

**Statistical Image Reconstruction Algorithms Using
Paraboloidal Surrogates for PET Transmission
Scans**

by

Hakan Erdoğan

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering: Systems)
in The University of Michigan
1999

Doctoral Committee:

Associate Professor Jeffrey A. Fessler, Chair

Professor W. Leslie Rogers

Professor Andrew E. Yagle

Senior Associate Research Scientist Neal H. Clinthorne

ABSTRACT

Statistical Image Reconstruction Algorithms Using Paraboloidal Surrogates for PET Transmission Scans

by

Hakan Erdoğan

Chair: Jeffrey A. Fessler

Positron Emission Tomography (PET) is a diagnostic imaging tool that provides images of radioactive substances injected into the body to trace biological functions. The radioactive substance emits a positron which annihilates with an electron to produce two 511 keV photons traveling in approximately opposite directions to be coincidentally detected by two detectors. Many photons are absorbed or scattered, reducing the number of detected emission events.

Attenuation correction is crucial for quantitatively accurate PET reconstructions. PET transmission scans are performed to estimate attenuation parameters which are in turn used to correct the emission scans for attenuation effects. The noise in estimating the attenuation parameters propagates to the emission images affecting their quality and quantitative correctness. Thus, attenuation image reconstruction is extremely important in PET.

Conventional methods of attenuation correction are suboptimal and ignore the Poisson nature of the data. We propose to use penalized likelihood image reconstruction techniques for transmission scans. Current algorithms for transmission tomography have two important problems: 1) they are not guaranteed to converge, 2) if they converge, the convergence is slow. We develop new fast and monotonic opti-

mization algorithms for penalized likelihood image reconstruction based on a novel paraboloidal surrogates principle. We present results showing the speed of the new optimization algorithms as compared to previous ones. We apply the algorithms to PET data obtained from an anthropomorphic thorax phantom and real patient data.

A transmission scan performed after the patient is injected is called a post-injection transmission scan. Post-injection transmission scans are desirable since the patient throughput is increased and motion artifacts are reduced as compared to pre-injection scans. However, there are emission counts contaminating the measurements. We include emission contamination in the post-injection transmission measurement statistical model and obtain better images as compared to conventional subtraction based approaches. We also present a joint estimation technique to estimate attenuation and emission images simultaneously from transmission and emission scans.

We analyze noise propagation from transmission scans to emission images for some sequential image reconstruction methods. The results show that transmission noise affects emission image quality heavily, especially when short transmission scans are utilized.

© Hakan Erdoğan 1999
All Rights Reserved

I dedicate this thesis to my wife Meryem and my son Ahmed Melih for their love,
patience and support during this work.

ACKNOWLEDGEMENTS

I would like to express my thanks to my supervisor Dr. Jeff Fessler for his excellent guidance and encouragement, his patience in discussing my questions and listening and respecting my ideas throughout my Ph.D. work. It was his financial and professional support that enabled me to finish this work.

I would also like to thank Scientific and Technical Research Council of Turkey (TUBITAK) for their financial support by scholarship for the first year of my study. Dearest thanks also goes to Dr. Emad Ebbini for his general encouragement and his help with teaching assistantships early in my Ph.D. work. My supervisor Dr. Jeff Fessler supported me with research assistantship from grants funded by National Institute of Health since 1995 and I thank them for that support.

I would like to express my gratitude to Dr. Ken Lange for his help and discussions in developing the PSCD algorithm and Gene Gualtieri for suggesting the need for an OSTR algorithm, Web Stayman for implementing LBFGS algorithm for comparison with PSCD, and professors Alfred Hero, Les Rogers, Andrew Yagle, Neal Clinthorne for serving in my committee and helping with their ideas.

I extend my gratitude to my colleagues: Osama Haddadin, Robinson Piramuthu, Anne Sauve, Claudio Simon, Web Stayman, Nan Sotthivirat, Mehmet Yavuz, Dan Yu with whom I shared the same or neighboring offices. It is the friendships and company of my fellow Turkish graduate students and their families that deserve credit in giving moral support to me and my family while I was away from my country. I thank all of them dearly as well.

Finally, I thank my family, my wife Meryem and my little son Ahmed Melih, for their loving support and patience. I cherish the color they brought to my life and thank for the sacrifice they made during this hard period in my life.

TABLE OF CONTENTS

DEDICATION	ii
ACKNOWLEDGEMENTS	iii
LIST OF TABLES	viii
LIST OF FIGURES	ix
LIST OF APPENDICES	xii
CHAPTERS	
1 Introduction	1
1.1 Motivation	1
1.2 Background and Significance	6
1.3 Organization of Dissertation	12
1.4 Original Contributions	13
2 Physics of PET and Statistical Modeling	14
2.1 Computed Tomography and FBP	14
2.1.1 Geometric Projection Model	16
2.1.2 Filtered Backprojection	17
2.2 PET Physics and Instrumentation	18
2.2.1 Radiotracers and Coincidence Detection	18
2.2.2 Detectors and Detection System	20
2.2.3 Background Events: Scattered and Random Coinci- dences	23
2.2.4 Attenuation	25
2.3 Measurement Model and Conventional Image Reconstruction	26
2.3.1 Transmission Scans	27
2.3.2 Emission Scans	31
2.3.3 Conventional Image Reconstruction	32
3 Statistical Image Reconstruction	33
3.1 Maximum Likelihood	34

3.2	Penalized Likelihood	36
3.2.1	Potential Functions	37
3.2.2	Resolution	40
3.2.3	Algorithms	41
4	Paraboloidal Surrogates Coordinate Descent Algorithms for Transmission Tomography	45
4.1	Introduction	45
4.2	The Problem	49
4.3	Paraboloidal Surrogates Algorithms	51
4.3.1	Maximum Curvature	54
4.3.2	Algorithms for Minimizing the Paraboloidal Surrogate	55
4.3.3	Coordinate Descent Applied to the Surrogate Function	55
4.3.4	Convergence and Convergence Rate	60
4.3.5	Optimum Curvature	62
4.3.6	Relationship to Newton-type Optimization Algorithms	63
4.3.7	Precomputed Curvature	64
4.4	Results	66
4.5	Conclusion	73
5	Paraboloidal Surrogates Ordered Subsets Algorithms for Transmission Tomography	80
5.1	Introduction	80
5.2	The Problem	83
5.3	The SPS Algorithm	84
5.3.1	The Likelihood Part	84
5.3.2	The Penalty Part	87
5.3.3	The SPS Algorithm	88
5.3.4	Ordered Subsets	90
5.3.5	Precomputed Denominator OSTR	92
5.4	Phantom Data Results	94
5.4.1	Reconstructions	94
5.4.2	Mean Squared and Segmentation Errors	96
5.5	Patient Data Results	99
5.6	Conclusion	101
6	Image Reconstruction Techniques for PET with Post-injection Transmission Scans	107
6.1	Post-Injection Transmission Scans	107
6.1.1	Rod Sources and Sinogram Windowing	109
6.1.2	The Model	110
6.1.3	Resolution	112
6.1.4	Non-statistical Methods	114
6.2	Sequential Statistical Methods	116
6.2.1	One-step Sequential Estimation	118

6.2.2	Alternating Estimation	119
6.3	Joint Estimation	120
6.3.1	Introduction	120
6.3.2	The Problem	120
6.3.3	The Method	122
6.4	Conventional versus One Step Sequential Methods	123
6.4.1	The Phantom and The Parameter Values	124
6.4.2	Results	124
6.5	Conventional versus Joint Methods	127
7	Noise Analysis and Scan Time Optimization	131
7.1	Introduction	131
7.2	ACF Estimation	133
7.3	Emission Image Reconstruction	134
7.4	Emission Covariance Estimates	134
7.5	Penalized-Likelihood Attenuation Reconstruction	138
7.5.1	Resolution	138
7.5.2	Covariance Approximations	140
7.6	Results	142
7.7	Conclusion	144
8	Conclusions and Future Work	146
8.1	Summary	146
8.2	Conclusions	147
8.3	Future Work	148
	APPENDICES	150
	BIBLIOGRAPHY	161

LIST OF TABLES

Table

4.1	Algorithm outline for a paraboloidal surrogates algorithm with coordinate descent (PSCD). The curvature choice shown here is the maximum second derivative (PS,M,CD).	59
4.2	Comparison of CPU times, number of iterations to converge, floating point operations and memory accesses for the PS algorithms versus CD, GD and FS methods. Convergence in this table means $\Phi(\mu^0) - \Phi(\mu^n) > 0.999 [\Phi(\mu^0) - \Phi(\mu^*)]$ where $\Phi(\mu^*)$ is the smallest objective value obtained in 30 iterations among all the methods. The floating point operations and memory accesses only in the order of M are shown for each method.	72
5.1	OSTR algorithm outline	92

LIST OF FIGURES

Figure	
2.1	Illustration of 2D projections and their organization. 15
2.2	System model for the geometric projections. 17
2.3	An illustration of true, attenuated, scattered and random coincidences. 20
2.4	A transaxial view (upper portion) and a top view (lower portion) of a PET scanner. Showing A) rod sources for attenuation correction, B) the septa used for scatter reduction, C) detector blocks D) photomultiplier tubes and E) End shields. 21
2.5	A block detector consisting of 8x8 array of crystals coupled to four PMTs. 22
2.6	Emission image reconstructed (a) without and (b) with attenuation correction from noiseless data. 30
3.1	Plots of various potential functions $\psi(t)$ 40
4.1	One-dimensional illustration of the optimization transfer principle. Instead of minimizing $\Phi(\mu)$, we minimize the surrogate function $\phi(\mu; \mu^n)$ at the n th iteration. Here, the surrogate function ϕ_2 has a smaller curvature and is wider than ϕ_1 , thus it has a bigger step size and hence faster convergence rate to the local minimum μ^* 52
4.2	Illustration of the tangent parabolas lying above a potential function. 57
4.3	This figure illustrates the optimum curvature and the maximum curvature surrogate functions and their derivatives for $b_i = 100, y_i = 70, r_i = 5$, and $l_i^n = 2.5$ 63
4.4	This figure illustrates the quadratic surrogates with the optimum curvature and the Newton curvature. 65
4.5	Denominators d_j^n versus iterations for different PSCD curvatures. . . 67
4.6	Optimum curvature denominators d_j^{100} versus precomputed denominators. 68
4.7	Newton's curvature denominators d_j^{100} versus precomputed denominators. 69

4.8	(a) FBP reconstruction of phantom data from 7-h transmission scan, (b) FBP reconstruction from 12-min transmission scan, and (c) Penalized-likelihood reconstruction from 12-min transmission scan using 12 iterations of the “optimum curvature” PSCD algorithm.	76
4.9	Comparison of objective function decrease $\Phi(\mu^0) - \Phi(\mu^n)$ versus iteration number n of monotonic PS methods with coordinate descent and LBFGS methods for real phantom data.	77
4.10	Same as Figure 4.9, but x-axis is CPU seconds on a DEC AlphaStation 600 5-333 MHz.	77
4.11	Comparison of the speed of the proposed PS algorithms with the fastest algorithm that was introduced before: grouped coordinate descent with 3x3 groups.	78
4.12	Comparison of objective function decrease $\Phi(\mu^0) - \Phi(\mu^n)$ versus iterations of monotonic PS and FS methods with coordinate descent. Note $r_i = 0$ in this simulation.	78
4.13	Comparison of objective function decrease $\Phi(\mu^0) - \Phi(\mu^n)$ versus CPU time of monotonic PS and FS methods with coordinate descent. Note $r_i = 0$ in this simulation.	79
5.1	Comparison of additive form with optimum curvature (AO), with pre-computed curvature (AP) and multiplicative form with optimum curvature (MO) SPS algorithms for PL image reconstruction.	88
5.2	Maximum Likelihood $L(\mu^0) - L(\mu^n)$	95
5.3	Penalized-likelihood $\Phi(\mu^0) - \Phi(\mu^n)$	96
5.4	Image obtained by hard segmentation of the FBP image reconstructed from the 14-hour scan assumed as the true image.	97
5.5	Normalized mean squared errors versus iterations for various methods of reconstruction.	98
5.6	Segmentation errors versus iterations for various methods.	99
5.7	Reconstructed (left) and segmented (right) attenuation map images using various methods. The numbers in parentheses show the iteration number. The bar plots show the relative NMSE and segmentation error levels. The middle lines in the right hand side bars for ML methods indicate the segmentation error reduction after median filtering.	100
5.8	FBP (denoted E-FBP) and quadratically penalized PL (denoted E-PL) emission image reconstructions with attenuation correction factors obtained using conventional (denoted T-FBP) and nonquadratic penalty PL (denoted T-PL) attenuation map reconstructions using the OSTR algorithm. Transaxial images of slice 36 (of total 47 slices) are presented. Emission scan was 15 minutes and transmission scan was 12 minutes. Emission images were all reconstructed at the same resolution of 6.6 mm FWHM.	103
5.9	Coronal images (slice 64) of the patient data with parameters same as Figure 5.8.	104

5.10	Sagittal images (slice 75) of the patient data with parameters same as Figure 5.8.	104
5.11	Same as Figure 5.9 but with a 2 minute transmission scan which is obtained by thinning the original 12 minute transmission scan. . . .	105
5.12	Coronal images of the patient data with parameters same as Figure 5.11.	106
5.13	Sagittal images of the patient data with parameters same as Figure 5.11.	106
6.1	Transaxial rod geometry in a rotating rod scanner with three rods. . .	110
6.2	Sinogram window illustration for a fixed position of a single rod. This window is changed real time as the rod moves.	111
6.3	The artifacts caused by mismatch of resolutions. Attenuation is reconstructed from noiseless data using quadratic penalty with $\log_2 \beta = 7.6$ resulting in a psf of FWHM 14 mm. The emission image is reconstructed by PL again from noiseless data with ACFs from μ -map using quadratic penalty with $\log_2 \beta = 3.0$ yielding a psf of FWHM 6 mm. .	113
6.4	Bias vs standard deviation trade-offs for attenuation map reconstructions. Labels indicate different reconstruction methods.	125
6.5	Bias vs standard deviation trade-offs for FBP emission reconstructions. Labels indicate the attenuation map that is used.	126
6.6	Bias vs standard deviation trade-offs for quadratic penalty emission reconstructions. Labels indicate the attenuation map that is used. . .	127
6.7	Attenuation Maps obtained through various reconstructions from a single noisy measurement realization.	128
6.8	FBP reconstructions obtained from a single realization.	129
6.9	MPL-Q reconstructions obtained from the same noisy data as the FBP reconstructions.	130
6.10	Images obtained from joint and conventional image reconstruction methods.	130
7.1	Standard deviation of the sum over the heart region estimates versus emission scan time fraction for conventional and statistical ACF computations.	143
7.2	Standard deviation estimates of the sum over the a rectangular region in the heart versus emission scan time fraction for conventional and statistical ACF computations.	144
7.3	Standard deviation estimates of a single pixel value versus emission scan time fraction for conventional and statistical ACF computations.	145
A.1	Plot of $\dot{q}(l; l^c)$ and $\dot{h}(l)$ for maximum curvature.	152
B.1	Plot of $\dot{q}(l; l^c)$ and $\dot{h}(l)$ for optimum curvature.	156

LIST OF APPENDICES

APPENDIX

A	Proof of Maximum Curvature	151
B	Proof of Optimum Curvature	155

CHAPTER 1

Introduction

1.1 Motivation

Computed tomography is the generic name given to the class of non-invasive imaging techniques in which one uses line integral measurements through objects to reconstruct a three dimensional image of a parameter of interest. Measurements are obtained by detecting radiations from radiating sources. Possible sources include electromagnetic radiation ranging from microwaves to high energy X-rays and a wide range of acoustic waves. Computed tomography is mostly used for diagnostic medical imaging. For example in X-ray CT, X-ray transmission sources are used to obtain attenuated measurements and this data is used to reconstruct anatomical images of the attenuation map of the patient.

Emission Computed Tomography is a medical imaging technique that yields functional information about physiological processes as opposed to X-ray CT which provides anatomical images. In emission tomography, radiating sources are injected into the body. Single photon emission computed tomography (SPECT) and positron emission tomography (PET) are two emission tomography modalities. In SPECT, the radioactive nucleus radiates a single photon and these photons are detected by collimated detectors. We focus on PET in this thesis.

In positron emission tomography (PET), a chemical compound labeled with a positron emitting radioisotope is injected into the bloodstream of the patient. The

purpose is to obtain an image of the concentration of that chemical compound which is related to a biological function in the patient body. The radiotracer nucleus emits a positron (positively charged electron) to change from a metastable state to a stable state. The emitted positron annihilates with a nearby electron to form a pair of 511 keV photons propagating in opposite directions. When a pair of photons are detected almost simultaneously outside the subject, it is known that a positron-electron annihilation took place in the line joining two detectors. This simultaneous detection is termed coincidence detection. Each coincidence detection increments a counter that represents the line integral specified by two detectors. After a certain scan time, line integral measurements of the radioisotope density is obtained.

However, all annihilations that result in a photon pair heading towards two detectors are not detected. Often, one or both of the photons get scattered or absorbed by the patient body resulting in no detection. The survival probability of the annihilation event is determined by the length and the type of the tissue that the photons traverse. This effect is called attenuation. Attenuation is different for different tissue types, hence the measurements should be compensated for attenuation for each ray (or line integral). According to Beer's law, the survival probability of a photon going through a uniform medium can be determined by

$$\alpha = e^{-\mu L},$$

where μ is the linear attenuation coefficient of the medium and L is the length. Obtaining a continuous linear attenuation coefficient image (attenuation map for short) is enough to determine the survival probabilities for each line integral (or ray) of PET emission measurements.

A good way to estimate attenuation is to perform a "transmission scan" with an outside radiating source before or after the radiotracer is injected. Ring or rod sources that contain long half life positron emitters are used in PET transmission scans.

The attenuation effects on the final emission image quality are particularly severe in thorax and abdomen imaging where the survival probability of photon pairs can be as low as 2%. There are potential new clinical applications for creating PET

images of the thorax, such as detection of breast cancer and lung tumors. Hence, the improvement of attenuation correction methods is very important.

If the noise level is negligibly small, Radon transform and Fourier slice theorem can be used to reconstruct images from projections. The highly accepted method that is based on the two principles mentioned above is the analytical method called filtered backprojection (FBP). FBP can be used to reconstruct emission images from emission scans and attenuation maps from transmission scans. However, in PET the statistical variability of the measurements is high because of the count limited nature of the detection process. Radioisotope dosage cannot exceed a certain level for safety and PET detector system has count-rate limitations for reliable detection of events. In addition, long scan times may also be impractical especially for whole body and dynamic PET studies. The lower the counts in the data the lower the signal to noise ratio for the measurements and the higher the percent variability. FBP ignores the counting noise present in the data. Especially for low count scans, FBP yields unacceptable images with streak artifacts. FBP treats all the measurements equally even though they all have different noise levels. Transmission and emission scan measurements are well modeled as Poisson random variables. Statistical image reconstruction methods use this information to reconstruct better images. Moreover, the measurements usually have to be compensated for factors like scatter, randoms, detector efficiencies, deadtime and attenuation. These factors themselves need to be estimated and the estimates are mostly noisy. Conventional FBP based methods use simple processing such as subtraction and multiplication to correct for these factors which might result in increased error and noise in the data. Decreasing the cut-off frequency in the filter of FBP will reduce the amount of the noise, but it results in a loss of resolution. Statistical methods include these factors in the measurement model. Some of these factors such as randoms and scatter can also be estimated by statistical means to reduce noise in them [84].

Statistical image reconstruction methods include maximum likelihood (ML) and penalized likelihood (PL) estimation. PL methods can also be cast as Bayesian or maximum *a posteriori* probability (MAP) estimation with Gibbs Markov random field

(MRF) priors. A fundamental problem in statistical image reconstruction algorithms is the excessive computational cost. Since no closed form solutions are available, iterative algorithms are employed. Even though the computational power of microprocessors is increasing, there is still a need of reduction in computation time for these methods to be used in commercially available scanners. Currently, some scanners ship with statistical image reconstruction for emission scans, but attenuation correction is still performed using conventional methods.

The benefits of using statistical methods to reconstruct emission images have been demonstrated clearly in the literature, *e.g.* [85, 47]. Attenuation maps can also be estimated using statistical methods, however using statistical methods for the transmission case has not been studied as extensively as the emission case. We focus on the transmission problem in this thesis.

The PL formulation regularizes the ML problem by including a roughness penalty term in the objective function. The ML problem can be considered a special case of the PL problem (by setting the penalty to zero). Many algorithms for ML and PL reconstruction of emission and attenuation images have been proposed.

In the first part of this thesis, we develop new fast and monotonic algorithms for PL reconstruction of attenuation maps from transmission scans. The statistical model includes background events and nonnegativity constraints can be enforced easily. The algorithms are based on a notion we call “paraboloidal surrogates”. This notion leads to many fast and/or monotonic algorithms which are further explored in chapters 4 and 5.

Conventionally, attenuation correction factors (ACFs) are computed from a transmission scan that precedes the radiotracer injection, thereby increasing total scan time. Reducing the scan time is crucial to increase the patient throughput and to make attenuation corrected whole body PET scans possible. For this purpose, post-injection measurements have been proposed that use rotating rod sources and sino-gram windowing to acquire transmission scans with some emission contamination. Post-injection scans also eliminate the problem of misregistration between emission and transmission reconstructions. Simultaneous emission/transmission (SET) scans

are also possible using sinogram windowing.

In the second part of this thesis, we focus on post-injection transmission scans. Conventional methods of reconstruction used in post-injection scans are suboptimal. Simple subtraction of emission data from the transmission sinogram and FBP reconstruction of emission image ignore the Poisson nature of the measured data. Although these methods give satisfactory results for brain scans [21], they perform much worse in the thorax where attenuation is more nonuniform and its correction is more important. We present new methods of image reconstruction which include the emission contamination in the statistical model. Various image reconstruction methods based on sequential estimation of transmission and emission images as well as joint estimation of both parameters are explored.

We also analyze the noise properties of the reconstructed emission image for post-injection transmission scenario by using approximations to predict the variance of the total uptake in a region of interest in the emission image for conventional and some statistical methods introduced. Previous studies on this problem were based on noise equivalent counts (NEC) criteria which only considered the conventional image reconstruction method. Our approximations separate the effects of emission and transmission scans on the emission image variance. This separation enables us to optimize scan times for emission and transmission scans under a total scan time constraint that would minimize the variance in a region of interest.

The goals of this research can be summarized as follows:

- Develop a framework of algorithms based on a new paraboloidal surrogates idea.
- Using the paraboloidal surrogates develop new fast and monotonic algorithms for PL image reconstruction in transmission tomography.
- Develop new statistical models, techniques and algorithms for the post-injection transmission scenario in PET scans.
- Analyze the noise properties of the reconstructed emission images, particularly to optimize the scan times of transmission and emission scans under a total scan time constraint.

1.2 Background and Significance

PET was introduced in 1970's [93, 109] as a novel tomographic imaging method based on the coincidence detection of photon couples generated by radioactive decay from radiotracers distributed in the human body. Image reconstruction was done using filtered backprojection (FBP) as it was done in X-ray CT. High counting noise in PET caused poor quality and streak artifacts in images reconstructed by FBP.

Initially, PET was used mostly for brain scans and attenuation correction was done by calculating the attenuation correction factors (ACFs) assuming the head as an elliptic region with a constant attenuation coefficient of soft tissue inside [58]. However, this approach is not accurate enough for thorax scans, due to the nonuniform attenuation properties of bone, lungs and soft tissue. Thus, most PET centers have adopted the measured attenuation correction method, in which one precedes the emission scan with a transmission scan that measures the unique attenuation characteristics of each patient for each slice of interest [61].

The conventional method of attenuation correction in PET using measured transmission scans consists of two steps: first compute the attenuation correction factors (ACFs) by smoothing the ratio of the blank scan to the transmission scan data, then multiply the emission measurements by the ACFs for each ray. Therefore, no attenuation map reconstruction is needed. However, it is advantageous to reconstruct attenuation maps and reproject them to get ACFs for improved noise performance¹ [79], for anatomical localization [1] and for reprojection in fully 3D PET studies [116]. Conventionally, to obtain the attenuation map, one first divides the blank scan measurements by transmission measurements to get a noisy estimate of the ACFs. After division, the logarithm of the ratio is taken to obtain line integrals of the attenuation map. Next, one applies the FBP method to the logarithmic data to reconstruct the attenuation map. This approach yields positively biased estimates of the attenuation map when the transmission counts are small [35]. After possibly some processing in

¹Simple reconstruction by FBP would yield the same results as the conventional method [88]. Some processing in the image domain such as segmentation or statistical methods of image reconstruction should be employed for better noise performance.

the image domain (like segmentation), reconstructed attenuation maps are usually reprojected to obtain line integrals and ACFs in the sinogram domain. Other approaches to reconstruct attenuation maps include penalized weighted least squares (PWLS) and penalized likelihood (PL) methods. PWLS reconstructs negatively biased images for low-count scans [35]. PL methods are preferable because they avoid the systematic bias problem and due to the algorithms we present in this thesis, the computational requirements are almost the same as the PWLS method. Now, we review the statistical image reconstruction algorithms for transmission and emission scans.

In early 1980's, Shepp and Vardi [100] and Lange and Carson [70] published seminal papers on image reconstruction in PET using a statistical model for data and maximizing the resulting likelihood using the expectation maximization (EM) algorithm which was introduced in 1977 [26]. Lange and Carson [70] discussed both the transmission and emission cases. The EM algorithm resulted in a closed form solution for the M-step in the emission case. However, for the transmission case, the M-step did not yield a closed form solution. Lange and Carson [70] suggested an approximate update for the M-step.

The algorithms presented in [100, 70] have some limitations: (1) the model does not fully account for the factors such as randoms and scatter, (2) due to ill-conditioned nature of the ML problem, the maximizing images are unacceptably noisy, and (3) convergence of the EM algorithm is very slow.

The algorithm in [70] was based on a simplified statistical model in which only attenuation and detector efficiencies could be included. Later, a randoms term was included in the statistical model [94] and the problem of scatter was addressed in the context of iterative image reconstruction in PET [22, 90]. In this thesis, we include the background terms in our transmission and emission models. These background terms contain randoms, scatter, cosmic radiation and any other extra term that is not directly due to a true annihilation. Nonlinear Bayesian estimation of randoms and scatter terms was discussed in [84, 85]. We focus on the algorithms based on the general measurement model and assume the factors other than attenuation were

estimated or measured appropriately in both the transmission and emission cases.

Maximum likelihood problem in PET is ill-conditioned, *i.e.* small changes in the data result in big changes in the reconstructed images. Thus, the maximizing images are very noisy. The EM algorithm is usually initialized with a uniform image and after a certain number of iterations, the image quality starts to deteriorate. Methods to overcome the noise problem include: stopping rules to stop iterations before convergence [118, 55], regularization using sieves [104], adding smoothing steps in between iterations [101], penalized weighted least squares [33], separable (nonsmoothness) priors [69, 74] and penalized-likelihood (PL) or maximum *a posteriori* probability (MAP) or Bayesian methods with smoothness priors [48, 54]. Studies [9] have shown that the PL methods outperform sieves. In this thesis, we focus on PL image reconstruction, but the algorithms we introduce are also applicable to the ML problem by setting a parameter to zero. The PL formulation also enables one to include other prior or side information (*e.g.* from a registered MRI scan) about the image in the model [115].

The roughness penalty in the PL formulation is chosen in general to encourage local smoothness (discourage roughness) in the image. Quadratic penalties result in a globally blurred image. The edges in the image are also blurred the same amount as the homogenous regions. Nonquadratic penalties may be designed to encourage smoothness within homogenous regions but allow relatively sharp edges. For transmission scans, nonquadratic penalties are preferred because attenuation maps are generally composed of large homogenous regions with sharp boundaries.

The convergence rate of the EM algorithm [70] is not satisfactory for either the transmission or emission cases (it is worse for the transmission case [89]). For the PL formulation, because of the coupling among neighboring pixels introduced by the penalty term, the maximization step (M-step) of the EM algorithm has no closed form solution even in the emission case. There have been many studies to improve the convergence rate as well as modify the EM algorithm to handle the penalty term. The following approaches were used for the emission problem. Green's one step late (OSL) [51, 52] algorithm replaces the gradient of the penalty term with values from the previous iteration to decouple the updates. However, this modification of the EM

algorithm is not guaranteed to converge to the true optimizer. Another approach is called generalized EM (GEM) which replaces the M-step with an updating procedure which guarantees monotonic changes in the objective function at each iteration [54]. If the penalty term is convex, another approach by De Pierro [24] uses convexity to find separable surrogates for the penalty part which decouples the M-step and yields a closed form update. Alternatively, by using a separate hidden data space for each parameter, the SAGE algorithm [44, 45] intrinsically uncouples the parameter updates and uses less informative complete data spaces to improve convergence speed. These algorithms are provably convergent algorithms for emission tomography.

The least informative data space is the measurement space itself. Thus, some researchers worked on optimizing the objective function directly instead of using the EM formalism. This approach yielded faster algorithms than the ones based on EM algorithm. Sauer and Bouman [98, 6] introduced the iterative coordinate descent (CD) method which optimizes the objective function by updating each parameter sequentially. This method also goes by other names such as iterated conditional modes (ICM) for MAP segmentation [2] and Gauss-Seidel (GS) for linear systems [123]. This method easily accommodates nonquadratic penalties and nonnegativity constraints. The optimization at each step is reduced to one dimensional function optimization which is simpler than multidimensional functions. Usually, coordinate descent is implemented with a Newton Raphson update step for each parameter. This implementation is not guaranteed to be monotonic (it is locally convergent for convex objective functions). Coordinate descent optimization is applicable to both emission and transmission tomography. Mumcuoglu [84, 85] uses conjugate gradient algorithms to optimize the objective function directly for both emission and transmission cases. It is harder to impose nonnegativity constraints with that method. For the transmission PET, attempts at developing an EM algorithm have not resulted in attractive closed form updates [70]. Approximate EM algorithms were proposed [7, 89], but the computational cost remains high. Lange derived an algorithm for transmission tomography called the convex algorithm [71] which simultaneously updates all the parameters and is faster than the transmission EM algorithm [70].

Although the coordinate descent algorithm is very fast, it has two drawbacks. The algorithm requires column access to the system matrix and it is not parallelizable. The first problem kept it from being widely used in the medical imaging community because of implementation difficulty. The EM method and simultaneous update algorithms are more popular because they can be implemented using projector/backprojector subroutines without precomputing and column accessing the system matrix.

It is possible to parallelize the coordinate descent by updating a group of pixels at a time instead of a single one [42, 99]. This approach is called grouped coordinate descent (GD) [42, 38]. In the case of transmission tomography, updating groups of pixels also reduces the amount of computation due to many exponentiations. Thus, the GD [42] method is also much faster even in a serial computer as compared to CD. However, the GD method is also more complicated to implement.

Another acceleration of emission EM was achieved by updating the parameters by using a subset of the data at a time [60]. These subsets are ordered and cover all the projection space. The algorithm goes through each subset to finish an iteration. This approach leads to a fast algorithm called ordered subsets EM (OSEM) [60]. However, the OSEM algorithm does not converge and it eventually cycles through points. Byrne introduced a modification of OSEM called RBI-EM [11, 12] which is convergent in the case of unrealistic consistent data. By adding appropriate relaxation parameters to OSEM, Browne and De Pierro [8] achieved a fast and convergent algorithm called RAMLA for the ML problem in the emission case. Manglos [77] applied the ordered subsets idea to transmission EM algorithm for the ML problem. We apply the ordered subsets idea to a new simultaneous update algorithm for PL image reconstruction called separable paraboloidal surrogates (SPS) which is much faster than the transmission EM algorithm.

When the background terms are included in the model, the PL objective function for the transmission problem is not convex, whereas it is convex for the emission case. So, this makes it harder to find a convergent algorithm for the transmission problem. In this thesis, we present algorithms for transmission tomography that are guaranteed

to be monotonic even with nonzero background events.

Earlier transmission scans were performed pre-injection. After the transmission scan is performed, the patient was injected with the radiotracer. After a long time, up to 50 minutes for some studies, the emission scan was performed. The waiting period is required for the radiotracer to get distributed to places where it will be metabolised in the body. Only after this waiting period could one perform the emission scan. There are some problems with this approach, such as unnecessary idle time in the scanner and artifacts due to motion of the patient while waiting. To avoid these problems, one can perform the transmission scan after injection right before or after the emission scan is done. These scans are called *post-injection transmission scans* [14, 21]. The drawback is that post-injection transmission scans are contaminated by emission counts. Conventional method of dealing with emission contamination is either to ignore it or subtract the scaled emission data (acquired separately) from the transmission scan [14]. It was also suggested that emission contamination can be estimated by performing a transmission scan with sinogram windowing but without the rods present [103]. Some researchers used the “other” sinogram window to simultaneously acquire emission data with the transmission data leading to simultaneous transmission-emission scans [114, 78] and some utilized this simultaneous emission scan to estimate emission contamination [57]. The use of transmission rod sources and sinogram windowing of the measurements reduce the emission contamination considerably to a much smaller percentage as compared to the case when ring sources are used. We focus on the post-injection transmission scans and discuss ways of correcting for the emission contamination. Both conventional and various statistical methods are explored.

