

**Statistical Tomographic Image Reconstruction Methods for  
Randoms-Precorrected PET Measurements**

by

**Mehmet Yavuz**

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
(Electrical Engineering:Systems)  
in The University of Michigan  
2000

Doctoral Committee:

Associate Professor Jeffrey A. Fessler, Chair  
Professor Alfred Hero  
Professor W. Leslie Rogers  
Professor Andrew E. Yagle

**This version is formatted single spaced to save paper when printing.  
It is *not* the official archived version.**

## ABSTRACT

Statistical Tomographic Image Reconstruction Methods for Randoms-Precorrected PET Measurements

by  
Mehmet Yavuz

Chair: Jeffrey A. Fessler

Medical imaging systems such as positron emission tomography (PET) and electronically collimated single positron emission tomography (SPECT) record particle emission events based on timing coincidences. These systems record accidental coincidence (AC) events simultaneously with the true coincidence events. Similarly in low light-level imaging, thermoelectrons generated by photodetector are indistinguishable from photoelectrons generated by photo-conversion, and their effect is similar to the AC events.

During PET emission scans, accidental coincidence (AC) events occur when photons that originate from separate positron-electron annihilations are mistakenly recorded as having arisen from the same annihilation. In PET, generally a significant portion of the collected data consists of AC events that are a primary source of background noise. Also, during PET transmission scans, photons that originate from different transmission sources cause AC events. In PET, the measurements are usually pre-corrected for AC events by *real-time* subtraction of the delayed window coincidences. Randoms subtraction compensates in mean for accidental coincidences, but destroys the Poisson statistics.

We develop statistical image reconstruction methods for randoms pre-corrected PET measurements using penalized maximum likelihood (ML) estimation. We introduce two new approximations to the complicated exact log-likelihood of the pre-corrected measurements: one based on a “shifted Poisson” model, and the other based on saddle-point approximations to the measurement probability mass function (pmf). We compare estimators based on the new models to the conventional data-weighted least squares (WLS) and conventional maximum likelihood (based on the ordinary Poisson (OP) model) using experiments, simulations and analytic approximations.

For transmission scans, we demonstrate that the proposed methods avoid the systematic bias of the WLS method, and lead to significantly lower variance than the conventional OP method. We also investigate the propagation of noise from the reconstructed attenuation maps into the emission images. Interestingly, the noise improvements in the emission images with the new methods are even greater than the improvements in the attenuation maps themselves. To corroborate the *empirical* studies, we develop *analytical* approximations to the reconstructed image covariance and we also develop analytical approximations for the propagation of noise from attenuation maps into the reconstructed emission images. The results of the analytic approximations are shown to be in good agreement with the experimental results supporting the improvements with the new methods.

Similarly, for the emission reconstructions, we demonstrate that the proposed methods lead to significantly lower variance than the conventional OP method and also avoid systematic positive bias of the OP method. Although the SP model is shown to be slightly biased for emission scans with very low count rates, the saddle-point model is free of any systematic bias and performs almost identically to the exact log-likelihood. Also, we investigate the bias-variance trade-offs of the models in 1-D by analyzing how close they perform to the “uniform” Cramer-Rao bounds.

The new methods offer improved image reconstruction in PET through more realistic statistical modeling, yet with negligible increase in computation over the conventional OP method.

© Mehmet Yavuz 2000  
All Rights Reserved

To my wife Sema

## ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor Professor Jeffrey A. Fessler for his enlightening and constructive guidance throughout my graduate study. His understanding, encouragement and moral support helped me at all stages of my graduate work, and made my Ph.D. research a lively learning experience.

I would like to thank TUBITAK for their financial support with scholarship for the first year of my graduate study. I would also like to thank to my advisor Professor Jeffrey Fessler, Professor Les Rogers and National Institute of Health for supporting me financially with research assistantship.

I would also like to express my gratitude to Professor Alfred Hero, Professor Les Rogers and Professor Andrew Yagle for serving in my committee and helping me with their ideas, Neal Clinthorne for his helpful suggestions, Web Stayman for his help with the modified quadratic penalty and my colleagues Hakan Erdogan, Web Stayman, Steve Titus and many others for sharing ideas and friendship.

Finally, I wish to thank to my parents, my brother, and my dear wife Sema for their loving support and encouragement.

# TABLE OF CONTENTS

<b>DEDICATION</b> . . . . .	ii
<b>ACKNOWLEDGEMENTS</b> . . . . .	iii
<b>LIST OF TABLES</b> . . . . .	vii
<b>LIST OF FIGURES</b> . . . . .	viii
<b>LIST OF APPENDICES</b> . . . . .	xiii
<b>CHAPTERS</b>	
1 Introduction . . . . .	1
1.1 Background and Motivation . . . . .	1
1.2 Organization of the Thesis . . . . .	4
1.3 Original Contributions . . . . .	5
2 Positron Emission Tomography (PET) Imaging . . . . .	6
2.1 Tomographic Image Reconstruction . . . . .	6
2.1.1 Filtered Backprojection . . . . .	8
2.2 PET Physics and System Description . . . . .	9
2.2.1 PET Imaging . . . . .	9
2.2.2 Attenuation . . . . .	10
2.2.3 Accidental Coincidence Events . . . . .	12
2.2.4 Scattered Events . . . . .	14
2.3 System and Measurement Model . . . . .	14
2.3.1 Emission Scan . . . . .	14
2.3.2 Transmission Scan . . . . .	16
2.4 Statistical Image Reconstruction Methods . . . . .	17
2.4.1 Maximum likelihood . . . . .	17
2.4.2 Penalized Maximum Likelihood . . . . .	18
3 Exact Log-Likelihood and Approximations . . . . .	22
3.1 Measurement Model . . . . .	22
3.2 Exact Log-Likelihood . . . . .	25
3.2.1 Infinite Summation Form of Exact Log-Likelihood . . . . .	25
3.2.2 Bessel Function Form of Exact Log-Likelihood . . . . .	25
3.3 Simple Approximations to the Likelihood . . . . .	27
3.3.1 Quadratic Approximations . . . . .	27

3.3.2	Ordinary Poisson (OP) Approximation . . . . .	29
3.3.3	Shifted Poisson (SP) Approximation . . . . .	29
3.4	Saddle-point (SD) Approximation . . . . .	30
3.5	Exact Log-likelihood for Prompt Coincidence Data . . . . .	36
4	PET Transmission Scans . . . . .	37
4.1	Introduction . . . . .	37
4.2	Exact Log-Likelihood . . . . .	38
4.3	Approximations to the Exact Log-Likelihood . . . . .	39
4.3.1	Quadratic Approximations . . . . .	39
4.3.2	Ordinary Poisson (OP) Approximation . . . . .	39
4.3.3	Shifted Poisson (SP) Approximation . . . . .	39
4.3.4	Saddle-point (SD) Approximation . . . . .	40
4.4	1-D Bias-Variance Analysis . . . . .	40
4.5	1-D Simulations . . . . .	42
4.6	Concavity and Convergence . . . . .	44
4.7	Log-likelihood Maximization : Coordinate Ascent Type Algorithms . . . . .	44
4.8	2-D Simulations . . . . .	46
4.8.1	Simulations . . . . .	46
4.8.2	Resolution <i>vs</i> Standard Deviation . . . . .	48
4.8.3	Comparison of SP and SD Models with Exact Model . . . . .	49
4.8.4	Zero-thresholding the Data . . . . .	50
4.8.5	Estimates of the AC Rates . . . . .	51
4.9	Experimental Results . . . . .	62
4.10	Covariance Approximations for Transmission Tomography . . . . .	66
4.11	Noise Propagation Into Emission Reconstruction . . . . .	67
4.11.1	Theory . . . . .	67
4.11.2	Results . . . . .	70
4.12	Conclusions . . . . .	72
5	PET Emission Scans . . . . .	79
5.1	Introduction . . . . .	79
5.2	Exact Log-Likelihood . . . . .	79
5.3	Approximations to the Exact Log-Likelihood . . . . .	81
5.3.1	Quadratic Approximations . . . . .	81
5.3.2	Ordinary Poisson (OP) Approximation . . . . .	81
5.3.3	Shifted Poisson (SP) Approximation . . . . .	81
5.3.4	Saddle-point (SD) Approximation . . . . .	81
5.4	Exact Log-likelihood for Prompt Data . . . . .	82
5.5	Concavity and Convergence . . . . .	82
5.6	Log-likelihood Maximization . . . . .	83
5.6.1	EM Type Algorithms . . . . .	83
5.6.2	Paraboloid Surrogates Maximization Algorithm . . . . .	85
5.7	1D Simulations . . . . .	87
5.8	2D Simulations . . . . .	87
5.9	Cramer-Rao Bounds . . . . .	106
5.9.1	Estimation of Bias Gradient . . . . .	107
5.9.2	Simulations . . . . .	107

5.10	Conclusions . . . . .	110
6	Conclusions and Future Work . . . . .	112
6.1	Conclusions . . . . .	112
6.2	Future Work . . . . .	114
	<b>APPENDICES . . . . .</b>	<b>116</b>
	<b>BIBLIOGRAPHY . . . . .</b>	<b>137</b>

## LIST OF TABLES

### Table

3.1	Sample mean, variance and $3^{rd}$ , $4^{th}$ , $5^{th}$ order central moments of different models compared with those of the exact distribution. . . . .	24
4.1	Local impulse response and the local sample standard deviation for the central pixel. . . . .	48
4.2	Empirical percent standard deviation and the approximate analytical percent standard deviation of emission reconstruction using ACFs based on the OP method and SP method (using both empirical transmission variance and approximate transmission variance and plug-in transmission variance) for different regions shown in Fig. 4.25. Last column shows the empirical percent noise of the regions due to only emission noise for two million counts per emission scan. . . . .	78

## LIST OF FIGURES

<b>Figure</b>	
2.1	Object $f(x, y)$ and its projection $p_\theta(R)$ at angle $\theta$ . . . . . 7
2.2	Transaxial view and a cross-section view (Section X-X) of a cylindrical PET scanner. A) Rod sources for transmission scan, B) Collimators for scatter rejection (septa), C) Detector crystals, D) Photomultiplier tubes. . . . . 9
2.3	Diagram of a PET detector system with coincidence detection between different detectors. . . . . 10
2.4	Different forms of attenuation in PET: A) Photoelectric absorption, B) Single scattering, C) Multiple scattering. . . . . 11
2.5	Photons $\gamma_1$ and $\gamma_2$ are attenuated through distances $L_1 - L$ and $L - L_2$ respectively. . . . . 12
2.6	Diagram of an Accidental Coincidence event . . . . . 12
2.7	Geometric system model showing the contribution of $j$ th pixel to the $n$ th detector pair. . . . . 15
3.1	Comparison of Gaussian, ordinary Poisson, shifted Poisson and Saddle Point models (-) (with the moments matched to the moments of precorrected measurements), with the empirical distribution (o) of precorrected measurements. From top to bottom: a) Gaussian model. b) Ordinary Poisson (OP) model. c) shifted Poisson (SP) model. d) Saddle Point (SD) approximation that will be introduced in section 3.4. . . . . 24
3.2	Comparison of exact log-likelihood function with objective functions of different models as a function of single projection across the reconstructed image. The proposed shifted Poisson model agrees with exact log-likelihood better than the quadratic and OP models. . . . . 31
3.3	Deformation of the contour $C^+$ in complex plane into a vertical line $C_0$ through saddle point $x_o$ and a semicircle $C_1$ around the left half plane at infinity. The singularities of the integrand are at $z = 0$ and $z = \infty + j0$ for $k \geq 0$ . . . . . 32
3.4	Representative comparison of exact log-likelihood function with objective functions of different models as a function of line integral $l_n(\mu)$ . Randoms rate is 5%. The proposed saddle-point approximation agrees with exact log-likelihood significantly better than the other models. . . . . 35
4.1	Comparison of analytical approximations and empirical results for bias and variance. Upper figure shows that WLS estimator is systematically negatively biased especially for low counts. Lower figure shows that ordinary Poisson model yields higher standard deviation than both other estimators. . . . . 43

4.2	Simulated abdomen attenuation map. . . . .	46
4.3	Horizontal profile through the sample mean images for abdomen phantom. The WLS method has a systematic negative bias. However, the ordinary Poisson (OP), shifted Poisson (SP) , saddle-point (SD) , exact (EX) and prompt (PR) methods are free of this systematic negative bias. . . . .	47
4.4	Histogram of the ratio of standard deviations of different methods over the OP method in reconstructions of the abdomen phantom. The ordinary Poisson (OP) method yields, on the average, 15% higher standard deviation than the shifted Poisson (SP) , saddle-point (SD) and exact (EX) methods, and 39% more standard deviation than the prompt (PR) method. . . . .	52
4.5	Simulated thorax attenuation map. . . . .	53
4.6	Profile through the reconstructed image of the exact log-likelihood (EX) method using 3.6 million counts transmission scan. Profiles near zero attenuation level correspond to the difference of the profiles between the EX method and the SP and the SD methods. It can be seen that there is some noticeable difference between the reconstructions with SP and EX method. . . . .	53
4.7	$E_1$ error norm between the exact log-likelihood (EX) method and the SP and SD methods for each noisy realization. . . . .	54
4.8	$E_2$ error norm between the exact log-likelihood (EX) method and the SP and SD methods for each noisy realization. . . . .	54
4.9	$E_\infty$ error norm between the exact log-likelihood (EX) method and the SP and SD methods for each noisy realization. . . . .	55
4.10	Comparison of analytical approximations and empirical results for “zero-thresholded” data. Upper figure shows that ordinary Poisson model is negatively biased compared to Fig. 4.1, due to thresholding. . . . .	55
4.11	Horizontal profile through the sample mean images for abdomen phantom, obtained by using zero-thresholded data. The ordinary Poisson model leads to systematic negative bias, especially for interior regions of the reconstructed image. The shifted Poisson model estimator is free of systematic bias. . . . .	56
4.12	Histograms of the ratio of standard deviations for abdomen phantom, obtained by using zero-thresholded data. The ordinary Poisson model still leads to higher standard deviation than the shifted Poisson model, (on the average 11%). . . . .	57
4.13	Separately collected sinograms (160 radial bins and uniformly spaced 192 angles). Clockwise from the upper left: (a) Delayed events of blank scan. (b) Delayed events of transmission scan. (c) Prompt events of transmission scan. (d) Prompt events of blank scan. . . . .	58
4.14	Phantom used in the PET system for transmission scan. . . . .	59
4.15	Scatter plot of delayed coincidence event of blank and transmission scans. . . . .	59
4.16	Horizontal profile through the sample mean images for abdomen phantom using constant AC rates. The constant AC rates approximation does not introduce any systematic bias to the estimators. . . . .	60
4.17	Histograms of the ratio of standard deviations of shifted Poisson estimators, for abdomen phantom. Using the constant AC rates approximation slightly increases the variance of the SP and SD estimators. . . . .	61

4.18	Reconstruction of attenuation map for the slice of interest from 5 hour transmission scan. . . . .	62
4.19	Horizontal profile 66 through the sample mean images for abdomen phantom. The WLS method has a systematic negative bias. The ordinary Poisson (OP) and shifted Poisson (SP) methods appear free of this systematic negative bias. . . . .	63
4.20	Sample standard deviation image of SP method from 100 transmission scans. . . . .	64
4.21	Ratio of sample standard deviation images of OP method to SP method from 100 transmission scans. . . . .	65
4.22	Histogram of the ratio of standard deviations in reconstructed attenuation maps. The ordinary Poisson (OP) method yields, on the average, about 11% higher standard deviation than the proposed shifted Poisson (SP) method. . . . .	65
4.23	Empirical standard deviation (with error bars) and the approximate standard deviation of OP method for pixels along horizontal profile 90 through the attenuation map. . . . .	68
4.24	Empirical standard deviation (with error bars) and the approximate standard deviation of SP method for pixels along horizontal profile 90 through the attenuation map. . . . .	69
4.25	Emission phantom with several rectangular regions for noise computation. . . . .	71
4.26	Empirical sample mean of emission images reconstructed with ACFs based on 100 different estimates of $\hat{\mu}_{SP}$ . . . . .	72
4.27	Sample standard deviation image of emission reconstruction with ACFs based on SP method. . . . .	73
4.28	Ratio of sample standard deviation images of emission reconstruction with ACFs based on OP method and SP method. . . . .	74
4.29	Histogram of the ratio of standard deviations in the reconstructed emission images with ACFs based on OP model and SP model. Attenuation correction factors based on the OP model yielded, about 20% higher standard deviation than the SP model on average. . . . .	75
4.30	Empirical standard deviation (with error bars) and the approximate standard deviation of OP method (using both empirical transmission variance and approximate transmission variance) for pixels along horizontal profile 90 through the reconstructed emission images. . . . .	76
4.31	Empirical standard deviation (with error bars) and the approximate standard deviation of SP method (using both empirical transmission variance and approximate transmission variance) for pixels along a horizontal profile 90 through the reconstructed emission images. . . . .	77
5.1	Sample mean of OP, SP and SD methods from 300 realizations where $\lambda_{true} = 1$ . . . . .	88
5.2	Sample standard deviation of OP, SP and SD methods from 300 realizations where $\lambda_{true} = 1$ . . . . .	89
5.3	Sample mean of OP, SP, SD and Exact methods from 300 realizations (with nonuniform $g_n$ and $r_n$ ) where $\lambda_{true} = 1$ . . . . .	90
5.4	Sample standard deviation of OP, SP, SD and Exact methods from 300 realizations (with nonuniform $g_n$ and $r_n$ ) where $\lambda_{true} = 1$ . . . . .	91
5.5	Results of OP, SP, SD and Exact methods with <i>noise free</i> data (with nonuniform $g_n$ and $r_n$ ) where $\lambda_{true} = 1$ . . . . .	92
5.6	Simulated emission phantom. . . . .	92

5.7	Sample mean images of different methods from 300 realization with 50,000 counts per scan. . . . .	93
5.8	Sample standard deviation images of different methods from 300 realization with 50,000 counts per scan. . . . .	94
5.9	Profile through the sample mean images of different methods from 300 realization with 50,000 counts per scan. . . . .	95
5.10	Profile through the sample standard deviation images of different methods from 300 realization with 50,000 counts per scan. . . . .	95
5.11	Histogram of the bias of different methods compared to PR method with 50,000 counts per scan. . . . .	96
5.12	Histogram of the ratio of standard deviation of different methods to the standard deviation of PR method with 50,000 counts per scan. . . . .	96
5.13	Reconstructed emission image (SD method) from 500,000 counts. . . . .	97
5.14	Sample mean images of different methods from 300 realization with 500,000 counts per scan. . . . .	98
5.15	Sample standard deviation images of different methods from 300 realization with 500,000 counts per scan. . . . .	99
5.16	Profile through the sample mean images of different methods from 300 realization with 500,000 counts per scan. . . . .	100
5.17	Profile through the sample standard deviation images of different methods from 300 realization with 500,000 counts per scan. . . . .	100
5.18	Histogram of the bias of different methods compared to PR method with 500,000 counts per scan. . . . .	101
5.19	Histogram of the ratio of standard deviation of different methods to the standard deviation of PR method with 500,000 counts per scan. . . . .	101
5.20	Reconstructed emission image (SD method) from 5,000,000 counts per scan. . . . .	102
5.21	Sample mean images of different methods from 100 realization with 5,000,000 counts per scan. . . . .	102
5.22	Sample standard deviation images of different methods from 100 realization with 5,000,000 counts per scan. . . . .	103
5.23	Profile through the sample mean images of different methods from 100 realization with 5,000,000 counts per scan. . . . .	104
5.24	Profile through the sample standard deviation images of different methods from 100 realization with 5,000,000 counts per scan. . . . .	104
5.25	Histogram of the bias of different methods compared to PR method with 5,000,000 counts per scan. . . . .	105
5.26	Histogram of the ratio of standard deviation of different methods to the standard deviation of PR method with 5,000,000 counts per scan. . . . .	105
5.27	The normalized uniform CR bound. . . . .	107
5.28	Bias versus standard deviation comparison of different estimators together with standard error bars. For almost all the cases the error bars are smaller than plotting symbols. The OP and SP models are observed to be positively biased especially for low count rates. . . . .	109

5.29	Performance of different estimators at different count levels compared to normalized uniform CR bound (with standard error bar curves). The plots for the OP method also include standard error bars. The error bars are not included for the other methods since they are very similar to error bars on the OP method. For all the count levels the OP method is observed to be further away from the uniform CR bound. . . . .	110
E.1	Plots of $f_d(x)$ and $g_d(x)$ and their sum for $x > 1$ . . . . .	132
E.2	Plots of $f_d(x)$ and $g_d(x)$ and their sum for $x < -1$ . . . . .	135

## LIST OF APPENDICES

### APPENDIX

A	Update Orders for Sequential Iterative Algorithms . . . . .	117
B	Taylor's Series Approximation of SP model . . . . .	122
C	Bias and Variance Analysis . . . . .	124
D	Evaluation of the Conditional Expectation . . . . .	129
E	Concavity Analysis of the SD Model . . . . .	130

# CHAPTER 1

## Introduction

### 1.1 Background and Motivation

Although strictly speaking, medical imaging began with Röntgen's discovery of X-rays in 1895, contemporary medical imaging began in 1970's with the advent of X-ray computerized tomography. In computerized tomography (CT) two dimensional (2-D) or three dimensional (3-D) images of the object can be reconstructed using the line integral measurements through the object. Although there are numerous non-medical applications of CT such as non-destructive testing, underground cross-borehole imaging, electron microscopy etc., CT is mostly used in medical imaging. For example in X-ray CT, X-ray projection measurements are collected at different positions around the patient to reconstruct anatomical images of the X-ray attenuation maps of the patient.

Emission CT provides physiologic functional images as opposed to the anatomical images provided by X-ray CT. Such functional images can be acquired by imaging the decay of radio-isotopes bound to molecules with known biological properties. In emission tomography radio-pharmaceuticals are administered to the patient either by injection, ingestion or inhalation to tag a specific biochemical function. The nuclide may emit single photons in the form of  $\gamma$ -rays or it may emit positrons (which then annihilate to produce two 511 keV photons). As long as the photons have enough energy to escape from the body in sufficient numbers, images of *in vivo* distribution of the pharmaceutical distribution can be generated. Two fundamental modalities of emission CT are: single photon emission computed tomography (SPECT) and positron emission tomography (PET). In SPECT imaging the radio-pharmaceutical radiates a single photon and these photons are detected by collimated detectors to perform tomographic image reconstructions [85].

In a PET study, a radio-pharmaceutical (which is a chemical compound tagged with positron emitting isotope) is administered to the patient. The scan is started after a delay to allow for the transport and uptake by the organ(s) of interest. When the radio-isotope decays, a positron is emitted that combines with a nearby electron generating two photons. These two annihilation photons each with an energy of 511 keV ( $= m_e c^2$ ) are generated traveling nearly in opposite directions. The concurrent generation of the two  $\gamma$ -ray photons and their travel in nearly opposite directions (with a velocity close to the speed of light) makes it possible to identify the annihilation event within a short coincidence time window (around 10ns) through two detectors on opposite sides. Thus, when two photons are detected by two detectors within the coincidence timing window, it is known that a positron-electron annihilation took place along the line, strip or tube joining the two detec-

tors. The total number of “coincidence events” detected by a pair of detectors constitutes a measure of integrated radioactivity (line integral measurement) along the line joining the two detectors. From a complete set of line-integral measurements obtained at different views around the patient, the activity distribution within the slice or volume can be reconstructed using tomographic image reconstruction algorithms.

However, not all of the emitted photon pairs heading towards the detectors are detected due to their interactions with the patient body in the form of absorption or scattering. The dominant form of interaction of photons at 511 keV with human body is Compton scattering. In Compton scattering the photon interacts with a bound electron such that it is deflected from its path and loses some of its energy. Most scattered photons are scattered out of the field of view or absorbed, and never detected. The effect of these interactions is called “attenuation”. To correct for the effects of attenuation, most PET centers perform transmission scans to measure the unique attenuation characteristics of each patient. Transmission scans are usually performed using ring or rod sources around the patient that contain positron emitting radioactive materials. A good way to compute attenuation correction factors in PET is to perform reconstruction of attenuation maps using transmission scan data and then to “re-project” these attenuation maps. Also, SPECT systems with external transmission sources are becoming increasingly available where reconstructed attenuation maps can be used for quantitative SPECT [65].

From a mathematical standpoint, the solution to the problem of reconstructing a function from its projections dates back to Radon’s original work in 1917 [76]. A practical implementation for tomographic reconstruction called filtered back projection (FBP) [58] is routinely used for X-ray CT. FBP can also be used to reconstruct emission images and attenuation maps in PET. However, derivation of FBP is based on noise free ideal projection measurements. Whereas, in PET the measurements are usually highly noisy due to count limited nature of the PET process (since the radio-isotope dose injected to the patient can not exceed a certain level because of patient safety). Moreover, PET detector systems have certain count rate limitations, and long scans are usually not practical due to patient motion — especially in applications such as whole body and dynamic PET. The direct application of FBP method with ramp filter to PET emission and transmission data results in unacceptably noisy images. Windowing or reducing the cut-off frequency of the ramp filter used in the FBP method reduces the amount of noise but results in loss of resolution [23, 50, 80]. Non-stationary sinogram processing [30, 63] and image post-processing methods [73] have shown some promise to improve image quality.

In the absence of the effects of random coincidences, PET transmission and emission measurements are well modeled as Poisson random variables [102]. Statistical image reconstruction (SIR) methods such as maximum likelihood (ML) estimation or penalized maximum likelihood (PML) estimation (which can also be viewed as maximum a posteriori probability (MAP) estimation with Markov prior) methods take the appropriate measurement statistics into account. SIR methods have been shown to result in improved image quality for PET and SPECT. Unfortunately, there is no closed form solution for the ML or the PML estimators neither for the emission nor for the transmission reconstruction. Hence, iterative algorithms are used which require excessive computation. However, recent advancements in fast algorithms enable the increasing use of SIR methods in PET centers and some commercial scanners are now equipped with SIR methods.

Systems such as PET [64] and electronically collimated SPECT [15, 17] record events based on timing coincidences. These systems record accidental coincidence (AC) events simultaneously with the true coincidence events. Similarly in low light-level imaging, thermo-

electrons generated by photodetector are indistinguishable from photoelectrons generated by photo-conversion and they have a similar effect as the AC events [74].

During PET emission scans, accidental coincidence (AC) events occur when photons that originate from separate positron-electron annihilations are mistakenly recorded as having arisen from the same annihilation. Generally a significant portion of the collected data is AC events: typically in 2-D PET with septa, 5% to 30% of the detected events are accidental coincidences (even more AC events occur in some exams such as brain studies with O-15) and it is a primary source of background noise in PET [53, 74, 86]. In septaless 3-D PET, even higher AC rates are common. Moreover, AC rates increase as the *square* of the amount of radio-isotope injected to the patient, while true coincidences increase only *linearly* with the radio-isotope concentration. This count rate limitation, along with detector deadtime, determines the upper limit on the injected radio-isotope dose for many PET studies. In PET transmission scans, photons that originate from different transmission sources (rod or ring sources around the patient) cause AC events. The ratio of total AC events to “true” coincidence events are usually small in transmission scans compared to emission scans. However, the effect of AC events becomes severe for regions of high attenuation coefficients such as thorax or abdomen, because projections through such regions result in low true coincidence rates. These low count rates can easily become comparable to AC rates.

Thus AC events are a primary source of background noise in PET and should be compensated appropriately both for the emission and transmission scans. One can use the “singles” method [9] for estimating AC events, however this approach is not widely used because of the necessity for additional hardware and moreover the singles rate often varies during data acquisition [72]. Although there are other suggested techniques (see Section 2.2.3), because of hardware, software and data storage limitations (and historical momentum) most PET centers collect and archive only randoms pre-corrected data. We recommend separate acquisition and storage of delayed coincidences whenever feasible. However, in practice most PET center archive and use only randoms pre-corrected data. Even most of the latest commercial PET scanners do not have the option of using prompt and delayed coincidence data separately in their reconstruction algorithms.

In randoms pre-correction the AC rates are estimated by delayed-window coincidences and data are pre-corrected for AC events by *real-time* subtraction<sup>1</sup> (see Section 3.1). This method also has the potential to be applied to electronically collimated SPECT. Real time subtraction of delayed coincidences compensates for the average of AC events, but *destroys* the Poisson statistics [53].

Moreover, negative values can result during the real-time subtraction of delayed coincidences (especially in 3-D PET where counts per each coincidence detector pair can be very low). These negative values would cause conventional penalized maximum likelihood algorithms to diverge in emission reconstructions. Setting the negative values to zero alleviates this problem but introduces a systematic positive bias in the resulting emission images [74, 96].

Since the introduction of the ML-EM [26, 61, 81] algorithm for PET, statistical image reconstruction methods have been based on idealized PET system with Poisson statistical model, and ignored the effects of AC events. Although, randoms pre-correction method clearly violates the Poisson statistics of the measurements, this problem has been largely ignored in the PET SIR literature. Numerous papers have been published simply ignoring

---

<sup>1</sup>After the real time pre-correction, one usually does not have access to the delayed coincidences separately, but usually has access to the total delayed coincidences for the whole scan.

the AC events and the randoms pre-correction. In most of the commercial PET scanners (with and without SIR image reconstruction tools) image reconstruction is done using randoms pre-corrected data. The purpose of this thesis is to provide accurate statistical models and image reconstruction techniques for PET measurements with pre-subtracted delayed coincidences.

## 1.2 Organization of the Thesis

The organization of the thesis is as follows. In Chapter 2, we describe the PET imaging system and physics. Different system components and models are explained. Then, image reconstruction techniques for emission and transmission scans are described.

In Chapter 3 we describe the statistical model for the randoms pre-corrected PET data and demonstrate how the randoms pre-correction renders data in a non-Poisson way. We present the probability distribution functions corresponding to the different approximation methods and compare their central moments with the randoms pre-corrected data. Then we derive the “exact” log-likelihood of the randoms pre-corrected data for the maximum likelihood (ML) and penalized maximum likelihood (PML) estimation. Since the exact log-likelihood is complicated, we develop and compare several approximations to the exact log-likelihood. For completeness, we first review the data-weighted least squares (WLS) method and the log-likelihood for the ordinary Poisson (OP) model. Then, we introduce a new “shifted” Poisson (SP) model [97]. This SP model is based on the idea of matching both the first and second-order moments of the model to the underlying statistics of the pre-corrected data. Although both the WLS and SP models match two moments to the underlying statistics, in the data WLS model the second moment is fixed independent of the unknown parameters to be estimated (*i.e.*: the image), whereas in SP model the moments vary with the measurement model appropriately. This difference is shown to be a very important difference between the two models and the corresponding log-likelihood function of the SP model is shown to have better agreement with the exact log-likelihood than the WLS and ordinary Poisson (OP) models. Then, we introduce a new saddle-point (SD) approximation [98, 99] for the probability mass function (pmf) of the pre-corrected measurements. This SD model is based on the idea of making a second order Taylor series approximation to the exact pmf in the  $z$ -transform domain (*i.e.*: on the probability generating function) and then carrying out the inverse transform. The corresponding log-likelihood function to the SD model is shown to have the best agreement with the exact log-likelihood (compared to all of the previous approximations), and its performance is shown to be almost identical to the exact log-likelihood method.

In Chapter 4, we first develop maximization algorithms for the SP and the SD methods, and then present representative performance results from computer simulations and experimental transmission scans [100]. The results show that the WLS method leads to systematic negative bias in the reconstructed attenuation maps and the OP method results in higher variance than the proposed SP and SD methods. We also investigate the propagation of noise from the reconstructed attenuation maps into emission images reconstructed using the FBP method. Interestingly, the difference in variances in the emission images with the new methods are shown to be even greater than in the attenuation maps. To corroborate the *empirical* studies, we develop *analytical* approximations to the reconstructed image covariance based on the techniques developed in [34]. Using these analytic approximations we show that the OP method yields more noisy images compared to the proposed methods

and we use the approximations to explain the negative bias of the WLS method. We also develop analytical approximations for the propagation of noise from attenuation maps into reconstructed emission images. The results of the analytic approximations are shown to be in good agreement with the experimental results.

In Chapter 5 we concentrate on the emission problem. We analyze the concavity of the proposed objective functions and develop appropriate maximization algorithms to be used in the image reconstructions with the proposed methods. We show that the proposed approximations result in reconstructions that are free of systematic bias and lead to images with less noise compared to the ordinary Poisson (OP) model. Although the SP model is slightly biased for emission scans with very low count rates, the SD model is free of any systematic bias and performs almost identically to the exact log-likelihood. Lastly, we study the bias-variance trade-offs of the models by analyzing how close they perform to the uniform Cramer-Rao bounds and show that the proposed SP and SD models perform very close to the uniform bounds as opposed to the ordinary Poisson model.

Finally, Chapter 6 contains conclusions and suggested directions for future research.

### 1.3 Original Contributions

The original contributions of this research are summarized in the following.

- The statistical model and the corresponding log-likelihood function is derived for randoms pre-corrected PET measurements.
- A novel approximation to the complicated exact log-likelihood of pre-corrected PET measurements is introduced [97] based on a “shifted Poisson” (SP) model. The model is shown to offer improved image quality compared to the conventional WLS and ordinary Poisson (OP) methods through more realistic statistical modeling. The method is very practical and easy to implement, and requires only negligible increase in computation.
- Another original approximation is introduced [98,99] based on a saddle-point (SD) approximation to the probability mass function (pmf) of the pre-corrected PET measurements. The corresponding log-likelihood function has the best agreement with the exact log-likelihood and it performs almost identically as the exact log-likelihood.
- Concavity of the proposed SP and SD methods are analyzed, and maximization algorithms are developed for PET transmission and emission imaging.
- Through experimental and simulation PET studies [100] and analytic approximations, it is demonstrated that the new methods offer improved image quality both in emission and transmission scans.
- Analytic approximations are developed [100] for the covariance of the reconstructed transmission images and for the propagation of noise from attenuation maps into reconstructed emission images. The results of the analytic approximations are shown to be in good agreement with experimental results and support the image quality improvements of the proposed methods.
- Bias-variance trade-offs of the models are investigated by analyzing how close they perform to the uniform Cramer-Rao bounds.

## CHAPTER 2

### Positron Emission Tomography (PET) Imaging

In positron emission tomography (PET) imaging, 2-D or 3-D tomographic images of radioactivity distribution within the patient are generated. During emission scans, coincidence events are detected with detector pairs around the patient which indicate a positron-electron annihilation took place along the line joining the two detectors. The total number of coincidence events detected by a pair of detectors constitutes a measure of integrated radioactivity (line integral measurement) along the line or strip (or tube in 3-D) joining the two detectors. From a complete set of line integral measurements obtained from detectors at different views surrounding the patient, the activity distribution within the slice or volume can be reconstructed using tomographic image reconstruction algorithms. Also in PET transmission scans, after proper operations performed on the data (normalization, log, etc.), the problem reduces to tomographic image reconstruction from line integral measurements.

We will first review the mathematics of tomographic reconstruction which is based on an idealized, continuous and noise-free data. Then we will review the PET system physics and measurement models. Lastly, we will describe statistical image reconstruction (SIR) methods that accurately take into account the statistics of the noisy measurements in PET.

#### 2.1 Tomographic Image Reconstruction

In this section we will describe the tomographic reconstruction mathematics from idealized, continuous and noise-free line integral measurements. Assume the object to be reconstructed is represented by a 2-D function  $f(x, y)$  as shown in Figure 2.1, where each line integrated through  $f(x, y)$  is parameterized with  $\theta$  and  $R$  as :

$$p_{\theta}(R) = \int_{s \in \text{line}} f(x, y) ds \quad (2.1)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \delta(x \cos \theta + y \sin \theta - R) dx dy . \quad (2.2)$$

This function  $p_{\theta}(R)$  is called as the Radon transform of the function  $f(x, y)$ .

The 1-D Fourier transform of the function  $p_{\theta}(R)$  as a function of  $R$  is given by:

$$S_{\theta}(\nu) = \int_{-\infty}^{\infty} p_{\theta}(R) e^{-j2\pi\nu R} dR \quad (2.3)$$

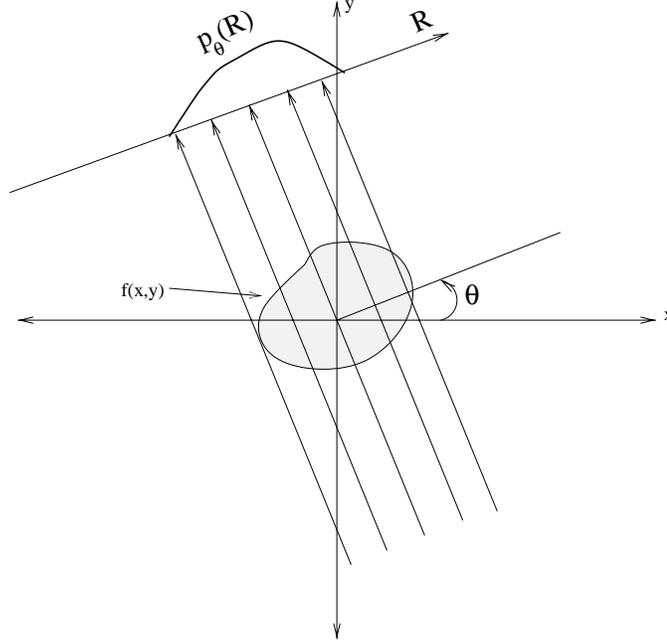


Figure 2.1: Object  $f(x, y)$  and its projection  $p_\theta(R)$  at angle  $\theta$

$$= \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \delta(x \cos \theta + y \sin \theta - R) dx dy \right] e^{-j2\pi\nu R} dR \quad (2.4)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-j2\pi\nu(x \cos \theta + y \sin \theta)} dx dy . \quad (2.5)$$

This expression is recognized to be equal to  $F(\nu \cos \theta, \nu \sin \theta)$  where  $F(u, v)$  is the 2-D Fourier transform of the original 2-D function  $f(x, y)$ :

$$F(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-j2\pi\nu(ux+vy)} dx dy . \quad (2.6)$$

This results is called the ‘‘Fourier Slice Theorem’’ [58], which can be stated as: the 1-D Fourier transform of a projection of an image  $f(x, y)$  at an angle  $\theta$  gives the values of the 2-D Fourier transform of the image along a line at an angle  $\theta$  in the 2-D Fourier transform domain.

Thus, if we collect projections of the image at all angles and then perform 1-D Fourier transform of each projection then we can construct the 2-D Fourier transform of the original image. And then the image can be reconstructed by simply taking the 2-D inverse Fourier transform. However in practical implementation 2-D fast Fourier transform (FFT) algorithms require the data to be on a rectangular grid while the projection Fourier transform gives data only along radial lines. Thus, one usually needs to perform some kind of interpolation from radial lines to rectangular 2-D grid, which usually results in some degradation especially at high frequencies.

### 2.1.1 Filtered Backprojection

A more practical and popular method for tomographic reconstruction is the filtered backprojection method (FBP) method which can be simply derived by re-writing the previous equations in a different form. We can write the 2-D inverse Fourier transform of the image  $f(x, y)$  in polar coordinates as:

$$f(x, y) = \int_0^{2\pi} \int_{-\infty}^{\infty} F(\nu, \theta) e^{j2\pi\nu(x \cos \theta + y \sin \theta)} \nu d\nu d\theta \quad (2.7)$$

$$= \int_0^{\pi} \int_{-\infty}^{\infty} F(\nu, \theta) |\nu| e^{j2\pi\nu R} d\nu d\theta, \quad (2.8)$$

with  $R = x \cos \theta + y \sin \theta$ .

Then using the Fourier Slice Theorem we substitute the 1-D Fourier transform of the projection at angle  $\theta$  *i.e.*:  $S_{\theta}(\nu)$  for  $F(\nu, \theta)$  as:

$$f(x, y) = \int_0^{\pi} \int_{-\infty}^{\infty} S_{\theta}(\nu) |\nu| e^{j2\pi\nu R} d\nu d\theta \quad (2.9)$$

$$= \int_0^{\pi} T_{\theta}(x \cos \theta + y \sin \theta) d\theta \quad (2.10)$$

with

$$T_{\theta}(R) = \int_{-\infty}^{\infty} S_{\theta}(\nu) |\nu| e^{j2\pi\nu R} d\nu. \quad (2.11)$$

Equations (2.10, 2.11) show that the image  $f(x, y)$  can be found by first *filtering* the projections with a ramp filter  $|\nu|$ , and then integrating these *filtered projection* values at the coordinate  $(x \cos \theta + y \sin \theta)$  over all projection angles  $\theta$ . In practice since there is only a finite number of projection angles, one uses summations to approximate the integration<sup>1</sup> and this operation is called the “backprojection”. This method of image reconstruction is called filter backprojection (FBP) [50, 58, 70, 80]. For 3-D tomographic reconstruction, modifications of FBP method [29] or similar methods can be used [24, 25].

As can be seen from its derivation, FBP is a deterministic method that ignores the noise in the measurements. Its wide use in clinics is due to historical reasons of computational simplicity despite its suboptimal image quality with noisy data such as in PET. In PET emission scans (especially for the ones with low count rates) FBP method can result in excessive noise amplification, streak artifacts and negative reconstruction values (which is not physically possible). Smoothing or reducing the cut-off frequency of the ramp filter  $|w|$  can reduce the amount of noise in the reconstructed images, but results in loss of resolution. Next we will review the PET system physics and data measurement models, and we will describe statistical image reconstruction (SIR) algorithms that accurately take into account the statistics of the noisy measurements in PET.

---

<sup>1</sup>In practice one also needs to perform some interpolation to compute  $T_{\theta}(x \cos \theta + y \sin \theta)$  for particular  $(x, y)$  values from discrete values of projection  $T_{\theta}(R)$ .

## 2.2 PET Physics and System Description

Although other PET geometries exist, such as hexagonal systems, typical PET systems consist of a cylindrical ring of detectors as shown in Figure 2.2 [72].

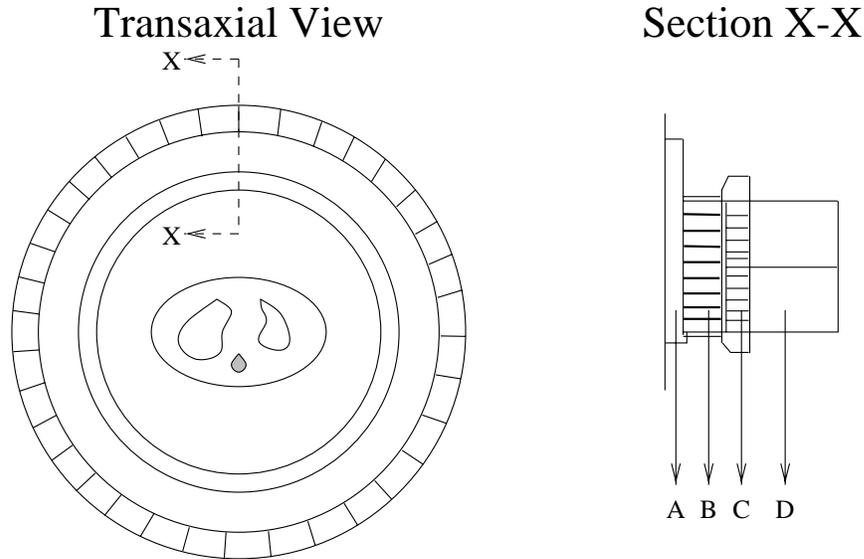


Figure 2.2: Transaxial view and a cross-section view (Section X-X) of a cylindrical PET scanner. A) Rod sources for transmission scan, B) Collimators for scatter rejection (septa), C) Detector crystals, D) Photomultiplier tubes.

The detectors have lead collimators (septa) to shield detectors from any radiation from upper or lower slices out of field of view<sup>2</sup>. Many PET scanners can either be operated in this *slice-collimated* mode or in *fully 3-D* mode<sup>3</sup>.

### 2.2.1 PET Imaging

An advantage of PET over other nuclear medical imaging systems is the availability of many positron-emitting radionuclides with low atomic number [13] that can be used as biologically relevant substances for human body. Frequently used radionuclides in PET are C-11, N-13, O-15 and F-18. These radionuclides have strong physiological relation to human body since C, N, O are the major components of organic molecules<sup>4</sup>. These radionuclides have short half-lives : 20.3 min. for C-11, 9.9 min. for N-13, 2.0 min. for O-15 and 110 min. for F-18. Many PET centers have on-site cyclotrons where these molecules are produced and rapidly coupled to biomolecules.

In PET, the aim is to determine the concentration and location of positron emitting radio-label in the desired cross section of the human body. When the radio-isotope decays, it emits a positron which annihilates with an electron after traveling a short distance of a

<sup>2</sup>Some septa configurations let the collection of coincidence events between neighboring upper or lower slices as well.

<sup>3</sup>In fully 3-D mode, septa is removed and coincidence events can be collected by all possible detector pairs in 3-D.

<sup>4</sup>Although F is not a major organic component, F-18 is used in FDG studies where FDG is an analog to true glucose that partially follows part of the metabolism pathway of glucose.

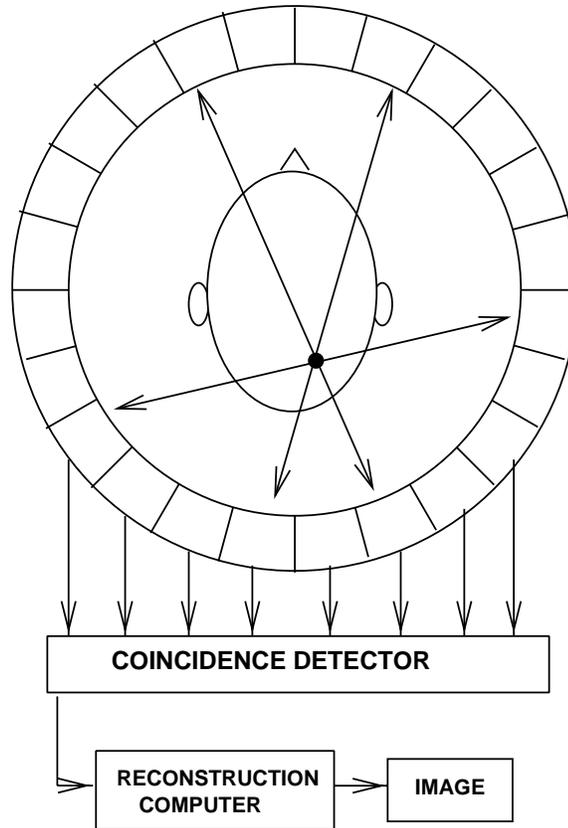


Figure 2.3: Diagram of a PET detector system with coincidence detection between different detectors.

few mm.'s. This annihilation of masses creates two  $\gamma$ -ray photons of 511keV ( $m_e c^2$ ) which propagate nearly  $180^\circ$  from one another as shown in Figure 2.3. If two photons are detected within a coincidence timing window (which is in the order of 10ns), then it is decided that an annihilation event (true event) has occurred along the line joining the detectors. Summing such events results in quantities that approximate line integrals through the radio-isotope distribution (or more truly positron annihilation distribution).

The accuracy of the spatial location of radio-isotope decay detected by detector pairs is limited by two physical properties: first the angular uncertainty in the direction of emitted photons (since they do not travel exactly at  $180^\circ$  from one another) and secondly the short distance traveled by the emitted positron before annihilation with an electron. This distance is usually a couple millimeters depending on the kinetic energy of the emitted positron for a given radionuclide [13].

### 2.2.2 Attenuation

The emitted photons either interact with the body tissue or pass through unaffected. The interaction of photons with the body occur in the form of photoelectric absorption (shown as A in Fig. 2.4) or Compton scattering (shown as B and C in Fig. 2.4). The dominant form of interaction for photons at 511 keV is Compton scattering. Compton scattering is caused by the collision between  $\gamma$ -ray photon and a loosely bound electron

in an outer shell [64]. When the photon interacts with an electron, its path is deflected and it loses some energy. Most of the scattered photons are scattered through an oblique angle from the plane of detectors resulting undetected<sup>5</sup> and this reduction in the number of photons (that would otherwise reach the detectors) is called “attenuation”.

## Attenuation in PET

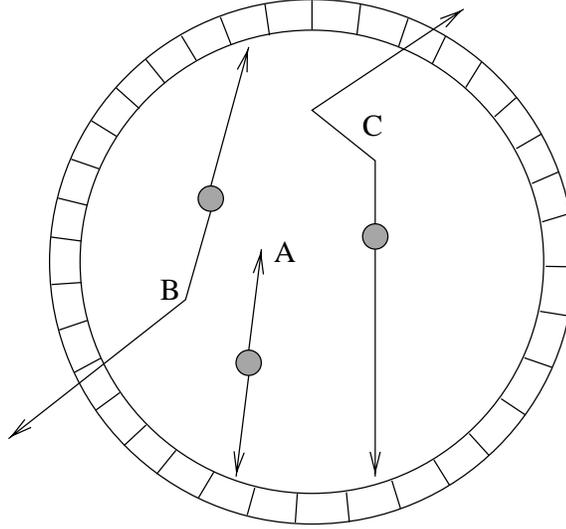


Figure 2.4: Different forms of attenuation in PET: A) Photoelectric absorption, B) Single scattering, C) Multiple scattering.

Assume that there has been an annihilation event at the point S in Figure 2.5 and two  $\gamma$ -ray photons  $\gamma_1$  and  $\gamma_2$  are released towards detectors  $D_1$  and  $D_2$  respectively. Because of attenuation according to Beer’s law, the probability of  $\gamma_1$  reaching  $D_1$  and  $\gamma_2$  reaching  $D_2$  are given by  $P_1$  and  $P_2$  respectively as follows:

$$P_1 = e^{-\int_{L_1}^{L_2} \mu(x) dx} , \quad P_2 = e^{-\int_{L_2}^L \mu(x) dx} ,$$

where  $\mu(x)$  is the linear attenuation coefficient at 511 keV as a function of position along the line joining two detectors. The probability that this particular annihilation event will be recorded, i.e. both photons will be detected is called “survival probability” [58] and it is given by:

$$\begin{aligned} P_s &= P_1 P_2 \\ &= e^{-\int_{L_1}^{L_2} \mu(x) dx} e^{-\int_{L_2}^L \mu(x) dx} \\ &= e^{-\int_{L_1}^L \mu(x) dx} . \end{aligned}$$

This result shows that survival probability is same independent of the position of the annihilation along the line joining  $D_1$  and  $D_2$  and it is equal to the attenuation that a photon beam of 511 keV would experience while propagating from  $L_1$  to  $L_2$ . Therefore in PET it is

<sup>5</sup>In septaless 3-D PET, considerable portion of scattered photons can be detected by detectors at upper or lower slices.

possible to correct for attenuation by the use of transmission scans with external radio-active sources which will be described later.

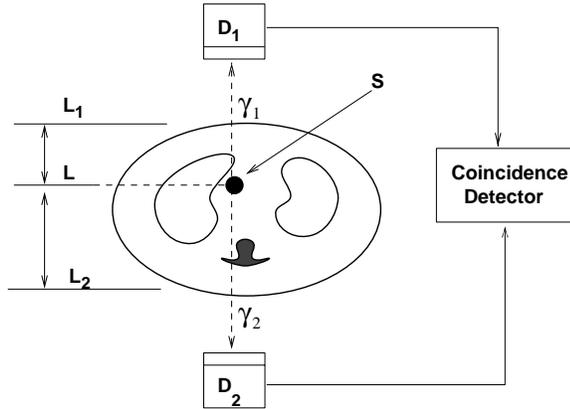


Figure 2.5: Photons  $\gamma_1$  and  $\gamma_2$  are attenuated through distances  $L_1 - L$  and  $L - L_2$  respectively.

### 2.2.3 Accidental Coincidence Events

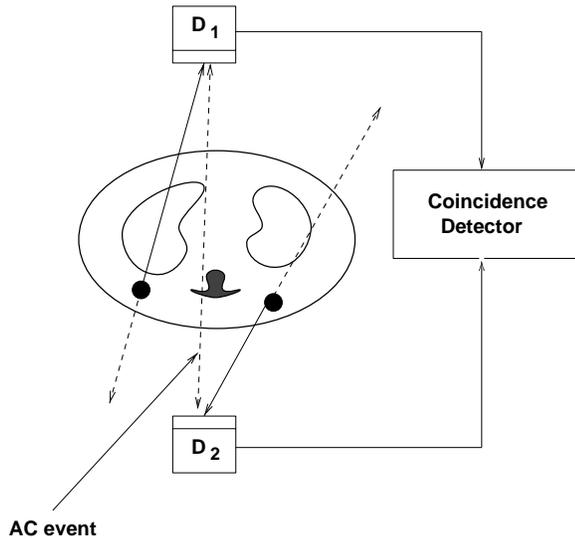


Figure 2.6: Diagram of an Accidental Coincidence event

Accidental coincidence (AC) events (or random coincidences) occur when photons that arise from separate annihilations are mistakenly recorded as having arisen from the same annihilation as shown in Figure 2.6. Most of the scattered photons leave the detector plane undetected and even if a photon makes it to the detector the detection at the scintillation crystal occurs with a probability less than one. Thus for many of the annihilation events only one of the photons will be detected. These events are called “singles”. AC events occur when two singles event from separate annihilations are recorded by two detectors

in the same coincidence window, so that they are mistakenly recorded as true coincidence events. In PET measurements, AC events are a primary source of background noise and usually 5% to 50% of the detected events are AC events.

Let  $R_i$  and  $R_j$  be the singles rate at detector  $i$  and  $j$  respectively and let  $\tau$  be the duration of the coincidence timing window. For each single event at detector  $i$ , on the average  $\tau R_j$  events will be detected at detector  $j$ . Thus the rate of AC events between detector  $i$  and  $j$  with first photon detected at detector  $i$  is  $\tau R_i R_j$ . With the similar argument for AC events with first photon detected at detector  $j$ , the rate of AC events between detector  $i$  and  $j$  is given by:

$$R_{AC} = 2\tau R_i R_j . \quad (2.12)$$

Singles rate are proportional to the amount of radio isotope injected to the patient. Thus in PET AC rate is proportional to the *square* of the amount of isotope in the field of view, while true coincidences are only *linearly* proportional to the amount of radio-isotope. This count rate limitation, along with detector deadtime, determines the upper limit on the injected radio isotope dose for many PET studies.

If one has access to singles measurements of each detector then (2.12) can be used to estimate the mean of AC events [9,69]. However this approach (singles method) is not widely used because of the necessity for additional hardware<sup>6</sup>. Mumcuoglu *et. al.* have developed a Bayesian estimation method to compute the mean of randoms with a method that requires the knowledge of the intrinsic detector efficiencies for the AC events [68]. However, working with experimental data they found that the intrinsic detector efficiencies for AC events are different than those for true coincidences, which limited the practical utility of their method [69]. Moreover, the singles rate will vary during data acquisition [72], and this is not modeled by (2.12).

The arrival time of photons due to AC events are uniformly distributed *locally*<sup>7</sup> in time while those of true coincidences fall within the timing window. Thus a simple method that is used in practice is to collect data in a second coincidence timing window that is offset in time (beyond resolution of the true coincidence timing window) such that it collects no true coincidences. This method yields data with approximately the same mean as AC events in true coincidence window, since the singles events are distributed uniform locally in time. This method has the advantage over the singles method that one does not need to consider the difference in detector efficiencies between singles events and true coincidence events. Moreover, this method is simple to implement and it can be performed in hardware before the data is stored. And lastly, the method appropriately takes into account the temporal variations in the AC events during data acquisition. Therefore, in most PET scans, the AC rates are estimated using delayed-window coincidences and the data are precorrected for AC events by real-time subtraction. Real-time subtraction of delayed window coincidences compensates in mean for AC events but destroys the Poisson statistics [53]. To avoid this problem, one needs to maintain the coincidence and randoms measurements as two separate sinograms [74]. If one could collect separate sinograms for the randoms measurements, then one could consider jointly estimating the AC means and the PET image (emission or

---

<sup>6</sup>In most PET scanner singles rate is available for each detector block, but the individual singles rate for each detector element is not.

