

Emission Image Reconstruction for Randoms-Precorrected PET Allowing Negative Sinogram Values

Sangtae Ahn*, *Student Member, IEEE*, and Jeffrey A. Fessler, *Senior Member, IEEE*

Abstract—Most positron emission tomography (PET) emission scans are corrected for accidental coincidence (AC) events by real-time subtraction of delayed-window coincidences, leaving only the randoms-precorrected data available for image reconstruction. The real-time randoms pre-correction compensates in mean for AC events but destroys the Poisson statistics. The exact log-likelihood for randoms-precorrected data is inconvenient, so practical approximations are needed for maximum likelihood or penalized-likelihood image reconstruction. Conventional approximations involve setting negative sinogram values to zero, which can induce positive systematic biases, particularly for scans with low counts per ray. We propose new likelihood approximations that allow negative sinogram values without requiring zero-thresholding. With negative sinogram values, the log-likelihood functions can be nonconcave, complicating maximization; nevertheless, we develop monotonic algorithms for the new models by modifying the separable paraboloidal surrogates and the maximum-likelihood expectation-maximization (ML-EM) methods. These algorithms ascend to local maximizers of the objective function. Analysis and simulation results show that the new shifted Poisson (SP) model is nearly free of systematic bias yet keeps low variance. Despite its simpler implementation, the new SP performs comparably to the saddle-point model which has shown the best performance (as to systematic bias and variance) in randoms-precorrected PET emission reconstruction.

Index Terms—Accidental coincidences, maximum-likelihood reconstruction, positron emission tomography (PET), randoms-precorrected PET.

I. INTRODUCTION

ACCIDENTAL coincidence (AC) events, also known as randoms, are a primary source of background noise in positron emission tomography (PET) [1]. AC events occur when two photons that arise from separate positron emissions are detected within a coincidence timing window and recorded as having originated from the same emission [2], [3]. Quantitative PET studies require correction for AC events.

Manuscript received August 4, 2003; revised January 27, 2004. This work was supported in part by the National Science Foundation (NSF) under Grant BES-9982349, in part by the National Institutes of Health (NIH) under Grant CA-60711, and in part by the Department of Energy (DOE) under Grant DE-FG02-87ER60561. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was R. H. Huesman. *Asterisk indicates corresponding author*

*S. Ahn is with the Electrical Engineering and Computer Science Department, University of Michigan, Ann Arbor, MI 48109-2122 USA (e-mail: sangtaea@umich.edu).

J. A. Fessler is with the Electrical Engineering and Computer Science Department, University of Michigan, Ann Arbor, MI 48109-2122 USA (e-mail: fessler@umich.edu).

Digital Object Identifier 10.1109/TMI.2004.826046

Usually, PET systems detect coincidence events during “prompt” windows and “delayed” windows [4], [5]. The delayed coincidences represent AC events (or randoms), and the prompt coincidences represent true coincidences contaminated by AC events (plus Compton scatter events). In most PET scans, the prompt data are precorrected for the effects of AC events by real-time subtraction of the delayed coincidences [1]. The subtraction compensates for the AC events in terms of the mean but increases the variance of the data [6]. Ideally, scanners would maintain both prompt and randoms sinograms. One could then estimate the mean of AC events from the randoms sinogram [6]–[9] and incorporate these estimates into an appropriate model for the prompt measurement [3], [6], [8] to estimate unknown parameters (radioactivity for emission scans and attenuation coefficients for transmission scans). However, because of data storage limitations and historical momentum, most PET centers store the randoms-precorrected data only [5]. This paper focuses on the problem of reconstructing emission images by considering the measurement statistics based on *only* randoms-precorrected data *without* access to separate prompt and randoms sinograms. We do assume that a rough estimate of the randoms contribution is available, such as can be computed from the block singles rates that are often available [6].

Whereas both (prompt and randoms) sinograms are well approximated as being Poisson distributed [10], the randoms-precorrected data do not follow Poisson statistics. The exact log-likelihood of precorrected data is inconvenient to maximize. Several practical approximations to the exact log-likelihood have been investigated [5], [11]–[14]. A shifted Poisson (SP) model and a saddle-point (SD) model are such approximations [5], [11]. For *transmission scans* both SP and SD models have been shown to outperform conventional ordinary Poisson (OP) and weighted least squares models in terms of systematic bias and variance [5], [11], [12]. In *transmission* image reconstruction, the SP model seems more attractive than the SD model since its implementation is simpler but their performance is comparable.

For *emission scans* the SP and SD models again lead to lower variance than the OP model [13]. However, SP suffers from a positive systematic bias for low counts per ray, albeit generally less than OP, while SD seems to be free of such a bias [13]. The systematic bias is caused by zeroing negative sinogram values (note randoms-precorrected data can be negative) [13]; it can affect contrast and quantitative studies adversely. The zero-thresholding of negative values for SP and OP was a natural choice

since those models are based on Poisson approximations. Moreover, negative sinogram values can cause reconstruction algorithms like classic ML-EM to diverge. Negative sinogram values also cause the Poisson log-likelihood to become nonconcave, and it is difficult to develop algorithms that globally maximize a nonconcave objective function. By contrast, in a *transmission* case, negative values do not cause reconstruction algorithms to diverge; in fact, they help ensure concavity of the log-likelihood that otherwise could be nonconcave [14, Sec. 4.6].

To eliminate the positive systematic bias in *emission scans*, we propose new SP and OP models that allow negative sinogram values, departing from the conventional tendency to zero-threshold them [13], [15]–[17]. We will henceforth call our new methods “SP[−]” and “OP[−]” to differentiate from the conventional ones with zero-thresholding that will be called, in this paper, “SP⁺” and “OP⁺.” In contrast to some previous methods that allow the pixel values to be negative [18], here we enforce the usual nonnegativity constraint in the *image domain* but allow the *sinogram* values to be negative for SP[−] and OP[−]. We show that negative sinogram values in emission scans need *not* cause divergence of appropriate algorithms for SP[−] and OP[−]. Although negative values can cause the likelihood for SP[−] or OP[−] to be nonconcave, one can achieve at least a locally optimal reconstruction by employing algorithms that increase the objective function monotonically. We use the “optimization transfer principle” [19] to derive two monotonic algorithms that allow negative values: separable paraboloidal surrogates (SPS) and a variant of maximum likelihood expectation maximization (ML-EM). Our practical experience is that the locally optimal reconstruction obtained by monotonic algorithms are very good regardless of initializations.

We show analytically that our new SP[−] model is nearly free of systematic bias (as is the new OP[−] model) and leads to less variance than other methods including OP[−] and filtered back-projection (FBP); this is corroborated by simulation results in Section VII. In other words, the new SP[−] model, our recommended method for randoms-precorrected emission image reconstruction, is comparable, in spite of its simpler implementation, to SD which has shown the best performance in terms of systematic bias and variance.

Section II reviews the statistical model for precorrected measurements and its exact log-likelihood. Section III describes conventional approximation models and our new ones for emission scans. We analyze the systematic bias due to zero-thresholding and the asymptotic variances in Sections IV and V. Section VI provides monotonic algorithms for the new models and Section VII gives simulation results.

II. MEASUREMENT MODEL AND EXACT LOG-LIKELIHOOD

Let $\mathbf{Y} = [Y_1 \dots Y_N]'$ denote the *precorrected* measurements for PET emission scans, where $'$ denotes vector and matrix transpose. The precorrected measurement for the i th bin is

$$Y_i = Y_i^{\text{prompt}} - Y_i^{\text{delay}} \quad (1)$$

where Y_i^{prompt} and Y_i^{delay} are the number of coincidences detected within the prompt and delayed windows, respectively.

The prompts and delays can be modeled reasonably as independent Poisson random variables [10] as follows:

$$Y_i^{\text{prompt}} \sim \text{Poisson} \left\{ \sum_{j=1}^p a_{ij} \lambda_j^{\text{true}} + r_i + s_i \right\} \quad (2)$$

$$Y_i^{\text{delay}} \sim \text{Poisson} \{r_i\} \quad (3)$$

where $a_{ij} \geq 0$ is the entry in the system matrix \mathbf{A} incorporating scan geometry, attenuation, detector efficiencies, etc.; $\lambda_j^{\text{true}} \geq 0$ is the activity at the j th voxel; and $r_i \geq 0$ and $s_i \geq 0$ are the means of AC events and scatters, respectively. We assume that $\mathbf{r} = [r_1 \dots r_N]'$ and $\mathbf{s} = [s_1 \dots s_N]'$ are *known*¹ in order to focus on the problem of estimating the unknown activity $\boldsymbol{\lambda}^{\text{true}} = [\lambda_1^{\text{true}} \dots \lambda_p^{\text{true}}]'$ based on the *non-Poisson distributed* measurements \mathbf{Y} without access to Y_i^{prompt} and Y_i^{delay} . In other words, we investigate the “upper bound” of performance of each reconstruction method that needs estimates of \mathbf{r} and \mathbf{s} . In [9], the effects of randoms estimates on bias for various reconstruction methods were investigated. We also assume $r_i > 0$ for all i for simplicity; the analysis and algorithms are easily adopted to include rays where $r_i = 0$.