Noise analysis for conventional image reconstruction in emission tomography relies mostly on noise equivalent counts (NEC) concept [108]. In count limited tomography, variance of the prompt data is equal to the mean. So, the data counts itself is an estimate of the variance. In the conventional method, corrections to the sinogram data often yield changes in the variance of the data. The changed variance can be estimated by NEC easily. Stearns [107] improved this analysis to include the noise

propagation from transmission scans to emission images. This approach was used to obtain scan time optimization for transmission and emission scans [3]. On the other hand, Fessler obtained approximations for the mean and variance of the images reconstructed by penalized likelihood methods [37]. We use these approximations and others to analyze the noise properties of emission and attenuation images where attenuation is reconstructed by conventional or PL methods and emission images are reconstructed by FBP. We evaluate our approximations by comparing them with empirical results.

1.3 Organization of Dissertation

The dissertation is organized as follows. In Chapter 2, the physics of a PET imaging system is described and a realistic model for the measured data is introduced for pre-injection transmission and emission scans. Different factors in the model are explained and estimation of some of the factors is explained briefly. Chapter 3 describes statistical image reconstruction methods such as maximum likelihood (ML) and penalized likelihood (PL). A general form of regularizing penalty is given and resolution properties are explained briefly. Some basic algorithms for ML and PL are introduced in this chapter as well. In Chapter 4, a new fast and monotonic algorithm called paraboloidal surrogates coordinate descent (PSCD) for transmission tomography PL image reconstruction is developed. Results illustrating speed and convergence of new algorithms are presented. Chapter 5 introduces a new simultaneous update algorithm and its ordered subsets version. Ordered subsets is shown to improve the speed of the algorithm. Phantom and real patient data are processed and reconstructed images are presented. Chapter 6 is devoted to post-injection transmission scans and image reconstruction techniques for it. Sequential reconstruction of transmission and emission data with conventional and statistical methods are described. A joint estimation approach is also developed in this chapter. Some preliminary results are shown. Next, Chapter 7 develops new variance prediction formulas for post-injection scans for some image reconstruction methods. Results comparing pre-

dicted and empirical variances are shown. Finally, Chapter 8 contains conclusions and directions for future work related to this research.

1.4 Original Contributions

The original contributions made by this research are summarized in the following.

- A novel image reconstruction method called Paraboloidal Surrogates (PS) is presented. This method gives a framework for designing algorithms for a general class of objective functions. By simplifying the PL image reconstruction problem at each iteration, PS enables faster algorithms to be developed.
- An optimum curvature is developed for the paraboloidal surrogate which would result in the fastest algorithm while preserving monotonicity.
- A new sequential update algorithm based on PS is presented (PSCD).
- A new simultaneous update algorithm based on PS is presented (SPS).
- Ordered subsets acceleration of SPS is presented (OSTR).
- New techniques for sequential reconstruction of attenuation and emission images are presented for a generalized model of post-injection transmission scans.
- Joint estimation of attenuation and emission images from a joint objective function is presented.
- Variance predictions for some sequential post-injection transmission and emission image reconstruction techniques are developed.

CHAPTER 2

Physics of PET and Statistical Modeling

Positron emission tomography (PET) is a diagnostic method that creates 3D tomographic images of the distribution of positron emitting radionuclides in the human body. The radiolabeled compounds used include substrates, ligands, drugs, antibodies, neurotransmitters and other biomolecules that are tracers for specific biological processes. Thus, the resulting PET images can be considered images of these biochemical or physiological processes. These images are often called “functional images”.

PET has been used extensively in neurological research for brain activation studies, and there is a growing interest in the clinic for it, especially for detecting and staging of cancer. The major disadvantage of PET is its high cost and requirement of an on-site cyclotron to produce the radiotracers.

PET image reconstruction is based on tomography. We first describe tomography in general terms and move on to PET measurement system after that.

2.1 Computed Tomography and FBP

The term tomography refers to the process of generating images of an object from its projections. In 2D planar tomography, line integrals along many angles and radial positions are gathered and the original image is reconstructed from the projection data. A line integral (or ray) can be characterized by its angle and radial position. A set of rays at a fixed angle is termed a projection. The process of

transforming a continuous 2D function into its projections (sets of line integrals) is known as the Radon transform [63]. In practice, due to the limitations of the detection systems, the projections are performed for a finite number of angles and at each angle the measurements are “binned” into a finite number of bins with different radial positions. The discretized projections are illustrated in Figure 2.1. The measurements are traditionally organized in a 2D array form where the projection angles form the vertical axis and radial bin positions form the horizontal axis. This “array” is called the sinogram, since a single point source in the original image approximately traces a sine wave in the projection domain.

For noiseless and continuous data, the Fourier slice theorem gives a relation between the Fourier transform of the projections and 2D Fourier transform of the image. This relation leads to a well known analytic image reconstruction method called filtered backprojection (FBP) [63, 76].

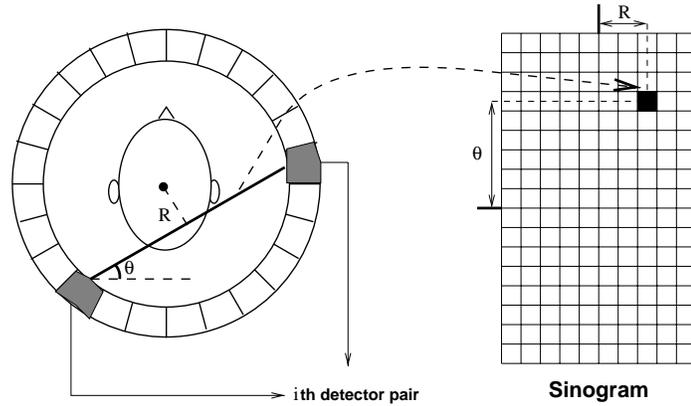


Figure 2.1: Illustration of 2D projections and their organization.

FBP is an image reconstruction method that has long been used in medical tomographic reconstruction, such as in X-ray CT [76]. FBP is a deterministic tool which ignores the counting noise in PET scan data. However, it has found a place in PET applications due to historical reasons and the speed of calculation, despite the suboptimal reconstructed image quality.

Although not needed to perform FBP, we first describe the forward geometric projection model used in our image reconstructions.

2.1.1 Geometric Projection Model

The geometric projections are binned into a set of bins with N_θ angles and N_r radial positions. Assume the continuous 3D image is represented by $\lambda(\underline{x}) : \mathbb{R}^3 \rightarrow \mathbb{R}$ and the projections at angle θ_m and radial position r_n can be written in terms of the original image as follows:

$$p_i \triangleq p_{\theta_m}(r_n) = \int f_i(\underline{x})\lambda(\underline{x})d\underline{x}, \quad m = 1 \dots N_\theta, \quad n = 1 \dots N_r, \quad i = m * N_r + n.$$

Here i indexes all the projections at every angle and radial position. In this equation $f_i(\underline{x})$ denotes the unitless geometric point response function of the i th ray.

Since a scanner can collect only a finite number of measurements, for reconstruction purposes, we must also represent the radiotracer distribution $\lambda(\underline{x})$ by a finite parameterization, say in terms of a set of basis functions:

$$\lambda(\underline{x}) = \sum_{j=1}^p \lambda_j b_j(\underline{x}),$$

where $\lambda = [\lambda_1, \dots, \lambda_p]'$ is the vector of unknown parameters that must be computed from measurements. Typically b_j is the indicator function for the j th voxel and we refer to λ_j as the j th voxel value hereafter.

Then, the relation between p and λ can be given by:

$$\begin{aligned} p_i(\lambda) &= [\mathbf{G}\lambda]_i \triangleq \sum_{j=1}^p g_{ij}\lambda_j, \quad \text{and} \\ g_{ij} &= \int f_i(\underline{x})b_j(\underline{x})d\underline{x}, \end{aligned} \tag{2.1}$$

where, g_{ij} relate voxel values to projections. In emission tomography, g_{ij} is the geometric probability that an emission from inside voxel j gets detected by detector pair i . In a single slice reconstruction (2-D problem), g_{ij} can be assumed to be proportional to the intersection of the strip characterizing the i th projection and the j th pixel. This is illustrated in Figure 2.2.

In PET, more realistic and complex geometric system models might be introduced considering these factors: 1) positron range of the radiotracer, 2) The variation of probability of detection across the strip, 3) The strip width reduction as the rays get far from the center.

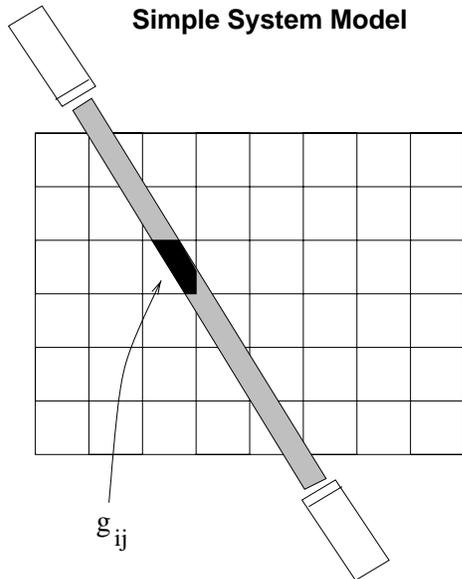


Figure 2.2: System model for the geometric projections.

We use the same system model (matrix g_{ij}) for emission and transmission PET reconstructions in this thesis. However, we must note that the nature of the projections are different in these two scans and that it might be necessary to use different geometric models for each one. For the sake of simplicity, we also use the same notation g_{ij} to present the theory although the theory is general and different geometric system matrices can be used for each scan. Before performing an FBP on an emission or transmission scan, there are some corrections on the data that should be performed to obtain the pure geometric projections first. Notably, for the transmission scans, the logarithm of the data should be taken since the attenuation map is related to the measured data with an exponentiation. We will discuss the PET measurement models in detail in this chapter.

2.1.2 Filtered Backprojection

FBP is performed in discrete space as follows. Let, $p_{\theta_m}[k]$ denote the projections at angle θ_m and the integers $k = -N_r/2 \dots N_r/2 - 1$. Each $p_{\theta_m}[k]$ is filtered by a ramp filter and then the resulting sinogram data is backprojected to reconstruct the image. The latter operation called backprojection is performed by smearing the

projection data back into the image domain. The following formula represents the filtered backprojection operation:

$$\begin{aligned} \text{Filtering step: } \tilde{p}_{\theta_m}[k] &= p_{\theta_m}[k] * h[k], \quad h[k] = \mathcal{F}^{-1}(|\omega|M(\omega)) \\ \text{Backprojection step: } \hat{\lambda}_j &= \frac{\pi}{N_\theta} \sum_{m=0}^{N_\theta} \tilde{p}_{\theta_m}[r_j \cos(\phi_j - \theta_m)/\Delta r] \end{aligned}$$

where $r_j = \sqrt{x_j^2 + y_j^2}$ and $\phi_j = \tan^{-1}(y_j/x_j)$, where (x_j, y_j) are the center coordinates of the j th pixel. If the value $r_j \cos(\phi_j - \theta_m)$ is not an integer multiple of the ray spacing (Δr) , the values of $\tilde{p}_{\theta_m}[k]$ are linearly interpolated.

Using the ramp filter directly is impractical, because being a highpass filter, it accentuates noise. Usually, the filter is cut-off at half-Nyquist rate and windowed with a smooth window $M(\omega)$ to reduce noise and get a reasonable reconstruction. However as a tradeoff, reducing noise by filtering degrades the resolution of the reconstructed image.

2.2 PET Physics and Instrumentation

2.2.1 Radiotracers and Coincidence Detection

Chemically traceable radionuclides exist that are biologically relevant substances for the body. These radionuclides can be produced in a small hospital based cyclotron and rapidly coupled to biomolecules. Examples of these radiotracers include C-11, N-13, O-15, F-18 with 20, 12, 2 and 110 minutes of half lives respectively. The biological molecules synthesized with these radiotracers include fluorodeoxyglucose (FDG), NH_3 , CO_2 , H_2O , etc. These radionuclides emit positrons to return to the stable state. The emitted positron annihilates with an electron after traveling a distance that depends on the radionuclide [19]. The distance traveled is called the *positron range* which depends on the energy of the positron. The range is 0.22 mm FWHM for F-18 and 0.28 mm FWHM for C-11 in water [105]. This event produces two 511 keV gamma photons which travel at an angle of about 180° to preserve momentum and energy.

If both these photons are detected by two detectors, the sinogram bin corresponding to the line of response (LOR) associated with the detector pair is incremented. The detectors should detect the two events in a coincidence window of a short duration τ (around 12 ns) to make sure that two photons originate from the same annihilation. Thus, the detector system in PET is named coincidence detection system [110]. A few PET systems are also capable of roughly measuring the arrival time difference during coincidence detection which provides specific information about the position of the annihilation along the line of response (time-of-flight PET). However, most PET systems only provide the information that the annihilation occurred somewhere along the line. Measurements of such coincidences therefore, correspond to approximations to parallel line integrals of the radiotracer distribution. However, due to scattering and absorption of the photons, not all true annihilations can be detected. This effect is called *attenuation*. Additionally, there are some undesired coincidences which deceive the system to record an annihilation at a ray when actually there were no annihilation along that line of detection. These events are scattered and random coincidences. When traveling through a medium, a photon could be scattered by an electron and change direction. Although a scattered photon loses some energy, it can still be detected by PET detection system. When a scattered photon is detected along with its coincident photon, that line of coincidence does not represent a true coincidence. Such coincidences are called *scatter*. Similarly, if two annihilations happen within the same time window and each one loses one photon due to the scatter of photon out of plane, then the resulting coincidence is called a *random* or accidental coincidence and is again undesired. Unless the data are compensated for such undesired events or the model is designed to include these effects, the resulting reconstruction would not be quantitatively correct. These effects are illustrated in Figure 2.3.

The explanation above is given in a simple way. Actually, PET systems are much more complex. We will describe some details in the following sections.

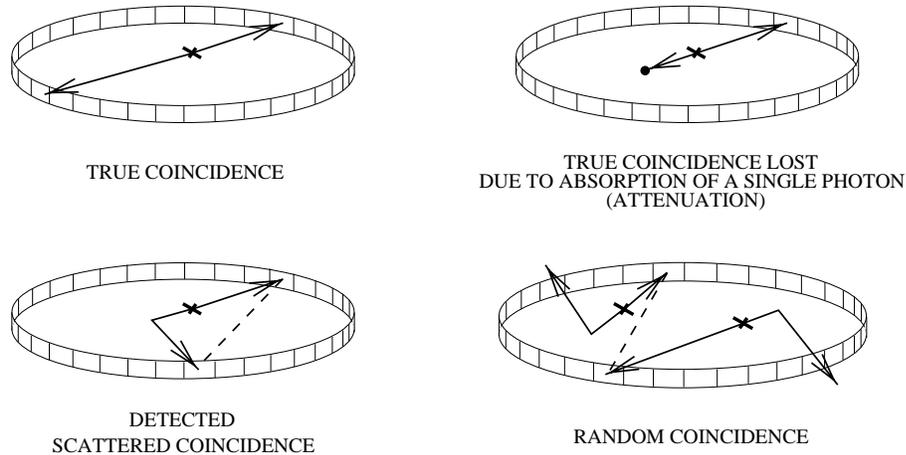


Figure 2.3: An illustration of true, attenuated, scattered and random coincidences.

2.2.2 Detectors and Detection System

Although conceptually PET system detectors can be viewed as a single ring, actual PET systems utilize cylindrically arranged blocks of detectors to achieve volumetric measurements. In current PET systems, detectors are arranged in 2D modular arrays of detectors because the modular approach has distinct advantages over discrete detectors in terms of cost, coincidence timing, energy resolution, sensitivity and axial resolution [83]. The detectors of most current PET cameras are made from blocks of cube shaped large crystals of BGO (bismuth-germanate) segmented into an array of elements. The BGO crystals are scintillators that stop the incoming photon and produce light in proportion to the energy of the incoming photon. This array is coupled to a group (usually 4) of photo multiplier tubes (PMTs) that amplify this light to a detectable level whose output channels are used in Anger-type positioning logic for detector element identification [91]. This geometry is shown in Figures 2.4 and 2.5. For cost reasons, thick sodium iodide (NaI) detectors are also used in some current systems, although they do not stop photons at 511 keV as well as BGO crystals.

When a photon is detected, the analog-digital circuitry produces a signal representing the arrival time, energy and position within the block. If the energy of the incoming photon is within the energy window, then a binning address is formed and the relevant data is sent to the coincidence processing unit. If two valid photons are

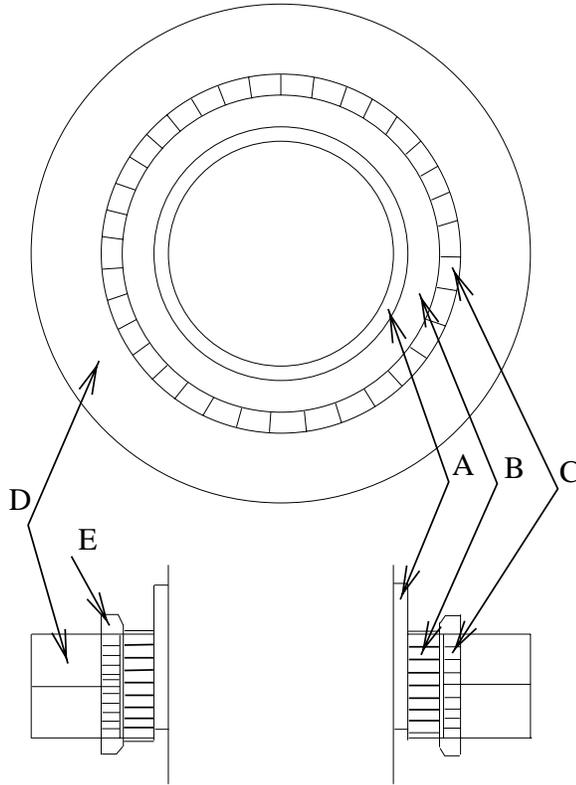


Figure 2.4: A transaxial view (upper portion) and a top view (lower portion) of a PET scanner. Showing A) rod sources for attenuation correction, B) the septa used for scatter reduction, C) detector blocks D) photomultiplier tubes and E) End shields.

detected between any two detector blocks within the time window (12 ns), then this event is counted as a coincidence for the corresponding detector pair. A certain duration of time (about $2 \mu s$) is required for the detector block and the circuitry to get ready for the next photon due to pulse integration, position calculation and energy discrimination [49]. During this period, the detector block is dead for the incoming photons. This is called the *deadtime* problem. Typically, deadtime is a function of the single photon rate at the detector blocks or *singles* rate. Deadtime correction factors have to be known for accurate reconstruction. Deadtime is especially a problem when the count rates are high in which case many true coincidences go undetected. Therefore, deadtime is the dominant effect that limits the injection dose [91].

Once a pair of photons reaches the detectors, correct detection probability is determined by the geometric and intrinsic detector efficiencies, deadtime of the detector

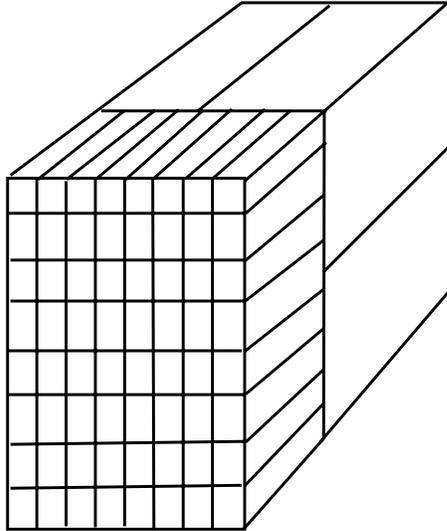


Figure 2.5: A block detector consisting of 8x8 array of crystals coupled to four PMTs.

block, intercrystal scatter and crystal penetration [85]. Efficiency of a detector depends on its position within the block. For example, detector efficiency is lower for a crystal that is located in the edge of a block, because the probability that a photon will be scattered out of the block and go undetected is higher for edge crystals [91]. This is called intrinsic detector efficiency. The stopping powers of detectors also decrease with the obliqueness of the incoming photon angle due to the decrease in the effective depth of the crystal. This is called the geometric detector efficiency. Scattering inside a crystal (intercrystal scatter) and penetration of a photon through a detector to another neighboring one (crystal penetration) are also factors affecting the detection. We ignore these last two effects, but they are important in high resolution systems. Detector efficiencies are determined by a long calibration measurement of a plane source. These efficiencies are included in the emission measurement model. For the transmission scans, a “blank” transmission scan serves as a calibration measurement. The efficiency values for scattered, accidental and true coincidences are different because scattered events have lower energies, and random and scattered events arrive at a wide range of angles as opposed to nearly the same angle for true events for a given detector pair.

It is possible to acquire volumetric data between detector pairs across planes with

a cylindrical set of detectors. This is called 3D PET. In 3D PET, efficiency increases significantly, but scatter, randoms and deadtime rates also increase, thus the image reconstruction is harder [116]. Therefore, in PET systems, there is parallel mechanical shielding (septa) between slices in front of the detectors. Septa absorb most of the photons travelling in the non-transaxial directions and thus only coincidence data in transaxial directions are acquired. This is called direct plane data. To increase efficiency and SNR, ± 1 plane difference coincidence data can also be acquired by adjusting septa accordingly and considered as parallel data set. This is called cross plane data.

2.2.3 Background Events: Scattered and Random Coincidences

Scatter

Compton scatter is the main cause of photon attenuation at 511 keV in biological tissues [111]. The energy discrimination cannot reduce the scatter component to zero since most of the scattered photons do not lose a lot of energy. Using longer septa might also reduce scatter but will also decrease sensitivity [111]. In case of septaless 3D PET, scatter is a big problem since many off-plane scatters will also be detected by the cylindrically placed detectors. The contribution of scattered data in PET measurements can be as high as 12-20 % with septa. The effect of the scatter term, if ignored, is to under and over estimate attenuation and emission parameters in transmission and emission scans respectively. This bias is not necessarily uniform, so the scatter terms should be estimated and included in the statistical model.

Scatter estimation is an important topic by itself. Many researchers have worked on scatter estimation. The simplest approach is to extrapolate projections outside the body. Convolution subtraction smooths the data with a spatially varying kernel to obtain the scatter component [15]. Dual energy window idea was also used to estimate scatter. Most recent scatter estimation techniques use the Klein-Nishina formula for estimating scatter from an initial reconstruction of the attenuation map

[90, 84, 85].

Randoms

Detections of single photons by detectors are called *singles*. If two singles from two different annihilations are detected in the same coincidence window of duration τ seconds (about 12 ns) with their “partners” either absorbed or scattered out of the plane, this event is called an accidental coincidence (or random event). These randoms or background events are also a part of the measurements and should be corrected for.

There is a relation between the singles rate and the randoms which is given by:

$$R_i = 2\tau S_{i_1} S_{i_2}, \quad (2.2)$$

where R_i is the randoms rate in the i th LOR, S_{i_1} and S_{i_2} are the singles rates in each detector involved in the i th LOR, and τ is the coincidence window duration. A random event is caused by two unrelated singles striking two detectors in the same time interval.

The randoms component is a function of the unknown image in general. Typically, the data are precorrected for randoms by subtracting a randoms sinogram from the transmission or emission sinogram. The randoms sinogram is collected by detecting “delayed” coincidences using a delayed coincidence circuit with the same resolving time as the coincidence sinogram. That is, if two single events detected by two elements of the i th detector pair are separated by a certain time delay (for example 2τ), they increment the random coincidence sinogram. Delayed coincidences can only be due to randoms, not true annihilations or scattered events, because of the delay between the arrival times of two photons. Single event occurrences are uniformly distributed over time in a PET scan. Therefore, the number of delayed coincidences for a detector pair is an estimate of the mean number of coincident random events in that pair. We assume the delayed coincidence or randoms sinogram is a Poisson random variable because of the counting nature of detection (this is approximately correct) [124].

In most scanners, the subtraction of the randoms sinogram is performed automatically during data acquisition. This correction compensates the data in its mean for the effect of the randoms. However, the randoms corrected sinogram is the difference of two independent Poisson random variables whose variance is equal to the sum of their means. In other words, randoms subtraction compensates in mean for randoms, but the variance of the measurement is also increased by an amount equal to the mean of the randoms. Moreover, the measurements are no longer Poisson. For these measurements, Yavuz and Fessler [121, 122] have developed a shifted Poisson model which approximates the probability distribution of subtracted data better than Gaussian models [33]. However, to avoid this problem, we must maintain the true measurements and randoms measurements as two separate sinograms.

The randoms component within the model can be estimated or approximated by many methods. Joint estimation of the image and the randoms sinogram was proposed [72]. Mumcuoglu used ML and MAP techniques to estimate randoms from the delayed sinogram using the singles model (2.2) [84]. A simple technique is to estimate randoms from the delayed sinogram by smoothing it since randoms are known to be very slowly varying factors. We could even assume a constant uniform randoms contribution as described in [94] because of the same reason. For the transmission case, the randoms in the blank scan and the transmission scan were shown to be very close [121], so the high count blank scan randoms can be used to estimate the mean of randoms in the transmission scan.

We mostly assume randoms are known fixed values in our simulations and reconstructions. Sometimes we use the blank scan randoms in the transmission model.

2.2.4 Attenuation

At 511 keV, the photons suffer from two different interactions: absorption and Compton scatter. Although absorption is negligible at this energy, scattering causes the photon not to arrive to its original destination [91]. Most scattered photons are not detected at all. Thus, these interactions affect the detection probability of

annihilations and their effect is termed *attenuation*. This is one of the biggest factors that has to be corrected for in PET emission scans. Both photons must be detected to have a coincidence event. The probability that a photon pair will survive along a ray is termed *survival probability*. This can be found from:

$$\alpha_i(\mu) = e^{-\int f_i(\underline{x})\mu(\underline{x})d\underline{x}},$$

where $\mu(\underline{x})$ is the linear attenuation coefficient (attenuation map value) of the tissue at position \underline{x} , and $f_i(\underline{x})$ is the geometric factor that represents the contribution of the attenuation map at position \underline{x} to the total attenuation for the i th ray. We parametrize the attenuation image similar to (2.1) and we get:

$$\alpha_i(\mu) = e^{-\sum_{j=1}^p g_{ij}\mu_j}. \quad (2.3)$$

The attenuation correction factors (ACFs) are defined as $\gamma_i(\mu) = 1/\alpha_i(\mu)$. They are the multiplicative factors to correct for the effects of attenuation in the emission scan data. The survival probabilities are independent from the location of annihilation along the line, so we can correct for the attenuation by processing the sinogram data. Transmission scans are used for attenuation correction in PET which we discuss in detail in section 2.3.1.

2.3 Measurement Model and Conventional Image Reconstruction

The measured data in both transmission and emission PET are due to annihilations that are a result of radioactive decay. The number of decays (thus annihilations) that occurs inside a voxel follows a Poisson probability with the mean proportional to the radiotracer concentration within the voxel. Then, the number of photon-pair detections from voxel j by the i th detector pair follows a Poisson distribution as well. We assume the direction of an emitted photon-pair is uniformly and randomly distributed over 3D space, thus all measurements are assumed to be independent Poisson random variables in PET scans. The above statement ignores the deadtime and

background events. Because of how deadtime affects the PET system and because of the measurement of randoms, the Poisson distribution assumption might not be exactly valid. However, even with deadtime and randoms, the measurements follow a probability distribution function (pdf) which can be very closely approximated by a Poisson pdf [124]. So, for all practical purposes, we assume the measurements in transmission and emission PET are Poisson. Let y_i be the random variable representing ray i and \bar{y}_i be its mean as a function of the underlying emission or transmission image, then the joint conditional density of the random vector $y = [y_1, \dots, y_N]$ can be given by:

$$P(y) = \prod_i e^{-\bar{y}_i} \frac{\bar{y}_i^{y_i}}{y_i!}.$$

Next, we describe the models for the mean of the data \bar{y} for transmission and emission scans.

2.3.1 Transmission Scans

Early “calculated” methods of attenuation correction employed simple geometric approximations to attenuating medium (*e.g.* a uniform ellipse), which is inaccurate for thorax scans because of attenuation coefficient differences between air, lung, soft tissue and bone.

Currently, “measured” attenuation correction methods provide better estimates of ACFs. Measuring the attenuation requires a separate scan called *transmission scan*. Transmission PET data is collected using an external radiotracer source. Usually, transmission sources contain Ge-68, a positron emitting radioisotope which has a half life of 270 days. A fixed ring source was used in earlier scanners [56]. Rotating rod sources [112] were also proposed. Current scanners with transmission rod sources use one to three rotating rods in the system. For the rotating rod sources, only the sinogram elements that are collinear with the rods are considered at a given time and the rest of the measurements are discarded. As the rods move, the sinogram window utilized is also changed. This is called sinogram windowing [62]. The major advantages of the rod sources are that the scatter and random component in the transmission

data is greatly reduced. If the transmission scan is done post-injection, the emission contamination is also reduced greatly. Besides, it is also possible to utilize the (normally) discarded window data to acquire simultaneous transmission emission (SET) scans [114, 78]. Recently, some scanners employ singles transmission measurements where the transmission scans are done in singles mode by using higher energy Cs-137 single photon emitting sources [65, 102]. Although these scans can achieve very high counts in a very short time, they have a much higher scatter percentage as compared to coincidence transmission scans. The model we describe for the transmission measurements is valid for ring sources, rod sources and singles sources for PET as well as transmission SPECT and X-ray CT.

To obtain attenuation correction factors (ACFs) or to reconstruct an attenuation map, it is necessary to collect both blank and transmission sinograms. For calibration purposes, a long blank transmission scan should be done when the patient is not in the scanner. This *blank scan* reflects the activity of the source and also includes effects of the detector efficiencies and the attenuation due to rods. Blank scans can be made virtually noise free because relatively long scanning times can be used (*e.g.* they can be done every night). On the other hand, transmission scan data is collected in a limited time with the patient in position before or after the radiotracer injection. So, typically transmission scans are much shorter than the blank scan. In the clinic, there is also a tendency to perform the transmission scans shorter than the emission scans since the emission scan is seen as the important one and the transmission scan is seen as a method to estimate nuisance parameters. This makes the transmission scans much noisier and statistical methods outperform the conventional methods with a wider margin when the scan counts are low. We analyze the post-injection case in Chapter 6 and focus on pre-injection transmission in this section.

The ratio of the blank sinogram to the transmission sinogram roughly gives the ACFs if we ignore background counts. However, correction factors based on the ratio of the blank and transmission sinograms provide noisy estimates of the true attenuation factors. Linear smoothing of the transmission data reduces noise but introduces bias and does not provide sufficient accuracy [79]. Reconstruction of the

attenuation map, followed by reprojection, can improve the accuracy of the estimated correction factors if some additional information about the attenuation image is used. Simply using FBP followed by reprojection of the attenuation map is essentially equivalent to direct linear filtering of the attenuation correction factors and should not be expected to yield further improvements. There are some other advantages of reconstructing the attenuation map. For example, if the patient moves in between the transmission and emission scans, two images can be re-registered before reprojection of the attenuation map. Attenuation maps can also be used to estimate scatter in emission and transmission scans [84]. Finally, attenuation maps provide anatomical boundaries that are often not visible in emission images. This information can be used when interpreting emission images or for registration of the PET images with other anatomical modalities such as MR or CT [92, 82].

Some researchers performed segmentation [120, 80] of attenuation maps before reprojecting. Segmentation misclassification might cause errors in the ACFs resulting in artifacts in emission images. From our perspective, penalized-likelihood methods of image reconstruction based on statistical modeling and prior information about the image (that it is locally smooth) are more appealing due to effective handling of the statistics and natural incorporation of penalty functions in the objective function. The main focus of chapters 4 and 5 is to develop fast and monotonic algorithms for penalized-likelihood transmission image reconstruction.

We model the mean of the transmission scan data as the following:

$$\bar{y}_i^T = \tau^T d_i^T [b_i \alpha_i(\mu) + r_i^T]. \quad (2.4)$$

The superscript T indicates that the terms belong to the transmission scan. Here, $\alpha_i(\mu)$ are the survival probabilities and are given by (2.3), b_i are the blank scan rates (time adjusted and deadtime corrected), r_i^T represent the background event rates such as randoms and scatter, and τ^T and d_i^T represent the scan time and live time fraction (1 - deadtime fraction) respectively. The blank scan rates b_i can be obtained from the blank scan by:

$$b_i = \frac{\tau^T}{\tau^B d_i^B} B_i,$$

where B_i are the blank scan counts, τ^B is the blank scan time and d_i^B are live time fractions for each ray for the blank scan. Since B_i 's are acquired for a very long time, we assume that b_i correctly represents mean blank scan rates, even though there is Poisson counting noise present in the blank scan.

In conventional attenuation correction strategy, we obtain an estimate of the survival probabilities by the following formula:

$$\hat{\alpha}_i = \text{smooth} \left\{ \frac{y_i^T / (\tau^T d_i^T) - r_i^T}{b_i} \right\}. \quad (2.5)$$

Smoothing is done to reduce the level of noise. It is reported to be better to smooth in the radial direction of the sinogram only [27].

Some whole body scans are performed without a transmission scan due to emission contamination and time constraints. However, for quantitatively accurate emission images, attenuation correction is a must. Especially in the thorax due to nonuniform attenuation, transmission scans are more important.