<sup>7</sup>While AC rates can be well modeled as uniformly distributed *locally* in time, they vary during the whole scan duration due to radio-active decay. Thus real-time subtraction of delayed coincidences compensates for the AC events appropriately. However, other methods such as the “singles” method does not compensate for the low frequency variation of AC events during the scan duration.

transmission) from the two separate sinograms. Or, one could exploit spatial smoothness of the AC events to estimate the AC means from the delayed coincidences and then use these estimates in the ML estimation [74]. However even if a PET system allows one to collect randoms (delayed coincidences) sinogram separately, this process would double the storage space for the acquired data. In practice because of software, hardware and data storage limitations (and historical momentum), most PET centers collect and archive only the randoms precorrected data. Even most of the latest commercial PET scanners do not have the option to use randoms separately in their image reconstruction algorithms but use randoms pre-corrected data instead.

### 2.2.4 Scattered Events

Although most of the scattered photons leave the detector ring undetected, some of them are still detected in coincidence with their photon pairs. These events are mispositioned because photon paths are not collinear.

Scattered photons lose some of their energy through Compton interaction, thus “energy discrimination” can be used at the detector to reject a significant portion of the scattered photons. There has been work in terms of estimating and rejecting the scattered events [57, 69, 71]. In our models we assume that the mean of scattered events is known.

## 2.3 System and Measurement Model

The photons generated by radioactivity decay follow a Poisson process. In PET the generated photons are detected with a probability less than one due to attenuation, detector efficiency etc., all of which can be well modeled as Bernoulli process. Moreover, a Poisson process which is “thinned” by independent Bernoulli trials remains Poisson [64]. Since PET measurements are based on a counting process (both emission and transmission scans) and since the emitted photon pairs are uniformly distributed in all directions in 3-D, a reasonable model for the collected measurements is independent Poisson probability distribution function (pdf). Although in practice the measurement distributions are not exactly Poisson because of detector deadtime, they can still be very well modeled as Poisson [102]. However, if data is processed with randoms subtraction (for AC correction) then the measurements are no longer Poisson distributed as explained in detail in Chapter 3. In this section we will describe the system model and the models for the *mean* of the collected measurements for emission and transmission scans.

### 2.3.1 Emission Scan

As mentioned earlier, the aim in PET is to obtain an estimate of the spatial distribution of the radio-isotope  $\lambda(x)$  inside the body. Since there are finite number of detectors, usually  $\lambda(x)$  is represented by a finite parameterization:

$$\lambda(x) = \sum_{j=1}^P \lambda_j b_j(x) \tag{2.13}$$

where  $\lambda_j$  is the unknown mean radioactivity in the  $j$ th voxel and  $b_j(x)$  is the indicator function with the  $j$ th voxel as its support.

The mean of the  $n$ th detector pair measurement can be written as:

$$\bar{y}_n(\lambda) = \sum_{j=1}^P g_{nj} \lambda_j + s_n^E \quad (2.14)$$

where  $s_n^E$  denotes the mean of the scattered events detected by  $n$ th detector pair and  $g_{nj} = c_n a_{nj}$  denotes the contribution of the annihilations in the  $j$ th voxel to the  $n$ th detector pair measurements with  $c_n$  denoting detector dependent factors such as calibration factors for scan time, attenuation factors, detector efficiencies and deadtime correction factors, and  $a_{nj}$  is the geometric contribution of the  $j$ th voxel to the  $n$ th detector pair<sup>8</sup> which can be computed as shown in Figure 2.7.

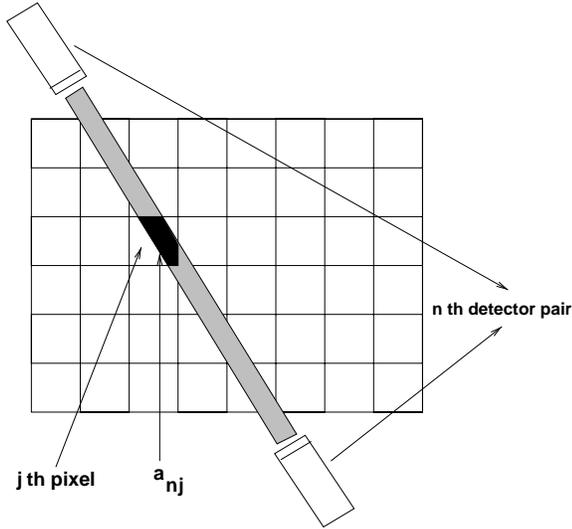


Figure 2.7: Geometric system model showing the contribution of  $j$ th pixel to the  $n$ th detector pair.

The classical method used to reconstruct  $\lambda_j$  parameters from the projection measurements is filtered backprojection (FBP) method. Although FBP is computationally simple, it is derived without any statistical information and results in noisy images compared to statistical image reconstruction methods. For PET many authors have proposed algorithms based on the assumption that the measurements have Poisson distribution with the given mean (2.14). However as described in Chapter 3, real time correction for random coincidences renders the data non-Poisson. Thus, Poisson distribution idealization ignores the AC events precorrection, and we will refer to this approach as “ordinary” Poisson (OP) model.

As mentioned earlier, to get accurate images using PET, one needs to correct for the effects of attenuation. For this purpose, a transmission scan is performed prior to emission scan which will be described next.

---

<sup>8</sup>In transmission scans for notational simplicity we use the same notation  $a_{nj}$  (2.15) for the geometric factor that represents the contribution of the attenuation factor of  $j$ th pixel to the total attenuation in  $n$ th detector measurement. And we assume proportionality constants between emission (2.14) and transmission (2.15) scans can be included in  $c_n$  factors.

### 2.3.2 Transmission Scan

To compute accurate estimates of the radioactivity distribution within a patient using positron emission tomography (PET), the effects of *attenuation* should be taken into account in a quantitative manner. One simple method for attenuation correction is the use of geometrical shapes, namely finding the edge contours and then using the contour information to find the attenuation length of each projection or line integral, for subsequent correction of the emission data. In this method, attenuation coefficients are customarily assumed to be constant within the boundary. A more accurate method is to use transmission scan information obtained through external positron sources surrounding the patient. Most PET centers have adopted a *measured* attenuation correction method where one precedes the emission scan with a transmission scan that measures the unique attenuation characteristics of each patient over the slice of interest [55]. PET and SPECT transmission scans are measurements of the correction factors rather than being the primary medical interest. Thus it is desirable to minimize the durations of transmission scans. Short scans suffer from statistical noise, leading to unwanted errors in the reconstructed emission image [66, 69]. Smoothing of the transmission data before computing the attenuation correction factors leads to resolution mismatch between transmission and emission data [10, 11].

In practice, reconstruction of attenuation maps requires a finite dimensional representation of the image. We assume that the images can be adequately represented as a set of pixels, each with constant linear attenuation coefficient  $\mu_j$ . If we let  $\mu = [\mu_1, \dots, \mu_P]'$  denote the vector of unknown linear attenuation coefficients (having units of inverse length), then the total attenuation between  $n$ th detector pair is determined by the sum of the attenuation coefficients  $\mu_j$  multiplied by their respective area (or volume in 3-D) of intersection  $a_{nj}$  with the  $n$ th projection strip (as shown in Figure 2.7), *i.e.*: the total attenuation between  $n$ th detector pair is

$$l_n(\mu) = \sum_{j=1}^P a_{nj} \mu_j, \quad (2.15)$$

again, the  $a_{nj} \geq 0$  factors have units of length and describe the tomographic system geometry. Then, the mean of the  $n$ th detector pair measurements in the transmission sinogram is approximately

$$\bar{y}_n(\mu) = b_n e^{-l_n(\mu)} + s_n^T \quad (2.16)$$

where the  $b_n > 0$  factors denote the blank scan counts and  $s_n^T$  factors denote the mean of the transmission scattered events.

The conventional method for attenuation correction in PET using measured transmission scans consists of two steps: first compute attenuation correction factors from the ratio of *blank scan*<sup>9</sup> measurements to the *transmission scan* measurements; then multiply the emission measurements by attenuation correcting factors in sinogram space. Therefore often *no* attenuation map is needed. However, there are several reasons why reconstruction of an attenuation map can be important. First, correction factors based on the ratio of blank scan and transmission scan measurements provide noisy and biased estimates of the true attenuation correction factors. Reconstruction of an attenuation map, followed by re-projection, can improve the accuracy of the estimated correction factors provided additional

---

<sup>9</sup>A blank scan is a transmission scan without the patient in the scanner that is acquired for the purpose of calibrating the measurements.

information, in the form of statistical model for the data, is introduced [66, 69]. There are other advantages of reconstruction of the attenuation map. For example, if the patient moves between the transmission and emission scans, they can be re-registered before reprojection for the computation of the attenuation correction factors. In addition, the attenuation maps provide anatomical landmarks that are often not visible in the emission images [2]. Finally, 2-D attenuation maps can be used for reprojection to form 3-D attenuation correction factors in septaless PET [12, 67, 90] and also attenuation maps can be used for quantitative SPECT [65].

The conventional method for reconstructing attenuation maps from transmission measurements consists of two steps: 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 *filtered backprojection* (FBP) method. As shown by Fessler [32], FBP and *data weighted least-squares* methods lead to systematic biases at low count rates. To eliminate this bias problem, one can use statistical methods which require no logarithm [32].

In transmission scans the photons that originate from different transmission sources (ring sources [52] or rotating rod or sector sources [89] around the patient) cause AC events. The ratio of total AC events to “true” events is usually small in transmission scans compared to emission scans. Nevertheless, the effect of AC events becomes severe for regions of high attenuation coefficients, because projections through such regions result in low true coincidence rates. These low count rates can become comparable to AC rates. Similar to emission scans, the real time correction of AC events destroys the Poisson statistics of the transmission scan measurements. Thus statistical image reconstruction methods which assume pure Poisson statistics (OP models) are *suboptimal*.

## 2.4 Statistical Image Reconstruction Methods

Next we will describe statistical image reconstruction algorithms that accurately take into account the statistics of the noisy measurements in PET.

### 2.4.1 Maximum likelihood

The Poisson nature of photon emission process invites statistical signal processing techniques for image reconstruction. For statistical image reconstruction (SIR) one needs an objective function which measures how well the data fits to the parameterized model. The estimation solution is the parameter vector that maximizes the objective function. In maximum likelihood (ML) estimation, one chooses the parameter values that maximizes the probability density function (pdf) after the observed value of the data is substituted into the pdf, which is also referred as likelihood function of our data, *i.e.*: in emission case the ML estimate is:

$$\begin{aligned}\hat{\lambda} &= \arg \max_{\lambda \geq 0} p(y; \lambda) \\ &= \arg \max_{\lambda \geq 0} \log p(y; \lambda)\end{aligned}$$

where  $p(y; \lambda)$  is the probability mass function (pmf) of projection measurement  $y = [y_1 \dots y_N]$  which includes  $\lambda$  as a parameter and  $\log p(y; \lambda) = L(y; \lambda)$  is called log-likelihood.

As mentioned previously, for emission and transmission tomography the conventional method (OP) is to assume the measurements are Poisson distributed with means  $\bar{y}_n(\lambda)$

(2.14) and  $\bar{y}_n(\mu)$  (2.16) respectively. Although this method ignores the real time correction for random coincidences (which renders the data non-Poisson) many authors have proposed algorithms based on this assumption. The corresponding OP log-likelihood function for emission tomography is:

$$L_{\text{OP}}(\lambda) = \sum_{n=1}^N y_n \log \bar{y}_n(\lambda) - \bar{y}_n(\lambda). \quad (2.17)$$

This function is concave for  $y_n \geq 0$ . However, the real-time correction of the data can lead to some negative values in the precorrected data, thus one needs to zero-threshold the measurements to guarantee convergence to a unique maximum. The effects of zero thresholding in the resultant image is discussed in Section (4.8.4).

There is no closed form expression for the maximizer  $\hat{\lambda} = \arg \max_{\lambda \geq 0} L_{\text{OP}}(\lambda)$ . Moreover the large size of the system matrix  $A = \{g_{ij}\}$  makes it computationally impractical even to compute the linear least square estimate directly. Thus one needs to use iterative methods for computing  $\hat{\lambda}$ .

One simple approach is to apply coordinate-ascent directly to  $L_{\text{OP}}(\lambda)$ . Equating the partial derivatives to zero:

$$\frac{\partial}{\partial \lambda_k} L_{\text{OP}}(\lambda) = - \sum_{n=1}^N g_{nk} + \sum_{n=1}^N g_{nk} \frac{y_n}{\sum_j g_{nj} \lambda_j + s_n^E} = 0, \quad k = 1, \dots, P$$

it is seen that this equation has no analytic solution. A line search method would evaluate the above expression multiple times, which would be computationally expensive.

Since the introduction of EM algorithm for tomography [61, 81], it has been used widely for ML tomographic reconstruction. In this method, the observed measurements are supplemented with a complete (unobserved) data space. Then at each iteration one calculates the conditional expectation of the complete data space (E step) and analytically maximizes the expectation with respect to unknown parameters (M step). EM algorithm results in the following iteration for each  $\lambda_k$  ( $k = 1, \dots, P$ ):

$$\lambda_k^{i+1} = \frac{\lambda_k^i}{\sum_n g_{nk}} \sum_{n=1}^N \frac{g_{nk} y_n}{\sum_j g_{nj} \lambda_j + s_n^E}, \quad i = 1, 2, \dots \quad (2.18)$$

This EM algorithm converges globally if initialized with a non-zero image, but convergence rate is usually very slow [37, 61].

### 2.4.2 Penalized Maximum Likelihood

As image reconstruction is inherently ill conditioned, the maximizer of the log-likelihood (2.17) is excessively noisy [84]. To address this problem, several methods have been proposed: aborting the iterations before convergence [94], post-smoothing the ML image (which is a special case of method of sieves) [84], and adding a roughness penalty to the log-likelihood function (penalized ML). Penalized ML (PML) can also be viewed as a maximum a posteriori (MAP) estimate with Gauss-Markov prior [45, 47]. PML method was shown to yield better results than post-smoothing [8]. Moreover, with post-smoothing the problem of slow convergence of the EM algorithm still remains, whereas PML algorithms converge more rapidly because the penalty function improves the conditioning of the reconstruction

problem. And lastly, PML method enables one to include space-variant penalties reflecting the prior anatomical boundary information. The PML objective function can be written as:

$$\Phi(\lambda) = L_{OP}(\lambda) - \beta R(\lambda),$$

where  $\beta$  controls the level of smoothing and  $R(\lambda)$  is the roughness penalty. For reducing noise, the usual penalty which discourage neighboring pixels from having disparate values is:

$$R(\lambda) = \frac{1}{2} \sum_j \sum_{k \in \mathcal{N}_j} w_{jk} \psi(\lambda_j - \lambda_k) \quad (2.19)$$

where  $\mathcal{N}_j$  is neighborhood of pixels near pixel  $j$  and  $\psi(x)$  is a symmetric convex function and  $w_{jk} = w_{kj}$ . The *quadratic* penalty of  $\psi(x) = \frac{1}{2}x^2$  leads to oversmoothing while *non-quadratic* penalties require additional parameters to be specified. Ordinarily,  $w_{jk} = 1$  for horizontal and vertical neighboring pixels,  $w_{jk} = \frac{1}{\sqrt{2}}$  for diagonal neighboring pixels and  $w_{jk} = 0$  otherwise. These choices of  $w_{jk}$ 's result in shift-invariant penalty, *i.e.*:  $R(\lambda)$  is independent of the translations of the image. Fessler and Rogers [43] showed that the penalties of the form (2.19) with conventional choices of  $w_{jk} = 1$ 's result in nonuniform resolution in the reconstructed images. Their analysis demonstrates that the influence of the smoothing penalty for a given pixel depends on the noise variance of the detector measurements whose rays intersect with that given pixel. This effect results in a different “effective” smoothing parameter for each pixel.

To achieve uniform (shift invariant) resolution Fessler and Rogers developed the modified quadratic penalty [43]:

$$R(\lambda) = \frac{1}{2} \sum_j \sum_{k \in \mathcal{N}_j} w_{jk} \kappa_j \kappa_k \psi(\lambda_j - \lambda_k) \quad (2.20)$$

with

$$\kappa_j = s_j \sqrt{\frac{\sum_n a_{nj} q_n(\lambda)}{\sum_n a_{nj}^2}}, \quad (2.21)$$

$$q_n(\lambda) = \frac{c_n^2}{\bar{y}_n(\lambda)} \quad (2.22)$$

and  $g_{nj} = c_n a_{nj} s_j$  such that  $c_n$ 's represent ray dependent factors (attenuation, detector efficiency and deadtime, etc.),  $a_{nj}$ 's represent object-independent geometric response (Figure 2.7) and  $s_j$ 's represent pixel-dependent factors (such as spatial variation in sensitivity). For practical implementation of  $q_n(\lambda)$  <sup>10</sup> a data (*i.e.*:  $y_n$ ) estimated form of (2.22) is suggested as:

$$\hat{q}_n = \frac{c_n^2}{\max\{y_n, 10\}} \quad (2.23)$$

---

<sup>10</sup>For the “prompt” (PR) and “shifted Poisson” (SP) models that are explained in Chapter 3,  $q_n(\lambda)$  can be shown to be equal to  $\frac{c_n^2}{\bar{y}_n(\lambda) + r_n}$  and  $\frac{c_n^2}{\bar{y}_n(\lambda) + 2r_n}$  respectively with  $r_n$ 's denoting the mean of AC events.

where maximization ensures that the denominator is not very close to zero. Then  $q_n(\lambda)$  term in (2.21) is replaced with  $\hat{q}_n$  which is then used in  $R(\lambda)$ . The  $\kappa_j$  terms in the penalty function  $R(\lambda)$  (2.20) cancel out the data dependence of the image resolution at different spatial locations resulting in approximately space invariant (uniform) resolution.

For the transmission scans the  $q_n(\lambda)$  is simply replaced with<sup>11</sup> :

$$q_n(\mu) = \bar{y}_n(\mu) . \quad (2.24)$$

For the PML case, EM algorithm is more difficult to apply. This is because the maximization step of EM has no closed form due to the coupling of the penalty term. Generalized EM (GEM) [48] replaces the true maximization step of EM with a few iterations of coordinate-ascent method. An alternate approach is the simultaneous update algorithm by De Pierro [20, 21] which is more parallelizable than GEM and it is globally convergent. This method decomposes both the log-likelihood and the penalty function using the convexity principle. De Pierro showed that decomposition of the log-likelihood using convexity principle can be used to derive EM algorithm instead of using a statistical framework [20]. One step late (OSL) method of Green [46] overcomes the problem of coupled equations at each iteration by substituting the parameter estimates from the previous iteration into the derivative of the penalty. However this method is not guaranteed to converge, thus one needs to include a line search [60]. Although conjugate gradient methods have rapid convergence for quadratic optimization, usually one needs some form of preconditioner and enforcing non-negativity of the solution is possible but difficult [68]. Space alternating generalized EM (SAGE) [40–42] is a generalized EM type algorithm which updates parameters sequentially by alternating between small hidden-data spaces [41]. As SAGE uses separate hidden data spaces for each parameter, not only the maximization is simplified but convergence rate is also improved compared to EM. SAGE was shown to converge faster than many other monotonic algorithms [40] due to its sequential nature. Moreover, sequential updates of SAGE can handle non-negativity constraints easily. The recently developed paraboloid surrogates algorithm of Fessler and Erdoğan [28, 38], which uses optimum curvatures for the parabolic surrogate functions at each iteration, was shown to converge even faster than the SAGE method.

The OP model is suboptimal for randoms precorrected measurements. The data weighted least squares (DWLS) [31], which is another suboptimal method based essentially on a quadratic approximation to the log-likelihood, can be used as an alternative to OP model. For DWLS objective function Bouman and Sauer have proposed a sequential coordinate-ascent (Gauss Seidel) method [5, 6]. The convergence rate of the sequential Gauss Seidel algorithm was analyzed by Sauer and Bouman using a novel frequency analysis method [7, 79]. Gauss Seidel is a special case of successive over-relaxation (+SOR) algorithm. +SOR was applied to emission reconstruction and the convergence properties were analyzed by Fessler [31]. Unlike simultaneous update methods, for sequential iterative methods the “update order” of the image pixels affects the convergence rate of the algorithm [7, 31, 79]. In Appendix A, we have analyzed the effects of different update orders on the convergence rate properties of the sequential algorithm, as a function of spatial frequency. Although the analysis is carried out for DWLS objective function, one can expect to extend the results to other sequential algorithms like SAGE.

---

<sup>11</sup>For the transmission case  $q_n(\mu)$  can be shown to be equal to  $\frac{(\bar{y}_n(\mu) - r_n)^2}{\bar{y}_n(\mu)}$  and  $\frac{\bar{y}_n(\mu)^2}{\bar{y}_n(\mu) + 2r_n}$  for the PR and the SP models respectively with  $r_n$ 's denoting the mean of AC events.

For transmission tomographic reconstruction, EM does not result in closed form expressions even for ML case [61]. Lange has adopted De Pierro’s convexity method to the transmission problem, which uses a simultaneous update [59, 62]. Although DWLS model leads to systematic bias for transmission image reconstruction for low count scans [32, 34], sequential coordinate ascent methods for DWLS were shown to converge rapidly [32, 79]. However, these methods require the computation of the exponential of the system matrix at each iteration, which is computationally expensive. Moreover these algorithms are not easily parallelizable. Grouped coordinate ascent (GCA) algorithms were suggested [39]) as an alternative to balance the convergence rate and computation per iteration. This method updates pixels in groups which reduces the number of operations per iteration. On the other hand, by choosing pixels in each group well separated spatially, the algorithm does not suffer from slow convergence. GCA uses modifications of De Pierro’s convexity method to compute additively separable surrogate functions. At each iteration within each group a few subiterations of 1D Newton-Rapson method are performed which results in monotonic increase in the log-likelihood. GCA method was shown to have fast convergence in terms of computation time, easily accommodate a nonnegativity constraint, and is easily parallelizable.

## CHAPTER 3

### Exact Log-Likelihood and Approximations

#### 3.1 Measurement Model

The statistical model describes the distribution of each measurement about its mean, and consequently determines a measure of similarity between the actual measurements and the calculated projections (2.16). Since the introduction of an ML-EM [26, 61, 81] algorithm for PET, statistical image reconstruction methods have been based on idealized PET systems with a Poisson statistical model, and ignored the effects of AC events. Several papers have attempted to incorporate AC effects into the Poisson framework for emission tomography [69, 74].

AC rates can be shown to be equal to the multiplication of singles rate and twice the coincidence timing window length as described in section (2.12). However this approach (singles method) [9, 69] is not widely used because of the necessity for additional hardware (since singles rate is not directly available for most PET scanners). Also, methods that require the knowledge of the intrinsic detector efficiencies have limited practical use, since the intrinsic detector efficiencies for the AC events are different than those for the true coincidences [68, 69]. Also, the singles rate can often vary during the course of data acquisition [72].

In conventional PET scans, the data are precorrected for the AC events by *real-time* subtraction of the delayed-window coincidences [53]. The system detects coincidence events during two time windows: “prompt” window and “delayed” window. For each coincidence event in the prompt window, the corresponding sinogram bin is incremented. The statistics of these increments should be well approximated by a Poisson process [102]. However, for coincidence events within the second delayed window, the corresponding sinogram bin is decremented, so the resultant “precorrected” measurements are *not* Poisson. Since prompt events and delayed events are independent Poisson processes [102], the precorrected measurements<sup>1</sup> correspond to the difference of two independent Poisson random variables with variance equal to the sum of the means of the two random variables. In other words, random subtraction compensates in mean for the AC events, but it also increases the variance of the measurement by an amount equal to the mean of the AC events.

As mentioned previously the mean of the projection measurements are  $\bar{y}_n(\lambda)$  (2.14) and  $\bar{y}_n(\mu)$  (2.16) for emission and transmission tomography respectively. In the following

---

<sup>1</sup>Although after real time precorrection one does not have access to the delayed events separately, usually total number of AC events is available at the end of the scan. This information can be used with our proposed methods described in later sections.

analysis we will use a general notation for both transmission and emission tomography with  $\theta = [\theta_1, \dots, \theta_p]'$  denoting the vector of unknown parameters to be estimated and  $\bar{y}_n(\theta)$  denoting the mean of precorrected measurements. We will emphasize the difference in formulations for transmission and emission tomography whenever it occurs.

Let  $Y = [Y_1, \dots, Y_N]'$  denote the vector of the *precorrected* measurements where “ ’ ” denotes vector and matrix transpose. The precorrected measurement for the  $n$ th coincidence detector pair is:

$$Y_n = Y_n^{\text{prompt}} - Y_n^{\text{delay}}, \quad (3.1)$$

where  $Y_n^{\text{prompt}}$  and  $Y_n^{\text{delay}}$  are the number of coincidences within the prompt and delayed windows, respectively. We assume that  $Y_n^{\text{prompt}}$  and  $Y_n^{\text{delay}}$  are statistically independent Poisson random variables [102] with means  $\bar{y}_n^{\text{p}}$  and  $\bar{y}_n^{\text{d}}$  respectively as:

$$\text{E} \{Y_n^{\text{prompt}}\} = \bar{y}_n^{\text{p}}(\theta) = \bar{y}_n(\theta) + r_n \quad (3.2)$$

$$\text{E} \{Y_n^{\text{delay}}\} = \bar{y}_n^{\text{d}} = r_n, \quad (3.3)$$

where the  $r_n \geq 0$  factors denote the mean of AC events.

Since  $Y_n^{\text{prompt}}$  and  $Y_n^{\text{delay}}$  are statistically independent and Poisson:

$$\begin{aligned} \text{E} \{Y_n\} &= \bar{y}_n^{\text{p}}(\theta) - \bar{y}_n^{\text{d}} = \bar{y}_n(\theta), \\ \text{Var} \{Y_n\} &= \bar{y}_n^{\text{p}}(\theta) + \bar{y}_n^{\text{d}} = \bar{y}_n(\theta) + 2r_n. \end{aligned}$$

To illustrate the inaccuracy of the ordinary Poisson measurement model for  $Y_n$ 's, we have performed a small Monte Carlo simulation similar to [31]. The circles in Fig. 3.1 show a simulated histogram for  $Y_n$  generated by a pseudo-random number generator in accordance with the distribution described above (for 300,000 realizations) where  $\bar{y}_n^{\text{p}} = 8$  and  $\bar{y}_n^{\text{d}} = r_n = 1$  (corresponding to 12.5% randoms). Fig. 3.1a shows the approximation based on Gaussian distribution model with mean  $(\bar{y}_n^{\text{p}} - \bar{y}_n^{\text{d}})$  and variance  $(\bar{y}_n^{\text{p}} + \bar{y}_n^{\text{d}})$ . Fig. 3.1b shows the ordinary Poisson (OP) model where approximation is based on a Poisson model with mean  $(\bar{y}_n^{\text{p}} - \bar{y}_n^{\text{d}})$ , the ideal mean. Fig. 3.1c shows the approximation based on a Poisson model with mean  $(\bar{y}_n^{\text{p}} + \bar{y}_n^{\text{d}})$  and then shifted by  $-2\bar{y}_n^{\text{d}}$ : so that resultant approximation corresponds to a model with mean and variance that match both first and second order moments of  $Y_n$ . This approximation corresponds to our proposed “shifted” Poisson (SP) model and it has a better agreement with the precorrected measurement  $Y_n$  than the previous two models. Lastly, Fig. 3.1d shows the Saddle Point (SD) approximation (with best agreement with the exact distribution) which will be introduced in section 3.4.

To make the comparison between different models and the exact distribution more quantitative, we computed different order moments of each model and displayed the results in Table (3.1). We also computed the exact moments of the pre-corrected distribution based on Romani's [78] method as described in [56, p. 191-192]. It is seen that the OP model only matches the mean and 3<sup>rd</sup> order central moment of the exact distribution. The Gaussian model matches mean, variance and partly 4<sup>th</sup> order central moment, while resulting in zero 3<sup>rd</sup> and 5<sup>th</sup> central moments. The SP model matches the mean and variance, and partly the higher order moments. Lastly, the SD method that will be introduced in section 3.4 matches all moments fairly accurately. More importantly, as it will be shown in sections 3.3 and 3.4, the second moments (variance) of the SP and the SD models change with  $\theta$  appropriately, while the variance of Gaussian model is “fixed” independent of  $\theta$ .

Moments	Exact (Theory)	Exact	Gaussian	OP	SP	SD
Mean	7	7.00	7.00	7.00	7.00	7.00
Variance	9	9.00	9.00	7.00	9.00	9.00
3 <sup>rd</sup> Cent. Moment	7	6.98	0.00	7.00	8.99	6.98
4 <sup>th</sup> Cent. Moment	252	252.3	242.9	154.0	251.9	252.3
5 <sup>th</sup> Cent. Moment	637	631.5	0.000	497.0	817.8	638.7

Table 3.1: Sample mean, variance and 3<sup>rd</sup>, 4<sup>th</sup>, 5<sup>th</sup> order central moments of different models compared with those of the exact distribution.

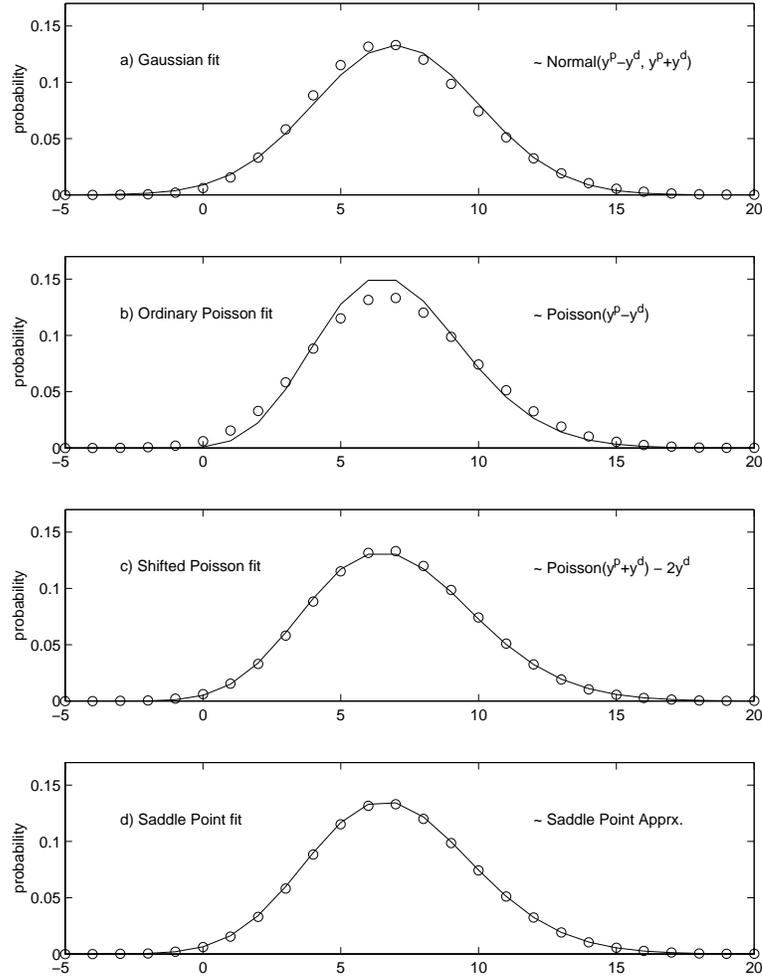


Figure 3.1: Comparison of Gaussian, ordinary Poisson, shifted Poisson and Saddle Point models (-) (with the moments matched to the moments of precorrected measurements), with the empirical distribution (o) of precorrected measurements. From top to bottom: a) Gaussian model. b) Ordinary Poisson (OP) model. c) shifted Poisson (SP) model. d) Saddle Point (SD) approximation that will be introduced in section 3.4.

## 3.2 Exact Log-Likelihood

In this section we will derive the probability distribution and log-likelihood for the random pre-corrected measurements. Let  $y = [y_1, \dots, y_N]'$  be a realization of statistically independent random variables  $Y$  given in (3.1). Under the usual assumption of independence between different rays, one can express the exact distribution of  $Y$  using total probability:

$$P(Y = y; \theta) = \prod_{n=1}^N P(Y_n = y_n; \theta) \quad (3.4)$$

$$\begin{aligned} &= \prod_{n=1}^N \sum_{m=0}^{\infty} P(Y_n = y_n | Y_n^{\text{delay}} = m; \theta) P(Y_n^{\text{delay}} = m) \\ &= \prod_{n=1}^N \sum_{m=0}^{\infty} P(Y_n^{\text{prompt}} = y_n + m; \theta) P(Y_n^{\text{delay}} = m). \end{aligned} \quad (3.5)$$

Since, both  $Y_n^{\text{prompt}}$  and  $Y_n^{\text{delay}}$  are statistically independent and Poisson distributed:

$$P(Y = y; \theta) = \prod_{n=1}^N \sum_{m=[-y_n]_+}^{\infty} \frac{[\bar{y}_n^{\text{P}}(\theta)]^{y_n+m} e^{-\bar{y}_n^{\text{P}}(\theta)}}{(y_n + m)!} \frac{r_n^m e^{-r_n}}{m!}, \quad (3.6)$$

where  $[x]_+ = x$  if  $x > 0$  and is 0 otherwise.

### 3.2.1 Infinite Summation Form of Exact Log-Likelihood

Using the pmf distribution (3.6) the exact log-likelihood for  $\theta$  can be written as:

$$\begin{aligned} L(\theta) &= \log P(Y = y; \theta) \\ &= \sum_{n=1}^N \log \left( \sum_{m=[-y_n]_+}^{\infty} \frac{[\bar{y}_n^{\text{P}}(\theta)]^{y_n+m} r_n^m}{(y_n + m)! m!} \right) - (\bar{y}_n^{\text{P}}(\theta) + r_n). \end{aligned} \quad (3.7)$$

### 3.2.2 Bessel Function Form of Exact Log-Likelihood

The infinite summations form of the pmf (3.6) of the difference of two Poisson random variables can also be expressed using modified Bessel functions [19, 82]. In this section we describe this type implementation of the exact log-likelihood. Since numerical approximations to Bessel functions are available in many programming libraries, this alternative implementation of exact log-likelihood may be useful in some applications.

The exact log-likelihood function (3.7) can be written as:

$$L(\theta) = \sum_{n=1}^N \log (v_n(\bar{y}_n^{\text{P}}(\theta), r_n)) - (\bar{y}_n^{\text{P}}(\theta) + r_n), \quad (3.8)$$

with

$$v_n(\bar{y}_n^{\text{P}}(\theta), r_n) = \begin{cases} \sum_{m=0}^{\infty} \frac{[\bar{y}_n^{\text{P}}(\theta)]^{y_n+m} r_n^m}{(y_n + m)! m!}, & y_n \geq 0 \\ \sum_{k=0}^{\infty} \frac{[\bar{y}_n^{\text{P}}(\theta)]^k r_n^{k-y_n}}{k! (k - y_n)!}, & y_n < 0. \end{cases} \quad (3.9)$$

For  $r_n = 0$ , the exact log-likelihood equals to the trivial OP log-likelihood (3.20), thus in the following we concentrate on the case where  $r_n > 0$ . For  $y_n \geq 0$ , from (3.9):

$$\begin{aligned}
v_n(\bar{y}_n^{\text{P}}(\theta), r_n) &= \bar{y}_n^{\text{P}}(\theta)^{y_n} \sum_{m=0}^{\infty} \frac{(-1)^m \left( \frac{2i\sqrt{\bar{y}_n^{\text{P}}(\theta) r_n}}{2} \right)^{2m}}{m! (m + y_n)!} \\
&= \left( \frac{1}{i} \sqrt{\frac{\bar{y}_n^{\text{P}}(\theta)}{r_n}} \right)^{y_n} \left( \frac{2i\sqrt{\bar{y}_n^{\text{P}}(\theta) r_n}}{2} \right)^{y_n} \sum_{m=0}^{\infty} \frac{(-1)^m \left( \frac{2i\sqrt{\bar{y}_n^{\text{P}}(\theta) r_n}}{2} \right)^{2m}}{m! (m + y_n)!} \\
&= \left( \frac{1}{i} \sqrt{\frac{\bar{y}_n^{\text{P}}(\theta)}{r_n}} \right)^{y_n} J_{y_n} \left( 2i\sqrt{\bar{y}_n^{\text{P}}(\theta) r_n} \right), \tag{3.10}
\end{aligned}$$

where  $i = \sqrt{-1}$  and  $J_n(\cdot)$  is the Bessel function of the first kind of order  $n$ :

$$J_n(\beta) = \sum_{m=0}^{\infty} \frac{(-1)^m \left( \frac{\beta}{2} \right)^{n+2m}}{m! (m + n)!} \tag{3.11}$$

$$= \left( \frac{\beta}{2} \right)^n \sum_{m=0}^{\infty} \frac{(-1)^m \left( \frac{\beta}{2} \right)^{2m}}{m! (m + n)!}. \tag{3.12}$$

Note that the argument of the Bessel function in (3.10) is complex — a feature not available in many Bessel programming subroutines.

For  $y_n < 0$ , from (3.9) :

$$\begin{aligned}
v_n(\bar{y}_n^{\text{P}}(\theta), r_n) &= r_n^{-y_n} \sum_{k=0}^{\infty} \frac{(-1)^k \left( \frac{2i\sqrt{\bar{y}_n^{\text{P}}(\theta) r_n}}{2} \right)^{2k}}{k! (k - y_n)!} \\
&= \left( \frac{1}{i} \sqrt{\frac{r_n}{\bar{y}_n^{\text{P}}(\theta)}} \right)^{-y_n} \left( \frac{2i\sqrt{\bar{y}_n^{\text{P}}(\theta) r_n}}{2} \right)^{-y_n} \sum_{k=0}^{\infty} \frac{(-1)^k \left( \frac{2i\sqrt{\bar{y}_n^{\text{P}}(\theta) r_n}}{2} \right)^{2k}}{k! (k - y_n)!} \\
&= \left( \frac{1}{i} \sqrt{\frac{r_n}{\bar{y}_n^{\text{P}}(\theta)}} \right)^{-y_n} J_{(-y_n)} \left( 2i\sqrt{\bar{y}_n^{\text{P}}(\theta) r_n} \right). \tag{3.13}
\end{aligned}$$

Thus, using (3.10) and (3.13) we can rewrite  $v_n(\bar{y}_n^{\text{P}}(\theta), r_n)$  as:

$$v_n(\bar{y}_n^{\text{P}}(\theta), r_n) = \begin{cases} \left( \frac{1}{i} \sqrt{\frac{\bar{y}_n^{\text{P}}(\theta)}{r_n}} \right)^{y_n} J_{y_n} \left( 2i\sqrt{\bar{y}_n^{\text{P}}(\theta) r_n} \right), & y_n \geq 0 \\ \left( \frac{1}{i} \sqrt{\frac{r_n}{\bar{y}_n^{\text{P}}(\theta)}} \right)^{-y_n} J_{(-y_n)} \left( 2i\sqrt{\bar{y}_n^{\text{P}}(\theta) r_n} \right), & y_n < 0. \end{cases} \tag{3.14}$$

In our PET simulations and experimental studies for the exact log-likelihood we use the above implementation interchangeably with the truncated implementation (of the infinite

summations) of the exact log-likelihood (3.7). However, one must be careful about the numerical stability of the above Bessel function implementation since the power term  $(.)^{y_n}$  and the Bessel term  $J_{y_n}(\cdot)$  increase very quickly with  $y_n$ .

Since the exact log-likelihood function is complicated because of the infinite summations (3.7) and complex Bessel functions (3.8), (3.14), in the light of the Monte Carlo simulations that we have performed previously, the following two sections develop tractable yet accurate approximations to  $L(\theta)$ .

### 3.3 Simple Approximations to the Likelihood

In this section, we first review the conventional approximations to the exact log-likelihood  $L(\theta)$ : the weighted least square (WLS) model and the conventional OP model. Then we introduce the new shifted Poisson SP model [97].

#### 3.3.1 Quadratic Approximations

The conventional quadratic approximation to the exact log-likelihood function results in the weighted least squares objective function  $L_{\text{WLS}}(\theta)$ . As mentioned in [56, p. 192], Fisz [44] also analyzed the difference between the Gaussian distribution and the pmf of difference of two Poisson random variables.

#### Weighted Least Squares with Data Weighting

For transmission tomography the data-weighted least squares (DWLS) objective function is [32, 79]:

$$L_{\text{WLS}}(\mu) = -\frac{1}{2} \sum_{n=1, y_n > s_n^T}^N (l_n(\mu) - \hat{l}_n)^2 \frac{1}{\hat{\sigma}_{l_n}^2}, \quad (3.15)$$

where  $\hat{l}_n = \log\left(\frac{b_n}{y_n - s_n^T}\right)$  is the method-of-moments estimate of the *line integral of the attenuation*  $l_n(\mu)$  and  $\hat{\sigma}_{l_n}^2 = \frac{y_n + 2r_n}{(y_n - s_n^T)^2}$ . The  $n$ th weighting factor  $\hat{\sigma}_{l_n}^2$  is a *data* estimated variance of  $\hat{l}_n(y_n)$  based on a second-order Taylor expansion around  $\hat{l}_n(\bar{y}_n)$  (Appendix B). This weighting is critical for the DWLS method. The errors corresponding to projections with large values of  $y_n$  are weighted more heavily. These projections pass through less dense regions and consequently have higher SNR values. Rays where  $y_n \leq s_n^T$  are excluded from the sum since  $\hat{l}_n$  is undefined.

For emission tomography the DWLS objective function is [31]:

$$L_{\text{WLS}}^{\text{E}}(\lambda) = -\frac{1}{2} \sum_{n=1}^N (\bar{y}_n(\lambda) - y_n)^2 \frac{1}{\hat{\sigma}_{y_n}^2}, \quad (3.16)$$

where  $\hat{\sigma}_{y_n}^2 = \max\{y_n + 2r_n, c\}$  is the *data* estimated variance of the emission measurements and  $c$  is a small positive integer. These weighting factors are critical to the DWLS method. Generally an important benefit of statistical image reconstruction methods over FBP method is the nonuniform weighting of the measurements, where the weighting factors reflect the relative information content of each measurement [31]. The ML-EM algorithm (2.18) implicitly incorporates such a weighting by dividing each measurement by its predicted value before backprojecting. This is in complete contrast with FBP methods, since

FBP treats all measurements equally<sup>2</sup>, despite large variations in counts and correction factors. Similar to ML-EM method, the DWLS method accounts for the relative information of each measurement through the weights  $\hat{\sigma}_{l_n}^2$  and  $\hat{\sigma}_{y_n}^2$ . However, these weights are usually suboptimal since they are directly driven from experimental data instead of some parametric relation with the unknown image. Although these weights become more accurate with increased count rates, one might need to incorporate some smoothing and thresholding methods for low count rates [33].

Alternatively, the choice of  $\hat{\sigma}_{l_n}^2 = 1$  and  $\hat{\sigma}_{y_n}^2 = 1$  in the above objective functions results in the unweighted least-squares (ULS) approach, which leads to much higher variance.

The familiar form of DWLS objective function invites quadratically penalized linear least squares estimation method such as:

$$\hat{\lambda}^{WLS} = [\mathbf{G}'\Sigma\mathbf{G} + \beta\mathbf{R}]^{-1} \mathbf{G}'\Sigma^{-1}y, \quad (3.17)$$

with  $\mathbf{G} = [g_{ij}]$  and  $\Sigma$  a diagonal matrix with weights and  $\mathbf{R}$  the quadratic penalty such as (2.19). However, this kind of “direct” least squares estimation is usually computationally impractical due to the large size of the system matrix  $\mathbf{G}$  for PET. Furthermore conventional linear least square estimates can result in negative pixel values which are physically not possible. And, lastly nonquadratic penalties can not be incorporated in the linear least squares form. Thus, one usually needs to implement iterative algorithms for maximizing the DWLS objective function.

### Weighted Least Squares with Parameter Dependent Weighting

Although fast maximization algorithms exist for the data weighted least squares (DWLS) objective function, the data based weighting is suboptimal (especially at low counts) and it can result in bias in the reconstructed images. To overcome this problem one can use weighting factors which depend on the parameter to be estimated [1]. The parameter dependent weight factors for the WLS estimation are:  $\hat{\sigma}_{l_n}^2(\mu) = \frac{\bar{y}_n(\mu) + 2r_n}{(\bar{y}_n(\mu) - s_n^T)^2}$  in (3.15) for “line integral” WLS objective function and  $\hat{\sigma}_{y_n}^2(\lambda) = \bar{y}_n(\lambda) + 2r_n$  in (3.16) for “measurement” WLS objective function. Since these objective functions have parameter dependent weights we call them as “line integral - parametric weighted least squares” (L-PWLS) and “measurement - parametric weighted least squares” (M-PWLS) objective functions respectively. With this kind of *parameter* dependent weighting the model matches the second moment appropriately. This approach is a special case of quasi-likelihood estimation [22, 95].

Although this approach can result in better estimates than the DWLS, we do not pursue this method because of increased computational requirements [1]. Moreover, for the transmission problem L-PWLS is not guaranteed to be concave<sup>3</sup>. The SP method that will be introduced in Section 3.3.3 also matches first and second moments appropriately, and moreover the SP model fits to the asymmetric pmf distribution of pre-corrected data better than the Gaussian model (Fig. 3.1). Thus, in our 2-D simulations and experimental studies we concentrate on the computationally efficient DWLS method and we refer to it simply as WLS method.

---

<sup>2</sup>FBP can also be thought as an unweighted least squares reconstruction with appropriate penalty function [33].

<sup>3</sup>As also can be observed from Fig 3.2 which will be described in section 3.3.3.

### 3.3.2 Ordinary Poisson (OP) Approximation

The conventional approach is to assume (approximate) that  $\{Y_n\}_{n=1}^N$  are distributed as independent Poisson random variables with mean  $\bar{y}_n(\theta)$  (2.16), *i.e.*:

$$P(Y = y; \theta) \approx \prod_{n=1}^N P_{\text{OP}}(Y_n = y_n; \theta) \quad (3.18)$$

$$= \prod_{n=1}^N \frac{[\bar{y}_n(\theta)]^{y_n} e^{-\bar{y}_n(\theta)}}{y_n!}. \quad (3.19)$$

The log-likelihood corresponding to this OP approximation is [61]:

$$L_{\text{OP}}(\theta) = \sum_{n=1}^N y_n \log \bar{y}_n(\theta) - \bar{y}_n(\theta) \quad (3.20)$$

disregarding the constants independent of  $\theta$ .

As mentioned previously this approximation only matches the first order and third order moments of the data, thus it is clearly a suboptimal approach for  $r_n > 0$ . This model becomes accurate only as  $r_n \rightarrow 0$ . However, the OP model is the conventional method PET reconstruction and thus we include this model in our studies for comparison purposes.

### 3.3.3 Shifted Poisson (SP) Approximation

In the light of Fig. 3.1c, a better approach is to match both the first and second order moments by approximating the quantities<sup>4</sup>  $\{Y_n + 2r_n\}_{n=1}^N$  as having Poisson distributions with means  $\{\bar{y}_n(\theta) + 2r_n\}$ :

$$P(Y = y; \theta) \approx \prod_{n=1}^N P_{\text{SP}}(Y_n = y_n; \theta) \quad (3.21)$$

$$= \prod_{n=1}^N \frac{[\bar{y}_n(\theta) + 2r_n]^{y_n + 2r_n} e^{-(\bar{y}_n(\theta) + 2r_n)}}{c(\bar{y}_n(\theta), 2r_n) \Gamma(y_n + 2r_n + 1)}, \quad (3.22)$$

where we define the constant

$$c(\bar{y}, 2r) = e^{-(\bar{y} + 2r)} \sum_{k=\lceil -2r \rceil}^{\infty} \frac{(\bar{y} + 2r)^{k+2r}}{\Gamma(k + 2r + 1)}$$

to ensure that the pmf  $P_{\text{SP}}(y)$  sums to one (where we define  $\lceil x \rceil = k$  with  $k$  being the smallest integer such that  $k \geq x$ ) and  $\Gamma(x)$  is the gamma function:

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt.$$

Note that the gamma function satisfies the recurrence relation  $\Gamma(x + 1) = x\Gamma(x)$  and when  $x$  is an integer the gamma function is just the familiar factorial function, but offset by one, *i.e.*:  $\Gamma(k + 1) = k!$  [75].

---

<sup>4</sup>In practice we use  $\hat{r}_n$ 's, see Section 4.8.5.

To simplify the corresponding log-likelihood to this SP approximation, we ignore the dependence<sup>5</sup> of  $c(\bar{y}_n(\theta), 2r_n)$  on  $\theta$ . This leads to our SP objective function:

$$L_{\text{SP}}(\theta) = \sum_{n=1}^N (y_n + 2r_n) \log(\bar{y}_n(\theta) + 2r_n) - (\bar{y}_n(\theta) + 2r_n). \quad (3.23)$$

For the transmission problem we can write the above objective function as

$$L_{\text{SP}}(\mu) = \sum_{n=1}^N h_n(l_n(\mu))$$

where  $h_n(l) = (y_n + 2r_n) \log(b_n e^{-l} + s_n^T + 2r_n) - (b_n e^{-l} + s_n^T + 2r_n)$ . In Appendix B we show that for transmission problem  $L_{\text{WLS}}(\mu)$  corresponds to the summations of second order Taylor series expansion of  $h_n(l_n(\mu))$  about  $h_n(\hat{l}_n)$  where  $\hat{l}_n = \log\left(\frac{b_n}{y_n - s_n^T}\right)$ .

Although both WLS and SP methods match two moments, in WLS the second moment of  $\hat{l}_n(y_n)$  is “fixed” independently of  $\theta$ , whereas in the SP model the moments vary with  $\bar{y}_n(\theta)$  appropriately. This turns out to be a very important difference between the two models as will be observed in the next sections.

Fig. 3.2 compares the actual log-likelihood function and the approximations for transmission problem as a function of a single projection across the reconstructed image. It is observed that  $L_{\text{SP}}(\theta)$  agrees fairly well with the exact log-likelihood  $L(\theta)$ , however quadratic objective function  $L_{\text{WLS}}(\theta)$  (DWLS) and OP model objective function  $L_{\text{OP}}(\theta)$  exhibit a noticeable departure from the exact log-likelihood function. The parametric weighted least squares models: L-PWLS and M-PWLS are also included for comparison purposes<sup>6</sup>.

### 3.4 Saddle-point (SD) Approximation

An alternative to the previous approximations for the exact pmf (3.6) of precorrected measurements is to make second order Taylor series approximations in the  $z$ -transform domain (i.e. on the probability generating function) and then to carry out the inverse transform. Snyder *et al.* [49, 83] have applied the saddle-point approximation to the distribution of the sum of independent Gaussian and Poisson random variables. Here we apply the saddle-point method to the distribution of the difference of two independent Poisson random variables. We performed a quadratic<sup>7</sup> approximation to the probability generating function and then carried out the inverse transform to find the pmf. We will show that our saddle-point approach leads to a more accurate, yet tractable approximation than the previously introduced models.

Let  $U \sim \text{Poisson}(\alpha)$ ,  $V \sim \text{Poisson}(\beta)$  and  $Y = U - V$  with pmf's  $P_U(k)$ ,  $P_V(k)$  and  $P_Y(k)$  respectively. When  $U$  and  $V$  are independent, the generating function of  $Y$  is:

$$G_Y(z) = \sum_k z^k P_Y(k) = G_U(z) G_V(z^{-1})$$

<sup>5</sup>It can be shown that  $1 \geq c(\bar{y}, 2r) > (1 - e^{-(\bar{y}+2r)})$  which approaches to unity as  $\bar{y}$  or  $2r$  increase.

<sup>6</sup>As mentioned in Section 3.3.1, both L-PWLS and M-PWLS models are computationally more expensive compared to the DWLS method. Moreover, the L-PWLS objective function is not guaranteed to be globally concave. In our 2-D simulations and experimental studies we concentrate on the computationally efficient DWLS method and we refer to this method simply as WLS method.

<sup>7</sup>The quadratic form of the probability generating function approximation shows resemblance to the Gram-Charlier [56] approximation used for the probability distribution functions.

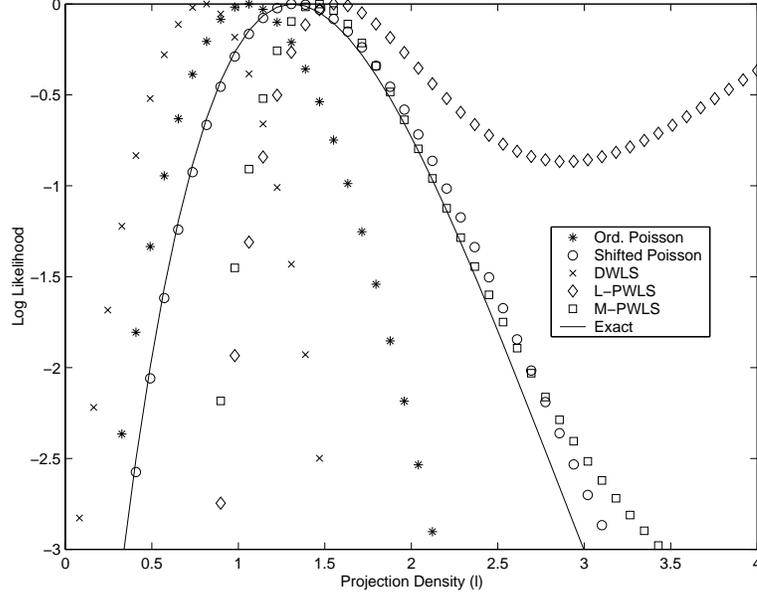


Figure 3.2: Comparison of exact log-likelihood function with objective functions of different models as a function of single projection across the reconstructed image. The proposed shifted Poisson model agrees with exact log-likelihood better than the quadratic and OP models.

where  $G_U(z) = \exp(\alpha(z - 1))$  and  $G_V(z) = \exp(\beta(z - 1))$ . In terms of the generating function,  $P_Y(k)$  is given by the contour integral

$$P_Y(k) = \frac{1}{2\pi j} \oint_{C^+} z^{-k-1} G_Y(z) dz = \frac{1}{2\pi j} \oint_{C^+} e^{\Phi_k(z)} dz, \quad (3.24)$$

where  $j = \sqrt{-1}$  and the contour  $C^+$  must lie in the region of convergence of  $G_Y(z)$  and enclose the origin, and

$$\begin{aligned} \Phi_k(z) &= -(k+1) \log(z) + \alpha(z-1) + \beta(z^{-1} - 1) \\ \frac{d\Phi_k(z)}{dz} &= \Phi_k^{(1)}(z) = -\frac{(k+1)}{z} + \alpha - \frac{\beta}{z^2} \\ \frac{d^2\Phi_k(z)}{dz^2} &= \Phi_k^{(2)}(z) = \frac{(k+1)}{z^2} + \frac{2\beta}{z^3}. \end{aligned}$$

We observe that  $\Phi_k(z)$  (and hence the integrand  $e^{\Phi_k(z)}$ ) is convex for  $z \in \Re$ ,  $z > 0$  and  $k \geq 0$ . The integrand has a minimum at  $x_o \in \Re$ ,  $x_o > 0$  which is called the saddle-point, *i.e.*:

$$\Phi_k^{(1)}(x_o) = -\frac{(k+1)}{x_o} + \alpha - \frac{\beta}{x_o^2} = 0 \quad \text{and } x_o > 0$$

which yields

$$x_o = \frac{(k+1) + v_k}{2\alpha} = \frac{2\beta}{-(k+1) + v_k}, \quad (3.25)$$

where  $v_k = x_o^2 \Phi_k^{(2)}(x_o) = \sqrt{(|k| + 1)^2 + 4\alpha\beta}$ .

