial Y_n^2} g_{\text{LS}}(\bar{Y}) = \frac{2}{(NC)^2} - \frac{1}{NC b_n e^{-\theta}}.$$

Substituting into (31) and (32), we obtain

$$E\{\hat{\theta}_{\text{LS}}\} \approx \theta - \frac{1}{2C} + \frac{1}{NC} \quad (35)$$

$$\text{Var}\{\hat{\theta}_{\text{LS}}\} \approx \frac{1}{NC}. \quad (36)$$

For the WC estimator, from (29)  $g_{\text{WC}}(\bar{Y}) = \theta$  and

$$\frac{\partial}{\partial Y_n} g_{\text{WC}}(\bar{Y}) = -\frac{1}{(NC)}$$

$$\frac{\partial^2}{\partial Y_n^2} g_{\text{WC}}(\bar{Y}) = \frac{1}{(NC)^2}.$$

Substituting into (31) and (32), we obtain

$$E\{\hat{\theta}_{\text{WC}}\} \approx \theta + \frac{1}{2NC} \quad (37)$$

$$\text{Var}\{\hat{\theta}_{\text{WC}}\} \approx \frac{1}{NC}. \quad (38)$$

Note that these two approximations are identical to the expressions (33) and (34) for the ML estimate!

For the UA estimator, from (30)  $g_{\text{UA}}(\bar{Y}) = \theta$  and

$$\frac{\partial}{\partial Y_n} g_{\text{UA}}(\bar{Y}) = -\frac{1}{N b_n e^{-\theta}}$$

$$\frac{\partial^2}{\partial Y_n^2} g_{\text{UA}}(\bar{Y}) = \frac{1}{N (b_n e^{-\theta})^2}.$$

Substituting into (31) and (32), we obtain

$$E\{\hat{\theta}_{\text{UA}}\} \approx \theta + \frac{1}{2N e^{-\theta}} \sum_n \frac{1}{b_n} \quad (39)$$

$$\text{Var}\{\hat{\theta}_{\text{UA}}\} \approx \frac{1}{N^2 e^{-\theta}} \sum_n \frac{1}{b_n}. \quad (40)$$

Since the function  $1/x$  is convex,  $(1/N) \sum_n (1/b_n) \geq [(1/N) \sum_n b_n]^{-1}$ , with equality if and only if the  $b_n$  are

all equal. Thus we have the following lower bounds for the bias and variance of  $\hat{\theta}_{UA}$ :

$$E\{\hat{\theta}_{UA}\} \geq \theta + \frac{1}{2C} \quad (41)$$

$$\text{Var}\{\hat{\theta}_{UA}\} \geq \frac{1}{NC}. \quad (42)$$

For PET systems, the  $b_n$  terms are never equal, and in fact can be quite disparate. Thus the lower bound in (42) will not be achieved. This partially explains why FBP images are noisier than images reconstructed by the statistical methods (cf. Fig. 6).

Fig. 1 shows plots of the bias terms (33), (35), (37), and (41) versus mean counts  $C$ , for a system with  $N = 20$ . From these equations and the figure, we see that the maximum likelihood estimator and the weighted cubic estimator are nearly unbiased, the weighted least squares estimator has a systematic negative bias, and the unweighted average estimator has a systematic positive bias. The solid lines in the figures are the formulas, whereas the symbols denote empirical results from 500 realizations of (26) and (27)–(30). These results demonstrate that the systematic biases of PWLS and FBP shown in Fig. 7 are not simply artifacts of that simulation, but are intrinsic to those methods.

The source of the significant negative bias for  $\hat{\theta}_{LS}$  can be seen in (28). That expression is a weighted sum, weighted by the noisy measurement  $Y_n$ . If  $Y_n$  is larger than  $\bar{Y}_n$ , then  $\log(b_n/Y_n)$  will be smaller than  $\theta$ , but the corresponding weight in the summation is larger. Thus *the sum gives more weight to the under-estimated terms*, leading to the negative bias in (35). In contrast, the ML estimate first averages the counts before taking the logarithm (27), which leads to negligible bias, as seen in Fig. 1.

Although this problem is a scalar simplification of transmission tomography, the basic conclusion, that the under-estimated log ratios are weighted disproportionately, applies to the general problem as well.