Let $\mathbf{y} = [y_1 \dots y_N]'$ be an observed realization of \mathbf{Y} . Since the measurements are independent, one can obtain the exact log-likelihood, ignoring constants independent of $\boldsymbol{\lambda}$, as in [5], [11]

$$L(\boldsymbol{\lambda}; \mathbf{Y}) = \sum_{i=1}^N h_i^{\text{EX}}(l_i(\boldsymbol{\lambda})) \quad (4)$$

with

$$l_i(\boldsymbol{\lambda}) = \sum_{j=1}^p a_{ij} \lambda_j \quad (5)$$

and

$$h_i^{\text{EX}}(l) = \log \left(\sum_{m=[-y_i]_+}^{\infty} \frac{(l + r_i + s_i)^{y_i+m} r_i^m}{(y_i + m)! m!} \right) - (l + 2r_i + s_i) \quad (6)$$

where $[x]_+ = \max\{x, 0\}$. For notational simplicity, we omit an argument indicating the dependence of h_i^{EX} on y_i in (4) and (6).

For penalized-likelihood (PL) reconstruction, one must find a maximizer of the objective function

$$\Phi(\boldsymbol{\lambda}; \mathbf{Y}) = L(\boldsymbol{\lambda}; \mathbf{Y}) - R(\boldsymbol{\lambda}) \quad (7)$$

over a nonnegativity constraint on the *image* $\boldsymbol{\lambda}$, where R is a regularization term that controls a trade-off of resolution and noise in the reconstructed image. The exact log-likelihood function (4) is inconvenient to maximize although it can be expressed without the infinite summations in (6) using Bessel functions [14, Sec. 3.2]. Section III describes practical approximations to the exact log-likelihood.

¹Even in a case where one does not have access to the delayed events separately, the total number of AC events or the block singles rates are often available at the end of the scan and can be used to estimate AC rates [6], [20]. Indeed, approximate models like SP and SD are known to be robust to errors in estimating AC rates [11]. Regarding scatter estimation and correction, see [6], [21], [22] for example.

III. APPROXIMATIONS TO EXACT LOG-LIKELIHOOD

A. Ordinary Poisson Approximation

A simple approach that does not need an estimate of AC events \mathbf{r} is to approximate the measurements as Poisson random variables as follows:

$$Y_i \stackrel{\text{OP}}{\sim} \text{Poisson} \left\{ \sum_{j=1}^p a_{ij} \lambda_j^{\text{true}} + s_i \right\}. \quad (8)$$

This model matches the first moment of Y_i only. The log-likelihood L^{OP^-} corresponding to this ‘‘OP’’ approximation² is of the form (4) with

$$h_i^{\text{OP}^-}(l) = y_i \log(l + s_i) - (l + s_i). \quad (9)$$

We assume $s_i > 0$ in (9); otherwise, negative values y_i would cause reconstruction algorithms to diverge since $h_i^{\text{OP}^-}(0) = +\infty$ for $y_i < 0$ and $s_i = 0$. To avoid such divergence, past studies of the OP approach [13], [16] for emission scans have used zero-thresholded values as follows:

$$h_i^{\text{OP}^+}(l) = [y_i]_+ \log(l + s_i) - (l + s_i) \quad (10)$$

called the ‘‘OP’’ approximation in this paper. (Note the slightly different use of terms from [13].) The zero-thresholding is natural in view of the nonnegative nature of Poisson random variables in (8). Moreover, it guarantees the concavity of $h_i^{\text{OP}^+}$, and hence the existence and uniqueness of the PL reconstruction under mild conditions [23]. However, zero-thresholding destroys the first moment matching in (8), and the increase of the precorrected data by zero-thresholding causes the estimators to have a positive systematic bias since emission data is linearly related to activity in the mean. Section IV shows that the seemingly unnatural use of negative sinogram values in the Poisson framework can alleviate the systematic bias problem of OP^+ .

B. Shifted Poisson Approximation

An improved approximation is to match both the first and second moments as follows:

$$Y_i + 2r_i \stackrel{\text{SP}}{\sim} \text{Poisson} \left\{ \sum_{j=1}^p a_{ij} \lambda_j^{\text{true}} + s_i + 2r_i \right\} \quad (11)$$

where in practice one must use an estimate \hat{r}_i . This ‘‘SP’’ approximation³ leads to a log-likelihood function L^{SP^-} of the form (4) with

$$h_i^{\text{SP}^-}(l) = (y_i + 2r_i) \log(l + s_i + 2r_i) - (l + s_i + 2r_i). \quad (12)$$

Similarly, its conventional zero-thresholded version L^{SP^+} uses [13]

$$h_i^{\text{SP}^+}(l) = [y_i + 2r_i]_+ \log(l + s_i + 2r_i) - (l + s_i + 2r_i). \quad (13)$$

The zero-thresholding again ensures the concavity of L^{SP^+} but also causes positive systematic bias, albeit generally less than that of OP^+ since it is more likely that $y_i < 0$ than $y_i + 2r_i < 0$. Section IV describes the details.

²The minus sign signifies that this approximation allows negative precorrected data $y_i < 0$.

³The minus sign signifies that this approximation allows $y_i + 2r_i < 0$.

C. Saddle-Point Approximation

Another approach is to make a second-order Taylor series approximation in the z -transform domain to the probability generating function and then carry out the inverse transform [24], [25]. The log-likelihood L^{SD} corresponding to this SD approximation [13] is of the form (4) with

$$h_i^{\text{SD}}(l) = y_i \log \left(\frac{l + s_i + r_i}{z_i + u_i(l)} \right) - l + u_i(l) - \frac{1}{2} \log u_i(l) \quad (14)$$

where

$$z_i = \begin{cases} y_i + 1, & \text{for } y_i \geq 0 \\ y_i - 1, & \text{for } y_i < 0 \end{cases}$$

and

$$u_i(l) = \sqrt{z_i^2 + 4(l + r_i + s_i)r_i}.$$

The SD model for emission image reconstruction is free of systematic bias and leads to lower variance than OP^+ [13]. Indeed, in all cases studied to date, the SD model has shown the best performance for randoms-preccorrected PET emission reconstruction. We observe those properties empirically in Section VII. However, the new SP^- , despite its simpler implementation, performs comparably to SD.

D. Log-Likelihood for Prompt Data

If one has access to the prompt data $\mathbf{Y}^{\text{prompt}}$, then one can use the log-likelihood for the prompt data in the form (4) with

$$h_i^{\text{PR}}(l) = y_i^{\text{prompt}} \log(l + s_i + r_i) - (l + s_i + r_i).$$

We include this PR model for comparing the bias and variance of the methods for randoms-preccorrected data in Section VII. Since $\mathbf{Y}^{\text{prompt}}$ has lower variance than \mathbf{Y} , it serves as a baseline for comparing algorithms.

IV. EFFECTS OF ZERO-THRESHOLDING ON BIAS

The sinogram zero-thresholding in (10) and (13) increases the mean values of the data. This section analyzes the effects of this shift.

First, we focus on a single ray to investigate the properties of OP^+ and SP^+ . Let Y be a precorrected measurement modeled as the difference of two independent Poisson random variables as follows:

$$Y \triangleq \text{Poisson}\{\theta + r\} - \text{Poisson}\{r\} \quad (15)$$

where θ and r denote the mean number of trues (possibly including scatters) and AC events, respectively. The normalized effective mean of trues are given by

$$m^{\text{OP}^+} = \frac{E\{[Y]_+\}}{\theta} \text{ for } \text{OP}^+$$

and

$$m^{\text{SP}^+} = \frac{E\{[Y + 2r]_+\} - 2r}{\theta} \text{ for } \text{SP}^+$$

as a function of θ . Note m^{OP^+} and m^{SP^+} would be unity without zero-thresholding. We calculated these expectations

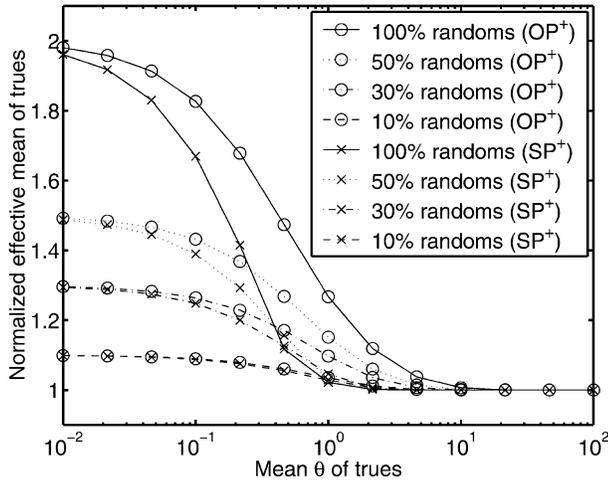


Fig. 1. Effective means of trues increased by zero-thresholding for OP⁺ and SP⁺. In this figure, randoms fractions in % denote (mean of randoms)/(mean of trues).

using the Bessel function expression for the probability mass function for Y [14, Sec. 3.2] as follows:

$$P(Y = y; \theta) = \sum_{m=[-y]_+}^{\infty} \frac{(\theta + r)^{y+m} e^{-(\theta+r)} r^m e^{-r}}{(y+m)! m!}$$

$$= \frac{e^{-(\theta+2r)}}{|y|!} \left(\sqrt{\frac{\theta+r}{r}} \right)^y J_{|y|} \left(2i \sqrt{(\theta+r)r} \right)$$

where $i = \sqrt{-1}$ and $J_n(\cdot)$ is the Bessel function of the first kind of order n [26, p. 575]. Fig. 1 shows the results, from which we infer that 1) for counts per ray higher than 10 (or 1), there is little effect of zero-thresholding for OP⁺ (or SP⁺) as long as randoms amount to less than 100% of trues, and 2) OP⁺ generally leads to a higher bias than SP⁺. For extremely low counts or low AC rates, OP⁺ and SP⁺ yield similar results.