Following [49], we deform the contour  $C^+$  in (3.24) into a vertical line  $C_0$  through saddle point  $x_o$ , as  $z = x_o + jy$ ,  $-\infty < y < \infty$  and a semicircle  $C_1$  around the left half plane at infinity, Fig. 3.3. This contour is permissible for  $k \geq 0$ , since the only singularities of the integrand are at  $z = 0$  and  $z = \infty + j0$ . If  $|z| \rightarrow \infty$  for  $\Re[z] < x_o$  then  $e^{\Phi_k(z)} \rightarrow 0$ . Hence the contribution of the semicircle around the left half plane at infinity vanishes and (3.24) reduces to

$$P_Y(k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{\Phi_k(x_o + jy)} dy. \quad (3.26)$$

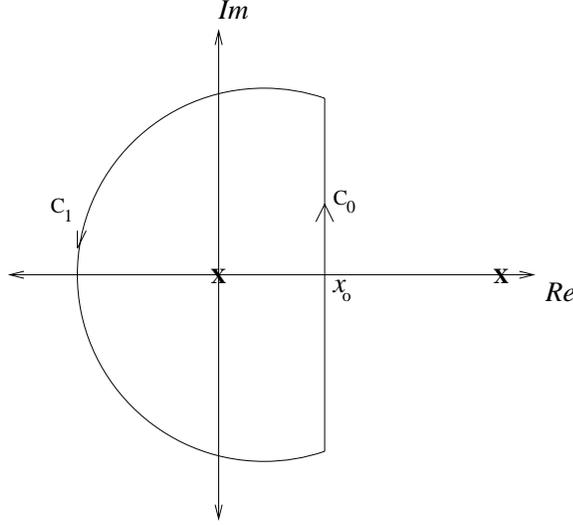


Figure 3.3: Deformation of the contour  $C^+$  in complex plane into a vertical line  $C_0$  through saddle point  $x_o$  and a semicircle  $C_1$  around the left half plane at infinity. The singularities of the integrand are at  $z = 0$  and  $z = \infty + j0$  for  $k \geq 0$ .

Expanding  $\Phi_k(z)$  in Taylor's series around  $z = x_o$ , one obtains:

$$\begin{aligned} \exp[\Phi_k(z)] &= \exp \left[ \Phi_k(x_o) + \frac{1}{2} \Phi_k^{(2)}(x_o)(z - x_o)^2 \right. \\ &\quad \left. + \sum_{l=3}^{\infty} \frac{1}{l!} \Phi_k^{(l)}(x_o)(z - x_o)^l \right] \\ &= \exp \left[ \Phi_k(x_o) + \frac{1}{2} \Phi_k^{(2)}(x_o)(z - x_o)^2 \right] F(z; x_o) \end{aligned}$$

since  $\Phi_k^{(1)}(x_o) = 0$  and with

$$F(z; x_o) = \exp \left[ \sum_{l=3}^{\infty} \frac{1}{l!} \Phi_k^{(l)}(x_o)(z - x_o)^l \right].$$

Using series expansion of exponential functions we can write  $F(z; x_o)$  :

$$F(z; x_o) = \left[ 1 + \frac{\Phi_k^{(3)}(x_o)}{6}(z - x_o)^3 + \left( \frac{\Phi_k^{(3)}(x_o)}{6}(z - x_o)^3 \right)^2 + \dots \right]$$

$$\begin{aligned}
& \cdot \left[ 1 + \frac{\Phi_k^{(4)}(x_o)}{24}(z - x_o)^4 + \left( \frac{\Phi_k^{(4)}(x_o)}{24}(z - x_o)^4 \right)^2 + \dots \right] \\
& \dots \\
& = \left[ 1 + \frac{\Phi_k^{(3)}(x_o)}{6}(z - x_o)^3 + \frac{\Phi_k^{(4)}(x_o)}{24}(z - x_o)^4 + \frac{\Phi_k^{(5)}(x_o)}{120}(z - x_o)^5 + \right. \\
& \quad \left( \left( \frac{\Phi_k^{(3)}(x_o)}{6} \right)^2 + \frac{\Phi_k^{(6)}(x_o)}{6!} \right) (z - x_o)^6 + \dots \\
& \quad \left. b_k (z - x_o)^k + \dots \right],
\end{aligned}$$

where  $b_k$  coefficients can be computed by a straightforward process of combining terms of same power in the above expression.

Thus the integral (3.26) becomes

$$P_Y(k) = \frac{e^{\Phi_k(x_o)}}{2\pi} \int_{-\infty}^{\infty} e^{\frac{1}{2}\Phi_k^{(2)}(x_o)(jy)^2} F(x_o + jy; x_o) dy \quad (3.27)$$

$$\begin{aligned}
& = \frac{e^{\Phi_k(x_o)}}{2\pi} \int_{-\infty}^{\infty} e^{\frac{1}{2}\Phi_k^{(2)}(x_o)(jy)^2} \left[ 1 + \frac{\Phi_k^{(3)}(x_o)}{6}(jy)^3 + \dots \right] dy \\
& = \frac{e^{\Phi_k(x_o)}}{\sqrt{2\pi\Phi_k^{(2)}(x_o)}} [1 + R] \\
& = \frac{x_o^{-k} e^{v_k - \alpha - \beta}}{\sqrt{2\pi v_k}} [1 + R], \quad (3.28)
\end{aligned}$$

where

$$R = \frac{\Phi_k^{(4)}(x_o)}{8 [\Phi_k^{(2)}(x_o)]^2} + \dots + (-1)^k \frac{(2k)!}{2^k k!} \frac{b_{2k}}{[\Phi_k^{(2)}(x_o)]^k} + \dots$$

since the odd terms vanish. Using the algorithm by Rice [77], the residuum  $R$  can be written as:

$$R = \frac{1}{24(k+1)} \left[ \frac{-5 + 12\sqrt{1+\eta} - 9(1+\eta)}{(1+\eta)^{3/2}} \right] + O \left[ \left( \frac{1}{k+1} \right)^2 \right],$$

where  $\eta = \frac{4\alpha\beta}{(k+1)^2}$ . The residuum asymptotically goes to zero as  $k \rightarrow \infty$  and more importantly we have observed empirically that the approximation error is negligibly small even for very small values of  $k$ . Neglecting  $R$  in (3.28) results in our saddle-point approximation for the pmf  $P_Y(k)$  as:

$$P_Y(k) \simeq P_Y^s(k) = \frac{x_o^{-k} e^{v_k - \alpha - \beta}}{\sqrt{2\pi v_k}}, \quad k \geq 0. \quad (3.29)$$

For  $k < 0$  the integrand in (3.24) is not guaranteed to be convex for  $z > 0$ . Moreover, the integrand does not vanish along the semicircle around the left half plane at infinity.

Thus we use the change of variables  $w = 1/z$  in (3.24), so that:

$$P_Y(k) = \frac{1}{2\pi j} \oint_{C^+} w^{k-1} G_Y(w^{-1}) dw = \frac{1}{2\pi j} \oint_{C^+} e^{\bar{\Phi}_k(w)} dw \quad (3.30)$$

where

$$\bar{\Phi}_k(w) = (k-1) \log(w) + \alpha(w^{-1} - 1) + \beta(w - 1).$$

Following similar steps as the case for  $k \geq 0$ , the saddle point approximation for  $k < 0$  can be shown to be :

$$P_Y(k) \simeq P_Y^s(k) = \frac{w_o^k e^{v_k - \alpha - \beta}}{\sqrt{2\pi v_k}}, \quad k < 0 \quad (3.31)$$

where

$$w_o = \frac{-(k-1) + v_k}{2\beta} = \frac{2\alpha}{(k-1) + v_k}.$$

Thus, combining (3.29) and (3.31) and disregarding constants independent of  $\theta$ , the saddle-point (SD)<sup>8</sup> approximation for the log-likelihood (3.7) is:

$$\begin{aligned} L_{SD}(\theta) &= \sum_{n=1}^N \log P_Y^s(y_n; \bar{y}_n(\theta)) \\ &= \sum_{n=1}^N h_n^s(\theta), \end{aligned} \quad (3.32)$$

where

$$h_n^s(\theta) \triangleq \begin{cases} y_n \log \left( \frac{\bar{y}_n(\theta) + r_n}{y_n + 1 + u_n(\theta)} \right) - t_n(\theta), & y_n \geq 0 \\ y_n \log \left( \frac{\bar{y}_n(\theta) + r_n}{y_n - 1 + u_n(\theta)} \right) - t_n(\theta), & y_n < 0 \end{cases}, \quad (3.33)$$

with

$$t_n(\theta) \triangleq \bar{y}_n(\theta) + u_n(\theta) - \frac{1}{2} \log u_n(\theta) \quad (3.34)$$

$$u_n(\theta) \triangleq \sqrt{(|y_n| + 1)^2 + 4(\bar{y}_n(\theta) + r_n)r_n}, \quad (3.35)$$

and disregarding constants independent of  $\theta$ . At first it might appear that for  $y_n < 0$ , as  $r_n \rightarrow 0$  the denominator  $(y_n - 1 + u_n(\theta)) \rightarrow 0$  resulting in computational instability. However, this does not happen since  $P(y_n < 0) \rightarrow 0$  as  $r_n \rightarrow 0$ . Moreover, in none of the experimental and simulation studies did we encounter such a problem.

The approximation (3.32) is considerably simpler than the exact log-likelihood (3.7), since no infinite sums or factorials are needed. Nevertheless, it is remarkably accurate as shown below. Also, one can observe that as  $r_n \rightarrow 0$ ,  $L_{SD}(\theta) \rightarrow [y_n \log \bar{y}_n(\theta) - \bar{y}_n(\theta)] = L_{OP}(\theta)$  (to within constants independent of  $\theta$ ), which is expected because for  $r_n = 0$  the ordinary Poisson model is appropriate.

---

<sup>8</sup>We caution the reader that the notation used for the shifted Poisson model is ‘‘SP’’ and the notation used for the saddle-point method is ‘‘SD’’.

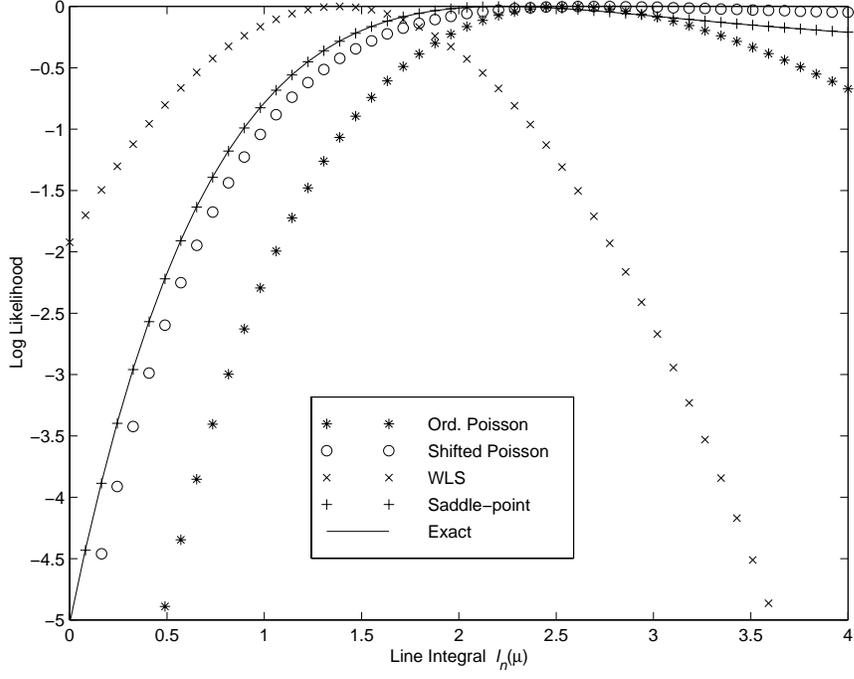


Figure 3.4: Representative comparison of exact log-likelihood function with objective functions of different models as a function of line integral  $l_n(\mu)$ . Randoms rate is 5%. The proposed saddle-point approximation agrees with exact log-likelihood significantly better than the other models.

Fig. 3.4 shows a representative comparison of the exact log-likelihood function and the approximations as a function of  $\mu$  (for noisy set of measurements). The  $L_{\text{WLS}}(\mu)$  is particularly poor, in part because of the condition  $y_n > 0$  in (3.15). Although  $L_{\text{SP}}(\mu)$  fits the exact log-likelihood better than  $L_{\text{WLS}}(\mu)$  and  $L_{\text{OP}}(\mu)$ , clearly  $L_{\text{SD}}(\mu)$  has the best agreement with the exact log-likelihood  $L(\mu)$ . In a large number of additional comparisons not shown due to space considerations, we have observed that  $L_{\text{SD}}(\mu)$  agrees remarkably well with the exact log-likelihood  $L(\mu)$  and clearly better than the other models.

Partial derivatives of  $h_n^{\text{SD}}(\theta)$  with respect to the mean  $\bar{y}_n(\theta)$  can be written as:

$$\frac{\partial h_n^{\text{SD}}(\bar{y})}{\partial \bar{y}} = \begin{cases} \frac{y_n}{\bar{y} + r_n} - 1 + \frac{2r_n}{u_n(\bar{y})} \left[ -\frac{y_n}{y_n + 1 + u_n(\bar{y})} + 1 - \frac{1}{2u_n} \right], & y_n \geq 0 \\ \frac{y_n}{\bar{y} + r_n} - 1 + \frac{2r_n}{u_n(\bar{y})} \left[ -\frac{y_n}{y_n - 1 + u_n(\bar{y})} + 1 - \frac{1}{2u_n} \right], & y_n < 0 \end{cases} \quad (3.36)$$

$$\frac{\partial^2 h_n^{\text{SD}}(\bar{y})}{\partial \bar{y}^2} = \begin{cases} \frac{-y_n}{(\bar{y} + r_n)^2} + \frac{4r_n^2}{u_n^3(\bar{y})} \left[ \frac{1 - u_n(\bar{y})}{u_n(\bar{y})} + \frac{y_n(1 + y_n + 2u_n(\bar{y}))}{(y_n + 1 + u_n(\bar{y}))^2} \right], & y_n \geq 0 \\ \frac{-y_n}{(\bar{y} + r_n)^2} + \frac{4r_n^2}{u_n^3(\bar{y})} \left[ \frac{1 - u_n(\bar{y})}{u_n(\bar{y})} + \frac{y_n(-1 + y_n + 2u_n(\bar{y}))}{(y_n - 1 + u_n(\bar{y}))^2} \right], & y_n < 0 \end{cases} \quad (3.37)$$

and the partials w.r.t.  $\theta$  can be found using chain rule. For the emission case  $\frac{\partial \bar{y}_n(\lambda)}{\partial \lambda_j} = g_{nj}$ ,

thus it is trivial to apply chain rule to the above expressions. For transmission case, deriving the partial derivatives w.r.t.  $\mu$  requires some tedious algebra, which is shown at the end of Appendix C. Using these closed-form expressions for the partial derivatives, one can easily apply coordinate-ascent type algorithms for maximizing the above  $L_{SD}(\theta)$ .

### 3.5 Exact Log-likelihood for Prompt Coincidence Data

If one has access to the prompt data  $y_n^p$  separately, then the exact log-likelihood  $L_{PR}(\mu)$  can be written as [32, 61]:

$$L_{PR}(\theta) = \sum_{n=1}^N (y_n^p + r_n) \log(\bar{y}_n(\theta) + r_n) - (\bar{y}_n(\theta) + r_n). \quad (3.38)$$

As mentioned earlier we recommend separate acquisition of prompt and delayed coincidence data whenever possible. However, because of hardware, software and data storage limitations most PET centers use *only* randoms precorrected data. We include  $L_{PR}(\theta)$  here since we compare its results with the methods for randoms-precorrected data in the next chapters.

## CHAPTER 4

### PET Transmission Scans

#### 4.1 Introduction

To obtain accurate images of the radioactivity distribution within a patient using emission computed tomography, one must correct for the effects of attenuation [54] and accidental coincidences [9]. Reconstructing images of attenuation distributions (attenuation maps) from noisy transmission scans has desirable properties (see Section 2.3.2) such as better noise performance in attenuation correction [66, 69] and anatomical localization [2]. In transmission scans, photons that originate from different transmission sources (rod or ring sources) cause most AC events. The effect of AC events is most severe for rays with low true coincidence rates [18], such as those traversing the abdomen or thorax.

In this chapter we derive accurate statistical image reconstruction methods for PET transmission scan measurements with pre-subtracted delayed coincidences. For completeness, we review the exact log-likelihood and the previously developed approximations (Chapter 3): WLS, OP, SP and SD in the transmission imaging context. We develop maximization algorithms for SP and SD methods and present representative performance results from computer simulations and experimental transmission scans. The results show that the WLS method leads to systematic negative bias in the reconstructed attenuation maps and the OP method results in higher variance than the proposed SP and SD methods.

In addition to evaluating the attenuation maps themselves, we also investigate the propagation of noise from the reconstructed attenuation maps into emission images reconstructed with the FBP method. Interestingly, the difference in variances in the emission images with the new methods is even greater than in the attenuation maps themselves.

To corroborate the *empirical* studies described above, we also develop *analytical* approximations to the reconstructed image covariance based on the techniques developed in [34]. These covariance approximations are shown to agree well with the empirical variance computed from the experimental PET transmission scans. Using these approximations we show that the OP method always yields more noisy images compared to the SP method. Also, the analytic approximations are used to explain the negative systematic bias of the WLS method. One can use these approximations to help choose the values of important parameters such as regularization parameters, instead of computationally expensive multiple numerical simulations.

We also develop analytical approximations for the propagation of noise from attenuation maps into reconstructed emission images. To isolate the effect of transmission noise in the resultant emission image, we consider noise-free emission measurements and develop

approximations for the covariance of the emission images reconstructed using ACFs computed from noisy attenuation maps. These approximations describe the propagation of noise from attenuation maps into emission reconstruction and they can be used for guidance in determining the transmission scan durations for a given noise constraint on emission reconstruction. We also show that the predicted variances agree with the empirical results from the experimental PET transmission scans.

## 4.2 Exact Log-Likelihood

As previously described, in PET the data are precorrected for AC events by *real-time* subtraction of delayed window coincidences [53].

Let  $Y = [Y_1, \dots, Y_N]'$  denote the vector of *precorrected* transmission scan measurements, where “ $'$ ” denotes vector and matrix transpose. The precorrected measurement for the  $n$ th coincidence detector pair is:

$$Y_n = Y_n^{\text{prompt}} - Y_n^{\text{delay}}, \quad (4.1)$$

where  $Y_n^{\text{prompt}}$  and  $Y_n^{\text{delay}}$  are the number of coincidences within the prompt and delayed windows, respectively. Let  $\mu = [\mu_1, \dots, \mu_M]'$  denote the vector of unknown linear attenuation coefficients. For transmission scans, we assume that  $Y_n^{\text{prompt}}$  and  $Y_n^{\text{delay}}$  are statistically independent Poisson random variables [102] with means  $\bar{y}_n^{\text{p}}$  and  $\bar{y}_n^{\text{d}}$  respectively as:

$$\mathbb{E} \{ Y_n^{\text{prompt}} \} = \bar{y}_n^{\text{p}}(\mu) = b_n e^{-l_n(\mu)} + s_n^T + r_n \quad (4.2)$$

$$\mathbb{E} \{ Y_n^{\text{delay}} \} = \bar{y}_n^{\text{d}} = r_n, \quad (4.3)$$

where  $l_n(\mu) = \sum_{j=1}^P a_{nj} \mu_j$  is the total attenuation between  $n$ th detector pair. The  $a_{nj} \geq 0$  factors have units of length and describe the tomographic system geometry. The  $b_n > 0$  factors denote the blank scan counts and the  $r_n \geq 0$  factors denote the mean of AC events and  $s_n^T$  denote the mean of the scattered events.

Since  $Y_n^{\text{prompt}}$  and  $Y_n^{\text{delay}}$  are statistically independent and Poisson:

$$\mathbb{E} \{ Y_n \} = \bar{y}_n^{\text{p}}(\mu) - \bar{y}_n^{\text{d}} = b_n e^{-l_n(\mu)} + s_n^T \triangleq \bar{y}_n(\mu), \quad (4.4)$$

$$\text{Var} \{ Y_n \} = \bar{y}_n^{\text{p}}(\mu) + \bar{y}_n^{\text{d}} = b_n e^{-l_n(\mu)} + s_n^T + 2r_n. \quad (4.5)$$

Let  $y = [y_1, \dots, y_N]'$  be an observed realization of  $Y$  in (4.1). Since the measurements are independent, one can express the exact log-likelihood as follows [99]:

$$L(\mu) = \sum_{n=1}^N h_n(l_n(\mu), y_n), \quad (4.6)$$

where, ignoring constants independent of  $\mu$  throughout:

$$h_n(l_n(\mu), y_n) \triangleq \log \left( \sum_{m=[-y_n]_+}^{\infty} \frac{[\bar{y}_n^{\text{p}}(\mu)]^{y_n+m}}{(y_n+m)!} \frac{r_n^m}{m!} \right) - (\bar{y}_n^{\text{p}}(\mu) + r_n), \quad (4.7)$$

where  $[x]_+ = x$  if  $x > 0$  and is 0 otherwise.

Since image reconstruction is ill conditioned, we combine a roughness penalty  $R(\mu)$  with the log-likelihood to form a penalized-likelihood objective function as described in Section 2.4.2:

$$\Phi(\mu) = L(\mu) - \beta R(\mu). \quad (4.8)$$

The goal is to estimate  $\mu$  by maximizing  $\Phi(\mu)$  over the nonnegative cone:

$$\hat{\mu} = \arg \max_{\mu \geq 0} \Phi(\mu). \quad (4.9)$$

Since the exact log-likelihood function (4.7) contains infinite summations and the summations start from  $[-y_n]_+$ , next we describe tractable yet accurate approximations to the exact log-likelihood.

### 4.3 Approximations to the Exact Log-Likelihood

In this section, for completeness we briefly review the four practical approximations to  $L(\mu)$  for transmission tomography imaging (described in more detail in Chapter 3) : the WLS model, the conventional OP model, the proposed SP model approximation and lastly the proposed SD model approximation. All log-likelihood approximations have the form (4.6) for different choices for  $h_n(l, y_n)$ .

#### 4.3.1 Quadratic Approximations

A quadratic approximation to the exact log-likelihood function [32, 79] leads to the data-weighted least squares objective function  $L_{\text{WLS}}(\mu)$  of the form (4.6) with

$$h_n^{\text{WLS}}(l, y_n) = \begin{cases} -\frac{1}{2}(l - \hat{l}_n)^2 \frac{1}{\hat{\sigma}_n^2}, & y_n > 0 \\ 0, & y_n \leq 0, \end{cases} \quad (4.10)$$

where  $\hat{l}_n = \log\left(\frac{b_n}{(y_n - s_n^T)}\right)$  is the method-of-moments estimate of  $l_n(\mu)$ . The weighting factor  $\hat{\sigma}_n^2 = \frac{y_n + 2r_n}{(y_n - s_n^T)^2}$  is an estimate of the variance of  $\hat{l}_n(y_n)$  based on a second-order Taylor expansion of  $\hat{l}(\cdot)$  around  $\bar{y}_n$  (Appendix B). The residuals corresponding to projections with large values of  $y_n$  are weighted more heavily in (4.10). These rays pass through less dense objects and consequently have higher SNR values.

#### 4.3.2 Ordinary Poisson (OP) Approximation

The conventional approach is to ignore the random coincidences by assuming that  $\{Y_n\}_{n=1}^N$  are distributed as independent Poisson random variables with means  $\bar{y}_n(\mu)$  given by (4.4). The log-likelihood  $L_{\text{OP}}(\mu)$  corresponding to this OP approximation is of the form (4.6) with

$$h_n^{\text{OP}}(l, y_n) = y_n \log(b_n e^{-l} + s_n^T) - (b_n e^{-l} + s_n^T). \quad (4.11)$$

#### 4.3.3 Shifted Poisson (SP) Approximation

A better approach is to match both the first and second moments by approximating the random variables  $\{Y_n + 2r_n\}_{n=1}^N$  as having Poisson distributions with means  $\{\bar{y}_n(\mu) + 2r_n\}$ .

This idea leads to the SP approximation  $L_{\text{SP}}(\mu)$  derived in Section 3.3.3 [97–99] of the form (4.6) with

$$h_n^{\text{SP}}(l, y_n) = (y_n + 2r_n) \log(b_n e^{-l} + s_n^T + 2r_n) - (b_n e^{-l} + s_n^T + 2r_n). \quad (4.12)$$

Although both  $L_{\text{WLS}}$  and  $L_{\text{SP}}$  match two moments of the measurement distribution, in WLS the second moment of  $\hat{l}_n(y_n)$  is “fixed” to  $\hat{\sigma}_n^2$  independently of  $\mu$ , whereas in the SP model the moments vary with  $\bar{y}_n(\mu)$  appropriately.

#### 4.3.4 Saddle-point (SD) Approximation

For transmission reconstruction, the SD approximation derived in Section 3.4 which is based on second order Taylor series approximation for the exact pmf in the  $z$ -transform domain, is of the form (4.6) with

$$h_n^{\text{SD}}(l, y_n) = \begin{cases} y_n \log\left(\frac{b_n e^{-l} + s_n^T + r_n}{y_n + 1 + u_n(l)}\right) - t_n(l), & y_n \geq 0 \\ y_n \log\left(\frac{b_n e^{-l} + s_n^T + r_n}{y_n - 1 + u_n(l)}\right) - t_n(l), & y_n < 0 \end{cases} \quad (4.13)$$

where from (3.35)

$$\begin{aligned} t_n(l) &= b_n e^{-l} + u_n(l) - \frac{1}{2} \log u_n(l), \\ u_n(l) &= \sqrt{(|y_n| + 1)^2 + 4(b_n e^{-l} + s_n^T + r_n)r_n}. \end{aligned}$$

## 4.4 1-D Bias-Variance Analysis

Two useful measures of the performance of a given reconstruction method are the bias:

$$\begin{aligned} \text{Bias}\{\hat{\mu}\} &= \text{E}\{\hat{\mu} - \mu^{\text{true}}\} \\ &= \text{E}\{\hat{\mu}\} - \mu^{\text{true}} \end{aligned} \quad (4.14)$$

and the variance:

$$\begin{aligned} \text{Var}\{\hat{\mu}\} &= \text{E}\{\hat{\mu} - \text{E}\{\hat{\mu}\}\}^2 \\ &= \text{E}\{\hat{\mu}^2\} - [\text{E}\{\hat{\mu}\}]^2. \end{aligned} \quad (4.15)$$

Evaluations of the expectations to determine these quantities is difficult without an explicit expression for the estimator as a function of measured data  $[y_1, \dots, y_N]$ . Unfortunately, for the tomographic problem there are no closed form expressions for the estimators. To analyze the bias and variance of each estimator (WLS, OP, SP and SD) analytically, we used the analytic approximations suggested by Fessler for tomographic imaging [34]. Assuming that the objective function  $\Phi(\theta, Y)$  has a unique global maximum  $\hat{\theta}$  for any measurement  $Y$  and that the maximum can be found by zeroing the partial derivatives of  $\Phi(\theta, Y)$ , then there exists an implicit function  $f(Y) = [f_1(Y) \dots f_P(Y)] = \hat{\theta}$  that maps the measurement  $Y$  into an estimate  $\hat{\theta}$ . The implicitly defined function  $f(Y)$  can rarely be found analytically and one usually implements an iterative method for maximizing  $\Phi(\theta, Y)$ . The absence of an explicit analytic expression of the form  $\hat{\theta} = f(Y)$  makes it difficult to study the mean

and the variance of the estimator  $\hat{\theta}$ , so often one needs to perform numerical simulations. In the following, we obtain approximate analytic expressions for the mean and variance of different estimators.

If  $\bar{Y}_n$  denotes the mean of measurement  $Y_n$ , then the first and second-order Taylor expansion of  $f(Y)$  around  $\bar{Y}$  results in the approximation of covariance and mean of  $\hat{\theta}$  respectively as:

$$\begin{aligned}\text{Cov}\{\hat{\theta}\} &\approx [-\nabla^{20}\Phi(\check{\theta}, \bar{Y})]^{-1} \nabla^{11}\Phi(\check{\theta}, \bar{Y}) \text{Cov}(Y) [\nabla^{11}\Phi(\check{\theta}, \bar{Y})]^T [-\nabla^{20}\Phi(\check{\theta}, \bar{Y})]^{-1} \\ \text{E}\{\hat{\theta}\} &\approx f(\bar{Y}) + \frac{1}{2} \sum_n \sum_m \frac{\partial^2}{\partial Y_n \partial Y_m} f(\bar{Y}) \text{Cov}(Y_n Y_m)\end{aligned}$$

where  $\check{\theta} = f(\bar{Y})$ . (The first and second order partial derivatives of  $f(Y)$  at  $\bar{Y}$  were determined in [34] by applying the chain rule.)

Using above equations one can find approximate expressions for the variance and the mean of the estimators:  $\hat{\theta}_{WLS} = \arg \max_{\theta} L_{WLS}(\theta)$ ,  $\hat{\theta}_{OP} = \arg \max_{\theta} L_{OP}(\theta)$ ,  $\hat{\theta}_{SP} = \arg \max_{\theta} L_{SP}(\theta)$  and  $\hat{\theta}_{SD} = \arg \max_{\theta} L_{SD}(\theta)$ . For this purpose we considered a highly simplified version of transmission tomography where the unknown is a scalar parameter, *i.e.*:  $p = 1$ . This simplified problem provides insight into the estimator bias and variance without the undue notation of the multi-parameter case. Since the measurements are statistically independent, for the scalar transmission problem the above approximations reduce to:

$$\text{Var}\{\hat{\mu}\} \approx \sum_n \left[ \frac{\partial}{\partial Y_n} f(\bar{Y}) \right]^2 \text{Var}(Y_n) \quad (4.16)$$

$$\text{E}\{\hat{\mu}\} \approx f(\bar{Y}) + \frac{1}{2} \sum_n \frac{\partial^2}{\partial Y_n^2} f(\bar{Y}) \text{Var}(Y_n) \quad (4.17)$$

The approximate expressions for bias and variance of the above estimators are derived in Appendix C:

$$\text{Var}\{\hat{\mu}_{WLS}\} \approx \left[ \sum_{n=1}^N \frac{a_n^2 \bar{y}_n(\mu^{\text{true}})^2}{(\bar{y}_n(\mu^{\text{true}}) + 2r_n)} \right]^{-1} \quad (4.18)$$

$$\text{E}\{\hat{\mu}_{WLS}\} \approx \mu^{\text{true}} + \frac{\sum_{n=1}^N a_n^3 (\bar{y}_n(\mu^{\text{true}}) + 2r_n)}{\left( \sum_{n=1}^N a_n^2 \bar{y}_n(\mu^{\text{true}}) \right)^2} - \frac{1}{2} \frac{\sum_{n=1}^N a_n \left( \frac{\bar{y}_n(\mu^{\text{true}}) + 2r_n}{\bar{y}_n(\mu^{\text{true}})} \right)}{\sum_{n=1}^N a_n^2 \bar{y}_n(\mu^{\text{true}})} \quad (4.19)$$

$$\text{Var}\{\hat{\mu}_{OP}\} \approx \frac{\sum_{n=1}^N a_n^2 (\bar{y}_n(\mu^{\text{true}}) + 2r_n)}{\left( \sum_{n=1}^N a_n^2 \bar{y}_n(\mu^{\text{true}}) \right)^2} \quad (4.20)$$

$$\text{E}\{\hat{\mu}_{OP}\} \approx \mu^{\text{true}} + \frac{1}{2} \text{Var}\{\hat{\mu}_{OP}\} \frac{\sum_{n=1}^N a_n^3 \bar{y}_n(\mu^{\text{true}})}{\sum_{n=1}^N a_n^2 \bar{y}_n(\mu^{\text{true}})} \quad (4.21)$$

$$\text{Var}\{\hat{\mu}_{SP}\} \approx \text{Var}\{\hat{\mu}_{WLS}\} \approx \left[ \sum_{n=1}^N \frac{a_n^2 \bar{y}_n(\mu^{\text{true}})^2}{(\bar{y}_n(\mu^{\text{true}}) + 2r_n)} \right]^{-1} \quad (4.22)$$

$$\text{E}\{\hat{\mu}_{SP}\} \approx \mu^{\text{true}} + \frac{1}{2} \text{Var}\{\hat{\mu}_{SP}\} \frac{\sum_{n=1}^N a_n^3 \frac{\bar{y}_n(\mu^{\text{true}})^2}{\bar{y}_n(\mu^{\text{true}}) + 2r_n}}{\sum_{n=1}^N a_n^2 \frac{\bar{y}_n(\mu^{\text{true}})^2}{\bar{y}_n(\mu^{\text{true}}) + 2r_n}} \quad (4.23)$$

where  $\mu^{\text{true}}$  denotes the true value of the attenuation coefficient and  $\bar{y}_n(\mu)$  is the mean value of pre-corrected measurements as defined in (4.4). As described in Appendix C,  $\text{E}\{\hat{\mu}_{WLS}\}$

is based on a simplified WLS objective function  $L_{\text{WLS}}(\mu)$  with the weighting factor  $\hat{\sigma}_n^2 = \frac{1}{\bar{y}_n}$ . The analytic approximations for SD objective function are quite complicated algebraically (C.7, C.8, C.9) and they are presented in Appendix C.

If one expresses the time dependence of the mean of randoms precorrected data explicitly as:

$$\bar{y}_n(\mu) = T \left( \bar{b}_n e^{-a_n \mu} + s_n^T \right),$$

where  $T$  is the total transmission scan duration and  $\bar{b}_n$  denotes the blank scan count rate per time, then one can see from the above analytic approximations that as  $T \rightarrow \infty$ ,  $E\{\hat{\mu}\} \rightarrow \mu^{\text{true}}$  and  $\text{Var}\{\hat{\mu}\} \rightarrow 0$  for all the models.

Letting  $o_n = a_n^2 \bar{y}_n(\mu^{\text{true}})$  and  $t_n = a_n^2 (\bar{y}_n(\mu^{\text{true}}) + 2r_n)$ , one can rewrite (4.20) and (4.22) as:

$$\frac{1}{\text{Var}\{\hat{\mu}_{OP}\}} \approx \frac{(\sum_n o_n)^2}{\sum_n t_n}, \quad \frac{1}{\text{Var}\{\hat{\mu}_{SP}\}} \approx \sum_n \frac{o_n^2}{t_n}$$

Let  $\underline{u}, \underline{v} \in \mathbb{R}^n$  such that  $u_n = \frac{o_n}{\sqrt{t_n}}$ ,  $v_n = \sqrt{t_n}$ . Using Cauchy-Schwarz inequality:  $|\underline{u}^T \underline{v}| \leq \|\underline{u}\|_2 \|\underline{v}\|_2$ ,

$$\begin{aligned} \sum_n o_n &\leq \left( \sum_n \frac{o_n^2}{t_n} \right)^{\frac{1}{2}} \left( \sum_n t_n \right)^{\frac{1}{2}} \\ &\left( \sum_n \frac{o_n^2}{t_n} \right)^{-1} \leq \frac{\sum_n t_n}{(\sum_n o_n)^2}, \end{aligned}$$

so that to within the accuracy of (4.16):

$$\text{Var}\{\hat{\mu}_{SP}\} \leq \text{Var}\{\hat{\mu}_{OP}\}, \quad (4.24)$$

with equality if and only if  $r_n/\bar{y}_n$  ratios are equal. For PET systems, these ratio terms are never constant, and in fact can be quite disparate. Thus we have shown the following result: the variance of the SP estimator will always be lower than the variance of the OP estimator.

## 4.5 1-D Simulations

Fig. 4.1 shows plots of the bias and variance terms (4.18)-(4.23) and simulation results, as a function of mean counts per detector,  $(1/N) \sum_{n=1}^N \bar{y}_n(\mu^{\text{true}})$ , for the 1D problem described above. In these simulations:  $\mu^{\text{true}} = 1$ ,  $N = 20$ , the  $a_n$  factors are uniformly distributed between 0 to 4, the  $r_n$  factors correspond to 10% uniform field of random coincidences, the  $b_n$  factors were generated using pseudo-random log-normal variates with a standard deviation of 0.3 to account for the detector efficiency variations. The solid lines denote the analytical predictions, whereas the symbols denote empirical results from 500 realizations. The agreement between the analytic approximations and simulation (especially for high count rates) shows that one can use formulas (4.18)-(4.23) to compare analytically the bias and variance introduced by different estimators.

From Figure 4.1, one can observe that WLS has a systematic negative bias. The source of this significant negative bias of the WLS method can be understood from the expression

of approximate WLS estimate  $\tilde{\mu}_{WLS}$  (C.6) derived in Appendix C which is rewritten here for convenience:

$$\tilde{\mu}_{WLS} = \tilde{f}_{WLS}(y) = \frac{\sum_{n=1, y_n > s_n^T}^N a_n^2 y_n \left( \frac{1}{a_n} \log \frac{b_n}{y_n - s_n^T} \right)}{\sum_{n=1, y_n > s_n^T}^N a_n^2 y_n}. \quad (4.25)$$

The expression for  $\tilde{\mu}_{WLS}$  is seen to be a weighted sum, weighted by the noisy measurements  $y_n$  (as also pointed out by Fessler [32] for the case of  $a_n = 1$ ). When  $y_n$  is larger than  $\bar{y}_n(\mu)$ , the  $\left( \frac{1}{a_n} \log \frac{b_n}{y_n - s_n^T} \right)$  term becomes smaller than  $\mu^{\text{true}}$ , however the corresponding weight in the summation will be larger. Thus, during the summation the *under-estimated* terms are weighted more heavily, resulting in a negative bias for  $\hat{\mu}_{WLS}$ .

Figure 4.1 shows that both the OP model and the SP model yield nearly unbiased estimates, while SP model yields a smaller bias. In terms of standard deviation, we see that the OP model leads to higher standard deviation than both the WLS and the SP models, in agreement with the above inequality (4.24).

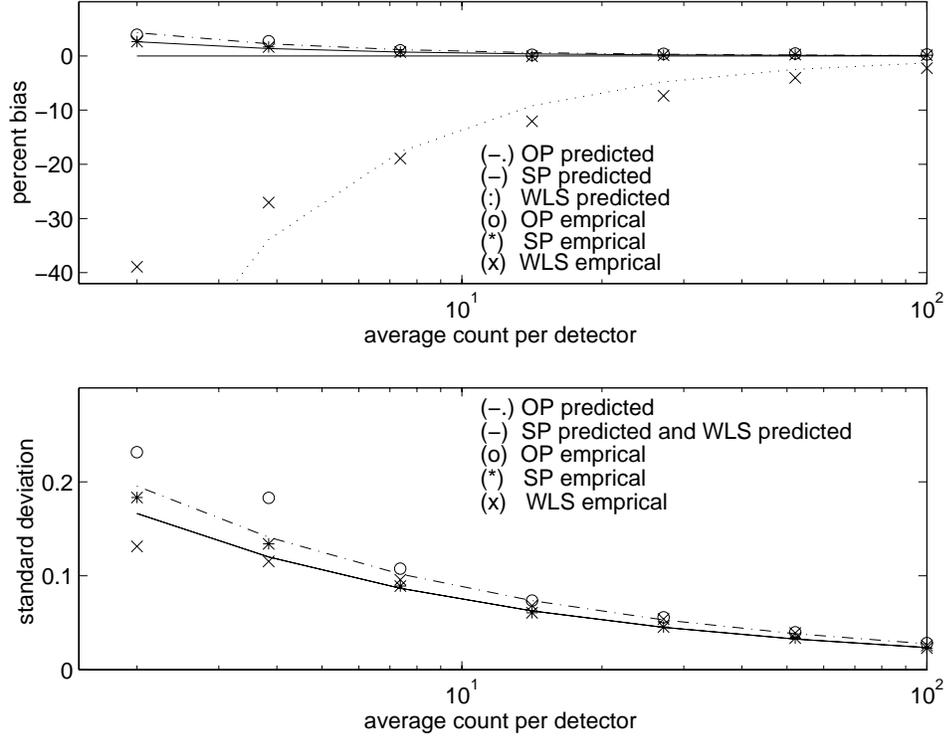


Figure 4.1: Comparison of analytical approximations and empirical results for bias and variance. Upper figure shows that WLS estimator is systematically negatively biased especially for low counts. Lower figure shows that ordinary Poisson model yields higher standard deviation than both other estimators.

## 4.6 Concavity and Convergence

The second partial derivatives of the OP (4.11) and the SP (4.12) model approximations and the PR log-likelihood (3.38) can be written as:

$$-\frac{\partial^2}{\partial \mu_j \partial \mu_k} L(\mu) = \sum_{n=1}^N a_{nj} a_{nk} \left[ 1 - \frac{x_n (s_n^T + d_n)}{(\bar{y}_n(\mu) + d_n)^2} \right] b_n e^{-l_n(\mu)}, \quad (4.26)$$

with

$$d_n \triangleq \begin{cases} 0, & \text{OP} \\ 2r_n, & \text{SP} \\ r_n, & \text{PR} \end{cases} \quad (4.27)$$

and

$$x_n \triangleq \begin{cases} y_n, & \text{OP} \\ y_n + 2r_n, & \text{SP} \\ y_n^p, & \text{PR.} \end{cases} \quad (4.28)$$

Although OP is globally concave, the SP and PR objective functions are only locally concave over the set:

$$\{\mu : (\bar{y}_n(\mu) + d_n)^2 \geq x_n (s_n^T + d_n), \forall n\}. \quad (4.29)$$

Thus it is difficult to establish global convergence for any maximization algorithm for the SP and the PR models [32]. Luckily, in PET and SPECT the  $r_n$  and  $s_n^T$  values are fairly small compared to  $y_n$  values and one usually operates in or near the concave region of log-likelihood (4.29). The recently developed paraboloid surrogates algorithm of Erdoğan and Fessler [28] is particularly attractive because it guarantees monotonicity even for nonzero  $r_n$  factors.

For the exact log-likelihood (EX) the complicated form of the expressions (4.7, 3.14) make it difficult to perform a rigorous concavity analysis. For the SD model we perform a concavity analysis for the emission case in Section E, however for the transmission case the concavity analysis proves to be algebraically tedious because of complicated form of the partial derivatives (Appendix C).

Previous convergence proofs for transmission algorithms depended heavily on the assumption of  $r_n = 0$  *i.e.*: OP model [61, 62]. Further investigation is necessary for the convergence properties of transmission reconstruction algorithms for randoms precorrected data and for the PR model with nonzero  $r_n$ 's. In practice we initialize the iteration with FBP image and always observe monotonic increase in the log-likelihood.

## 4.7 Log-likelihood Maximization : Coordinate Ascent Type Algorithms

In coordinate-ascent (CA) type iterative maximization one simply updates the parameters only one at a time always using the most recent value of each parameter at the next iteration. Sequential CA methods were shown to converge rapidly in transmission tomography [32, 79]. In addition to fast convergence, CA methods enable decoupling of parameter updates and one can easily enforce non-negativity constraints.

However, in transmission tomography there is no closed form maximizer of  $\Phi(\mu)$  over  $\mu_j$  (even for OP, SP and PR models) even with all the parameters  $\{\mu_i : i \neq j\}$  are fixed. For the quadratic penalty one can use the 1-D Newton's method to sequentially update all the parameters as [32]:

$$\mu_j^{new} = \left[ \mu_j^{old} + \omega \frac{\frac{\partial}{\partial \mu_j} \Phi(\mu^{old})}{-\frac{\partial}{\partial \mu_j^2} \Phi(\mu^{old})} \right]_+ \quad (4.30)$$

where  $\omega \in (0, 1]$  is relaxation parameter, and  $[x]_+ = x$  if  $x > 0$  and is 0 otherwise.

Although the CA method converges rapidly, it is also computationally expensive for transmission tomography, since one needs to compute  $K$  exponentials<sup>1</sup> during each iteration where  $K$  is the number of nonzero  $a_{ij}$ 's.

A grouped coordinate ascent (GCA) algorithm was suggested [39] as an alternative to balance the convergence rate and computation per iteration. This method updates pixels in groups, which reduces the number of operations per iteration. On the other hand, by choosing pixels in each group well separated spatially, the algorithm does not suffer from slow convergence. For a subset of pixels  $S = \{1, \dots, p\}$ , the GCA algorithm monotonically increases the objective function at the  $i$ th iteration by finding  $\mu_S^{i+1}$  such as:

$$\Phi(\mu_S^{i+1}, \mu_{\tilde{S}}^i) \geq \Phi(\mu_S^i, \mu_{\tilde{S}}^i) = \Phi(\mu^i), \quad (4.31)$$

where  $\tilde{S}$  is the compliment of  $S$ . To achieve this purpose, GCA method uses the separable surrogate function  $\phi(\mu_S; \mu_{\tilde{S}}^i)$  that satisfies :

$$\Phi(\mu_S, \mu_{\tilde{S}}^i) - \Phi(\mu^i) \geq \phi(\mu_S; \mu^i) - \phi(\mu_{\tilde{S}}^i; \mu^i). \quad (4.32)$$

Fessler *et al.* [39] developed the following additively separable surrogate function using a generalization of De Pierro's transfer idea [20, 21]:

$$\phi(\mu_S; \mu^i) = \sum_{j \in S} \phi_j(\mu_j; \mu^i), \quad (4.33)$$

with

$$\phi_j(\mu_j; \mu^i) = \sum_n \alpha_{nj} h_n \left( \frac{a_{nj}}{\alpha_{nj}} (\mu_j - \mu_j^i) + \sum_k a_{nk} \mu_k^i \right) - \beta R(\mu) \quad (4.34)$$

and

$$\alpha_{nj} = a_{nj} / \sum_{k \in S} a_{nk} . \quad (4.35)$$

Since,  $\phi(\mu_S; \mu^i)$  is additively decoupled (*i.e.*: each  $\phi_j$  in (4.33) depends on one  $\mu_j$  only), one can use a couple iterations of the 1-D Newton's method similar to (4.30) for maximization of each subpixel group  $S$ . GCA with subgroups of few pixel ( $p \approx 3$ ) were shown to converge very rapidly in terms of CPU time [39]. For this class of algorithms one needs to evaluate first and second order derivatives of the log-likelihood at each iteration (second derivative can also be approximated [39] for speeding up the algorithm). Both of the

---

<sup>1</sup>According to (4.30) the update for  $\mu_j$  requires the computation of  $e^{-a_{nj} \mu_j^{new}}$  for each ray  $n \in \{0, 1, \dots, N\}$ . And one needs update each pixel  $j \in \{0, 1, \dots, P\}$  to complete one iteration.

proposed methods (SP and SD) have closed form expressions for the derivatives of the log-likelihoods (see Appendix C and (3.36), (3.37)), which enables one to easily modify the GCA type maximization algorithm. We use the fast GCA method [39] in our 2-D simulations and the experimental studies. In the SP method, the additional computation is negligibly small compared to the the OP method. The SD algorithm was observed to require around 20% more CPU time. However it should be mentioned that no effort was taken to optimize the algorithm for the SD method.

The recently developed monotonic paraboloid surrogates CA algorithm by Erdoğan and Fessler [28], which uses the optimum curvature for the surrogate functions at each iteration, converges even faster than the GCA method. We use this method [38] for the maximization of the objective functions in our 2-D *emission* reconstruction studies (in Chapter 5).

## 4.8 2-D Simulations

To study bias and variance properties of the estimators based on the described approximations, we performed 2-D simulations. Next we describe these simulations and the quantitative results.

### 4.8.1 Simulations

In the 2-D simulation, for  $\mu$  we used the synthetic attenuation map shown in Fig. 4.2, which represents a human *abdomen* with linear attenuation coefficient 0.0096/mm. The image was a 128 by 128 array of 4.7 mm pixels. We simulated a PET transmission scan with 192 radial bins and 256 angles uniformly spaced over 180 degrees. The  $a_{nj}$  factors correspond to 3.1 mm wide strip integrals on 3.1 mm center-to-center spacing. The  $b_n$  factors were generated using pseudo-random log-normal variates with standard deviation of 0.3 to simulate detector pairs with nonuniform detector efficiencies and scaled so that  $\sum_n \bar{y}_n$  was 3.6 million counts. The  $r_n$  factors corresponded to a uniform field of 10% random coincidences. Pseudo-random transmission measurements were generated according to (4.2) and (4.3). For regularization, we used the *modified* quadratic penalty [43] introduced in Section 2.4.2. This penalty improves the resolution uniformity and enables us to match the spatial resolution of different methods.



Figure 4.2: Simulated abdomen attenuation map.

We generated 150 independent realizations of the transmission measurements. For each measurement realization, an estimate of the attenuation map was reconstructed using 20 iterations of the grouped-coordinate ascent algorithms [39] (Section 4.7) applied to the WLS

(4.10), OP (4.11), SP (4.12), SD (4.13), EX (4.7) and PR (3.38) objective functions. Since we have closed form expressions for all the objective functions (and their partial derivatives), we were able to modify the GCA method (4.7) for the maximization of each objective function. For the exact log-likelihood (EX) we performed a very precise implementation by truncating the infinite summations. Although this method is not practical in terms of its computational requirements, it still serves for the purpose of evaluating the performance of the exact log-likelihood method. In our simulations, we initialized the iterations with a FBP image and always observed monotonic increase in the log-likelihood for all methods.

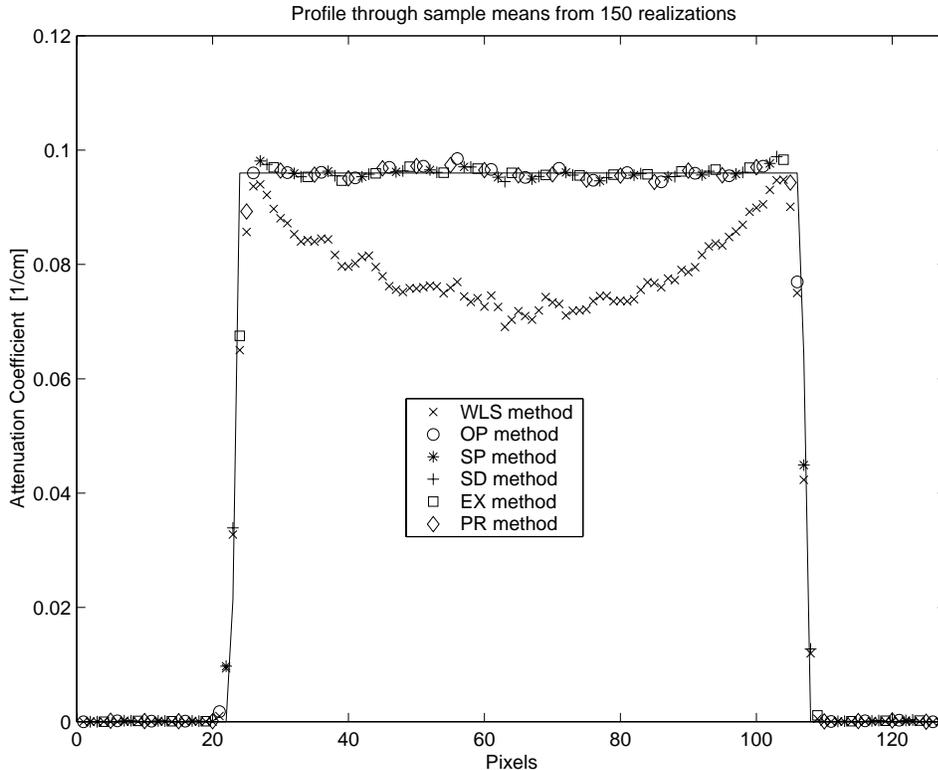


Figure 4.3: Horizontal profile through the sample mean images for abdomen phantom. The WLS method has a systematic negative bias. However, the ordinary Poisson (OP), shifted Poisson (SP), saddle-point (SD), exact (EX) and prompt (PR) methods are free of this systematic negative bias.

We computed both the sample mean and sample standard deviation images for all methods. Fig. 4.3 shows horizontal profiles through the sample mean images. These profiles show that WLS is *systematically* negatively biased [32], whereas all the other methods are free of systematic bias.

To study the variance, we computed the *ratio* of the sample standard deviation images of different estimators, over all the interior pixels. Fig. 4.4 shows the histogram of the standard deviation ratios. The OP model yields, on the average, 15% higher standard deviation than the SP, SD and EX models. In other words, to achieve the same noise level, the OP method would require about 32% greater scan time. Also, the OP model yields, on the average 39% higher standard deviation than the PR model. It should be mentioned that in these simulations PR method is the idealized method where we assume that one has

Estimator	FWHM (pixels)			% Std. Dev.
	horizontal	vertical	average	
FBP	2.66	2.68	2.67	18.20 $\pm$ 1.05
OP	2.13	3.22	2.67	9.94 $\pm$ 0.57
SP	1.94	3.40	2.67	7.70 $\pm$ 0.44
SD	1.93	3.41	2.67	7.94 $\pm$ 0.45

Table 4.1: Local impulse response and the local sample standard deviation for the central pixel.

access to the means of random rates (*i.e.*:  $r_n$ ), but in practice one needs to estimate these quantities from noisy measurements of delayed windows. Thus, the results reported here with PR method shows the upper bounds on the performance of the PR method.

Although the standard deviation values could be decreased by using higher count rates, the ratio of standard deviations of different estimators will remain approximately same for higher count rates [34]. This follows from the fact that analytic approximations (4.20)-(4.22) will be more accurate with increasing count rates, and these approximations show that for a set of fixed system parameters, the ratio of standard deviation of different estimators remains constant independent of the count rate.

We also performed additional simulations using a digital thorax phantom (shown in Figure 4.5) with nonuniform attenuation. The reductions in noise with the proposed methods were comparable [97].

These results show that using randoms pre-corrected data instead of prompt data increases the noise in the reconstructed images. However, if one is using the randoms pre-corrected data (as currently done in most PET centers) then both SP and SD methods perform very close to the exact log-likelihood (EX) and both of them result in less noise than OP method. We will show a more detailed comparison between SP,SD and EX methods in Section 4.8.3.

#### 4.8.2 Resolution vs Standard Deviation

It is well known in tomographic image reconstruction that one can compromise between resolution and noise in reconstructed images. In the simulations reported here, we have used the modified quadratic penalty [43], which matches the spatial resolution of both least squares based and Poisson based estimators. To show that the noise reduction with the proposed SP and SD methods does not come with the price of lower resolution (compared to the OP method), we have investigated the local resolution and standard deviation of a pixel at the center of the abdomen phantom.

We computed the linearized local impulse response [43] of different estimators at the central pixel of the abdomen phantom. Table 4.1 shows the full width half maximum (FWHM) values of local impulse response functions and the local sample standard deviation for the central pixel estimates. The table also reports the standard errors for the sample standard deviation estimates. These results show that the reductions in the standard deviations are truly due to the improved statistical modeling rather than resolution differences.

Although the local impulse response functions are asymmetric with respect to the horizontal and vertical axis, the “average” resolution of each method is matched. As expected the non-statistical FBP method yields much higher standard deviation than statistical methods. The standard deviations of the proposed SP and SD estimators are about 27% lower

than the OP method.

The asymmetry of the local impulse responses is caused partly by the eccentricity of the abdomen phantom in Figure 4.2, [43]. In Table 4.1 the resolution of the SP and SD models are observed to be more asymmetric than the OP model. In order to investigate this effect we performed additional simulations using a circularly symmetric disk phantom which yields a symmetric impulse response at the center. For the central pixel (where all methods have the same impulse response) the reductions in standard deviation with the proposed SP and SD methods were around 24% compared to OP method. Recently Stayman and Fessler have developed an improved modified penalty which not only results in *uniform* resolution but also *symmetric* impulse response [87, 88]. We repeated our simulations with this improved penalty and observed very similar noise improvements with both of the SP and SD methods compared to the OP method as reported in Section 4.8.1.

### 4.8.3 Comparison of SP and SD Models with Exact Model

We observed very close agreement between the exact log-likelihood method (EX) and the SD approximation both from log-likelihood plots (Figure 3.4) and 1-D simulations. Therefore we were expecting the SD method to perform better than the SP method. However, for the 2-D simulations reported here (Section 4.8) we did not observe any statistically significant difference between the SD and the SP methods. To make a more detailed analysis of the performance of these methods, we compared the results of the reconstructed images from each noisy realization. Figure 4.6 shows a profile through the reconstruction of the EX method from simulated transmission data of 3.6 million counts as described in Section 4.8. The figure also displays the difference between the EX method and the SP and SD methods. The difference between the SD method and the EX method is virtually zero, while there is some noticeable difference between the SP method and the EX method.