#### APPENDIX B SMOOTHNESS PENALTIES

This appendix synthesizes our method [22], [34] for specifying the  $w_{jk}$  factors in (3). With  $u_n$  defined as in (20), define

$$g_j = \frac{\sum_n a_{nj}^2 u_n}{\sum_n a_{nj}^2}, \quad j = 1, \dots, p.$$

One can show that  $u_n$  is an estimate of the inverse the variance of  $\hat{l}_n$  defined in (19), and if these variances  $u_n$  were all equal, then  $g_j$  would be a constant independent of  $j$ . In general however,  $g_j$  will depend on  $j$ . Define

$$w'_{jk} = \begin{cases} 1, & \text{if } j \text{ and } k \text{ are horizontal neighbors} \\ \frac{1}{\sqrt{2}}, & \text{if } j \text{ and } k \text{ are diagonal neighbors} \\ 0, & \text{otherwise} \end{cases}$$

and

$$w_{jk} = w'_{jk} \sqrt{g_j g_k}.$$

This modified choice for  $w_{jk}$  provides approximately spatially-invariant image resolution and also allows one to relate the smoothing parameter  $\beta$  directly to image resolution [22], [34].

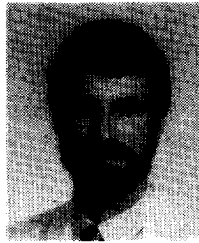
#### ACKNOWLEDGMENT

The author gratefully acknowledges the comments of E. Ficaró.

#### REFERENCES

- [1] K. Lange and J. A. Fessler, "Globally convergent algorithms for maximum a posteriori transmission tomography," *IEEE Trans. Image Processing*, this issue, pp. 1430–1438.
- [2] S. C. Huang, E. J. Hoffman, M. E. Phelps, and D. E. Kuhl, "Quantitation in positron emission computed tomography: 2 effects of inaccurate attenuation correction," *J. Comput. Assist. Tomog.*, vol. 3, no. 6, pp. 804–814, Dec. 1979.
- [3] R. H. Huesman *et al.*, "Orbiting transmission source for positron tomography," *IEEE Trans. Nucl. Sci.*, vol. 35, no. 1, pp. 735–739, Feb. 1988.
- [4] E. P. Ficaró *et al.*, "Clinical application of attenuation correction for cardiac SPECT studies," *Euro. J. Nucl. Med., Euro. Assoc. Nucl. Med. Cong.*, vol. 20, no. 10, pp. 143, Oct. 1993.
- [5] R. J. Jaszcak *et al.*, "Fast transmission CT for determining attenuation maps using a collimated line source, rotatable air-copper-lead attenuators and fan-beam collimation," *J. Nucl. Med.*, vol. 34, no. 9, pp. 1577–1586, Sept. 1993.
- [6] S. H. Manglos, F. D. Thomas, G. M. Gagne, and B. J. Hellwig, "Phantom study of breast tissue attenuation in myocardial imaging," *J. Nucl. Med.*, vol. 34, no. 6, pp. 992–996, June 1993.
- [7] C. H. Tung *et al.*, "Nonuniform attenuation correction using simultaneous transmission and emission converging tomography," *IEEE Trans. Nucl. Sci.*, vol. 39, no. 4, pp. 1134–1143, Aug. 1992.
- [8] S. L. Bacharach *et al.*, "Three dimensional registration of cardiac PET attenuation scans," *J. Nucl. Med., Abs. Book*, vol. 33, no. 5, p. 881, May 1992.
- [9] D. W. Townsend *et al.*, "Fully three-dimensional reconstruction for a PET camera with retractable septa," *IEEE Trans. Med. Imaging*, vol. 10, no. 4, pp. 505–512, Dec. 1991.
- [10] S. R. Meikle, M. Dahlbom, and S. R. Cherry, "Attenuation correction using count-limited transmission data in positron emission tomography," *J. Nucl. Med.*, vol. 34, no. 1, pp. 143–150, Jan. 1993.
- [11] S. H. Manglos *et al.*, "A quantitative comparison of attenuation-weighted backprojection with multiplicative and iterative postprocessing attenuation compensation in SPECT," *IEEE Trans. Med. Imaging*, vol. 7, no. 2, pp. 127–134, June 1988.
- [12] A. J. Rockmore and A. Macovski, "A maximum likelihood approach to transmission image reconstruction from projections," *IEEE Trans. Nucl. Sci.*, vol. 24, no. 3, pp. 1929–1935, June 1977.
- [13] K. Sauer and C. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. Signal Processing*, vol. 41, no. 2, pp. 534–548, Feb. 1993.
- [14] C. Bouman and K. Sauer, "Fast numerical methods for emission and transmission tomographic reconstruction," in *Proc. 27th Conf. Inform. Sci. Syst.*, Johns Hopkins Univ., Baltimore, MD, 1993, pp. 611–616.
- [15] ———, "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Trans. Image Processing*, to be published.
- [16] K. Lange, "An overview of Bayesian methods in image reconstruction," in *Proc. SPIE, Digital Image Synthesis Inverse Opt.*, vol. 1351, 1990, pp. 270–287.
- [17] K. Lange, M. Bahn, and R. Little, "A theoretical study of some maximum likelihood algorithms for emission and transmission tomography," *IEEE Trans. Med. Imaging*, vol. 6, no. 2, pp. 106–114, June 1987.
- [18] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comp. Assist. Tomog.*, vol. 8, no. 2, pp. 306–316, Apr. 1984.
- [19] J. M. Ollinger, "Maximum likelihood reconstruction of transmission images in emission computed tomography via the EM algorithm," *IEEE Trans. Med. Imaging*, vol. 13, no. 1, pp. 89–101, Mar. 1994.
- [20] N. T. Ranger, C. J. Thompson, and A. C. Evans, "The application of a masked orbiting transmission source for attenuation correction in PET," *J. Nucl. Med.*, vol. 30, no. 6, pp. 1056–1068, June 1989.
- [21] E. P. Ficaró, J. A. Fessler, W. L. Rogers, and M. Schwaiger, "Comparison of Americium-241 and Technicium-99m as transmission sources for