Next, we consider a one-parameter example for the OP approach to illustrate how allowing negative values can be helpful for reducing systematic positive bias. Let the measurements be the difference of two independent Poisson random variables as follows:

$$Z_i \triangleq \text{Poisson}\{a_i \theta^{\text{true}} + s_i + r_i\} - \text{Poisson}\{r_i\} \quad (16)$$

for $i = 1, \dots, N$. Setting $s_i = 0$, ML estimates based on OP⁻ and OP⁺ models [see (4) with (9) and (10)] with an image nonnegativity constraint have the following analytical solutions:

$$\hat{\theta}^{\text{OP}^-} = \frac{\left[\sum_{i=1}^N z_i \right]_+}{\sum_{i=1}^N a_i} \quad (17)$$

and

$$\hat{\theta}^{\text{OP}^+} = \frac{\sum_{i=1}^N [z_i]_+}{\sum_{i=1}^N a_i}. \quad (18)$$

Note the zero-thresholding in (17) is due to the *image-domain* nonnegativity constraint and not a primary source of the positive

bias, whereas the zero-thresholding $[z_i]_+$ in (18) is in the *sino-gram domain* from (10). In view of Fig. 1, $\hat{\theta}^{\text{OP}^+}$ is more biased than $\hat{\theta}^{\text{OP}^-}$. For instance, if $\theta^{\text{true}} = 1$, $a_i = 1$, $r_i = 0.5$, and $N = 10$, then the estimator biases can be computed using Fig. 1 with (17) and (18) as follows: $b^{\text{OP}^-} = E\{\hat{\theta}^{\text{OP}^-}\} - \theta^{\text{true}} = 1.014 - 1 = 0.014$ and $b^{\text{OP}^+} = E\{\hat{\theta}^{\text{OP}^+}\} - \theta^{\text{true}} = 1.152 - 1 = 0.152$. So the OP⁻ model reduces significantly the positive bias in OP⁺. This example suggests that when the rays passing through a particular voxel have low counts but high AC rates, OP⁺ will yield a higher positive systematic bias than OP⁻. The comparison of SP⁻ and SP⁺ would be similar although there are no closed-form estimators for SP⁻ and SP⁺ like (17) and (18).

For *high counts* per ray cases, sinogram zero-thresholding is not problematic since the probability of negative values is greatly reduced. Section V investigates the asymptotic behavior of the estimators for high counts.

V. ASYMPTOTIC ANALYSIS

This section analyzes the *asymptotic* bias and covariance of OP⁻, SP⁻, and SD estimators for *high counts* (per ray) cases. The purpose of the analysis is to compare the estimator properties rather than to accurately predict estimator behavior. (The prediction of the mean and covariance of PL or ML estimators could be conducted following [27]; see [12] for such analysis for randoms-precorrected PET *transmission* scans). We focus on ML estimators for simplicity. We do not consider OP⁺ and SP⁺ since OP⁺ and SP⁺ should behave quite similarly to OP⁻ and SP⁻, respectively, for high counts per ray.

A. Asymptotic Unbiasedness and Asymptotic Covariance

Let the *precorrected* measurement $\mathbf{Y}^n = [Y_1^n \dots Y_N^n]'$ be such that

$$Y_i^n \triangleq \text{Poisson}\{n(l_i(\lambda^{\text{true}}) + r_i + s_i)\} - \text{Poisson}\{nr_i\} \quad (19)$$

for $i = 1, \dots, N$ where l_i is defined in (5) and $n \in \mathbb{N}$ represents a factor proportional to the number of total counts or the scan time. Define $\bar{\mathbf{Y}}^n \triangleq \mathbf{Y}^n/n$, then⁴

$$\bar{\mathbf{Y}}^n \xrightarrow{p} \bar{\mathbf{Y}} \quad (20)$$

as $n \rightarrow \infty$ by the weak law of large numbers [28, p. 112] where $\bar{\mathbf{Y}} = E\{\mathbf{Y}^1\} = \mathbf{A}\lambda^{\text{true}} + \mathbf{s}$ and “ \xrightarrow{p} ” denotes convergence in probability. Also,

$$\sqrt{n}(\bar{\mathbf{Y}}^n - \bar{\mathbf{Y}}) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \text{Cov}\{\mathbf{Y}^1\}) \quad (21)$$

as $n \rightarrow \infty$ by the central limit theorem [29, p. 61] where $\text{Cov}\{\mathbf{Y}^1\} = \text{diag}\{l_i(\lambda^{\text{true}}) + s_i + 2r_i\}$ and “ $\xrightarrow{\mathcal{L}}$ ” denotes convergence in law (or distribution). Because $h_i^{\text{OP}^-}$ in (9) is affine in \mathbf{Y} , we can write the OP⁻ estimate based on \mathbf{Y}^n as follows:

$$\arg \max_{\lambda \geq 0} L^{\text{OP}^-}(\lambda; \mathbf{Y}^n) = \arg \max_{\lambda \geq 0} \left\{ \frac{L^{\text{OP}^-}(\lambda; \mathbf{Y}^n)}{n} \right\}$$

$$= \arg \max_{\lambda \geq 0} L^{\text{OP}^-}(\lambda; \bar{\mathbf{Y}}^n) \quad (22)$$

⁴Note \mathbf{Y}^n is identical to the sum of n iid random vectors each of which is identical to \mathbf{Y}^1 .

$$\triangleq \hat{\boldsymbol{\lambda}}^{\text{OP}^-}(\bar{\mathbf{Y}}^n) \quad (23)$$

where L^{OP^-} is of the form (4) with (9). One can show $L^{\text{OP}^-}(\boldsymbol{\lambda}^{\text{true}}; \bar{\mathbf{Y}}) \geq L^{\text{OP}^-}(\boldsymbol{\lambda}; \bar{\mathbf{Y}})$ for all $\boldsymbol{\lambda} \geq \mathbf{0}$ since $h_i^{\text{OP}^-}(l)$ attains a maximum over $l \geq 0$ at $l = [y_i - s_i]_+$. We assume that the $N \times p$ system matrix \mathbf{A} has full column rank, ensuring uniqueness of the noiseless reconstruction $\hat{\boldsymbol{\lambda}}^{\text{OP}^-}(\bar{\mathbf{Y}}) = \boldsymbol{\lambda}^{\text{true}}$.

One can easily show that $\nabla^{10} L^{\text{OP}^-}(\boldsymbol{\lambda}^{\text{true}}; \bar{\mathbf{Y}}) = \mathbf{0}$ and that

$$\nabla^{20} L^{\text{OP}^-}(\boldsymbol{\lambda}^{\text{true}}; \bar{\mathbf{Y}}) = \mathbf{A}' \text{diag} \left\{ \frac{1}{\sum_{j=1}^p a_{ij} \lambda_j^{\text{true}} + s_i} \right\} \mathbf{A}$$

is positive definite since \mathbf{A} has full column rank where $\nabla^{10} = [\partial/\partial\lambda_1 \dots \partial/\partial\lambda_p]'$ and ∇^{20} denote the column gradient operator and the Hessian operator, respectively. Then $\hat{\boldsymbol{\lambda}}^{\text{OP}^-}(\cdot)$ is continuously differentiable at $\bar{\mathbf{Y}}$ by the implicit function theorem [30, p. 668].

Since $\hat{\boldsymbol{\lambda}}^{\text{OP}^-}(\cdot)$ is continuous at $\bar{\mathbf{Y}}$, one can show [28, p. 124]

$$\hat{\boldsymbol{\lambda}}^{\text{OP}^-}(\bar{\mathbf{Y}}^n) \xrightarrow{p} \hat{\boldsymbol{\lambda}}^{\text{OP}^-}(\bar{\mathbf{Y}}) = \boldsymbol{\lambda}^{\text{true}}$$

as $n \rightarrow \infty$, in view of (20). In other words, the OP^- estimator (and OP^+ as well) is *asymptotically unbiased*.