To make a more quantitative comparison we computed the normalized  $E_1$ ,  $E_2$  and  $E_\infty$  norms of the differences between the EX method and the SP and SD methods for all interior pixels in the reconstructed image as:

$$E_1 = \frac{1}{N} \sum_{j: \mu_j \in W} \frac{|\hat{\mu}_j^{\text{method}} - \hat{\mu}_j^{\text{EX}}|}{\mu_j^{\text{true}}} \quad (4.36)$$

$$E_2 = \frac{1}{N} \sqrt{\sum_{j: \mu_j \in W} \frac{|\hat{\mu}_j^{\text{method}} - \hat{\mu}_j^{\text{EX}}|^2}{(\mu_j^{\text{true}})^2}} \quad (4.37)$$

$$E_\infty = \max_{j: \mu_j \in W} \left\{ \frac{|\hat{\mu}_j^{\text{method}} - \hat{\mu}_j^{\text{EX}}|}{\mu_j^{\text{true}}} \right\} \quad (4.38)$$

with  $W$  representing all the interior pixels and  $\hat{\mu}_j^{\text{method}}$  being either  $\hat{\mu}_j^{\text{SP}}$  or  $\hat{\mu}_j^{\text{SD}}$ .

Figures 4.7 - 4.9 show the  $E_1$ ,  $E_2$  and  $E_\infty$  error norms of the SP and the SD methods compared to the exact log-likelihood (EX) method as a function of noisy data realization (with 3.6 million counts as described in Section 4.8). For all the error norms the SP method results in 40 to 80 times more error than the SD method compared to the EX method.

Thus, it can be said that for each noisy realization the SD method is performing very close to the exact log-likelihood as compared to the SP method. However, for the 2-D simulations reported here this difference does not seem to make significance in the final

ensemble statistics and the SP method performs as well as the SD and the EX methods. Thus the SP method is particularly attractive since it requires comparable computation to the OP method but has reduced variance.

#### 4.8.4 Zero-thresholding the Data

Real-time subtraction of the delayed coincidence events from prompt events can lead to some negative values in the precorrected data. Since the mean of precorrected measurements is nonnegative, a natural choice might be to threshold the negative values in the precorrected data to zero before applying the maximization algorithm. The modified form of the objective functions (4.11),(4.12) for the zero thresholded data are:

$$h_n^{\text{OP}^+}(l, y_n) = [y_n]_+ \log(b_n e^{-l} + s_n^T) - (b_n e^{-l} + s_n^T), \quad (4.39)$$

$$h_n^{\text{SP}^+}(l, y_n) = [y_n + 2r_n]_+ \log(b_n e^{-l} + s_n^T + 2r_n) - (b_n e^{-l} + s_n^T + 2r_n). \quad (4.40)$$

Since the “thresholding function”  $[y_n]_+$  is not differentiable at  $y_n = 0$ , it is difficult to derive accurate analytic approximations for the mean and variance of the different estimators above. However, one can explain intuitively the overall effect of zero-thresholding as follows: setting negative precorrected data values to zero increases the mean of the precorrected data. For transmission problem the data is exponentially related to attenuation coefficients *i.e.*:  $y_n \sim b_n e^{-\sum_{j=1}^P a_{nj} \mu_j}$ , thus the increase in the mean value of the precorrected data causes the estimator to introduce a systematic *negative bias* for the estimated attenuation coefficients.

Fig. 4.10 shows plots of bias and variance terms for the 1D transmission system described in Fig. 4.1, using zero-thresholded data. The solid lines denote the formulas (4.18)-(4.23), whereas the symbols denote empirical results from 500 realizations. Fig. 4.10 shows the systematic negative bias resulting from the thresholding of the data. One can observe that while OP estimator suffers from a systematic negative bias, SP estimator is still nearly unbiased. This is due to the fact that the precorrected data is already shifted by  $2r_n$  before zero-thresholding. As a result, the number of negative values in the precorrected data are reduced dramatically. The standard deviation of the OP estimator is reduced slightly, however standard deviation of the SP estimator remains similar to non zero-thresholded case.

To study further the effects of zero-thresholding the data, we performed additional 2-D transmission simulations, using the *abdomen* phantom (Fig. 4.2) and the PET system described previously. Similar to non zero-thresholded case, we generated 150 independent realizations of the transmission measurements as mentioned previously, but this time using zero-thresholded data. We computed both the sample mean and sample standard deviation images for both the OP model and the SP estimators.

Fig. 4.11 displays horizontal profiles through the sample mean images. These profiles show that the OP estimator is negatively biased, especially for interior regions of the reconstructed image. This is due to the fact that projections through the interior regions of the object have lower count rates, and for lower count rates the OP model yields more systematic bias as can be seen from Fig. 4.10.

Fig. 4.12 shows the histogram of the ratio of sample standard deviation images of both estimators. It can be seen that the OP model still leads to higher standard deviation (on the average 11%) than the SP model. This result shows an additional advantage of the SP

model. Namely, SP estimator is not only nearly unbiased but also has a smaller standard deviation than the OP estimator, even for zero-thresholded data.

#### 4.8.5 Estimates of the AC Rates

One needs to know the mean of the AC events ( $r_n$ ) to compute  $L_{SP}(\mu)$  and  $L_{SD}(\mu)$ . Since the  $r_n$  terms are not readily available from the real (precorrected) data, some estimates of the randoms must be used.

Fig. 4.13 displays prompt and delayed coincidence sinograms for a blank scan and transmission scan. The transmission scan measurements were obtained using the phantom with the attenuation map shown in Fig. 4.14. We observe that the delayed coincidence sinograms of transmission scan and blank scan are similar. Fig. 4.15 displays the scatter plot of real delayed coincidence sinograms for blank scan and transmission scan data. Each point in the plot corresponds to a specific detector pair. The similarity of both delayed coincidence measurements suggests that one can acquire the delayed coincidence events during the blank scan and use them (after properly normalizing for different scan durations) as an estimate of the AC rates for transmission scans performed on the same PET system.

To test the robustness of the SP and SD estimators to the errors in estimates of AC rates, we performed simulations using the *abdomen* phantom and the PET system described previously. We used the average of the  $r_n$  values,  $\bar{r} = (1/N) \sum_n^N r_n$ , as an estimate of the AC event rates in the objective functions  $L_{SP}(\mu)$  and  $L_{SD}(\mu)$ . Similar to the previous simulations, we generated 150 independent realizations of the transmission measurements and then computed the sample mean and sample standard deviation images for the SP and SD estimators.

Fig. 4.16 displays horizontal profiles through the sample mean images. This profile (obtained by using constant AC rates) is observed to be unbiased just as in Figure 4.3 which was obtained using true AC rates. Thus, we conclude that this constant AC rates approximation does not introduce any systematic bias to the estimators.

Lastly, Fig. 4.17 shows the histogram of the ratio of the sample standard deviations of the SP and SD estimators with true AC rates and with constant AC rates approximation. It can be seen that using the constant AC rates approximation only slightly (around 1% – 2%) increases the standard deviation of the estimators. The resulting standard deviations are still much less than the OP model estimator. These results demonstrate that both the SP and SD approximations are robust to the errors in the  $r_n$  estimates.

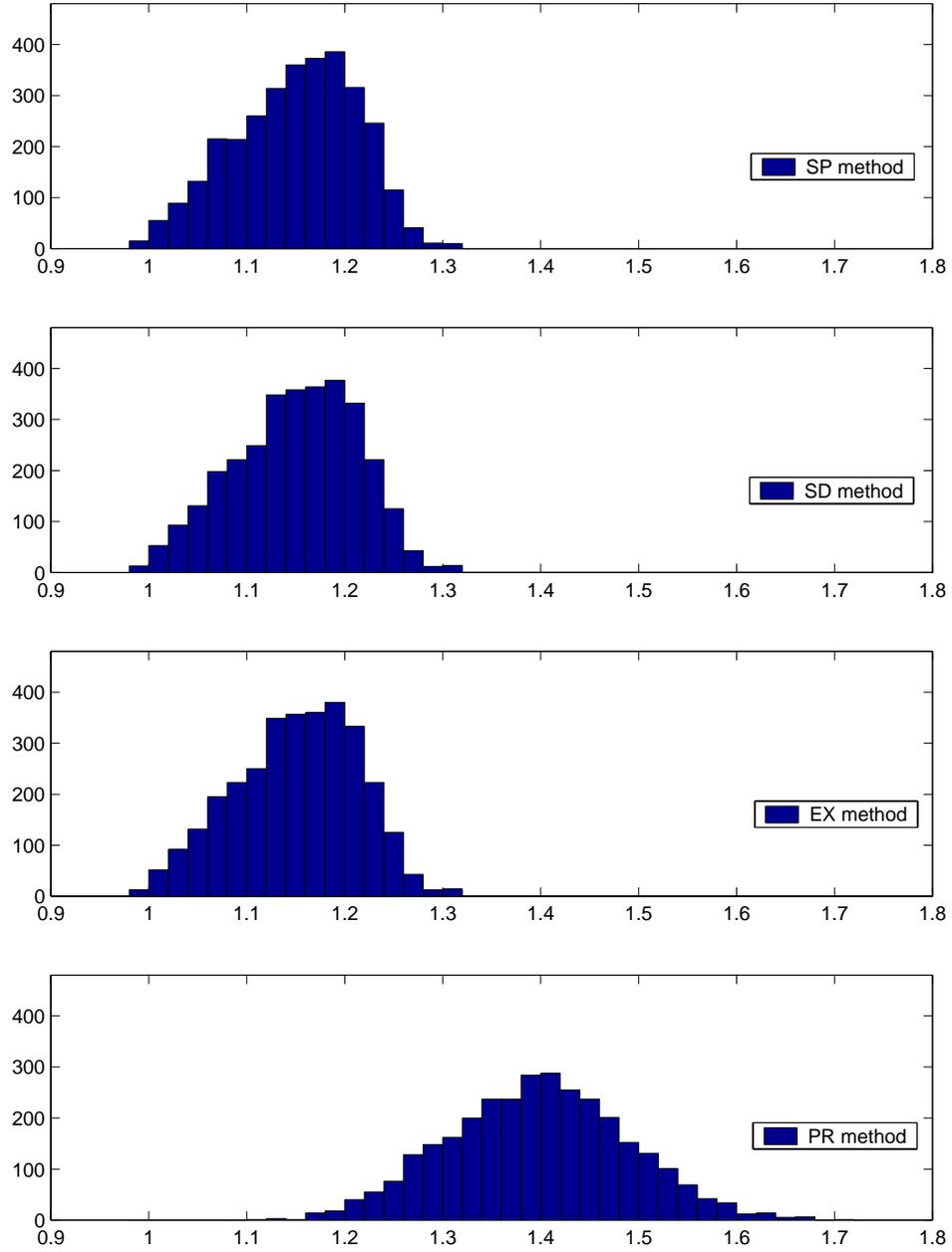


Figure 4.4: Histogram of the ratio of standard deviations of different methods over the OP method in reconstructions of the abdomen phantom. The ordinary Poisson (OP) method yields, on the average, 15% higher standard deviation than the shifted Poisson (SP), saddle-point (SD) and exact (EX) methods, and 39% more standard deviation than the prompt (PR) method.

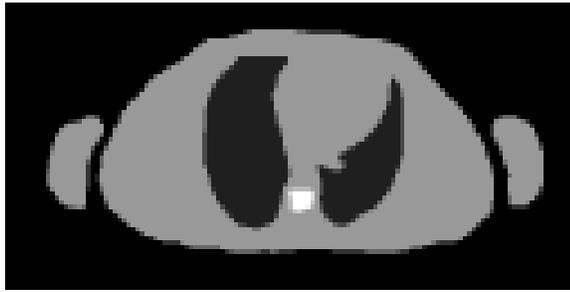


Figure 4.5: Simulated thorax attenuation map.

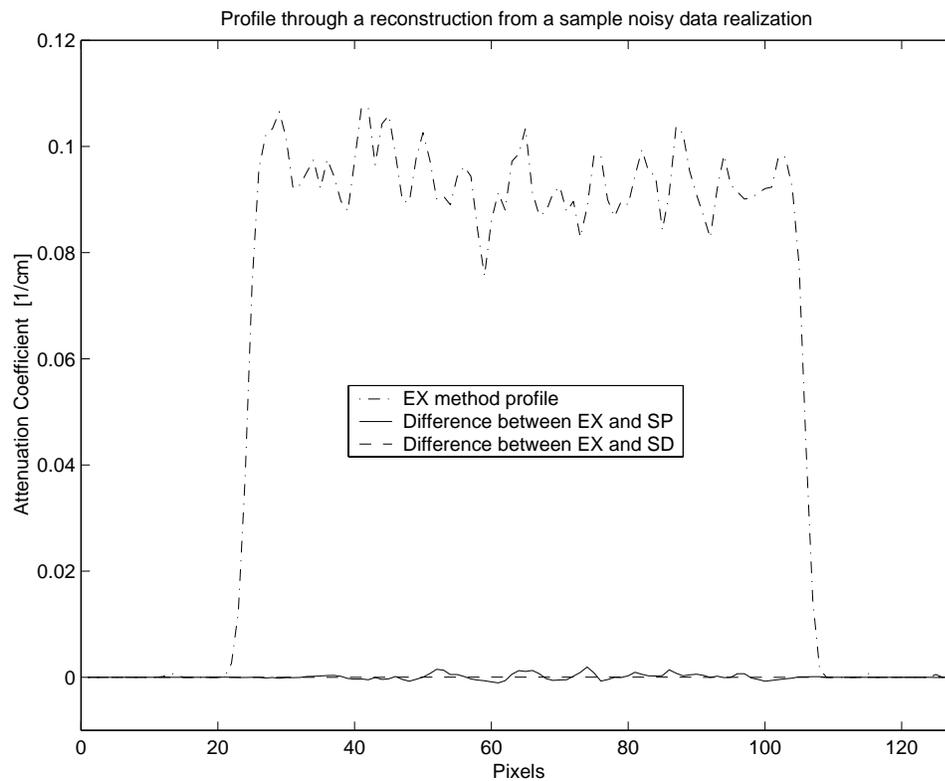


Figure 4.6: Profile through the reconstructed image of the exact log-likelihood (EX) method using 3.6 million counts transmission scan. Profiles near zero attenuation level correspond to the difference of the profiles between the EX method and the SP and the SD methods. It can be seen that there is some noticeable difference between the reconstructions with SP and EX method.

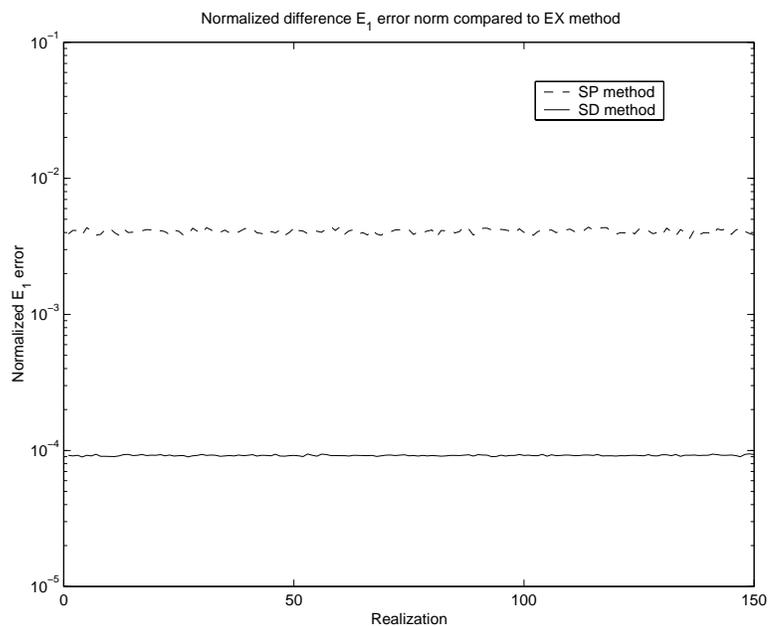


Figure 4.7:  $E_1$  error norm between the exact log-likelihood (EX) method and the SP and SD methods for each noisy realization.

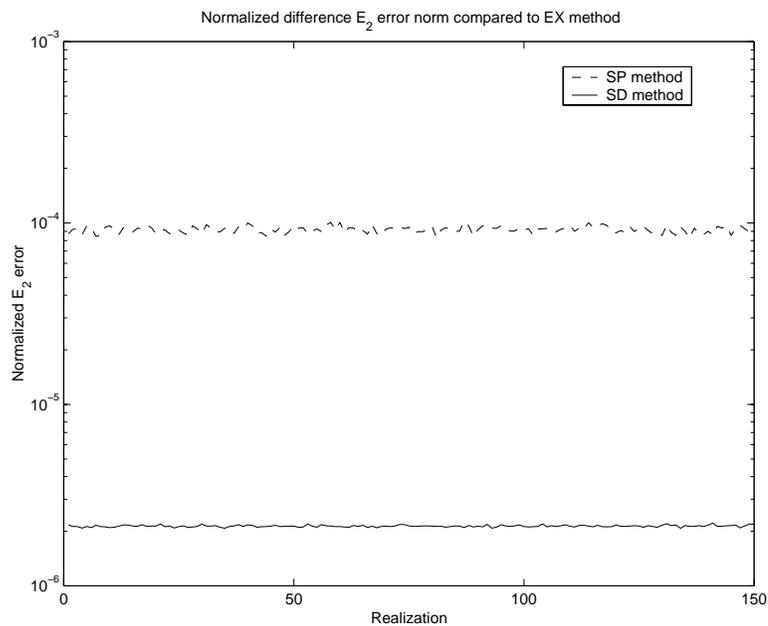


Figure 4.8:  $E_2$  error norm between the exact log-likelihood (EX) method and the SP and SD methods for each noisy realization.

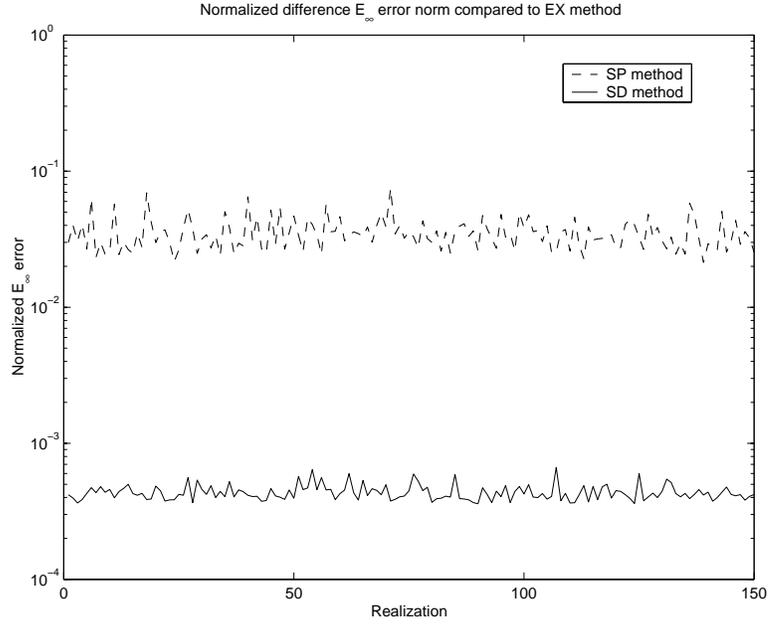


Figure 4.9:  $E_\infty$  error norm between the exact log-likelihood (EX) method and the SP and SD methods for each noisy realization.

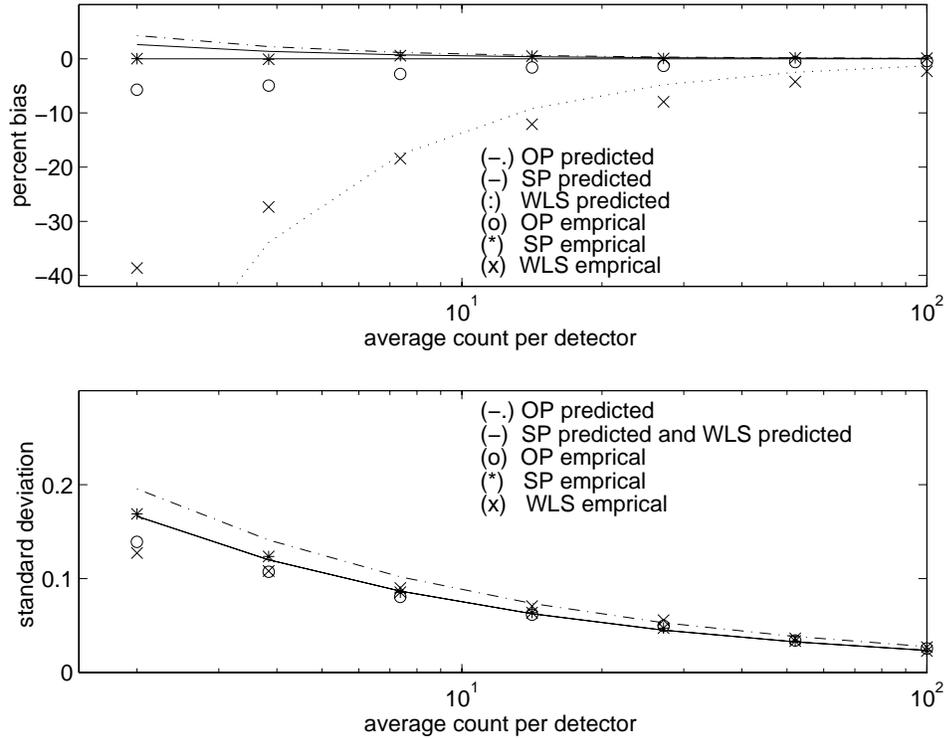


Figure 4.10: Comparison of analytical approximations and empirical results for “zero-thresholded” data. Upper figure shows that ordinary Poisson model is negatively biased compared to Fig. 4.1, due to thresholding.

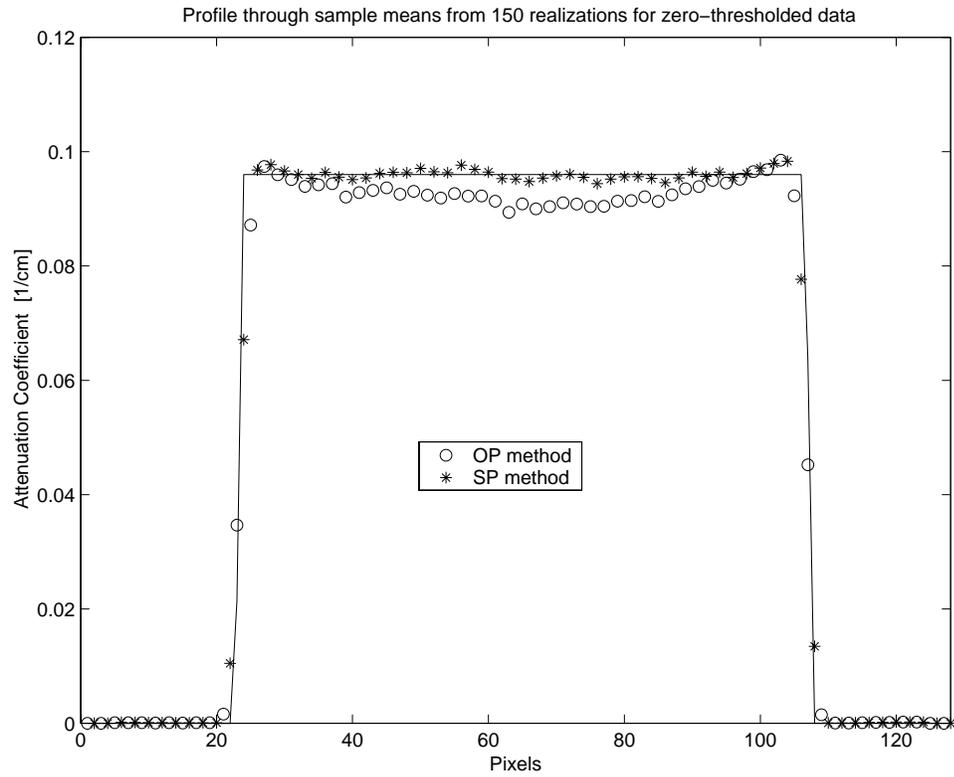


Figure 4.11: Horizontal profile through the sample mean images for abdomen phantom, obtained by using zero-thresholded data. The ordinary Poisson model leads to systematic negative bias, especially for interior regions of the reconstructed image. The shifted Poisson model estimator is free of systematic bias.

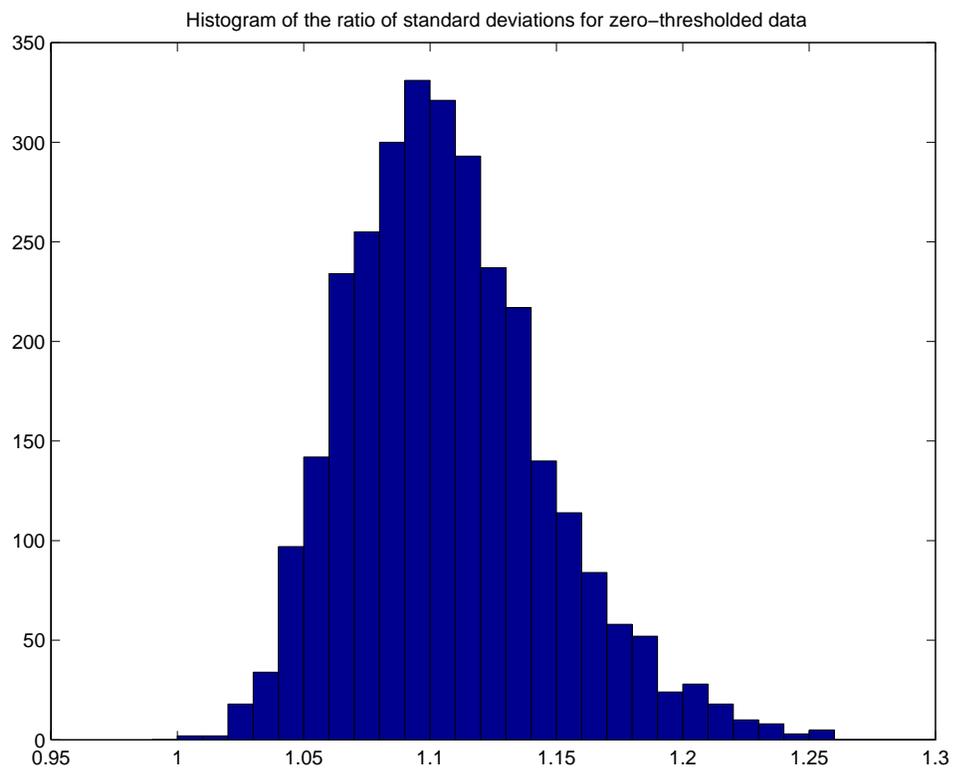


Figure 4.12: Histograms of the ratio of standard deviations for abdomen phantom, obtained by using zero-thresholded data. The ordinary Poisson model still leads to higher standard deviation than the shifted Poisson model, (on the average 11%).

## Sinograms of transmission and blank scans

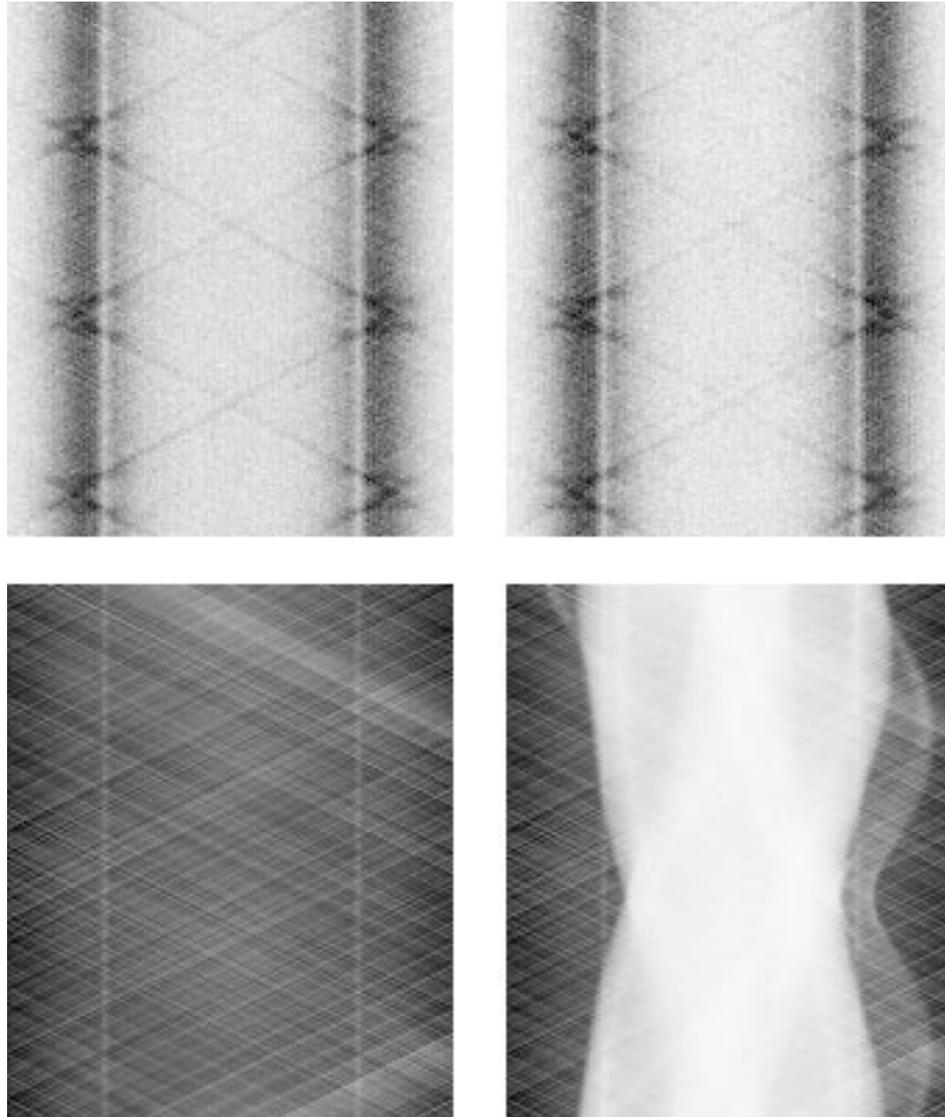


Figure 4.13: Separately collected sinograms (160 radial bins and uniformly spaced 192 angles). Clockwise from the upper left: (a) Delayed events of blank scan. (b) Delayed events of transmission scan. (c) Prompt events of transmission scan. (d) Prompt events of blank scan.

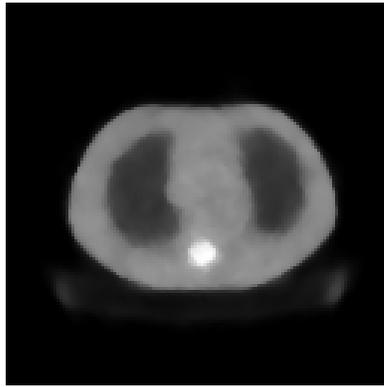


Figure 4.14: Phantom used in the PET system for transmission scan.

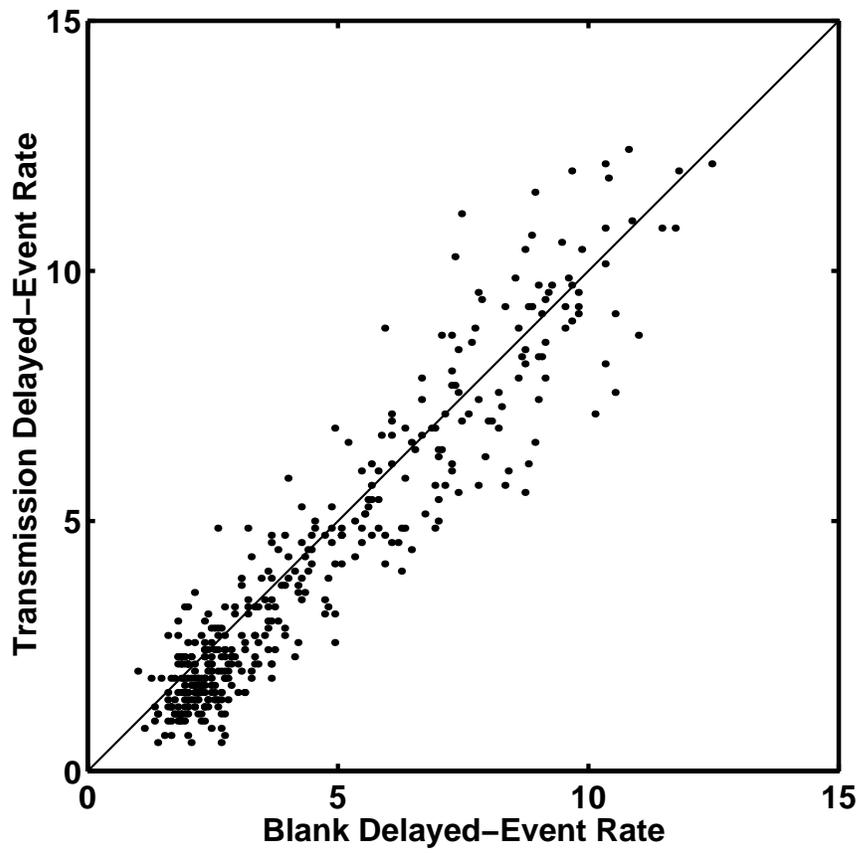


Figure 4.15: Scatter plot of delayed coincidence event of blank and transmission scans.

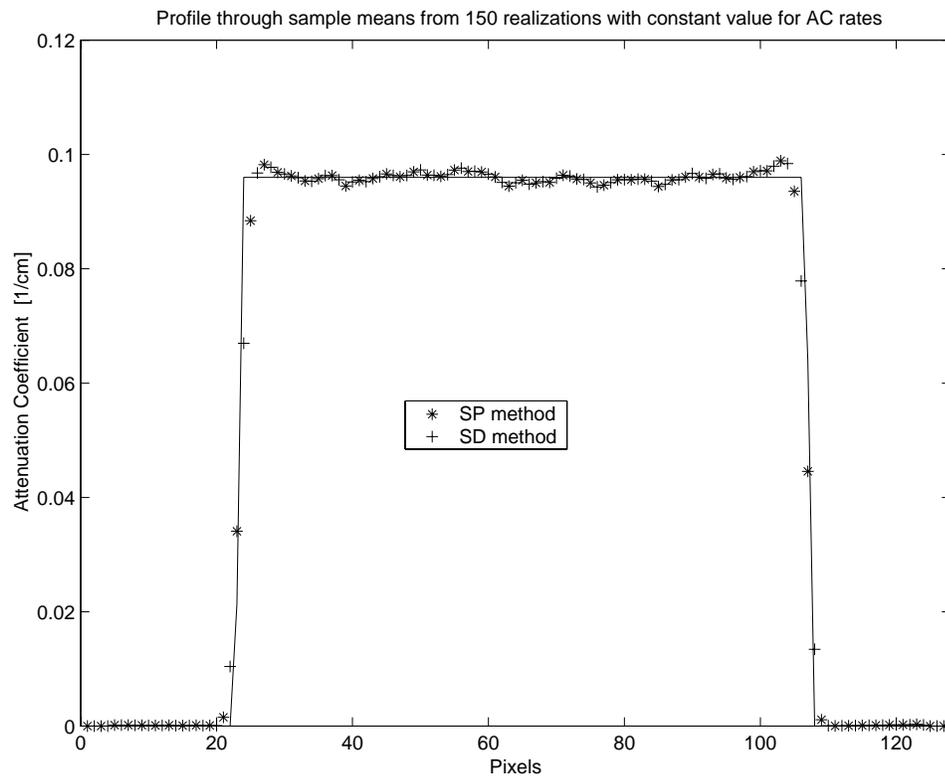


Figure 4.16: Horizontal profile through the sample mean images for abdomen phantom using constant AC rates. The constant AC rates approximation does not introduce any systematic bias to the estimators.

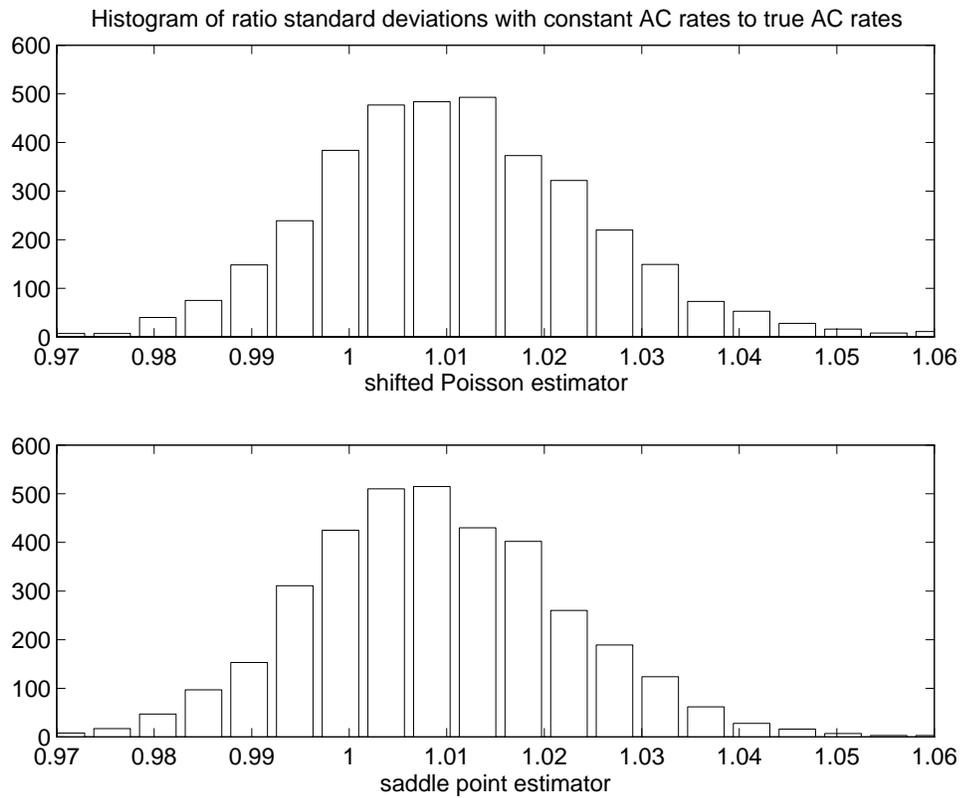


Figure 4.17: Histograms of the ratio of standard deviations of shifted Poisson estimators, for abdomen phantom. Using the constant AC rates approximation slightly increases the variance of the SP and SD estimators.

## 4.9 Experimental Results

We applied penalized-likelihood estimators based on the approximations presented in Section 4.3 to reconstruct attenuation maps from transmission scans acquired with a Siemens/CTI 931 PET scanner. To study the bias and variance properties of these estimators, we collected 100 two-minute transmission scans of an anthropomorphic thorax phantom (Data Spectrum, North Carolina). Fig. 4.18 shows the reconstructed attenuation map of the slice of interest from a 5 hour transmission scan. In each two-minute scan there were about 4.5M prompt coincidence events and 0.7M delayed events for the slice of interest and the acquired data was already randoms pre-corrected in hardware with standard delayed window coincidences method. The sinograms had 192 radial bins and 256 angles uniformly sampled over 180 degrees. We approximated the system geometry with 3.1 mm wide strip integrals and 3.1 mm ray spacing. The reconstructed images were 128 by 128 with 4.7 mm pixels. For regularization, we used the modified quadratic penalty [43] described in Section 2.4.2. This penalty improves the resolution uniformity and enables matching of the spatial resolutions of different methods. We matched the resolution of the reconstructed transmission images for all methods to 2.65 pixels FWHM.

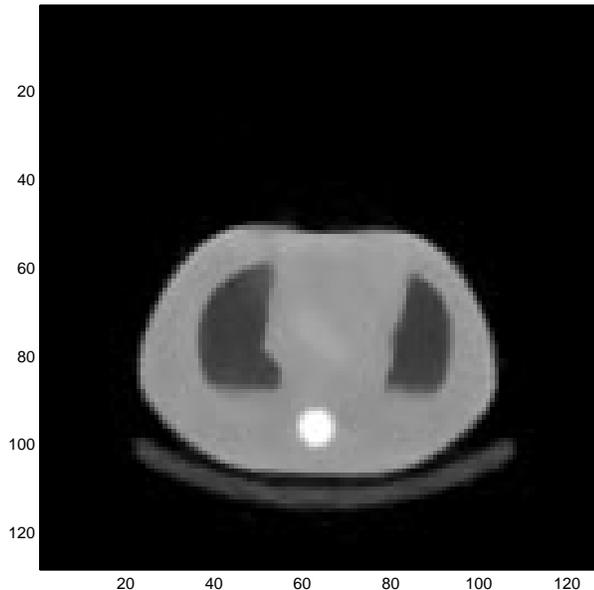


Figure 4.18: Reconstruction of attenuation map for the slice of interest from 5 hour transmission scan.

We applied EX, WLS, OP, SP and SD log-likelihood models (4.7, 4.10, and 4.11, 4.12, 4.13) to the experimental randoms pre-corrected transmission scans. The empirical results from this study are consistent with the previous simulation results: a large bias for the WLS method, and lower variance for the SP, SD and EX methods. Similar to Section 4.8.3, although the individual images reconstructed by the SP method and the EX method differed slightly, we observed very close agreement between the SD method and extremely precise *truncated* exact log-likelihood for each reconstruction. However, the differences between SP, SD and EX models in the ensemble means and variances were insignificant. Thus we concentrate on the simple SP log-likelihood model in this section.

Previously we have shown that a time-scaled version of delayed-coincidence events acquired during the blank scan is a good estimate for the  $r_n$  factors. (Even using a single scalar constant works fairly well as shown Figure 4.17.) Note that these estimates of the  $r_n$  factors are used essentially for estimating the variance of the randoms pre-corrected data in (4.5), not for performing randoms pre-correction. In our experiments the  $r_n$  factors were *not* available neither for the transmission nor for the blank scans, since the data was already pre-corrected for the randoms. Thus, to estimate the  $r_n$  factors for use in (4.12), we simply scaled the blank scan so that its sum corresponded to the total number of AC events (this scalar is available in the transmission scan file header) with no additional processing. Despite this possibly being a suboptimal approach, the SP method still yielded lower variance attenuation maps than the OP method.

For each transmission scan an estimate of the attenuation map was reconstructed using 20 iterations of the grouped-coordinate ascent algorithms (described in Section 4.7) applied to the objective functions (4.10), (4.11), and (4.12). In our simulations, we initialized the iterations with a resolution-matched FBP image and always observed monotonic increase in the objective function  $\Phi(\mu)$  for all cases. However, as a cautionary note it should be mentioned that we have no theoretical guarantee for the transmission problem that all methods will converge to the global maximum.

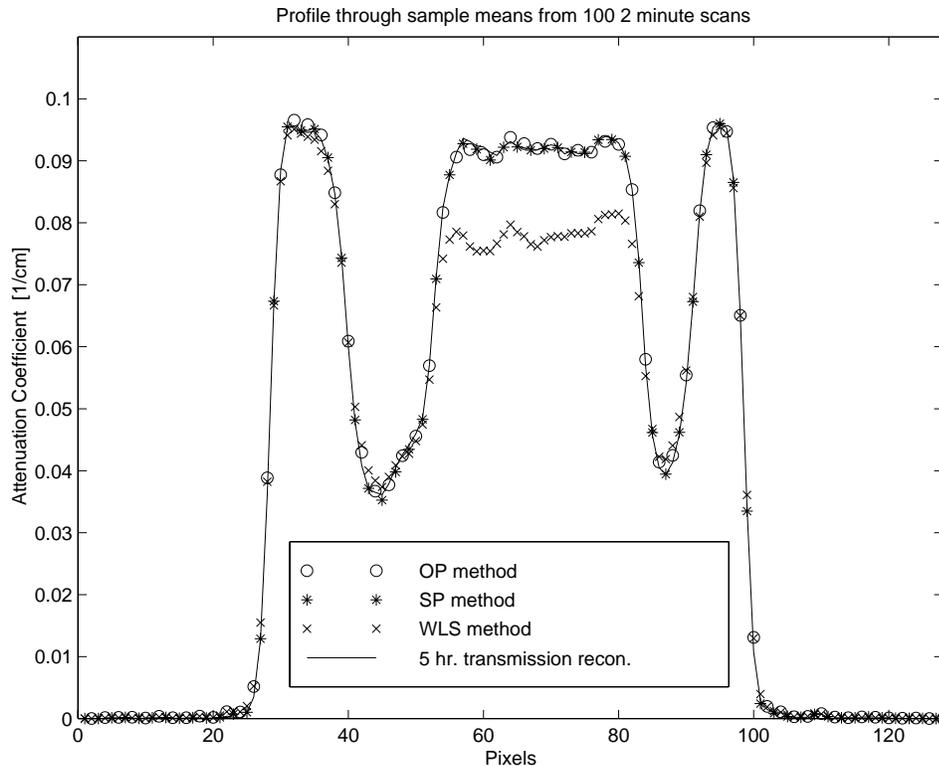


Figure 4.19: Horizontal profile 66 through the sample mean images for abdomen phantom. The WLS method has a systematic negative bias. The ordinary Poisson (OP) and shifted Poisson (SP) methods appear free of this systematic negative bias.

We computed both the sample mean and sample standard deviation images for the methods. Fig. 4.19 shows horizontal profiles of the sample mean images. These profiles show

that WLS is *systematically* negatively biased [32], whereas the OP and SP models appear free of such systematic bias. The logarithm required by the WLS method negatively biases the reconstructed transmission images (as described in Section 4.5) and this bias increases as counts decrease. Since the rays traversing the center of the transmission phantom have the lowest counts, these regions show the largest negative bias. (The overshoot at the edges is due to the quadratic penalty used in the reconstruction. Even with noiseless data, this blurring effect will still be present.).

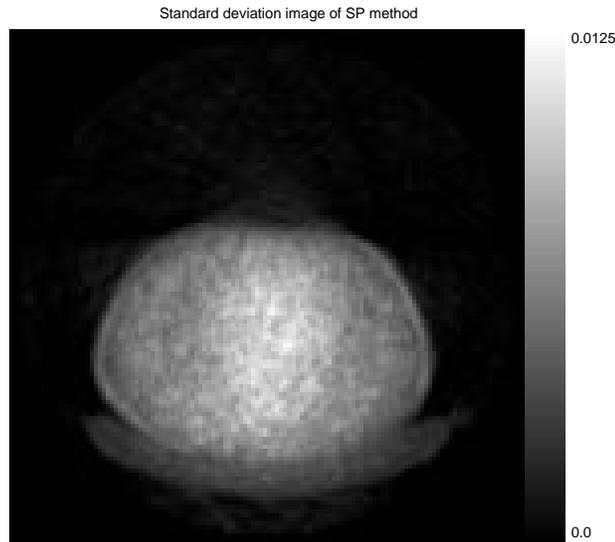


Figure 4.20: Sample standard deviation image of SP method from 100 transmission scans.

Fig. 4.20 shows the sample standard deviation image for the SP method. To study the variance, we computed the *ratio* of the sample standard deviation image of the OP method to the SP method, shown in Fig. 4.21. Fig. 4.22 shows the histogram of the standard deviation ratios over all interior pixels. The OP model yields, on the average, about 11% higher standard deviation than the SP model. Although the absolute standard deviation values could be decreased by using longer scan durations, we expect *relative* standard deviations of the OP and SP estimators to remain approximately constant for higher counts [34, 99]. This follows from the fact that the 1-D analytic approximations (4.20)-(4.22) and the 2-D analytic approximation (which will be introduced in the next section) become more accurate with increasing counts, and these approximations predict that the SP method will have less noise than the OP method.

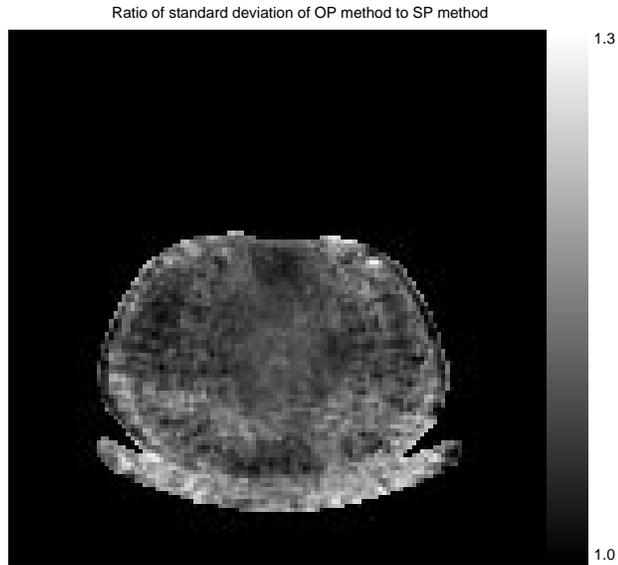


Figure 4.21: Ratio of sample standard deviation images of OP method to SP method from 100 transmission scans.

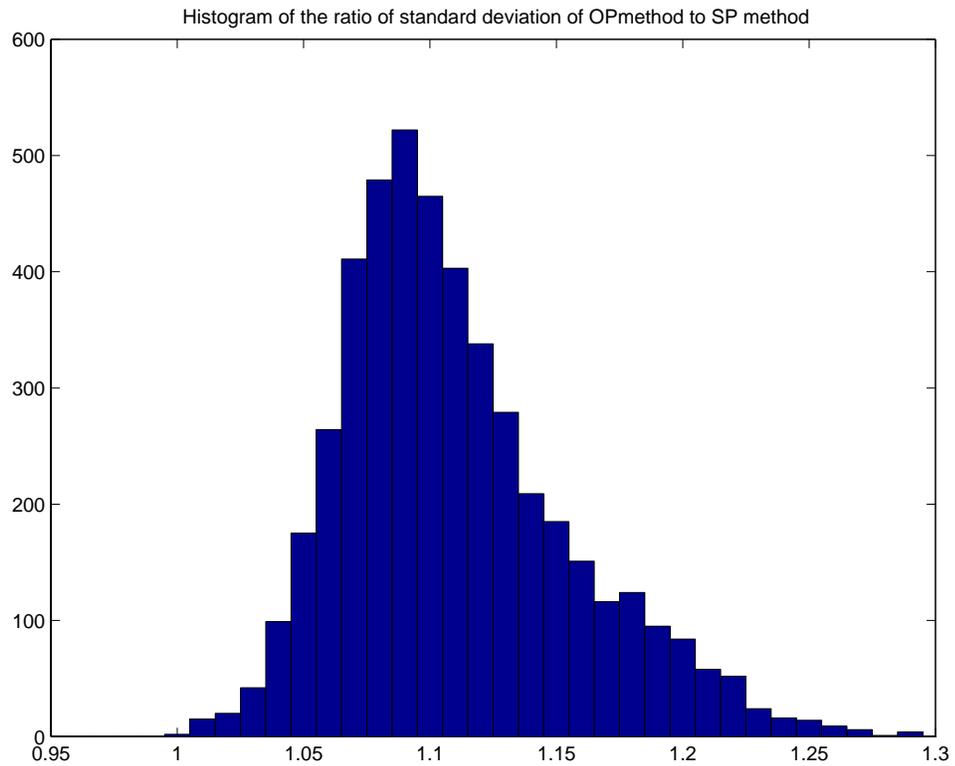


Figure 4.22: Histogram of the ratio of standard deviations in reconstructed attenuation maps. The ordinary Poisson (OP) method yields, on the average, about 11% higher standard deviation than the proposed shifted Poisson (SP) method.

## 4.10 Covariance Approximations for Transmission Tomography

One can use analytic approximations proposed in [34] to predict the covariance of penalized-likelihood reconstruction methods without exhaustive simulations. In [34] these approximations were shown to agree with empirical results from simulated PET scans (without randoms precorrection) even for the highly nonlinear transmission reconstruction methods. Here, we apply the covariance approximation presented in [34] to the OP and SP methods and compare the results with experimental randoms precorrected transmission data.

We can express both the OP (4.11) and SP (4.12) log-likelihood approximations in the form (4.6) with

$$h_n(l, y_n) = (y_n + d_n) \log(b_n e^{-l} + s_n^T + d_n) - (b_n e^{-l} + s_n^T + d_n) \quad (4.41)$$

and

$$d_n \triangleq \begin{cases} 0, & \text{OP} \\ 2r_n, & \text{SP} \end{cases} . \quad (4.42)$$

Combining the log-likelihood approximation with a roughness penalty forms the penalized log-likelihood objective function  $\Phi(\mu)$  as in (4.8).

A first-order Taylor expansion of  $\hat{\mu}(Y) = \arg \max_{\mu \geq 0} \Phi(\mu, Y)$  around  $\bar{Y} \triangleq \mathbb{E}\{Y\}$  leads to the following approximation for the covariance of  $\hat{\mu}$  [34]:

$$\begin{aligned} \text{Cov}\{\hat{\mu}\} &\approx \left[ -\nabla^{20}\Phi(\check{\mu}, \bar{Y}) \right]^{-1} \nabla^{11}\Phi(\check{\mu}, \bar{Y}) \text{Cov}\{Y\} \\ &\cdot \left[ \nabla^{11}\Phi(\check{\mu}, \bar{Y}) \right]' \left[ -\nabla^{20}\Phi(\check{\mu}, \bar{Y}) \right]^{-1}, \end{aligned} \quad (4.43)$$

where

$$\check{\mu} \triangleq \arg \max_{\mu} \Phi(\mu, \bar{Y}). \quad (4.44)$$

Following [34]:

$$-\nabla^{20}\Phi(\check{\mu}, \bar{Y}) = \mathbf{H} \triangleq \mathbf{A}' \text{diag}\{u_n\} \mathbf{A} + \beta \mathbf{R}(\check{\mu}) \quad (4.45)$$

$$\nabla^{11}\Phi(\check{\mu}, \bar{Y}) = -\mathbf{A}' \text{diag}\{c_n\}, \quad (4.46)$$

where  $\mathbf{A} = \{a_{nj}\}$  is the sparse system matrix, and

$$u_n \triangleq \left( 1 - \frac{(s_n^T + d_n)(\bar{y}_n(\mu^{\text{true}}) + d_n)}{(\bar{y}_n(\check{\mu}) + d_n)^2} \right) b_n e^{-l_n(\check{\mu})}, \quad (4.47)$$

$$c_n \triangleq \frac{b_n e^{-l_n(\check{\mu})}}{\bar{y}_n(\check{\mu}) + d_n}, \quad (4.48)$$

and  $\mathbf{R}(\theta) = \nabla^2 R(\theta)$ . Substituting (4.45), (4.46) and  $\text{Cov}\{Y\} = \text{diag}\{\bar{y}_n(\mu^{\text{true}}) + 2r_n\}$  into (4.43) yields the following approximation for the estimator covariance:

$$\text{Cov}\{\hat{\mu}\} \approx \mathbf{H}^{-1} \mathbf{A}' \text{diag}\{v_n\} \mathbf{A} \mathbf{H}^{-1} \quad (4.49)$$

with

$$v_n \triangleq \frac{(b_n e^{-l_n(\check{\mu})})^2 (\bar{y}_n(\mu^{\text{true}}) + 2r_n)}{(\bar{y}_n(\check{\mu}) + d_n)^2}, \quad (4.50)$$

with  $\bar{y}_n(\mu) = b_n e^{-l_n(\mu)} + s_n^T$  as in (4.4).

For the experimental transmission data we predicted the variance of  $\hat{\mu}_{\text{OP}}$  and  $\hat{\mu}_{\text{SP}}$  using the above approximations. In our implementation, we ignored the scattered events and followed the ‘‘plug-in’’ approach of [34], by replacing each  $\bar{y}_n(\check{\mu})$  and  $\bar{y}_n(\mu^{\text{true}})$  in (4.49) with the corresponding sample mean of the 100 transmission sinograms<sup>2</sup>. We used the preconditioned conjugate gradient method [16, 36] to compute selected diagonal elements of (4.49). Overall computation for computing the variance of each pixel was roughly equivalent to one maximization of  $\Phi(\mu)$ . Fig. 4.23 and Fig. 4.24 show the comparison of the empirical standard deviation and the approximate standard deviation of pixels through a horizontal cross section through the attenuation map for the OP method and the SP method. The predicted variance agrees well with the empirical results both for the OP and the SP methods. These results show that even for two-minute transmission scans analytical approximations can be used reliably. For longer scans with higher transmission counts the agreement should be even better [34].