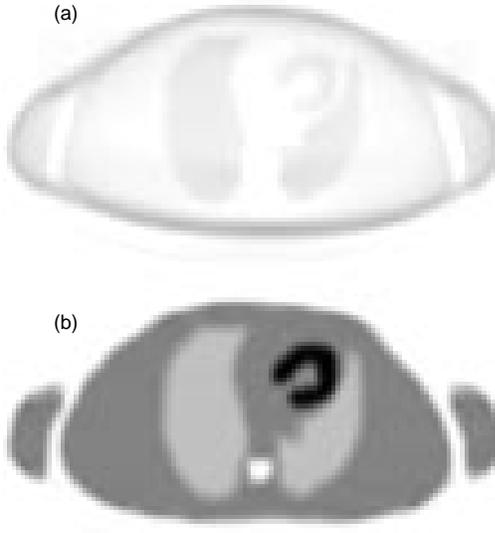


Figure 2.6: Emission image reconstructed (a) without and (b) with attenuation correction from noiseless data.

In Fig. 2.6, we can see two images reconstructed with and without attenuation correction from noiseless data. In the uncorrected image, the radiotracer distribution

is quantitatively incorrect. The structures inside the body appear to have lower radiotracer density than the outside and the lungs appear to have higher dose as compared to attenuation corrected one.

Attenuation correction can be done by dividing the i th emission measurement by $\hat{\alpha}_i$. Next, we describe the emission scan model and conventional attenuation correction.

2.3.2 Emission Scans

The emission scan model is similar to transmission scan one but the measurements are related to the line integrals of the radio-tracer activity in the scanner in an affine way. More precisely, in an emission scan, we model the mean of the recorded events y_i^E as:

$$\begin{aligned}\bar{y}_i^E &= \tau^E d_i^E [\varepsilon_i \alpha_i(\mu) p_i(\lambda) + r_i^E], \\ p_i(\lambda) &= \sum_{j=1}^p g_{ij} \lambda_j,\end{aligned}$$

where τ^E is the scan time, $\alpha_i(\mu)$ are the survival probabilities as defined in (2.3), $p_i(\lambda)$ are the average number of coincident annihilations per second that can potentially¹ be detected by the i th detector pair, r_i^E are the rate of accidental coincidences (randoms) and scattered events. ε_i denote the detector efficiencies for true events for the i th detector pair. Finally, d_i^E denotes the probability of an event not being lost due to detector deadtime (live time fraction).

To reconstruct the emission image, first we need to estimate the factors other than $p_i(\lambda)$ in equation (2.6) if they are unknown, such as survival probabilities, detector efficiencies, and background event rates. We *assume* the background events, detector efficiencies and live time fractions are known exactly or approximated reliably. Thus, only the survival probabilities need to be estimated. We focus on it in the next section.

¹if there had been no scatter, attenuation etc.

2.3.3 Conventional Image Reconstruction

Conventional attenuation correction is performed by dividing the emission measurements by survival probabilities. The survival probabilities are smoothed to reduce noise. Emission scans are also smoothed before attenuation correction. If the emission measurements are directly divided to the smoothed survival probabilities or the emission data is smoothed by a different filter, the reconstructed emission images have some undesired artifacts, as we elaborate more in Chapter 6. Therefore, the emission scan measurements need to be smoothed with the same filter as the survival probabilities [16]. Thus, we get the following formulae for attenuation correction and FBP reconstruction of emission image:

$$\begin{aligned}\hat{p}_i &= \frac{\text{smooth} \left\{ \frac{1}{\varepsilon_i} \left(\frac{y_i^E}{(\tau^E d_i^E)} - r_i^E \right) \right\}}{\hat{\alpha}_i}, \\ \hat{\lambda} &= \text{FBPramp} \{ \hat{p} \},\end{aligned}$$

where $\hat{\alpha}_i$ was found in (2.5). Note that in this case, the Nyquist-windowed ramp filter is used directly in FBP, because the smoothing is done before filtering.

We explore statistical and hybrid methods for attenuation correction and emission image reconstruction in Chapter 6 in detail.

CHAPTER 3

Statistical Image Reconstruction

In current clinical practice, tomographic image reconstruction is mostly based on Radon transform theory, Fourier slice theorem and filtered backprojection. These analytical methods perform very well on low noise data and extensive research was performed on developing better filters for filtered backprojection to deal with noise. However, if the noise is structured like in the case of count limited tomography, methods based on statistical modeling and iterative optimization of statistically derived objective functions yield much better images in terms of quality and quantitative correctness. Iterative methods were first used for X-ray CT to solve deterministic linear equations. Arithmetic reconstruction technique (ART) was probably the first iterative algorithm to be used. Weighted least squares was first suggested for 3D SPECT [50]. Maximum likelihood reconstruction based on Poisson likelihood for emission tomography was first proposed by Rockmore and Macovski [95]. The same authors proposed the ML technique for transmission tomography one year later [96]. EM algorithm for ML estimation was introduced in 1977 [26]. Shepp and Vardi applied the EM algorithm for ML image reconstruction in emission tomography [100]. Lange and Carson [70] improved on their work by also applying the technique to the transmission case. We present the maximum likelihood and penalized-likelihood image reconstruction formulation in this chapter.

3.1 Maximum Likelihood

PET measurements are based on a counting process. Thus, it is reasonable to assume that the measurements are independently distributed Poisson random variables. Let y be the measurements from an emission or transmission scan. For both scans, the measurements can be well modeled as Poisson random vectors with mean \bar{y} as a function of the underlying image to be estimated.

$$y_i \sim \text{Poisson} \{ \bar{y}_i(\theta) \}, \quad i = 1 \dots N, \quad (3.1)$$

where N is the number of detector pairs (or LORs), $\bar{y}_i(\theta)$ is the mean value of the i th measurement in terms of the emission or attenuation parameters θ . Then, the log-likelihood function for estimating θ from y is:

$$L(\theta; y) = \sum_i (y_i \log \bar{y}_i(\theta) - \bar{y}_i(\theta) - \log y_i!).$$

We model the measurement means for the transmission case as follows:

$$\bar{y}_i^T(\mu) = b_i e^{-[\mathbf{G}\mu]_i} + r_i^T. \quad (3.2)$$

This model is different than the one in (2.4) because scan time and deadtime correction terms are included in b_i and r_i^T for simplicity. Here b_i are the time-adjusted and deadtime corrected total blank scan counts, r_i^T are the total background counts and $[\mathbf{G}\mu]_i \triangleq \sum_{j=1}^p g_{ij} \mu_j$, where g_{ij} represent the geometric projection factors which should represent the contribution of the average attenuation coefficient at pixel j to the total attenuation effecting the measurement i .

On the other hand, we model the measurement means for the emission case by the following affine model:

$$\bar{y}_i^E(\lambda) = \sum_{j=1}^p a_{ij} \lambda_j + r_i^E. \quad (3.3)$$

Here λ_j denotes the count rate in the j th pixel, $a_{ij} = c_i g_{ij}$ is the contribution of the j th pixel counts to the i th measurement mean. g_{ij} is the geometric probability that an emission from pixel j is detected by the detector pair i in ideal conditions. For

notational simplicity here, we assume that g_{ij} terms in (3.3) and (3.2) are proportional and the proportionality constant can be included in the c_i factors which contain calibration factors for scan time, detector efficiencies, attenuation factors and possibly deadtime correction factors as well. r_i^E represents the total detected random and scattered event counts. Note that these representations are simpler than the one we used in Chapter 2 since some factors are incorporated into the b_i , r_i^T , a_{ij} and r_i^E terms. This is done for notational convenience.

The log-likelihood is concave for the emission case and has a nonnegative unique maximizer at which the gradient of the log-likelihood is zero. However for the transmission case, when $r_i^T \neq 0$, the log-likelihood is not concave and theoretically there might be multiple local maxima or a continuous region of maxima. In both cases, there is no closed form expression for the maximizer image. Instead, one resorts to iterative methods for computing them. The EM algorithm [70, 100] computes the ML estimate in an easy way. For the emission problem, each iteration of the EM algorithm is given by:

$$\lambda_j^{n+1} = \lambda_j^n \frac{1}{\sum_i a_{ij}} \sum_i a_{ij} \frac{y_i^E}{\sum_{j'} a_{ij'} \lambda_{j'}^n + r_i}. \quad (3.4)$$

The iterations as above converge to the true maximizer of the likelihood if initialized with a positive image. The convergence is very slow, typically requiring hundreds of iterations. Each iteration monotonically increases the log-likelihood. The updates are done simultaneously rather than sequentially.

For the transmission case, there is no closed form solution for the maximization step (M-step) of the EM algorithm. Some approximate solutions can be applied at each iteration [70]. Even with the approximations, the EM algorithm for transmission scans is also very slow and requires hundreds of iterations as well [89].

Usually the maximizer of the log-likelihood is extremely noisy due to the ill-posedness of the problem. Tomographic reconstruction is ill-posed meaning small changes in data results in big changes in the reconstruction. Some form of regularization is required. Methods to regularize the ML problem are: stopping the iterations before convergence [118], post-smoothing the image [104], and adding a roughness

penalty term to the log-likelihood [43]. The final methods are called penalized likelihood (PL) methods. They can also be viewed as maximum *a posteriori* probability (MAP) estimation with a Gauss-Markov (or Gibbs) prior [53] where the log-prior corresponds to the penalty function.

3.2 Penalized Likelihood

PL methods have some distinct advantages over other methods: 1) Adding a penalty term regularizes the problem and causes iterative algorithms to converge faster; 2) Local smoothness penalties add negligible computational cost to the per iteration costs; 3) PL methods give the opportunity to use spatially variant penalties that reflect prior anatomical boundary information; 4) These methods has been shown to outperform other methods [9] in image quality. Thus, we focus on PL methods in this thesis.

The objective function for the PL problem for estimating the parameters θ from the measurements y is:

$$\Phi(\theta) = -L(\theta; y) + \beta R(\theta).$$

Here, the penalty term is $R(\theta)$ and β is the parameter that controls the level of smoothing. We seek to minimize this objective function in the nonnegative orthant since the images we would like to reconstruct are known to be nonnegative:

$$\hat{\theta} = \arg \min_{\theta \geq 0} \Phi(\theta). \quad (3.5)$$

$R(\theta)$ is usually chosen to be a function which penalizes differences between neighboring pixels. It is traditionally expressed as:

$$R(\theta) = \frac{1}{2} \sum_m \sum_{j \in \mathcal{N}_m} w_{jm} \psi(\theta_j - \theta_m), \quad (3.6)$$

where \mathcal{N}_m is the neighborhood of voxel m , w_{jm} represent the weights for each term. $\psi(t)$ is called the potential or cost function associated with each difference. For a quadratic penalty, the potential function is quadratic, $\psi(t) = t^2/2$ in (3.6). The “weights” are usually chosen to be $w_{jm} = 1$ for 1st order neighbors and $w_{jm} =$

$1/\sqrt{2}$ for 2nd order ones. In 2D, a first order neighborhood includes north, south, east and west neighbors. In 3D, a first order neighborhood has 6 elements. Second order neighborhoods also include diagonal neighbors, having 8 elements in 2D and 26 elements in 3D.

The roughness penalty can be expressed in general as:

$$R(\theta) = \sum_{k=1}^K \psi_k([\mathbf{C}\theta - z]_k). \quad (3.7)$$

The form of the penalty in (3.7) is more general than (3.6). In this form, \mathbf{C} is a $K \times p$ matrix, and z is a K -vector, where K is the number of “soft constraints” of type $[\mathbf{C}\theta]_k \approx z_k$. $[\mathbf{C}\theta]_k$ is a shorthand notation for the k th entry of the vector $\mathbf{C}\theta$ or $[\mathbf{C}\theta]_k = \sum_{j=1}^p c_{kj}\theta_j$. ψ_k are the weighted potential functions associated with each soft constraint.

To represent (3.6) with (3.7), we have to have \mathbf{C} as the generalization of this 1-D differencing matrix form:

$$\mathbf{C} = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ & & \ddots & \ddots & \\ 0 & 0 & \dots & -1 & 1 \end{bmatrix},$$

and $z = 0$.

In (3.6), the neighbor pairs are counted twice, hence the $1/2$ term in front. However, the pairs are counted once in (3.7) and the number of distinct pairs is K . Specifically, in 2D, $K = 2p$ if we are using a 1st order neighborhood and $K = 4p$ for a 2nd order neighborhood (ignoring edge effects). The form (3.7) of the penalty term covers a broader range of penalties such as second order difference penalties [73] and more.

3.2.1 Potential Functions

In choosing a penalty function, the ability to model smooth regions as well as edges is a major concern. Since we use gradient based optimization techniques, it is

advantageous to use penalty functions with continuous first derivatives. It is desirable to have a convex potential function to ensure the convexity of the overall objective function if the log-likelihood is concave. If the objective function is convex, there is only one local minimum which is also the global minimum. If gradient based techniques reach a stationary point of the objective function, then it is the global optimum. The convexity of the potential function $\psi(t)$ ensures the convexity of the penalty function $R(\theta)$. The potential functions are also chosen to be symmetric because they act as a norm and there is no need to differentiate between negative or positive differences. We consider penalties derived from the following functions: Quadratic, Lange, Generalized Gaussian and Huber potential functions.

- **Quadratic Penalty:** The potential function is quadratic:

$$\psi(t) = t^2/2.$$

The effect of this potential is to smooth the reconstructed image globally. The quadratically penalized objective function is much easier to optimize than others due to its simple form. The penalty function can be written compactly as $R(\theta) = \frac{1}{2}\theta' \mathbf{C}' \mathbf{C} \theta$. The edges in the image are also smoothed out when a quadratic penalty is used. If edge preservation is desired, a nonquadratic penalty should be used. Nonquadratic penalties can be chosen to produce sharper edges. In the transmission case, the attenuation map is composed of homogeneous regions with sharp boundaries. So, it is desirable to use edge-preserving penalties in that case. However, in the emission case, there is often not a clear separation of homogeneous regions. Moreover, nonquadratic penalties (as described next) have an “edge size” parameter which is easy to determine in the transmission case but difficult in the emission case. Another disadvantage of nonquadratic penalties in the emission case is that they tend to produce blocky piecewise smooth regions and seem to result in artifacts that might be misinterpreted as tumors when actually there is none. Due to these problems with the nonquadratic penalties, quadratic penalties are desirable in the emission case. For the transmission scans, nonquadratic penalties should be used, which we explain

next.

- **Lange’s Penalty:**

$$\psi(t) = \delta^2 [|t/\delta| - \log(1 + |t/\delta|)].$$

This penalty was introduced in [68]. This nonquadratic penalty behaves like the absolute value penalty for $|t| \gg \delta$, but like the quadratic penalty for $|t| \ll \delta$. δ needs to be predetermined. It is chosen to be less than the minimum edge difference in the reconstructed image so that the neighbor differences at the edges are less penalized than the differences in the homogeneous regions. We used this penalty function in most of our reconstructions in this thesis.

- **Generalized Gaussian MRF (GGMRF) penalties:** These potential functions were first introduced by Bouman and Sauer [5] and are defined as:

$$\psi(t) = |t|^p,$$

where $1 \leq p \leq 2$. When $p = 2$, this is equivalent to the quadratic penalty. The GGMRF penalties have two nice properties when $p = 1$ (the absolute value potential) [5]: 1) invariance to scaling of data, 2) they do not differentiate between slow monotonic changes and abrupt changes, and consequently does not penalize the presence of edges in the image. However, GGMRF priors are challenging for gradient-based optimization techniques since the second derivative is infinite at $t = 0$ for $p < 2$. The reconstructions with $p = 1$ usually result in blocky images. To avoid these problems, one can round the tip of the potential function at $t = 0$ by replacing the potential with a quadratic function for small values which we consider next.

- **Generalized Huber Penalty:** Huber penalty was introduced by Huber [59] to circumvent problems with the absolute value penalty. Here, we generalize it for p values other than $p = 1$:

$$\psi(t) = \begin{cases} \frac{1}{2}p\delta^{p-2}t^2, & |t| \leq \delta, \\ |t|^p - (1 - p/2)\delta^p, & |t| > \delta, \end{cases}$$

for $1 \leq p \leq 2$. The potential function is quadratic for $|t| < \delta$ and it is parallel to $|t|^p$ for $|t| > \delta$. The first derivatives from left and right match at $|t| = \delta$, thus it has a continuous first derivative. This penalty is not scale invariant and the choice of δ depends on the scale as for the Lange penalty. The generalized Huber penalty can be seen as a regularization of GGMRF penalty.

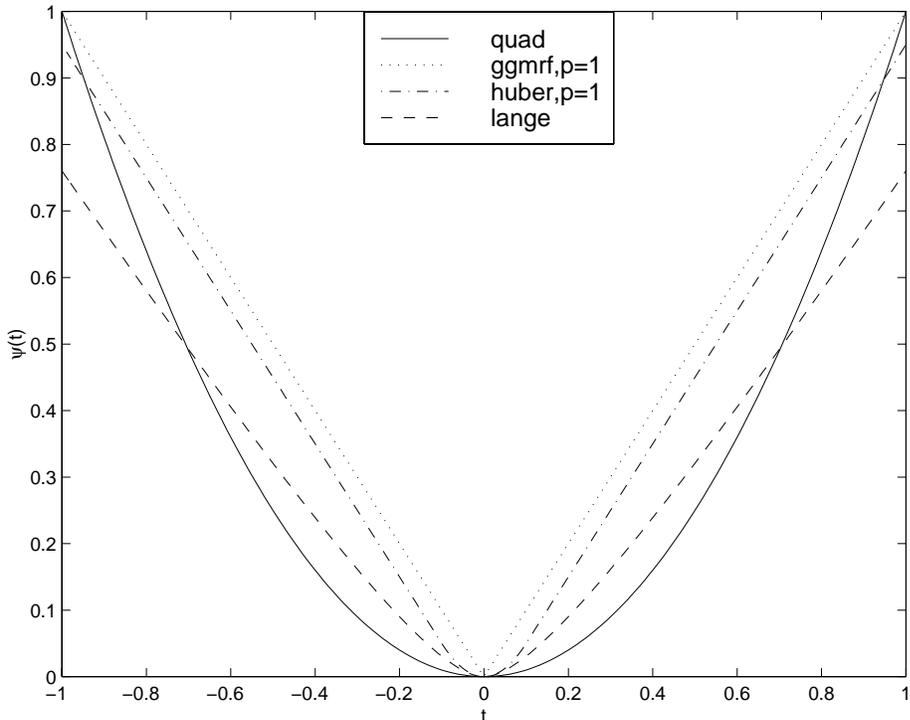


Figure 3.1: Plots of various potential functions $\psi(t)$.

Representative plots for the above potential functions are given in Figure 3.1. Many other convex and nonconvex penalty functions (or priors) were introduced some of which are Geman & McClure [48], Green’s log-cosh [51] and others in [68].

3.2.2 Resolution

The images reconstructed with penalized likelihood exhibit spatially varying resolution properties [46] when spatially invariant weights w_{jk} are used in (3.6). This is obvious with nonquadratic penalties since they smooth the image more in homogeneous regions and less in the edges. However, even with quadratic penalties, the

above statement is correct. High intensity regions are smoothed more than the low intensity regions and the point spread functions are not circularly symmetric [46]. The weights w_{jk} can be modified to obtain a more uniform resolution over all the image. Fessler’s certainty based penalty [46] yields more uniform resolution in terms of the average FWHM of the point spread function over the image. But, it still has non-uniform resolution in that the psf is not circularly symmetric but the level contours look like ellipses whose orientation are image dependent and space-variant. Stayman and Fessler have recently proposed a new modification to the quadratic penalty [106] which yields more circularly symmetric uniform resolution properties.

3.2.3 Algorithms

There is no closed form solution to the penalized likelihood problem (3.5). The original EM algorithm cannot be directly applied because the penalty term adds coupling to the M-step of the EM algorithm. Adding a regularizing penalty term to the log-likelihood improves the condition of the problem and causes algorithms to converge faster. There are many algorithms proposed for this problem.

The algorithms can be classified in many ways. One possible classification is in regards to the monotonicity of the algorithm. A monotonic algorithm is one that is guaranteed to decrease the objective function at each iteration. There are three types: 1) Intrinsically monotonic: Monotonicity is guaranteed by the algorithm, 2) Forced monotonic: Can be made monotonic by doing a line search, 3) Nonmonotonic algorithms. Monotonic algorithms are desirable since convergence is guaranteed at least to a local optimum and the algorithm is guaranteed not to diverge. We present a monotonic algorithm in Chapter 4.

Another classification is the type of update an algorithm performs. Four categories are possible in this case:

1. Simultaneous Update Algorithms:

$$\theta^{n+1} = f(\theta^n, y).$$

In a simultaneous update algorithm, all the parameters are updated at once

using the previous iteration values and the measurements. The EM algorithm is an example for simultaneous update algorithms. Simultaneous updates are parallelizable since each voxel's update is decoupled from all the other ones. Simultaneous update algorithms can further be divided into two subgroups:

- (a) Separable simultaneous updates: $\theta_j^{n+1} = f_j^1(\theta_j^n, y)$.
- (b) Nonseparable simultaneous updates: $\theta_j^{n+1} = f_j^2(\theta^n, y)$.

In addition to EM, separable simultaneous update algorithms include De Pierro's modification of EM for PL [24] and Lange's Convex algorithm for transmission tomography [71]. The conjugate gradient algorithm is an example for nonseparable simultaneous update algorithms [85]. Nonnegativity is enforced easily for separable simultaneous updates, whereas it is very difficult for nonseparable ones.

2. Sequential Update Algorithms:

for $j = 1, \dots, p$

$$\theta_j^{n+1} = f_j(\theta^*, y)$$

end

In these algorithms, the parameters are updated sequentially and the latest estimate of the parameters are used in determining the next iteration value for the current one. We use θ^* to denote the latest values for the parameters. Coordinate descent is a very good example of a sequential update algorithm. Sequential updates are easy to perform since minimization is reduced to a one dimensional problem for each voxel. Nonnegativity is enforced easily by setting the iterates back to zero if they go negative. In sequential update algorithms, usually the high frequencies converge faster [98] so that if one starts with an image that has correct low frequencies (like a smooth FBP image), the iterations converge to the optimizer much faster than other algorithms. Another example of a sequential update algorithm is SAGE [44]. In SAGE, less informative

hidden data spaces for each pixel are used and each pixel is updated sequentially using the EM idea. It converges very fast for the emission problem. Sequential update algorithms are harder to implement since they require column access to the system matrix \mathbf{G} .

3. Group Update Algorithms:

for $k = 1, \dots, N_S$

$$\theta_{S_k} = f_k(\theta^*, y)$$

end

In group update algorithms, a group of variables S_k are updated at a time using the latest parameter values θ^* . This is a style in between simultaneous and sequential updates and may have a lot of benefits. They are parallelizable and require less computation per iteration than sequential ones if implemented carefully. Grouped descent [42, 99] is the algorithm that is in this category. These algorithms also require column access to \mathbf{G} and are fairly hard to implement.

4. Ordered Subsets/Block Iterative/Row Action Algorithms:

for $k = 1, \dots, N_S$

$$\theta^{n,k} = f_k(\theta^{n,k-1}, y_{S_k})$$

end

$$\theta^{n+1,0} = \theta^{n,N_S}$$

In an ordered subsets algorithm, all the parameters are updated using a subgroup of data at a time. Usually, these algorithms are approximations and are not guaranteed to converge. OSEM algorithm is the first of these algorithms [60]. RBI-EM modifies the OSEM to make it converge in the consistent case [10]. RAMLA is a row action algorithm and adds relaxation parameters to each update and proves convergence for a certain class of relaxation parameters in the ML problem [8]. OSC algorithm [64] is the ordered subsets version of the convex algorithm for transmission tomography. Although most of these algo-

rithms are not guaranteed to converge, the resultant image quality seems to be acceptable. They are fast algorithms and are very easy to implement.

We introduce new algorithms for transmission tomography penalized likelihood image reconstruction in this thesis. A new class of sequential update algorithms called PSCD is introduced in Chapter 4 which includes an intrinsically monotonic algorithm. A new class of separable simultaneous update algorithms called SPS and their ordered subsets versions called OSTR are introduced in Chapter 5. The paraboloidal surrogates idea can also be used to obtain algorithms for emission scans [41]. However, our focus in this thesis is the transmission problem. We consider both post-injection transmission and emission cases in Chapter 6 to obtain a joint estimation algorithm based on paraboloidal surrogates.

CHAPTER 4

Paraboloidal Surrogates Coordinate Descent Algorithms for Transmission Tomography

4.1 Introduction

¹Attenuation correction is required for quantitatively accurate image reconstruction in emission tomography. The accuracy of this correction is very important in both PET and SPECT [58]. Transmission scans are performed to measure the attenuation characteristics of the object and to determine attenuation correction factors (ACFs) for emission image reconstruction. Conventional smoothing methods for ACF computation are simple and fast, but suboptimal [28, 84]. For low-count transmission scans, statistical reconstruction methods provide lower noise ACFs. However, a drawback of statistical methods is the slow convergence (or possible divergence) of current reconstruction algorithms. This chapter describes fast and monotonic algorithms for penalized-likelihood reconstruction of attenuation maps from transmission scan data. These reconstructed attenuation maps can be reprojected to calculate lower noise ACFs for improved emission image reconstruction.

Statistical methods for reconstructing attenuation maps from transmission scans are becoming increasingly important in thorax and whole-body PET imaging, where lower counts and short scan times are typical. 3-D PET systems also require attenu-

¹This chapter is based on [30].

ation correction, which can be done by reprojecting 2-D attenuation maps. SPECT systems with transmission sources are becoming increasingly available where statistical algorithms can be efficiently used for attenuation map reconstructions. For low-count transmission scans, the non-statistical FBP reconstruction method systematically overestimates attenuation map coefficients, whereas data-weighted least squares methods (WLS) for transmission reconstruction are systematically negatively biased [35]. By accurate statistical modeling, penalized-likelihood reconstruction of attenuation maps eliminates the systematic bias and yields lower variance relative to linear methods. Hence, we focus on penalized-likelihood image reconstruction rather than WLS in our work.

There are many reconstruction algorithms based on the Poisson model for transmission measurements. The expectation maximization (EM) algorithm [26], which led to a simple M-step for the emission problem, does not yield a closed form expression for the M-step in the transmission case [70]. Modifications of the transmission ML-EM algorithm [89, 66, 7] as well as algorithms that directly optimize the penalized-likelihood objective [6, 98, 42, 99, 84] have been introduced. Some of these algorithms seem to converge rapidly in the convex case.

However, up to now, no practically realizable monotonic (or convergent) algorithm has been found for the penalized-likelihood problem when the objective is not convex. The negative log-likelihood is nonconvex when there are “background” counts in the data. This is unavoidable in PET and SPECT, due to the accidental coincidences in PET and emission crosstalk² in SPECT. The assumption of no background counts may be reasonable in some X-ray CT applications.

In this chapter, we present a new algorithm which is guaranteed to be monotonic even when the objective function is nonconvex. This algorithm depends on paraboloidal surrogate functions for the log-likelihood which transform the problem into a simpler quadratic optimization problem at each iteration. The transformed problem at each iteration is similar to a Penalized Weighted Least Squares (PWLS)

²Even though different photon energies are used in simultaneous emission/transmission SPECT imaging, some emission events are recorded in the transmission energy window due to Compton scatter and finite energy resolution.

problem, and thus has a familiar and simple form. This quadratic problem need not be solved exactly; an algorithm that monotonically decreases the surrogate function suffices. Since evaluating the gradient and Hessian of the surrogate function is much less costly, the CPU time per iteration is greatly reduced as compared to algorithms that directly attempt to minimize the objective function, such as coordinate descent. Remarkably, the convergence rate is comparable to other direct algorithms. For non-convex objective functions, monotonicity alone does not guarantee convergence to the global minimizer when local minima exist, but it does ensure that the estimates do not diverge since the likelihood is bounded. Whether the transmission log likelihood has multiple local minima is an open question.

The “surrogate” or “substitute” function idea is not new to the tomographic reconstruction area. EM algorithms can be viewed as providing a surrogate function for the log-likelihood function by means of a statistically more informative “complete” data set which is unobservable [26]. The conditional expectation of the log-likelihood function for this new space is often easier to maximize, having a closed form for the emission case. This statistical construction of surrogate functions is somewhat indirect and seems to yield a limited selection of choices. De Pierro has developed surrogate functions for nonnegative least squares problems based solely on convexity arguments, rather than statistics [23]. Our proposed approach is similar in spirit.

The EM algorithm did not result in a closed form M-step for the transmission case [70], so direct minimization of the objective function became more attractive. Cyclic Newtonian coordinate descent (CD,NR) [98] has been used effectively in transmission tomography. However, coordinate descent based on Newton’s iteration for each pixel is not guaranteed to be monotonic. Furthermore, an iteration of Newton-based coordinate descent requires at least M exponentiations and $17M$ floating point operations³, where M is the (very large) number of nonzero elements in the system matrix \mathbf{G} in (4.1) below. These exponentiations and floating point operations constitute a significant fraction of the CPU time per iteration. Recently, Zheng *et al.*

³This can be reduced to $9M$ floating point operations if the denominator terms are precomputed similar to section 4.3.7.

introduced a “functional substitution” (FS) method [125, 97] which is proven to be monotonic for transmission scans with no background counts ($r_i = 0$ in (4.1) below). Like coordinate descent, FS algorithm cyclically updates the coordinates of the image vector, *i.e.* the attenuation map values for each pixel. However, instead of minimizing the original complex objective function with respect to each parameter, FS algorithm minimizes a 1-D parabolic surrogate function. The minimization of the surrogate is guaranteed to monotonically decrease the original objective function if the derivative of the negative log-likelihood is concave (which is true when $r_i = 0$) [125, 97]. On the other hand, the FS algorithm requires at least $2M$ exponentiations and $17M$ floating point operations⁴ per iteration, which means that the guarantee of monotonicity comes at a price of significantly increased computation time per iteration for that method. Furthermore, the FS algorithm is not monotonic in the nonconvex case of interest in PET and SPECT, where $r_i \neq 0$.

De Pierro [24] has used a surrogate function for the penalty part of the penalized-likelihood problem for convex penalties. The surrogate function idea was also used in several algorithms which update a group of pixel values at a time instead of sequential update of each pixel. Examples of these types of algorithms are the convex algorithm of [71] which updates all pixels simultaneously and the grouped coordinate ascent (GCA) algorithm of [42, 125] which updates a subset of pixels at a time. The surrogate functions used in these algorithms were obtained using De Pierro’s convexity trick [24] to form a separable function that is easier to minimize than the non-separable original objective function. The convergence rates per iteration decrease due to the higher curvature of these surrogate functions, but these algorithms require less computation per iteration as compared to single coordinate descent [98] and are parallelizable. Furthermore, it is trivial to impose the nonnegativity constraint with an additively separable surrogate function [42].

In this work, we propose to use a global surrogate function for the original objective function. This global surrogate function is *not* separable, but has a simple quadratic form. The method is based on finding 1-D parabolic functions that are tangent to

⁴Precomputation of the denominator terms in FSCD would destroy monotonicity.

and lie above each of the terms in the log-likelihood, similar to Huber’s method for robust linear regression [59]. Whereas Huber considered strictly convex cost functions, we extend the method to derive provably monotonic algorithms even for nonconvex negative log-likelihood functions. Remarkably, these algorithms require *less* CPU time to converge than the fastest algorithm introduced before (GCA of [42]) and as an additional advantage, they are proven to be monotonic. We call the new approach to image reconstruction the “Paraboloidal Surrogates” (PS) method.

In the rest of this chapter, we describe the problem, develop the new algorithm, and present representative performance results on real PET transmission data.

4.2 The Problem

The measurements in a photon-limited application such as PET or SPECT are modeled appropriately as Poisson random variables. In transmission tomography, the means of the prompt coincidences are related exponentially to the projections (or line integrals) of the attenuation map through Beer’s Law [70]. In addition, the measurements are contaminated by extra “background” counts due mostly to random coincidences and scatter in PET and emission crosstalk in SPECT. Thus, it is realistic to assume the following model:

$$y_i \sim \text{Poisson}\{b_i e^{-[\mathbf{G}\boldsymbol{\mu}]_i} + r_i\}, \quad i = 1, \dots, N, \quad (4.1)$$

where N is the number of measurements, μ_j is the average linear attenuation coefficient in voxel j for $j = 1, \dots, p$, and p denotes the number of voxels. The notation $[\mathbf{G}\boldsymbol{\mu}]_i = \sum_{j=1}^p g_{ij}\mu_j$ represents the i th “line integral” of the attenuation map $\boldsymbol{\mu}$, and $\mathbf{G} = \{g_{ij}\}$ is the $N \times p$ system matrix. We assume that $\{b_i\}$, $\{r_i\}$ and $\{g_{ij}\}$ are known nonnegative constants⁵, where r_i is the mean number of background events, b_i is the blank scan factor, and y_i represents the number of transmission events counted by

⁵The assumption that the background counts r_i are known nonnegative constants is an approximation. In PET, we estimate the r_i ’s by smoothing the delayed coincidences from the transmission scan [85]. Alternatively, one can use time scaled delayed coincidences from a blank scan (which are less noisy due to longer scan times) as the r_i factors [121] or use Bayesian estimation techniques to estimate r_i ’s from delayed coincidences [84, 85].

the i th detector (or detector pair in PET).