Next, we investigate the asymptotic variance for OP^- . Since $\hat{\boldsymbol{\lambda}}^{\text{OP}^-}(\cdot)$ is continuously differentiable in a neighborhood of $\bar{\mathbf{Y}}$, it can be shown by the Delta method [29, p. 61], in view of (21), that

$$\sqrt{n} \left(\hat{\boldsymbol{\lambda}}^{\text{OP}^-}(\bar{\mathbf{Y}}^n) - \boldsymbol{\lambda}^{\text{true}} \right) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}^{\text{OP}^-})$$

with

$$\boldsymbol{\Sigma}^{\text{OP}^-} = \nabla \hat{\boldsymbol{\lambda}}^{\text{OP}^-}(\bar{\mathbf{Y}}) \text{Cov}\{\bar{\mathbf{Y}}^1\} [\nabla \hat{\boldsymbol{\lambda}}^{\text{OP}^-}(\bar{\mathbf{Y}})]'$$

where $\nabla = [\partial/\partial Y_1 \dots \partial/\partial Y_N]$ denotes the row gradient operator. The gradient $\nabla \hat{\boldsymbol{\lambda}}^{\text{OP}^-}(\bar{\mathbf{Y}})$ of the implicitly defined function (23) can be computed as in [27]. Some manipulation yields the asymptotic covariance as follows:

$$\boldsymbol{\Sigma}^{\text{OP}^-} = \mathbf{F}_{\text{OP}^-}^{-1} \mathbf{A}' \text{diag} \left\{ \frac{l_i(\boldsymbol{\lambda}^{\text{true}}) + s_i + 2r_i}{(l_i(\boldsymbol{\lambda}^{\text{true}}) + s_i)^2} \right\} \mathbf{A} \mathbf{F}_{\text{OP}^-}^{-1} \quad (24)$$

where

$$\mathbf{F}_{\text{OP}^-} = \mathbf{A}' \text{diag} \left\{ \frac{1}{l_i(\boldsymbol{\lambda}^{\text{true}}) + s_i} \right\} \mathbf{A}.$$

Similarly, one can show that the SP^- method (and SP^+ as well) is also *asymptotically unbiased* and that its asymptotic covariance is

$$\boldsymbol{\Sigma}^{\text{SP}^-} = \left[\mathbf{A}' \text{diag} \left\{ \frac{1}{l_i(\boldsymbol{\lambda}^{\text{true}}) + s_i + 2r_i} \right\} \mathbf{A} \right]^{-1}. \quad (25)$$

To analyze the SD case, one needs the following approximation that, from (14), is valid for large n :

$$\frac{L^{\text{SD}}(\boldsymbol{\lambda}; \mathbf{Y}^n)}{n} \approx L^{\text{SD}'}(\boldsymbol{\lambda}; \bar{\mathbf{Y}}^n) \quad (26)$$

with $L^{\text{SD}'}(\boldsymbol{\lambda}; \mathbf{Y}) = \sum_{i=1}^N h_i^{\text{SD}'}(l_i(\boldsymbol{\lambda}))$ where

$$h_i^{\text{SD}'}(l) = y_i \log \left(\frac{l + s_i + r_i}{y_i + \tilde{u}_i(l)} \right) - l + \tilde{u}_i(l)$$

and

$$\tilde{u}_i(l) = \sqrt{y_i^2 + 4(l + r_i + s_i)r_i}.$$

The SD estimate can be written as follows:

$$\begin{aligned} \arg \max_{\boldsymbol{\lambda} \geq \mathbf{0}} L^{\text{SD}}(\boldsymbol{\lambda}; \mathbf{Y}^n) &= \arg \max_{\boldsymbol{\lambda} \geq \mathbf{0}} \left\{ \frac{L^{\text{SD}}(\boldsymbol{\lambda}; \mathbf{Y}^n)}{n} \right\} \\ &\approx \arg \max_{\boldsymbol{\lambda} \geq \mathbf{0}} L^{\text{SD}'}(\boldsymbol{\lambda}; \bar{\mathbf{Y}}^n) \\ &\triangleq \hat{\boldsymbol{\lambda}}^{\text{SD}'}(\bar{\mathbf{Y}}^n). \end{aligned}$$

Since 1) $L^{\text{SD}'}(\boldsymbol{\lambda}; \bar{\mathbf{Y}})$ has the unique maximizer (over $\boldsymbol{\lambda} \geq \mathbf{0}$), 2) $\hat{\boldsymbol{\lambda}}^{\text{SD}'}(\bar{\mathbf{Y}}) = \boldsymbol{\lambda}^{\text{true}}$ (note $L^{\text{SD}'}$ can be shown to be strictly concave) and 3) the approximation (26) becomes more accurate as n increases, it can be shown that the SD method is also *asymptotically unbiased*. By similar manipulations, one can obtain the asymptotic covariance for SD as follows:

$$\boldsymbol{\Sigma}^{\text{SD}} \approx \boldsymbol{\Sigma}^{\text{SD}'} = \left[\mathbf{A}' \text{diag} \left\{ \frac{1}{l_i(\boldsymbol{\lambda}^{\text{true}}) + s_i + 2r_i} \right\} \mathbf{A} \right]^{-1} \quad (27)$$

which is equal to (25).

Both SP^- and SD are *asymptotically efficient* in the following sense. Noting $\bar{\mathbf{Y}}^n$ is asymptotically normal with mean $\bar{\mathbf{Y}}$ and covariance $\text{Cov}\{\mathbf{Y}^1\}/n$ from (21), one can obtain the Cramér-Rao bound from the asymptotic normal likelihood as follows:

$$\begin{aligned} B(\boldsymbol{\lambda}^{\text{true}}) &\approx \frac{1}{n} \left[\mathbf{A}' \text{Cov}\{\mathbf{Y}^1\}^{-1} \mathbf{A} \right]^{-1} \\ &= \frac{1}{n} \left[\mathbf{A}' \text{diag} \left\{ \frac{1}{l_i(\boldsymbol{\lambda}^{\text{true}}) + s_i + 2r_i} \right\} \mathbf{A} \right]^{-1}. \end{aligned}$$

Now one can see SP^- and SD asymptotically achieve this bound from (25) and (27). Note that the exact CR bound appears intractable due to form of (6).

The reasons that SP^- and SD are asymptotically efficient are the following. First, for SP^- , the precorrected data are modeled as the Poisson approximation in (11) that matches the first and second moments, so the SP model approaches the asymptotic normal distribution in (21) of the precorrected data in (19) for large n . Intuitively, this suggests that SP estimators approach ML estimators for large n , and consequently, they are asymptotically efficient. Next, noting the SD approximation of a normal variate is exact, one could also expect SD estimators to approach ML estimators asymptotically; so, they should also be asymptotically efficient.

To summarize, we have shown that all OP^- , SP^- and SD are asymptotically unbiased, and we have derived their asymptotic covariances (24), (25) and (27). We have also shown that SP^- and SD are asymptotically efficient.

B. Comparison of Covariances

We compare the asymptotic variances of OP^- and SP^- (equivalently, SD) estimators for a simple one-parameter example introduced in (16) (see [5], [11] for a similar comparison in a *transmission* case). Using (24) and (25), one

obtains asymptotic variances (or approximate variances for high counts) for OP^- and SP^- :

$$\text{Var}\{\hat{\theta}^{OP^-}\} = \frac{\sum_{i=1}^N a_i^2 \frac{a_i \theta^{\text{true}} + s_i + 2r_i}{(a_i \theta^{\text{true}} + s_i)^2}}{\left(\sum_{i=1}^N \frac{a_i^2}{a_i \theta^{\text{true}} + s_i}\right)^2}$$

and

$$\text{Var}\{\hat{\theta}^{SP^-}\} = \left(\sum_{i=1}^N \frac{a_i^2}{a_i \theta^{\text{true}} + s_i + 2r_i}\right)^{-1}.$$

Using the Schwartz inequality [31, pp. 107], one can show

$$\text{Var}\{\hat{\theta}^{OP^-}\} \geq \text{Var}\{\hat{\theta}^{SP^-}\} \quad (28)$$

where equality holds if and only if the $(a_i \theta^{\text{true}} + s_i)/(a_i \theta^{\text{true}} + s_i + 2r_i)$ ratios are equal, which is impossible as long as $r_i > 0$ and $s_i > 0$. Therefore, the inequality (28) is strict; the variance of OP^- is higher than that of SP^- (or SD). This is corroborated by empirical results for a multi-parameter case in Section VII.

VI. RECONSTRUCTION ALGORITHMS

After choosing a suitable likelihood approximation, one needs an algorithm to maximize the corresponding objective function for ML or PL estimation. It is straightforward to use globally convergent (and monotonic) algorithms⁵ such as SPS [32] and ML-EM [34], [35] for PR, OP^+ , SP^+ and SD, all of which have concave log-likelihoods. However, the new OP^- and SP^- models can have nonconcave log-likelihood functions when negative sinogram values are present. The algorithms need some modifications to ensure monotonicity for the nonconcave case as well. Monotonicity is one of the most desirable properties to enable at least a locally optimal reconstruction.

A large class of monotonic iterative algorithms (including SPS and ML-EM) are based on the ‘‘optimization transfer principle’’: at each iteration we choose a surrogate function that is easier to maximize than the original objective function, and then maximize that surrogate. To ensure monotonicity, the surrogate function is chosen so that increasing the surrogate guarantees the increase of the original objective function (for sufficient conditions for such surrogates, see [19], [36], [37]).

The idea for extending the algorithms to allow negative y_i 's is to choose a linear surrogate when a marginal log-likelihood is convex. That is, for OP^- , if $y_i < 0$, a tangent line to $h_i^{OP^-}$ at a current iterate l_i^n in projection domain

$$q_i^{OP^-}(l; l_i^n) = \dot{h}_i^{OP^-}(l_i^n)(l - l_i^n) + h_i^{OP^-}(l_i^n) \quad (29)$$

is a proper surrogate for $h_i^{OP^-}$ in light of [36, Eq. 7] since $q_i^{OP^-}$ lies below for all $l \geq 0$ due to convexity of $h_i^{OP^-}$, as illustrated in Fig. 2.