## 4.11 Noise Propagation Into Emission Reconstruction

In this section we derive approximate expressions to analyze the propagation of noise from the attenuation maps through the ACFs into the reconstructed emission images. This analysis describes the effects of transmission noise on the final emission images, which may assist studies of the tradeoff between emission and transmission scan times, *e.g.* [4, 27]. Dahlbom and Hoffman [18] have analyzed emission image noise for the special case of uniform density disk phantom (assuming both emission and transmission images are reconstructed using FBP method). The covariance approximations presented here apply to arbitrary objects, for attenuation maps reconstructed by penalized-likelihood estimators with quadratic regularization.

### 4.11.1 Theory

To isolate the effects of transmission noise on the resultant emission images, we consider noiseless emission measurements and we consider the FBP method for reconstructing emission images after correcting for attenuation using noisy attenuation maps. We assume the noiseless emission measurements are:

$$z_n = e^{-l_n(\mu^{\text{true}})} p_n, \quad (4.51)$$

where

$$p_n = \sum_{k=1}^M g_{nk} \lambda_k$$

---

<sup>2</sup>Although replacing  $\bar{y}_n(\check{\mu})$  and  $\bar{y}_n(\mu^{\text{true}})$  in (4.49) with the sample mean of the transmission sinograms is impractical, it enables us to compute quickly the approximations for many pixels in the reconstructed image. In Section 4.11 we present the results of variance approximations for a set of pixels for noise propagation into emission images using the *true* plug-in approach (where we replace  $\bar{y}_n(\check{\mu})$  and  $\bar{y}_n(\mu^{\text{true}})$  with noisy measurements). There we show that the predictions again agree well with empirical standard deviation values.

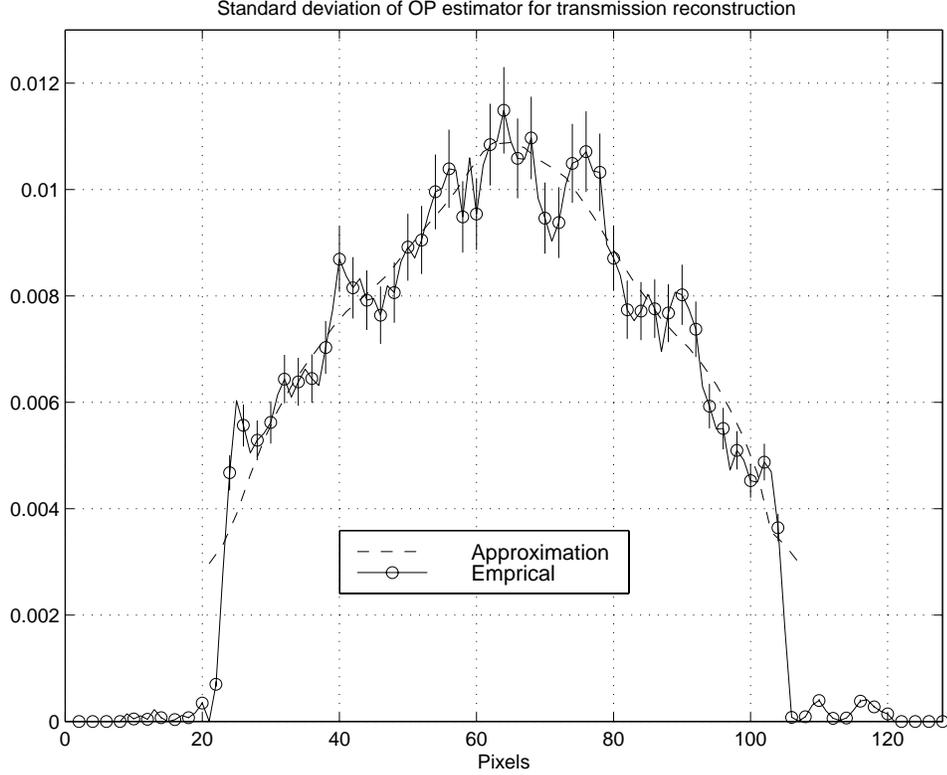


Figure 4.23: Empirical standard deviation (with error bars) and the approximate standard deviation of OP method for pixels along horizontal profile 90 through the attenuation map.

is the attenuation-free projection of the emission image and where  $\lambda = [\lambda_1 \dots \lambda_P]'$  denotes the vector of radio-isotope concentration.  $\mathbf{G} = \{g_{nk}\}$  represents the tomographic system response including the geometric system model, ray dependent factors (*e.g.* detector efficiency factors, dead-time, radio-isotope decay) and pixel dependent factors such as spatial variations in sensitivity. And  $e^{-l_n(\mu^{\text{true}})}$  (with  $l(\mu^{\text{true}}) = \mathbf{A}\mu^{\text{true}}$ ) represents the survival probability for the  $n$ th ray. The noiseless emission measurements  $z_n$  (4.51) are corrected for attenuation using ACFs based on the attenuation map estimates  $\hat{\mu}$ . If one directly corrects the emission measurements for attenuation by multiplication, the resultant images have some artifacts because of the resolution mismatch between emission and transmission sinograms [10, 11]. Thus, one needs to smooth the emission sinogram to the same resolution as the survival probabilities. We can write the attenuation-corrected emission sinogram as follows:

$$\hat{z}_n = e^{l_n(\hat{\mu})} \text{smooth} \left\{ e^{-l_n(\mu^{\text{true}})} p_n \right\}. \quad (4.52)$$

For FBP reconstruction of the emission images we consider the constrained least-squares (CLS) window corresponding to (50) of [33] :

$$\frac{\text{sinc}(ku) / \text{sinc}(u)}{\text{sinc}^2(ku) + \alpha u^3}, \quad u \in \left[0, \frac{1}{2}\right], \quad (4.53)$$

where  $u$  denotes spatial frequency in cycles per radial sample,  $k$  is the ratio of the strip width to the pixel size of the system model, and  $\alpha$  is linearly related to  $\beta$  below [33]. (The

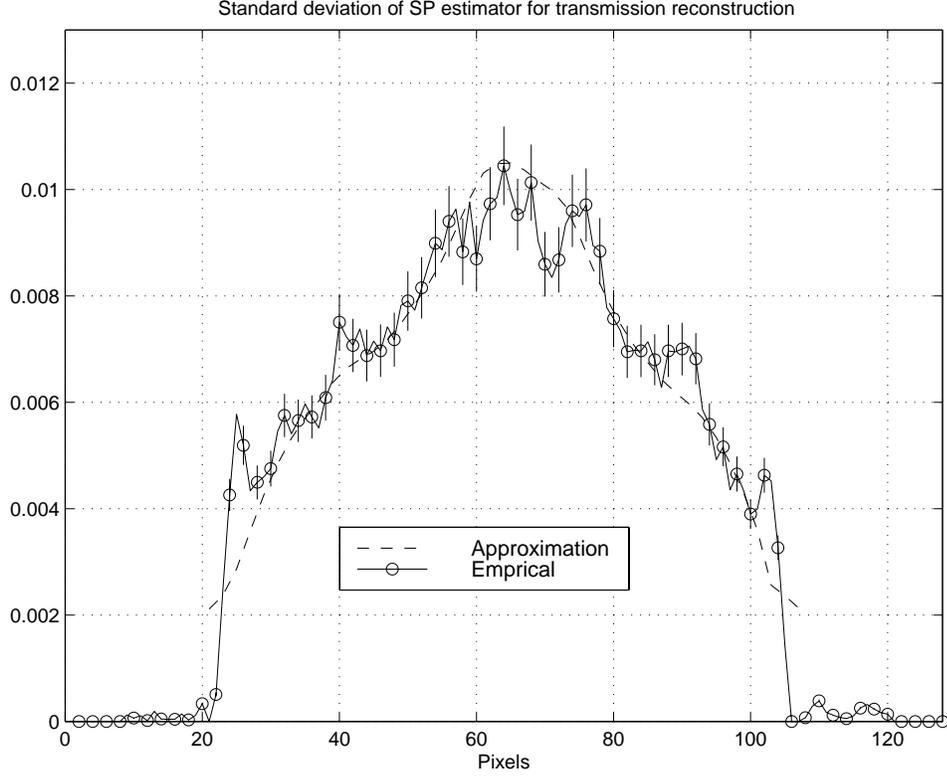


Figure 4.24: Empirical standard deviation (with error bars) and the approximate standard deviation of SP method for pixels along horizontal profile 90 through the attenuation map.

detector response is a rectangular function with frequency response  $\text{sinc}(ku)$ .) Dividing by  $\text{sinc}(u)$  in the numerator compensates for the linear interpolation step of the FBP method. The FBP algorithm with the above smoothing window (4.53) is essentially equivalent to quadratically penalized unweighted least-squares (QPULS) estimator without the nonnegativity constraint [33]. The QPULS estimator is defined as [35]:

$$\begin{aligned}\hat{\lambda}^{\text{QPULS}} &= \arg \min_{\lambda} \|\hat{z} - \mathbf{G}\lambda\|^2 + \beta\lambda' \mathbf{R}_o \lambda \\ &= [\mathbf{G}'\mathbf{G} + \beta\mathbf{R}_o]^{-1} \mathbf{G}'\hat{z},\end{aligned}\quad (4.54)$$

with

$$\mathbf{R}_o[j, k] = \begin{cases} \sum_l w_{jl}, & k = j \\ -w_{jk}, & k \neq j, \end{cases}\quad (4.55)$$

where  $w_{jk} = 1$  for horizontal and vertical neighboring pixels and 0 otherwise. Since this estimator is linear, its covariance is:

$$\text{Cov} \left\{ \hat{\lambda}^{\text{QPULS}} \right\} = [\mathbf{G}'\mathbf{G} + \beta\mathbf{R}_o]^{-1} \mathbf{G}' \text{Cov} \{ \hat{z} \} \mathbf{G} [\mathbf{G}'\mathbf{G} + \beta\mathbf{R}_o]^{-1}.\quad (4.56)$$

We must find  $\text{Cov} \{ \hat{z} \}$  to complete the above approximation. For simplicity we first make the following approximation:

$$\text{smooth} \left\{ e^{-ln(\mu^{\text{true}})} p_n \right\} \approx e^{-ln(\bar{\mu})} \text{smooth} \{ p_n \},\quad (4.57)$$

where  $\check{\mu}$  is defined in (4.44). We plug this into (4.52):

$$\hat{z}_n \approx e^{l_n(\hat{\mu})} e^{-l_n(\bar{\mu})} \text{smooth}\{p_n\}, \quad (4.58)$$

and approximate  $\text{Cov}\{\hat{z}\}$  as:

$$\text{Cov}\{\hat{z}\} \approx \mathbf{D} \text{Cov}\{t(\hat{\mu})\} \mathbf{D}', \quad (4.59)$$

where  $t(\hat{\mu}) \triangleq [t_1(\hat{\mu}) \dots t_N(\hat{\mu})]'$  with  $t_n(\hat{\mu}) = e^{l_n(\hat{\mu}) - l_n(\bar{\mu})}$  and  $\bar{\mu} \triangleq \text{E}\{\hat{\mu}\}$  and  $\mathbf{D} \triangleq \text{diag}\{\text{smooth}\{p_n\}\}$ .

Using first-order Taylor expansion around  $\bar{\mu}$  we approximate<sup>3</sup>  $\text{Cov}\{t(\hat{\mu})\}$  as:

$$\text{Cov}\{t(\hat{\mu})\} \approx \mathbf{A} \text{Cov}\{\hat{\mu}\} \mathbf{A}'. \quad (4.60)$$

Finally, plugging (4.59) and (4.60) into (4.56) yields

$$\text{Cov}\{\hat{\lambda}^{\text{QPULS}}\} \approx [\mathbf{G}'\mathbf{G} + \beta\mathbf{R}_o]^{-1} \mathbf{G}'\mathbf{D}\mathbf{A} \text{Cov}\{\hat{\mu}\} \mathbf{A}'\mathbf{D}\mathbf{G} [\mathbf{G}'\mathbf{G} + \beta\mathbf{R}_o]^{-1}. \quad (4.61)$$

The variance of the estimated total activity within a region of interest (ROI), *i.e.*  $\hat{\theta}_e = e' \hat{\lambda}^{\text{QPULS}}$ , is simply:

$$\text{Var}\{\hat{\theta}_e\} = e' \text{Cov}\{\hat{\lambda}^{\text{QPULS}}\} e, \quad (4.62)$$

where  $e$  is a column vector of length  $M$  that equals unity for the pixels in the region of interest and zero elsewhere. To within the accuracy of the preceding approximations, (4.61) shows the first-order propagation of the noise from the attenuation map  $\hat{\mu}$  into the emission reconstruction, and (4.59, 4.60) also show that  $\text{Cov}\{\hat{\mu}\}$  is scaled quadratically by the attenuation-free emission projections  $p_n$  (4.51) before propagating into emission image covariance (since it is sandwiched between  $\mathbf{D}$  matrices).

### 4.11.2 Results

We simulated noiseless emission measurements (4.51) for the emission phantom shown in Fig. 4.25, using the same system specifications as the experimental transmission data. (The rectangular regions numbered 1 through 5 are regions of interest used at the end of this section). The spine, lungs, soft tissue, and heart had relative radioactivity concentrations of 0, 1, 2 and 4 respectively. The effects of attenuation were included in (4.51) by calculating survival probabilities from an attenuation map reconstructed from a five-hour transmission scan. To reconstruct this attenuation map we used the very precise saddle-point (SD) approximation (4.13) along with an edge-preserving penalty function [39].

After smoothing the noiseless emission measurements to match the resolution of the transmission data [10, 11], we applied ACFs computed from the noisy attenuation map estimates  $\hat{\mu}_{\text{OP}}$  and  $\hat{\mu}_{\text{SP}}$  that were reconstructed from each experimental transmission scan. We reconstructed emission images using FBP with the CLS window (4.53). Fig. 4.26 shows the sample mean emission image with ACFs based on the SP method computed from 100 two-minute transmission scans as explained in Section 4.10. (The mean image of OP method is not shown since it was very similar to that of the SP method.)

Fig. 4.27 shows the sample standard deviation image of the 100 emission reconstructions with ACFs based on the SP method. To study the noise due to different methods, we

---

<sup>3</sup>We have found empirically that standard the deviation of the quantities  $l_n(\hat{\mu}) - l_n(\bar{\mu})$  were around 0.06. This empirical finding with our noisy experimental data justifies the Taylor series approximation.

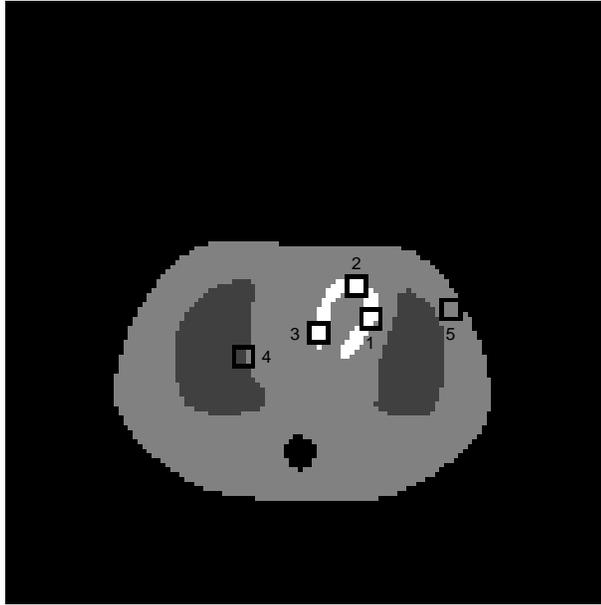


Figure 4.25: Emission phantom with several rectangular regions for noise computation.

computed the *ratio* of sample standard deviation images of emission reconstruction with ACFs based on the OP method and the SP method, shown in Fig. 4.28. Fig. 4.29 shows the histogram of the standard deviation ratios, over all interior pixels. Attenuation correction based on the OP model yielded about 20% higher standard deviation than the SP model on average.

To assess the accuracy of our analytical approximations, we compared to empirical variances described above to the variances predicted by (4.62). We used the preconditioned conjugate gradient method to compute (4.62) for a set of pixels in the reconstructed emission image. We determined the elements of  $\text{Cov}\{\hat{\mu}\}$  in (4.61) two different ways: one way used the approximation (4.49); the other way used the empirical covariance of the 100 independent attenuation map reconstructions<sup>4</sup>. Although replacing  $\text{Cov}\{\hat{\mu}\}$  with an empirical covariance is impractical for routine use, it helps establish the accuracy of approximation (4.62). Figures 4.30 and 4.31 compare the empirical standard deviations and the approximate standard deviations of pixels along a horizontal profile through the emission images. The analytical approximations for transmission noise propagation agree well with the empirical results, and confirm the reduction in noise for SP method compared to OP method.

Table 4.2 shows the percent standard deviation of the activity within the five different 3 by 3 pixel ROIs shown in Fig. 4.25 for the reconstructed images, with ACFs based on the OP method and SP method. For each ROI, we also implemented the practical plug-in

<sup>4</sup>Instead of computing the empirical covariance directly from the independent attenuation map reconstructions, we used the following computationally more efficient method. It can be seen from (4.61) and (4.62) that  $\text{Var}\{\hat{\theta}_e\} = \text{Var}\{S\}$  where  $S \triangleq c'\hat{\mu}$  and  $c' \triangleq e'[\mathbf{G}'\mathbf{G} + \beta\mathbf{R}_o]^{-1}\mathbf{G}'\mathbf{D}\mathbf{A}$ . Using the preconditioned conjugate gradient method [16, 36], we pre-compute the row vector  $c'$  only once and then compute the scalar  $S$  for each independent attenuation map reconstruction. And then finally the sample variance of  $S$  is computed.

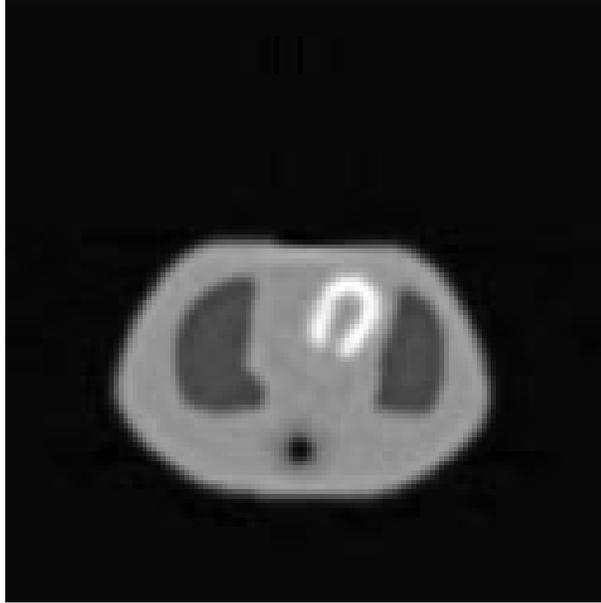


Figure 4.26: Empirical sample mean of emission images reconstructed with ACFs based on 100 different estimates of  $\hat{\mu}_{SP}$ .

approach for computing (4.49), (which is then used in (4.62) for predicting the variance of the reconstructed emission image pixels.) In this plug-in approach, we replaced each  $\bar{y}_n(\check{\mu})$  and  $\bar{y}_n(\mu^{\text{true}})$  in (4.49) with the corresponding noisy sinogram element  $y_n$ . We computed variance approximation (4.62) for each of the 100 sinograms. Table 4.2 shows the sample means (and standard errors) of the plug-in predicted variances for each ROI. The OP model yields 8% to 23% higher standard deviation than SP model, and all the analytical approximations agree well with empirical standard deviation values.

For comparison purposes we simulated 100 noisy emission sinograms having an average of 2M counts per scan, and performed FBP reconstruction of the emission images. For the ACFs we used the empirical mean of the transmission scans, to ensure that only emission noise affected the reconstructions. (Since the emission noise is inversely proportional to the square root of the total counts per scan, one could also predict emission noise for other count levels.) Table 4.2 shows the empirical standard deviations for different ROIs due to emission noise. These simulations illustrate the relative effects of emission and transmission noise<sup>5</sup>.

## 4.12 Conclusions

AC events are a primary source of background noise in positron emission tomography. After the AC events are precorrected, the measurement statistics are no longer Poisson and the exact log-likelihood is complicated. For transmission scans, WLS method and PML method based on ordinary Poisson (OP) model lead to systematic bias and higher

<sup>5</sup>Although transmission scans contained about 3.6M counts per scan, most of the counts were from detector pairs whose line of responses do not intersect with the patient which yield un-attenuated high counts.

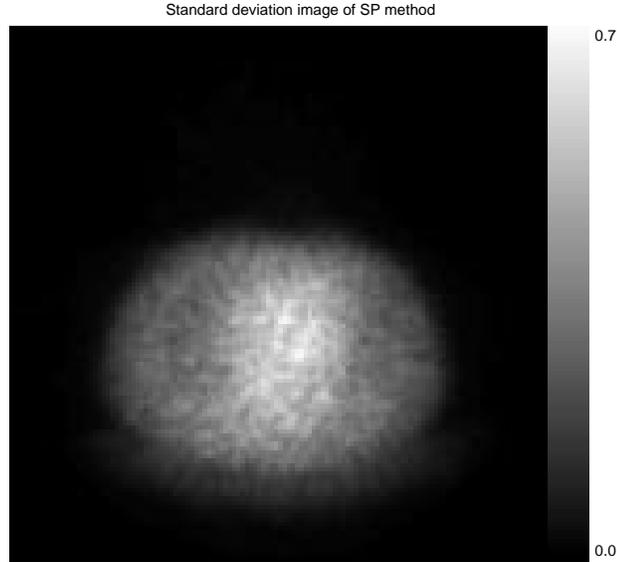


Figure 4.27: Sample standard deviation image of emission reconstruction with ACFs based on SP method.

variance, respectively, compared to our proposed shifted Poisson (SP) and saddle point (SD) models for the measurement statistics. Approximations, simulations and experimental studies show that the new approximation agrees closely with the exact log-likelihood model of the randoms pre-corrected measurements. Both the SP method and the SD method are free of systematic bias and yield reduced standard deviation (about 10 – 15%) compared to the OP model (at matched spatial resolution).

Although the individual images reconstructed by the SP method and the EX method differed slightly, we observed very close agreement between the SD method and *truncated* exact log-likelihood for each reconstruction. However, the differences between SP, SD and EX models were statistically insignificant (based on the ensemble means and variances). Thus SP method is particularly attractive since it requires comparable computation to the OP method but has reduced variance.

We applied the covariance approximations to the attenuation map estimates from the OP method and the SP method, and demonstrated that these approximations agree with the empirical results from the experimental PET transmission scans. These approximations can be used to determine the variance of transmission reconstruction to investigate parameters of interest (*e.g.* regularization parameters) and can supplement simulations. The approximations also showed that the SP method yields less noisy images compared to the OP method.

We also developed approximations to analyze the propagation of noise from attenuation maps into emission reconstruction. For this purpose we assumed noiseless emission measurements and developed approximations for the covariance of emission reconstruction with ACFs computed from noisy attenuation maps. The approximations agree with the empirical results and describe the propagation of noise from attenuation maps into emission reconstruction.

Both the approximations and the empirical results show the interesting property that when the transmission scan noise is propagated into the emission images, the relative dif-

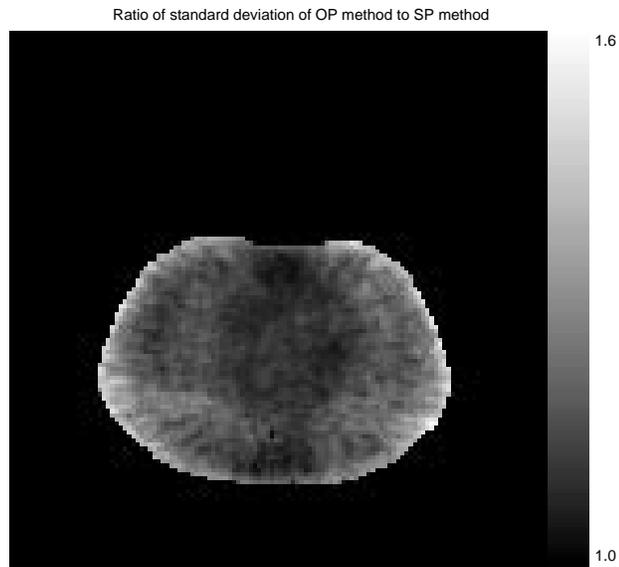


Figure 4.28: Ratio of sample standard deviation images of emission reconstruction with ACFs based on OP method and SP method.

ferences in the variances between the OP model and the proposed SP and SD models, can be even greater than when one considers the noise in the attenuation maps alone.

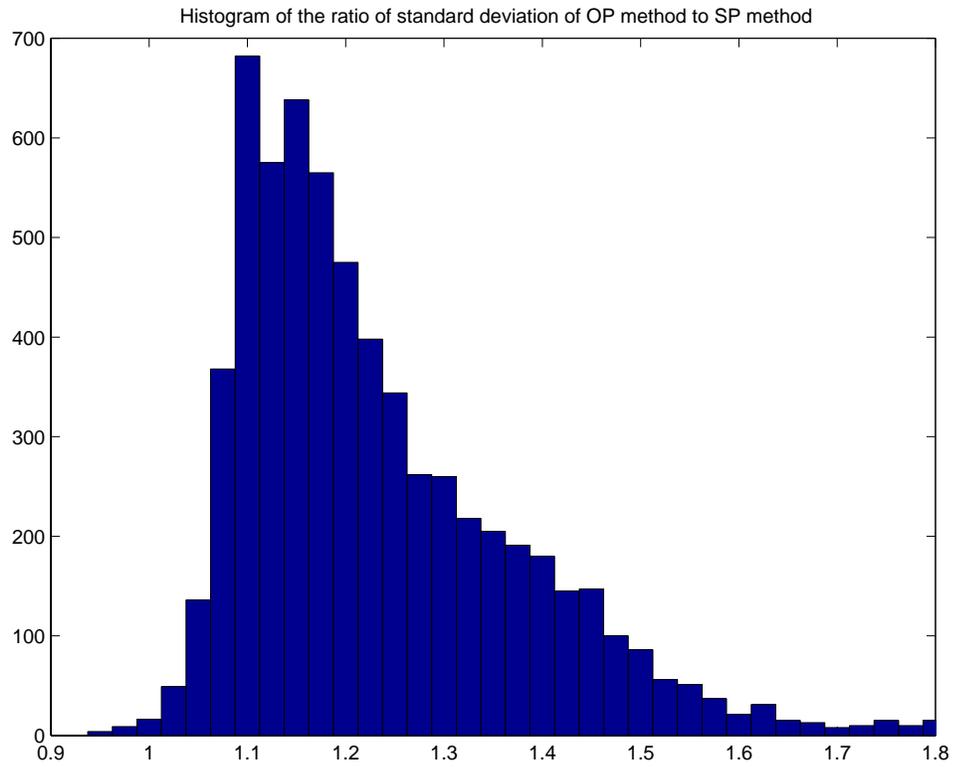


Figure 4.29: Histogram of the ratio of standard deviations in the reconstructed emission images with ACFs based on OP model and SP model. Attenuation correction factors based on the OP model yielded, about 20% higher standard deviation than the SP model on average.

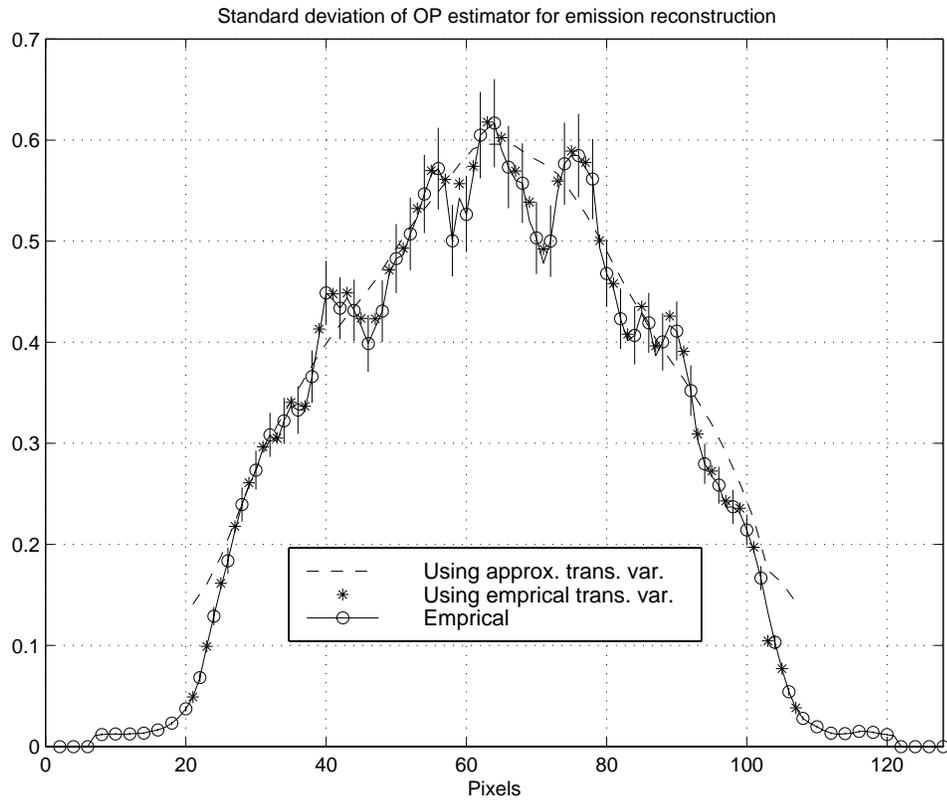


Figure 4.30: Empirical standard deviation (with error bars) and the approximate standard deviation of OP method (using both empirical transmission variance and approximate transmission variance) for pixels along horizontal profile 90 through the reconstructed emission images.

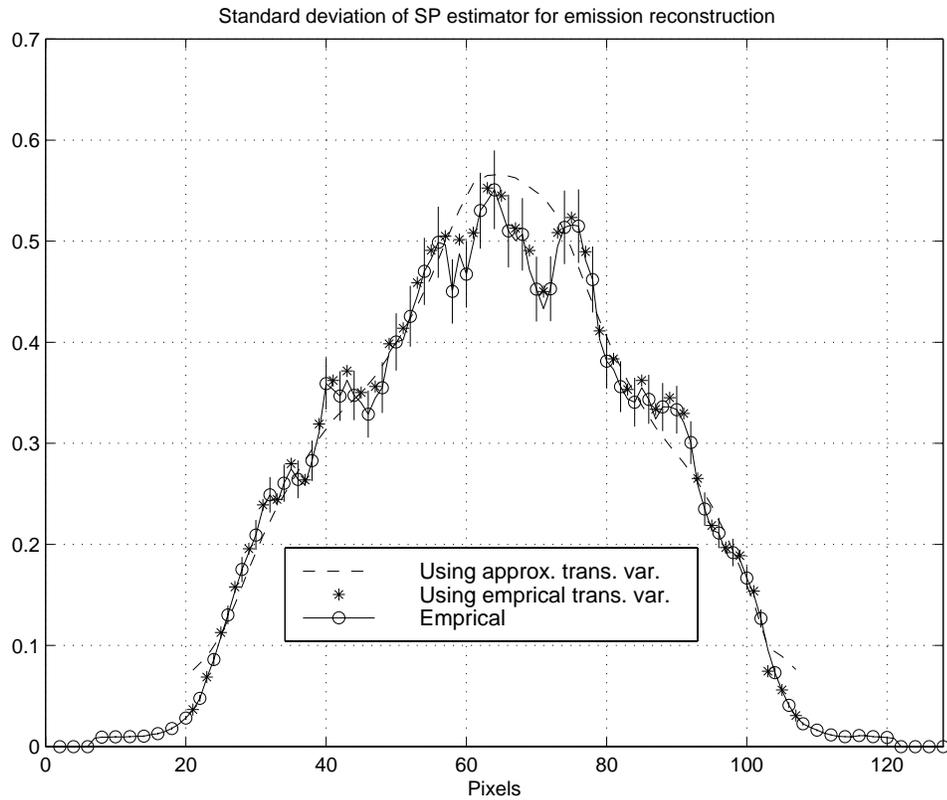


Figure 4.31: Empirical standard deviation (with error bars) and the approximate standard deviation of SP method (using both empirical transmission variance and approximate transmission variance) for pixels along a horizontal profile 90 through the reconstructed emission images.

Region	OP Method				
	Empr. Std.	App. Std. (wt. emp tr var)	App. Std. (wt. app tr var)	App. Std. (wt. plug-in )	
1	11.35	11.56	12.28	12.23 $\pm$ 0.14	
2	12.04	12.14	10.82	10.74 $\pm$ 0.14	
3	16.87	17.09	14.74	15.07 $\pm$ 0.22	
4	25.55	25.72	23.86	23.59 $\pm$ 0.27	
5	8.89	8.89	9.63	9.74 $\pm$ 0.10	
Region	SP Method				Emission Noise
	Empr. Std.	App. Std. (wt. emp tr var)	App. Std. (wt. app tr var)	App. Std. (wt. plug-in )	
1	10.20	10.39	11.34	10.88 $\pm$ 0.10	2.60
2	10.93	10.98	9.80	9.39 $\pm$ 0.09	2.12
3	15.68	15.91	14.32	13.99 $\pm$ 0.16	2.79
4	24.85	25.34	23.53	22.54 $\pm$ 0.20	4.66
5	7.30	7.35	7.61	7.49 $\pm$ 0.07	2.47

Table 4.2: Empirical percent standard deviation and the approximate analytical percent standard deviation of emission reconstruction using ACFs based on the OP method and SP method (using both empirical transmission variance and approximate transmission variance and plug-in transmission variance) for different regions shown in Fig. 4.25. Last column shows the empirical percent noise of the regions due to only emission noise for two million counts per emission scan.

## CHAPTER 5

### PET Emission Scans

#### 5.1 Introduction

In PET emission scans, generally a significant portion of the collected data is accidental coincidence (AC) events and it is a primary source of background noise [53,74,86]. Moreover, AC rates increase as the *square* of the amount of radio-isotope injected to the patient, while true coincidences increase only *linearly* with the radio-isotope concentration. This count rate limitation, along with detector deadtime determines the upper limit on the injected radio-isotope dose for many PET studies. Most PET scans are compensated for AC events by real-time subtraction of delayed-window coincidences. Real time subtraction of delayed coincidences compensates for the average of AC events, but also destroys the Poisson statistics [53] (Chapter 3). Moreover, negative values result during the real-time subtraction of delayed coincidences. These negative values would cause conventional penalized maximum likelihood algorithms to diverge. Setting the negative values to zero alleviates this problem but introduces a systematic positive bias in the resulting images [74,96].

In this chapter we briefly review the measurement model, the exact log-likelihood and the approximations to the exact log-likelihood (described in Chapter 3) in the context of PET emission scans with randoms pre-corrected measurements. We analyze the concavity of the proposed objective functions and develop appropriate maximization algorithms to be used in the image reconstructions with the proposed methods. We show that the proposed approximate statistical models result in reconstructions free of systematic bias and lead images with less noise compared to ordinary Poisson (OP) model for the randoms pre-corrected data [100]. Although the SP model is shown to be slightly biased for emission scans with very low count rates, the SD model is free of any systematic bias and perform almost identically as the exact log-likelihood. Lastly, we study the bias-variance trade-offs of the new methods by analyzing how close they perform to the uniform Cramer-Rao bounds [51,91].

#### 5.2 Exact Log-Likelihood

In conventional PET scans the system detects coincidence events during two time windows: “prompt” window and “delayed” window, and the data are pre-corrected for AC events by *real-time* subtraction of delayed window coincidences [53]. Each such pre-corrected measurement is the difference of two independent Poisson random variables, which compensates in mean for AC events, but which also increases the measurement variance.

Let  $Y = [Y_1, \dots, Y_N]'$  denote the vector of *precorrected* measurements, where “ ’ ” denotes vector and matrix transpose. The precorrected measurement for the  $n$ th coincidence detector pair is:

$$Y_n = Y_n^{\text{prompt}} - Y_n^{\text{delay}}, \quad (5.1)$$

where  $Y_n^{\text{prompt}}$  and  $Y_n^{\text{delay}}$  are the number of coincidences within the prompt and delayed windows, respectively. Let  $\lambda = [\lambda_1, \dots, \lambda_P]'$  denote the vector of unknown radio-isotope concentration. For emission scans, we assume that  $Y_n^{\text{prompt}}$  and  $Y_n^{\text{delay}}$  are statistically independent Poisson random variables with means  $\bar{y}_n^{\text{p}}$  and  $\bar{y}_n^{\text{d}}$  respectively as:

$$\mathbb{E} \{ Y_n^{\text{prompt}} \} = \bar{y}_n^{\text{p}}(\lambda) = \sum_{j=1}^P g_{nj} \lambda_j + s_n^E + r_n \quad (5.2)$$

$$\mathbb{E} \{ Y_n^{\text{delay}} \} = \bar{y}_n^{\text{d}} = r_n, \quad (5.3)$$

where  $\mathbf{G} = \{g_{nj}\}$  represents the geometric system response and ray-dependent factors such as attenuation and detector efficiency and the  $r_n > 0$  factors denote the mean of the AC events.

Since  $Y_n^{\text{prompt}}$  and  $Y_n^{\text{delay}}$  are statistically independent and Poisson:

$$\mathbb{E} \{ Y_n \} = \bar{y}_n^{\text{p}}(\lambda) - \bar{y}_n^{\text{d}} = \sum_{j=1}^P g_{nj} \lambda_j + s_n^E \triangleq \bar{y}_n(\lambda), \quad (5.4)$$

$$\text{Var} \{ Y_n \} = \bar{y}_n^{\text{p}}(\lambda) + \bar{y}_n^{\text{d}} = \sum_{j=1}^P g_{nj} \lambda_j + s_n^E + 2r_n. \quad (5.5)$$

Let  $y = [y_1, \dots, y_N]'$  be an observed realization of  $Y$  in (5.1). Since the measurements are independent, one can express the exact log-likelihood as follows:

$$L(\lambda) = \sum_{n=1}^N h_n(l_n(\lambda)), \quad (5.6)$$

with

$$l_n(\lambda) = \sum_{j=1}^P g_{nj} \lambda_j, \quad (5.7)$$

and ignoring constants independent of  $\lambda$  throughout:

$$h_n(l_n(\lambda)) \triangleq \log \left( \sum_{m=[-y_n]_+}^{\infty} \frac{[\bar{y}_n^{\text{p}}(\lambda)]^{y_n+m}}{(y_n+m)!} \frac{r_n^m}{m!} \right) - (\bar{y}_n^{\text{p}}(\lambda) + r_n), \quad (5.8)$$

where  $[x]_+ = x$  if  $x > 0$  and is 0 otherwise.

Since image reconstruction is ill conditioned, we combine a roughness penalty  $R(\lambda)$  with the log-likelihood to form a penalized-likelihood objective function:

$$\Phi(\lambda) = L(\lambda) - R(\lambda). \quad (5.9)$$

The goal is to estimate  $\lambda$  by maximizing  $\Phi(\lambda)$  over the nonnegative cone:

$$\hat{\lambda} = \arg \max_{\lambda \geq 0} \Phi(\lambda). \quad (5.10)$$

The exact log-likelihood function (5.8) has a complicated form because of the lower and upper summation limits. Next we describe approximations to the exact log-likelihood.

### 5.3 Approximations to the Exact Log-Likelihood

In this section, we briefly review the four practical approximations to  $L(\lambda)$  (described in Chapter 3) : the WLS model, the conventional OP model, the proposed SP model approximation and lastly the newly proposed SD model approximation. All log-likelihood approximations have the form (5.6) for different choices for  $h_n(l, y_n)$ .

#### 5.3.1 Quadratic Approximations

Quadratic approximation to the exact log-likelihood function [31]: leads to the data-weighted least squares objective function  $L_{\text{WLS}}(\lambda)$  of the form (5.6) with

$$h_n^{\text{WLS}}(l) = \begin{cases} -\frac{1}{2} \sum_{n=1}^N (l + s_n^E - y_n)^2 \frac{1}{\hat{\sigma}_n^2}, & y_n > 0 \\ 0, & y_n \leq 0, \end{cases} \quad (5.11)$$

where  $\hat{\sigma}_n^2 = y_n + 2r_n$  is the data estimated variance of emission measurements.

#### 5.3.2 Ordinary Poisson (OP) Approximation

The conventional approach is to ignore the random coincidences by assuming that  $\{Y_n\}_{n=1}^N$  are distributed as independent Poisson random variables with means  $\bar{y}_n(\lambda)$  given by (5.2). The log-likelihood  $L_{\text{OP}}(\lambda)$  corresponding to this OP approximation is of the form (5.6) with

$$h_n^{\text{OP}}(l) = [y_n]_+ \log(l + s_n^E) - (l + s_n^E), \quad (5.12)$$

where  $[x]_+ = x$  if  $x > 0$  and is 0 otherwise. This thresholding ensures concavity of the OP objective function as will be described in Section 5.5.

#### 5.3.3 Shifted Poisson (SP) Approximation

A better approach is to match both the first and second moments by approximating the random variables  $\{Y_n + 2r_n\}_{n=1}^N$  as having Poisson distributions with means  $\{\bar{y}_n(\lambda) + 2r_n\}$ . This idea leads to the SP approximation  $L_{\text{SP}}(\lambda)$  [97–99] (derived in Section 3.3.3) of the form (5.6) with

$$h_n^{\text{SP}}(l) = [y_n + 2r_n]_+ \log(l + s_n^E + 2r_n) - (l + s_n^E + 2r_n), \quad (5.13)$$

where again the zero thresholding of  $(y_n + 2r_n)$  is necessary to guarantee that the objective function is concave.

#### 5.3.4 Saddle-point (SD) Approximation

An even better approximation, which is based on second order Taylor series approximation for the exact pmf, is derived previously in Section 3.4. For emission tomography this saddle point (SD) approximation [98, 99] is of the form (5.6) with:

$$h_n^{\text{SD}}(l) = y_n \log \left( \frac{l + s_n^E + r_n}{z_n + u_n(l)} \right) - l + u_n(l) - \frac{1}{2} \log(u_n(l)) \quad (5.14)$$

where from (3.35)

$$z_n = \begin{cases} y_n + 1, & y_n \geq 0 \\ y_n - 1, & y_n < 0 \end{cases}, \quad (5.15)$$

and

$$u_n(l) = \sqrt{z_n^2 + 4(l + s_n^E + r_n)r_n}. \quad (5.16)$$

## 5.4 Exact Log-likelihood for Prompt Data

If one has access to the prompt data  $y_n^P$  (5.1) separately, then the exact log-likelihood  $L_{\text{PR}}(\mu)$  can be written in the form (5.6) with [61]:

$$h_n^{\text{PR}}(l) = y_n^P \log(l + s_n^E + r_n) - (l + s_n^E + r_n). \quad (5.17)$$

The reason we include the exact log-likelihood model for prompt data is to simply be able to compare the bias and variance results with the methods for randoms-precorrected data.

## 5.5 Concavity and Convergence

In this section we analyze the concavity of the various log-likelihood approximations to study their converge properties for the emission reconstruction.

The second partial derivatives of the OP (5.12) and the SP (5.12) objective functions and the PR log-likelihood (5.17) can be written as:

$$-\frac{\partial^2}{\partial \lambda_j \partial \lambda_k} L(\lambda) = \sum_{n=1}^N g_{nj} g_{nk} \frac{x_n}{(\bar{y}_n(\lambda) + d_n)^2}, \quad (5.18)$$

with

$$d_n \triangleq \begin{cases} 0, & \text{OP} \\ 2r_n, & \text{SP} \\ r_n, & \text{PR} \end{cases} \quad (5.19)$$

and

$$x_n \triangleq \begin{cases} \lfloor y_n \rfloor_+, & \text{OP} \\ \lfloor y_n + 2r_n \rfloor_+, & \text{SP} \\ y_n^P, & \text{PR.} \end{cases} \quad (5.20)$$

Thus, it can be seen that the methods are globally concave for  $x_n > 0$ , hence the zero thresholds in (5.12, 5.13).

Since the ‘‘thresholding function’’  $\lfloor y_n \rfloor_+$  is not differentiable at  $y_n = 0$ , it is difficult to derive accurate analytic approximations for the mean and variance of the different estimators above. However, one can explain intuitively the overall effect of zero-thresholding as follows: setting negative precorrected data values to zero increases the mean of the precorrected data. For the emission problem the data is linearly related to emission rates (i.e.:  $y_n \approx \sum_{j=1}^P g_{nj} \lambda_j$ ), thus the increase in the mean value of the precorrected data causes the estimator to introduce a systematic *positive bias* for the estimated emission rates [74,96].

Concavity proof of the SD method is very detailed and it is investigated in Appendix E. In Appendix E we prove that  $h_n^{\text{SD}}(l)$ 's are concave for  $l \in [0, \infty)$ . We also investigate the convexity of derivatives of the  $h_n^{\text{SD}}(l)$ 's since we use the paraboloid surrogates maximization algorithm of Fessler and Erdođan [38] which requires certain convexity conditions of the derivatives of the  $h_n^{\text{SD}}(l)$ 's [28].

## 5.6 Log-likelihood Maximization

In this section we briefly review the maximization algorithms from the point of view of the proposed objective functions for randoms-precorrected emission measurements and derive appropriate maximization methods. We will first review the EM type algorithms which are commonly used for emission tomography, and show how they can be applied to the proposed approximations. Next, we will describe the application of paraboloid surrogates maximization method to the SD method.

### 5.6.1 EM Type Algorithms

The expectation maximization (EM) algorithm is an iterative technique for computing ML estimates [26], which is especially useful when direct calculation of ML estimates are intractable. In this method, the observed measurements are supplemented with a complete (unobserved) data space. Then at each iteration, one calculates the conditional expectation of the complete data space and simultaneously maximizes the expectation with respect to unknown parameters. Since its introduction, [61, 81] EM method has been used widely to compute ML estimates in emission tomography.

Space-alternating generalized EM (SAGE) algorithm [40–42] is a generalized EM type algorithm which updates parameters sequentially by alternating between small hidden-data spaces [41]. As SAGE uses separate hidden data spaces for each parameter, not only the maximization is simplified but convergence rate is also improved compared to EM. In this section we will follow the notation in [41] for different SAGE algorithms for emission tomography reconstruction. We will derive the algorithms for randoms-precorrected measurements.

For emission tomography the number of coincidences within the prompt window for the  $n$ th coincidence detector pair is:

$$Y_n^{\text{prompt}} = \sum_{k=1}^p N_{nk} + R_n + S_n,$$

where  $N_{nk}$  denote the number of emissions from  $k$ th pixel that are detected by the  $n$ th detector pair within the prompt window and  $R_n$  and  $S_n$  denote the number of AC events and scattered events respectively, detected by the  $n$ th coincidence detector pair within the prompt window.  $N_{nk}$ ,  $R_n$  and  $S_n$  are statistically independent Poisson random variables:

$$\begin{aligned} N_{nk} &\sim \text{Poisson}(g_{nk}\lambda_k) \\ R_n &\sim \text{Poisson}(r_n) \\ S_n &\sim \text{Poisson}(s_n^E) \end{aligned}$$

where  $\lambda_k$  and  $g_{nk}$  are as defined by (2.13, 2.14) and  $r_n \geq 0$  and  $s_n^E \geq 0$  factors denote the mean of AC events and scattered events respectively.

The number of coincidence events within the delayed window for the  $n$ th coincidence detector pair are also distributed Poisson, with mean  $r_n$ :

$$Y_n^{\text{delay}} \sim \text{Poisson}(r_n)$$

Then the precorrected measurements (5.1) for the  $n$ th detector pair is :

$$Y_n = \sum_{k=1}^p N_{nk} + R_n + S_n - Y_n^{\text{delay}}.$$

**EM ALGORITHM:**

The complete-data space for EM algorithm [61] is the set of unobservable random variates

$$X^1 = \{\{N_{nk}\}_{k=1}^P, \{R_n\}, \{S_n\}, \{Y_n^{\text{delay}}\}\}_{n=1}^N$$

For this complete-data space, the conditional expectation of the log-likelihood of  $X^1$  (ignoring constant terms independent of  $\lambda$ ):

$$\begin{aligned} Q^1(\lambda; \lambda^i) &= \mathbb{E} \left\{ \log \left( p_{X^1}(X^1; \lambda) \mid Y = y; \lambda^i \right) \right\} \\ &= \sum_{n=1}^N \sum_{k=1}^P \left( -g_{nk} \lambda_k + \bar{N}_{nk} \log(g_{nk} \lambda_k) \right) \end{aligned}$$

where

$$\begin{aligned} \bar{N}_{nk} &= \mathbb{E} \left\{ N_{nk} \mid Y_n = y_n; \lambda^i \right\} \\ &= \mathbb{E} \left\{ \mathbb{E} \left\{ N_{nk} \mid Y_n = y_n, Y_n^{\text{delay}} = y_n^{\text{delay}}; \lambda^i \right\} \mid Y_n = y_n; \lambda^i \right\} \\ &= \frac{g_{nk} \lambda_k^i}{(\bar{y}_n(\lambda^i) + r_n)} \mathbb{E} \left\{ \sum_{k=1}^P N_{nk} + R_n + S_n \mid Y_n = y_n; \lambda^i \right\} \\ &= g_{nk} \lambda_k^i \frac{p_{Y_n}(y_n - 1; \lambda^i)}{p_{Y_n}(y_n; \lambda^i)} \end{aligned} \quad (5.21)$$

where  $\bar{y}_n(\lambda) = \sum_{k=1}^P g_{nk} \lambda_k + s_n^E$  as in (5.4) and  $p_{Y_n}(\cdot; \lambda)$  is the pmf of  $n$ th precorrected measurement as in (3.4) with  $\lambda$  as a parameter. The evaluation of the last conditional mean in the above expressions is derived in Appendix D.

The maximization of  $Q^1(\cdot; \lambda^i)$  analytically leads to the iterative ML-EM algorithm for  $\lambda = [\lambda_1 \dots \lambda_P]$ :

$$\lambda_k^{i+1} = \frac{\lambda_k^i}{\sum_{n=1}^N g_{nk}} \sum_{n=1}^N g_{nk} \frac{p_{Y_n}(y_n - 1; \lambda^i)}{p_{Y_n}(y_n; \lambda^i)}, \quad \text{for } i = 1, 2, \dots \quad (5.22)$$

As mentioned before the exact pmf  $p_{Y_n}(\cdot; \lambda)$  (3.4) contains infinite summations and it is computationally impractical. Thus we will plug in the previous approximations (OP, SP and SD approximations) for the exact pmf in the above algorithm.

**OP ML-EM Algorithm:**

The conventional OP assumption for the pre-corrected events (3.18) leads to the iterative update:

$$\begin{aligned} \lambda_k^{i+1} &= \frac{\lambda_k^i}{\sum_{n=1}^N g_{nk}} \sum_{n=1}^N g_{nk} \frac{p_{Y_n}^{\text{OP}}(y_n - 1; \lambda^i)}{p_{Y_n}^{\text{OP}}(y_n; \lambda^i)} \\ &= \frac{\lambda_k^i}{\sum_{n=1}^N g_{nk}} \sum_{n=1}^N g_{nk} \frac{(\bar{y}_n(\lambda^i))^{y_n-1} e^{-\bar{y}_n(\lambda^i)}}{(y_n - 1)!} \frac{y_n!}{(\bar{y}_n(\lambda^i))^{y_n} e^{-\bar{y}_n(\lambda^i)}} \\ &= \frac{\lambda_k^i}{\sum_{n=1}^N g_{nk}} \sum_{n=1}^N g_{nk} \frac{y_n}{\bar{y}_n(\lambda^i)}, \quad \text{for } i = 1, 2, \dots \end{aligned}$$

The above algorithm is the conventional ML-EM algorithm 2.18 [61], which is derived assuming that the measurements have Poisson distribution. As mentioned earlier in Section

5.5, randoms pre-correction can result in negative  $y_n$  values and one needs to zero threshold these values to guarantee the global convergence of the algorithm which results in the iterative update:

$$\lambda_k^{i+1} = \frac{\lambda_k^i}{\sum_{n=1}^N g_{nk}} \sum_{n=1}^N g_{nk} \frac{\lfloor y_n \rfloor_+}{\bar{y}_n(\lambda^i)}, \text{ for } i = 1, 2, \dots \quad (5.23)$$

**SP ML-EM Algorithm:**

Plugging in the  $p_{Y_n}^{\text{SP}}(\cdot; \lambda^i)$  (3.21) for the exact pmf in the EM algorithm (5.22) leads to SP ML-EM algorithm:

$$\begin{aligned} \lambda_k^{i+1} &= \frac{\lambda_k^i}{\sum_{n=1}^N g_{nk}} \sum_{n=1}^N g_{nk} \frac{p_{Y_n}^{\text{SP}}(y_n - 1; \lambda^i)}{p_{Y_n}^{\text{SP}}(y_n; \lambda^i)} \\ &= \frac{\lambda_k^i}{\sum_{n=1}^N g_{nk}} \sum_{n=1}^N g_{nk} \frac{\lfloor y_n + 2r_n \rfloor_+}{\bar{y}_n(\lambda^i) + 2r_n}, \text{ for } i = 1, 2, \dots \end{aligned} \quad (5.24)$$

which is very similar to previous ML-EM algorithms except the  $2r_n$  terms (which account for the shift in the model). Thus the computational requirements of the above SP ML-EM algorithm is approximately same with ML-EM.

**SD ML-EM Algorithm:**

Lastly, the SD approximation for the  $p_{Y_n}^{\text{SD}}(\cdot; \lambda^i)$  (3.29,3.31) for the exact pmf in the EM algorithm (5.22) leads to the iterative SD ML-EM algorithm:

$$\begin{aligned} \lambda_k^{i+1} &= \frac{\lambda_k^i}{\sum_{n=1}^N g_{nk}} \sum_{n=1}^N g_{nk} \frac{p_{Y_n}^{\text{SD}}(y_n - 1; \lambda^i)}{p_{Y_n}^{\text{SD}}(y_n; \lambda^i)} \\ &= \frac{\lambda_k^i}{\sum_{n=1}^N g_{nk}} \sum_{n=1}^N g_{nk} \frac{y_n + \text{sign}\{y_n\} + u_n(\lambda^i; y_n)}{2(\bar{y}_n(\lambda^i) + r_n)} F(\lambda^i; y_n), \text{ for } i = 1, 2, \dots \end{aligned}$$

where  $u_n(\lambda; k) = \sqrt{(|k| + 1)^2 + 4(\bar{y}_n(\lambda) + r_n)r_n}$  and  $F(\lambda; k) = \frac{e^{u_n(\lambda; k-1)}}{e^{u_n(\lambda; k)}} \sqrt{\frac{u_n(\lambda; k)}{u_n(\lambda; k-1)}}$ .

As shown previously SD approximation agrees with the exact pmf better than the other models. Moreover as EM algorithm simultaneously updates the parameters, the above update equation is only slightly more expensive than the previous update equations in terms of computation requirements.

The above methods can be extended to the SAGE type maximization methods [41]. We originally planned to use the fast SAGE maximization method for our 2-D reconstructions. However, recently a method which is even faster than SAGE was introduced by Fessler and Erdoğan [38] for emission reconstructions, called paraboloid surrogates maximization algorithm. Thus in our reconstruction studies we used this method which we will describe next.

### 5.6.2 Paraboloid Surrogates Maximization Algorithm

Similar to the surrogates function idea in grouped coordinate ascent method of Section 4.7, one maximizes a surrogate function (which is parabolic) in the paraboloid surrogates coordinate ascent (PSCA) method [38]. Namely at the  $i$ th iteration we find:

$$\lambda^{i+1} = \arg \max_{\lambda \geq 0} Q(\lambda; \lambda^i) - \beta R(\lambda). \quad (5.25)$$

The surrogate function  $Q(\lambda; \lambda^i)$  is composed such that the log-likelihood function  $L(\lambda)$  converges to the true maximizer. This is achieved by forming a summation of 1-D surrogate functions as [38]:

$$Q(\lambda; \lambda^i) \triangleq \sum_{n=1}^N q_n([\mathbf{G}\lambda]_n; [\mathbf{G}\lambda^i]_n) \quad (5.26)$$

with  $[\mathbf{G}\lambda]_n = \sum_j g_{nj}\lambda_j$ . Since  $h_n$ 's are concave and their first derivatives are convex for the OP, SP and PR methods, one can use the following parabolic surrogate functions [38]:

$$q_n(l; l_n^i) = h_n(l_n^i) + \dot{h}_n(l_n^i)(l - l_n^i) + -\frac{1}{2}n_n(l_n^i)(l - l_n^i)^2, \quad (5.27)$$

with

$$n_n(l) = \begin{cases} \frac{2}{l} [\dot{h}_n(l) - h_n(0) - l\ddot{h}_n(l)], & l > 0, \\ -\ddot{h}_n(l), & l = 0. \end{cases} \quad (5.28)$$

Then we use the fast coordinate ascent method (4.30) for maximizing the parabolic function  $Q$ .