We seek to find a statistical estimate of the attenuation map μ which “agrees” with the data and is anatomically reasonable. For this purpose, a natural approach is to use a likelihood-based estimation strategy. The log-likelihood function for the independent transmission data is:

$$L(\mu) = \sum_{i=1}^N \left\{ y_i \log(b_i e^{-[\mathbf{G}\mu]_i} + r_i) - (b_i e^{-[\mathbf{G}\mu]_i} + r_i) \right\},$$

ignoring constant terms. The log-likelihood depends on the parameter vector μ through only its projections $[\mathbf{G}\mu]_i$ and can be expressed in the following form:

$$-L(\mu) = \sum_{i=1}^N h_i([\mathbf{G}\mu]_i), \quad (4.2)$$

where the contribution of the i th measurement to the negative log-likelihood is given by:

$$h_i(l) \triangleq (b_i e^{-l} + r_i) - y_i \log(b_i e^{-l} + r_i). \quad (4.3)$$

The proposed algorithm exploits the additive form of (4.2). Directly minimizing $-L(\mu)$ (maximum likelihood) results in a very noisy estimate $\hat{\mu}$ due to the ill-posed nature of the problem. However, it is well known that the attenuation map in the body consists of approximately locally homogeneous regions. This property has formed the basis of many segmentation methods for transmission scans [80]. Rather than applying hard segmentation, we add to the negative log-likelihood a penalty term which encourages piecewise smoothness in the image, resulting in the penalized-likelihood image reconstruction formulation as given below:

$$\hat{\mu} = \arg \min_{\mu \geq 0} \Phi(\mu), \quad \Phi(\mu) = -L(\mu) + \beta R(\mu). \quad (4.4)$$

Our goal is to develop an algorithm for finding the minimizing $\hat{\mu}$ with minimal CPU time.

We consider roughness penalties $R(\mu)$ that can be expressed in the following very general form [38, 24]:

$$R(\mu) = \sum_{k=1}^K \psi_k([\mathbf{C}\mu]_k), \quad (4.5)$$

where the ψ_k 's are potential functions acting as a norm on the “soft constraints” $C\mu \approx 0$ and K is the number of such constraints. The functions ψ_k we consider are convex, symmetric, nonnegative, differentiable and satisfy some more conditions that are listed in Section 4.3.3. The β in equation (4.4) is a parameter which controls the level of smoothness in the final reconstructed image. For more explanation of the penalty function, see [38].

The objective function defined in (4.4) is not convex when there are nonzero background counts ($r_i \neq 0$) in the data. In this realistic case, there is no guarantee that there is a single global minimum. However, some practical algorithms exist that seem to work very well, yet none of them are proven to be monotonic. In this chapter we introduce an algorithm that is monotonic even when Φ is not convex. The new approach is based on successive paraboloidal surrogate functions and will be explained in the rest of the chapter.

4.3 Paraboloidal Surrogates Algorithms

The penalized-likelihood objective function $\Phi(\mu)$ has a complex form that precludes analytical minimization. Thus, iterative methods are necessary for minimizing $\Phi(\mu)$. Our approach uses the optimization transfer idea proposed by De Pierro [23, 24], summarized as follows. Let μ^n be the attenuation map estimate after the n th iteration. We would like to find a “surrogate” function⁶ $\phi(\mu; \mu^n)$ which is easier to minimize or to monotonically decrease than $\Phi(\mu)$. This approach transforms the optimization problem into a simpler problem at each iteration, as illustrated in Figure 4.1. The following “monotonicity” condition on the surrogate function is sufficient to ensure that the iterates $\{\mu^n\}$ monotonically decrease Φ :

$$\Phi(\mu) - \Phi(\mu^n) \leq \phi(\mu; \mu^n) - \phi(\mu^n; \mu^n), \quad \forall \mu \geq 0. \quad (4.6)$$

We restrict ourselves to differentiable surrogate functions, for which the following

⁶We use the notation $\phi(\mu; \mu^n)$ to emphasize that the surrogate is a function of μ once μ^n is fixed and it changes for each μ^n , following the Q function notation of the EM algorithm [26].

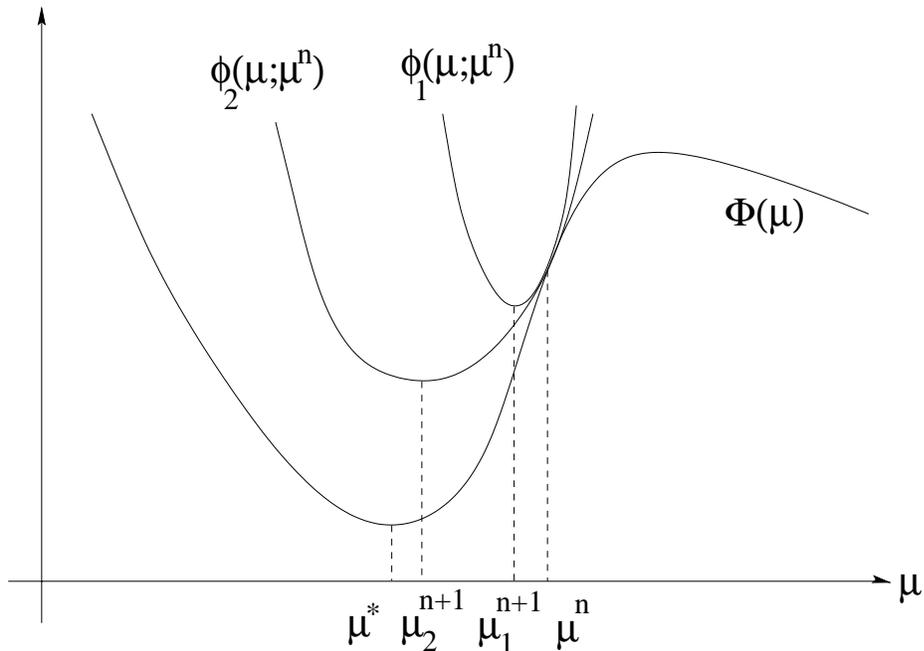


Figure 4.1: One-dimensional illustration of the optimization transfer principle. Instead of minimizing $\Phi(\mu)$, we minimize the surrogate function $\phi(\mu; \mu^n)$ at the n th iteration. Here, the surrogate function ϕ_2 has a smaller curvature and is wider than ϕ_1 , thus it has a bigger step size and hence faster convergence rate to the local minimum μ^* .

conditions are sufficient⁷ to ensure (4.6):

1. $\phi(\mu^n; \mu^n) = \Phi(\mu^n)$
2. $\left. \frac{\partial \phi}{\partial \mu_j}(\mu; \mu^n) \right|_{\mu=\mu^n} = \left. \frac{\partial \Phi}{\partial \mu_j}(\mu) \right|_{\mu=\mu^n}, \forall j = 1, \dots, p$ (4.7)
3. $\phi(\mu; \mu^n) \geq \Phi(\mu)$ for $\mu \geq 0$.

Figure 4.1 illustrates two different surrogate functions that are tangent to the original objective at the current iterate and lie above it for all feasible values of the parameters.

The EM algorithm [70] provides a statistical method for constructing surrogate functions $\phi(\mu; \mu^n)$ satisfying the above conditions. However, in the transmission tomography problem, the natural EM surrogate is difficult to minimize and leads to slow convergence. In this work, we construct a simpler surrogate using ordinary calculus

⁷The second condition follows from the other two conditions for differentiable surrogate functions.

rather than statistical techniques.

The log-likelihood function (4.2) has a certain kind of dependence on the parameters μ , namely through their projections $\mathbf{G}\mu$. The negative log-likelihood is the sum of individual functions h_i , each of which depends on a single projection only. We can exploit this form of the likelihood function by selecting a 1-D surrogate function for each of the one-dimensional h_i functions in the projection (l) domain. The overall sum of these individual 1-D functions will be an appropriate surrogate for the likelihood part of the objective.

Let $l_i^n = [\mathbf{G}\mu^n]_i$ denote the estimate of the i th line integral of the attenuation coefficient at the n th iteration. We choose the following quadratic form for the surrogate functions q_i :

$$q_i(l; l_i^n) \triangleq h_i(l_i^n) + \dot{h}_i(l_i^n)(l - l_i^n) + \frac{1}{2}c_i(l_i^n)(l - l_i^n)^2, \quad (4.8)$$

where $c_i(l_i^n)$ is the curvature of the parabola q_i and \dot{h} denotes first derivative of h . This construction ensures that $q_i(l_i^n; l_i^n) = h_i(l_i^n)$ and $\dot{q}_i(l_i^n; l_i^n) = \dot{h}_i(l_i^n)$ similar to (4.7). To ensure monotonicity, we must choose the curvatures $c_i(l_i^n)$ to satisfy the following inequality at each iteration:

$$h_i(l) \leq q_i(l; l_i^n), \quad \text{for } l \geq 0. \quad (4.9)$$

After determining the parabolas, one can easily verify that the following function is a global surrogate function for the objective $\Phi(\mu)$ which satisfies the properties in (4.7):

$$\phi(\mu; \mu^n) = Q(\mu; \mu^n) + \beta R(\mu), \quad (4.10)$$

where

$$\begin{aligned} Q(\mu; \mu^n) &\triangleq \sum_{i=1}^N q_i([\mathbf{G}\mu]_i; l_i^n) \\ &= \Phi(\mu^n) + \mathbf{d}_h(l^n)' \mathbf{G}(\mu - \mu^n) + \frac{1}{2}(\mu - \mu^n)' \mathbf{G}' \mathbf{D}(c_i(l_i^n)) \mathbf{G}(\mu - \mu^n) \end{aligned} \quad (4.11)$$

where the column vector $\mathbf{d}_h(l^n) \triangleq [\dot{h}_i(l_i^n)]_{i=1}^N$, \mathbf{x}' denotes the transpose of \mathbf{x} , and $\mathbf{D}(c_i(l_i^n))$ is the $N \times N$ diagonal matrix with diagonal entries $c_i(l_i^n)$ for $i = 1, \dots, N$.

The surrogate function $\phi(\mu; \mu^n)$ in (4.10) consists of the sum of a paraboloid (*i.e.* a quadratic form) and the convex penalty term. An algorithm that decreases the

function ϕ will also monotonically decrease the objective function if the inequality in (4.9) holds. The general paraboloidal surrogates (PS) method can be outlined as follows:

for each iteration n

determine $c_i(l_i^n)$ and consequently $\phi(\mu; \mu^n)$

find a $\mu^{n+1} \geq 0$ that decreases (or minimizes) $\phi(\mu; \mu^n)$

end.

The key design choices in the general method outlined above are:

1. The different ways of choosing the curvatures $c_i(l_i^n)$'s which would satisfy (4.9).
2. The algorithm to monotonically decrease $\phi(\mu; \mu^n)$ defined in (4.10) for $\mu \geq 0$.

Each combination of choices leads to a different algorithm, as we elaborate in the following sections.

4.3.1 Maximum Curvature

A natural choice for $c_i(l_i^n)$ is the maximum second derivative in the feasible region for the projections. This “maximum curvature” ensures that (4.9) holds, which follows from the generalized mean value theorem for twice differentiable functions (page 228, [17]). The feasible region for the projections is $[0, \infty)$ due to the nonnegativity constraint. Hence, the choice

$$c_i(l_i^n) = \max_{l \in [0, \infty)} \{\ddot{h}_i(l)\} \quad (4.13)$$

is guaranteed to satisfy (4.9). We show in Appendix A that the closed form expression for $c_i(l_i^n)$ is:

$$c_i(l_i^n) = \left[\left(1 - \frac{y_i r_i}{(b_i + r_i)^2} \right) b_i \right]_+ \quad (4.14)$$

where $[x]_+ = x$ for $x > 0$ and zero otherwise. Thus, it is trivial to compute the $c_i(l_i^n)$ terms in this case. The choice (4.14) for the curvature $c_i(l_i^n)$ does not depend on the

iteration n , so it is a constant. We refer to this choice as the “maximum curvature” (PS,M,CD).

Having specified the curvatures $\{c_i(l_i^n)\}$, the paraboloidal surrogate $Q(\mu; \mu^n)$ in (4.12) is now fully determined. Next we need an algorithm that decreases or minimizes the surrogate function $\phi(\mu; \mu^n)$.

4.3.2 Algorithms for Minimizing the Paraboloidal Surrogate

In the absence of the nonnegativity constraint, in principle one could minimize the surrogate function $\phi(\mu; \mu^n)$ over μ by zeroing its gradient. The column gradient of $\phi(\mu; \mu^n)$ with respect to μ is given by

$$\nabla_{\mu}\phi(\mu; \mu^n) = \mathbf{G}'\mathbf{d}_h(l^n) + \mathbf{G}'\mathbf{D}(c_i(l_i^n))\mathbf{G}(\mu - \mu^n) + \beta\nabla R(\mu). \quad (4.15)$$

If $R(\mu)$ is a quadratic form, *i.e.* $R(\mu) = \frac{1}{2}\mu'\mathbf{R}\mu$, then we can analytically zero the gradient, yielding the iteration:

$$\mu^{n+1} = \mu^n - [\mathbf{G}'\mathbf{D}(c_i(l_i^n))\mathbf{G} + \beta\mathbf{R}]^{-1}\nabla_{\mu}\phi(\mu^n). \quad (4.16)$$

There are three problems with the above iteration. It does not enforce the non-negativity constraint, the matrix inverse is impractical to compute exactly, and it is limited to quadratic penalty functions. To overcome these limitations, we instead apply a monotonic coordinate descent iteration to decrease $\phi(\mu; \mu^n)$.

4.3.3 Coordinate Descent Applied to the Surrogate Function

To apply coordinate descent to monotonically decrease the surrogate function $\phi(\mu; \mu^n)$, we need a quadratic function that majorizes (*i.e.* lies above $\forall \mu \geq 0$) the function $\phi(\mu; \mu^n)$ at each pixel. We treat the likelihood part and the penalty part separately. Let $\hat{Q}_j^n(\mu_j) \triangleq Q([\hat{\mu}_1, \dots, \hat{\mu}_{j-1}, \mu_j, \hat{\mu}_{j+1}, \dots, \hat{\mu}_p]; \mu^n)$ and $\hat{R}_j^n(\mu_j) \triangleq R([\hat{\mu}_1, \dots, \hat{\mu}_{j-1}, \mu_j, \hat{\mu}_{j+1}, \dots, \hat{\mu}_p])$, where $\hat{\mu}$ denotes the current estimate of the parameter μ . Then we must select curvatures d_j^n and \hat{p}_j that satisfy the following:

$$\hat{Q}_j^n(\mu_j) = Q(\hat{\mu}; \mu^n) + \dot{Q}_j^n(\hat{\mu})(\mu_j - \hat{\mu}_j) + \frac{1}{2}d_j^n(\mu_j - \hat{\mu}_j)^2 \quad (4.17)$$

$$\hat{R}_j^o(\mu_j) \leq \hat{R}_j(\mu_j) \triangleq R(\hat{\mu}) + \dot{R}_j(\hat{\mu})(\mu_j - \hat{\mu}_j) + \frac{1}{2}\hat{p}_j(\mu_j - \hat{\mu}_j)^2, \quad \forall \mu_j \geq 0, \quad (4.18)$$

where $\hat{Q}_j^n(\mu_j)$ and $\hat{R}_j^o(\mu_j)$ are treated as functions of μ_j only. Equality is achievable in (4.17) since the likelihood surrogate $\hat{Q}_j^n(\mu_j)$ is quadratic. For the penalty part $\hat{R}_j^o(\mu_j)$, we must find a quadratic function $\hat{R}_j(\mu_j)$ that lies above it, by appropriate choice of \hat{p}_j as considered below.

The derivative of the likelihood surrogate parabola at $\hat{\mu}_j$ is (from (4.11))

$$\dot{Q}_j^n(\hat{\mu}) \triangleq \left. \frac{\partial}{\partial \mu_j} \hat{Q}_j^n(\mu_j) \right|_{\mu_j = \hat{\mu}_j} = \sum_{i=1}^N g_{ij} \dot{q}_i(\hat{l}_i),$$

where from (4.8)

$$\dot{q}_i(\hat{l}_i) = \dot{h}_i(l_i^n) + c_i(l_i^n)(\hat{l}_i - l_i^n), \quad (4.19)$$

where $\hat{l}_i = \sum_{j=1}^N g_{ij} \hat{\mu}_j$, and

$$\dot{h}_i(l) = \left(\frac{y_i}{b_i e^{-l} + r_i} - 1 \right) b_i e^{-l}. \quad (4.20)$$

From (4.8) and (4.11), the curvature of the parabola $\hat{Q}_j^n(\mu_j)$ is obviously:

$$d_j^n \triangleq \sum_{i=1}^N g_{ij}^2 c_i(l_i^n). \quad (4.21)$$

From (4.5), the derivative of the penalty part at $\hat{\mu}_j$ is

$$\dot{R}_j(\hat{\mu}) \triangleq \left. \frac{\partial}{\partial \mu_j} \hat{R}_j^o(\mu_j) \right|_{\mu_j = \hat{\mu}_j} = \sum_{k=1}^K c_{kj} \dot{\psi}_k([\mathbf{C}\hat{\mu}]_k).$$

We must obtain a parabolic surrogate $\hat{R}_j(\mu_j)$ that satisfies (4.18). We assume the potential functions $\psi_k(\cdot)$ satisfy the following conditions:

- ψ is symmetric
- ψ is everywhere differentiable (and therefore continuous)
- $\dot{\psi}(t) = d/dt \psi(t)$ is non-decreasing (and hence ψ is convex)
- $\omega_\psi(t) \triangleq \dot{\psi}(t)/t$ is non-increasing for $t \geq 0$
- $\omega_\psi(0) = \lim_{t \rightarrow 0} \dot{\psi}(t)/t$ is finite and nonzero *i.e.* $0 < \omega_\psi(0) < \infty$.

In the context of robust regression, Huber showed (Lemma 8.3 on page 184 in [59], also [38]) that for potential functions ψ_k that satisfy the conditions above, we can find

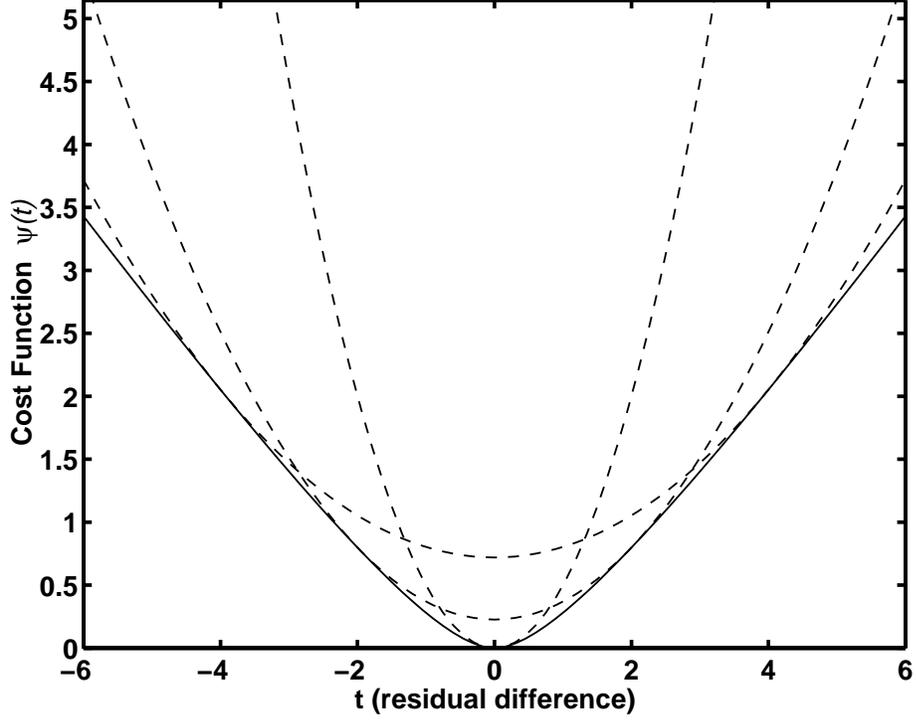


Figure 4.2: Illustration of the tangent parabolas lying above a potential function.

a parabola $\hat{\psi}_k(t)$ that lies above $\psi_k(t)$, $\forall t \in \mathbb{R}$. This parabola $\hat{\psi}_k(t)$ is tangent to the potential function at the current point $\hat{t}_k \triangleq [\mathbf{C}\hat{\mu}]_k$ and at $-\hat{t}_k$ and has the curvature $\omega_{\psi_k}(\hat{t}_k)$ where $\omega_{\psi}(\cdot)$ was defined above. The surrogate parabola is given by:

$$\hat{\psi}_k(t) = \psi_k(\hat{t}_k) + \dot{\psi}_k(\hat{t}_k)(t - \hat{t}_k) + \frac{1}{2}\omega_{\psi_k}(\hat{t}_k)(t - \hat{t}_k)^2,$$

and is illustrated in Figure 4.2. Thus, the following is a surrogate parabola for the penalty part of the objective function:

$$\hat{R}_j(\mu_j) = \sum_{k=1}^K \hat{\psi}_k([\mathbf{C}\mu]_k) \Big|_{\mu_m = \hat{\mu}_m, \forall m \neq j}. \quad (4.22)$$

The curvature of the parabola $\hat{R}_j(\mu_j)$ is:

$$\hat{p}_j \triangleq \sum_{k=1}^K c_{kj}^2 \omega_{\psi_k}([\mathbf{C}\hat{\mu}]_k). \quad (4.23)$$

Combining the above surrogate parabolas (4.17) and (4.22), the minimization step

of the coordinate descent for pixel j is simply:

$$\hat{\mu}_j^{\text{new}} = \arg \min_{\mu_j \geq 0} \hat{Q}_j^n(\mu_j) + \beta \hat{R}_j(\mu_j) = \left[\hat{\mu}_j - \frac{\dot{Q}_j^n(\hat{\mu}) + \beta \dot{R}_j(\hat{\mu})}{d_j^n + \beta \hat{p}_j} \right]_+. \quad (4.24)$$

This is an update that monotonically decreases the value of $\phi(\cdot; \mu^n)$ and consequently the value of $\Phi(\cdot)$. One iteration is finished when all pixels are updated via (4.24) in a sequential order. We usually update the paraboloidal surrogate function after one iteration of coordinate descent (CD), but one could also perform more than one CD iteration per surrogate. We call this method the Paraboloidal Surrogates Coordinate Descent (PSCD) method.

The PSCD algorithm with the curvatures obtained from (4.14) is outlined in Table 4.1. In this table, the algorithm flow is given for the general case where $c_i(l_i^n)$ may change at each iteration. However, the curvatures $c_i(l_i^n)$ given in (4.26) in Table 4.1 are constant throughout the iterations. If one uses fixed $c_i(l_i^n)$ values which do not depend on n as in (4.26), then the d_j^n terms can be precomputed and the algorithm should be reorganized to take this computational advantage into account.

Another computational advantage of curvatures that do not depend on the iterations is as follows. If we define $\tilde{q}_i = \dot{q}_i / \sqrt{c_i}$ and $\tilde{g}_{ij} = g_{ij} \sqrt{c_i}$, then the update in (4.29) will be simplified to:

$$\tilde{q}_i := \tilde{q}_i + \tilde{g}_{ij}(\mu_j^{\text{work}} - \hat{\mu}_j),$$

which decreases the computation time devoted to back and forward projections per iteration by about 20% for implementations using precomputed system matrices. The equations (4.28) and (4.30) should also be modified to use the new variables. We have not implemented this faster version for this work.

The algorithm in Table 4.1 requires roughly double the floating point operations required for one forward and one backprojection per iteration. The gradient of the original log-likelihood with respect to the projections $\{\dot{h}_i(l_i^n)\}_{i=1}^N$ and the curvature terms $c_i(l_i^n)$ are computed only once per iteration⁸. The gradient of the surrogate

⁸In contrast to PSCD algorithm, when coordinate descent (CD,NR) is applied to the original objective function, new gradients and curvatures must be computed after each pixel is updated. These computations involve expensive exponentiations and floating point operations which increase the CPU time required for original coordinate descent.

Initialize: $\hat{\mu} = \text{FBP}\{\log(b_i/(y_i - r_i))\}_{i=1}^N$ and $\hat{l}_i = \sum_{j=1}^p a_{ij}\hat{\mu}_j$, $\forall i = 1, \dots, N$
for each iteration $n = 0, \dots, \text{Niter} - 1$

$$\dot{q}_i = \dot{h}_i = \left(\frac{y_i}{b_i e^{-\hat{l}_i} + r_i} - 1 \right) b_i e^{-\hat{l}_i}, \text{ for } i = 1, \dots, N \quad (4.25)$$

$$c_i = \max_{l \geq 0} \ddot{h}_i(l) = \left[\left(1 - \frac{y_i r_i}{(b_i + r_i)^2} \right) b_i \right]_+, \text{ for } i = 1, \dots, N \quad (4.26)$$

$$c_i := \begin{cases} c_i, & c_i > \epsilon \\ \epsilon, & c_i \leq \epsilon \end{cases} \quad (4.27)$$

repeat one or more times

for $j = 1, \dots, p$

$$\begin{aligned} \dot{Q}_j &= \sum_{i=1}^N a_{ij} \dot{q}_i, & d_j &= \sum_{i=1}^N a_{ij}^2 c_i \\ \mu_j^{\text{old}} &= \hat{\mu}_j \end{aligned} \quad (4.28)$$

for a couple sub-iterations

$$\hat{\mu}_j := \left[\hat{\mu}_j - \frac{\dot{Q}_j + d_j(\hat{\mu}_j - \mu_j^{\text{old}}) + \beta \sum_{k=1}^K c_{kj} \dot{\psi}([\mathbf{C}\hat{\mu}]_k)}{d_j + \beta \sum_{k=1}^K c_{kj}^2 \omega_{\psi_k}([\mathbf{C}\hat{\mu}]_k)} \right]_+$$

end

$$\dot{q}_i := \dot{q}_i + a_{ij} c_i (\hat{\mu}_j - \mu_j^{\text{old}}) \forall i \text{ s.t. } a_{ij} \neq 0 \quad (4.29)$$

end

end

$$\hat{l}_i := \hat{l}_i + \frac{\dot{q}_i - \dot{h}_i}{c_i}, \text{ for } i = 1, \dots, N \quad (4.30)$$

end

Table 4.1: Algorithm outline for a paraboloidal surrogates algorithm with coordinate descent (PSCD). The curvature choice shown here is the maximum second derivative (PS,M,CD).

paraboloid uses \hat{q}_i terms which can be updated easily as shown in (4.29) in the algorithm. This implementation does not update the projections \hat{l}_i after each pixel update since they are only needed in the outer loop (4.25). The projections are computed in (4.30) after updating all pixels. The update (4.30) requires $c_i^n > 0$ to work. In (4.27), we constrain the curvature value to some small value $\epsilon > 0$ (which obviously does not hurt monotonicity) so that (4.30) can be evaluated for all $i = 1, \dots, N$. However, ϵ should not be very small since it will cause undesirable numerical precision errors. Storage requirements are also modest for the proposed algorithm. A single copy of the image and four sinograms for \hat{l}_i , c_i , \hat{h}_i and \hat{q}_i need to be stored in addition to data vectors y_i, b_i, r_i .

In the following, we discuss the convergence rate of the algorithm, which provides motivation for obtaining better curvatures.

4.3.4 Convergence and Convergence Rate

In the absence of background events, *i.e.* when $r_i = 0$, the penalized-likelihood objective Φ is convex and our proposed PSCD algorithm is globally convergent. This is a fairly straightforward consequence of the proof in [45] for convergence of SAGE, so we omit the details.

However when $r_i \neq 0$, little can be said about global convergence due to the possibility that there are multiple minima or a continuous region of minima. Our practical experience suggests that local minima are either unlikely to be present, or are quite far from reasonable starting images, since all experiments with multiple initializations of the algorithm yielded the same limit within numerical precision. The PSCD algorithm is monotonic even with the nonconvex objective function. One can easily show that every fixed point of the algorithm is a stationary point of the objective function and vice versa. Thus, it is comforting to know that the algorithm will converge to a local minimum and will not blow up.

The convergence rate of the proposed algorithm with the “maximum curvature” choice is suboptimal. The curvatures $c_i(l_i^n)$ are too conservative and the paraboloids

are unnecessarily narrow. Intuitively, one can deduce that smaller $c_i(l_i^n)$ values will result in faster convergence. The reason for this is that the lower the curvature, the wider the paraboloid and the bigger the step size as can be seen in Fig. 4.1. To verify this intuition, we analyze the convergence rate of the algorithm. For simplicity, we assume that a quadratic penalty is used in the reconstruction and that the surrogate function $\phi(\mu; \mu^n)$ (4.10) is minimized exactly.

Let $\hat{\mu}$ be the unconstrained minimizer of the original objective function. At step n , by zeroing the gradient of (4.10), we get the simple Newton-like update in (4.16). By Taylor series, for $\mu^n \approx \hat{\mu}$, we can approximate the gradient of the objective function as: $\nabla\Phi(\mu^n) \approx \mathbf{H}(\hat{\mu})(\mu^n - \hat{\mu})$, where $\mathbf{H}(\hat{\mu})$ is the Hessian of Φ at $\hat{\mu}$. Define $\mathbf{N}(c) = \mathbf{G}'\mathbf{D}(c_i)\mathbf{G} + \beta\mathbf{R}$, then from (4.16):

$$\begin{aligned} \mu^{n+1} - \hat{\mu} &\approx \mu^n - \hat{\mu} - [\mathbf{N}(c)]^{-1} \mathbf{H}(\hat{\mu})(\mu^n - \hat{\mu}) \\ &= (\mathbf{I} - [\mathbf{N}(c)]^{-1} \mathbf{H}(\hat{\mu}))(\mu^n - \hat{\mu}). \end{aligned} \quad (4.31)$$

This equation describes how the convergence rate of the proposed algorithm is affected by different c_i choices. We use the results from [39] to evaluate the convergence rate. Let $\mathbf{N}(c^1)$ and $\mathbf{N}(c^2)$ be two matrices corresponding to curvature vectors c^1 and c^2 respectively with $c_i^1 < c_i^2, \forall i$. Then obviously $\mathbf{N}(c^2) - \mathbf{N}(c^1)$ is positive definite and it follows from Lemma 1 in [39] that the algorithm corresponding to c^1 has a lower root-convergence factor and thus converges faster than the algorithm corresponding to c^2 .

Therefore, to optimize the convergence rate, we would like the $c_i(l_i^n)$ values to be as small as possible while still satisfying (4.9). The optimal choice for the curvatures is the solution to the following constrained optimization problem for each i :

$$c_i(l_i^n) = \min \left\{ c \geq 0 : h_i(l) \leq h_i(l_i^n) + \dot{h}_i(l_i^n)(l - l_i^n) + \frac{1}{2}c(l - l_i^n)^2 \quad \forall l \geq 0 \right\}. \quad (4.32)$$

This choice yields the fastest convergence rate while still guaranteeing monotonicity. In the following section, we discuss the solution to (4.32).

4.3.5 Optimum Curvature

The curvature that satisfies (4.32) is not trivial to find for general functions $h_i(\cdot)$. However, the marginal negative log-likelihood functions for each projection (h_i defined in (4.3)) in transmission tomography have some nice properties. We show the following in Appendix B. The parabola that is:

1. tangent to h_i at the current projection l_i^n , and
2. intersects h_i at $l = 0$,

is guaranteed to lie above $h_i(l) \forall l \geq 0$. This claim is true only when the curvature $c_i(l_i^n)$ of q_i is nonnegative. If the curvature obtained by the above procedure is negative, then we set $c_i(l_i^n)$ to zero⁹. When $c_i(l_i^n) = 0$, the q_i function is the line which is tangent to the h_i curve at the current projection value l_i^n .

The curvature of the parabola described above is¹⁰ :

$$c_i(l_i^n) = \begin{cases} \left[2 \frac{h_i(0) - h_i(l_i^n) + \dot{h}_i(l_i^n)(l_i^n)}{(l_i^n)^2} \right]_+, & l_i^n > 0, \\ \left[\ddot{h}_i(0) \right]_+, & l_i^n = 0. \end{cases} \quad (4.33)$$

We prove in Appendix B that this curvature is the optimum curvature that satisfies (4.32). The nonnegativity constraint plays an important role in the proof. If nonnegativity is not enforced, the projections at an iteration may go negative and the curvature (4.33) will not guarantee monotonicity anymore. Fig. 4.3 illustrates this surrogate parabola with the ‘‘optimum curvature’’ (4.33). In Table 4.1, the curvature computation in (4.26) should be changed to (4.33) to implement PSCD method with the optimum curvature (PS,O,CD).

⁹In fact, any nonnegative $c_i(l_i^n)$ will ensure monotonicity, hence the ϵ in (4.27).

¹⁰When l_i^n is nonzero but small, due to numerical precision, (4.33) might turn out to be extremely large during computation. If $c_i(l_i^n) > \left[\ddot{h}_i(0) \right]_+$ (which theoretically should not happen but practically happens due to limited precision), then we set $c_i(l_i^n)$ to be equal to the maximum second derivative $\left[\ddot{h}_i(0) \right]_+$ which eliminates the problem.