The same principle applies to SP^- when $y_i + 2r_i < 0$. We derive modified SPS and ML-EM applicable to OP^- and SP^- , using a linear surrogate (29) when needed.

⁵Ordered subsets algorithms [32], [33] can also be used with the aim of accelerating convergence speeds at the expense of monotonicity or global convergence.

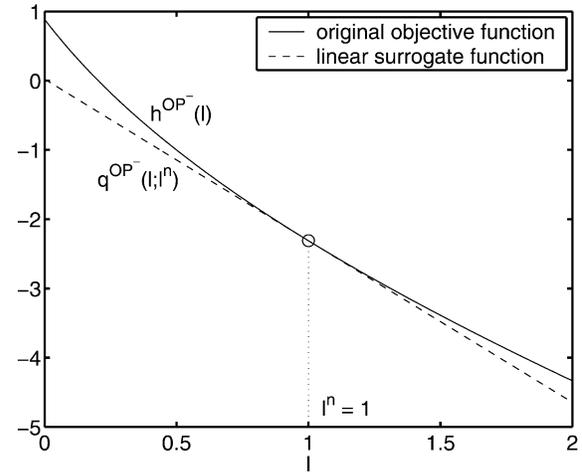


Fig. 2. Illustration of a linear surrogate q^{OP^-} [see (29)] at $l^n = 1$ for an OP^- log-likelihood h^{OP^-} for a negative value $y_i < 0$. The concave surrogate q^{OP^-} lies below the objective h^{OP^-} that is convex. One can see that $q^{OP^-}(l; l^n) \geq q^{OP^-}(l^n; l^n)$ implies that $h^{OP^-}(l) \geq h^{OP^-}(l^n)$ for $l \geq 0$.

A. SPS for OP^- and SP^- (and SD)

We consider the PL objective function Φ in (7) with a quadratic penalty for simplicity:

$$R(\boldsymbol{\lambda}) = \frac{\beta}{2} \sum_{j=1}^p \sum_{k \in \mathcal{N}_j} \omega_{jk} \frac{(\lambda_j - \lambda_k)^2}{2} \quad (30)$$

where $\beta \geq 0$ is a regularization parameter that controls the smoothness of the reconstructed image, \mathcal{N}_j denotes the neighborhood of the j th pixel, and ω_{jk} is a weighting factor. A monotonic SPS method for OP^- and SP^- is readily derived following [32] with (29). The resulting algorithm differs only slightly from the ordinary SPS algorithm in [32], and uses the following iteration:

$$\lambda_j^{n+1} = \left[\lambda_j^n + \frac{1}{d_j^{\text{SPS}}(\boldsymbol{\lambda}^n)} \frac{\partial \Phi(\boldsymbol{\lambda}^n)}{\partial \lambda_j} \right]_+ \quad (31)$$

with

$$d_j^{\text{SPS}}(\boldsymbol{\lambda}) = \sum_{i=1}^N a_{ij} a_i c_i(l_i(\boldsymbol{\lambda})) + 2\beta \sum_{k \in \mathcal{N}_j} \omega_{jk} \quad (32)$$

where $a_i = \sum_{j=1}^p a_{ij}$ and

$$c_i(l) = \begin{cases} 2 \frac{[h_i(l) - h_i(0) - l \dot{h}_i(0)]}{l^2}, & l > 0, x_i > 0 \\ -\ddot{h}_i(0), & l = 0, x_i > 0 \\ 0, & x_i \leq 0, \end{cases} \quad (33)$$

in which h_i represents $h_i^{OP^-}$ for OP^- and $h_i^{SP^-}$ for SP^- , and we define

$$x_i \triangleq \begin{cases} y_i, & \text{for } OP^- \\ y_i + 2r_i, & \text{for } SP^-. \end{cases} \quad (34)$$

The only difference from the ordinary SPS method (using optimum curvatures) in [32] is that here we set c_i to zero for $x_i < 0$ in (33) (note that $x_i < 0$ never occurs for previous zero-thresholding or prompt models). So one can easily modify existing codes to apply OP^- or SP^- . Nonquadratic penalties are included as in [32].

Being constructed by the optimization transfer principle, the iteration (31) increases the objective function Φ every iteration.

Since the step (32) requires an ‘‘extra’’ backprojection, we often forego strict monotonicity by replacing the curvatures c_i with the following precomputed values,

$$c_i(l) = \begin{cases} -\ddot{h}_i(\hat{l}_i), & x_i > 0 \\ 0, & x_i \leq 0 \end{cases}$$

where $\hat{l}_i = \arg \max_{l \geq 0} h_i(l) = [y_i - s_i]_+$. This allows d_j^{SPS} to be computed prior to iterating.

Paraboloidal surrogates algorithms for SD were developed in [14, sec. 5.6]. A monotonic SPS version has the form of (31) and (32) with the following curvatures,

$$c_i(l) = \begin{cases} -\ddot{h}_i^{\text{SD}}(l_i^*), & y_i = 0, l_i^* > 0 \\ -\ddot{h}_i^{\text{SD}}(l_i^{**}), & y_i = -1, l_i^{**} > 0 \\ t_i(l), & \text{otherwise} \end{cases}$$

and

$$t_i(l) = \begin{cases} 2 \frac{[h_i^{\text{SD}}(l) - h_i^{\text{SD}}(0) - l\dot{h}_i^{\text{SD}}(l)]}{l^2}, & l > 0 \\ -\ddot{h}_i^{\text{SD}}(0), & l = 0 \end{cases}$$

with $l_i^* = 7/9 - 4r_i(r_i + s_i)$ and $l_i^{**} = x_0^2 - 1 - r_i(r_i + s_i)$ where $x_0 \approx -1.1193219$ is a root of a polynomial [14, Appendix E]. In this case for SD, the following precomputed curvatures can be used for saving computation at the expense of monotonicity,

$$c_i(l) = -\ddot{h}_i^{\text{SD}}(\hat{l}_i)$$

where $\hat{l}_i = \arg \max_{l \geq 0} h_i^{\text{SD}}(l)$, or one could use a simple estimate $\hat{l}_i \approx [y_i - s_i]_+$.

B. Variation of ML-EM for OP⁻ and SP⁻

Following the derivation of ML-EM in a surrogates framework in [38], using (29), leads to a variation of ML-EM for OP⁻ and SP⁻. Although we used SPS rather than ML-EM variants for the results in Section VII, we provide the ML-EM variants for completeness as follows (see Appendix for derivation):

$$\lambda_j^{n+1} = \frac{\lambda_j^n}{d_j^{\text{EM}}(\boldsymbol{\lambda}^n)} \sum_{i=1}^N \frac{a_{ij}[x_i]_+}{\bar{x}_i(\boldsymbol{\lambda}^n)} \quad (35)$$

or

$$\lambda_j^{n+1} = \lambda_j^n + \frac{\lambda_j^n}{d_j^{\text{EM}}(\boldsymbol{\lambda}^n)} \frac{\partial L(\boldsymbol{\lambda}^n)}{\partial \lambda_j}$$

with

$$d_j^{\text{EM}}(\boldsymbol{\lambda}) = \sum_{i=1}^N a_{ij} \left(1 + \frac{[-x_i]_+}{\bar{x}_i(\boldsymbol{\lambda})} \right) \quad (36)$$

where x_i is defined in (34), and

$$\bar{x}_i(\boldsymbol{\lambda}) \triangleq l_i(\boldsymbol{\lambda}) + b_i \quad (37)$$

where

$$b_i \triangleq \begin{cases} s_i, & \text{for OP} \\ s_i + 2r_i, & \text{for SP.} \end{cases} \quad (38)$$

This variation of ML-EM reverts to classic ML-EM [35], [39] as a special case for nonnegative sinogram values. Regularization can also be incorporated as in [38]. The steps (36) require an extra backprojection each iteration compared to classic ML-EM [35], [39].

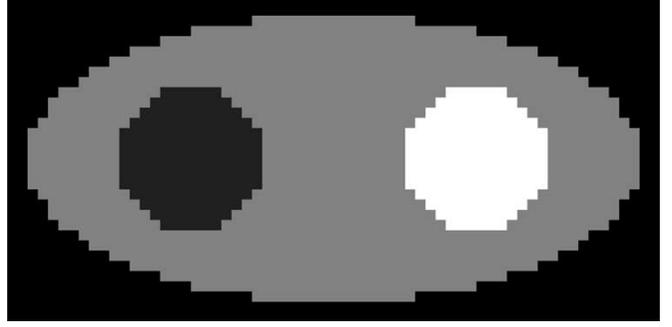


Fig. 3. Digital phantom used in simulations. The background, left cold disc, and right hot disc have relative emission activities of 2, 0.5, and 4, respectively.

VII. SIMULATIONS

A. Methods

To compare the bias and variance properties of the estimators (OP⁻, OP⁺, SP⁻, SP⁺, and SD), we simulated two-dimensional (2-D) PET emission scans. The PR model was also included for comparison purposes since in this simulation we had access to Y_i^{prompt} and Y_i^{delay} separately.