For the SD method: the results from Appendix E show that  $\ddot{h}^{\text{SD}}(l) < 0$  for  $l \in [0, \infty)$ , proving that  $h^{\text{SD}}(l)$  is *strictly concave*.

Also, the sign of  $h^{\text{SD}(3)}(l)$  for different regions of  $l$  is summarized below for convenience:

$$h^{\text{SD}(3)}(l) = \begin{cases} > 0, & y \leq -2 \\ < 0, & y = -1, r \leq \sqrt{x_o^2 - 1}, l < l_2^* \\ = 0, & y = -1, r \leq \sqrt{x_o^2 - 1}, l = l_2^* \\ > 0, & y = -1, l > \max[0, l_2^*] \\ < 0, & y = 0, r \leq \sqrt{7}/6, l < l_1^* \\ = 0, & y = 0, r \leq \sqrt{7}/6, l = l_1^* \\ > 0, & y = 0, l > \max[0, l_1^*] \\ > 0, & y \geq 1 \end{cases} \quad (5.29)$$

with  $x_o$ ,  $l_1^*$  and  $l_2^*$  as defined in (E.60), (E.34) and (E.64) respectively.

Thus,  $\dot{h}^{\text{SD}}(l)$  is convex for  $l \in [0, \infty)$  for

$$\mathbf{A)} \quad y \leq -2 \quad (5.30)$$

$$\mathbf{B)} \quad y = -1, r \geq \sqrt{x_o^2 - 1} \quad (5.31)$$

$$\mathbf{C)} \quad y = 0, r \geq \frac{\sqrt{7}}{6} \quad (5.32)$$

$$\mathbf{D)} \quad y \geq 1 \quad (5.33)$$

For the cases (**A** – **D**) one can use the optimum curvature for the paraboloid surrogate functions. For the remaining cases we use the maximum curvature  $\max_{l \in [0, \infty)} \{\ddot{h}_n^{\text{SD}}(l)\}$ . Based on the generalized mean value theorem for twice differentiable functions, the maximum curvature ensures monotonicity [14, 28].

Thus, at each iteration the paraboloid surrogate function to be maximized is defined as follows:

$$q_n(l; l_n^i) = h_n^{\text{SD}}(l_n^i) + \dot{h}_n^{\text{SD}}(l_n^i)(l - l_n^i) + -\frac{1}{2}n_n(l_n^i)(l - l_n^i)^2, \quad (5.34)$$

with

$$n_n(l) = \begin{cases} -\ddot{h}_n^{\text{SD}}(l_2^*), & y = -1, r \leq \sqrt{x_o^2 - 1}, \\ -\ddot{h}_n^{\text{SD}}(l_1^*), & y = 0, r \leq \frac{\sqrt{7}}{6}, \\ t_n(l), & \text{else} \end{cases} \quad (5.35)$$

and

$$t_n(l) = \begin{cases} \frac{2}{l^2} [h_n^{\text{SD}}(l) - h_n^{\text{SD}}(0) - l\dot{h}_n^{\text{SD}}(l)], & l > 0 \\ -\dot{h}_n^{\text{SD}}(l), & l = 0. \end{cases} \quad (5.36)$$

## 5.7 1D Simulations

To analyze the performance of the approximations first we performed 1D simulations, *i.e.*:  $\lambda$  scalar and  $P = 1$ . In the simulations the total number of *true* counts and *randoms* counts were fixed as  $\sum_{n=1}^N g_n \lambda = 100$  and  $\sum_{n=1}^N r_n = 50$ , and the  $g_n$  and  $r_n$  were constants, and  $\lambda_{\text{true}} = 1$ . Fig. 5.1 shows the computed sample mean values (from 300 realizations) of different estimators as a function of number of bins  $N$ . It is observed that as the number of bins gets larger, *i.e.*: as the number of counts per bin gets smaller, both the OP and SP method results in positive bias. This bias is due to the zero-thresholding of the data in (5.12) and (5.13). Zero-thresholding increases the mean value of the data and this results in a positive systematic bias since the data is linearly related to radio-isotope concentration  $\lambda$ . Fig. 5.2 shows the sample standard deviation of different estimators as a function of the number of bins.

We performed additional 1D simulations with nonuniform  $g_n$  and  $r_n$  values as well. Also, we implemented the exact log-likelihood using two different methods: first we implemented an extremely precise approximation to the exact log-likelihood based on the truncation of the infinite summations (5.8) and also we implemented the exact log-likelihood using Bessel functions (3.8). Fig. 5.3 shows the sample mean of each estimator for a total number of 100 true counts and 100 random counts. It is seen that both the SD method and the exact log-likelihood results in bias free estimate independent of the number of counts per bin. Fig. 5.4 shows the sample standard deviation of each estimator.

Lastly we performed simulations with noise free data for the same number of total counts per bin. It can be seen from Fig. 5.5 that the *noise free* data with *fractional* counts results in negative bias both for the SD and the exact log-likelihood methods. This result shows that statistical methods designed for noisy data may results in bias when applied to noise free data with low fractional counts per bin.

## 5.8 2D Simulations

To study bias and variance properties of the estimators based on the above approximations, we performed 2D simulations. For  $\lambda$  we used the synthetic emission phantom shown in Fig. 5.6. The spine, lungs, soft tissue, and heart had relative radioactivity concentrations of 0, 1, 2 and 4 respectively. The sinograms had 200 radial bins and 300 angles uniformly sampled over 180 degrees. We approximated the system geometry with 2.8 mm wide strip integrals and 2.8 mm ray spacing. The reconstructed images were 64 by 64 with 9 mm pixels. The  $r_n$  factors corresponded to a uniform field of 50% random coincidences.

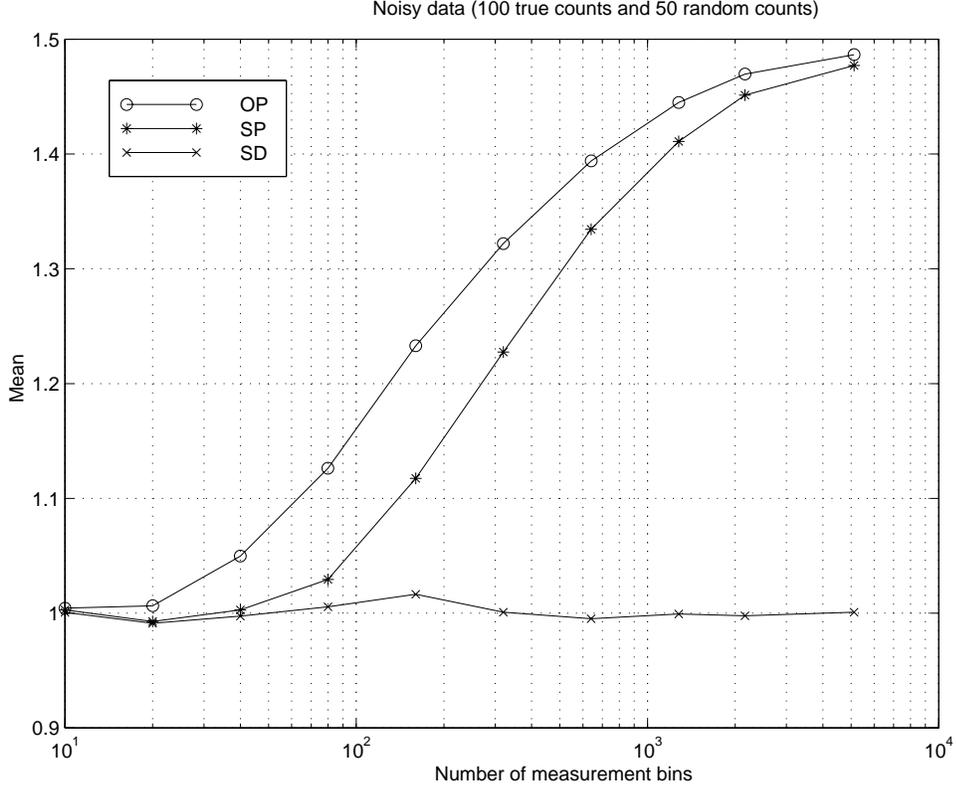


Figure 5.1: Sample mean of OP, SP and SD methods from 300 realizations where  $\lambda_{true} = 1$ .

We generated 300 pseudo-random emission measurements according to (5.2) and (5.3). For each realization, an estimate of the emission phantom was reconstructed using 30 iterations of the paraboloid surrogates algorithm [28, 38] applied to objective functions (5.12), (5.13) and (5.14).

For regularization, we used the modified quadratic penalty [43] introduced in Section 2.4.2. This penalty improves the resolution uniformity and enables matching of the spatial resolutions of different methods. When we used the practical implementation (2.23) we observed some resolution non-uniformity in the reconstructed images and some artifacts at the edges, especially for low count simulations. To overcome this problem, we re-projected the initial FBP reconstructed image to obtain improved estimates of  $\bar{Y}(\check{\theta})$  in [43] and also approximated  $\bar{Y}(\check{\theta})$  with  $\bar{Y}(\check{\theta})$ . This method resulted in artifact free and uniform resolution images. We matched the resolution of the reconstructed images for all methods to 1.9 pixels FWHM.

Since in these simulations we had access to  $Y_n^{\text{prompt}}$  and  $Y_n^{\text{delay}}$  values separately, we also performed conventional penalized maximum likelihood reconstruction with prompt counts (PR) for comparison purposes. In the PR case the data is not precorrected for randoms and we have access to  $r_n$  values separately. Thus, this method is expected to perform better than the randoms-precorrected methods<sup>1</sup>. We include this method in our simulations for

<sup>1</sup>In these simulations PR method is the idealized method where we assume that one has access to the means of randoms rates (*i.e.*:  $r_n$ ), but in practice one needs to estimate these quantities from noisy measurements of delayed windows. Thus, the results reported here with PR method shows the upper bounds

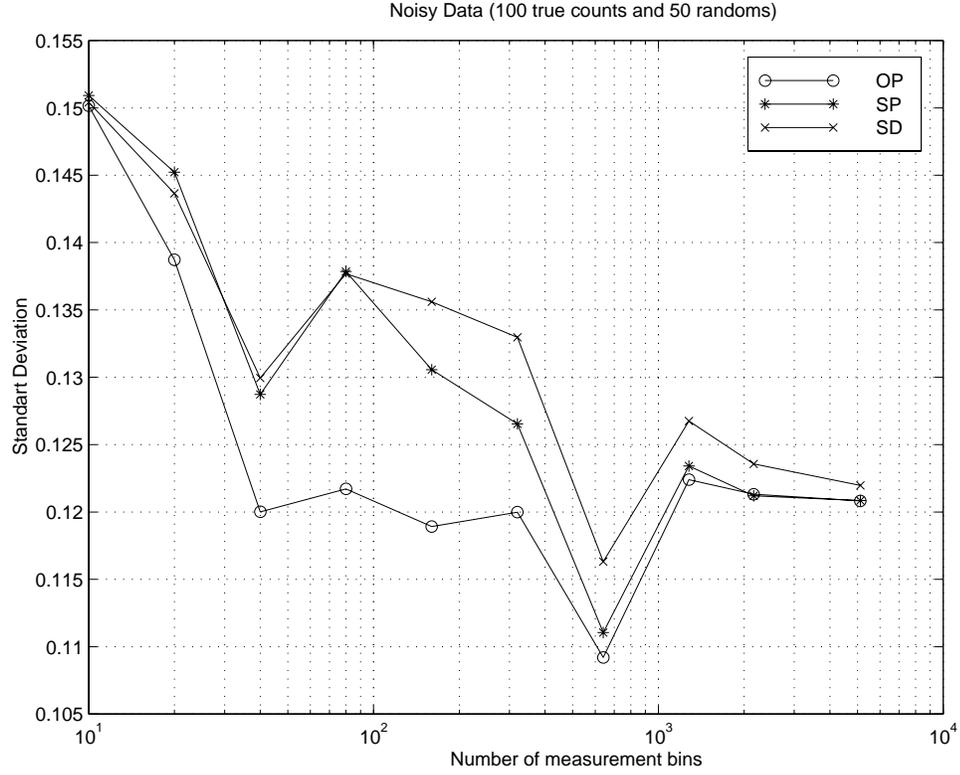


Figure 5.2: Sample standard deviation of OP, SP and SD methods from 300 realizations where  $\lambda_{true} = 1$ .

comparison purposes only.

Fig. 5.7 and 5.8 show the sample mean and standard deviation images of different methods for a total of 50,000 counts. And, Fig. 5.9 and 5.10 show the profiles through the sample mean and standard deviation images. Also, Fig. 5.11 shows the histogram of bias of different methods compared to the PR method (difference of the sample mean from the sample mean of the PR method) and Fig. 5.12 shows the histogram of the ratio of the standard deviation of different methods with respect to the PR method. The OP method results in severe bias and the SP results in some bias in the reconstructed images. However, the SP and SD methods yield similar amount of standard deviations.

---

on the performance of the PR method.

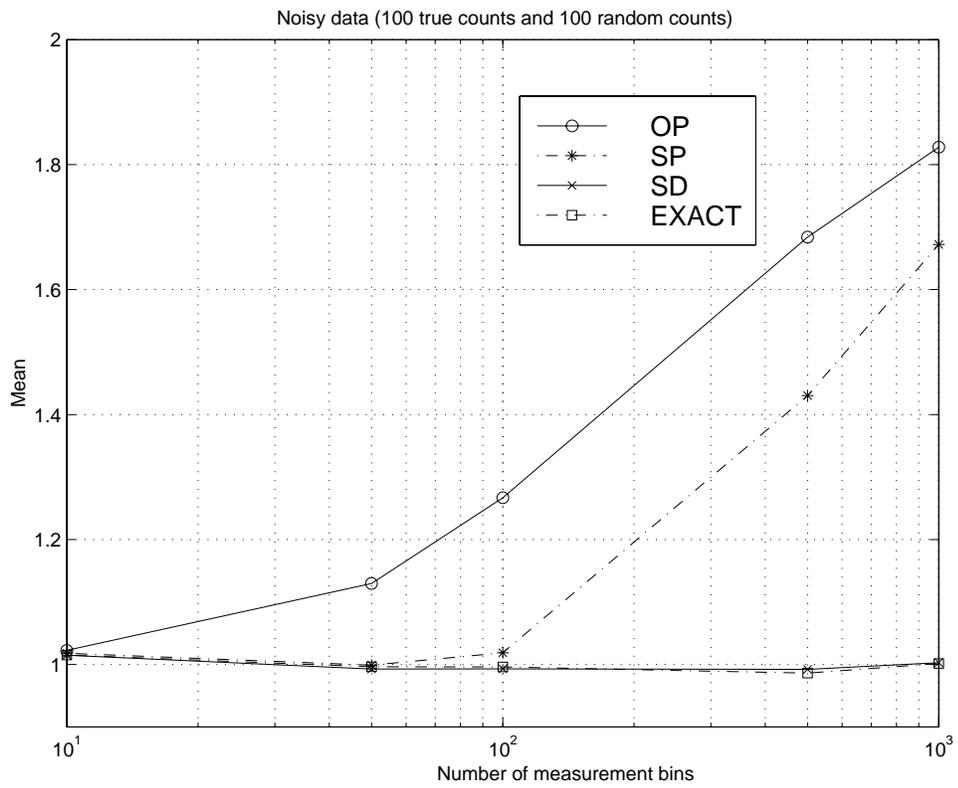


Figure 5.3: Sample mean of OP, SP, SD and Exact methods from 300 realizations (with nonuniform  $g_n$  and  $r_n$ ) where  $\lambda_{true} = 1$ .

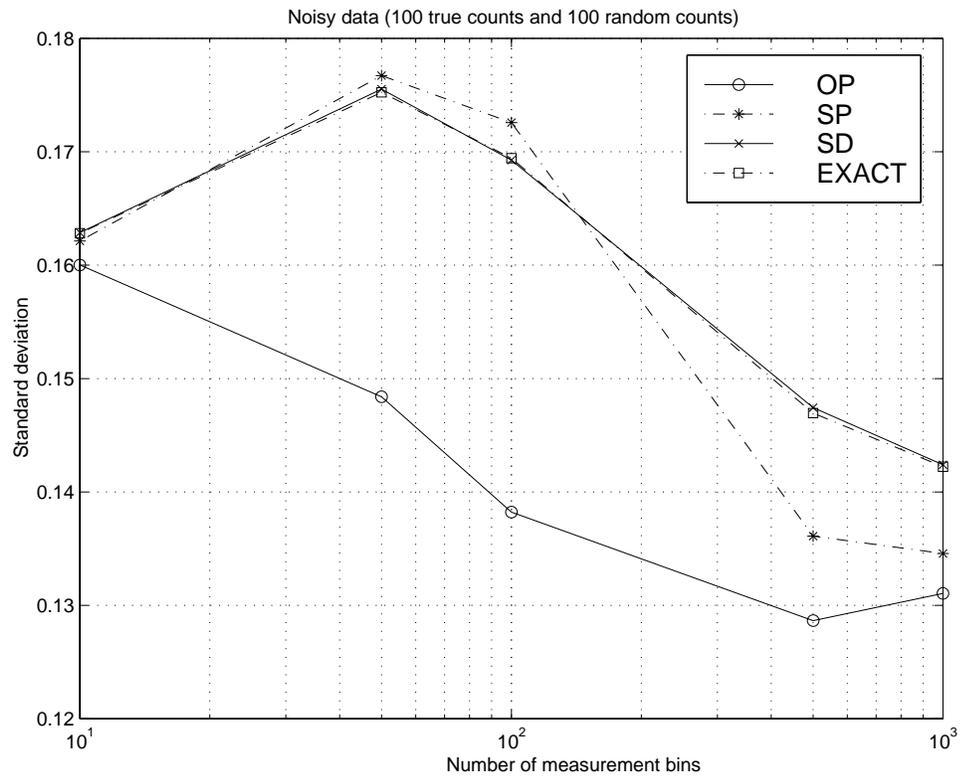


Figure 5.4: Sample standard deviation of OP, SP, SD and Exact methods from 300 realizations (with nonuniform  $g_n$  and  $r_n$ ) where  $\lambda_{true} = 1$ .

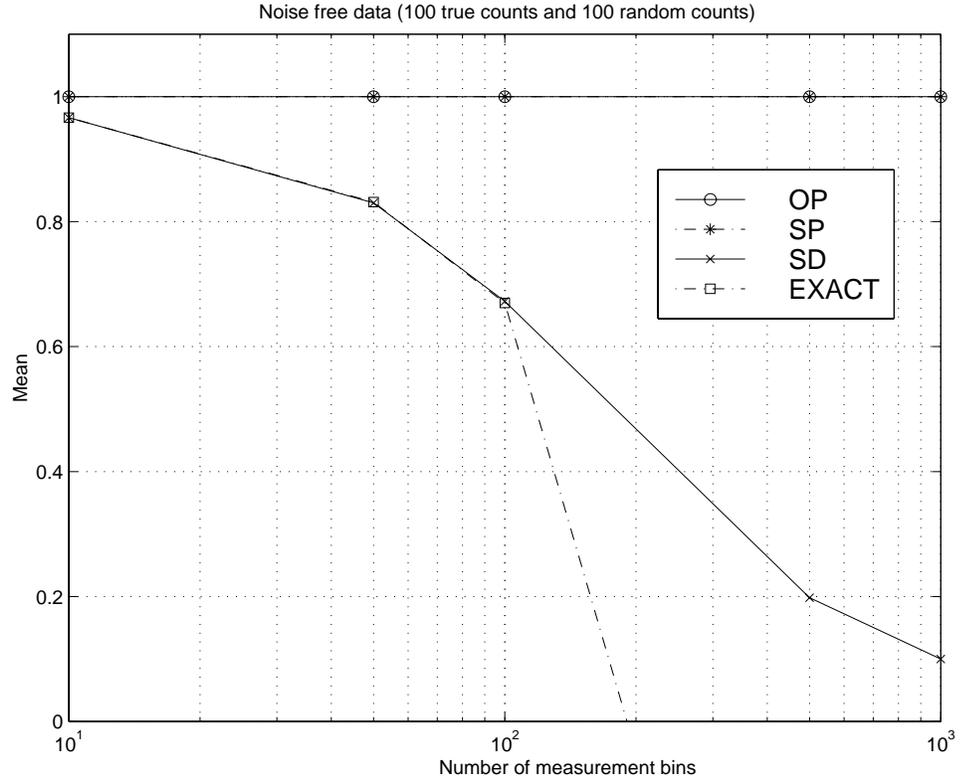


Figure 5.5: Results of OP, SP, SD and Exact methods with *noise free* data (with nonuniform  $g_n$  and  $r_n$ ) where  $\lambda_{true} = 1$ .

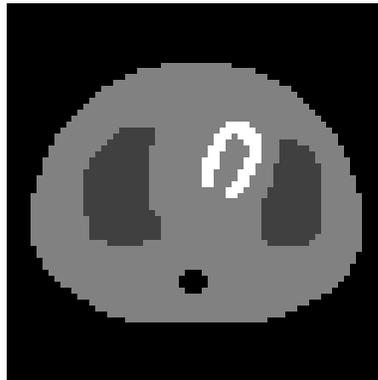


Figure 5.6: Simulated emission phantom.

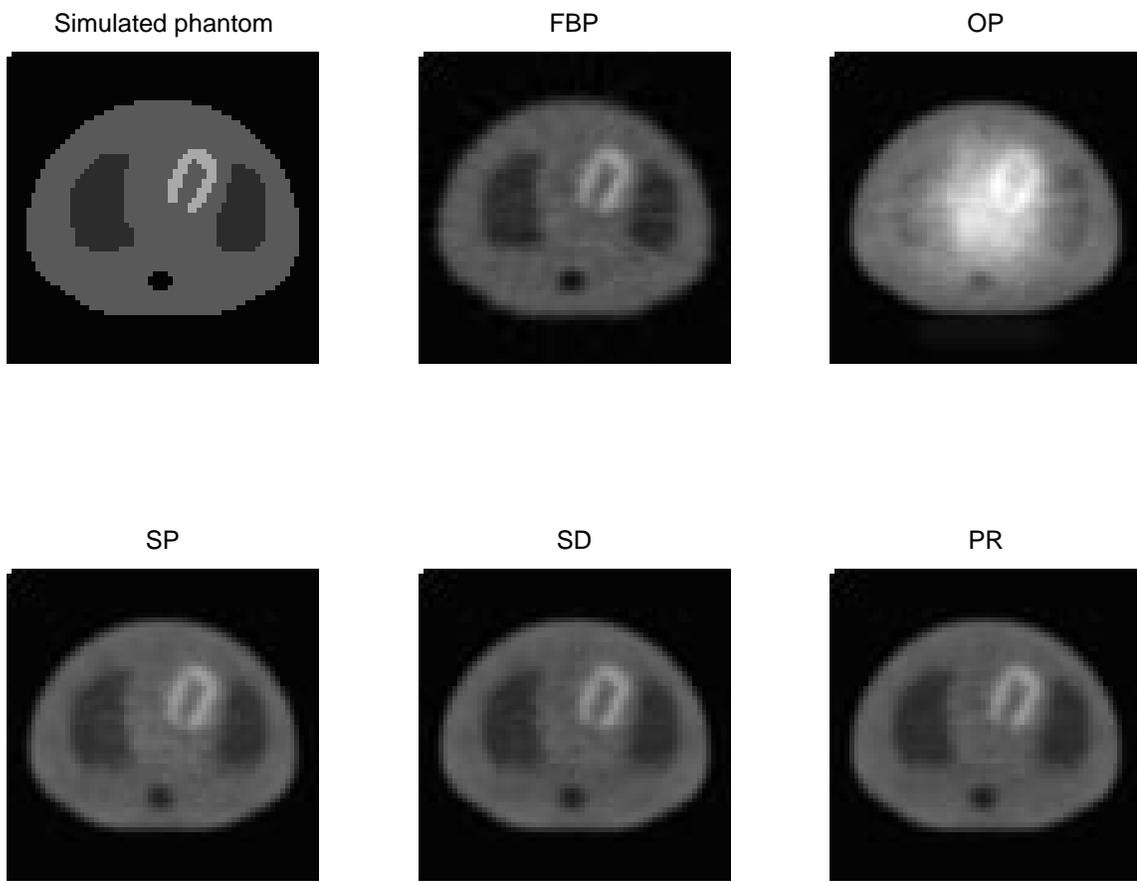


Figure 5.7: Sample mean images of different methods from 300 realization with 50,000 counts per scan.

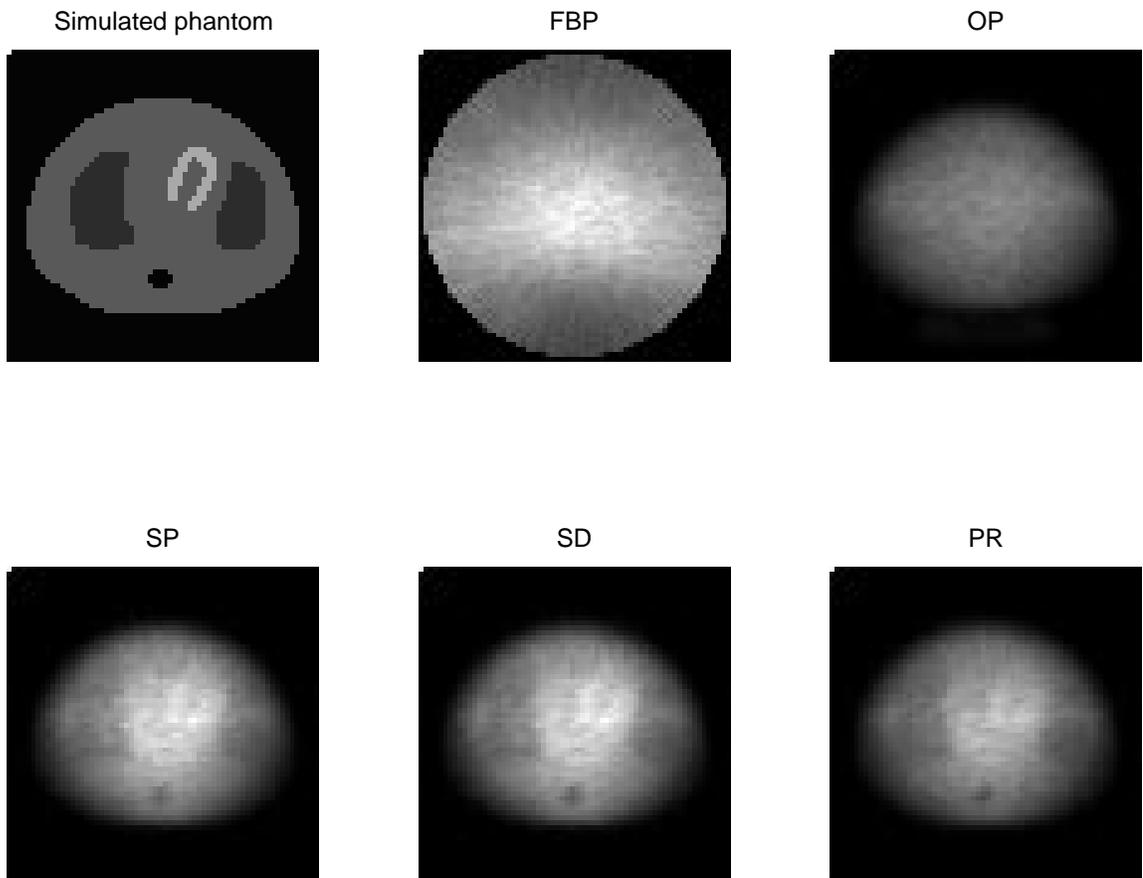


Figure 5.8: Sample standard deviation images of different methods from 300 realization with 50,000 counts per scan.

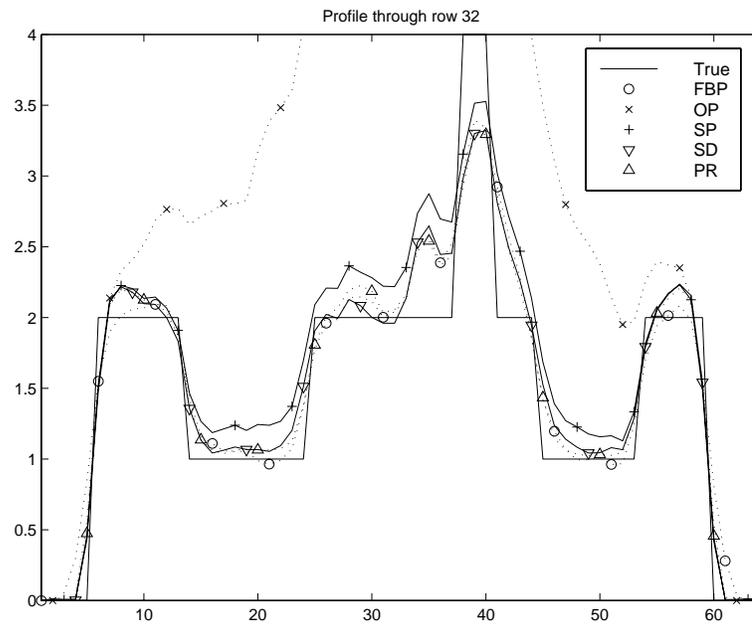


Figure 5.9: Profile through the sample mean images of different methods from 300 realization with 50,000 counts per scan.

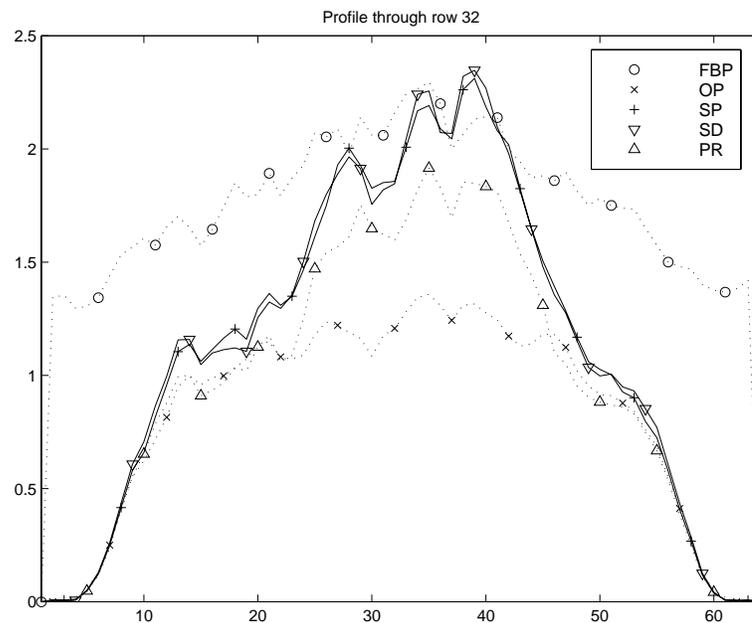


Figure 5.10: Profile through the sample standard deviation images of different methods from 300 realization with 50,000 counts per scan.

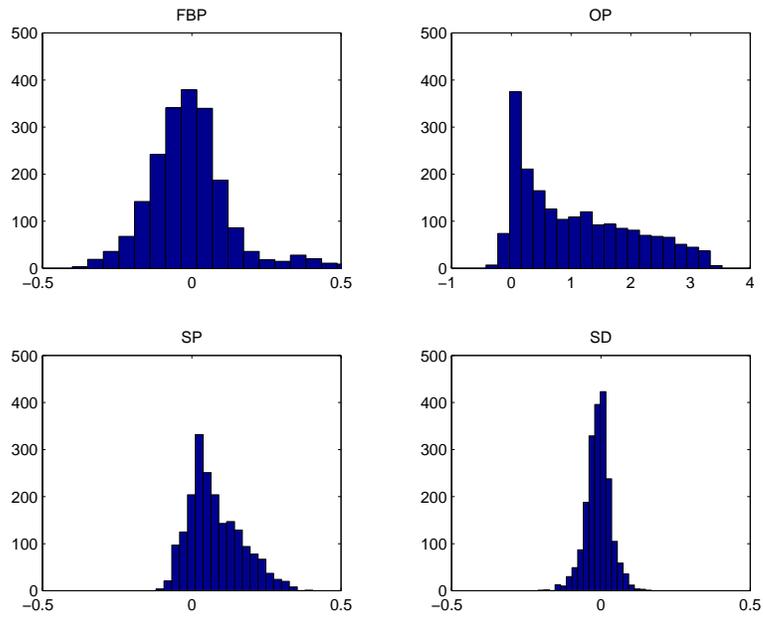


Figure 5.11: Histogram of the bias of different methods compared to PR method with 50,000 counts per scan.

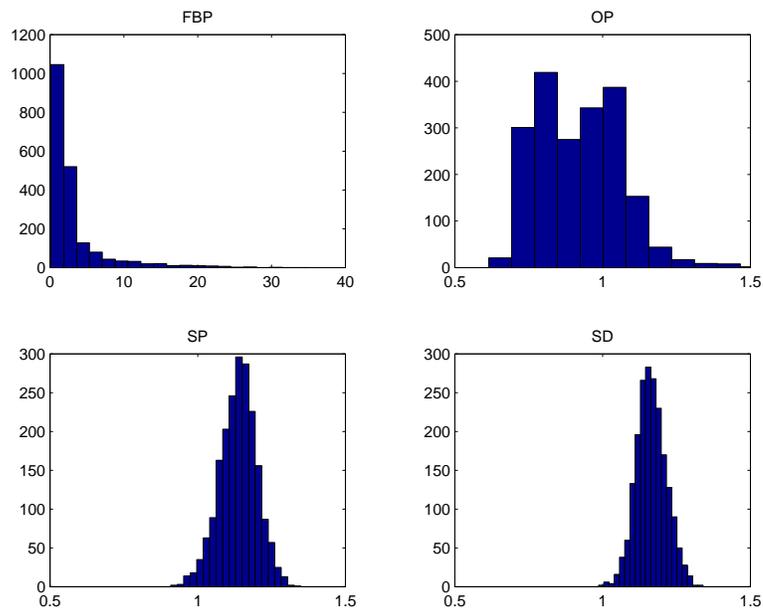


Figure 5.12: Histogram of the ratio of standard deviation of different methods to the standard deviation of PR method with 50,000 counts per scan.

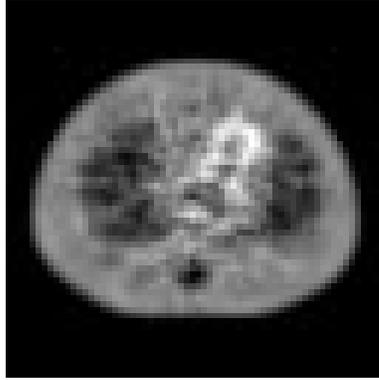


Figure 5.13: Reconstructed emission image (SD method) from 500,000 counts.

Also, Figs. 5.13 to 5.26 show the simulation results for 0.5 million and 5 million counts per scan. For 0.5 million counts per scan, the OP method still causes systematic bias. Moreover, as can be seen from histogram of standard deviation ratios in Fig. 5.19, on the average the OP method results in more standard deviation than both the SP and the SD methods. For 5 million counts per scan, all methods are free of systematic bias but the OP method results in larger standard deviation than both the SP and the SD methods.

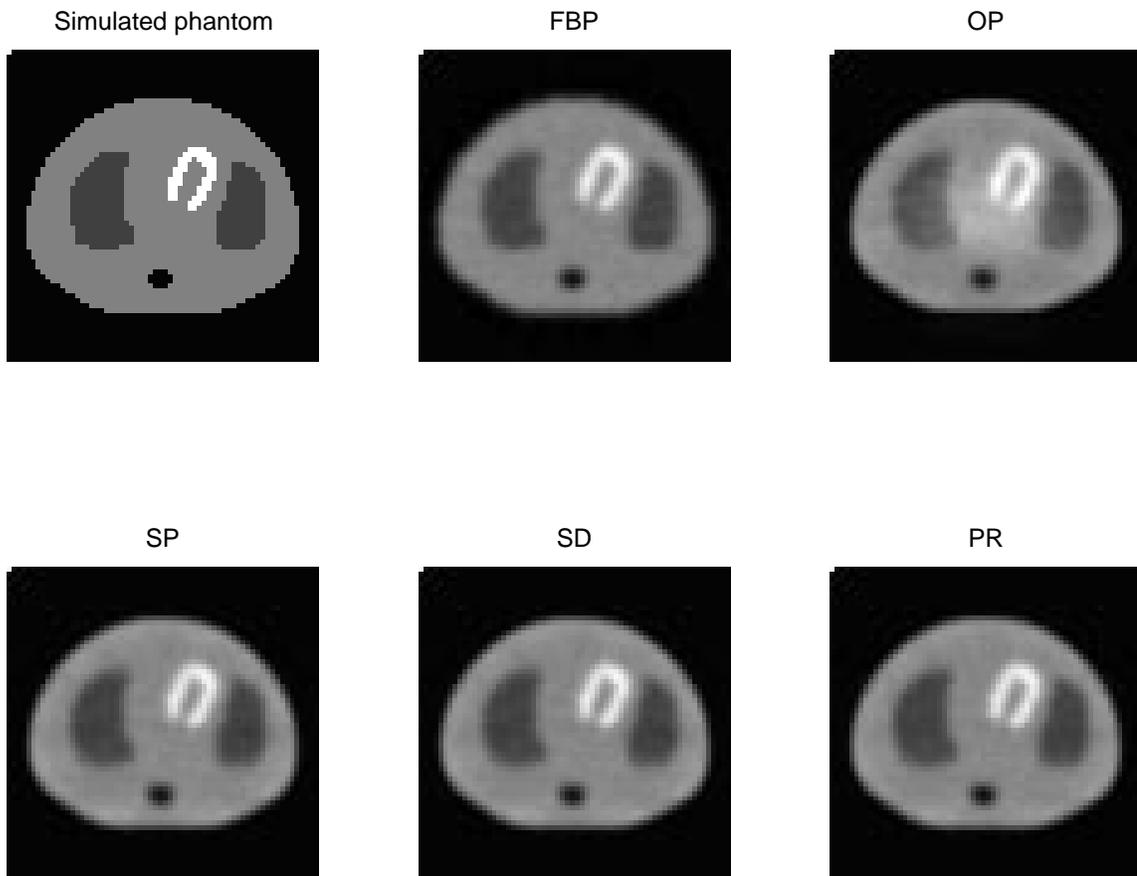


Figure 5.14: Sample mean images of different methods from 300 realization with 500,000 counts per scan.

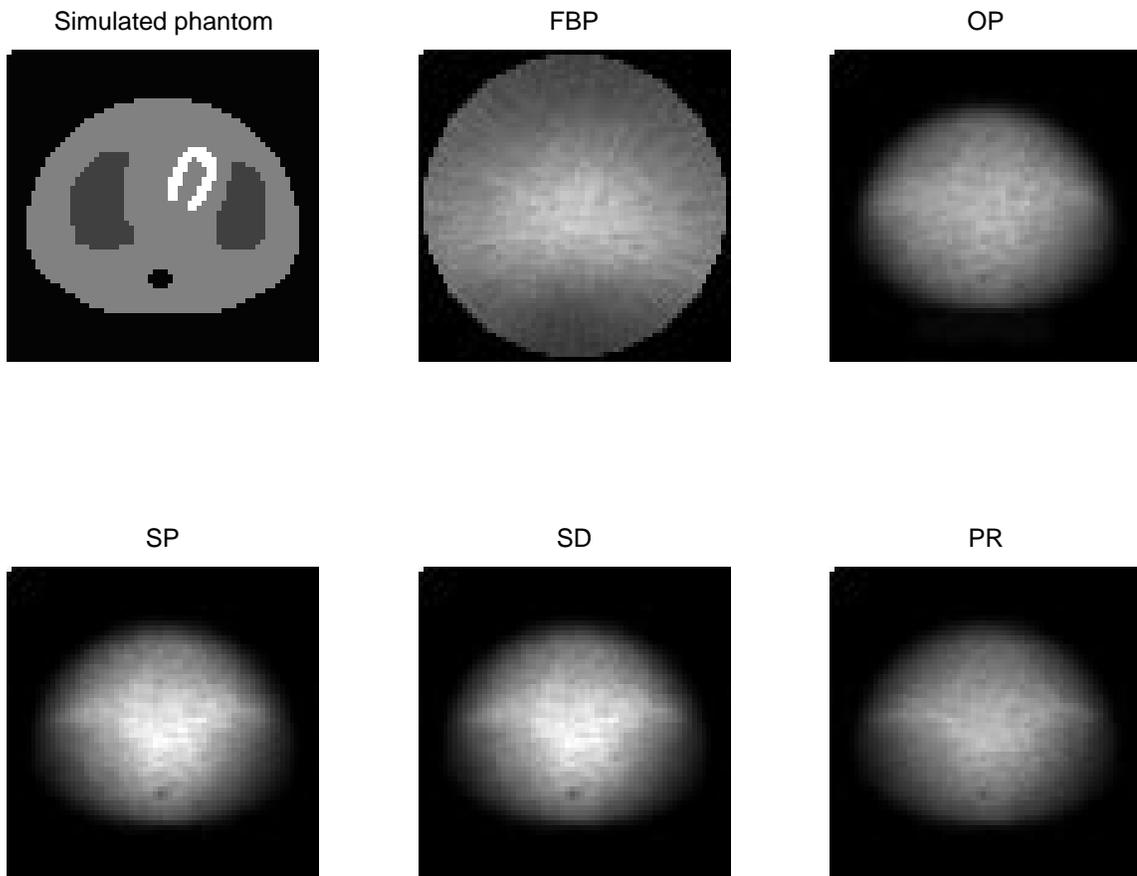


Figure 5.15: Sample standard deviation images of different methods from 300 realization with 500,000 counts per scan.

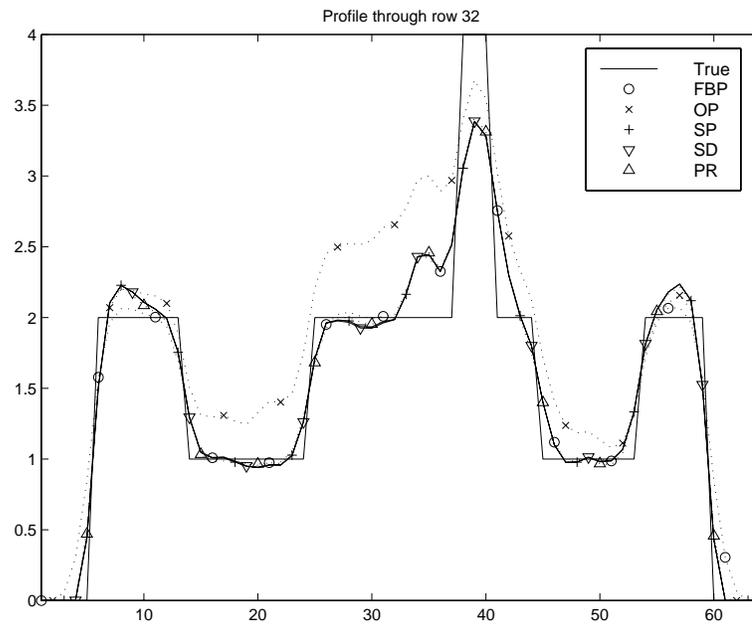


Figure 5.16: Profile through the sample mean images of different methods from 300 realization with 500,000 counts per scan.

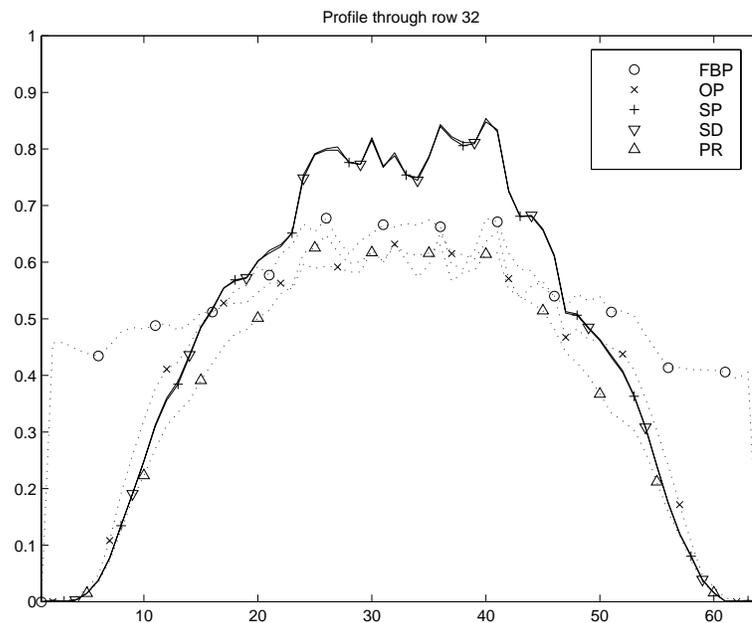


Figure 5.17: Profile through the sample standard deviation images of different methods from 300 realization with 500,000 counts per scan.

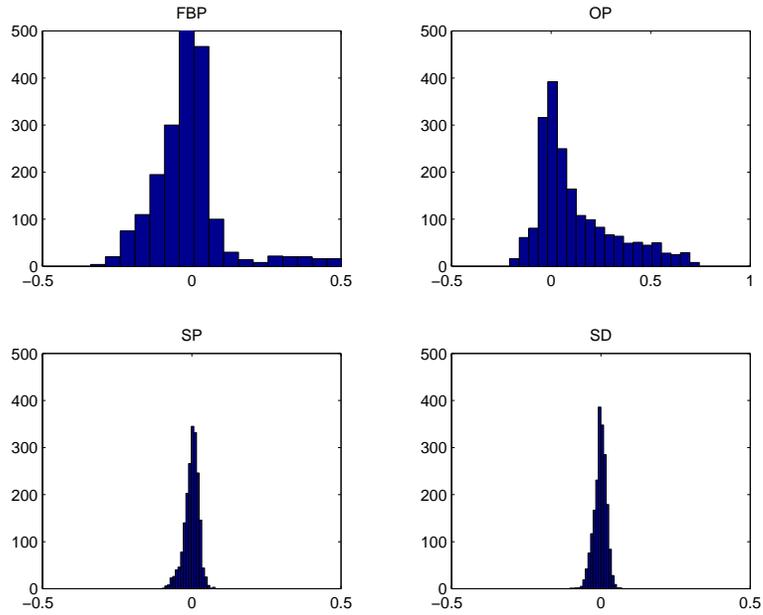


Figure 5.18: Histogram of the bias of different methods compared to PR method with 500,000 counts per scan.

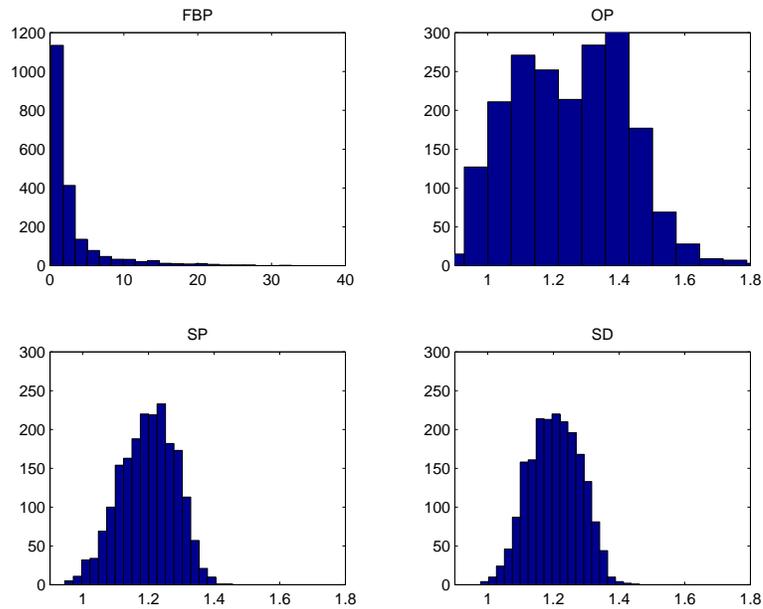


Figure 5.19: Histogram of the ratio of standard deviation of different methods to the standard deviation of PR method with 500,000 counts per scan.

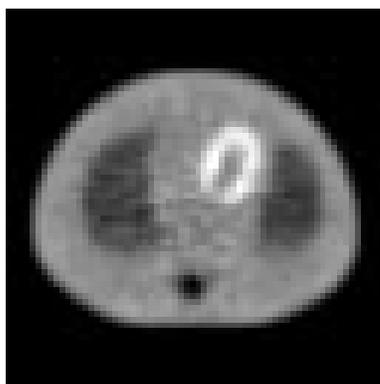


Figure 5.20: Reconstructed emission image (SD method) from 5,000,000 counts per scan.

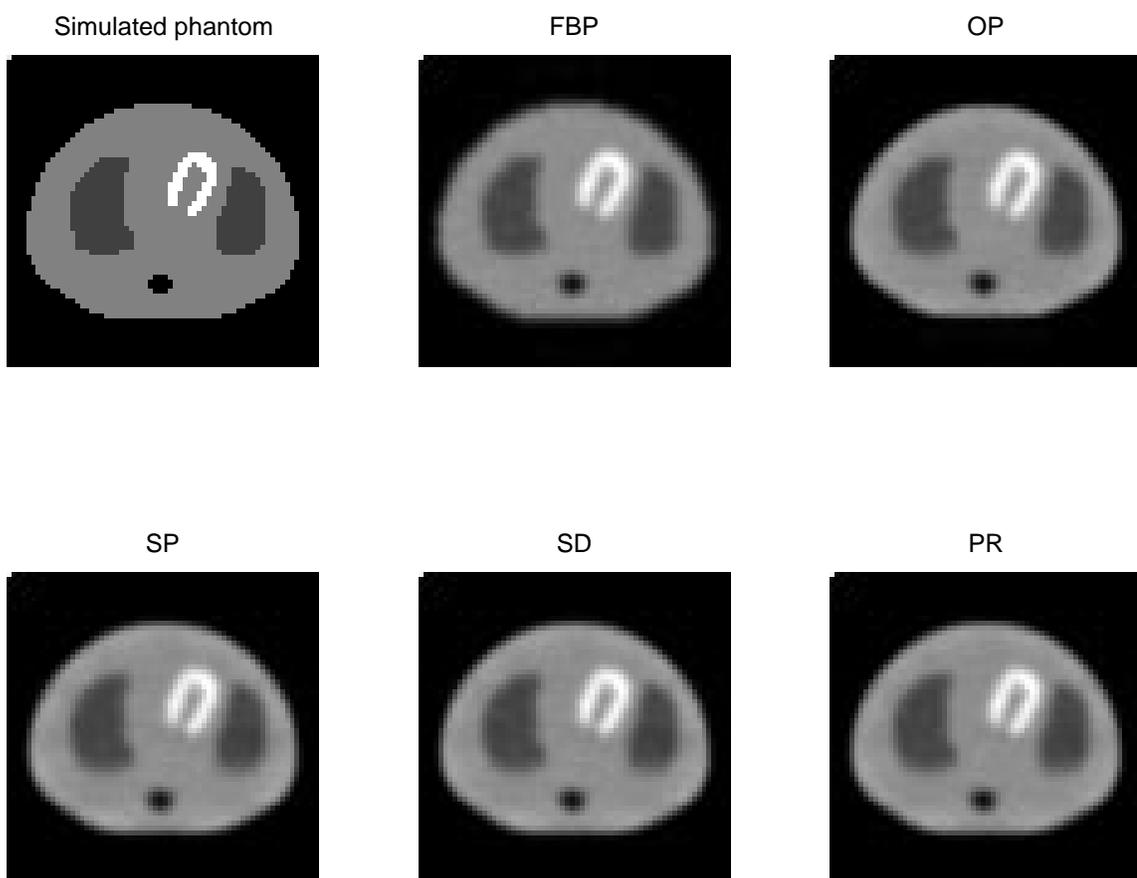


Figure 5.21: Sample mean images of different methods from 100 realization with 5,000,000 counts per scan.

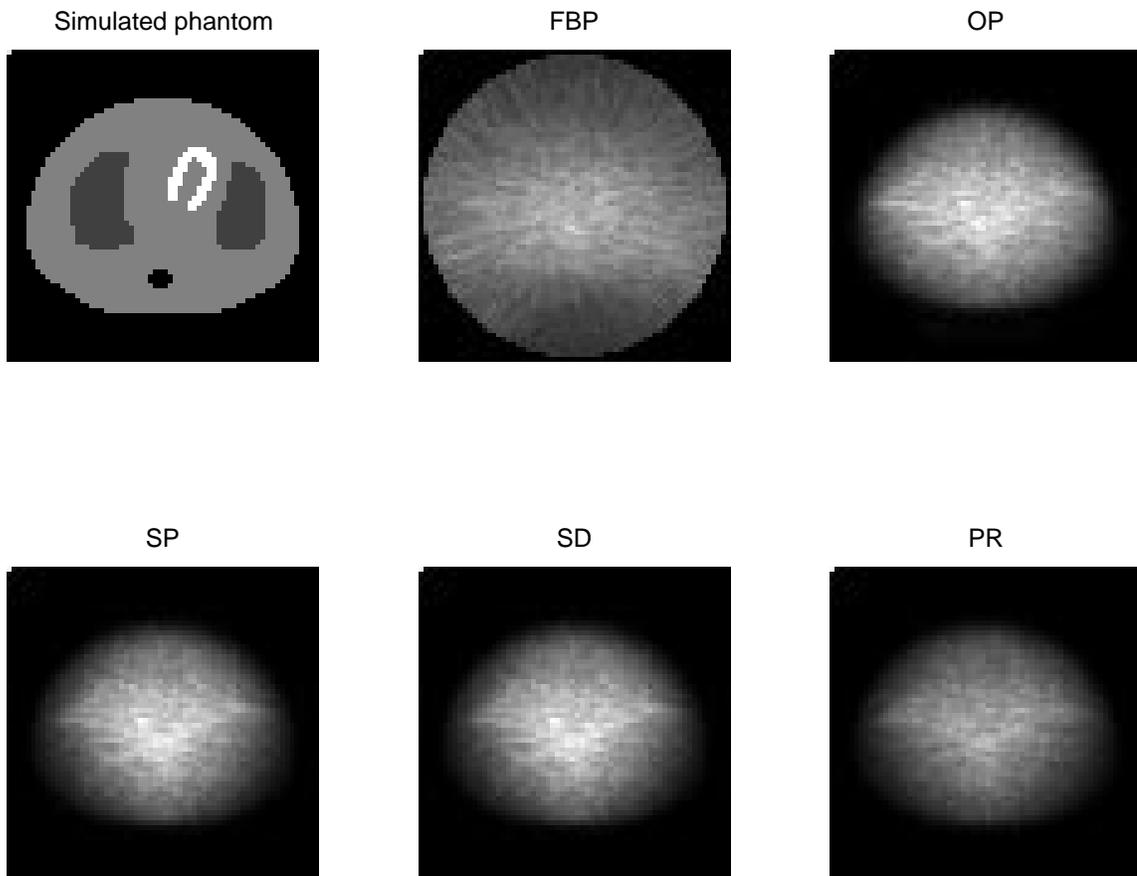


Figure 5.22: Sample standard deviation images of different methods from 100 realization with 5,000,000 counts per scan.