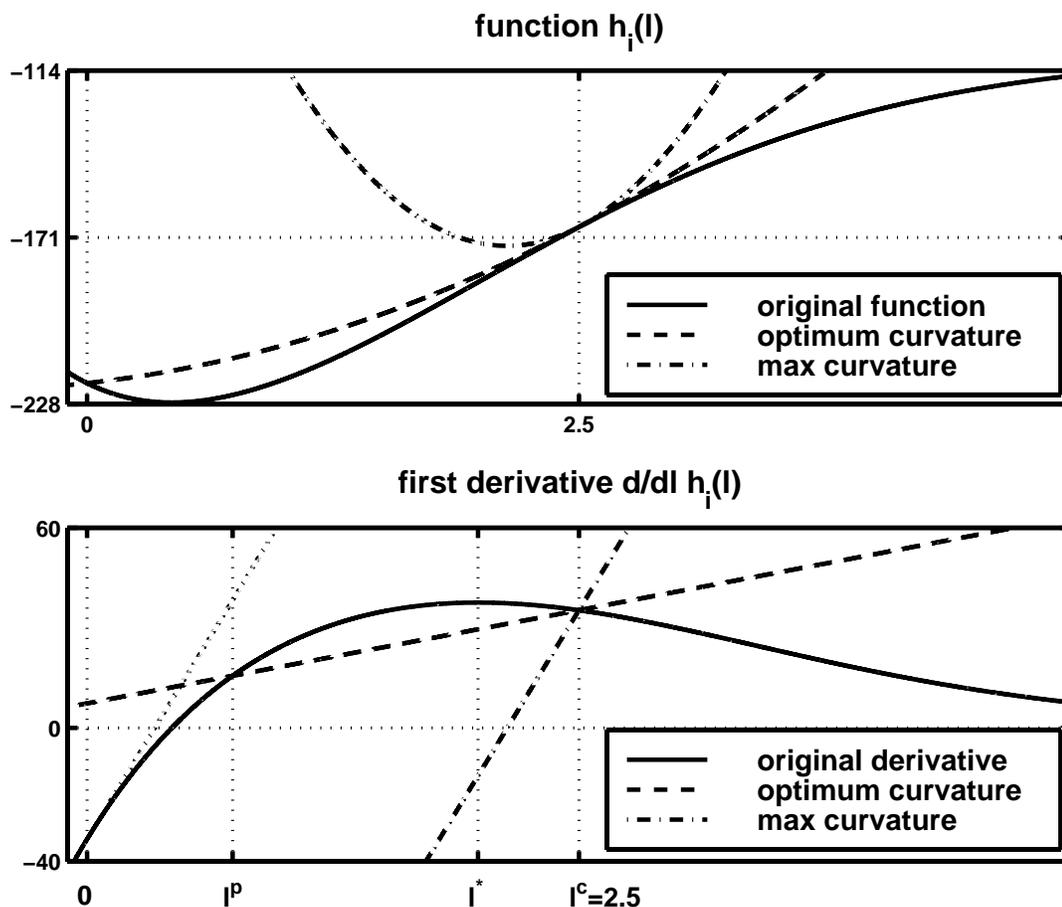


Figure 4.3: This figure illustrates the optimum curvature and the maximum curvature surrogate functions and their derivatives for $b_i = 100, y_i = 70, r_i = 5$, and $l_i^n = 2.5$.

4.3.6 Relationship to Newton-type Optimization Algorithms

A quadratic approximation of any objective function can be obtained by using the first three terms of the Taylor's expansion around a point. An iterative algorithm can be obtained by successively approximating the objective function with a quadratic function at each iteration using first three terms of the Taylor series expansion around the current parameter value and minimizing that quadratic function. Newton's method minimizes this approximating function at each iteration by performing an update similar to (4.16) with the following Newton's (second derivative) curvatures:

$$c_i(l_i^n) = \ddot{h}_i(l_i^n).$$

This method has drawbacks similar to those listed in Section 4.3.2, the major drawback being the computation of the inverse of the Hessian matrix. However, one can use the quadratic approximation obtained by the first three terms of Taylor’s expansion and apply coordinate descent on this quadratic function similar to what we did for the maximum or optimum curvature paraboloids to avoid the computational requirements of (4.16). This algorithm would not be guaranteed to be monotonic, but it would be locally convergent. Our PSCD algorithm is different than the Newton’s method in that we use a curvature that guarantees monotonicity by guaranteeing the quadratic surrogate to lie above the original objective function. Secondly, we do not directly minimize the quadratic objective using matrix inverses similar to (4.16), but we use monotonic coordinate descent to decrease the surrogate function. The parabolic surrogates q_i ’s obtained using the optimum curvature and Newton’s curvature are shown in Figure 4.4. The Newton parabola does not lie above the original function, hence does not guarantee monotonicity.

Bohning *et al.* [4] introduced a monotonic version of Newton’s method which is similar to our maximum curvature paraboloidal surrogates method. They propose to use an update similar to (4.16) as well. We improve on their idea by introducing the optimum curvature which changes at each iteration unlike the maximum curvature. Their method results in narrower quadratic surrogates as compared to our optimum curvature surrogates.

Quasi-Newton methods avoid the matrix inverse in (4.16) by updating an approximation of the Hessian inverse at each iteration, thus reducing the computational requirements. We compared our PSCD methods with a Quasi-Newton algorithm called LBFGS [126] in the results section.

4.3.7 Precomputed Curvature

By relaxing the monotonicity requirement, we can develop faster yet “almost always” monotonic algorithms. We can do this by choosing curvatures $c_i(l_i^n)$ in equation (4.8) such that $\dot{h}_i(l) = \dot{q}_i(l; l_i^n)$, but $h_i(l) \approx q_i(l; l_i^n)$, rather than requiring the

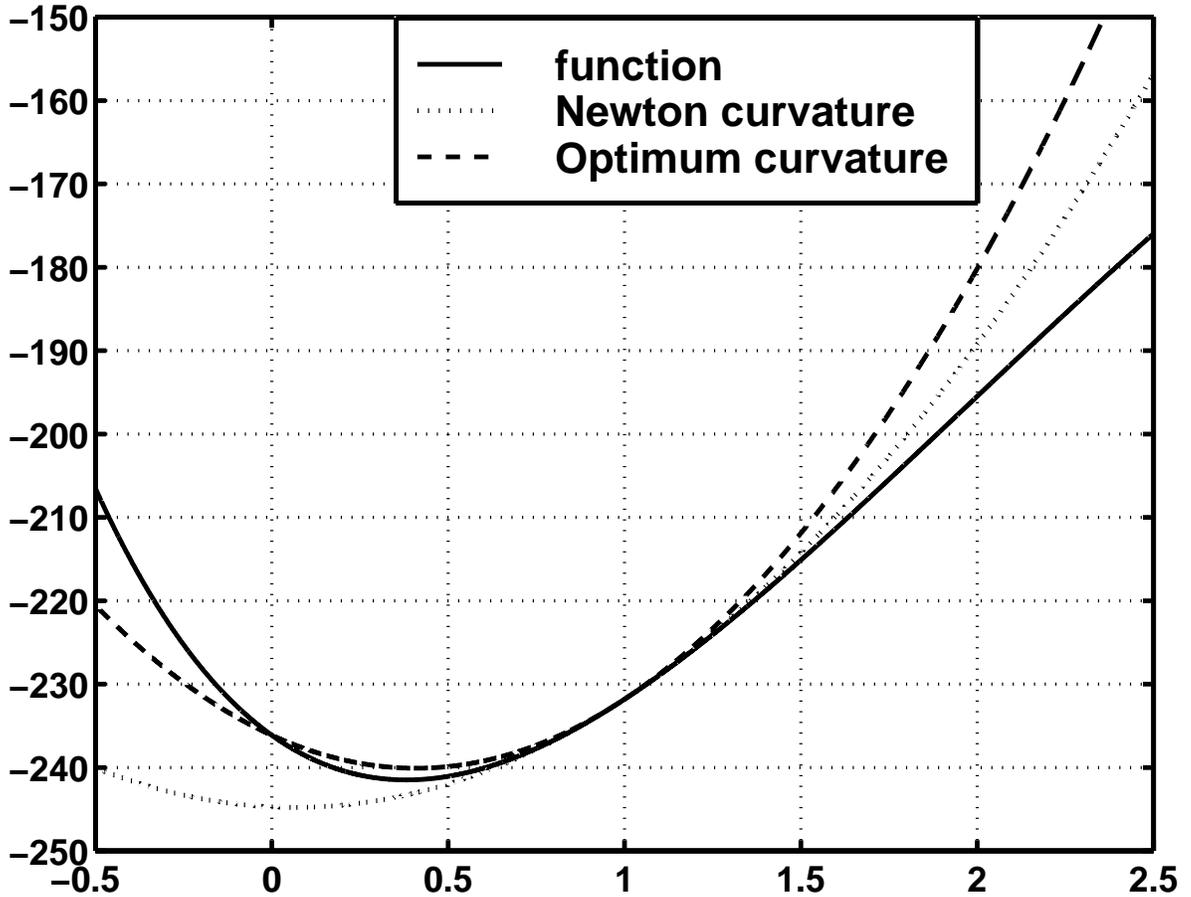


Figure 4.4: This figure illustrates the quadratic surrogates with the optimum curvature and the Newton curvature.

inequality (4.9). In this case, the paraboloids are quadratic “approximations” to the log-likelihood function at each iteration. A reasonable choice for the curvatures is:

$$c_i = \ddot{h}_i \left(\log \frac{b_i}{y_i - r_i} \right) = (y_i - r_i)^2 / y_i. \quad (4.34)$$

The value $l_i^{\min} = \log\left(\frac{b_i}{y_i - r_i}\right)$ is the point that minimizes the h_i function. These curvatures c_i in (4.34) are close approximations to the second derivative of h_i functions at the projection values $\mathbf{G}\hat{\mu}$ where $\hat{\mu}$ is the solution to the penalized-likelihood problem [42]. This is an approximation to the Newton curvatures. This is called the “fast denominator” approach in [42], since it features a one-time precomputed approximation to the curvature that is left unchanged during the iterations so that the denominator terms d_j^n (4.21) can be computed prior to iteration (similar to “maximum curvature”

in equation (4.14)). Computational benefits for iteration independent curvatures as summarized in Section 4.3.3 can be utilized. This approximation works well because we usually start the iterations with an FBP image μ^0 where projections $\mathbf{G}\mu^0$ are usually close to l^{\min} . Nevertheless, unlike with (4.33) monotonicity is not guaranteed with (4.34).

The PS method with the curvature (4.34) yields faster convergence than the other PS algorithms presented above. This method is related to the PWLS image reconstruction method [98, 33], but instead of making a one-time quadratic approximation to the log-likelihood function, the approximation is renewed at each iteration. Although the curvature of the paraboloid remains same, the gradient is changed to match the gradient of the original objective function at the current iterate. The nonnegativity constraint does not play an important role for the derivation, and this curvature may be used for algorithms where nonnegativity is not enforced. We refer to this curvature as “precomputed curvature” (PS,P,CD).

To test the relations between different curvatures, we plotted denominator d_j^m values in a reconstruction for a single pixel in Figure 4.5. This plot shows that maximum and optimum curvatures are more conservative as compared to precomputed denominator. We also compared d_j^m values obtained after 100 iterations of PSCD with optimum curvature and Newton’s curvature (second derivative) with the precomputed denominators in Figures 4.6 and 4.7 by scatter plots. These scatter plots show that precomputed denominators are very close to the Newton’s denominators (after 100th iteration), and that optimum curvature denominators are more conservative than the precomputed denominators to ensure monotonicity.

4.4 Results

To assess the effectiveness and speed of the new PS algorithms, we present results using real PET data. We acquired a 15-hour blank scan (b_i ’s) and a 12-min transmission scan data (y_i ’s) using a Siemens/CTI ECAT EXACT 921 PET scanner with rotating rod transmission sources [119]. The phantom used was an anthropomorphic

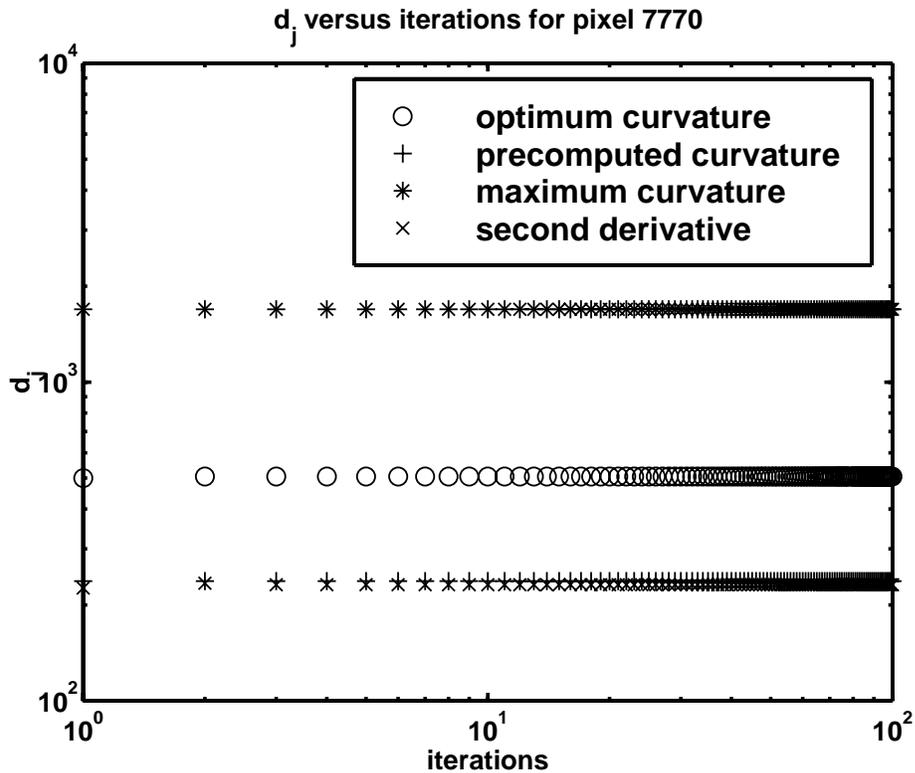


Figure 4.5: Denominators d_j^n versus iterations for different PSCD curvatures.

thorax phantom (Data Spectrum, Chapel Hill, NC). Delayed coincidence sinograms were collected separately in each scan. There were 0.920 million prompt coincidences for the reconstructed slice and the delayed coincidences were about 2.6% of the prompt ones. The blank and transmission scan delayed-coincidence sinograms were shown to be numerically close¹¹ [121], so we used a time-scaled version of blank scan delayed coincidences as the r_i factors with no other processing. The projection space was 160 radial bins and 192 angles, and the reconstructed images were 128×128 with 4.2 mm pixels. The system matrix $\{g_{ij}\}$ was computed by using 3.375 mm wide strip integrals with 3.375 mm spacing, which roughly approximates the system geometry [35].

We performed reconstructions of the phantom by FBP as well as various penalized-likelihood methods. For the penalty term in PL reconstructions, we used the following

¹¹This is due to the fact that singles rate is mostly affected by transmission rods.

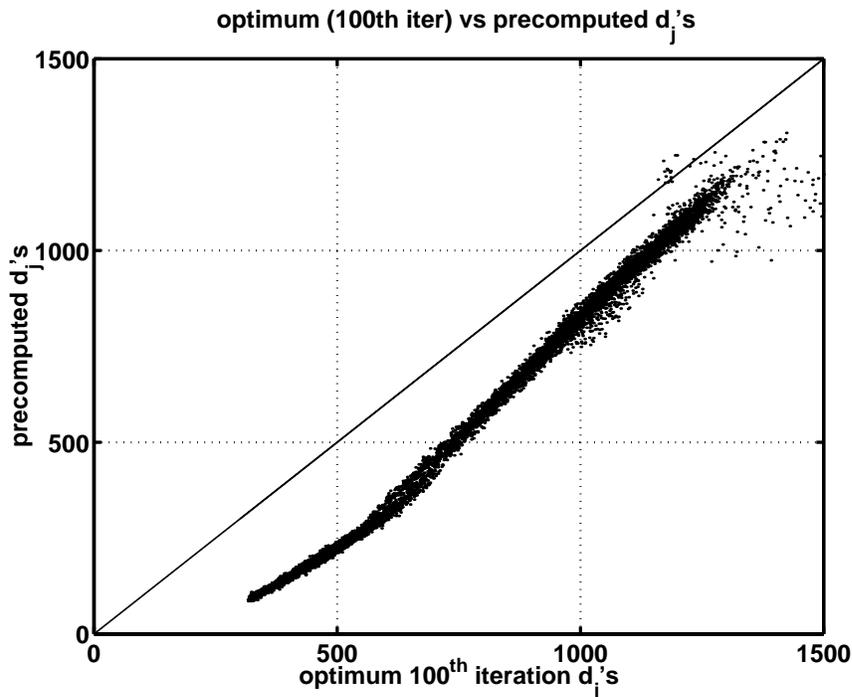


Figure 4.6: Optimum curvature denominators d_j^{100} versus precomputed denominators.

function:

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

which is a special case of (4.5). Here w_{jk} is normally equal to 1 for horizontal and vertical neighbors and $1/\sqrt{2}$ for diagonal neighbors. We used the modified w_{jk} 's described in [46] to achieve more uniform resolution. For the potential function, we used one of the edge-preserving nonquadratic cost functions that was introduced in [68]

$$\psi(x) = \delta^2 [|x/\delta| - \log(1 + |x/\delta|)].$$

This function acts like a quadratic penalty for small differences in neighboring pixels and is close to absolute value function for differences greater than δ . This nonquadratic function penalizes sharp edges less than quadratic functions. We used $\delta = 0.004 \text{ cm}^{-1}$ chosen by visual inspection. In the final reconstructed image, the horizontal and vertical neighbor differences are less than this δ in homogeneous regions (90% of all differences) which makes the curved part of the penalty effective in

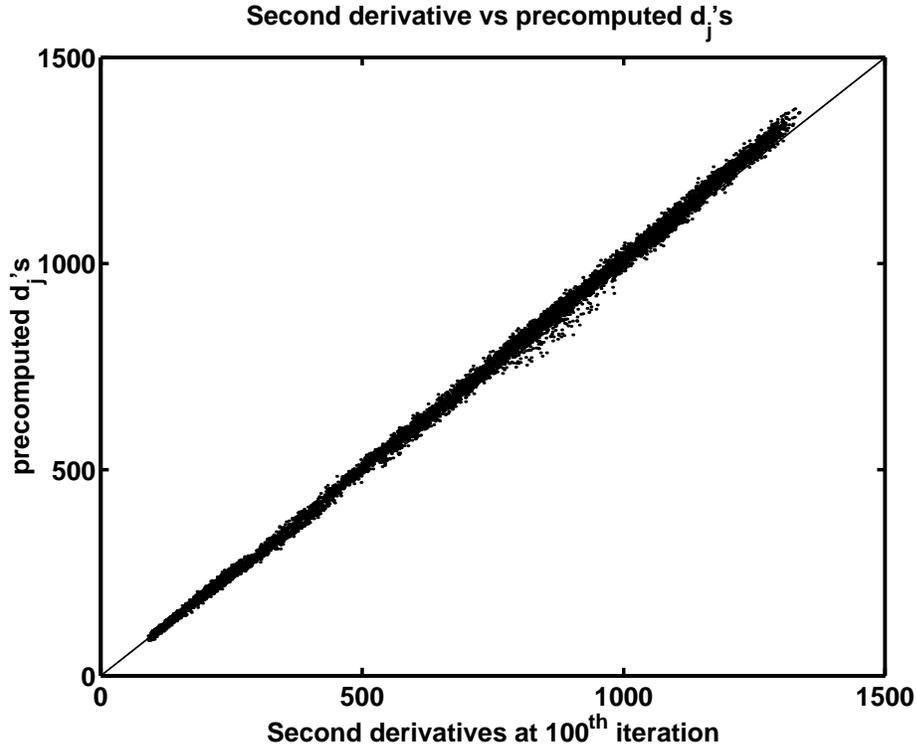


Figure 4.7: Newton’s curvature denominators d_j^{100} versus precomputed denominators.

those regions. However at edges, for which the differences are greater than δ , this penalty penalizes less than the quadratic one.

The PS algorithms described throughout this section are named using the following format: PS, \mathcal{C} ,CD. PS stands for paraboloidal surrogates as the general framework for the algorithms and CD stands for coordinate descent applied to the surrogate function. The letter \mathcal{C} in the format represents the curvature type $c_i(l_i^n)$. The types are: “M”, “O” and “P” for maximum second derivative curvature (4.14), optimum curvature (4.33) and precomputed curvature (4.34) respectively. The other algorithms we used for comparison in this section are as follows. LBFGS: a constrained Quasi-Newton algorithm [126], CD,P: coordinate descent with precomputed denominators and CD,NR: coordinate descent with Newton-Raphson denominators [98, 42] applied to objective function, GD,P: grouped descent with precomputed denominators [42].

Fig. 4.8 shows images reconstructed by FBP and statistical methods from a 12 minute scan. For comparison, an FBP reconstruction of a 7 hour scan is also shown.

For FBP images, we used 2D Gaussian sinogram smoothing with $(3 \text{ mm}, \pi/192)$ FWHM followed by a ramp filter to reconstruct the 7 hour scan, and $(12 \text{ mm}, \pi/48)$ FWHM for the 12 minute scan. Qualitatively, the statistical reconstruction looks better than the FBP image, having less noise and more uniform homogeneous regions. However, our focus here is not the image quality but the amount of time it takes the algorithms to converge to the minimizer image. Nevertheless, improved emission image quality is our ultimate goal. Statistical methods for transmission reconstruction yield better ACFs as compared to conventional methods and result in better emission images. Our goal here is to speed-up and stabilize statistical methods to make them usable routinely in clinic.

Fig. 4.9 shows that the proposed PSCD algorithms decreased Φ almost as much per iteration as the coordinate descent algorithm applied to Φ directly. This result is important because it shows that the surrogate paraboloids (especially with the optimum curvature) closely approximate the original log-likelihood. More importantly, in Fig. 4.10 the PSCD algorithms are seen to be much faster than coordinate descent in terms of the actual CPU time¹². One of the main overhead costs in coordinate descent is the computation of the log-likelihood gradient term after each pixel change [42]. In PSCD algorithm, the gradient of the surrogate function (\hat{q}_i 's) can be computed (updated) by a single multiplication (4.19). The “maximum curvature” method introduced in Section 4.3.1 precomputes the denominator terms (d_j^n) for the likelihood part since $c_i(l_i^n)$'s do not depend on the iterations. However, these $c_i(l_i^n)$'s are much larger than the optimal curvatures, so more iterations are required for PS,M,CD than PS,O,CD to converge.

We also compared the PSCD algorithms to the general purpose constrained Quasi-Newton algorithm (LBFGS) [126] in Figures 4.9 and 4.10. Although the LBFGS algorithm takes about 25% less CPU time (0.88 seconds) per iteration than PSCD algorithms, it did not converge as fast as the proposed algorithms. This shows that the algorithms such as PSCD which are tailored to our specific problem converge

¹²All CPU times are recorded on a DEC 600 5-333 MHz workstation with compiler optimization enabled.

faster than the general purpose Quasi-Newton method.

In Fig. 4.11, we consider the fastest previous algorithm we know of (*i.e.* GD with 3×3 groups with precomputed denominator [42]) and compare it to the fastest PS algorithms. The PSCD with “precomputed curvatures” (PS,P,CD) (introduced in Section 4.3.7) requires slightly less CPU time than GD,P to converge. Although the PS,P,CD algorithm is not provably monotonic, it is a reasonable approximation and we did not observe any non-monotonicity in our practical experience when initializing with an FBP image. The monotonic PS,O,CD method is shown in this plot as a baseline for comparison with Fig. 4.10.

In Figures 4.12 and 4.13, we present the results of a transmission scan simulation with zero background counts ($r_i = 0$) and compare the monotonic PSCD algorithm with the functional substitution (FS) method of Zheng *et al.*[125, 97]. The FS algorithm is proven to be monotonic when $r_i = 0$ in which case h_i is convex. However, the FSCD method requires considerably more computation per iteration than both CD and PSCD. The plot in Figure 4.13 shows that FSCD requires more CPU time than PSCD.

Table 4.2 compares the number of iterations and CPU seconds required to minimize the objective function by each method. The CPU times¹³, floating point operations and memory accesses (of order M only) per iteration are also tabulated, where M is the number of nonzero entries in system matrix \mathbf{G} . For comparison purposes, a single forward and backprojection requires about 0.78 CPU seconds. The CD and FS methods are significantly different from our proposed PSCD methods in the following respect. In our methods, the \dot{q}_i terms are kept updated for all i outside the projection loop in (4.29). In contrast, both CD and FS require \dot{h}_i terms within the backprojection loop, and these change with every pixel update so they must be computed on the fly within the backprojection loop. Thus that backprojection must access y_i, b_i, r_i, \hat{l}_i and the system matrix within the loop, and perform quite a few floating point operations (including the exponentiations) with them. Not only is there inher-

¹³The CPU times are computed on a DEC 600 5-333 MHz. We also compiled the code on a SUN Ultra 2 computer and got similar CPU time ratios for the algorithms. However, the ratios could differ on another architecture or with another compiler due to cache size and pipelining differences.

Real data, $r_i \neq 0$	monotonic		nonmonotonic				
	PS,M,CD	PS,O,CD	PS,P,CD	GD,P,3x3	CD,P	CD,NR	FSCD
methods							
iters for convergence	18	12	11	14	11	11	11
CPU s for convergence	23.3	17.4	15.1	18.1	44.3	52.3	56.2
CPU s per iteration	1.2	1.3	1.2	1.1	3.8	4.6	4.9
exponentiations per iteration	0	0	0	0	M	M	2M
add/subts per iteration	2M	3M	2M	2M	4M	6M	7M
mult/divs per iteration	3M	5M	3M	2M	5M	11M	10M
nonsequential accesses per backprojection	M	2M	M	M	4M	4M	4M
nonsequential accesses per forward projection	2M	2M	2M	M	M	M	M
system matrix accesses per iteration	2M	2M	2M	2M	2M	2M	2M

Table 4.2: Comparison of CPU times, number of iterations to converge, floating point operations and memory accesses for the PS algorithms versus CD, GD and FS methods. Convergence in this table means $\Phi(\mu^0) - \Phi(\mu^n) > 0.999 [\Phi(\mu^0) - \Phi(\mu^*)]$ where $\Phi(\mu^*)$ is the smallest objective value obtained in 30 iterations among all the methods. The floating point operations and memory accesses only in the order of M are shown for each method.

ently more floating point operations required for CD and FS, we suspect that the need to nonsequentially access parts of four sinogram-sized arrays, in addition to the system matrix, significantly degrades the ability of the CPU to pipeline operations. This leads to the dramatic differences in CPU time between PSCD and CD methods.

If a monotonic algorithm is required, the PSCD algorithm with the optimal curvature (PS,O,CD) is the fastest algorithm. The other algorithms are not guaranteed to be monotonic except PSCD with maximum curvature. Although PS,M,CD algorithm consumes less CPU time per iteration, it takes longer to converge since the curvatures result in an unnecessarily narrow surrogate function which causes small step sizes.

Among the nonmonotonic algorithms, another PS method, PSCD with precomputed curvatures (PS,P,CD) is the fastest. It converged in about 15 seconds with the real data used. The CPU time per iteration is the same as PS,M,CD since they both precompute the denominator (d_j^m) terms. Since the curvatures are smaller, this method decreases the objective very rapidly, nevertheless it is not guaranteed to be monotonic. However, as with the CD and GD with precomputed denominators [42], we have never observed any nonmonotonicity in practical applications with iterations started with an FBP image. The FSCD and CD algorithms consume a lot of CPU cycles per iteration and they are much slower than the proposed algorithms. The GD,P algorithm lowers the CPU requirements by decreasing the number of exponentiations, but it does not decrease the objective function as much per iteration as coordinate descent. Thus, it is also slightly slower than the PS,P,CD algorithm. This Table shows that PSCD algorithms are preferable for both monotonic and nonmonotonic transmission image reconstructions.

4.5 Conclusion

We have introduced a new class of algorithms for minimizing penalized-likelihood objective functions for transmission tomography. The algorithms are shown to be monotonic even with the nonconvex objective function. In the nonconvex case, there is no proof that these algorithms will find the global minimum but at least the algorithms

will monotonically decrease the objective function towards a local minimum. Practical experience suggests there are rarely multiple minima in this problem, but there is no proof. In the strictly convex case, the proposed algorithms are guaranteed to converge to the global minimum by a proof similar to that in [44].

Convergence is very important for algorithms for any optimization problem, particularly in medical applications. The PSCD algorithm is globally convergent when there are no background counts. Even when there are background counts, the new algorithm is guaranteed to monotonically decrease the objective function making the algorithm stable. Previous algorithms could not guarantee that property without expensive line searches. The robustness, stability and speed of the new algorithm renders it usable in routine clinical studies. Such use should increase the emission image quality as compared to conventional methods which use linear processing and FBP for reconstruction. Further “acceleration” is possible by ordered subsets [32], albeit without guaranteed monotonicity.

The algorithms we introduced are simple, easy to understand and fast. The simplicity in part is due to the additive form of (4.2), which is a direct consequence of independent measurements. Since the emission tomography log-likelihood has a very similar form due to independence of measurements, it is possible to apply the paraboloidal surrogates idea to the emission case as well to get faster algorithms [41]. Since the emission problem is convex regardless of the background events, PSCD will be a globally convergent algorithm for the emission problem.

It is possible to parallelize the PS algorithms by applying either grouped descent (GD) [42, 99] algorithm to the surrogate function, or by parallelizing the projection and backprojection operators [86] for each pixel. However, in a serial computer we found that PS method with GD update (PSGD) was not faster than the PSCD algorithm. This is due to the fact that the gradient updates in PSCD algorithm consume much less CPU time than the gradient evaluations in the original CD algorithm which require expensive exponentiations and floating point operations. Hence, grouped descent did not reduce the CPU time per iteration as much in PS method as in the direct method.

In our opinion, the PS,O,CD algorithm supersedes all of our previous methods [35, 71, 42], and is our recommended algorithm for penalized-likelihood transmission tomography. The PS,P,CD algorithm is a faster but nonmonotonic alternative which can be used for noncritical applications. A possible compromise would be to run a few iterations of PS,O,CD algorithm and then fix the curvatures and denominator terms (d_j^n) for the rest of the iterations to save computation time. Alternatively, one can run PS,P,CD algorithm and check the objective function $\Phi(\mu)$ after each iteration to verify that it has decreased. If the objective does not decrease (happens very rarely), then PS,O,CD algorithm can be applied to the previous iterate to ensure monotonicity. For medical purposes, we believe that a monotonic algorithm should be used to reduce the risk of diagnostic errors due to erroneous reconstructions. Fortunately, with the new proposed methods, monotonicity can be assured with only a minor increase in CPU time (17.2 versus 15.1 CPU seconds).

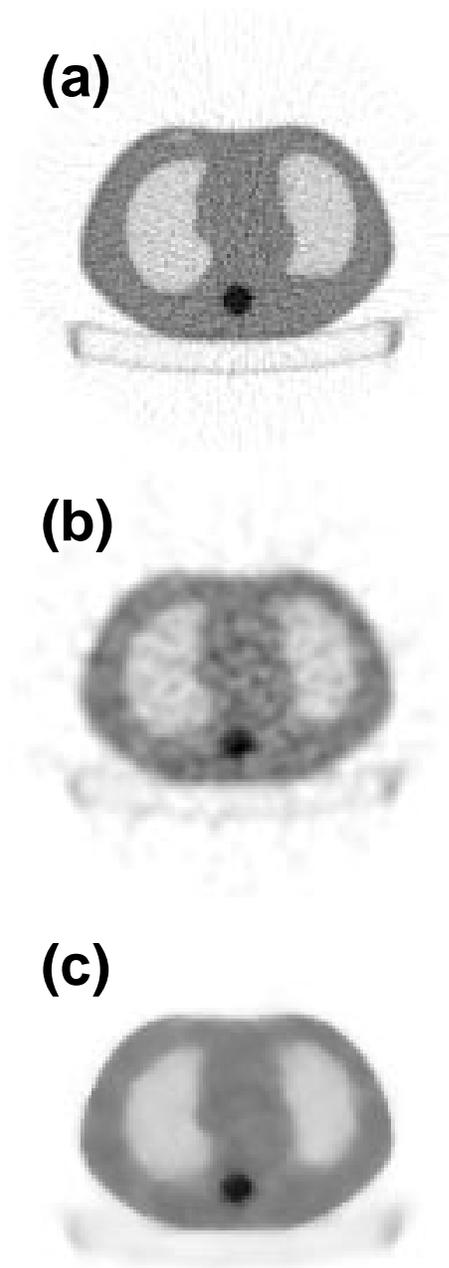


Figure 4.8: (a) FBP reconstruction of phantom data from 7-h transmission scan, (b) FBP reconstruction from 12-min transmission scan, and (c) Penalized-likelihood reconstruction from 12-min transmission scan using 12 iterations of the “optimum curvature” PSCD algorithm.