The synthetic emission phantom shown in Fig. 3 was used; its warm background, left cold disc, and right hot disc had relative emission activities of 2, 0.5, and 4, respectively. The sinograms had 192 radial bins and 120 angles uniformly sampled over 180 degrees. The system matrix was generated using ASPIRE [40]; the system geometry was approximated with 3-mm-wide strip integrals and 3-mm ray spacing. We simulated nonuniform detector efficiencies using pseudorandom log-normal variates with standard deviation of 0.3. Attenuation was not considered in this simulation. The reconstructed images were 64×32 with 9-mm pixels. The *known* r_i and s_i factors corresponded to a uniform field of 60% randoms and 10% scatters, respectively.⁶

The specific aim of the simulation was to compare biases for low counts and to compare variances for high counts. We performed two studies with $2K$ and $2M$ total counts. We generated 500 realizations of pseudorandom emission measurements according to (1) with (2) and (3). For each realization, images were reconstructed using 100 iterations of the SPS method for $2K$ counts, and using 40 iterations of the SPS method after 10 iterations of ordered subsets SPS (with 8 subsets) [32] for $2M$ counts. The FBP reconstruction for each realization served as an initial image for the iterations. The number of iterations was determined by looking at objective function values over iteration for a few realizations to ensure that convergence was reasonably achieved. For initial FBP reconstructions, a Hanning filter was used with such a cut-off frequency that their impulse responses were of 3 pixels full-width at half-maximum (FWHM).

For regularization, we used a second-order quadratic penalty (30) where ω_{jk} is 1 for horizontal or vertical neighborhoods, $1/\sqrt{2}$ for diagonal neighborhoods, and 0 otherwise. It is important to match the spatial resolution in reconstructed images for a fair comparison of different estimators. Penalty functions can be designed to achieve spatially uniform resolution [41]–[43]. However, in this simulation, we used a simpler hybrid technique

⁶The fractions in % denote (mean of randoms)/(mean of trues) and (mean of scatters)/(mean of trues), respectively.

[43] consisting of two steps: 1) for each method, we adjusted a global regularization parameter β so that the local impulse response⁷ at the center pixel was of 1.5 pixels FWHM, and then performed PL reconstructions; 2) we applied a 2-D Gaussian postsmoothing filter to the PL reconstructions so that the overall local impulse response (at the center pixel), which is the convolution of the postsmoothing filter and the original local impulse response (of 1.5 pixels FWHM), achieved a target resolution of 3 pixels FWHM.

This technique enables us to obtain reconstructions with various target resolutions by simply changing the postsmoothing filter. As the postsmoothing filter becomes wider (higher FWHM), the overall resolution becomes more uniform spatially since postsmoothing dominates the overall response. To check the spatial uniformity, the overall resolutions at every third pixel were investigated and it was found that, except the 2 pixel wide strip along the phantom boundary, each pixel achieved the target resolution (3 pixels FWHM) within 5% errors for all estimators—reasonably uniform resolution.

B. Results

Fig. 4(a) shows the profiles through the sample mean images of different estimators for $2K$ counts—very low counts. Both OP^+ and SP^+ showed large positive systematic biases particularly in the cold spot (pixels 12–25) and near the ends (phantom boundary). Zero-thresholding in sinogram domain contributes to the positive bias since the rays passing through those regions (cold spot and boundary) have low counts, as discussed in Section IV. Overall the systematic bias of OP^+ was slightly larger than that of SP^+ , as predicted in Section IV.

On the other hand, other methods (OP^- , SP^- , SD, and PR) seem reasonably free of such a bias. However, some positive biases are present in the cold spot (pixels 12–25) for OP^- , SP^- , SD, and PR. The positive bias in the cold spot is mainly due to the interaction of the image-domain nonnegativity constraints and the large variances, which causes the nonnegativity constraints to be active frequently and, consequently, increases image mean values. Note the bias is not due to zero-thresholding in sinogram domain since PR does not require any such thresholding. The coefficients of variation for those methods are larger than 100% in the cold spot in Fig. 4(b). In fact, the positive biases in the cold spot for OP^+ and SP^+ are caused by both sinogram-domain zero-thresholding and image-domain nonnegativity combined with large variances. Also, note small negative biases in background and hot regions for OP^- , SP^- , SD, and PR in Fig. 4(a). Our hypothesis is that the positive bias in the cold spot tends to decrease image values in other regions since the reconstruction methods try to make projections of image values close to given sinogram data. However, it is hard to analyze the effects of

⁷The approximate expression for the local impulse response, which could be interpreted as the *point spread function*, of an implicitly defined estimator was given in [41, Eq. 14]. It can be computed efficiently using 2-D fast Fourier transforms by assuming local shift-invariance as in [42, Eq. 9]. All resolutions in this paper (except those of simple linear FBP reconstruction) were computed as FWHM of the local impulse response (at a specific pixel) obtained using the methods in [41], [42]. One might doubt the feasibility of the approximate expressions in a low-counts-per-ray case where nonnegativity constraints are often active; however, in Section VII, (the sample means of) reconstructed images seem to have reasonably matched resolutions for different methods even for extremely low counts per ray.

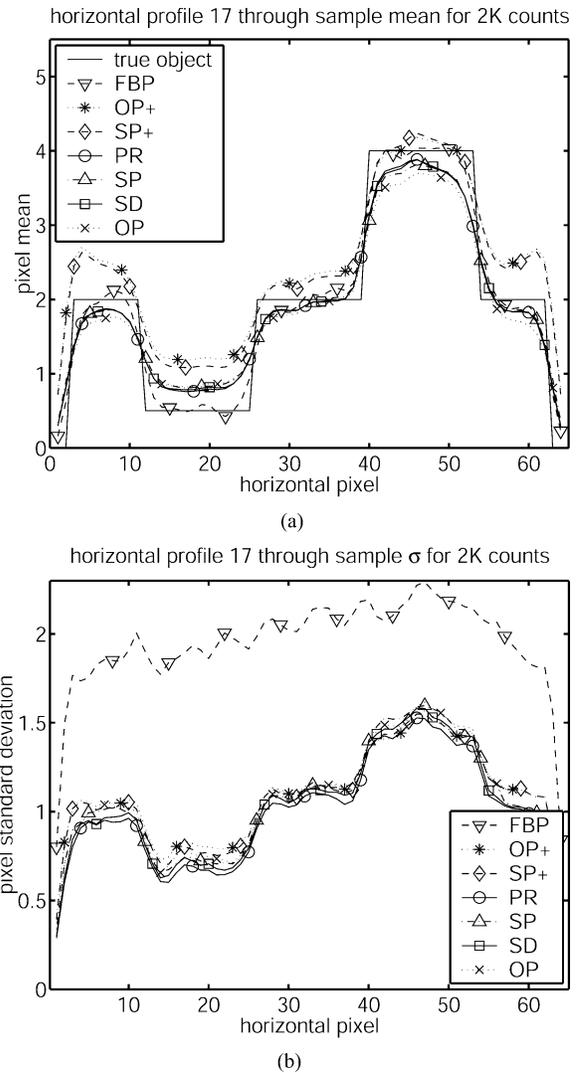


Fig. 4. (a) Horizontal profile through sample mean of estimators for $2K$ counts. (b) Horizontal profile through sample standard deviation of estimators for $2K$ counts.

image-domain nonnegativity constraints or to study them even experimentally since we would need, for comparison purposes, new models and algorithms (like NEG-ML in [18]) allowing negative image values. Further investigation is needed, and it would be interesting future work. To summarize, the results showed that both SP^- and SD were comparable to PR, the baseline reconstruction, and that they were free of systematic bias *caused by sinogram-domain zero-thresholding* which appeared in SP^+ and OP^+ .

As shown in Fig. 4(a), FBP was nearly unbiased since image-domain nonnegativity constraints are not imposed on the FBP reconstruction. However, it showed significantly large variances in Fig. 4(b). Although not shown here, we found that even if FBP is constrained by image nonnegativity, it still shows larger bias and variance than other methods [44].

Fig. 5(a) shows the profiles through the sample mean images of different estimators for $2M$ counts—high counts. All of the methods are seen to be unbiased, as predicted from the analysis in Section V. Fig. 5(b) shows profiles through the sample standard deviation images. FBP again showed the highest standard

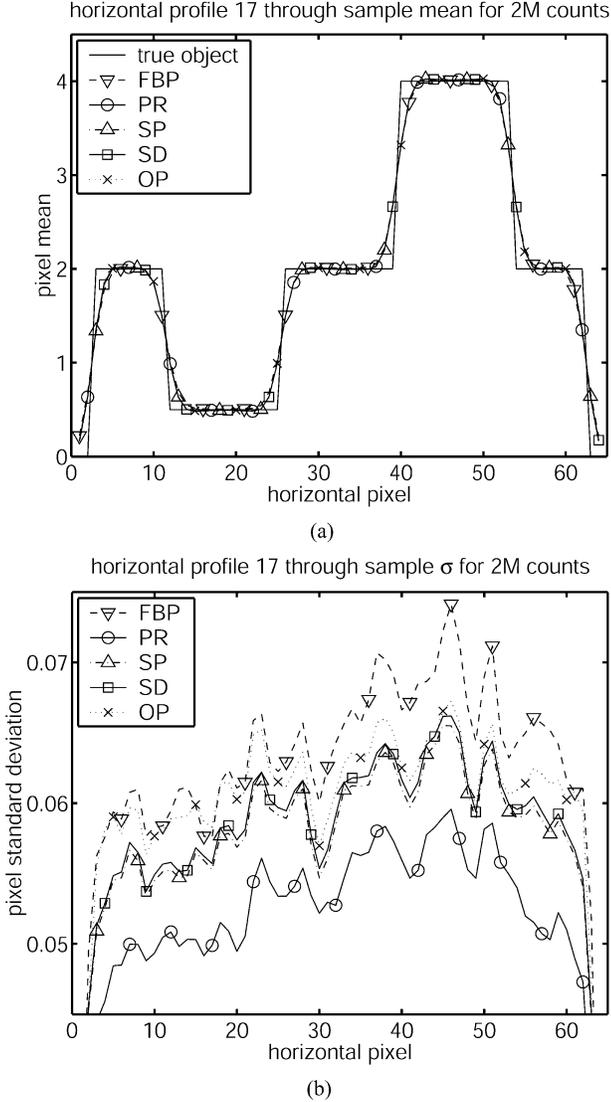


Fig. 5. (a) Horizontal profile through sample mean of estimators for $2M$ counts. (b) Horizontal profile through sample standard deviation of estimators for $2M$ counts.