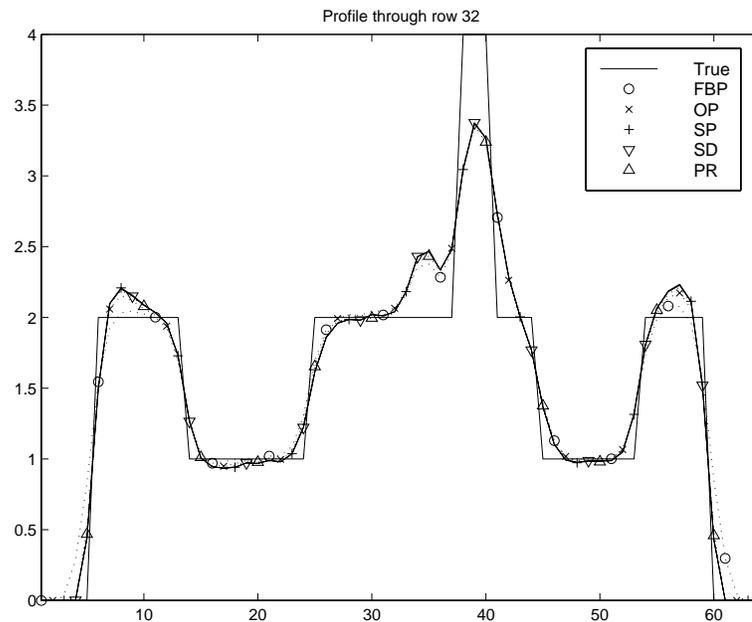


Figure 5.23: Profile through the sample mean images of different methods from 100 realization with 5,000,000 counts per scan.

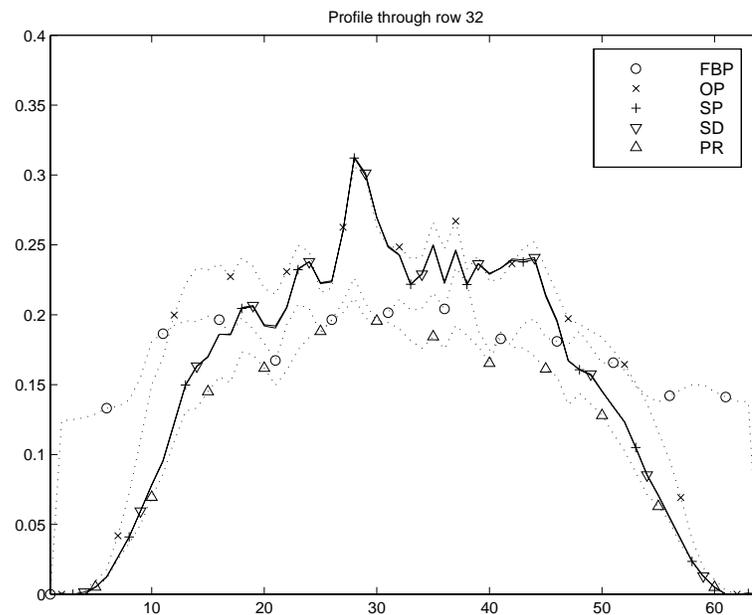


Figure 5.24: Profile through the sample standard deviation images of different methods from 100 realization with 5,000,000 counts per scan.

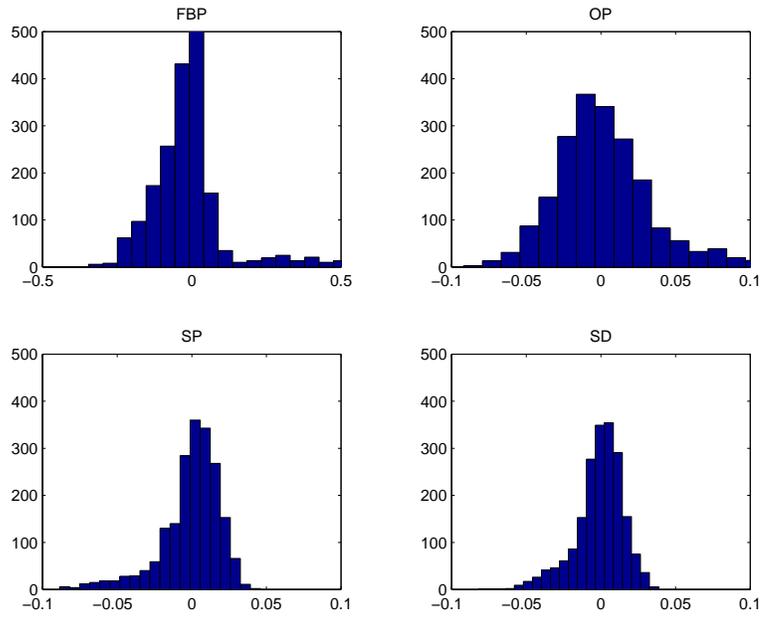


Figure 5.25: Histogram of the bias of different methods compared to PR method with 5,000,000 counts per scan.

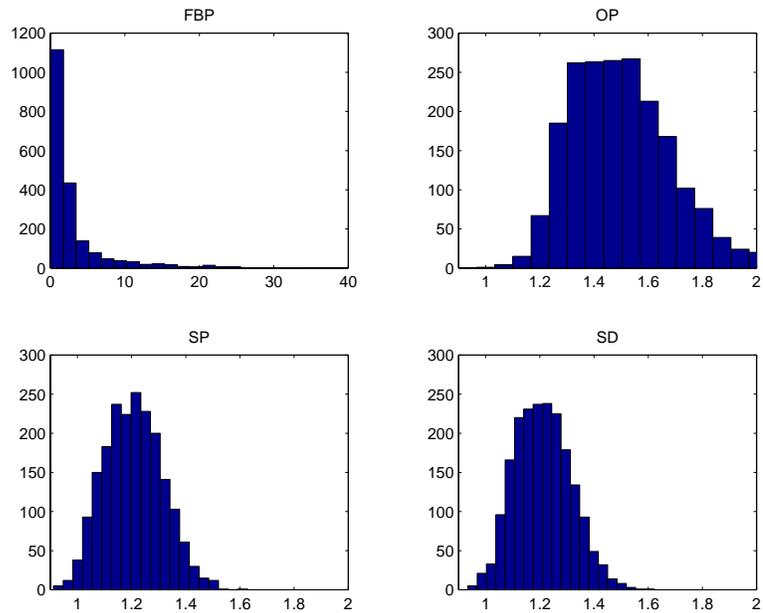


Figure 5.26: Histogram of the ratio of standard deviation of different methods to the standard deviation of PR method with 5,000,000 counts per scan.

## 5.9 Cramer-Rao Bounds

To study bias-variance trade-offs of the proposed methods and to see how close they perform to achievable bounds, we compared the performance of the proposed methods to the uniform Cramer-Rao bound [91]. Unlike conventional Cramer-Rao bounds, uniform Cramer-Rao bounds are applicable to biased estimators with unknown bias gradient length. For this purpose we compare 1D empirical results of the OP, SP, SD and exact methods to the 1D Cramer-Rao bounds. Note that for the 1D problem bias gradient length is relatively easy to interpret.

For the 1D estimation problem, *i.e.*:  $\lambda$  scalar and  $P = 1$ , the Cramer-Rao (CR) bound [93] on the variance of unbiased estimator  $\hat{\lambda}$  is given by:

$$\text{Var} \left\{ \hat{\lambda} \right\} \geq F_Y^{-1}, \quad (5.37)$$

where  $F_Y$  is the Fisher information matrix

$$F_Y \triangleq \text{E} \left\{ \left( \frac{\partial}{\partial \lambda} L(Y, \lambda) \right)^2 \right\}. \quad (5.38)$$

However, the CR lower bound is only applicable to unbiased estimators. Although, there is a *biased* CR bound [93] applicable to biased estimators, it is only applicable to estimators with a given bias gradient. A more general form of CR bound called as uniform CR bound has been developed [51, 92] that applies to all biased estimators whose bias gradient length satisfies:

$$\left| \frac{\partial}{\partial \lambda} b(\lambda) \right|^2 \leq \delta^2 < 1, \quad (5.39)$$

with bias defined as  $b(\lambda) = \text{E} \left\{ \hat{\lambda} \right\} - \lambda$ . A more general form of the following theorem is proven in [51].

*Theorem 1:* Let  $\hat{\lambda}$  be an estimator with bias  $b(\lambda)$  whose bias gradient satisfies (5.39). For nonzero  $F_Y$ , the variance of  $\hat{\lambda}$  is bounded as:

$$\text{Var} \left\{ \hat{\lambda} \right\} \geq B(\lambda, \delta), \quad (5.40)$$

where  $B(\lambda, \delta)$  is:

$$B(\lambda, \delta) = \kappa^2 \frac{F_Y}{(1 + F_Y)^2} \quad (5.41)$$

and  $\kappa$  is determined by the unique positive solution of :

$$g(\kappa) = \frac{1}{(1 + \kappa F_Y)^2} = \delta^2. \quad (5.42)$$

By tracing out the family of points  $(\delta, \sqrt{B(\lambda, \delta)})$  one can obtain a curve in the bias gradient ( $\delta$ ) - standard deviation ( $\sigma$ ) plane. This tracing can be achieved by continuously varying  $\kappa$  over the range  $(0, \infty)$  and plotting the curve using (5.41) and (5.42). Since  $B(\lambda, \delta)$  is a lower bound on  $\text{Var} \left\{ \hat{\lambda} \right\} = \sigma_{\hat{\lambda}}^2$ , below the curve defines the *unachievable* region where no realizable estimator exist. Figure 5.27 shows a  $\delta - \sigma$  tradeoff curve [51] plotted in terms of normalized standard deviation  $\sigma = \sqrt{\frac{B(\lambda, \delta)}{B(\lambda, 0)}}$ . If an estimator lies on the curve then lower variance can only be achieved at the price of increased bias gradient and vice versa.

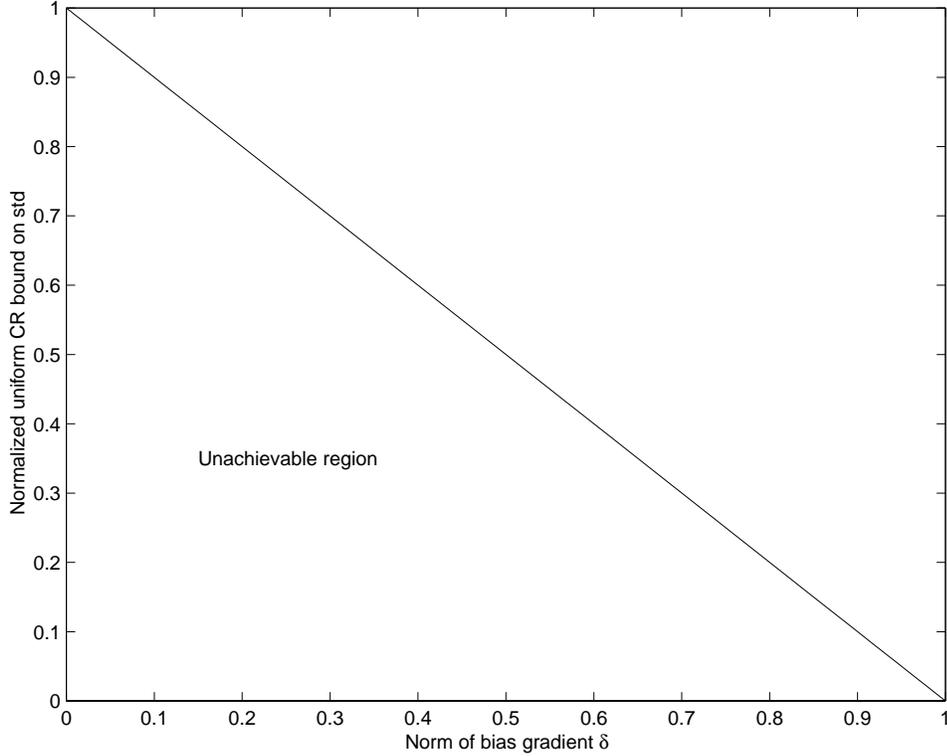


Figure 5.27: The normalized uniform CR bound.

### 5.9.1 Estimation of Bias Gradient

To compare a particular estimator to the uniform CR bound in Theorem 1, the variance and the bias gradient length need to be determined. Thus the particular estimator can be placed in the achievable region above the uniform CR curve.

Bias and variance are analytically *intractable* (even for this 1-D problem), both for the exact log-likelihood estimator and for the estimators based on approximate log-likelihood models. Thus, we experimentally determine sample mean and sample variance for a sequence of  $L$  repeated measurement realizations  $\{Y_j\}_{j=1}^L$ , *i.e.*: the sample variance is  $\hat{\sigma}^2 = \frac{1}{L-1} \sum_{j=1}^L (\hat{\lambda}(Y_j) - \bar{\hat{\lambda}})^2$  where  $\bar{\hat{\lambda}} = \frac{1}{L} \sum_{j=1}^L \hat{\lambda}(Y_j)$  is the estimator sample mean.

One can estimate the bias gradient by performing additional experiments with perturbation of parameter  $\lambda$ . A computationally more efficient method is developed in [51] which requires the same number of simulations as the sample mean and the sample variance estimates. The unbiased and consistent sample mean estimate of bias gradient is given as [51]:

$$\frac{\partial}{\partial \lambda} \widehat{b(\lambda)} = \frac{1}{L-1} \sum_{j=1}^L (\hat{\lambda}(Y_j) - \bar{\hat{\lambda}}) \left( \frac{\partial}{\partial \lambda} L(Y_j, \lambda) \right) - 1. \quad (5.43)$$

### 5.9.2 Simulations

We performed simulations to compare the performance each estimator with the uniform CR bounds. In the simulations the total number of measurement bins were  $N = 100$  and

the true value of radio-isotope concentration was  $\lambda = 1$ , and the  $c_n$  and the  $r_n$  factors were non-uniformly distributed with 50% random counts, *i.e.*:  $\sum c_n \lambda = \sum r_n$ . We performed simulations with different amount of average counts per bin as : 0.2, 0.5, 1, 2, 20 and 200. For each count level, we generated  $L = 10000$  realizations of the measurements  $\{Y_j\}_{j=1}^L$ . We applied each estimator : exact (5.8), OP (5.12), SP (5.13) and SD (5.14) to the multiple measurement realizations and computed the sample mean and the sample variance for each estimator. We also estimated the bias gradient length for each estimator using (5.43).

To compare the performance of each estimator to the bounds, we generated uniform CR bound curves at each count level using the following approach. It can be shown using (3.8, 3.9) that

$$\frac{\partial}{\partial \lambda} L(Y, \lambda) = \sum_{n=1}^N c_n \left[ \frac{v_n(y_n - 1, \lambda)}{v_n(y_n, \lambda)} - 1 \right], \quad (5.44)$$

since

$$\frac{\partial}{\partial \lambda} v_n(y_n, \lambda) = c_n v_n(y_n - 1, \lambda), \quad (5.45)$$

with

$$v_n(y_n, r_n) = \begin{cases} \sum_{m=0}^{\infty} \frac{\bar{y}_n(\lambda)^{y_n+m}}{(y_n+m)!} \frac{r_n^m}{m!}, & y_n \geq 0 \\ \sum_{k=0}^{\infty} \frac{\bar{y}_n(\lambda)^k}{k!} \frac{r_n^{(k-y_n)}}{(k-y_n)!}, & y_n < 0 \end{cases} \quad (5.46)$$

$$= \begin{cases} \left( \frac{\bar{y}_n(\lambda)}{i\sqrt{\bar{y}_n(\lambda) r_n}} \right)^{y_n} J_{y_n} \left( 2i\sqrt{\bar{y}_n(\lambda) r_n} \right), & y_n \geq 0 \\ \left( \frac{r_n}{i\sqrt{\bar{y}_n(\lambda) r_n}} \right)^{(-y_n)} J_{(-y_n)} \left( 2i\sqrt{\bar{y}_n(\lambda) r_n} \right), & y_n < 0 \end{cases}. \quad (5.47)$$

from (3.9, 3.14) and  $J_n$  is the Bessel function of the first kind of order  $n$ .

In the simulations we computed single trial Fisher information for each realization as:

$$F_{Y_j} = \left( \frac{\partial}{\partial \lambda} L(Y_j, \lambda) \right)^2 \quad (5.48)$$

and then computed the sample mean  $\widehat{F}_Y = \frac{1}{L} \sum_{j=1}^L F_{Y_j}$  to estimate the Fisher information matrix. Using this Fisher information estimate, we computed  $B(\lambda, \delta)$  and  $g(\kappa)$  in (5.41) and (5.42). Hence, we generated the uniform CR bound curves in the  $\sigma - \delta$  trade-off plane by varying  $\kappa$  over the range  $(0, \infty)$ .

Figure 5.28 shows the standard deviation versus bias of each estimator for 0.2, 0.5, 1, 2, 20 and 200 counts per bin. Lower counts correspond to higher standard deviation in the figure. The plots also show the standard error bars (plus and minus one standard deviation) for bias (horizontal lines) and standard deviation (vertical lines). For almost all the cases the error bars are smaller than the plotting symbols. Both the OP and the SP model estimators are observed to be highly positively biased especially at low count levels. This bias is essentially due to the zero thresholding of the data. Zero thresholding increases the mean of the data which results in a systematic positive bias since the data is linearly related to  $\lambda$ . At all count levels the exact and the SD models are observed to be virtually unbiased.

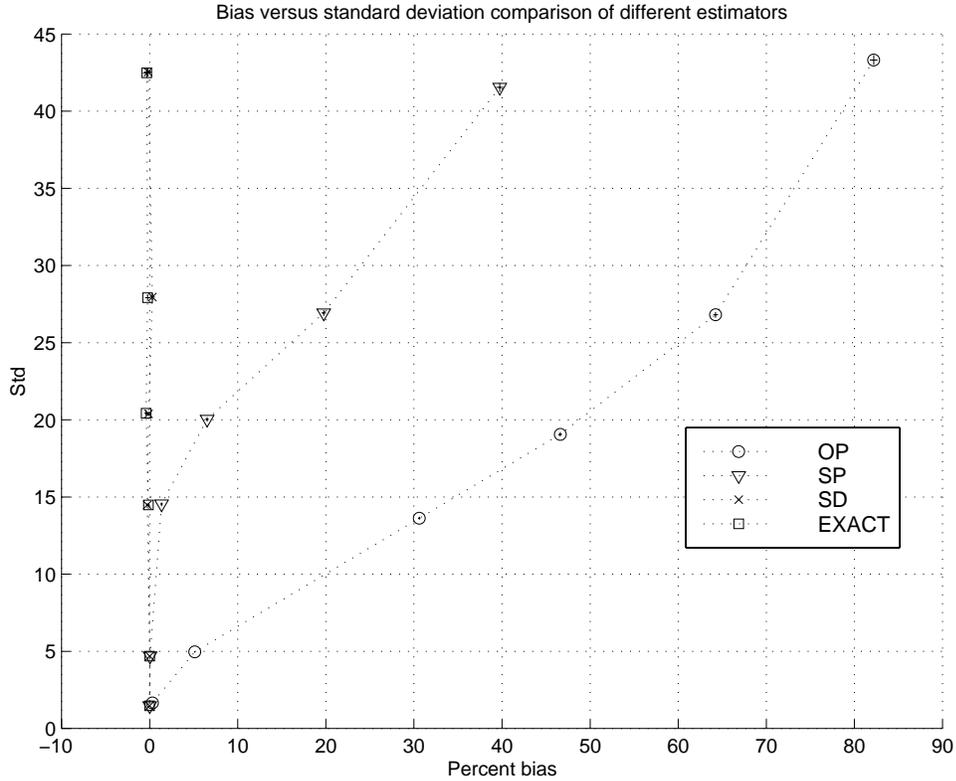


Figure 5.28: Bias versus standard deviation comparison of different estimators together with standard error bars. For almost all the cases the error bars are smaller than plotting symbols. The OP and SP models are observed to be positively biased especially for low count rates.

Figure 5.29 compares different estimators to uniform CR bound in the  $\sigma - \delta$  trade-off plane. We included the standard error bars for the uniform CR bound curve (shown in broken lines above and below the CR bound curve). To show all the results from different count levels in the same plot we applied a scaling factor (inversely proportional to the square root of the average counts per bin) to the standard deviation axis for the results at each count level such that the uniform CR bound curves overlap. The plots for the OP method also include standard error bars (plus and minus one standard deviation) for bias-gradient (horizontal lines) and standard deviation (vertical lines). The error bars for other methods are very similar to OP method and are not plotted in order not to clutter the figure. For all count levels the OP model is observed to be further away from the uniform bound than all the other estimators and thus it has the worst performance. The SP, SD and exact methods are all observed to be very close to the uniform CR bound curve (especially at higher counts). Although SP method results are a little further away from the CR bound compared to the SD and exact methods, this difference does not seem statistically significant based on the error bars.

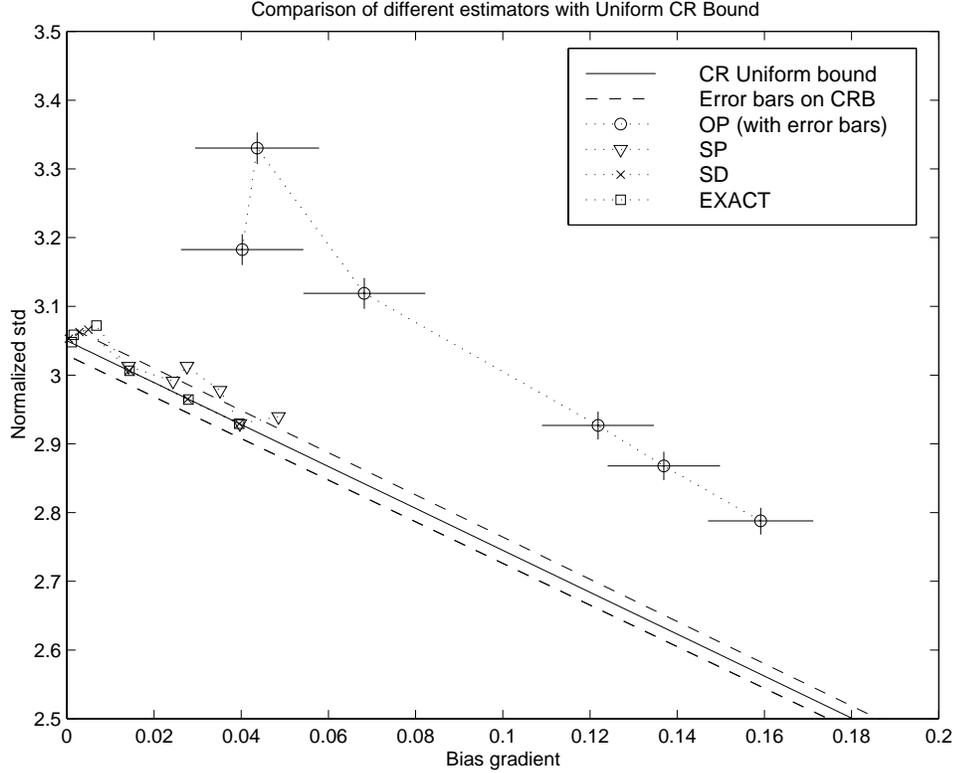


Figure 5.29: Performance of different estimators at different count levels compared to normalized uniform CR bound (with standard error bar curves). The plots for the OP method also include standard error bars. The error bars are not included for the other methods since they are very similar to error bars on the OP method. For all the count levels the OP method is observed to be further away from the uniform CR bound.

## 5.10 Conclusions

In PET emission scans, generally a significant portion of the collected data is accidental coincidence (AC) events and it is a primary source of background noise. Most PET scans are compensated for AC events by real-time subtraction of delayed-window coincidences. For the randoms pre-corrected data we analyzed the concavity of the objective functions and showed the data needs to be properly zero thresholded for the OP and SP methods to ensure convergence, whereas the SD model is globally concave without any necessity to zero thresholding. We developed appropriate maximization algorithms to be used in the image reconstructions with the proposed methods: first we introduced EM type maximization algorithms for the proposed methods, then we applied the paraboloid surrogates maximization algorithm.

To analyze the performance of the proposed methods first we performed 1-D simulations. The results with different count levels showed that the OP and SP methods result in systematic positive bias due to zero thresholding, while SD and exact log-likelihood methods result in bias free estimates at all count levels. However, our simulations with noise free emission measurement showed the interesting property that the SD method and exact log-likelihood can result in negative bias when there is fractional counts (less than one) per

bin. This phenomena requires further investigation.

We also performed 2-D simulations (with different count levels) which showed that the proposed models result in reconstructions that are free of any systematic bias and lead to images with less noise compared to ordinary Poisson (OP) model for the randoms pre-corrected data. The SP model is shown to be slightly biased for emission scans with very low count rates, whereas the SD model is free of any systematic bias and performs almost identically as the exact log-likelihood.

Lastly, we studied the bias-variance trade-offs of the models in 1-D by analyzing how close they perform to the uniform Cramer-Rao bounds. The analysis showed that the OP method results in further distance from the uniform bound, *i.e.* it results in worse performance than the proposed methods.

## CHAPTER 6

### Conclusions and Future Work

#### 6.1 Conclusions

AC events are a primary source of background noise in PET and should be compensated appropriately both for the emission and transmission scans. One can use the “singles” method [9] for estimating AC events, however this approach is not widely used because of the necessity for additional hardware and moreover usually singles rate vary during data acquisition. Thus, most PET centers use randoms pre-corrected data. In randoms pre-correction the AC rates are estimated by delayed-window coincidences and data are pre-corrected for AC events by *real-time* subtraction. Real time subtraction of delayed coincidences compensates for the average of AC events, but *destroys* the Poisson statistics [53]. Since the introduction of an ML-EM algorithm for PET more than 25 years ago [61,81], statistical image reconstruction methods have been based on idealized PET system with Poisson statistical model, and ignored the effects of AC events. Although, randoms pre-correction method *clearly* violates the Poisson statistics of the measurements, this problem has been largely ignored in the PET SIR literature. Numerous papers have been published simply ignoring the AC events and the randoms pre-correction. In most of the commercial PET scanners (with or without statistical image reconstruction tools) image reconstruction is done using randoms pre-corrected data.

We recommend separate acquisition and storage of delayed coincidences whenever feasible. However, in practice most PET center archive and use only randoms precorrected data because of software, hardware and data storage limitations (and historic momentum). In this thesis we developed accurate statistical models and image reconstruction techniques for PET measurements with pre-subtracted delayed coincidences. It may seem ironic that we developed complicated image reconstruction methods to solve a problem created by the data acquisition techniques employed by the PET scanners. However, this situation is a real problem that most PET centers face with everyday. In practice, almost all of the PET centers collect randoms pre-corrected data. Even most of the latest commercial PET scanners (with or without iterative statistical image reconstruction tools) use only random pre-corrected data in their image reconstruction. Thus, it can be said that for the foreseeable future PET centers will collect and achieve randoms pre-corrected data.

We introduced two new approximations to the complicated exact log-likelihood of the pre-corrected measurements in PET: one based on a “shifted Poisson” (SP) model, and the other based on saddle-point approximations to the measurement probability mass function (pmf) in Chapter 3. The SP model is based on the idea of matching both the first

and second-order moments of the model to the underlying statistics of the pre-corrected data [97]. Although both the WLS and SP models match two moments to the underlying statistics, in the data WLS model the second moment is fixed independent of the unknown parameters to be estimated (*i.e.*: the image), whereas in SP model the moments vary with the measurement model appropriately. This difference is shown to be a very important difference between the two models and the corresponding log-likelihood function of the SP model is shown to have better agreement with the exact log-likelihood than the conventional WLS and ordinary Poisson (OP) models. Moreover, the method is very practical and easy to implement, and requires only negligible increase in computation. The second method introduced (saddle-point (SD) model) [98,99] is a very precise approximation to the exact distribution of the pre-corrected measurements, based on the idea of making a second order Taylor series approximation to the exact pmf in the  $z$ -transform domain (*i.e.*: on the probability generating function) and then carrying out the inverse transform. The corresponding log-likelihood function to the SD model is shown to have the best agreement with the exact log-likelihood compared to all of the previous approximations and its performance is shown to be almost identical to the exact log-likelihood method.

We compared the estimators based on the new models to the conventional data WLS and conventional maximum likelihood (based on the ordinary Poisson (OP) model) using experiments, simulations and analytic approximations. We developed maximization algorithms for the SP and the SD methods and presented representative performance results from computer simulations and experimental transmission scans in Chapter 4. The results show that the WLS method leads to systematic negative bias in the reconstructed attenuation maps and the OP method results in higher standard deviation than the proposed SP and SD methods. Although the individual images reconstructed by the SP method and the EX method differed slightly, we observed very close agreement between the SD method and *truncated* exact log-likelihood for each reconstruction. However, the differences between SP, SD and EX models were statistically insignificant (based on the ensemble means and variances). Thus SP method is particularly attractive since it requires comparable computation to the OP method but has reduced variance.

Both for the SP and SD methods some form of the estimates of the mean of random coincidences needs to be used. We showed that the methods are very robust to the errors in these estimates (even using a single scalar constant works fairly well). Note that these estimates are used essentially for estimating the variance of the randoms pre-corrected data, not for performing randoms pre-correction. In our experiments the random coincidence factors were *not* available neither for the transmission nor for the blank scans, since the data were already pre-corrected for the randoms. Thus, to estimate the mean of randoms, we simply scaled the blank scan so that its sum corresponded to the total number of AC events (this scalar is available in the transmission scan file header) with no additional processing. Despite this possibly being a suboptimal approach, the SP method still yielded lower variance attenuation maps than the OP method.

We also investigated the propagation of noise from the reconstructed attenuation maps into the emission images reconstructed using the FBP method. Interestingly, the the difference in standard deviations in the emission images with the new methods were shown to be even greater than in the attenuation maps.

To corroborate the *empirical* studies, we developed *analytical* approximations to the reconstructed image covariance and we also developed analytical approximations for the propagation of noise from attenuation maps into the reconstructed emission images. The results of the analytic approximations are shown to be in good agreement with the experi-

mental results.

In Chapter 5 we concentrated on the emission problem. We analyzed the concavity of the objective functions and showed the data needs to be properly zero thresholded for the OP and SP methods to ensure convergence, whereas the SD model is globally concave without any necessity for zero thresholding. We developed appropriate maximization algorithms to be used in the image reconstructions with the proposed methods: first we introduced EM type maximization algorithms for the proposed methods, then we applied the paraboloid surrogates maximization algorithm.

With 1-D and 2-D simulations (with different count levels) we showed that the proposed models result in reconstructions that are free of any systematic bias and lead to images with less noise compared to ordinary Poisson (OP) model for the randoms pre-corrected data. Although the SP model is shown to be slightly biased for emission scans with very low count rates, the SD model is free of any systematic bias and performs almost identically to the exact log-likelihood. Also, we studied the bias-variance trade-offs of the models in 1-D by analyzing how close they perform to the uniform Cramer-Rao bounds. The analysis showed that the OP method results in further distance from the uniform bound, *i.e.*: it results in worse performance than the proposed methods.

The new methods offer improved image reconstruction in PET through more realistic statistical modeling, yet with negligible increase in computation over the conventional OP method.

## 6.2 Future Work

In this section we provide several suggestions for future work.

- In Appendix A, we have analyzed the effects of different update orders on the convergence rate properties of the sequential algorithms, as a function of spatial frequency. Although the analysis is carried out for WLS objective function, the results can be extended to other sequential algorithms like SAGE, GCA or ordered subsets EM.
- In our models we assumed that the scattered events are known. The introduced methods seem to be robust to the errors in the estimated values of scattered values. For example in experimental transmission scan studies we ignored the scattered events but the analytical approximations still performed fairly accurately. Nevertheless, this topic requires further investigation.
- For the transmission problem neither the SP objective function nor the PR objective function are concave. For the SD model we proved that it is concave for the emission case, but for the transmission case the concavity analysis proves to be algebraically tedious due to the complicated forms of the partial derivatives. For the exact log-likelihood the complicated form of the log-likelihood expressions make it difficult to perform a rigorous concavity analysis. Previous convergence proofs for transmission algorithms depended heavily on the assumption of  $r_n = 0$  (*i.e.*: OP model). Further investigation is necessary for the convergence properties of transmission algorithms for randoms precorrected data and for the PR model with nonzero  $r_n$ 's.
- We developed analytical approximations for the propagation of noise from attenuation maps into reconstructed emission images. To isolate the effect of transmission noise in the resultant emission image, we considered noise-free emission measurements and

develop approximations for the covariance of the emission images reconstructed with FBP method using ACFs computed from noisy attenuation maps. Further analysis for noisy emission measurements reconstructed with other reconstruction methods would be helpful.

- In our 1-D and 2-D emission simulations we observed that both the SD method and exact log-likelihood method are free of any systematic bias even for very low count rate scans. However, during the simulations with *noise-free* data, we noticed that these methods result in negative bias for low fractional counts. This “mysterious” behavior of the exact ML estimation method requires further investigation.
- Applications of the proposed methods to fully 3-D PET should show even further image quality improvement compared to conventional ML methods (based on OP model), since high AC rates and very low counts per sinogram bin are common in 3-D PET.
- General approaches of this thesis can be applicable to the case where another complicated distribution arise when detector deadtime effects are severe. It may be possible to extent related ideas to the detector deadtime problem. But utility of this is uncertain since Poisson assumption seems to be a reasonably good model for this effect.
- It would also be beneficial to extend the 1-D CRB analysis of the proposed methods to 2-D.
- Lastly, it may be worthwhile to investigate the applicability of the developed methods for electronically collimated SPECT.

## APPENDICES

## APPENDIX A

### Update Orders for Sequential Iterative Algorithms

Unlike simultaneous update methods, for sequential iterative methods the “update order” of the image pixels effects the convergence rate of the algorithm [7, 31, 79]. In this appendix, we analyze the effects of different update orders on the convergence rate properties of the sequential algorithm, as a function of spatial frequency. Although the analysis is carried out for PWLS objective function, one can expect to extend the results to other sequential algorithms like SAGE.

For the emission problem the discretized tomographic system can be modeled with the *system matrix*  $\mathbf{A}$  where an element  $g_{ij}$  of  $\mathbf{A}$  denotes the contribution of the annihilations in the  $j$ th pixel to the  $i$ th detector pair measurements. The mean of the  $i$ th detector pair measurement can be approximated as

$$\bar{y}_i = \sum_{j=1}^N g_{ij} \lambda_j \quad (\text{A.1})$$

where  $\lambda_j$  represents the annihilation activity in the  $j$ th pixel.

The WLS objective for emission tomography (3.16) with the penalty (PWLS) is:

$$\Phi(\lambda) = \frac{1}{2}(\hat{y} - \mathbf{A}\lambda)' \mathbf{\Sigma}^{-1}(\hat{y} - \mathbf{A}\lambda) + \beta R(\lambda) \quad (\text{A.2})$$

where  $\hat{y}$  is the measurement vector with  $\hat{y}_i$  corresponding to the  $i$ th detector pair measurement,  $\mathbf{\Sigma}$  is the diagonal weighting matrix with  $i$ th diagonal entry  $\sigma_i^2$ , being the estimated variance of  $\hat{y}_i$ . Finally  $R(\lambda)$  is the penalty function and  $\beta$  is the smoothing parameter. The basic penalty function we use in this appendix is the quadratic smoothness penalty:

$$R(\lambda) = \frac{1}{2} \lambda' \mathbf{R} \lambda. \quad (\text{A.3})$$

As shown in [31], the WLS objective function (A.2), with the above penalty (for  $\beta > 0$ ) leads to strictly convex objective function. If  $\mathbf{A}$  has full rank,  $\mathbf{R}$  is symmetric nonnegative definite, and the intersection of null spaces of  $\mathbf{A}$  and  $\mathbf{R}$  is empty then the corresponding unconstrained solution  $\hat{\lambda}$  satisfies

$$\nabla_{\lambda=\hat{\lambda}} \Phi = -\mathbf{A}' \mathbf{\Sigma}^{-1}(\hat{y} - \mathbf{A}\hat{\lambda}) + \beta \mathbf{R}\hat{\lambda} = 0 \quad (\text{A.4})$$

$$\mathbf{H}\hat{\lambda} = \mathbf{A}' \mathbf{\Sigma}^{-1} \hat{y} \quad (\text{A.5})$$

where  $\mathbf{H}$  is the Hessian :

$$\mathbf{H} = \nabla_{\lambda}^2 \Phi = \mathbf{A}' \mathbf{\Sigma}^{-1} \mathbf{A} + \beta \mathbf{R}. \quad (\text{A.6})$$

## A.1 Iterative Algorithm

Successive over-relaxation method (+SOR) is a computational efficient algorithm for minimizing the quadratic objective function subject to the nonnegativity constraint. +SOR is a coordinate descent algorithm, which sequentially updates one pixel at a time in order to minimize the objective function while holding remaining pixel values constant. Gauss-Seidel is a special case of SOR (when over-relaxation parameter is one) also known as ICM [3]. SOR algorithm without the nonnegativity constraint can be described in matrix form as follows [101],

We first decompose  $\mathbf{H}$  (A.6) as:

$$\mathbf{H} = \mathbf{L} + \mathbf{D} + \mathbf{L}' \quad (\text{A.7})$$

where  $\mathbf{L}$  and  $\mathbf{D}$  are *strictly lower triangular* and *diagonal* parts of  $\mathbf{H}$  respectively. Then SOR method corresponds to

$$\lambda^{i+1} = -(\mathbf{D} + \alpha\mathbf{L})^{-1}[(\alpha - 1)\mathbf{D} + \alpha\mathbf{L}']\lambda^i + (\mathbf{D} + \alpha\mathbf{L})^{-1}\alpha\mathbf{A}'\mathbf{\Sigma}^{-1}\hat{y} \quad (\text{A.8})$$

where  $i$  indexes iteration and  $\alpha \in (0, 2)$  is the relaxation parameter. The convergence behavior of such iterations is determined [101] by the eigenvalues of

$$\mathbf{G}_\alpha = -(\mathbf{D} + \alpha\mathbf{L})^{-1}[(\alpha - 1)\mathbf{D} + \alpha\mathbf{L}']. \quad (\text{A.9})$$

## A.2 Convergence Properties

To analyze the eigenvalues of  $\mathbf{G}_\alpha$  we consider the 1-D problem with the simplifying assumptions that  $\mathbf{\Sigma} = \mathbf{I}$  and  $\mathbf{A}'\mathbf{A}$ ,  $\mathbf{R}$  and  $\mathbf{H}$  are circulant matrices (adopting the 2-D assumptions in [31] and [79]). The later assumption implies  $\mathbf{G}_\alpha$  to be also a circulant matrix, which enables one to analyze its eigenvalues as a function of frequency (of the corresponding eigenvectors), by using 1-D Discrete Fourier Transform (DFT) of the kernel of  $\mathbf{G}_\alpha$ .

For the kernel of  $\mathbf{A}'\mathbf{A}$  we use the following analytic approximation to  $1/r$  as suggested in [31] :

$$f(r) = \begin{cases} \pi - r & r \in [0, 1] \\ 2(\arcsin(1/r) - (r - \sqrt{r^2 - 1})) & r > 1 \end{cases} \quad (\text{A.10})$$

and the quadratic penalty function  $\mathbf{R}$  has the kernel  $[-1 \ 2 \ 1]$ .

Let  $l(n)$  be the kernel of  $\mathbf{L}$  which corresponds to the causal part of the kernel of  $\mathbf{H}$  and let  $L(w)$  be the corresponding 1-D DFT. Since  $\mathbf{D}$  is diagonal,  $\mathbf{D} = d\mathbf{I}$  where  $\mathbf{I}$  is the identity matrix. Then the eigenvalues of  $\mathbf{G}_\alpha$  (A.9) as a function of frequency  $\omega$  are approximately

$$\lambda_\alpha(\omega) = -\frac{(\alpha - 1)d + \alpha L^*(\omega)}{d + \alpha L(\omega)} \quad (\text{A.11})$$

where  $*$  denotes complex conjugate corresponding to matrix transpose operation in (A.9).

To compare  $\lambda_\alpha(\omega)$  with the exact eigenvalues of  $\mathbf{G}_\alpha$ , we calculated the eigenvalues of  $\mathbf{G}_\alpha$  using MATLAB and considered the dominant frequency component of each eigenvector as the frequency content of the corresponding eigenvector. Comparing the exact eigenvalues with the analytic approximation for a  $\mathbf{G}_\alpha$  of size 128x128, relaxation parameter  $\alpha = 1$  and smoothness parameter  $\beta = 7$ , we observed that analytic approximation agrees well with the exact eigenvalues and, as shown by Sauer and Bouman [79], high frequency components converge fastest.

### A.2.1 Forward Backward Ordering

As noted previously SOR is a sequential algorithm, which enables one to alter the update order of the pixels for each iteration. The previous analysis corresponds to conventional ordering, *i.e.*: one updates  $1^{st}2^{nd} \dots P^{th}$  pixels sequentially in every iteration. Different update orderings change the form of  $\mathbf{G}_\alpha$  resulting in different convergence properties. We experimented with several different update orders and one that works well is forward-backward ordering (FBO) in which one begins from the first pixel and updates every other pixel skipping the next one and repeats the procedure from the end to the beginning for the remaining pixels, *i.e.*: update order is:  $1^{st}3^{rd} \dots P^{th} (P-1)^{st}(P-3)^{rd} \dots 4^{th}2^{nd}$  (when  $P$  is odd).

Such an ordering corresponds to the new linear system (compare with (A.5)):

$$(\mathbf{PHP}')(\mathbf{P}\lambda) = \mathbf{PA}'\Sigma^{-1}\hat{g} \quad (\text{A.12})$$

where  $\mathbf{P}$  is the permutation matrix such that

$$\mathbf{P}_{ij} = \begin{cases} 1 & j = (2i - 1) \quad i \leq \frac{P}{2} \\ 1 & j = 2(i - \frac{P}{2}) \quad \frac{P}{2} < i \leq P \\ 0 & \text{otherwise.} \end{cases} \quad (\text{A.13})$$

The new Hessian becomes  $\mathbf{H}_{fb} = \mathbf{PHP}'$ . Let the corresponding strictly lower triangular and diagonal parts to be  $\mathbf{L}_{fb}$  and  $\mathbf{D}_{fb}$  such that  $\mathbf{H}_{fb} = \mathbf{L}_{fb} + \mathbf{D}_{fb} + \mathbf{L}'_{fb}$ .

By analogy to (A.9) the convergence rate of FBO is determined by the eigenvalues of

$$\mathbf{G}_{fb} = -(\mathbf{D}_{fb} + \alpha\mathbf{L}_{fb})^{-1}[(\alpha - 1)\mathbf{D}_{fb} + \alpha\mathbf{L}'_{fb}]. \quad (\text{A.14})$$

It can be shown that the Hessian  $\mathbf{H}_{fb}$  has the form:

$$\mathbf{H}_{fb} = \begin{bmatrix} \mathbf{L}_d + \mathbf{D}_d + \mathbf{L}'_d & \mathbf{K}_d^a \\ \mathbf{K}_d^a & \mathbf{L}_d + \mathbf{D}_d + \mathbf{L}_d \end{bmatrix}. \quad (\text{A.15})$$

where  $\mathbf{D}_d = d \mathbf{I}_{(\frac{P}{2} \times \frac{P}{2})}$  is diagonal matrix with same diagonal entries as  $\mathbf{D}_{fb}$ .  $\mathbf{L}_d$  is the lower triangular matrix having the kernel  $l_d(n)$  corresponding to *down sampling* by 2 of the kernel  $l(n)$  (causal part of  $h(n)$ ), *i.e.*:  $l_d(n) = l(2n)$ .  $\mathbf{K}_d$  has the kernel  $k_d(n) = h(2n + 1)$  and  $\mathbf{K}_d^a$  corresponds to time reversal, *i.e.*:  $k_d(-n)$ .

The corresponding DTFT of  $l(n)$  and  $k_d(n)$  are

$$\begin{aligned} L_d(\omega) &= \frac{1}{2} [L(\frac{\omega}{2}) + L(\frac{\omega}{2} + \pi)] \\ K_d(\omega) &= \frac{1}{2} e^{-j\frac{\omega}{2}} [(L(\frac{\omega}{2}) - L(\frac{\omega}{2} + \pi)) + (L^*(\frac{\omega}{2}) - L^*(\frac{\omega}{2} + \pi))]. \end{aligned} \quad (\text{A.16})$$

Then  $\mathbf{G}_{fb}$  (14) can be written as:

$$\begin{aligned} \mathbf{G}_{fb} &= - \begin{bmatrix} \alpha\mathbf{L}_d + \mathbf{D}_d & 0 \\ \alpha\mathbf{K}_d^a & \alpha\mathbf{L}_d + \mathbf{D}_d \end{bmatrix}^{-1} \begin{bmatrix} (\alpha - 1)\mathbf{D}_d + \alpha\mathbf{L}'_d & \alpha\mathbf{K}_d^a \\ 0 & (\alpha - 1)\mathbf{D}_d + \alpha\mathbf{L}'_d \end{bmatrix} \\ &= - \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ -\mathbf{BA} & -\mathbf{BB} + \mathbf{A} \end{bmatrix}. \end{aligned} \quad (\text{A.17})$$

where

$$\begin{aligned}\mathbf{A} &= (\alpha\mathbf{L}_d + \mathbf{D}_d)^{-1} ((\alpha - 1)\mathbf{D}_d + \alpha\mathbf{L}'_d) \\ \mathbf{B} &= (\alpha\mathbf{L}_d + \mathbf{D}_d)^{-1} (\alpha\mathbf{K}_d^a)\end{aligned}\quad . \quad (\text{A.18})$$

To find the eigenvalues of  $\mathbf{G}_{fb}$ , one needs to solve  $\mathbf{G}_{fb}\underline{x} = \lambda\underline{x}$ . Breaking the eigenvector  $\underline{x}$  into two parts as:

$$\underline{x} = \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \end{bmatrix}$$

we obtain the set of equations:

$$\begin{aligned}\mathbf{A}\underline{x}_1^i + \mathbf{B}\underline{x}_2^i &= -\lambda_i\underline{x}_1^i \\ \mathbf{A}\underline{x}_2^i &= -\lambda_i\mathbf{B}\underline{x}_1^i - \lambda_i\underline{x}_2^i\end{aligned}\quad \text{for } i \in (0, 1, \dots, P-1). \quad (\text{A.19})$$

Our empirical results suggest that for each eigenvalue  $\lambda_i$  both of the eigenvectors  $\underline{x}_1^i$  and  $\underline{x}_2^i$  are composed of linear combinations of DFT basis vectors with corresponding frequencies  $\pm\omega_i$ , such as

$$\begin{aligned}\underline{x}_1^i &= k_{11}^i\underline{x}^{\omega_i} + k_{12}^i\underline{x}^{-\omega_i} \\ \underline{x}_2^i &= k_{21}^i\underline{x}^{\omega_i} + k_{22}^i\underline{x}^{-\omega_i}\end{aligned}\quad (\text{A.20})$$

where  $\underline{x}^{\omega_i}$  is the DFT basis vector as  $[e^{-j0} e^{-j\frac{2\pi}{P}i} \dots e^{-j\frac{2\pi(P-1)}{P}i}]$ .

Using the definitions of  $\mathbf{A}$  and  $\mathbf{B}$  (A.18) :

$$\begin{aligned}\mathbf{A}\underline{x}^{\omega_i} &= \left(\frac{(\alpha-1)d+\alpha L_d(\omega_i)^*}{d+\alpha L_d(\omega_i)}\right) \underline{x}^{\omega_i} = a(\omega_i)\underline{x}^{\omega_i} \\ \mathbf{B}\underline{x}^{\omega_i} &= e^{-j\omega_i} \left(\frac{K_d(\omega_i)}{d+\alpha L_d(\omega_i)}\right)^* \underline{x}^{-\omega_i} = b(\omega_i)\underline{x}^{-\omega_i}\end{aligned}\quad (\text{A.21})$$

the set of eigen equations (19) becomes

$$\begin{aligned}[(a(\omega_i) + \lambda_i)k_{11}^i] \underline{x}^{\omega_i} + [(a(\omega_i)^* + \lambda_i)k_{12}^i] \underline{x}^{-\omega_i} &= [-b(\omega_i)^*k_{22}^i] \underline{x}^{\omega_i} + [-b(\omega_i)k_{21}^i] \underline{x}^{-\omega_i} \\ [(a(\omega_i) + \lambda_i)k_{21}^i] \underline{x}^{\omega_i} + [(a(\omega_i)^* + \lambda_i)k_{22}^i] \underline{x}^{-\omega_i} &= \lambda_i [-b(\omega_i)^*k_{12}^i] \underline{x}^{\omega_i} + \lambda_i [-b(\omega_i)k_{11}^i] \underline{x}^{-\omega_i}.\end{aligned}$$

For  $i \neq 0$ ,  $\underline{x}^{\omega_i}$  and  $\underline{x}^{-\omega_i}$  are linearly independent, so it is required that:

$$\begin{aligned}(a(\omega_i) + \lambda_i)k_{11}^i &= -b(\omega_i)^*k_{22}^i, & (a(\omega_i)^* + \lambda_i)k_{12}^i &= -b(\omega_i)k_{21}^i \\ (a(\omega_i)^* + \lambda_i)k_{22}^i &= -\lambda_i b(\omega_i)k_{11}^i, & (a(\omega_i) + \lambda_i)k_{21}^i &= -\lambda_i b(\omega_i)^*k_{12}^i.\end{aligned}\quad (\text{A.22})$$

If one divides the left side of above equalities by the complex conjugate of the right side equalities and gets rid of the  $i$  dependence for notational simplicity, the resulting equalities are

$$\frac{(a(\omega) + \lambda)k_{11}}{a(\omega) + \lambda^*)k_{12}^*} = \frac{-b(\omega)^*k_{22}}{-b(\omega)^*k_{21}^*} = \frac{k_{22}}{k_{21}^*} \quad (\text{A.23})$$

$$\frac{(a(\omega)^* + \lambda)k_{22}}{a(\omega)^* + \lambda^*)k_{21}^*} = \frac{-\lambda b(\omega)k_{11}}{-\lambda^* b(\omega)k_{12}^*} = \frac{\lambda k_{11}}{\lambda^* k_{12}^*}. \quad (\text{A.24})$$

From the above equalities

$$\frac{(a(\omega)^* + \lambda)(a(\omega) + \lambda)\lambda^*}{(a(\omega) + \lambda^*)(a(\omega)^* + \lambda^*)\lambda} = 1 \quad (\text{A.25})$$

$$\text{Im}[(a(\omega)^* + \lambda)(a(\omega) + \lambda)\lambda^*] = 0 \quad (\text{A.26})$$

$$\text{Im}[|a(\omega)|^2\lambda^* + a(\omega)^*|\lambda|^2 + a(\omega)|\lambda|^2 + \lambda|\lambda|^2] = 0 \quad (\text{A.27})$$

$$|\lambda|^2 = |a(\omega)|^2. \quad (\text{A.28})$$

Using (A.21) for  $a(\omega)$ , one gets the relation between the eigenvalues of  $\mathbf{G}_{fb}$  and the  $L(\omega)$  as

$$|\lambda(\omega)| = \left| \frac{(\alpha - 1)d + \frac{\alpha}{2}[L(\frac{\omega}{2})^* + L(\frac{\omega}{2} + \pi)^*]}{d + \frac{\alpha}{2}[L(\frac{\omega}{2}) + L(\frac{\omega}{2} + \pi)]} \right|. \quad (\text{A.29})$$

We observed close agreement for the above analytic approximation with the exact eigenvalues. Having an analytic approximation as above helps one to predict the convergence rate properties of the algorithm without calculating the exact eigenvalues of the system (which is computationally very difficult for a real sized problem).

Comparing the convergence properties of both update orders, we observed that at lower frequencies FBO has smaller eigenvalues, which suggest that one can use FBO for the early stages of iterations to quickly fine-tune the low frequencies. Another observation was that the largest eigenvalue of FBO is smaller than that of regular ordering, which corresponds to a better asymptotic convergence rate for FBO.

## APPENDIX B

### Taylor's Series Approximation of SP model

For transmission problem, SP model objective function (3.23) can be rewritten as:

$$L_{SP}(\mu) = \sum_{n=1}^N h_n(l_n(\mu)), \quad (\text{B.1})$$

where

$$\begin{aligned} h_n(l) &= (y_n + 2r_n) \log g_n(l) - g_n(l) \\ g_n(l) &= b_n e^{-l} + s_n^T + 2r_n. \end{aligned}$$

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

$$h_n(l) \approx h_n(\hat{l}_n) + h_n^{(1)}(\hat{l}_n)(l_n - \hat{l}_n) + \frac{h_n^{(2)}(\hat{l}_n)}{2}(l_n - \hat{l}_n)^2 \quad (\text{B.2})$$

where  $h_n^{(i)}(l) = \frac{d^i}{dl^i} h_n(l)$  and

$$h_n^{(1)}(l) = \left[ 1 - \frac{y_n + 2r_n}{g_n(l)} \right] b_n e^{-l} \quad (\text{B.3})$$

$$-h_n^{(2)}(l) = \left[ 1 - \frac{(y_n + 2r_n)2r_n}{g_n^2(l)} \right] b_n e^{-l}. \quad (\text{B.4})$$

Let  $\hat{l}_n \triangleq \log\left(\frac{b_n}{y_n - s_n^T}\right)$ , which is the method-of moment estimate of the line integral of attenuation  $l_n$  (as used in (3.15)), this estimate yields  $g_n(\hat{l}_n) = y_n + 2r_n$ . Substituting  $\hat{l}_n$  into the above equations:

$$\begin{aligned} h_n(\hat{l}_n) &= (y_n + 2r_n) \log(y_n + 2r_n) - (y_n + 2r_n) \\ h_n^{(1)}(\hat{l}_n) &= 0 \\ -h_n^{(2)}(\hat{l}_n) &= \frac{(y_n - s_n^T)^2}{(y_n + 2r_n)}. \end{aligned}$$

Substituting into (B.2) results in the approximation:

$$h_n(l) \approx [(y_n + 2r_n) \log(y_n + 2r_n) - (y_n + 2r_n)] - \frac{1}{2} \frac{(y_n - s_n^T)^2}{(y_n + 2r_n)} (l_n - \hat{l}_n)^2. \quad (\text{B.5})$$

Since the first term in (B.5) is independent of  $l$  we can disregard it, and then substituting into  $L_{SP}(\mu)$  (B.1) results in the WLS approximation (3.15):

$$L_{WLS}(\mu) = -\frac{1}{2} \sum_{n=1, y_n > 0}^N (l_n(\mu) - \hat{l}_n)^2 \frac{(y_n - s_n^T)^2}{(y_n + 2r_n)}. \quad (\text{B.6})$$

## APPENDIX C

### Bias and Variance Analysis

In this appendix, we analyze bias and variance of the estimators for the 1D transmission problem, using the analytic approximations suggested by Fessler for tomographic imaging [34]. Assuming that the objective function  $L(\mu, y)$  has a unique global maximum  $\hat{\mu}$  for any measurement  $y$  and that the maximum can be found by zeroing the partial derivatives of  $L(\mu, y)$ , *i.e.*:

$$0 = \frac{\partial}{\partial \mu} L(\mu, y) \Big|_{\mu=\hat{\mu}}, \quad (\text{C.1})$$

then there exists an implicit function  $f(y) = [f_1(y) \dots f_P(y)] = \hat{\mu}$  that maps the measurement  $y$  into an estimate  $\hat{\mu}$ . From (C.1), the function  $f(y)$  must satisfy:

$$0 = \frac{\partial}{\partial \mu} L(\mu, y) \Big|_{\mu=f(y)} = \frac{\partial}{\partial \mu} L(f(y), y). \quad (\text{C.2})$$

Computing the first and second order derivatives of (C.2) with respect to  $y_n$  (by applying chain rule) and considering the special case  $y_n = \bar{y}_n$ , we obtain

$$\frac{\partial}{\partial y_n} f(\bar{y}) = \frac{\frac{\partial^2}{\partial \mu \partial y_n} L(\check{\mu}, \bar{y})}{-\left(\frac{\partial^2}{\partial \mu^2} L(\check{\mu}, \bar{y})\right)} \quad (\text{C.3})$$

and

$$\begin{aligned} \frac{\partial^2}{\partial y_n^2} f(\bar{y}) &= \frac{\frac{\partial}{\partial y_n} f(\bar{y}_n)}{-\left(\frac{\partial^2}{\partial \mu^2} L(\check{\mu}, \bar{y})\right)} \left( \frac{\partial^3}{\partial \mu^3} L(\check{\mu}, \bar{y}) \frac{\partial}{\partial y_n} f(\bar{y}_n) + 2 \frac{\partial^3}{\partial \mu^2 \partial y_n} L(\check{\mu}, \bar{y}) \right) \\ &+ \frac{\frac{\partial^3}{\partial \mu \partial y_n^2} L(\check{\mu}, \bar{y})}{-\left(\frac{\partial^2}{\partial \mu^2} L(\check{\mu}, \bar{y})\right)} \end{aligned} \quad (\text{C.4})$$

where  $\check{\mu} = f(\bar{y})$ .