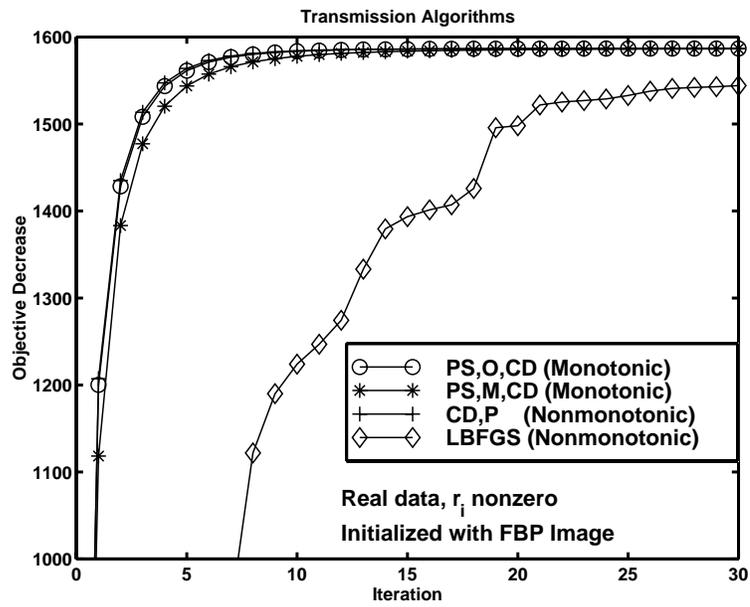


Figure 4.9: Comparison of objective function decrease $\Phi(\mu^0) - \Phi(\mu^n)$ versus iteration number n of monotonic PS methods with coordinate descent and LBFGS methods for real phantom data.

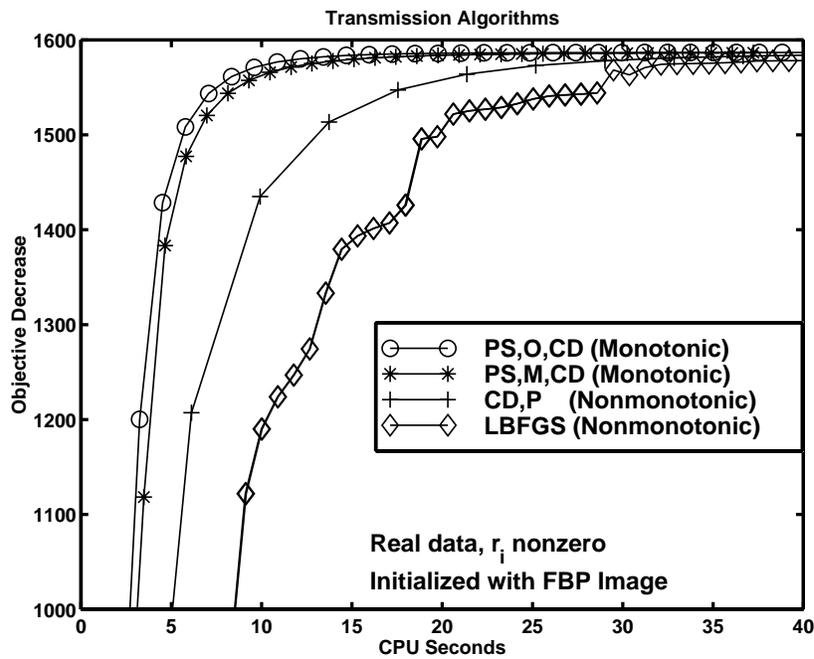


Figure 4.10: Same as Figure 4.9, but x-axis is CPU seconds on a DEC AlphaStation 600 5-333 MHz.

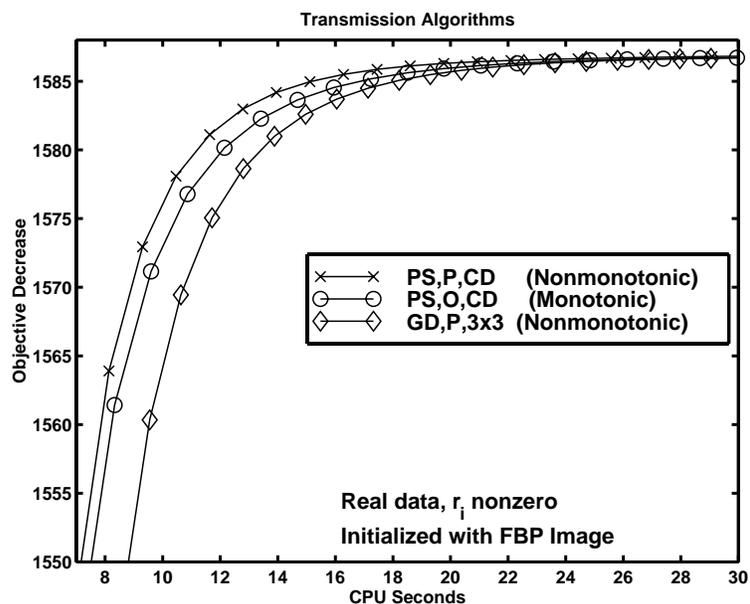


Figure 4.11: Comparison of the speed of the proposed PS algorithms with the fastest algorithm that was introduced before: grouped coordinate descent with 3x3 groups.

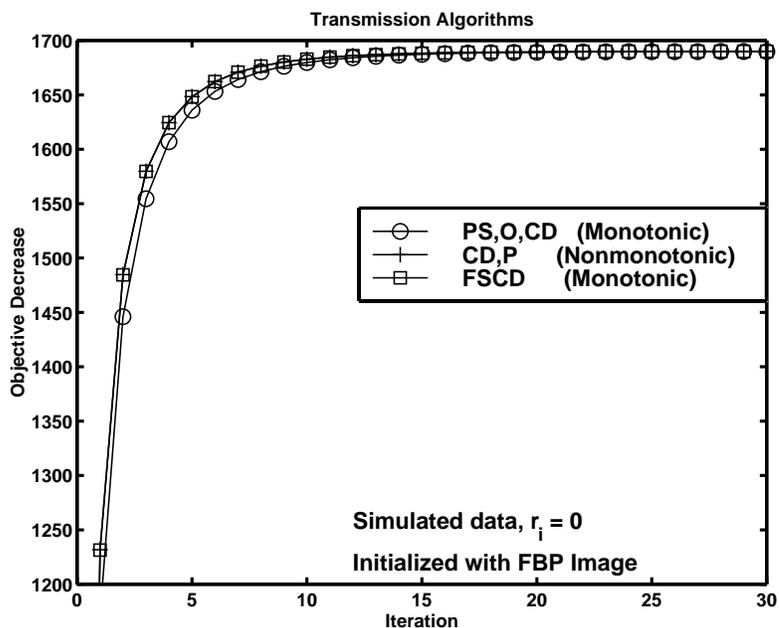


Figure 4.12: Comparison of objective function decrease $\Phi(\mu^0) - \Phi(\mu^n)$ versus iterations of monotonic PS and FS methods with coordinate descent. Note $r_i = 0$ in this simulation.

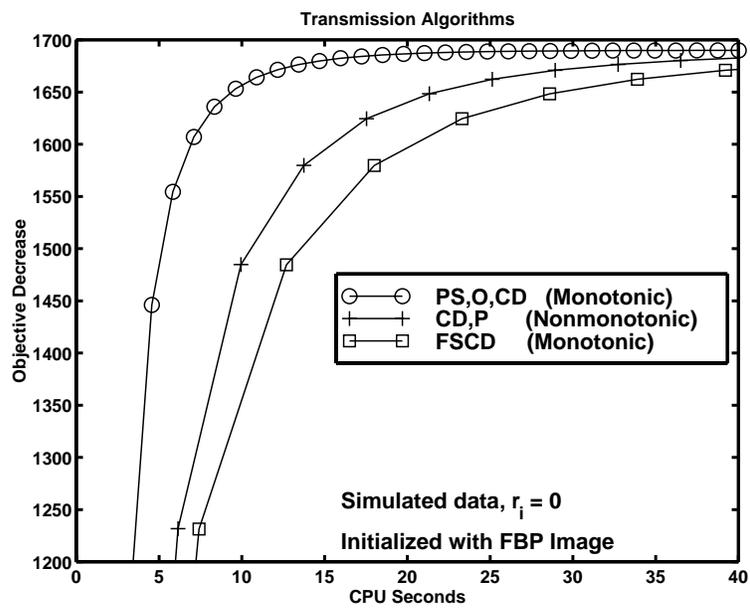


Figure 4.13: Comparison of objective function decrease $\Phi(\mu^0) - \Phi(\mu^n)$ versus CPU time of monotonic PS and FS methods with coordinate descent. Note $r_i = 0$ in this simulation.

CHAPTER 5

Paraboloidal Surrogates Ordered Subsets Algorithms for Transmission Tomography

5.1 Introduction

¹Attenuation is an important factor that should be corrected for in emission computed tomography. In modern PET and SPECT systems, transmission scans are performed in addition to emission scans to correct for the effects of attenuation. Statistical methods can be used to reconstruct attenuation maps, from which one can calculate attenuation correction factors (ACFs) to yield quantitatively accurate emission images.

Many algorithms exist for maximum likelihood (ML) and penalized likelihood (PL) transmission image reconstruction problems. Most of the recent ones [98, 42] are based on direct maximization of the objective function rather than on the famous expectation maximization (EM) algorithm [70] due to the fact that the EM algorithm for transmission reconstruction converges very slowly [89].

Recently, ordered subsets EM (OSEM) [60] for the emission problem has been used extensively in emission image reconstruction, primarily because of the following reasons.

- OSEM provides order-of-magnitude acceleration over EM in ML problems.

¹This chapter is based on [31].

- The reconstructed images look good after only a few iterations.
- OSEM is implemented by slightly modifying the well-known EM algorithm.
- OSEM is easily implemented with any type of system model.

Although the images seem to look good, the resolution and variance properties of OSEM are unclear. In addition it does not converge and may cycle. Due to its popularity, OSEM has even been applied to transmission data after taking its logarithm. In the results section, we show that this approach yields lower quality images than the ordered subsets transmission (OSTR) algorithm that we introduce in this chapter.

The ordered subsets principle can be applied to any algorithm which involves a sum over sinogram indices. The sums over all the sinogram indices are replaced by sums over a subset of the data and an ordered subsets version of the algorithm is obtained. However, it seems best to apply this idea to algorithms which update the parameters simultaneously at each iteration rather than to sequential update algorithms. Simultaneous update algorithms take smaller steps in the update direction than sequential update algorithms due to the requirement of a separable surrogate function which has higher curvature than a nonseparable one. Sequential update algorithms such as coordinate descent tend to update high frequencies faster [98]. When only a subset of the data is used, as in ordered subsets, there is no point in making high frequency details converge. For the algorithms that use only a portion of the data at each iteration such as ART, underrelaxation along the update direction helps the algorithm to converge [8].

We introduce a new simultaneous update algorithm called separable paraboloidal surrogates (SPS) algorithm in this chapter. A paraboloidal surrogate [30] is a quadratic function that is designed to lie above the negative log-likelihood. Using convexity [42], we get a separable quadratic function that lies above this paraboloid. Another separable surrogate can be obtained for the penalty part by using De Pierro's methods [23, 24]. The global separable surrogate function can be minimized by a simple simultaneous update.

The SPS algorithm has three advantages as compared to previous simultaneous update algorithms such as transmission EM algorithm [70] and Lange’s Convex algorithm [71] : 1) It requires fewer flops per iteration than the transmission EM algorithm and is comparable to the Convex algorithm, 2) SPS is derived for the PL problem which is a more general form than the ML problem, 3) SPS is guaranteed to be monotonic, even with nonzero background events.

The ordered subsets principle has been applied to other transmission ML algorithms. Manglos *et al.* [77] applied the ordered subsets idea to the transmission EM method for ML problem. Although ordered subsets accelerates the original transmission EM algorithm, it still converges slowly. Nuyts *et al.* [87] tested an ordered subsets version of an approximate simultaneous update algorithm they developed. Their algorithm disregards background counts (such as random coincidences in PET) and the convergence properties are unknown. Kamphuis and Beekman [64] applied the ordered subsets principle to Lange’s Convex algorithm to accelerate ML transmission image reconstruction, also ignoring the background counts.

In this work, we apply the ordered subsets principle to the SPS algorithm for both ML and PL transmission tomography problems. We show that ordered subsets accelerates the initial speed of the original SPS algorithm. However, OSTR is not guaranteed to be monotonic and does not converge to the true optimum for number of subsets greater than one. Browne and Depierro [8] developed a new algorithm called RAMLA which is similar to OSEM with a relaxation parameter incorporated to the algorithm. For a certain class of relaxation parameters, they prove that RAMLA converges to the true ML solution for emission tomography. It might be possible to obtain a convergent version of OSTR by similar means. However, our results show that, even without relaxation, the PL images reconstructed with OSTR are very similar to the ones obtained by convergent algorithms.

In the rest of the chapter, we first introduce the problem and the OSTR algorithm for general penalized-likelihood (PL) objective. Then, we present results on real PET transmission data with ML and PL reconstructions. We analyze the algorithms in terms of their mean squared error. We also perform hard segmentation on the

reconstructed images to analyze their tissue classification performance.

5.2 The Problem

For transmission scans, it is realistic to assume the following statistical model if the raw (prompt) measurements $\{y_i\}$ are available:

$$y_i \sim \text{Poisson}\{b_i e^{-[\mathbf{G}\mu]_i} + r_i\}, \quad i = 1, \dots, N, \quad (5.1)$$

where N is the number of measured rays, μ_j is the average linear attenuation coefficient in voxel j for $j = 1, \dots, p$, and p denotes the number of voxels. The notation $[\mathbf{G}\mu]_i = \sum_{j=1}^p g_{ij}\mu_j$ represents the line integral of the attenuation map μ , and $\mathbf{G} = \{g_{ij}\}$ is the $N \times p$ system matrix. We assume that $\{b_i\}, \{r_i\}$ and $\{g_{ij}\}$ are known nonnegative constants, where r_i is the mean number of background events, b_i is the blank scan count and y_i represents the number of coincident transmission events counted by the i th detector pair. Although we adopt PET terminology throughout, the algorithm is also applicable to SPECT and X-ray CT.

For most PET systems, the delayed coincidences are pre-subtracted from true (prompt) coincidences by the device hardware in an attempt to remove the background counts. The subtracted data is no longer Poisson [33, 121], but a difference of two Poisson random variables:

$$y_i^s \sim \text{Poisson}\{b_i e^{-[\mathbf{G}\mu]_i} + r_i\} - \text{Poisson}\{r_i\}. \quad (5.2)$$

In this case, the model (5.1) is not accurate. Yavuz and Fessler [121] showed that an accurate model is achieved by adding a sinogram which is a good estimate of twice the mean background events (r_i 's) to the subtracted data and assuming that this resultant random vector has the distribution:

$$y_i^s + 2r_i \sim \text{Poisson}\{b_i e^{-[\mathbf{G}\mu]_i} + 2r_i\}, \quad i = 1, \dots, N.$$

This ‘‘Shifted Poisson’’ model matches the mean and variance of the data and yields more accurate images than the PWLS method [121]. For the rest of the chapter, we

focus on the model (5.1). However extension to the Shifted Poisson model can easily be done by replacing y_i by $y_i^s + 2r_i$ and r_i by $2r_i$.

The negative log-likelihood function for the independent transmission data is:

$$-L(\mu) = \sum_{i=1}^N h_i([\mathbf{G}\mu]_i), \quad (5.3)$$

where $h_i(l) = b_i e^{-l} + r_i - y_i \log(b_i e^{-l} + r_i)$, ignoring the constant terms. Directly minimizing $-L(\mu)$ (ML method) results in a very noisy estimate $\hat{\mu}$. Segmentation of the attenuation map is commonly performed to reduce noise afterwards. Penalized-likelihood (PL) (or MAP) methods regularize the problem and reduce the noise by adding a roughness penalty to the objective function as follows:

$$\hat{\mu} = \arg \min_{\mu \geq 0} \Phi(\mu), \quad \Phi(\mu) = -L(\mu) + \beta R(\mu).$$

For simplicity we focus here on a roughness penalty R of this form:

$$R(\mu) = \frac{1}{2} \sum_{j=1}^p \sum_{k \in \mathcal{N}_j} w_{jk} \psi(\mu_j - \mu_k),$$

where \mathcal{N}_j represents a neighborhood of voxel j , ψ is a symmetric and convex function that penalizes neighboring pixel differences. The method easily generalizes to other forms of penalty functions.

In the following discussion, we use the PL formulation to derive the new algorithm. Setting $\beta = 0$ in the following discussion yields the ML estimator.

5.3 The SPS Algorithm

In this section, we describe a new simultaneous update algorithm called separable paraboloidal surrogates (SPS) algorithm.

5.3.1 The Likelihood Part

Nonseparable Paraboloidal Surrogate

We presented the paraboloidal surrogates algorithm for transmission tomography previously [30, 29]. We first find a one-dimensional surrogate parabola $q_i(l; l_i^n)$ that

is tangent to the marginal negative log-likelihood function $h_i(l)$ at the current iterate $l_i^n = [\mathbf{G}\mu^n]_i$ and lies above it for all $l > 0$. Then, we sum up these parabolas like (5.3) to obtain an overall (nonseparable) paraboloidal surrogate function for the negative log-likelihood as follows:

$$Q_1(\mu; \mu^n) \triangleq \sum_{i=1}^N q_i([\mathbf{G}\mu]_i; l_i^n) \geq -L(\mu), \quad \forall \mu \geq 0,$$

where

$$q_i(l; l_i^n) \triangleq h_i(l_i^n) + \dot{h}_i(l_i^n)(l - l_i^n) + \frac{1}{2}c_i(l_i^n)(l - l_i^n)^2.$$

The optimum curvature that provides the fastest convergence rate while preserving monotonicity was shown to be [30]

$$\begin{aligned} c_i(l_i^n) &= \begin{cases} \left[2 \frac{h_i(0) - h_i(l_i^n) + \dot{h}_i(l_i^n)(l_i^n)}{(l_i^n)^2} \right]_+, & l_i^n > 0 \\ \left[\ddot{h}_i(0) \right]_+, & l_i^n = 0 \end{cases}, \\ &= \begin{cases} \left[\frac{2}{(l_i^n)^2} \left\{ b_i(1 - e^{-l_i^n}) - y_i \log \frac{b_i + r_i}{\bar{y}_i^n} + l_i^n b_i e^{-l_i^n} \left(\frac{y_i}{\bar{y}_i^n} - 1 \right) \right\} \right]_+, & l_i^n > 0 \\ \left[b_i \left(1 - \frac{y_i r_i}{(b_i + r_i)^2} \right) \right]_+, & l_i^n = 0 \end{cases} \end{aligned} \quad (5.4)$$

where $\bar{y}_i^n = b_i e^{-l_i^n} + r_i$. This surrogate function $Q_1(\mu; \mu^n)$ and each $q_i(l; l_i^n)$ are naturally convex. Previously, we used coordinate descent to minimize this function [30]. That approach leads to a very fast and monotonic algorithm. However, the computational advantages only exist if the system matrix is precomputed and column accessible [34]. For implementations in which the system matrix is not precomputed (*e.g.* software that uses projector/backprojector subroutines which compute the g_{ij} on the fly), algorithms that update all the parameters simultaneously are preferable since column access to the system matrix is not needed. Moreover, simultaneous update algorithms parallelize more readily than sequential updates. A simultaneous update algorithm can be obtained from the above paraboloidal surrogate by finding another separable surrogate function that lies above it for all possible feasible parameter values μ . The additive separability of the surrogate will result in decoupling of the optimization problem for each parameter and each parameter μ_j can be updated

independently from the others. To obtain this separable function we use the convexity tricks employed in [23, 24].

Separable Surrogate

Lange [71] applied De Pierro's ideas [23, 24] to transmission tomography to get a separable function that is tangent to the negative log-likelihood and lies above it everywhere when it is convex. It can be based on rewriting the sum

$$[\mathbf{G}\mu]_i = \sum_{j=1}^p g_{ij}\mu_j = \sum_{j=1}^p \alpha_{ij} \left(\frac{g_{ij}}{\alpha_{ij}} (\mu_j - \mu_j^n) + [\mathbf{G}\mu^n]_i \right), \quad (5.5)$$

where

$$\sum_{j=1}^p \alpha_{ij} = 1, \quad \forall i \quad \text{and} \quad \alpha_{ij} \geq 0. \quad (5.6)$$

Using the convexity of q_i , we can show that:

$$q_i([\mathbf{G}\mu]_i; l_i^n) \leq \sum_{j=1}^p \alpha_{ij} q_i \left(\frac{g_{ij}}{\alpha_{ij}} (\mu_j - \mu_j^n) + [\mathbf{G}\mu^n]_i; l_i^n \right). \quad (5.7)$$

The form of the right hand side of (5.7) ensures that the function value and gradient of the left hand side are equal to those of the right hand side at the current iterate $\mu = \mu^n$. In other words the curves are tangent at the current iterate. One possible choice for α_{ij} that has been used in [23, 71] is:

$$\alpha_{ij} = \frac{g_{ij}\mu_j^n}{[\mathbf{G}\mu^n]_i}. \quad (5.8)$$

We call this choice of α_{ij} 's as "multiplicative" form. Using the inequality (5.7) with these α_{ij} 's, we get a separable surrogate function for $Q_1(\mu; \mu^n)$ as follows:

$$Q_2^M(\mu; \mu^n) = \sum_{i=1}^N \sum_{j=1}^p \alpha_{ij} q_i \left(\frac{[\mathbf{G}\mu^n]_i \mu_j}{\mu_j^n}; l_i^n \right).$$

This is the separable surrogate obtained using the "multiplicative" form (5.8), hence we use M in the superscript.

Another possible set of α_{ij} 's is given in [42]:

$$\alpha_{ij} = \frac{g_{ij}}{\gamma_i},$$

where $\gamma_i = \sum_{k=1}^p g_{ik}$ is the projection of an image of all ones. We call this choice the “additive” form, which results in a separable surrogate as follows:

$$Q_2(\mu; \mu^n) = \sum_{i=1}^N \sum_{j=1}^p \frac{g_{ij}}{\gamma_i} q_i \left(\gamma_i (\mu_j - \mu_j^n) + [\mathbf{G}\mu^n]_i; l_i^n \right).$$

The function $Q_2(\mu; \mu^n)$ is separable in j and quadratic, so that the exact minimization is reduced to minimization of p 1D parabolas each of which depend on one pixel value μ_j only.

The separable surrogate obtained from the multiplicative form has some problems with convergence speed due to the fact that the curvature is inversely proportional to the current iterate value μ_j^n :

$$\frac{\partial^2}{\partial^2 \mu_j} Q_2^M(\mu_j^n; \mu_j^n) = \frac{1}{\mu_j^n} \sum_{i=1}^N g_{ij} [\mathbf{G}\mu^n]_i c_i(l_i^n).$$

The surrogate parabola becomes infinitely thinner when μ_j^n gets close to zero and slows down the convergence for zero regions in the image. The convergence rates of the two algorithms based on multiplicative and additive forms (PL problem with optimum curvature (5.4)) are shown in Figure 5.1. This figure reveals that the additive form yields a faster algorithm than the multiplicative form does. Hence, we focus on the additive form for the rest of the chapter.

5.3.2 The Penalty Part

Section 5.3.1 derived separable surrogate functions for the log-likelihood function. A similar separable surrogate is needed for the penalty part $R(\mu)$ to obtain a simultaneous update for the PL objective function. We exploit the convexity of the potential function $\psi(t)$ to obtain the surrogate. For completeness, we repeat the arguments in [24, 71]:

$$\begin{aligned} \psi(\mu_j - \mu_k) &= \psi \left(\frac{1}{2} [2\mu_j - \mu_j^n - \mu_k^n] + \frac{1}{2} [-2\mu_k + \mu_j^n + \mu_k^n] \right) \\ &\leq \hat{\psi}_{jk}(\mu; \mu^n) \triangleq \frac{1}{2} \psi(2\mu_j - \mu_j^n - \mu_k^n) + \frac{1}{2} \psi(2\mu_k - \mu_j^n - \mu_k^n). \end{aligned} \quad (5.9)$$

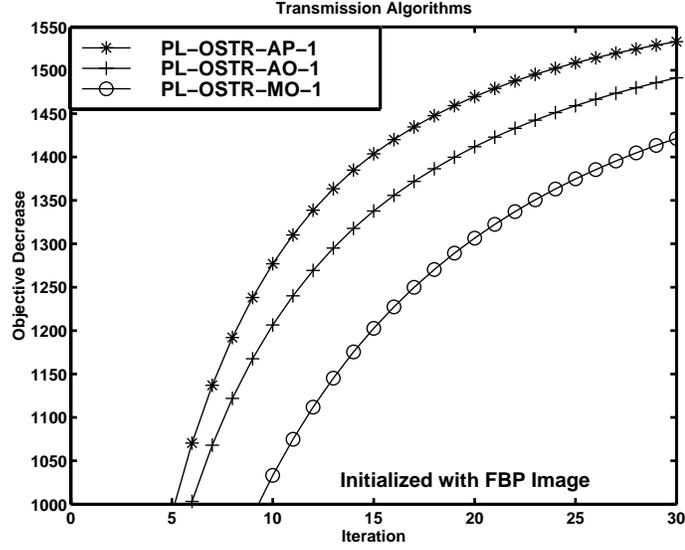


Figure 5.1: Comparison of additive form with optimum curvature (AO), with pre-computed curvature (AP) and multiplicative form with optimum curvature (MO) SPS algorithms for PL image reconstruction.

Using this inequality, one gets the following separable surrogate function for the penalty:

$$S(\mu; \mu^n) \triangleq \frac{1}{2} \sum_{j=1}^p \sum_{k \in \mathcal{N}_j} w_{jk} \hat{\psi}_{jk}(\mu; \mu^n) \geq R(\mu), \quad \forall \mu \in \mathbb{R}. \quad (5.10)$$

One can verify that this surrogate function is tangent to $R(\mu)$ at the current iterate and lies above it for all μ values. Furthermore, the curvature of the surrogate at the current iterate μ^n is exactly twice the curvature of the original penalty function.

5.3.3 The SPS Algorithm

We designed separable surrogate functions for both the likelihood and the penalty parts in the preceding sections. By combining those, we define the global surrogate function

$$\phi(\mu; \mu^n) \triangleq Q_2(\mu; \mu^n) + \beta S(\mu; \mu^n),$$

which satisfies $\phi(\mu; \mu^n) \geq -L(\mu) + \beta R(\mu) = \Phi(\mu)$, $\forall \mu \geq 0$, and is tangent to $\Phi(\mu)$ at current iterate μ^n , *i.e.*

$$\Phi(\mu^n) = \phi(\mu^n; \mu^n), \quad \nabla \Phi(\mu^n) = \nabla \phi(\mu^n).$$

We minimize (or decrease) the function $\phi(\mu; \mu^n)$ at each iteration and repeat the procedure iteratively,

$$\mu^{n+1} = \arg \min_{\mu \geq 0} \phi(\mu; \mu^n).$$

We call this algorithm separable paraboloidal surrogates (SPS) algorithm. One can show [30] that decreasing the surrogate function $\phi(\mu; \mu^n)$ also decreases the original objective function $\Phi(\mu)$. Hence, this algorithm is intrinsically monotonic. The minimization of $\phi(\mu; \mu^n)$ is easy. Due to the additive separability, the update for each parameter only involves the parameter itself and μ^n . When a quadratic penalty is used, *i.e.* $\psi(t) = t^2/2$ and the nonnegativity constraint is ignored, the maximization can be done exactly in a single step via Newton's algorithm as follows:

$$\mu^{n+1} = \mu^n - \mathbf{D}^{-1} \nabla' \Phi(\mu^n), \quad (5.11)$$

where $\nabla' \Phi(\mu^n)$ is the column gradient of Φ at μ^n and \mathbf{D} is a $p \times p$ diagonal matrix with diagonal entries

$$\mathbf{D}_{jj} = d_j^n + 2\beta \sum_k w_{jk}, \text{ for } j = 1 \dots p.$$

The factor 2 in the denominator comes from the curvature of the separable surrogate $S(\mu; \mu^n)$ in (5.10). The denominator terms d_j^n are:

$$d_j^n = \sum_{i=1}^N g_{ij} \gamma_i c_i(l_i^n). \quad (5.12)$$

For transmission tomography, it is advantageous to use edge-preserving non-quadratic penalties, such as [68]:

$$\psi(t) = \delta^2 [|t/\delta| - \log(1 + |t/\delta|)], \quad (5.13)$$

where $\delta > 0$ is predetermined. We used this penalty in our PL reconstruction results.

In the nonquadratic penalty case, exact minimization of $\phi(\mu; \mu^n)$ is not easy, but one can monotonically decrease the surrogate objective by doing one or more of the following Newton-Raphson type subiteration(s):

$$\hat{\mu}_j := \left[\hat{\mu}_j - \frac{\frac{\partial}{\partial \mu_j} \phi(\hat{\mu}; \mu^n)}{d_j^n + 2\beta \sum_{k \in \mathcal{N}_j} w_{jk} \omega_\psi(\hat{\mu}_j - \mu_k^n)} \right]_+, \quad (5.14)$$

where $\omega_\psi(t) = \dot{\psi}(t)/t$. The detailed explanation of the $\omega_\psi(t)$ function can be found in [30, 38]. The partial derivative of the surrogate ϕ with respect to μ_j can be found as:

$$\frac{\partial}{\partial \mu_j} \phi(\hat{\mu}; \mu^n) = \sum_{i=1}^N g_{ij} \dot{h}_i(l_i^n) + d_j^n (\hat{\mu}_j - \mu_j^n) + \beta \sum_{k \in \mathcal{N}_j} w_{jk} \dot{\psi}(\hat{\mu}_j - \mu_k^n), \quad (5.15)$$

where $\dot{h}_i(l) = \left(\frac{y_i}{b_i e^{-l} + r_i} - 1 \right) b_i e^{-l}$.

Next, we apply the ordered subsets idea to the simultaneous update algorithm developed above.

5.3.4 Ordered Subsets

The ordered subsets principle can be used with any algorithm that involves sums over sinogram indices. The SPS algorithm (5.14) contains sums over sinogram indices in computing the denominator d_j^n terms (5.12) and the gradient terms $\frac{\partial}{\partial \mu_j} \phi$ (5.15). We apply the ordered subsets idea to this algorithm.

Ordered subsets methods group projection data into an ordered sequence of subsets or blocks and processes each block at once. These blocks are usually chosen so that the projections within one block correspond to projections of the image with downsampled projection angles. It was reported [60] that it is best to order the subsets such that the projections corresponding to angles with maximum angular distance from previously used angles are chosen at each step. This accelerates convergence as compared to random or sequential ordering of the subsets. This is due to the fact that the rows of the system matrix corresponding to subsets are chosen to be as orthogonal as possible to previously used subsets.

Let M be the number of subsets chosen in the projection domain. Let S_1, \dots, S_M denote the subsets in the order selected. At step m the following objective function corresponding to the subset S_m should be minimized (or decreased):

$$\Phi_m(\mu) = M \left\{ \sum_{i \in S_m} h_i([\mathbf{G}\mu]_i) \right\} + \beta R(\mu). \quad (5.16)$$

The scaling of the negative log-likelihood function ensures that effective β value is independent of the number of subsets. Note that the original objective function can

be written in terms of the objective functions (5.16) as follows:

$$\Phi(\mu) = \sum_{m=1}^M \frac{1}{M} \Phi_m(\mu). \quad (5.17)$$

The success of the ordered subsets methods depends on the following approximation:

$$\Phi(\mu) \approx \Phi_m(\mu), \quad (5.18)$$

which should be reasonable if the subsets are chosen by subsampling the projection angles.

One iteration is completed when the algorithm cycles through all the projections by using all the subsets. An update performed using a single subset is called a subiteration. The modification of the SPS algorithm to incorporate ordered subsets idea is relatively easy. We call the resulting algorithm ordered subsets transmission (OSTR) algorithm. The algorithm outline is shown in Table 5.1.

The form of the update (5.22) requires the gradient and curvature associated with the penalty term to be computed for each subset. Although the contribution of that computation is only about 4 – 5% in SPS, it might be costly for a large number of subsets since it is repeated for each subset. Other approaches might be possible such as to consider the penalty function as one of the subsets and update it separately at the end of each cycle [25]. It might also be possible to break down the penalty term similar to the likelihood part to reduce computation at each iteration. The choice we made in this work follows naturally from the approximation (5.16) of the PL objective function. Further investigation is required to reduce this computation.