deviation and PR showed the lowest as expected. SP^- and SD showed similar performance, and OP^- led to higher standard deviation than both of them. These empirical results corroborate the analysis of asymptotic variance in Section V. For each pixel, we computed the ratios of the sample standard deviation of different methods to the sample standard deviation of PR (see [44] for histograms of the ratios), and the means (over the entire image) of the ratios were 1.20 for FBP, 1.16 for OP^- , 1.11 for OP^+ , and 1.12 for SD. This also supports the claim that both SP^- and SD lead to less variance than OP^- (and FBP).

Whereas SP^- performed comparably to SD, the computation time for SP^- reconstruction was shorter than SD by 3–20% (depending on curvature type and counts) for the image and sinogram size here in our C and MATLAB implementation. However, as the image and sinogram size increases, the difference in computation would become smaller since projection and back-projection operations will contribute more significantly to the computational cost.

VIII. CONCLUSION

We proposed new log-likelihood approximations (SP^- and OP^-) for randoms-precorrected PET emission image reconstruction by allowing negative sinogram values and also developed algorithms (SPS and ML-EM variants) for the new models. The new methods are free of the positive systematic bias that degrades SP^+ and OP^+ images. The positive biases appearing in SP^+ and OP^+ are more distinguishable in low counts per ray regions such as cold spots, the boundary of an object, or high attenuation regions rather than depending solely on total counts. Our new models seem particularly promising for fully three-dimensional PET emission scans where AC rates are high and photon counts per ray can be low, essentially for newer scanners with small crystals.

The new SP^- model yields less variance (than OP^- and FBP). Its performance is comparable to SD in terms of systematic bias and variance; yet its implementation is simpler. Indeed, when implemented with the usual ordered-subsets approach [23], the modified OS-SPS algorithm presented in this paper for the new SP^- model has essentially the same compute complexity as the popular OS-EM method for PET.

We recommend the PR method if the prompt and the randoms data are accessible separately; however, if only randoms-precorrected data are available, the new SP^- is our recommended method.

APPENDIX

DERIVATION OF A VARIATION OF ML-EM FOR OP^- AND SP^-

Define $L^+(\boldsymbol{\lambda}; \mathbf{Y}) \triangleq \sum_{i:x_i \geq 0} h_i(l_i(\boldsymbol{\lambda}))$ and $L^-(\boldsymbol{\lambda}; \mathbf{Y}) \triangleq \sum_{i:x_i < 0} h_i(l_i(\boldsymbol{\lambda}))$ with $h_i(l) = x_i \log(l + b_i) - (l + b_i)$ where l_i , x_i , and b_i are defined in (5), (34), and (38), respectively. Then, by concavity of \log [38]

$$\begin{aligned} L^+(\boldsymbol{\lambda}; \mathbf{Y}) &= \sum_{i:x_i \geq 0} x_i \log \left(\sum_{j=1}^p \frac{a_{ij} \lambda_j^n}{\bar{x}_i(\boldsymbol{\lambda}^n)} \frac{\lambda_j}{\lambda_j^n} \bar{x}_i(\boldsymbol{\lambda}^n) \right. \\ &\quad \left. + \frac{b_i}{\bar{x}_i(\boldsymbol{\lambda}^n)} \bar{x}_i(\boldsymbol{\lambda}^n) \right) - \bar{x}_i(\boldsymbol{\lambda}) \\ &\geq \sum_{i:x_i \geq 0} \sum_{j=1}^p \frac{x_i a_{ij} \lambda_j^n}{\bar{x}_i(\boldsymbol{\lambda}^n)} \log \lambda_j - a_{ij} \lambda_j + C^+ \\ &\triangleq Q^+(\boldsymbol{\lambda}; \boldsymbol{\lambda}^n) \end{aligned}$$

where \bar{x}_i is defined in (37) and C^+ is a constant with respect to $\boldsymbol{\lambda}$. On the other hand, since $h_i(\cdot)$ is convex for $x_i < 0$

$$\begin{aligned} L^-(\boldsymbol{\lambda}; \mathbf{Y}) &\geq \sum_{i:x_i < 0} \dot{h}_i(l_i(\boldsymbol{\lambda}^n))(l_i(\boldsymbol{\lambda}) - l_i(\boldsymbol{\lambda}^n)) + h_i(l_i(\boldsymbol{\lambda}^n)) \\ &= \sum_{i:x_i < 0} \sum_{j=1}^p \left(\frac{x_i}{\bar{x}_i(\boldsymbol{\lambda}^n)} - 1 \right) a_{ij} \lambda_j + C^- \\ &\triangleq Q^-(\boldsymbol{\lambda}; \boldsymbol{\lambda}^n) \end{aligned}$$

where C^- is a constant with respect to $\boldsymbol{\lambda}$. Since

$$\begin{aligned} Q(\boldsymbol{\lambda}; \boldsymbol{\lambda}^n) &\triangleq Q^+(\boldsymbol{\lambda}; \boldsymbol{\lambda}^n) + Q^-(\boldsymbol{\lambda}; \boldsymbol{\lambda}^n) \\ &\leq L^+(\boldsymbol{\lambda}; \mathbf{Y}) + L^-(\boldsymbol{\lambda}; \mathbf{Y}) = L(\boldsymbol{\lambda}; \mathbf{Y}) \end{aligned}$$

and $Q(\boldsymbol{\lambda}^n; \boldsymbol{\lambda}^n) = L(\boldsymbol{\lambda}^n; \mathbf{Y})$, one can show $Q(\cdot; \boldsymbol{\lambda}^n)$ is a proper surrogate [36] for $L(\cdot; \mathbf{Y})$. The surrogate $Q(\cdot; \boldsymbol{\lambda}^n)$ can be maximized by setting its derivative to zero, and as a result, its max-

imizer λ^{n+1} is calculated as (35). Because of the optimization transfer principle, this derivation ensures monotonicity.

The following is an alternative “intuitive” but not rigorous derivation. The partial derivatives of the log-likelihood function L^{SP-} or L^{OP-} at a nonnegative maximizer $\hat{\lambda}$ are, by the Karush–Kuhn–Tucker conditions [30, p. 310]

$$\frac{\partial}{\partial \lambda_j} L(\hat{\lambda}) = \sum_{i=1}^N a_{ij} \left(\frac{x_i}{\bar{x}_i(\hat{\lambda})} - 1 \right) \begin{cases} = 0, & \hat{\lambda}_j > 0 \\ \leq 0, & \hat{\lambda}_j = 0 \end{cases}$$

where x_i and \bar{x}_i are defined in (34) and (37), respectively. So, for $\hat{\lambda}_j > 0$

$$\sum_{i=1}^N a_{ij} = \sum_{i=1}^N \frac{a_{ij}}{\bar{x}_i(\hat{\lambda})} x_i = \sum_{i=1}^N \frac{a_{ij}}{\bar{x}_i(\hat{\lambda})} ([x_i]_+ - [-x_i]_+).$$

Moving the subtracted term to the other side (cf. [45], [46]) leads to the following:

$$\sum_{i=1}^N a_{ij} \left(1 + \frac{[-x_i]_+}{\bar{x}_i(\hat{\lambda})} \right) = \sum_{i=1}^N \frac{a_{ij} [x_i]_+}{\bar{x}_i(\hat{\lambda})}.$$

The ratio of these terms yields the multiplicative update (35). In other words, $\hat{\lambda}$ is a fixed point of the iteration (35), and so is it for $\hat{\lambda}_j = 0$.

ACKNOWLEDGMENT

The authors would like to thank M. Yavuz for encouraging us to resolve issues raised in [13], [14].