#### **OP Model:**

For the 1-D problem,  $l_n(\mu) = g_n \mu$ . Thus, the OP model objective function (3.20) can be written as:

$$L_{OP}(\mu, y) = \sum_{n=1}^N y_n \log \bar{y}_n(\mu) - \bar{y}_n(\mu) \quad (\text{C.5})$$

where  $\bar{y}_n = b_n e^{-g_n \mu} + s_n^T$ . In the following derivations,  $s_n^T$  is assumed to be zero for notational simplicity. The results for the models (OP, SP, SD) can easily be extended to the case  $s_n^T \neq 0$ . The corresponding derivatives are:

$$\begin{aligned}\frac{\partial}{\partial \mu} L_{OP}(\mu, y) &= \sum_{n=1}^N g_n \bar{y}_n(\mu) \left[ 1 - \frac{y_n}{\bar{y}_n(\mu)} \right] \\ -\frac{\partial^2}{\partial \mu^2} L_{OP}(\mu, y) &= \sum_{n=1}^N a_n^2 \bar{y}_n(\mu) \\ \frac{\partial^3}{\partial \mu^3} L_{OP}(\mu, y) &= \sum_{n=1}^N a_n^3 \bar{y}_n(\mu) \\ \frac{\partial^2}{\partial \mu \partial y_n} L_{OP}(\mu, y) &= -g_n \\ \frac{\partial^3}{\partial \mu \partial y_n^2} L_{OP}(\mu, y) &= \frac{\partial^3}{\partial \mu^2 \partial y_n} L_{OP}(\mu, y) = 0.\end{aligned}$$

Since we have omitted the penalty term from the objective function, estimator works perfectly with noiseless data  $\bar{y}_n$ , *i.e.*:  $\frac{\partial}{\partial \mu} L_{OP}(\mu_t, \bar{y}) = 0$ . Thus,  $\hat{\mu} = f(\bar{y}(\mu)) = \mu_t$  where  $\mu_t$  is the true value of the attenuation coefficient. Computing the values of above equalities at  $\mu_t$  and  $\bar{y}$ :

$$\begin{aligned}\frac{\partial}{\partial \mu} L_{OP}(\mu_t, \bar{y}) &= 0 \\ -\frac{\partial^2}{\partial \mu^2} L_{OP}(\mu_t, \bar{y}) &= \sum_{n=1}^N a_n^2 \bar{y}_n(\mu_t) \\ \frac{\partial^3}{\partial \mu^3} L_{OP}(\mu_t, \bar{y}) &= \sum_{n=1}^N a_n^3 \bar{y}_n(\mu_t) \\ \frac{\partial^2}{\partial \mu \partial y_n} L_{OP}(\mu_t, \bar{y}) &= -g_n \\ \frac{\partial^3}{\partial \mu \partial y_n^2} L_{OP}(\mu_t, \bar{y}) &= \frac{\partial^3}{\partial \mu^2 \partial y_n} L_{OP}(\mu_t, \bar{y}) = 0.\end{aligned}$$

Substituting into (C.3) and (C.4):

$$\begin{aligned}\frac{\partial}{\partial y_n} f_{OP}(\bar{y}) &= -\frac{g_n}{\sum_{n=1}^N a_n^2 \bar{y}_n(\mu_t)} \\ \text{and} \\ \frac{\partial^2}{\partial y_n^2} f_{OP}(\bar{y}) &= \frac{a_n^2 \sum_{n=1}^N a_n^3 \bar{y}_n(\mu_t)}{\left( \sum_{n=1}^N a_n^2 \bar{y}_n(\mu_t) \right)^3}.\end{aligned}$$

Lastly, substituting into (4.16) and (4.17) with  $\text{Var}(y_n) = \bar{y}_n(\mu_t) + 2r_n$ :

$$\begin{aligned}\text{Var}\{\hat{\mu}_{OP}\} &\approx \frac{\sum_{n=1}^N g_n^2 (\bar{y}_n(\mu_t) + 2r_n)}{\left( \sum_{n=1}^N g_n^2 \bar{y}_n(\mu_t) \right)^2}, \\ \text{E}\{\hat{\mu}_{OP}\} &\approx \mu_t + \frac{1}{2} \text{Var}\{\hat{\mu}_{OP}\} \frac{\sum_{n=1}^N g_n^3 \bar{y}_n(\mu_t)}{\sum_{n=1}^N g_n^2 \bar{y}_n(\mu_t)}.\end{aligned}$$

**SP Model:**

For SP objective function (3.23), the corresponding derivatives are:

$$\begin{aligned}
\frac{\partial}{\partial \mu} L_{SP}(\mu, y) &= \sum_{n=1}^N g_n \bar{y}_n(\mu) \left[ 1 - \frac{y_n + 2r_n}{\bar{y}_n(\mu) + 2r_n} \right] \\
-\frac{\partial^2}{\partial \mu^2} L_{SP}(\mu, y) &= \sum_{n=1}^N a_n^2 \bar{y}_n(\mu) \left[ 1 - \frac{(y_n + 2r_n)2r_n}{(\bar{y}_n(\mu) + 2r_n)^2} \right] \\
\frac{\partial^3}{\partial \mu^3} L_{SP}(\mu, y) &= \sum_{n=1}^N a_n^3 \bar{y}_n(\mu) \left[ 1 - \frac{(y_n + 2r_n)2r_n(2r_n - \bar{y}_n(\mu))}{(\bar{y}_n(\mu) + 2r_n)^3} \right] \\
\frac{\partial^2}{\partial \mu \partial y_n} L_{SP}(\mu, y) &= -g_n \frac{\bar{y}_n(\mu)}{\bar{y}_n(\mu) + 2r_n} \\
\frac{\partial^3}{\partial \mu \partial y_n^2} L_{SP}(\mu, y) &= 0 \\
\frac{\partial^3}{\partial \mu^2 \partial y_n} L_{SP}(\mu, y) &= a_n^2 \frac{\bar{y}_n(\mu) + 2r_n}{(\bar{y}_n(\mu) + 2r_n)^2}.
\end{aligned}$$

For SP estimator,  $\check{\mu} = f(\bar{y}(\mu)) = \mu_t$ , since  $\frac{\partial}{\partial \mu} L_{SP}(\mu_t, \bar{y}) = 0$ . Computing the values of above equalities at  $\mu_t$  and  $\bar{y}$ :

$$\begin{aligned}
\frac{\partial}{\partial \mu} L_{SP}(\mu_t, \bar{y}) &= 0 \\
-\frac{\partial^2}{\partial \mu^2} L_{SP}(\mu_t, \bar{y}) &= \sum_{n=1}^N a_n^2 \frac{\bar{y}_n^2(\mu_t)}{\bar{y}_n(\mu_t) + 2r_n} \\
\frac{\partial^3}{\partial \mu^3} L_{SP}(\mu_t, \bar{y}) &= \sum_{n=1}^N a_n^3 \frac{\bar{y}_n^2(\mu_t)(\bar{y}_n(\mu_t) + 6r_n)}{(\bar{y}_n(\mu_t) + 2r_n)^2} \\
\frac{\partial^2}{\partial \mu \partial y_n} L_{SP}(\mu_t, \bar{y}) &= -g_n \frac{\bar{y}_n(\mu_t)}{\bar{y}_n(\mu_t) + 2r_n} \\
\frac{\partial^3}{\partial \mu \partial y_n^2} L_{SP}(\mu_t, \bar{y}) &= 0 \\
\frac{\partial^3}{\partial \mu^2 \partial y_n} L_{SP}(\mu_t, \bar{y}) &= g_n \frac{\bar{y}_n(\mu_t) 2r_n}{(\bar{y}_n(\mu_t) + 2r_n)^2}.
\end{aligned}$$

Substituting into (C.3) and (C.4):

$$\frac{\partial}{\partial y_n} f_{SP}(\bar{y}) = -\frac{g_n \frac{\bar{y}_n(\mu_t)}{\bar{y}_n(\mu_t) + 2r_n}}{\sum_{n=1}^N a_n^2 \frac{\bar{y}_n^2(\mu_t)}{\bar{y}_n(\mu_t) + 2r_n}}$$

and

$$\frac{\partial^2}{\partial y_n^2} f_{SP}(\bar{y}) = \left( \frac{\partial}{\partial y_n} f_{SP}(\bar{y}) \right)^2 \left[ \frac{1}{\sum_{n=1}^N a_n^2 \frac{\bar{y}_n^2(\mu_t)}{\bar{y}_n(\mu_t) + 2r_n}} \sum_{m=1}^N a_m^3 \frac{\bar{y}_m^2(\mu_t)(\bar{y}_m(\mu_t) + 6r_m)}{(\bar{y}_m^2(\mu_t) + 2r_m)^2} - \frac{4r_n g_n}{(\bar{y}_n(\mu_t) + 2r_n)} \right].$$

Lastly, substituting into (4.16) and (4.17) with  $\text{Var}(y_n) = \bar{y}_n(\mu_t) + 2r_n$ :

$$\begin{aligned}\text{Var}\{\hat{\mu}_{SP}\} &\approx \left[ \sum_{n=1}^N \frac{g_n^2 \bar{y}_n(\mu_t)^2}{(\bar{y}_n(\mu_t) + 2r_n)} \right]^{-1} \\ \text{E}\{\hat{\mu}_{SP}\} &\approx \mu_t + \frac{1}{2} \text{Var}\{\hat{\mu}_{SP}\} \frac{\sum_{n=1}^N g_n^3 \frac{\bar{y}_n(\mu_t)^2}{\bar{y}_n(\mu_t) + 2r_n}}{\sum_{n=1}^N g_n^2 \frac{\bar{y}_n(\mu_t)^2}{\bar{y}_n(\mu_t) + 2r_n}}.\end{aligned}$$

**Quadratic Model:**

For the 1D problem, WLS objective function (3.15) reduces to:

$$L_{WLS}(\mu) = -\frac{1}{2} \sum_{n=1, y_n > s_n^T}^N \left( g_n \mu - \log \left( \frac{b_n}{y_n - s_n^T} \right) \right)^2 \frac{(y_n - s_n^T)^2}{y_n + 2r_n}.$$

For this specific case, one can find the function  $f(y)$  explicitly. Namely, by zeroing the derivative of the objective function, one obtains the WLS estimate of  $\mu$ :

$$\hat{\mu}_{WLS} = f_{WLS}(y) = \frac{\sum_{n=1, y_n > s_n^T}^N \log \left( \frac{b_n}{y_n - s_n^T} \right) \frac{g_n (y_n - s_n^T)^2}{y_n + 2r_n}}{\sum_{n=1, y_n > s_n^T}^N \frac{a_n^2 (y_n - s_n^T)^2}{y_n + 2r_n}}.$$

Then, for  $s_n^T = 0$ :

$$\frac{\partial}{\partial y_n} f_{WLS}(\bar{y}) = -\frac{g_n \frac{\bar{y}_n(\mu_t)}{\bar{y}_n(\mu_t) + 2r_n}}{\sum_{n=1}^N a_n^2 \frac{\bar{y}_n^2(\mu_t)}{\bar{y}_n(\mu_t) + 2r_n}} = \frac{\partial}{\partial y_n} f_{SP}(\bar{y})$$

which results in:

$$\text{Var}\{\hat{\mu}_{WLS}\} \approx \text{Var}\{\hat{\mu}_{SP}\} \approx \left[ \sum_{n=1}^N \frac{g_n^2 \bar{y}_n(\mu_t)^2}{(\bar{y}_n(\mu_t) + 2r_n)} \right]^{-1}.$$

To derive approximate expression for  $\text{E}\{\hat{\mu}_{WLS}\}$  we considered the simpler WLS estimator, using the approximation  $\frac{y_n^2}{y_n + 2r_n} \approx y_n$ , *i.e.*:

$$\tilde{\mu}_{WLS} = \tilde{f}_{WLS}(y) = \frac{\sum_{n=1, y_n > s_n^T}^N \log \frac{b_n}{y_n} g_n y_n}{\sum_{n=1, y_n > s_n^T}^N a_n^2 y_n}. \quad (\text{C.6})$$

Then,

$$\frac{\partial^2}{\partial y_n^2} \tilde{f}_{WLS}(\bar{y}) = \frac{2a_n^3}{\sum_{n=1}^N a_n^3 \bar{y}_n(\mu_t)} - \frac{g_n}{\bar{y}_n(\mu_t) \sum_{n=1}^N a_n^2 \bar{y}_n(\mu_t)}.$$

Substituting into (4.17) we obtain the approximation:

$$\text{E}\{\hat{\mu}_{WLS}\} \approx \mu_t + \frac{\sum_{n=1}^N g_n^3 (\bar{y}_n(\mu_t) + 2r_n)}{\left( \sum_{n=1}^N g_n^2 \bar{y}_n(\mu_t) \right)^2} - \frac{1}{2} \frac{\sum_{n=1}^N g_n \left( \frac{\bar{y}_n(\mu_t) + 2r_n}{\bar{y}_n(\mu_t)} \right)}{\sum_{n=1}^N g_n^2 \bar{y}_n(\mu_t)}.$$

**SD Model:**

For the 1D SD objective function (3.32), we note that

$$\frac{\partial u_n(\mu)}{\partial \mu} = \frac{2r_n}{u_n(\mu)} (-a_n \bar{y}_n(\mu)).$$

In the following we consider the case  $y_n \geq 0$  for notational simplicity, the expressions for  $y_n < 0$  can be derived in similar way. Using the partials defined in (3.36), (3.37) we obtain:

$$\begin{aligned} \frac{\partial L_{SD}(\mu, y)}{\partial \mu} = \\ \sum_{n=1}^N a_n \bar{y}_n(\mu) \left[ \frac{y_n}{\bar{y}_n(\mu) + r_n} - 1 + \frac{2r_n}{u_n(\mu)} \left( -\frac{y_n}{y_n + 1 + u_n(\mu)} + 1 - \frac{1}{2u_n} \right) \right], \end{aligned} \quad (\text{C.7})$$

$$\begin{aligned} -\frac{\partial^2 L_{SD}(\mu, y)}{\partial \mu^2} = \\ \sum_{n=1}^N a_n^2 \bar{y}_n(\mu) \left[ \frac{y_n}{\bar{y}_n(\mu) + r_n} - 1 + \frac{2r_n}{u_n(\mu)} \left( -\frac{y_n}{y_n + 1 + u_n(\mu)} + 1 - \frac{1}{2r_n} \right) \right] \\ + a_n^2 \bar{y}_n^2(\mu) \left[ \frac{y_n}{(\bar{y}_n(\mu) + r_n)^2} - \frac{4r_n^2 y_n (1 + y_n + 2u_n(\mu))}{u_n^3(\mu) (y_n + 1 + u_n(\mu))^2} - \frac{4r_n^2}{u_n^4} + \frac{4r_n^2}{u_n^3} \right] \end{aligned} \quad (\text{C.8})$$

and

$$\begin{aligned} \frac{\partial^2 L_{SD}(\mu, y)}{\partial \mu \partial y} = \\ a_n \bar{y}_n(\mu) \left[ -\frac{1}{\bar{y}_n(\mu) + r_n} + \frac{2r_n [u_n^2(\mu) + u_n(\mu) - 2y_n(y_n + 1) - \frac{y_n^3 + 2y_n^2 + y_n}{u_n}]}{[u_n^2(\mu) + u_n(\mu)y_n + u_n(\mu)]^2} \right. \\ \left. - \frac{2r_n(y_n + 1)}{u_n^4(\mu)} + \frac{2r_n(y_n + 1)}{u_n^3(\mu)} \right]. \end{aligned} \quad (\text{C.9})$$

One can substitute above expressions into (4.16) and (4.17) to obtain closed form expressions for bias and variance estimates for SD estimator.

## APPENDIX D

### Evaluation of the Conditional Expectation

Let  $U \sim \text{Poisson}(\alpha)$ ,  $V \sim \text{Poisson}(\beta)$  be independent and  $Y = U - V$  with pmf's  $p_U(k)$ ,  $p_V(k)$  and  $p_Y(k)$  respectively. We need to find  $E\{U | Y = y\}$ .

The conditional pmf of U :

$$\begin{aligned}
 P(U = k | Y = y) &= \frac{p_U(k) p_V(k - y)}{P(U - V = y)}, \\
 &= \frac{\frac{\alpha^k e^{-\alpha}}{k!} \frac{\beta^{k-y} e^{-\beta}}{(k-y)!}}{\sum_{m=\lfloor y \rfloor + 1}^{\infty} \frac{\alpha^m e^{-\alpha}}{m!} \frac{\beta^{m-y} e^{-\beta}}{(m-y)!}}, \quad \text{for } k \geq 0 \text{ and } y \leq k \\
 &= \frac{\frac{\alpha^k \beta^{k-y}}{k! (k-y)!}}{\sum_{m=\lfloor y \rfloor + 1}^{\infty} \frac{\alpha^m \beta^{m-y}}{m! (m-y)!}},
 \end{aligned}$$

which results in the conditional expectation:

$$\begin{aligned}
 E\{U | Y = y\} &= \frac{\sum_{k=\lfloor y \rfloor + 1}^{\infty} k \frac{\alpha^k \beta^{k-y}}{k! (k-y)!}}{\sum_{m=\lfloor y \rfloor + 1}^{\infty} \frac{\alpha^m \beta^{m-y}}{m! (m-y)!}}, \\
 &= \frac{\alpha \sum_{l=\lfloor y-1 \rfloor + 1}^{\infty} \frac{\alpha^l \beta^{l-(y-1)}}{l! (l-(y-1))!}}{\sum_{m=\lfloor y \rfloor + 1}^{\infty} \frac{\alpha^m \beta^{m-y}}{m! (m-y)!}}, \\
 &= \frac{\alpha P(U - V = y - 1)}{P(U - V = y)}, \\
 &= \frac{\alpha P(Y = y - 1)}{P(Y = y)},
 \end{aligned}$$

where going from first equality to the second we use the change of variables  $l = k - 1$ .

## APPENDIX E

### Concavity Analysis of the SD Model

In this section we prove that the SD log-likelihood approximation is concave, *i.e.*:  $h_n^{\text{SD}}(l)$ 's in (5.14) are concave for  $l \in [0, \infty)$ . We also investigate the convexity of derivatives of the  $h_n^{\text{SD}}(l)$ 's since we use the paraboloid surrogates maximization algorithm by Fessler and Erdoğan [38] which requires certain convexity conditions of the derivatives of the  $h_n^{\text{SD}}(l)$ 's [28].

In the following we drop the subscript  $n$  and the  $s_n^E$  factors for simplicity, *i.e.*:

$$h^{\text{SD}}(l) = y \log \left( \frac{l+r}{z+u(l)} \right) - (l) + u(l) - \frac{1}{2} \log(u(l)), \quad (\text{E.1})$$

where

$$z = \begin{cases} y+1, & y \geq 0 \\ y-1, & y < 0, \end{cases} \quad (\text{E.2})$$

and

$$u(l) = \sqrt{z^2 + 4(l+r)r}. \quad (\text{E.3})$$

The first and second derivatives of  $h_n^{\text{SD}}(l)$  can be written as:

$$\dot{h}^{\text{SD}}(l) = \left( \frac{y}{l+r} - 1 + \frac{2r}{u(l)} \left[ -\frac{y}{z+u(l)} + 1 - \frac{1}{2u(l)} \right] \right) \quad (\text{E.4})$$

$$\ddot{h}^{\text{SD}}(l) = \left( -\frac{y}{(l+r)^2} - \frac{4r^2}{u(l)^3} \left[ 1 - \frac{y(z+2u(l))}{(z+u(l))^2} - \frac{1}{u(l)} \right] \right). \quad (\text{E.5})$$

Since  $h^{\text{SD}}(l)$  is three times continuously differentiable, it is strictly concave if  $\ddot{h}^{\text{SD}}(l) < 0$ . Dropping the dependence of  $u(l)$  on  $l$  for notational simplicity,  $\ddot{h}^{\text{SD}}(l)$  can be rewritten as:

$$\ddot{h}^{\text{SD}}(l) = - \left( \frac{y}{(l+r)^2} + \frac{4r^2}{u^3} [1-K] \right) \quad (\text{E.6})$$

with

$$K = \frac{y(z+2u)}{(z+u)^2} + \frac{1}{u}, \quad (\text{E.7})$$

**The case  $y \geq 0$ :**

For  $y \geq 0$ , from (E.2)

$$y = z - 1 \quad (\text{E.8})$$

and let

$$x \triangleq \frac{u}{z}. \quad (\text{E.9})$$

Then K (E.7) becomes:

$$K = \frac{(z-1)(1+2x)}{z(1+x)^2} + \frac{1}{zx} \quad (\text{E.10})$$

$$= \frac{z(x+2x^2) - x - 2x^2 + 1 + 2x + x^2}{zx(x+1)^2} \quad (\text{E.11})$$

$$= \frac{z(x+2x^2) - x^2 + x + 1}{zx(x+1)^2} \quad (\text{E.12})$$

and

$$1 - K = \frac{zx^3 + x^2 - x - 1}{zx(x+1)^2} \quad (\text{E.13})$$

$$= \frac{x^3(z-1) + (x+1)^2(x-1)}{zx(x+1)^2} \quad (\text{E.14})$$

Since,  $(l+r) = \frac{(x^2-1)z^2}{4r}$  :

$$\ddot{h}^{\text{SD}}(l) = - \left[ \frac{(z-1)16r^2}{(x^2-1)^2z^4} + \frac{4r^2}{x^3z^3} \frac{x^3(z-1) + (x+1)^2(x-1)}{zx(x+1)^2} \right] \quad (\text{E.15})$$

$$= \frac{-4r^2}{z^4(x+1)^2} \left[ \frac{4(z-1)}{(x-1)^2} + \frac{x^3(z-1) + (x+1)^2(x-1)}{x^4} \right] \quad (\text{E.16})$$

$$= \frac{-4r^2}{z^4(x+1)^2} \left[ \frac{4x^4(z-1) + x^3(z-1)(x-1)^2 + (x+1)^2(x-1)^3}{(x-1)^2x^4} \right] \quad (\text{E.17})$$

$$= \frac{-4r^2}{z^4(x-1)^2x^4} \left[ (z-1)x^3 + (x-1)^3 \right] \quad (\text{E.18})$$

$$= \frac{-4r^2}{z^4} [f(l) + g(l)], \quad (\text{E.19})$$

with

$$f(l) \triangleq \frac{z-1}{(x-1)^2x} \quad (\text{E.20})$$

and

$$g(l) \triangleq \frac{(x-1)}{x^4}. \quad (\text{E.21})$$

Since  $z \geq 1$ ,  $r > 0$ , then

$$x = \frac{\sqrt{z^2 + 4(l+r)r}}{z} > 1. \quad (\text{E.22})$$

Thus for  $l \in [0, \infty)$ ,  $f(l) \geq 0$  and  $g(l) > 0$ , and consequently  $\ddot{h}^{\text{SD}}(l) < 0$ , proving that  $h^{\text{SD}}(l)$  is strictly concave for  $y \geq 0$ .  $\diamond$

To investigate the maximum of  $-\ddot{h}^{\text{SD}}(l)$  and convexity of  $\dot{h}^{\text{SD}}(l)$  in  $[0, \infty)$  we compute  $h^{\text{SD}(3)}(l)$ :

$$h^{\text{SD}(3)}(l) = \frac{-4r^2}{z^4} [f(l) + g(l)], \quad (\text{E.23})$$

with

$$\dot{f}(l) = \frac{-(z-1)(3x^2-4x+1)2r}{(x-1)^4 x^2 z^2 x} \quad (\text{E.24})$$

$$= \frac{-2r(z-1)(3x-1)}{z^2 x^3 (x-1)^3}. \quad (\text{E.25})$$

Also,

$$\dot{g}(l) = \frac{x^4 - (x-1)4x^3}{x^8} \frac{2r}{z^2 x} \quad (\text{E.26})$$

$$= \frac{-2r(3x-4)}{z^2 x^6}. \quad (\text{E.27})$$

Rewriting  $h^{\text{SD}(3)}(l)$  (E.23) :

$$h^{\text{SD}(3)}(l) = \frac{8r^3}{z^6 x^3} [f_d(l) + g_d(l)] \quad (\text{E.28})$$

with

$$f_d(l) \triangleq \frac{(z-1)(3x-1)}{(x-1)^3}, \quad (\text{E.29})$$

and

$$g_d(l) \triangleq \frac{(3x-4)}{x^3}. \quad (\text{E.30})$$

Fig. E.1 shows the plots of  $f_d(x)$  and  $g_d(x)$  for  $x > 1$ .

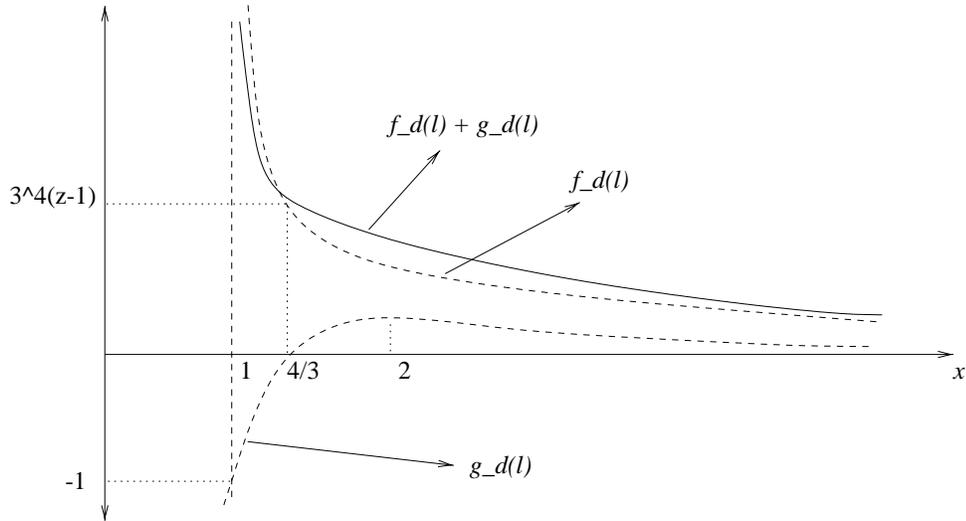


Figure E.1: Plots of  $f_d(x)$  and  $g_d(x)$  and their sum for  $x > 1$ .

**The case  $y = 0$ :**

It can be seen from (E.29) and (E.30) that for  $x > 1$ ,  $f_d(x) = 0$  and

$$g_d(x) = \begin{cases} < 0, & x \in [1, 4/3] \\ = 0, & x = 4/3 \\ > 0, & x \in [4/3, \infty) \end{cases} \quad (\text{E.31})$$

Thus  $h^{SD(3)}(l)$  has a zero crossing at  $l = l_1^*$ :

$$x = 4/3 = \frac{1}{z} \sqrt{z^2 + 4(l_1^* + r)r} \quad (\text{E.32})$$

$$\frac{16}{9} = 1 + 4(l_1^* r) + 4r^2 \quad (\text{E.33})$$

$$l_1^* = \frac{1}{4r} \left( \frac{7}{9} - 4r^2 \right) \quad (\text{E.34})$$

and

$$h^{SD(3)}(l) = \begin{cases} < 0, & r \leq \frac{\sqrt{7}}{6}, l < l_1^* \\ = 0, & r \leq \frac{\sqrt{7}}{6}, l = l_1^* \\ > 0, & l > \max[0, l_1^*] \end{cases} \quad (\text{E.35})$$

**The case  $y \geq 1$ :**

It can be seen from (E.28,E.29,E.30) and Fig. E.1 that  $h^{SD(3)}(l) > 0$  for  $x \in [1, \infty)$  and thus for  $l \in [0, \infty)$ .

**The case  $y < 0$ :**

For  $y < 0$ , from (E.2)

$$y = z + 1 \quad (\text{E.36})$$

and again let

$$x \triangleq \frac{u}{z}. \quad (\text{E.37})$$

Then K (E.7) becomes:

$$K = \frac{(z+1)(1+2x)}{z(1+x)^2} + \frac{1}{zx} \quad (\text{E.38})$$

$$= \frac{z(x+2x^2) + x + 2x^2 + 1 + 2x + x^2}{zx(x+1)^2} \quad (\text{E.39})$$

$$= \frac{z(x+2x^2) + 3x^2 + 3x + 1}{zx(x+1)^2} \quad (\text{E.40})$$

and

$$1 - K = \frac{zx^3 - 3x^2 - 3x - 1}{zx(x+1)^2} \quad (\text{E.41})$$

$$= \frac{x^3(z+1) - (x+1)^3}{zx(x+1)^2} \quad (\text{E.42})$$

Since,  $(l+r) = \frac{(x^2-1)z^2}{4r}$  :

$$\ddot{h}^{\text{SD}}(l) = - \left[ \frac{(z+1)16r^2}{(x^2-1)^2z^4} + \frac{4r^2}{x^3z^3} \frac{x^3(z+1) - (x+1)^3}{zx(x+1)^2} \right] \quad (\text{E.43})$$

$$= \frac{-4r^2}{z^4(x+1)^2} \left[ \frac{4(z+1)}{(x-1)^2} - \frac{x^3(z+1) - (x+1)^3}{x^4} \right] \quad (\text{E.44})$$

$$= \frac{-4r^2}{z^4(x+1)^2} \left[ \frac{4x^4(z+1) + x^3(z+1)(x-1)^2 - (x+1)^3(x-1)^2}{(x-1)^2x^4} \right] \quad (\text{E.45})$$

$$= \frac{-4r^2}{z^4(x-1)^2x^4} \left[ (z+1)x^3 - (x+1)(x-1)^2 \right] \quad (\text{E.46})$$

$$= \frac{-4r^2}{z^4} [f(l) + g(l)], \quad (\text{E.47})$$

with

$$f(l) = \frac{z+1}{(x-1)^2x} \quad (\text{E.48})$$

and

$$g(l) = \frac{-(x+1)}{x^4}. \quad (\text{E.49})$$

Since  $z < -1$ ,  $r > 0$ , then

$$x = \frac{\sqrt{z_n^2 + 4(l+r_n)r_n}}{z} < -1. \quad (\text{E.50})$$

Thus for  $l \in [0, \infty)$ ,  $f(l) > 0$  and  $g(l) > 0$ , and consequently  $\ddot{h}^{\text{SD}}(l) < 0$ , proving that  $h^{\text{SD}}(l)$  is strictly concave for  $y < 0$ .  $\diamond$

In order to investigate the maximum of  $-\ddot{h}^{\text{SD}}(l)$  and convexity of  $\dot{h}^{\text{SD}}(l)$  in  $[0, \infty)$  we compute  $h^{\text{SD}(3)}(l)$ :

$$h^{\text{SD}(3)}(l) = \frac{-4r^2}{z^4} [f'(l) + g'(l)], \quad (\text{E.51})$$

with

$$\dot{f}(l) = \frac{-(z+1)}{(x-1)^4x^2} \frac{(3x^2-4x+1)2r}{z^2x} \quad (\text{E.52})$$

$$= \frac{-2r(z+1)(3x-1)}{z^2x^3(x-1)^3}. \quad (\text{E.53})$$

Also,

$$\dot{g}(l) = \frac{-x^4 + (x+1)4x^3}{x^8} \frac{2r}{z^2x} \quad (\text{E.54})$$

$$= \frac{2r(3x+4)}{z^2x^6}. \quad (\text{E.55})$$

Rewriting  $h^{\text{SD}(3)}(l)$  (E.51) :

$$h^{\text{SD}(3)}(l) = \frac{8r^3}{z^6x^3} [f_d(l) + g_d(l)] \quad (\text{E.56})$$

with

$$f_d(l) \triangleq \frac{(z+1)(3x-1)}{(x-1)^3}, \quad (\text{E.57})$$

and

$$g_d(l) \triangleq \frac{-3x-4}{x^3}. \quad (\text{E.58})$$

Fig. E.2 shows the plots of  $f_d(x)$  and  $g_d(x)$  for  $x < -1$ .

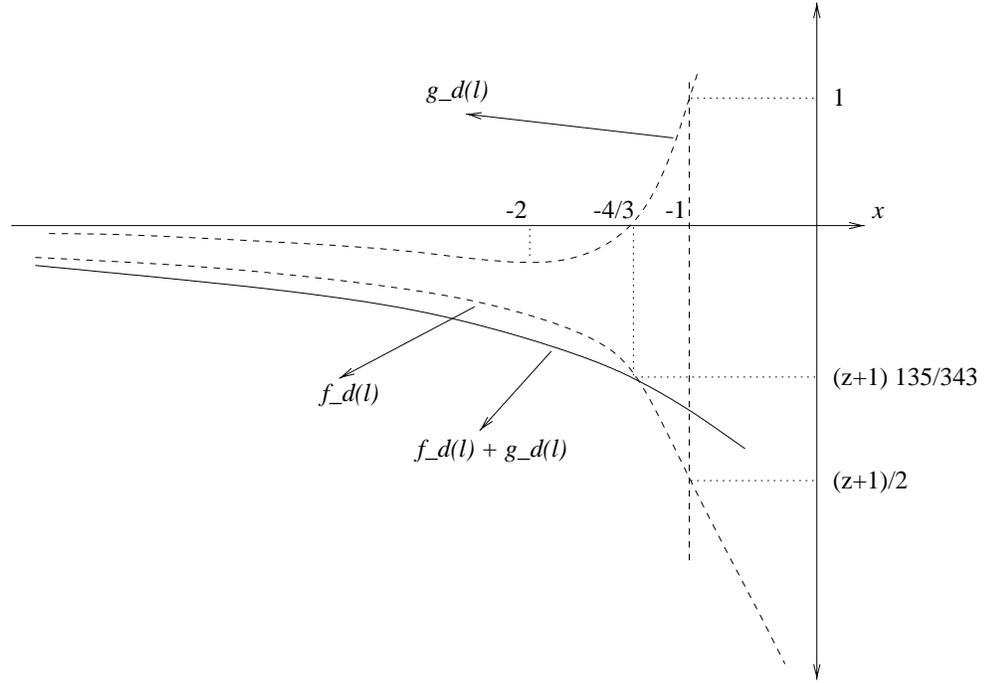


Figure E.2: Plots of  $f_d(x)$  and  $g_d(x)$  and their sum for  $x < -1$ .

Lastly we can rewrite  $h^{SD(3)}(l)$  (E.56) as:

$$h^{SD(3)}(l) = \frac{8r^3}{z^6 x^3} \left[ \frac{3zx^4 + (4-z)x^3 + 3x^2 - 9x + 4}{x^3(x-1)^3} \right] \quad (\text{E.59})$$

**The case  $y = -1$ :**

For this case, the only real root of  $h^{SD(3)}(x)$  (E.59) for  $x \leq -1$  can be found as

$$x_o = -1.1193219 \quad (\text{E.60})$$

Thus  $h^{SD(3)}(l)$  has a zero crossing at  $l = l_2^*$  as:

$$x = x_o = \frac{1}{z} \sqrt{z^2 + 4(l_2^* + r)r} \quad (\text{E.61})$$

$$x_o = \frac{1}{-2} \sqrt{4 + 4(l_2^* r) + 4r^2} \quad (\text{E.62})$$

$$x_o^2 = 1 + l_2^* r + r^2 \quad (\text{E.63})$$

$$l_2^* = \frac{x_o^2 - 1 - r^2}{r} \quad (\text{E.64})$$

and

$$h^{SD(3)}(l) = \begin{cases} < 0, & r \leq \sqrt{x_o^2 - 1}, l < l_2^* \\ = 0, & r \leq \sqrt{x_o^2 - 1}, l = l_2^* \\ > 0, & l > \max[0, l_2^*] \end{cases} \quad (\text{E.65})$$

**The case  $y \leq -2$ :**

It can be shown using (E.59), that  $h^{SD(3)}(l)$  does not have any real root for  $x \in (-\infty, -1]$  and thus  $h^{SD(3)}(l) > 0$  for  $l \in [0, \infty)$ .

## BIBLIOGRAPHY

## BIBLIOGRAPHY

- [1] J. M. M. Anderson, B. A. Mair, M. Rao, and C.-H. Wu, “Weighted least-squares reconstruction methods for positron emission tomography,” *IEEE Tr. Med. Im.*, vol. 16, no. 2, pp. 159–65, April 1997.
- [2] S. L. Bacharach, M. A. Douglas, R. E. Carson, P. J. Kalkowski, N. M. T. Freedman, P. Perrone-Filardi, and R. O. Bonow, “Three dimensional registration of cardiac PET attenuation scans,” *J. Nuc. Med. (Abs. Book)*, vol. 33, no. 5, pp. 881, May 1992.
- [3] J. Besag, “On the statistical analysis of dirty pictures,” *J. Royal Stat. Soc. Ser. B*, vol. 48, no. 3, pp. 259–302, 1986.
- [4] T. Beyer, P. E. Kinahan, and D. W. Townsend, “Optimization of transmission and emission scan duration in 3D whole-body PET,” *IEEE Tr. Nuc. Sci.*, vol. 44, no. 6, pp. 2400–7, December 1997.
- [5] C. Bouman and K. Sauer, “Fast numerical methods for emission and transmission tomographic reconstruction,” in *Proc. 27th Conf. Info. Sci. Sys., Johns Hopkins*, pp. 611–616, 1993.
- [6] C. A. Bouman and K. Sauer, “A unified approach to statistical tomography using coordinate descent optimization,” *IEEE Tr. Im. Proc.*, vol. 5, no. 3, pp. 480–92, March 1996.
- [7] W. L. Briggs, *A multigrid tutorial*, SIAM, Philadelphia, 1987.
- [8] C. S. Butler and M. I. Miller, “Maximum a Posteriori estimation for SPECT using regularization techniques on massively parallel computers,” *IEEE Tr. Med. Im.*, vol. 12, no. 1, pp. 84–89, March 1993.
- [9] 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. Comp. Assisted Tomo.*, vol. 10, no. 5, pp. 845–850, 1986.
- [10] A. Chatziioannou and M. Dahlbom, “Detailed investigation of transmission and emission data smoothing protocols and their effects on emission images,” in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 4, pp. 1568–72, 1994.
- [11] A. Chatziioannou and M. Dahlbom, “Detailed investigation of transmission and emission data smoothing protocols and their effects on emission images,” *IEEE Tr. Nuc. Sci.*, vol. 43, no. 1, pp. 290–4, February 1996.
- [12] S. R. Cherry, M. Dahlbom, and E. J. Hoffman, “High sensitivity, total body PET scanning using 3D data acquisition and reconstruction,” *IEEE Tr. Nuc. Sci.*, vol. 39, no. 4, pp. 1088–1092, August 1992.
- [13] Z. H. Cho, J. P. Jones, and M. Singh, *Foundations of medical imaging*, Wiley, New York, 1993.
- [14] P. G. Ciarlet, *Introduction to numerical linear algebra and optimisation*, Cambridge, Cambridge, 1982.
- [15] N. H. Clinthorne, C.-Y. Ng, C.-H. Hua, J. E. Gormley, J. W. Leblanc, S. J. Wilderman, and W. L. Rogers, “Theoretical performance comparison of a Compton-scatter aperture and parallel-hole collimator,” in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, 1996.

- [16] N. H. Clinthorne, T. S. Pan, P. C. Chiao, W. L. Rogers, and J. A. Stamos, "Preconditioning methods for improved convergence rates in iterative reconstructions," *IEEE Tr. Med. Im.*, vol. 12, no. 1, pp. 78–83, March 1993.
- [17] N. H. Clinthorne, S. J. Wilderman, J. E. Gormley, G. F. Knoll, D. K. Wehe, and W. L. Rogers, "Theoretical performance limits for electronically collimated single photon imaging systems," *J. Nuc. Med.*, vol. 37, no. 5, pp. 119, May 1996.
- [18] M. Dahlbom and E. J. Hoffman, "Problems in signal-to-noise ratio for attenuation correction in high resolution PET," *IEEE Tr. Nuc. Sci.*, vol. 34, no. 1, pp. 288–293, February 1987.
- [19] G. de Castro, "Note on differences of Bernoulli and Poisson variables," *Portugaliae Mathematica*, vol. 11, pp. 173–5, 1952.
- [20] A. R. De Pierro, "On the relation between the ISRA and the EM algorithm for positron emission tomography," *IEEE Tr. Med. Im.*, vol. 12, no. 2, pp. 328–333, June 1993.
- [21] A. R. De Pierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE Tr. Med. Im.*, vol. 14, no. 1, pp. 132–137, March 1995.
- [22] C. B. Dean, "A robust property of pseudo-likelihood estimation for count data," *Journal of Statistical Planning and Inference*, vol. 35, no. 3, pp. 309–17, 1993.
- [23] S. Deans, *The Radon transform and some of its applications*, Wiley, New York, 1983. 1993 2nd edition by Krieger Publishing Co., Malabar, Florida.
- [24] M. Defrise, R. Clask, and D. Townsend, "Solution to the three-dimensional image reconstruction problem from two-dimensional parallel projections," *J. Opt. Soc. Am. A*, vol. 10, no. 5, pp. 869–877, May 1993.
- [25] M. Defrise, P. E. Kinahan, D. W. Townsend, C. Michel, M. Sibomana, and D. F. Newport, "Exact and approximate rebinning algorithms for 3-D PET data," *IEEE Tr. Med. Im.*, vol. 16, no. 2, pp. 145–58, April 1997.
- [26] 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.
- [27] H. Erdoğan and J. A. Fessler, "Scan time optimization for post-injection PET scans," in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 3, pp. 1842–6, 1998.
- [28] H. Erdoğan and J. A. Fessler, "Monotonic algorithms for transmission tomography," *IEEE Tr. Med. Im.*, vol. 18, no. 9, pp. 801–14, September 1999.
- [29] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone beam algorithm," *J. Opt. Soc. Am. A*, vol. 1, no. 6, pp. 612–9, 1984.
- [30] J. A. Fessler, "Tomographic reconstruction using information weighted smoothing splines," in *Information Processing in Medical Im.*, H. H. Barrett and A. F. Gmitro, editors, volume 687 of *Lecture Notes in Computer Science*, pp. 372–86, Springer Verlag, Berlin, 1993.
- [31] J. A. Fessler, "Penalized weighted least-squares image reconstruction for positron emission tomography," *IEEE Tr. Med. Im.*, vol. 13, no. 2, pp. 290–300, June 1994.
- [32] J. A. Fessler, "Hybrid Poisson/polynomial objective functions for tomographic image reconstruction from transmission scans," *IEEE Tr. Im. Proc.*, vol. 4, no. 10, pp. 1439–50, October 1995.
- [33] J. A. Fessler, "Resolution properties of regularized image reconstruction methods," Technical Report 297, Comm. and Sign. Proc. Lab., Dept. of EECS, Univ. of Michigan, Ann Arbor, MI, 48109-2122, August 1995.

- [34] J. A. Fessler, "Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography," *IEEE Tr. Im. Proc.*, vol. 5, no. 3, pp. 493–506, March 1996.
- [35] J. A. Fessler, "Approximate variance images for penalized-likelihood image reconstruction," in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 2, pp. 949–52, 1997.
- [36] J. A. Fessler and S. D. Booth, "Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction," *IEEE Tr. Im. Proc.*, vol. 8, no. 5, pp. 688–99, May 1999.
- [37] J. A. Fessler, N. H. Clinthorne, and W. L. Rogers, "On complete data spaces for PET reconstruction algorithms," *IEEE Tr. Nuc. Sci.*, vol. 40, no. 4, pp. 1055–61, August 1993.
- [38] J. A. Fessler and H. Erdođan, "A paraboloidal surrogates algorithm for convergent penalized-likelihood emission image reconstruction," in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 2, pp. 1132–5, 1998.
- [39] J. A. Fessler, E. P. Ficaro, N. H. Clinthorne, and K. Lange, "Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction," *IEEE Tr. Med. Im.*, vol. 16, no. 2, pp. 166–75, April 1997.
- [40] J. A. Fessler and A. O. Hero, "Space-alternating generalized EM algorithms for penalized maximum-likelihood image reconstruction," Technical Report 286, Comm. and Sign. Proc. Lab., Dept. of EECS, Univ. of Michigan, Ann Arbor, MI, 48109-2122, February 1994.
- [41] J. A. Fessler and A. O. Hero, "Space-alternating generalized expectation-maximization algorithm," *IEEE Tr. Sig. Proc.*, vol. 42, no. 10, pp. 2664–77, October 1994.
- [42] J. A. Fessler and A. O. Hero, "Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms," *IEEE Tr. Im. Proc.*, vol. 4, no. 10, pp. 1417–29, October 1995.
- [43] J. A. Fessler and W. L. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction methods: Space-invariant tomographs," *IEEE Tr. Im. Proc.*, vol. 5, no. 9, pp. 1346–58, September 1996.
- [44] M. Fisz, "The limiting distribution of the difference of two Poisson random variables," *Zastosowania Matematyki*, vol. 1, pp. 41–5, 1953.
- [45] S. Geman and D. E. McClure, "Bayesian image analysis: an application to single photon emission tomography," in *Proc. of Stat. Comp. Sect. of Amer. Stat. Assoc.*, pp. 12–18, 1985.
- [46] P. J. Green, "Bayesian reconstructions from emission tomography data using a modified EM algorithm," *IEEE Tr. Med. Im.*, vol. 9, no. 1, pp. 84–93, March 1990.
- [47] T. Hebert and R. Leahy, "A Bayesian reconstruction algorithm for emission tomography using a Markov random field prior," in *Proc. SPIE 1092, Med. Im. III: Im. Proc.*, pp. 458–466, 1989.
- [48] T. Hebert and R. Leahy, "A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors," *IEEE Tr. Med. Im.*, vol. 8, no. 2, pp. 194–202, June 1989.
- [49] C. Helstrom, "Approximate evaluation of detection probabilities in radar and optical communications," *IEEE Tr. Aero. Elec. Sys.*, vol. 14, no. 4, pp. 630–40, July 1978.
- [50] G. T. Herman, *Image reconstruction from projections: The fundamentals of computerized tomography*, Academic Press, New York, 1980.
- [51] A. O. Hero, J. A. Fessler, and M. Usman, "Exploring estimator bias-variance tradeoffs using the uniform CR bound," *IEEE Tr. Sig. Proc.*, vol. 44, no. 8, pp. 2026–41, August 1996.
- [52] E. J. Hoffman, T. M. Guerrero, G. Germano, W. M. Digby, and M. Dahlbom, "PET system calibration and corrections for quantitative and spatially accurate images," *IEEE Tr. Nuc. Sci.*, vol. 36, no. 1, pp. 1108–1112, February 1989.

- [53] 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. Comp. Assisted Tomo.*, vol. 5, no. 3, pp. 391–400, 1981.
- [54] 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. Comp. Assisted Tomo.*, vol. 3, no. 6, pp. 804–814, December 1979.
- [55] R. H. Huesman, S. E. Derenzo, J. L. Cahoon, A. B. Geyer, W. W. Moses, D. C. Uber, T. Vuletich, and T. F. Budinger, "Orbiting transmission source for positron tomography," *IEEE Tr. Nuc. Sci.*, vol. 35, no. 1, pp. 735–739, February 1988.
- [56] N. L. Johnson, S. Kotz, and A. W. Kemp, *Univariate discrete distributions*, Wiley, New York, 1992.
- [57] W. F. Jones, W. M. Digby, W. K. Luk, M. E. Casey, and L. B. Byars, "Optimizing rod window width in positron emission tomography," *IEEE Tr. Med. Im.*, vol. 14, no. 2, pp. 266–270, June 1995.
- [58] A. C. Kak and M. Slaney, *Principles of computerized tomographic imaging*, IEEE Press, New York, 1988.
- [59] K. Lange, "An overview of Bayesian methods in image reconstruction," in *Proc. SPIE 1351, Dig. Im. Synth. and Inverse Optics*, pp. 270–287, 1990.
- [60] K. Lange, "Convergence of EM image reconstruction algorithms with Gibbs smoothing," *IEEE Tr. Med. Im.*, vol. 9, no. 4, pp. 439–446, December 1990. Corrections, T-MI, 10:2(288), June 1991.
- [61] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comp. Assisted Tomo.*, vol. 8, no. 2, pp. 306–316, April 1984.
- [62] K. Lange and J. A. Fessler, "Globally convergent algorithms for maximum a posteriori transmission tomography," *IEEE Tr. Im. Proc.*, vol. 4, no. 10, pp. 1430–8, October 1995.
- [63] Z. Liang and R. E. Coleman, "Restoration for detector response in high resolution PET image reconstruction," *J. Nuc. Med. (Abs. Book)*, vol. 33, no. 5, pp. 872, May 1992.
- [64] A. Macovski, *Medical imaging systems*, Prentice-Hall, New Jersey, 1983.
- [65] S. H. Manglos, R. J. Jaszczak, C. E. Floyd, L. J. Hahn, K. L. Greer, and R. E. Coleman, "A quantitative comparison of attenuation-weighted backprojection with multiplicative and iterative postprocessing attenuation compensation in SPECT," *IEEE Tr. Med. Im.*, vol. 7, no. 2, pp. 127–134, June 1988.
- [66] S. R. Meikle, M. Dahlbom, and S. R. Cherry, "Accuracy of attenuation correction PET due to transmission processing," *J. Nuc. Med. (Abs. Book)*, vol. 33, no. 5, pp. 862, May 1992.
- [67] S. R. Meikle, M. Dahlbom, and S. R. Cherry, "Attenuation correction using count-limited transmission data in positron emission tomography," *J. Nuc. Med.*, vol. 34, no. 1, pp. 143–150, January 1993.
- [68] E. U. Mumcuoglu, R. Leahy, S. R. Cherry, and Z. Zhou, "Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images," *IEEE Tr. Med. Im.*, vol. 13, no. 3, pp. 687–701, December 1994.
- [69] 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, September 1996.
- [70] F. Natterer, *The mathematics of computerized tomography*, Teubner-Wiley, Stuttgart, 1986.
- [71] J. M. Ollinger, "Model-based scatter correction for fully 3D PET," *Phys. Med. Biol.*, vol. 41, no. 1, pp. 153–76, January 1996.

- [72] J. M. Ollinger and J. A. Fessler, "Positron emission tomography," *IEEE Sig. Proc. Mag.*, vol. 14, no. 1, pp. 43–55, January 1997.
- [73] F. O'Sullivan, Y. Pawitan, and D. Haynor, "Reducing negativity artifacts in emission tomography: post-processing filtered backprojection solutions," *IEEE Tr. Med. Im.*, vol. 12, no. 4, pp. 653–663, December 1993.
- [74] D. G. Politte and D. L. Snyder, "Corrections for accidental coincidences and attenuation in maximum-likelihood image reconstruction for positron-emission tomography," *IEEE Tr. Med. Im.*, vol. 10, no. 1, pp. 82–89, March 1991.
- [75] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical recipes in C*, Cambridge Univ. Press, 1988.
- [76] J. Radon, "On the determination of functions from their integrals along certain manifold," *Berichte Sächs. Akad. Wiss. (Leipzig)*, vol. 69, pp. 262–78, 1917. Über die Bestimmung von Funktionen durch ihre Intergralwerte Langs gewisser Mannigfaltigkeiten.
- [77] 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, November 1968.
- [78] J. Romani, "Distribucion de la suma algebraica de variables de Poisson," *Trabajos de Estadisca*, vol. 7, pp. 175–81, 1956.
- [79] K. Sauer and C. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Tr. Sig. Proc.*, vol. 41, no. 2, pp. 534–548, February 1993.
- [80] L. A. Shepp and B. F. Logan, "The Fourier reconstruction of a head section," *IEEE Tr. Nuc. Sci.*, vol. 21, no. 3, pp. 21–43, June 1974.
- [81] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Tr. Med. Im.*, vol. 1, no. 2, pp. 113–122, October 1982.
- [82] J. G. Skellam, "The frequency distribution of the difference between two Poisson variates belonging to different populations," *J. Royal Stat. Soc.*, vol. 109, no. 3, pp. 296, 1946.
- [83] 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. Am. A*, vol. 12, no. 2, pp. 272–83, February 1995.
- [84] D. L. Snyder, M. I. Miller, L. J. Thomas, and D. G. Politte, "Noise and edge artifacts in maximum-likelihood reconstructions for emission tomography," *IEEE Tr. Med. Im.*, vol. 6, no. 3, pp. 228–238, September 1987.
- [85] J. A. Sorenson and M. E. Phelps, *Physics in nuclear medicine*, Saunders, Philadelphia, 2 edition, 1987.
- [86] T. J. Spinks, T. Jones, M. C. Gilardi, and J. D. Heather, "Physical performance of the latest generation of commercial positron scanner," *IEEE Tr. Nuc. Sci.*, vol. 35, no. 1, pp. 721–725, February 1988.
- [87] J. W. Stayman and J. A. Fessler, "Regularization for uniform spatial resolution properties in penalized-likelihood PET reconstruction," in *Biomedical Imaging: Beyond Diagnostics*, p. A4, 1999.
- [88] J. W. Stayman and J. A. Fessler, "Regularization for uniform spatial resolution properties in penalized-likelihood image reconstruction," *IEEE Tr. Med. Im.*, vol. 19, no. 6, pp. 601–15, June 2000.
- [89] C. J. Thompson et al., "A technique to reject scatter radiation in PET transmission scans," in *Proc. SPIE 671, Physics and Engineering of Computerized Multidimensional Im. and Processing*, pp. 244–253, 1986.

- [90] D. W. Townsend, A. Geissbuhler, M. Defrise, E. J. Hoffman, T. J. Spinks, D. L. Bailey, M. C. Gilardi, and T. Jones, "Fully three-dimensional reconstruction for a PET camera with retractable septa," *IEEE Tr. Med. Im.*, vol. 10, no. 4, pp. 505–512, December 1991.
- [91] M. Usman, A. O. Hero, and J. A. Fessler, "Bias-variance tradeoffs analysis using uniform CR bound for image reconstruction," in *Proc. IEEE Intl. Conf. on Image Processing*, volume 2, pp. 835–839, 1994.
- [92] M. Usman, A. O. Hero, and J. A. Fessler, "Uniform CR bound: implementation issues and applications," in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 3, pp. 1443–1447, 1994.
- [93] H. L. Van Trees, *Detection, estimation, and modulation theory*, Wiley, New York, 1968.
- [94] E. Veklerov and J. Llacer, "Stopping rule for the MLE algorithm based on statistical hypothesis testing," *IEEE Tr. Med. Im.*, vol. 6, no. 4, pp. 313–319, December 1987.
- [95] R. W. M. Wedderburn, "Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method," *Biometrika*, vol. 61, no. 3, pp. 439–47, 1974.
- [96] L. Xuan, C. Comtat, C. Michel, P. Kinahan, M. Defrise, and D. Townsend, "Comparison of 3D reconstruction with OSEM and FORE+OSEM for PET," in *Proc. of the 1999 Intl. Mtg. on Fully 3D Im. Recon. in Rad. Nuc. Med.*, pp. 39–42, 1999.
- [97] M. Yavuz and J. A. Fessler, "Objective functions for tomographic reconstruction from randoms-precorrected PET scans," in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 2, pp. 1067–71, 1996.
- [98] M. Yavuz and J. A. Fessler, "New statistical models for randoms-precorrected PET scans," in *Information Processing in Medical Im.*, J. Duncan and G. Gindi, editors, volume 1230 of *Lecture Notes in Computer Science*, pp. 190–203, Springer Verlag, Berlin, 1997.
- [99] M. Yavuz and J. A. Fessler, "Statistical image reconstruction methods for randoms-precorrected PET scans," *Med. Im. Anal.*, vol. 2, no. 4, pp. 369–378, 1998.
- [100] M. Yavuz and J. A. Fessler, "Penalized-likelihood estimators and noise analysis for randoms-precorrected PET transmission scans," *IEEE Tr. Med. Im.*, vol. 18, no. 8, pp. 665–74, August 1999.
- [101] D. M. Young, *Iterative solution of large linear systems*, Academic Press, New York, 1971.
- [102] D. F. Yu and J. A. Fessler, "Mean and variance of photon counting with deadtime," in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 3, pp. 1470–4, 1999.