The OSTR algorithm reduces to the SPS algorithm (5.14) when $M = 1$. Since the projections and backprojections are performed for only the elements of a single block, processing of each block in an OSTR algorithm with M subsets (OSTR- M) roughly takes $1/M$ of time that it would take for one iteration of the SPS algorithm for the ML problem. For PL problem, actually it would take more than $1/M$ of the time since the CPU time required for computing the gradient and curvatures of the penalty surrogate at each full iteration is multiplied by the number of subsets. Yet, one hopes that processing of one block increases the objective function as much as one iteration

Precompute d_j if possible
for each iteration $n = 1, \dots, \text{niter}$
for each subset $m=1, \dots, M$

$$\hat{l}_i = \sum_{j=1}^p g_{ij} \hat{\mu}_j, \quad \dot{h}_i = \left(\frac{y_i}{b_i e^{-\hat{l}_i} + r_i} - 1 \right) b_i e^{-\hat{l}_i}, \quad \forall i \in S_m \quad (5.19)$$

$$\mu^{\text{old}} = \hat{\mu}$$

for $j = 1, \dots, p$

$$\dot{L}_j = M \sum_{i \in S_m} g_{ij} \dot{h}_i \quad (5.20)$$

$$d_j = M \sum_{i \in S_m} g_{ij} \gamma_i c_i(\hat{l}_i) \quad (5.21)$$

$$\hat{\mu}_j := \left[\hat{\mu}_j - \frac{\dot{L}_j + \beta \sum_k w_{jk} \dot{\psi}(\hat{\mu}_j - \mu_k^{\text{old}})}{d_j + 2\beta \sum_k w_{jk} \omega_\psi(\hat{\mu}_j - \mu_k^{\text{old}})} \right]_+ \quad (5.22)$$

end

end

end

Table 5.1: OSTR algorithm outline

of the original algorithm. That is, the objective increase for M iterations of OSTR-1 should be close to that increase for one full iteration of OSTR- M . This intuition is verified in the initial iterations and for up to a reasonable number of subsets in the results section.

5.3.5 Precomputed Denominator OSTR

We obtained the OSTR algorithm above from a monotonic simultaneous update algorithm. However, the monotonicity is destroyed by doing ordered subsets iterations. So, the algorithm is monotonic only when one subset is used which is equivalent to SPS.

Since the monotonicity is destroyed at the end anyway, we can remove the condition that the surrogate paraboloid $Q_1(\mu; \mu^n)$ lie above the original objective function

and obtain a yet faster algorithm [30]. Our aim is to precompute the denominator terms d_j^n before iterating and save computation by not updating them. This “pre-computed curvature” idea was introduced in [42, 30] for algorithms that used all the data at once unlike the OS algorithms. We can generalize this idea to ordered subsets easily. First, we notice that we can replace the curvature $c_i(l_i^n)$ terms with the Newton curvatures $\ddot{h}_i(l_i^n)$ in (5.21) and obtain a new algorithm which is no longer guaranteed to be monotonic (even for single subset). We notice that the second derivative of h_i does not change very rapidly and the projections remain very close to the values $l_i^* \triangleq \log\left(\frac{b_i}{y_i - r_i}\right)$ during the iterations (which is actually the minimum for $h_i(l)$ over l). So, as a second approximation, we replace $\ddot{h}_i(l_i^n)$ with $\ddot{h}_i(l_i^*)$ ². The third approximation is to replace M times the sum of the curvatures $\ddot{h}_i(l_i^*)$ over the subset S_m in (5.21) with the sum over all sinogram indices $\{1, \dots, N\}$. This is an accurate approximation if the projections l_i^* vary slowly with respect to the projection angle and each subset is chosen by subsampling the projection angles. So, we can precompute and fix the denominator terms d_j^n by:

$$\begin{aligned} d_j^n &= M \sum_{i \in S_m} g_{ij} \gamma_i \ddot{h}_i(l_i^n) \\ &\approx d_j^* \triangleq \sum_{i=1}^N g_{ij} \gamma_i \ddot{h}_i \left(\log \left(\frac{b_i}{y_i - r_i} \right) \right) = \sum_{i=1}^N g_{ij} \gamma_i \frac{(y_i - r_i)^2}{y_i}. \end{aligned} \quad (5.23)$$

This approximation further reduces CPU time. The minimization step is similar to (5.14) but the gradient terms in (5.20) are computed using just a subset of the data. We also found that doing more than one subiteration (5.14) does not improve “convergence” and costs a lot computationally in the ordered subsets case. So, we only perform one subiteration to improve “convergence” and CPU time.

The algorithm looks very simple for the ML problem. The updates are done as follows in ML reconstruction using the fast precomputed denominator:

Precompute and store: $d_j^* = \sum_{i=1}^N g_{ij} \gamma_i (y_i - r_i)^2 / y_i$, where $\gamma_i = \sum_j g_{ij}$

² l_i^* can only be evaluated when $y_i > r_i$. Otherwise, $h_i(l)$ is a convex function which is monotonically decreasing as $l \rightarrow \infty$. When $r_i \geq y_i$, $\lim_{l \rightarrow \infty} \ddot{h}_i(l) = 0$, so we replace $\ddot{h}_i(l_i^*)$ with zero or a very small number in that case.

for each subset Compute: \hat{l}_i, \dot{h}_i as in (5.19) in Table 5.1

Update:

$$\mu_j := \left[\mu_j - \frac{M \sum_{i \in S_m} g_{ij} \dot{h}_i}{d_j^*} \right]_+ . \quad (5.24)$$

end

This ML-OSTR algorithm is very easy to implement using any kind of system matrix. Precomputed denominator is applicable to PL problem as well. Figure 5.1 shows that PL-OSTR with precomputed denominators converge faster than PL-OSTR with optimum curvature. We used this precomputed denominator approach for the results presented next.

5.4 Phantom Data Results

We acquired a 15-hour blank scan (b_i 's) and a 12-min transmission scan data (y_i 's) using a Siemens/CTI ECAT EXACT 921 PET scanner with rotating rod sources for transmission scans. The phantom used was an anthropomorphic thorax phantom (Data Spectrum, Chapel Hill, NC). Delayed coincidence sinograms were collected separately in each scan. There were 0.920 million prompt coincidences for the reconstructed slice and the delayed coincidences were about 2.6% of the prompt ones. The blank and transmission scan delayed-coincidence sinograms were shown to be numerically close [121], so we used a time-scaled version of blank scan delayed coincidences as the r_i factors with no other processing. The projection space was 160 radial bins and 192 angles, and the reconstructed images were 128×128 with 4.2 mm pixels. The system matrix g_{ij} was computed by using 3.375 mm wide strip integrals with 3.375 mm spacing, which roughly approximates the system geometry.

5.4.1 Reconstructions

The attenuation map was reconstructed for both ML and PL methods using OSTR algorithm with 1, 2, 4, 8, 16 and 32 subsets. In all OSTR reconstructions, precom-

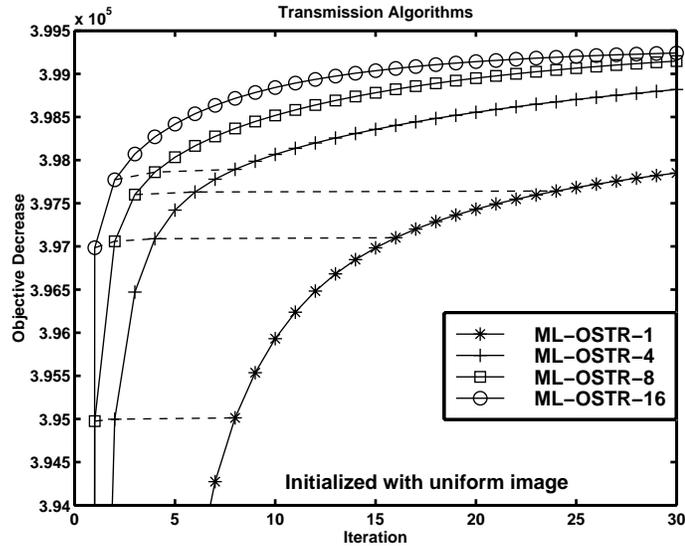


Figure 5.2: Maximum Likelihood $L(\mu^0) - L(\mu^n)$.

puted denominator (5.23) was used. For PSCD reconstructions, we used the optimum curvature (4.33). Figure 5.2 shows objective function decrease for the ML reconstructions initialized with a uniform image. The order-of-magnitude acceleration can be seen by the horizontal dashed lines in this plot for initial iterations. One iteration of ML-OSTR-16 decreases the objective almost as much as 16 iterations of ML-OSTR-1 and 4 iterations of ML-OSTR-4 for initial iterations. Although, when $M > 1$, the algorithm does not converge to the true ML solution, in practice one would only do a few iterations using ML-OSTR- M . In the ML problem, exact maximization is not desired since the ML image is extremely noisy.

Figure 5.3 shows objective function decrease versus iterations for PL reconstructions ($\beta = 2^{10}$ and nonquadratic Lange’s penalty (5.13) with $\delta = 0.004$). The iterations are initialized with an FBP image. There is a speed-up in using more subsets, but as the number of subsets increase, the order-of-magnitude acceleration does not hold. For example, one iteration of PL-OSTR-16 decreases the objective more than one iteration of PL-OSTR-32 (not shown). So, more than 16 subsets did not seem to improve “convergence” for this configuration and data. For comparison, the image is also reconstructed with the optimum curvature paraboloidal surrogates coordinate

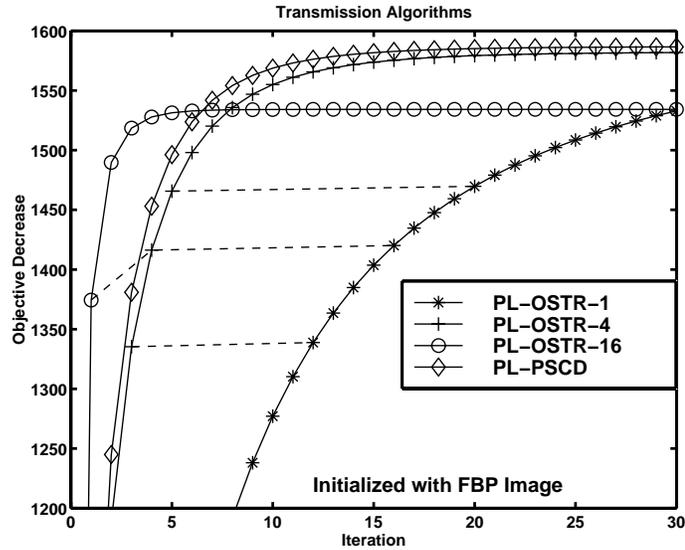


Figure 5.3: Penalized-likelihood $\Phi(\mu^0) - \Phi(\mu^n)$.

descent (PL-PSCD) method which is a fast monotonic algorithm [30]. The CPU times for one iteration of PL-PSCD and one iteration of PL-OSTR-1 are similar. It is clearly seen that PL-OSTR- M algorithms do not converge to the true minimum when $M > 1$. To assure convergence, one could sequentially decrease the number of subsets with each iteration.

5.4.2 Mean Squared and Segmentation Errors

The reconstructions were done using real data. We wished to find mean squared errors and segmentation errors on the reconstructed images. The true image of course was unknown. We acquired a long 14 hour scan of the thorax phantom which was almost noise free. We reconstructed the data with FBP with a sharp filter. Then, we performed a 4 level hard thresholding segmentation on this image with attenuation map parameters assumed to be average standard attenuation map values for air, soft tissue, lungs and bone. We obtained regions for each attenuation level. Then, we eroded these regions with a 5×5 window to get more conservative estimates of the regions and calculated the average value of the FBP image in these regions. These new values were assumed to be the true attenuation coefficient levels for the image (air

$= 0$, lungs = 0.035, soft tissue (water) = 0.093, bone (teflon) = 0.164 cm^{-1}). Then, the FBP image was segmented by thresholding using new levels to obtain the “true” phantom image shown in Figure 5.4. It is possible to use a PL image reconstructed from the 14 hour data as the “true” image as well. The reason we used FBP was not to bias the results in favor of the statistical methods.



Figure 5.4: Image obtained by hard segmentation of the FBP image reconstructed from the 14-hour scan assumed as the true image.

We computed normalized mean squared errors (NMSE) for each reconstruction method by comparing to the true phantom image in Figure 5.4. The reconstructed images were also hard-segmented with the thresholds found above and we evaluated their segmentation performance by counting the number of misclassified pixels.

We also applied the emission ML-OSEM algorithm to the logarithm of the transmission data $-\log \{(y_i - r_i)/b_i\}$. Although there is no theoretical basis for this approach, it has nevertheless been used by many groups. Our results show that this approach is inferior to the ML-OSTR method and that it should be avoided.

Figure 5.5 shows NMSE versus iterations for ML-OSTR, ML-OSEM, PL-OSTR and PL-PSCD methods. Figure 5.6 shows the percentage of segmentation errors versus iterations for the same methods. These results show that ML-OSTR algorithms get noisy after a certain number of iterations and that the iterations should be stopped before convergence. For this transmission scan, the ML-OSTR-16 algorithm should be stopped at the third iteration for lowest NMSE. ML-OSEM applied to the

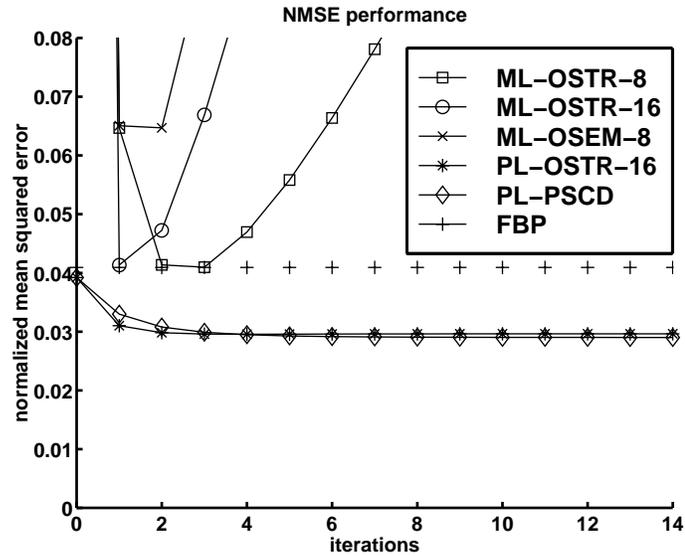


Figure 5.5: Normalized mean squared errors versus iterations for various methods of reconstruction.

logarithm of the transmission data is inferior in quality to all other methods we tried, regardless of number of subsets. PL reconstructions have better quality than ML reconstructions in terms of both lower mean squared errors and lower segmentation errors. Although PL-OSTR-16 algorithm does not converge to the minimum of Φ in Figure 5.3, remarkably it appears to be comparable to the convergent PL-PSCD algorithm in terms of NMSE and segmentation performance. In fact, the normalized mean squared difference between images reconstructed by PL-PSCD and PL-OSTR-16 is less than 0.015% at iteration 30 of each algorithm.

Figure 5.7 shows reconstructed images and their segmentations for FBP, ML-OSTR, ML-OSEM, PL-OSTR and PL-PSCD methods. Each image is the best among their kind. For example, to obtain the FBP image, we performed 20 different FBPs with Hanning windows with different cutoff frequencies and picked the one with lowest NMSE. ML-OSTR image is obtained by 8 subsets at 3 iterations. ML-OSEM image is obtained by 8 subsets at 2 iterations. We used the images obtained at 10th iteration of PL-PSCD and 4th iteration of PL-OSTR-16. The bars show the levels of NMSE and segmentation errors. We conclude that PL reconstruction images are much better than the images obtained using other methods.

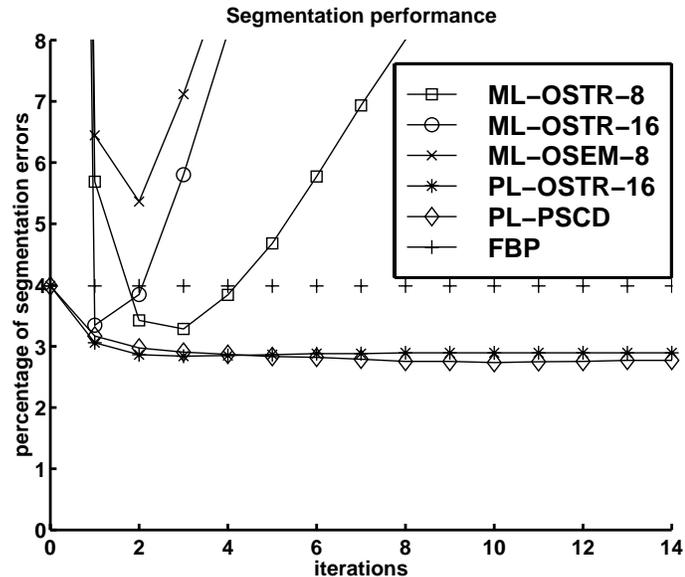


Figure 5.6: Segmentation errors versus iterations for various methods.

5.5 Patient Data Results

We applied the new OSTR algorithm to patient transmission and FDG emission data obtained from ECAT EXACT 921 scanner. A total of 47 slices were acquired for both transmission and emission scans with the delayed coincidences subtracted from the prompt data in both cases. Emission data was acquired for 15 minutes and had total prompt counts of 0.6 million per slice on the average. Transmission data was acquired for 12 minutes before injection and had total average prompt counts of 1.5 million per slice. The randoms were 14% for the emission scan and 6% for the transmission scan. We reconstructed emission images using ACFs obtained from the transmission scan. ACFs were computed using two different methods: 1) conventional (or FBP reconstructed and reprojected) and 2) Nonquadratic penalty PL-OSTR-16 reconstructions with precomputed denominators and 5 iterations. Attenuation maps were both post-smoothed axially with the same Gaussian shaped filter with 5 mm FWHM to reduce noise. Emission reconstructions were done with 1) FBP and 2) Quadratic penalty PL-SAGE [44]. The resolutions of the emission images were matched at 6.6 mm FWHM.

There was only a 12 minute transmission scan data available. The transmission

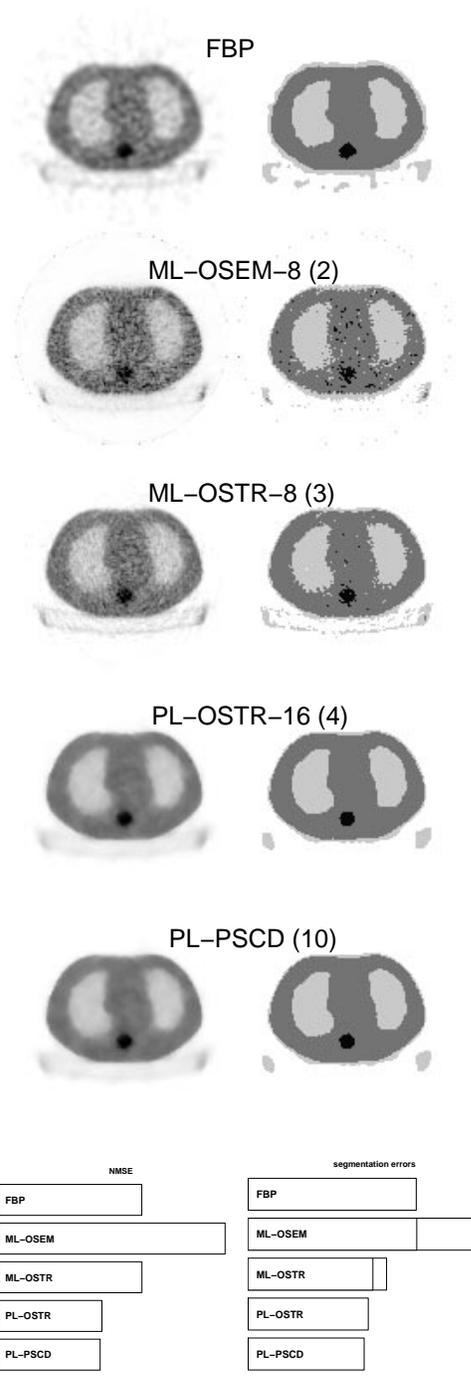


Figure 5.7: Reconstructed (left) and segmented (right) attenuation map images using various methods. The numbers in parentheses show the iteration number. The bar plots show the relative NMSE and segmentation error levels. The middle lines in the right hand side bars for ML methods indicate the segmentation error reduction after median filtering.

randoms were pre-subtracted, so we used the shifted Poisson model (5.2) for the data. The randoms were assumed uniform and the percentage of randoms were estimated from total delayed counts which was available in the file header. To obtain the 2 minute transmission data, we thinned the 12 minute transmission data $y_i^s + 2r_i$ by generating binomial random variables with parameters $n = \max\{0, y_i^s + 2r_i\}$ and $p = 2/12$. Here, n is an approximately Poisson random variable with mean \bar{n} . This binomial thinning approach yields a new (approximately) Poisson random variable with a reduced mean of $p\bar{n}$. The 2 minute scan randoms level was adjusted to 2 minutes as well. Transmission regularization parameter β was adjusted automatically by the method in [46] for the 2 minute scan as well to yield an image with similar resolution properties as the 12 minute image. We also used shifted Poisson model and uniform randoms estimate for the emission data. Scatter and deadtime effects were ignored.

In Figures 5.8, 5.9, and 5.10, we show transaxial, coronal and sagittal emission images reconstructed using ACFs obtained from a 12 minute transmission scan. The image obtained using the statistical method shows some improvement in the image quality over the conventional or hybrid methods. Figures 5.11, 5.12, and 5.13 show transaxial, coronal and sagittal emission images obtained from different combinations of image reconstruction methods with a 2 minute transmission scan. With the 2 minute transmission scan, the improvements in image quality are more apparent for the statistical methods as compared to the conventional methods. These images show that statistical image reconstruction in transmission scans is more important than that in emission scans especially for short transmission scan times.

5.6 Conclusion

We introduced a new ordered subsets algorithm for ML and PL image reconstruction in transmission tomography. Although the algorithm does not converge for number of subsets greater than one, it seems to rapidly decrease the objective function value in the early iterations. The images reconstructed from real PET data with ML

method are worse in quality than images reconstructed with PL method. However, ML-OSTR is superior to ML-OSEM applied to the logarithm of transmission data for this particular data. The new algorithm is easy to implement with any type of system model and does not require column access to the system matrix unlike sequential update algorithms such as coordinate descent. It is also easily parallelizable.

Kudo *et al.* [67] claim that for a general convex objective function, it is possible to obtain convergent ordered subsets algorithms by using appropriate relaxation schemes. The general form in [67] includes OSTR algorithm as a special case. So it might be possible to obtain convergent algorithms by incorporating a relaxation parameter to the OSTR algorithm.

We conclude that if an approximate minimum is acceptable due to practical time and programming constraints, then the OSTR algorithm offers faster convergence than prior methods. However, for guaranteed global convergence to the minimum, other methods must be used, such as [30].

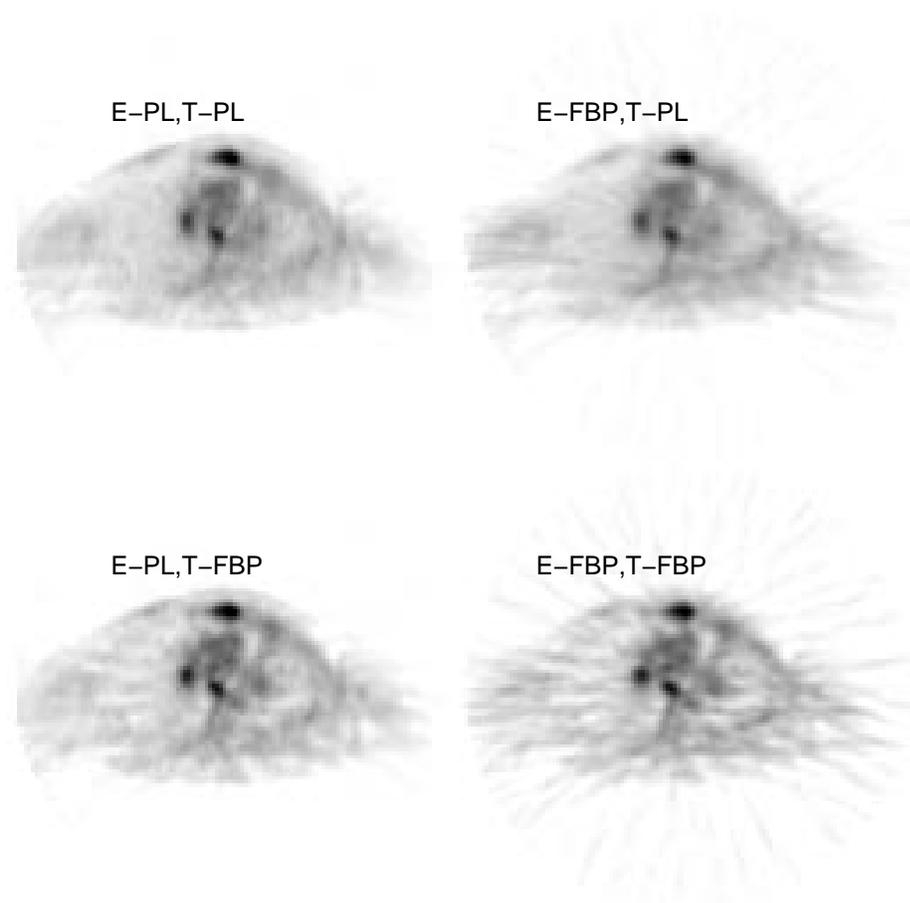


Figure 5.8: FBP (denoted E-FBP) and quadratically penalized PL (denoted E-PL) emission image reconstructions with attenuation correction factors obtained using conventional (denoted T-FBP) and nonquadratic penalty PL (denoted T-PL) attenuation map reconstructions using the OSTR algorithm. Transaxial images of slice 36 (of total 47 slices) are presented. Emission scan was 15 minutes and transmission scan was 12 minutes. Emission images were all reconstructed at the same resolution of 6.6 mm FWHM.

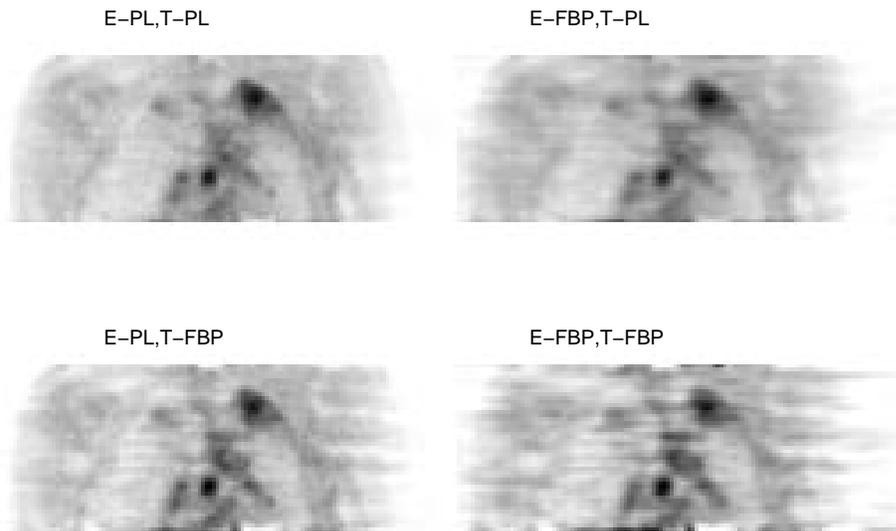


Figure 5.9: Coronal images (slice 64) of the patient data with parameters same as Figure 5.8.

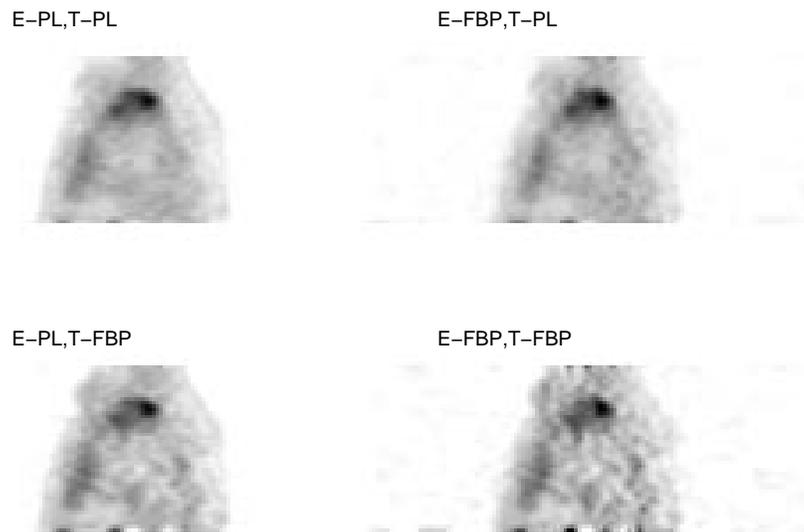


Figure 5.10: Sagittal images (slice 75) of the patient data with parameters same as Figure 5.8.

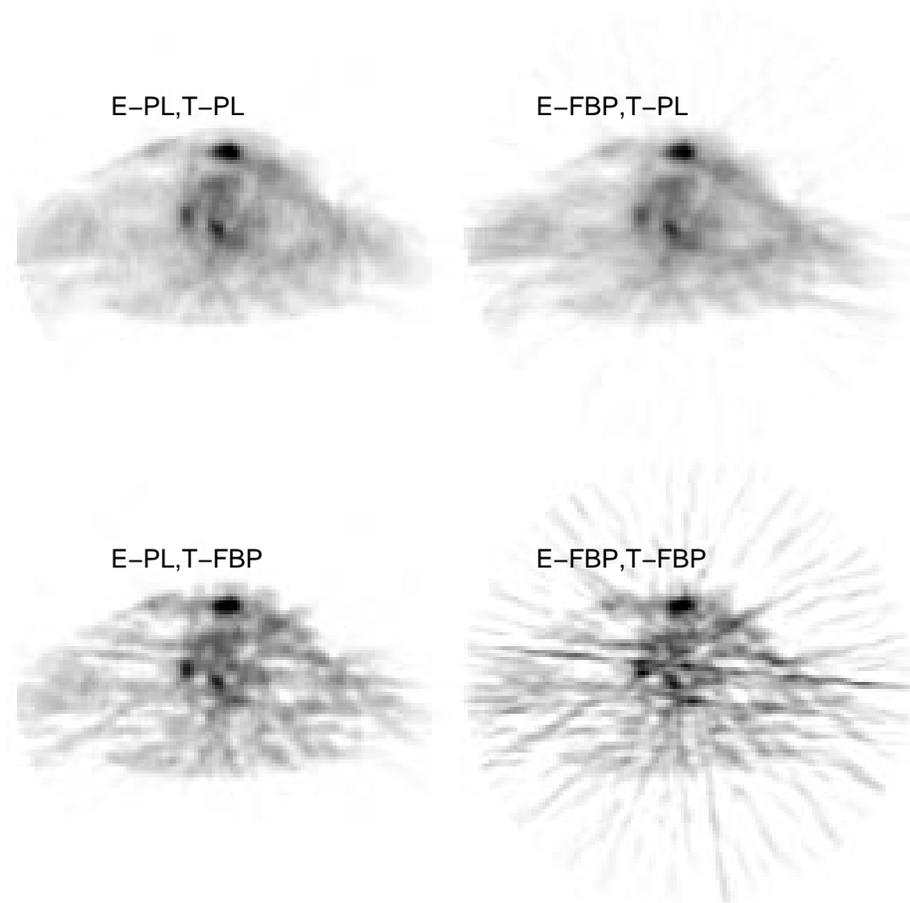


Figure 5.11: Same as Figure 5.9 but with a 2 minute transmission scan which is obtained by thinning the original 12 minute transmission scan.

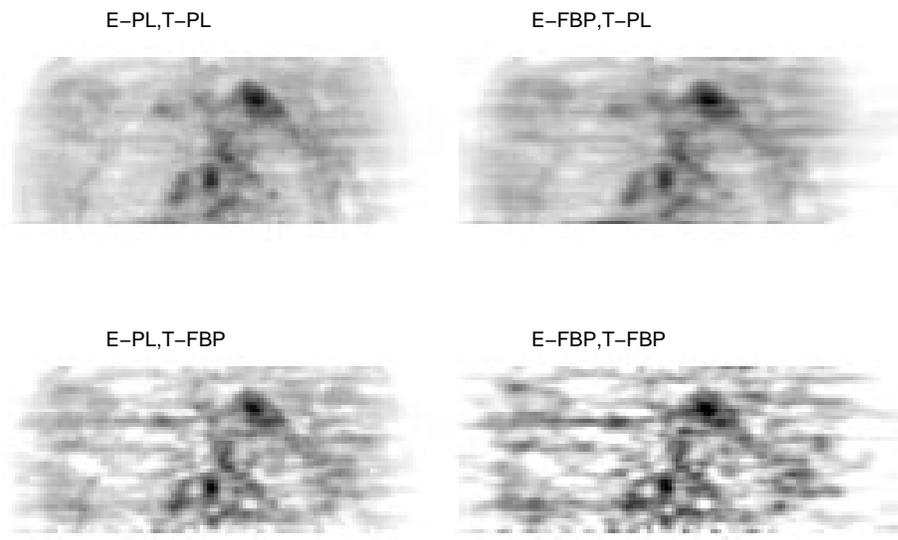


Figure 5.12: Coronal images of the patient data with parameters same as Figure 5.11.