REFERENCES

- [1] 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. Assist. Tomogr.*, vol. 5, no. 3, pp. 391–400, 1981.
- [2] J. M. Ollinger and J. A. Fessler, “Positron emission tomography,” *IEEE Signal Processing Mag.*, vol. 14, no. 1, pp. 43–55, Jan. 1997.
- [3] D. G. Politte and D. L. Snyder, “Corrections for accidental coincidences and attenuation in maximum-likelihood image reconstruction for positron-emission tomography,” *IEEE Trans. Med. Imag.*, vol. 10, pp. 82–89, Mar. 1991.
- [4] T. J. Spinks, T. Jones, M. C. Gilardi, and J. D. Heather, “Physical performance of the latest generation of commercial positron scanner,” *IEEE Trans. Nucl. Sci.*, vol. 35, pp. 721–725, Feb. 1988.
- [5] M. Yavuz and J. A. Fessler, “Statistical image reconstruction methods for randoms-precorrected PET scans,” *Med. Imag. Anal.*, vol. 2, no. 4, pp. 369–378, Dec. 1998.
- [6] E. Ü Mumcuoğlu, R. M. Leahy, and S. R. Cherry, “Bayesian reconstruction of PET images: Methodology and performance analysis,” *Phys. Med. Biol.*, vol. 41, no. 9, pp. 1777–1807, Sept. 1996.
- [7] 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. Tomogr.*, vol. 10, no. 5, pp. 845–850, 1986.
- [8] E. Ü Mumcuoğlu, R. Leahy, S. R. Cherry, and Z. Zhou, “Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images,” *IEEE Trans. Med. Imag.*, vol. 13, pp. 687–701, Dec. 1994.
- [9] D. Hogg, K. Thielemans, S. Mustafovic, and T. J. Spinks, “A study of bias for various iterative reconstruction methods in PET,” in *Proc. IEEE Nuclear Science Symp. Medical Imaging Conf.*, vol. 3, 2002, pp. 1519–1523.
- [10] D. F. Yu and J. A. Fessler, “Mean and variance of coincidence photon counting with deadtime,” *Nucl. Instr. Meth. Phys. Res. A.*, vol. 488, no. 1–2, pp. 362–374, Aug. 2002.
- [11] M. Yavuz and J. A. Fessler, “New statistical models for randoms-precorrected PET scans,” in *Information Processing in Medical Imaging*, J. Duncan and G. Gindi, Eds. Berlin, Germany: Springer-Verlag, 1997, vol. 1230, Lecture Notes in Computer Science, pp. 190–203.
- [12] —, “Penalized-likelihood estimators and noise analysis for randoms-precorrected PET transmission scans,” *IEEE Trans. Med. Imag.*, vol. 18, pp. 665–674, Aug. 1999.
- [13] —, “Maximum likelihood emission image reconstruction for randoms-precorrected PET scans,” in *Proc. IEEE Nuclear Science Symp. Medical Imaging Conf.*, vol. 2, 2000, pp. 15/229–15/233.
- [14] M. Yavuz, “Statistical tomographic image reconstruction methods for randoms-precorrected PET measurements,” Ph.D. dissertation, Univ. Michigan, Ann Arbor, 1999.
- [15] J. Qi and R. M. Leahy, “Resolution and noise properties MAP reconstruction for fully 3-D PET,” *IEEE Trans. Med. Imag.*, vol. 19, pp. 493–506, May 2000.
- [16] C. Michel, X. Liu, S. Sanabria, M. Lonnew, M. Sibomana, A. Bol, C. Comtat, P. E. Kinahan, D. W. Townsend, and M. Defrise, “Weighted schemes applied to 3-D-OSEM reconstruction in PET,” in *Proc. IEEE Nuclear Science Symp. Medical Imaging Conf.*, vol. 3, 1999, pp. 1152–1157.
- [17] C. Michel, M. Sibomana, A. Bol, X. Bernard, M. Lonnew, M. Defrise, C. Comtat, P. E. Kinahan, and D. W. Townsend, “Preserving Poisson characteristics of PET data with weighted OSEM reconstruction,” in *Proc. IEEE Nuclear Science Symp. Medical Imaging Conf.*, vol. 2, 1998, pp. 1323–1329.
- [18] J. Nuys, S. Stroobants, P. Dupont, S. Vleugels, P. Flamen, and L. Mortelmans, “Reducing loss of image quality due to the attenuation artifact in uncorrected PET whole body images,” *J. Nuc. Med.*, vol. 43, no. 8, pp. 1054–1062, Aug. 2002.
- [19] K. Lange, D. R. Hunter, and I. Yang, “Optimization transfer using surrogate objective functions,” *J. Computational Graphical Statist.*, vol. 9, no. 1, pp. 1–20, Mar. 2000.
- [20] J. Qi, R. M. Leahy, C. Hsu, T. H. Farquhar, and S. R. Cherry, “Fully 3-D Bayesian image reconstruction for the ECAT EXACT HR+,” *IEEE Trans. Nucl. Sci.*, vol. 45, pp. 1096–1103, June 1998.
- [21] C. C. Watson, D. Newport, M. E. Casey, R. A. De Kemp, R. S. Beanlands, and M. Schmand, “Evaluation of simulation-based scatter correction for 3-D PET cardiac imaging,” *IEEE Trans. Nucl. Sci.*, vol. 44, pp. 90–97, Feb. 1997.
- [22] J. M. Ollinger, “Model-based scatter correction for fully 3D PET,” *Phys. Med. Biol.*, vol. 41, no. 1, pp. 153–76, Jan. 1996.
- [23] S. Ahn and J. A. Fessler, “Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms,” *IEEE Trans. Med. Imag.*, vol. 22, pp. 613–626, May 2003.
- [24] S. O. Rice, “Uniform asymptotic expansions for saddle point integrals—Application to a probability distribution occurring in noise theory,” *Bell Syst. Tech. J.*, vol. 47, pp. 1971–2013, Nov. 1968.
- [25] D. L. Snyder, C. W. Helstrom, A. D. Lanterman, M. Faisal, and R. L. White, “Compensation for readout noise in CCD images,” *J. Opt. Soc. Amer. A*, vol. 12, no. 2, pp. 272–283, Feb. 1995.
- [26] G. Arfken, *Mathematical Methods for Physicists*, 3rd ed. Orlando, FL: Academic Press, 1985.
- [27] J. A. Fessler, “Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography,” *IEEE Trans. Image Processing*, vol. 5, pp. 493–506, Mar. 1996.
- [28] C. R. Rao, *Linear Statistical Inference and its Applications*. New York: Wiley, 1973.
- [29] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, 2nd ed. New York: Springer-Verlag, 1998.
- [30] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA: Athena Scientific, 1999.
- [31] G. Strang, *Linear Algebra and Its Applications*. New York: Academic Press, 1980.
- [32] H. Erdoğan and J. A. Fessler, “Ordered subsets algorithms for transmission tomography,” *Phys. Med. Biol.*, vol. 44, no. 11, pp. 2835–2851, Nov. 1999.
- [33] H. M. Hudson and R. S. Larkin, “Accelerated image reconstruction using ordered subsets of projection data,” *IEEE Trans. Med. Imag.*, vol. 13, pp. 601–609, Dec. 1994.
- [34] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *J. Roy. Statist. Soc. Ser. B*, vol. 39, no. 1, pp. 1–38, 1977.
- [35] L. A. Shepp and Y. Vardi, “Maximum likelihood reconstruction for emission tomography,” *IEEE Trans. Med. Imag.*, vol. MI-1, pp. 113–122, Oct. 1982.

- [36] H. Erdoğan and J. A. Fessler, "Monotonic algorithms for transmission tomography," *IEEE Trans. Med. Imag.*, vol. 18, pp. 801–814, Sept. 1999.
- [37] S. Sotthivirat and J. A. Fessler, "Statistical reconstruction for digital holography," *J. Opt. Soc. Amer. A*, to be published.
- [38] A. R. De Pierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE Trans. Med. Imag.*, vol. 14, pp. 132–137, Mar. 1995.
- [39] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comput. Assist. Tomogr.*, vol. 8, no. 2, pp. 306–316, Apr. 1984.
- [40] J. A. Fessler. (1995) ASPIRE 3.0 User's Guide. A Sparse Iterative Reconstruction Library. Commun. and Signal Proc. Lab., Dept. of Elec. Eng. Comput. Sci. Univ. Michigan, Ann Arbor, MI. [Online]. Available: <http://www.eecs.umich.edu/~fessler>
- [41] J. A. Fessler and W. L. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction methods: Space-invariant tomographs," *IEEE Trans. Image Processing*, vol. 5, pp. 1346–1358, Sept. 1996.
- [42] J. W. Stayman and J. A. Fessler, "Regularization for uniform spatial resolution properties in penalized-likelihood image reconstruction," *IEEE Trans. Med. Imag.*, vol. 19, pp. 601–615, June 2000.
- [43] —, "Compensation for nonuniform resolution using penalized-likelihood reconstruction in space-variant imaging systems," *IEEE Trans. Med. Imag.*, vol. 23, pp. 269–284, Mar. 2004.
- [44] S. Ahn and J. A. Fessler, "Statistical emission image reconstruction for randoms-precorrected PET scans using negative sinogram values," in *Proc. IEEE Nuclear Science Symp. Medical Imaging Conf.*, 2003.
- [45] H. Lantéri, M. Roche, O. Cuevas, and C. Aime, "A general method to devise maximum-likelihood signal restoration multiplicative algorithms with nonnegativity constraints," *Signal Processing*, vol. 81, no. 5, pp. 945–974, May 2001.
- [46] H. Lantéri, M. Roche, and C. Aime, "Penalized maximum likelihood image restoration with positivity constraints: Multiplicative algorithms," *Inverse Prob.*, vol. 18, no. 5, pp. 1397–1419, Oct. 2002.