- attenuation correction of Thallium-201 SPECT imaging of the heart," *J. Nucl. Med.*, vol. 35, no. 4, pp. 652-663, Apr. 1994.
- [22] J. A. Fessler and W. L. Rogers, "Uniform quadratic penalties cause nonuniform image resolution (and sometimes vice versa)," in *Conf. Rec. IEEE Nucl. Sci. Symp. Med. Imaging Conf.*, vol. 4, 1994, pp. 1915-1919.
- [23] K. Lange, "Convergence of EM image reconstruction algorithms with Gibbs smoothing," *IEEE Trans. Med. Imaging*, vol. 9, no. 4, pp. 439-446, Dec. 1990, corrections, June 1991.
- [24] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C*. Cambridge, MA: Cambridge Univ. Press, 1988.
- [25] J. A. Fessler, "Penalized weighted least-squares image reconstruction for positron emission tomography," *IEEE Trans. Med. Imaging*, vol. 13, no. 2, pp. 290-300, June 1994.
- [26] J. A. Fessler and A. O. Hero, "Space-alternating generalized EM algorithms for penalized maximum-likelihood image reconstruction," *Comm. Signal Processing Lab., Dept. Elec. Eng. Comput. Sci., Univ. of Michigan, Ann Arbor, MI, Tech. Rep. 286, Feb. 1994.*
- [27] J. A. Fessler and A. O. Hero, "Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms," *IEEE Trans. Image Processing*, this issue, pp. 1417-1429.
- [28] A. R. De Pierro, "On the relation between the ISRA and the EM algorithm for positron emission tomography," *IEEE Trans. Med. Imaging*, vol. 12, no. 2, pp. 328-333, June 1993.
- [29] ———, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE Trans. Med. Imaging*, vol. 14, no. 1, pp. 132-137, Mar. 1995.
- [30] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Stat. Soc. Ser. B*, vol. 39, no. 1, pp. 1-38, 1977.
- [31] J. A. Fessler, "EM and gradient algorithms for transmission tomography with background contamination," Cyclotron PET Facility, Univ. of Michigan, Tech. Rep. UM-PET-JF-94-1, Dec. 1994. Available on WWW from <http://www.engin.umich.edu/fessler>.
- [32] E. Mumcuoglu, R. Leahy, and S. Cherry, "A statistical approach to transmission image reconstruction from ring source calibration measurements in PET," in *Conf. Rec. IEEE Nucl. Sci. Symp. Med. Imaging Conf.*, 1992, pp. 910-912.
- [33] E. U. Mumcuoglu, R. Leahy, S. R. Cherry, and Z. Zhou, "Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images," *Signal Image Processing Inst., Univ. Southern Calif., Tech. Rep. 241, Sept. 1993.*
- [34] J. A. Fessler *et al.*, "Resolution properties of regularized image reconstruction methods," *Commun. Signal Processing Lab., Dept. Elec. Eng. Comput. Sci., Univ. of Michigan, Ann Arbor, Tech. Rep., Aug. 1995.*



**Jeffrey A. Fessler** (S'83-M'90) received the B.S.E.E. degree from Purdue University in 1985, the M.S.E.E. degree from Stanford University in 1986, and the M.S. degree in statistics from Stanford University in 1989. From 1985-1988, he was a National Science Foundation Graduate Fellow at Stanford, where he received the Ph.D. degree in electrical engineering in 1990.

From 1991 to 1992, he was a Department of Energy Alexander Hollaender Post-Doctoral Fellow in the Division of Nuclear Medicine at the University of Michigan. From 1993 to 1995, he was an Assistant Professor in Nuclear Medicine and the Bioengineering Program of the University of Michigan. Since 1995, he has been with the Department of Electrical Engineering and Computer Science of the University of Michigan. His research interests are in statistical aspects of medical imaging.