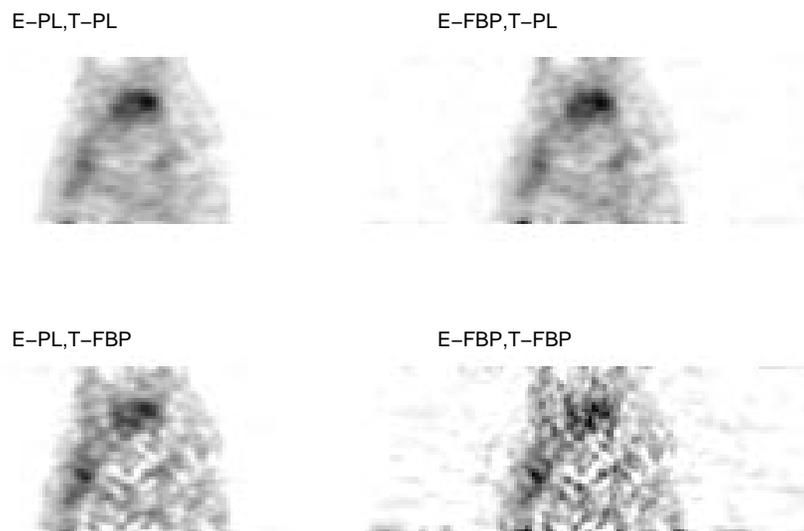


Figure 5.13: Sagittal images of the patient data with parameters same as Figure 5.11.

CHAPTER 6

Image Reconstruction Techniques for PET with Post-injection Transmission Scans

6.1 Post-Injection Transmission Scans

Conventional PET protocols utilize a transmission scan that precedes the radiotracer injection in order not to have interference of emission counts from the body. For example, in a cancer imaging study using FDG to detect lung tumors, the patient should be injected 50 minutes before scanning, in order for the radiotracer to get distributed in the body and obtain accurate readings. Since the transmission scan is done before injection, the patient must lie still in the scanner for this 50 minute period before starting an emission scan. Pre-injection transmission scans have a few disadvantages: 1) The patient throughput is reduced due to the waiting time; 2) It is inconvenient for the patient to lie in the scanner for a time longer than necessary; 3) The likelihood of patient motion is increased between transmission and emission scans due to misregistration of attenuation and emission images. Significant movement produces glaring artifacts, while subtle motions produce small, less easily detected, biases [14].

To overcome these problems, transmission scans can be done after the injection. In a post-injection protocol, the patient is injected before getting inside the scanner and can wait 50 minutes in a waiting room without any restriction. The emission and

transmission scans are done without any delay in between. This technique avoids the problems with the pre-injection transmission scans. There are other advantages of them as well. Whole-body emission scans are very important for detection and staging of cancer. In a whole-body scan, multiple bed positions are utilized with a very short scan time for each position to cover the whole body of the patient. In whole-body scans, transmission scans are often not performed due to time constraints. If better methods of image reconstruction are developed for post-injection transmission scans, a short one can be utilized in whole-body scans to reconstruct quantitatively correct whole-body emission images. Another use of post-injection transmission scans is determination of correct axial position in cancer imaging with a short post-injection transmission scan which would give anatomical information without a need to guess the position and without any loss of time. Hence, there is no doubt that post-injection protocols are clinically very useful and will increase patient throughput considerably. In return, one has to deal with the contamination of the transmission scan with undesirable emission counts.

There have been studies addressing post-injection transmission scans [14, 21, 117, 75], most of which employ the subtraction of a fraction of a separate emission scan from the post-injection transmission scan. The use of transmission rod sources and sinogram windowing of the measurements reduce the emission contamination considerably to a much smaller percentage as compared to the case when ring sources are used. Emission contamination was also estimated by performing a transmission scan with sinogram windowing without the rods present [103]. Some researchers utilized simultaneous emission scans to estimate emission contamination [57]. These methods subtract the estimated emission contamination as well. The subtraction method is shown to work satisfactorily on brain scans with rotating rod sources [21]. In a thorax scan however, the emission counts for some rays traversing the body will exceed the transmission counts. In this case, subtraction approach results in noisy attenuation correction factors and emission images with streak artifacts. We propose that statistical methods can overcome this problem by incorporating the emission contamination into the statistical measurement model.

Our methods will be useful for clinical post-injection studies of the thorax since statistical methods will decrease the effect of emission contamination. Better regularization of the attenuation map will also help us reconstruct more accurate attenuation maps in post-injection as well as pre-injection scans which will result in less noisy and more accurate emission images.

6.1.1 Rod Sources and Sinogram Windowing

Earlier transmission scans in PET were performed using ring sources. Using rotating rod or pin sources for transmission scanning were proposed [13] and applied [112] in the 1980s. The scanner monitors the source position as it rotates about the patient aperture and reject coincidences that do not intersect the source position [14]. This process, called *sinogram windowing*, removes most scattered and random coincidences from the transmission measurement. The sinogram window changes in real time according to the location of the rotating rods. Some researchers suggested to use the “rejected” data (which primarily contains emission counts) to acquire simultaneous emission/transmission (SET) scans [114, 113, 78]. The simultaneous emission data has some problems with a very high rate of scatter, randoms and deadtime. Currently, we do not use the simultaneous emission data and use a conventional emission scan either before or after the transmission scan. However, our model is very general and can incorporate simultaneous scans as well.

An illustration of the rod geometry is given in Figure 6.1. The lines of response that are collinear with the center of a rod at a given position can be found roughly using the following equation

$$d = r_{\text{rod}} \sin(\theta - \phi),$$

where r_{rod} is the radius of the rod orbit, d and θ characterize the projection (distance and angle) and ϕ is the rod angle at current position. This is a sine wave in sinogram domain, but the peaks are cut off because the rods are outside the FOV, so it is almost linear.

A more accurate equation which takes the nonuniform spacing of the constant

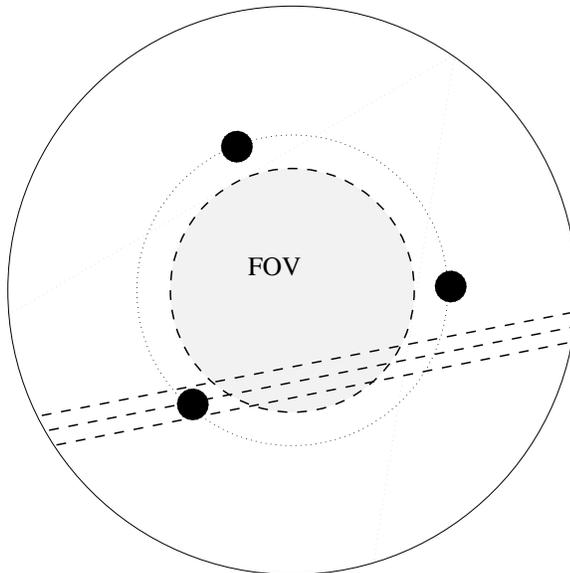


Figure 6.1: Transaxial rod geometry in a rotating rod scanner with three rods.

angle projection lines into account is (spacing is bigger in the middle of FOV, but smaller at the edges) [62]:

$$d = \frac{N}{2\pi} \arcsin \left[\frac{r_{\text{rod}}}{r_{\text{detector}}} \sin(\theta - \phi) \right],$$

where r_{detector} is the detector ring radius and N is the number of detectors per ring.

We take a radial window of width W around this curve to form the sinogram window. The data received only within this dynamic window are recorded since the rest does not contain true transmission counts. The optimal width which maximizes the noise equivalent counts (NEC) is reported to be $W = 5$ for ECAT EXACT scanners [62]. This sinogram window is shown in Figure 6.2 for a single rod. One to three rods are used in current scanners.

We describe the model for the post-injection measurements in the next section.

6.1.2 The Model

The models for the measurements are very similar to the ones in Chapter 2, but transmission scans include emission contamination.

Let $\lambda = [\lambda_1, \dots, \lambda_p]'$ denote the vector of unknown emission counts originating from image pixels, and $\mu = [\mu_1, \dots, \mu_p]'$ be the vector of linear attenuation coeffi-

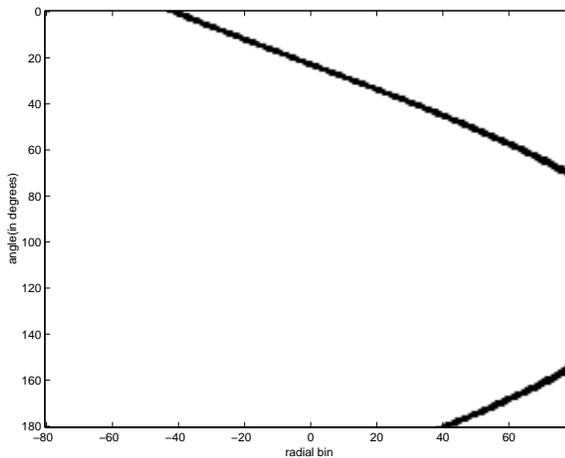


Figure 6.2: Sinogram window illustration for a fixed position of a single rod. This window is changed real time as the rod moves.

icients (having units of inverse length). Let $y^T = [y_1^T, \dots, y_N^T]'$ denote the vector of post-injection transmission scan counts, and $y^E = [y_1^E, \dots, y_N^E]'$ denote the vector of emission scan counts.

We define the survival probabilities as follows:

$$\alpha_i(\mu) = e^{-l_i(\mu)},$$

where $l_i(\mu)$ represents the line integral along projection i of the attenuation map μ :

$$l_i(\mu) = \sum_j g_{ij} \mu_j.$$

We also define the emission contamination count rate

$$\kappa_i(\lambda, \mu) = k_i \varepsilon_i \alpha_i(\mu) p_i(\lambda). \quad (6.1)$$

Here k_i is the fraction of emission counts contaminating the transmission data and includes some other factors, $p_i(\lambda)$ represents the geometric projections of the true emission image:

$$p_i(\lambda) = \sum_j g_{ij} \lambda_j,$$

and ε_i denotes the detector efficiencies.

We assume that the emission scan measurements y^E and the transmission scan measurements y^T are independent Poisson measurements with corresponding means:

$$\bar{y}_i^T(\mu, \lambda) = \tau^T d_i^T (b_i \alpha_i(\mu) + r_i^T + \kappa_i(\lambda, \mu)), \quad (6.2)$$

$$\bar{y}_i^E(\lambda, \mu) = \tau^E d_i^E (\varepsilon_i \alpha_i(\mu) p_i(\lambda) + r_i^E), \quad (6.3)$$

for $i = 1 \dots N$. Here, τ^T and τ^E are transmission and emission scan times respectively. d_i^T and d_i^E are deadtime correction factors for each scan. $l_i(\mu)$ and $p_i(\lambda)$ are geometric tomographic projections of parameters μ and λ . b_i , r_i^T and r_i^E are blank scan, transmission scan randoms and emission scan randoms count rates respectively. We assume $\{b_i\}$, $\{k_i\}$, $\{r_i^T\}$, $\{\varepsilon_i\}$, $\{r_i^E\}$ and $\{g_{ij}\}$ are known constants throughout this chapter.

k_i should contain these factors for accurate modeling of the emission contamination:

1. The proportion of time the i th ray is included in the transmission sinogram window during a whole scan. This value might be space variant depending on the design of the sinogram windows and typically the mean is less than 0.1.
2. The radioactive decay correction between emission and transmission scans. This might be less than or greater than one depending on the scan order.
3. Rod attenuation factors should also be included in k_i . This factor is less than one.

Our final goal is to estimate λ from the two set of measurements. But, an estimate for the attenuation map μ is also found in the process for statistical reconstruction methods.

6.1.3 Resolution

Generally, for reducing noise in the transmission scan, radial sinogram smoothing is employed. Although this reduces the noise in the attenuation correction factors, it introduces a resolution mismatch between two scans. It is well known that when

the transmission data is smoother than the emission data, the outcome is serious artifacts at the edges which can be misinterpreted as pathological changes [16, 79]. Particularly, the photon activity at or around the edges of the attenuation map will be underestimated at the high attenuation side, and overestimated at the low side, an artifact caused by oversmoothed attenuation map.

We found that using smooth ACFs in statistical methods causes a similar effect on the emission image, as might be expected. If we include a biased (blurred) estimate of survival probabilities e^{-l_i} in the emission data model, this bias propagates to the final image, causing visually disturbing images. This is illustrated in Figure 6.3.

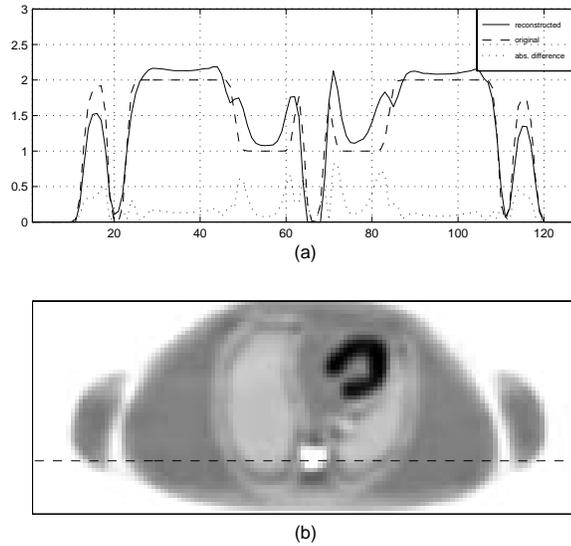


Figure 6.3: The artifacts caused by mismatch of resolutions. Attenuation is reconstructed from noiseless data using quadratic penalty with $\log_2 \beta = 7.6$ resulting in a psf of FWHM 14 mm. The emission image is reconstructed by PL again from noiseless data with ACFs from μ -map using quadratic penalty with $\log_2 \beta = 3.0$ yielding a psf of FWHM 6 mm.

It is possible to overcome the mismatch problem in non-statistical reconstructions by smoothing the emission sinogram to the resolution of the attenuation correction factors [16]. However, when one uses statistical methods, the same procedure cannot be done, since it will destroy the Poisson nature of the data. Consequently, for the statistical methods, it is necessary to use attenuation reconstructions having as small

bias as possible. This suggests use of nonquadratic penalties which result in sharper edges in the attenuation map [68].

6.1.4 Non-statistical Methods

Conventional Subtract/Smooth/FBP Method

To precisely estimate the attenuation map, we need to get an initial estimate of emission contamination in the transmission window. Conventional methods estimate this contamination from either a preceding emission scan or from the simultaneous emission data. The emission contamination estimate is subtracted from the transmission measurements in an effort to correct the data.

The conventional method of ACF computation is described as follows.

1. Estimate emission contamination using (6.1) (6.2) and (6.3) as:

$$k_i(y_i^E/(\tau^E d_i^E) - r_i^E).$$

2. Subtract the emission contamination rate and randoms estimate from the transmission data:
3. Divide the result by the blank scan to obtain a survival probability estimate ($\hat{\alpha}$).
4. Smooth $\hat{\alpha}$ to reduce noise:

$$\hat{\alpha}_i = \text{smooth} \left\{ \frac{(y_i^T/(\tau^T d_i^T) - r_i^T - k_i(y_i^E/(\tau^E d_i^E) - r_i^E))}{b_i} \right\}.$$

5. To reconstruct the emission image, we correct the emission data for the effects of attenuation as well as randoms and detector efficiencies. We smooth the randoms-corrected emission rates with a Gaussian 2-D kernel to achieve the same resolution as the $\hat{\alpha}_i$ in the sinogram domain. The kernel size is determined from the best Gaussian match to the survival probabilities $\hat{\alpha}$:

$$\hat{p}_i = \frac{\text{smooth}\{(1/\varepsilon_i)(y_i^E/(\tau^E d_i^E) - r_i^E)\}}{\hat{\alpha}_i}.$$

6. After attenuation is corrected by division, the emission image is reconstructed by FBP with a ramp filter. That is:

$$\hat{\lambda} = \text{FBP}_{\text{ramp}} \{\hat{p}\}.$$

This conventional “subtraction” method is the simplest way to reconstruct emission images.

Subtract/Segment/Reproject/Smooth/FBP Method

In this method [79, 78], the attenuation map is reconstructed by FBP from the $\log(\hat{\alpha}_i)$ and then segmented in the image domain. The attenuation maps consist of homogeneous regions because attenuation parameter μ depends on the tissue type. For example, in a thorax scan there are roughly three types of distinguishable tissue types: air (μ_{air}), lung (μ_{lung}) and soft tissue (μ_{st}) with $\mu_{air} < \mu_{lung} < \mu_{st}$. The segmentation is performed by the following procedure. The histogram of attenuation values in the image is calculated. The histogram contains peaks corresponding to lung and soft tissue. Then, a Gaussian function is mapped to each peak and two PDFs are calculated for two tissue types. Then, for each pixel value μ_j , the probabilities that it belongs to lung or soft tissue is calculated from the PDFs by $P(\mu_{lung}|\mu_j)$ and $P(\mu_{st}|\mu_j)$. The new value for pixel j is then calculated by [81]:

$$\begin{aligned} \mu'_j &= \mu_{air} \quad \text{if } \mu_j < \mu_{lung} \text{ and } P(\mu_{lung}|\mu_j) < 0.25, \\ \mu'_j &= \mu_{st} \quad \text{if } \mu_j > \mu_{st}, \\ \mu'_j &= \frac{\mu_{lung}P(\mu_{lung}|\mu_j) + \mu_sP(\mu_{st}|\mu_j)}{P(\mu_{lung}|\mu_j) + P(\mu_{st}|\mu_j)} \quad \text{otherwise.} \end{aligned}$$

After this soft-segmentation, the attenuation map is forward projected and smoothed with a one dimensional Gaussian kernel. The emission image is obtained by FBP after emission scan is divided by the survival probability estimates.

This method will misclassify some pixels for short transmission scan times since it does not use the local information about the pixels, so we have not implemented it.

6.2 Sequential Statistical Methods

Statistical methods have been used in emission and transmission tomography for more than a decade. However until recently, they have not been used extensively in the clinic because of problems like long reconstruction times, problems with modeling and resistance to change in the medical society. Most statistical methods have been applied to simulated data with sometimes optimistic assumptions. One of the assumptions in emission tomography is that the ACFs are known beforehand, whereas in reality they are determined from a transmission scan which results in noisy ACF estimates. This noise propagates to the final emission image and should not be ignored. We study the effects of this noise in this chapter and in Chapter 7.

In penalized likelihood estimation for both transmission and emission tomography, we utilize an objective function consisting of a log-likelihood function and a regularizing penalty function. Thus, the two formulations are similar. There might be two kinds of approaches to estimate the emission image. In the “sequential” approach, one estimates the attenuation map μ and the ACFs first from the transmission scan. We then use these ACFs in the emission log-likelihood and estimate the activity image, λ , from the emission counts. On the other hand, in a “joint estimation” approach, two sets of parameters μ and λ can be estimated at the same time from two scans. This might improve the results better especially in post-injection case where both scans depend on both parameters.

In a post-injection protocol and if we employ a “sequential” approach, both objective functions are directly or indirectly affected by both scans. The emission scan counts enter in the transmission log-likelihood due to emission contamination and the emission scan log-likelihood depends on the noisy ACF estimates which in turn depend on transmission scan counts.

We describe the problem formulation for emission and transmission case in the following for the sequential approach. The two formulations are very similar.

Let $y^E = [y_1^E, \dots, y_N^E]'$ and $y^T = [y_1^T, \dots, y_N^T]'$ be emission and post-injection transmission scan count vectors, and let $\mu = [\mu_1, \dots, \mu_p]'$ and $\lambda = [\lambda_1, \dots, \lambda_p]'$ be

attenuation and emission image vectors respectively.

To use a sequential statistical method to reconstruct the emission image, we need to plug in some estimates for the contamination and ACFs to avoid coupling in the optimizations. After plugging in those estimates, the mean values and the objective functions are only functions of the parameter that is to be estimated. The κ and α 's should be estimated from the data that is available and used in the Poisson model.

Let $\hat{\kappa}$ and $\hat{\alpha}$ be estimates of $\kappa(\lambda, \mu)$ and $\alpha(\mu)$ respectively. We *assume* that y^E and y^T are independent Poisson measurements with means:

$$\begin{aligned}\bar{y}_i^T(\mu, \hat{\kappa}) &= \tau^T d_i^T (b_i e^{-l_i(\mu)} + r_i^T + \hat{\kappa}_i), \\ \bar{y}_i^E(\lambda, \hat{\alpha}) &= \tau^E d_i^E (\varepsilon_i p_i(\lambda) \hat{\alpha}_i + r_i^E).\end{aligned}$$

With the equations above for the mean counts kept in mind, we present the formulation for the transmission problem as follows:

$$\begin{aligned}\hat{\mu} &= \arg \min_{\mu} \Phi_{\mu}(\mu; y^T, \hat{\kappa}), \\ \Phi_{\mu}(\mu; y^T, \hat{\kappa}) &= -\frac{1}{\tau^T} L_{\mu}(\mu; y^T, \hat{\kappa}) + \beta_{\mu} R_{\mu}(\mu),\end{aligned}$$

where

$$\begin{aligned}L_{\mu}(\mu; y^T, \hat{\kappa}) &= \sum_i y_i^T \log(\bar{y}_i^T(\mu, \hat{\kappa})) - \bar{y}_i^T(\mu, \hat{\kappa}), \\ R_{\mu}(\mu) &= \sum_{k=1}^K \psi_k^{\mu}([\mathbf{C}\mu - z]_k).\end{aligned}$$

Here $\mu = [\mu_1 \dots \mu_p]'$ are the attenuation map pixel values to be estimated, and $y^T = [y_1^T \dots y_N^T]'$ are the transmission scan counts. $\hat{\kappa}$ denote the estimated emission contamination vector. The roughness penalty function $R_{\mu}(\mu)$ has the same form as the penalty explained in detail in Chapter 3.

We use the algorithms we present in Chapters 4 and Chapter 5 to minimize this objective function. No special treatment is necessary for post-injection case once the emission contamination estimates are included in the algorithm as part of the randoms.

Similarly, the emission problem can be formulated as follows:

$$\begin{aligned}\hat{\lambda} &= \arg \min_{\lambda} \Phi_{\lambda}(\lambda; y^E, \hat{\alpha}), \\ \Phi_{\lambda}(\lambda; y^E, \hat{\alpha}) &= -\frac{1}{\tau^E} L_{\lambda}(\lambda; y^E, \hat{\alpha}) + \beta_{\lambda} R_{\lambda}(\lambda),\end{aligned}$$

where

$$\begin{aligned}L_{\lambda}(\lambda; y^E, \hat{\alpha}) &= \sum_i y_i^E \log(\bar{y}_i^E(\lambda, \hat{\alpha})) - \bar{y}_i^E(\lambda, \hat{\alpha}), \\ R_{\lambda}(\lambda) &= \sum_{k=1}^K \psi_k^{\lambda}([\mathbf{C}\lambda - z]_k).\end{aligned}$$

Here $\lambda = [\lambda_1, \dots, \lambda_p]'$ are the emission parameters (pixel values) to be estimated, and $y^E = [y_1^E, \dots, y_N^E]'$ are the emission scan counts. $\hat{\alpha}$ denote the survival probability vector which is usually estimated from the transmission scan.

It is straightforward to use the available algorithms with this formulation. We include the survival probability estimates $\hat{\alpha}_i$ as a multiplicative factor along with the detector efficiencies. We used a PSCD algorithm [41] to reconstruct emission images presented in this thesis.

6.2.1 One-step Sequential Estimation

As outlined in detail in the previous section, the easiest statistical estimation can be done by first estimating the attenuation map from the transmission scan by including emission contamination estimate in the model. Similarly, emission image can be estimated from the emission scan by including the survival probabilities in the statistical model. We can estimate the parameters μ and λ by the following sequential algorithm:

1. $\hat{\kappa} = \text{diag}\{k_i\} \left(\text{diag}\left\{\frac{1}{\tau^E d_i^E}\right\} \mathbf{B} y^E - r^E \right)$
2. $\hat{\mu} = \arg \max_{\mu} \Phi_{\mu}(\mu; y^T, \hat{\kappa})$
3. $\hat{\alpha} = e^{-[\mathbf{G}\hat{\mu}]_i}$
4. $\hat{\lambda} = \arg \max_{\lambda} \Phi_{\lambda}(\lambda; y^E, \hat{\alpha})$.

We refer to the above method as one-step sequential approach. Here, \mathbf{B} is a smoothing matrix applied to reduce noise, and $\mathbf{G} = \{g_{ij}\}$ is the system matrix for the transmission problem.

6.2.2 Alternating Estimation

In the one-step sequential approach, we notice that we can improve our first estimate of the emission contamination $\hat{\kappa}$ after we find $\hat{\lambda}$ at step 4. This suggests the following algorithm:

First perform a one-step sequential method as above and get $(\hat{\lambda}, \hat{\mu})$. Then:

for $k=1$ to $Niter$

1. $\hat{\kappa} = \text{diag} \{ \varepsilon_i k_i e^{-[\mathbf{G}\hat{\mu}]_i} \} \mathbf{G} \hat{\lambda}$
2. $\hat{\mu} = \arg \max_{\mu} \Phi_{\mu}(\mu; y^T, \hat{\kappa})$
3. $\hat{\alpha} = e^{-[\mathbf{G}\hat{\mu}]_i}$
4. $\hat{\lambda} = \arg \max_{\lambda} \Phi_{\lambda}(\lambda; y^E, \hat{\alpha})$

end

Since the emission contamination estimate is improved at each iteration, we expect better results with this method. This can also be seen as an ad hoc attempt to do joint estimation.

Another alternative is not to fix the $\hat{\mu}$ term in $\hat{\kappa}$ at step 1 and actually set it free for maximization at second step. This might be a little better as compared to above, but we do not expect considerable improvement.

This method is not guaranteed to converge to a reasonable emission image. In our practical experience, we have seen cases where for more than three iterations of this alternating sequential approach, the algorithm started to diverge and we obtained bad emission images.

6.3 Joint Estimation

6.3.1 Introduction

Nowadays, PET scans consist of two separate scans, namely transmission and emission. Transmission scans are performed to estimate the attenuation characteristics of the medium. The attenuation information gathered from transmission scans are used to correct for its effects on the emission data to reconstruct quantitatively accurate emission images. Conventional method consists of linear processing (smoothing) of transmission data to obtain attenuation correction factors (ACFs) and multiplying the smoothed emission data with these factors to correct for the effects of attenuation [16]. Statistical penalized-likelihood methods reconstruct the attenuation map image with a local smoothing penalty and reproject them to obtain ACFs. These ACFs are then used in the penalized-likelihood reconstruction of the emission data by incorporating them in the emission data statistical model [85]. Both of these methods employ a sequential approach. First, ACFs are obtained from transmission scans and then emission data is reconstructed using the ACFs.

In this section, we propose a different approach to image reconstruction which attempts to utilize all the information in transmission and emission scans. In the post-injection transmission scans, certain portion of emission counts contaminate the transmission scan, so there is information about the emission scan in the transmission scan. On the other hand, in an emission scan, there is information about the attenuation properties of the medium. So, to make optimal use of the information in these two scans, one can derive a joint objective function based on both scans to jointly estimate attenuation and emission parameters. This approach should yield better results than the standard sequential statistical estimation.

6.3.2 The Problem

We use the models in (6.2) and (6.3) for transmission and emission scan counts. Our final goal is to estimate λ from the measurements. However, the unknown atten-

uation map μ (or the ACFs) has to be estimated to get an accurate estimate of λ . The quality of the emission image should be *the* performance criteria of any algorithm. Joint estimation is theoretically more advantageous as compared to sequential methods since all the data is used to estimate all the unknown parameters. So we approach the problem as a joint estimation problem. In this method, we maximize one joint objective function to find the optimum values for μ and λ . We simply concatenate the measurements y^E and y^T to form the measurement vector and also λ and μ to form the parameter vector. Since, emission and transmission counts are statistically independent from each other, a joint penalized likelihood objective function can be written by summing up individual log-likelihoods and the individual penalty terms.

$$\begin{aligned} \begin{bmatrix} \hat{\lambda} \\ \hat{\mu} \end{bmatrix} &= \arg \min_{\lambda, \mu} \Phi \left(\begin{bmatrix} \lambda \\ \mu \end{bmatrix}; \begin{bmatrix} y^E \\ y^T \end{bmatrix} \right), \\ \Phi \left(\begin{bmatrix} \lambda \\ \mu \end{bmatrix}; \begin{bmatrix} y^E \\ y^T \end{bmatrix} \right) &= \Phi^T(\mu, \lambda; y^T) + \Phi^E(\mu, \lambda; y^E), \end{aligned}$$

where

$$\Phi^T(\mu, \lambda; y^T) = \sum_{i=1}^N h_i^T(l_i(\mu), p_i(\lambda)) + \beta_\mu R_\mu(\mu)$$

and

$$\Phi^E(\mu, \lambda; y^E) = \sum_{i=1}^N h_i^E(l_i(\mu), p_i(\lambda)) + \beta_\lambda R_\lambda(\lambda),$$

where we view the marginal negative log-likelihood functions h_i^T and h_i^E as a function of the projections l_i and p_i . The objective function only depends on the parameters λ and μ through their projections p_i and l_i :

$$h_i^T(l_i, p_i) = \bar{y}_i^T(l_i, p_i) - y_i^T \log \bar{y}_i^T(l_i, p_i)$$

and

$$h_i^E(l_i, p_i) = \bar{y}_i^E(l_i, p_i) - y_i^E \log \bar{y}_i^E(l_i, p_i).$$

Note that the mean values of two measurements \bar{y}_i^E and \bar{y}_i^T both contain the emission and attenuation projections l_i and p_i in them. In general the objective is nonconvex and the global minimization is very hard.

6.3.3 The Method

We propose to minimize¹ the objective function Φ by alternately updating the emission and attenuation images. We make use of the paraboloidal surrogates idea presented in Chapter 4 to obtain an algorithm that monotonically decreases the objective function assuring convergence to at least a local minimum. First we make this observation: Once either λ or μ is fixed, the form of the functions h_i^T and h_i^E are similar to their counterparts in statistical (penalized likelihood) estimation for the other parameter. We use this observation to derive the following algorithm.

We describe the algorithm using induction. Say, $\mu = \mu^n$ and $\lambda = \lambda^n$ are the current estimates of two parameters obtained after iteration n . We fix the terms λ^n at their current value and allow only the terms μ to change. We denote the current values of the projections as $p_i^n \triangleq p_i(\lambda^n)$ and $l_i^n \triangleq l_i(\mu^n)$. The form of the mean values for both scans when the λ terms are fixed and assumed constant is:

$$\bar{y}_i^S = A_i^S e^{-l_i} + B_i^S, \text{ for } S \in \{T, E\}.$$

where $A_i^S = A_i^S(p_i^n)$ and B_i^S are constants independent of l_i . Furthermore $A_i^S > 0$ and $B_i^S \geq 0$ for both scans. These conditions satisfy the conditions in Theorem 1 in Appendix B, and we can find surrogate parabolas $q_i^T(l_i)$ and $q_i^E(l_i)$ that lie above $h_i^T(l_i)$ and $h_i^E(l_i)$ and tangent to them at the current projection l_i^n . The sum of these two parabolas $q_i(l_i) \triangleq q_i^T(l_i) + q_i^E(l_i)$ is still a parabola. Once the curvature and gradient of the parabola is determined, they can be fed into the paraboloidal surrogates coordinate descent (PSCD) algorithm to update the attenuation parameters to obtain the next iterate μ^{n+1} .

Similarly, we now fix the attenuation map values μ^{n+1} and allow only the λ 's to change. Then, the form of the means for both scans is as follows:

$$\bar{y}_i^S = C_i^S p_i + D_i^S, \text{ for } S \in \{T, E\}. \quad (6.4)$$

Here once again $C_i^S = C_i^S(l_i^{n+1})$ and D_i^S are constants independent of p_i . The objective function viewed as only a function of λ (or p_i 's) is convex, and strictly convex if

¹Or at least achieve a local minimum.

$y_i^S > 0$. Hence, the form of (6.4) makes it possible for $h_i^E(p_i)$ and $h_i^T(p_i)$ (viewed as functions of p_i only) to satisfy the conditions of Theorem 1 in [30]. Hence, similar to the attenuation parameter update, one can obtain parabolas that lie above these h functions and tangent to them at the current iterate p_i^n [41]. After the parabolas are obtained, it is easy to implement a PSCD algorithm similar to [41].

This joint estimation algorithm is easy to implement and results in a very fast algorithm. Once the gradient and curvatures of the parabolas are determined, the problem turns into a PWLS optimization problem and the computations of updates become very fast [30, 41].

We present images reconstructed with this joint method from simulation data in section 6.5.

6.4 Conventional versus One Step Sequential Methods

To compare the proposed statistical methods with conventional FBP based methods, we have carried out some numerical simulations. We reconstructed images with sequential statistical methods and conventional methods as well as hybrid methods. We think of the sequential reconstruction as two steps: 1. Finding the survival probabilities (or ACFs), 2. Finding the emission image. For both steps we have many alternatives, statistical or not. We performed reconstructions using various combinations of our alternatives for each. The alternatives for step 1 are:

1. RAW: Ignore emission contamination and find ACFs by linear methods.
2. SUB: Subtract emission contamination and find ACFs by linear methods.
3. MPL-Q: PL reconstruction with quadratic penalty and reprojection to find ACFs.
4. MPL-N: PL reconstruction with non-quadratic penalty.

And the alternatives we tried for step 2 are:

1. FBP: FBP reconstruction.
2. MPL-Q: Quadratic penalty PL image reconstruction.

6.4.1 The Phantom and The Parameter Values

We used a synthetic attenuation map and emission distribution shown in top left corners of Figures 6.7 and 6.8 as μ^{true} and λ^{true} . The attenuation map represents a human thorax cross section with linear attenuation coefficients 0.16 cm^{-1} , 0.096 cm^{-1} , 0.025 cm^{-1} , for bone, soft tissue and lungs, respectively. The emission image represents sample activity in the same cross section with values 1, 2 and 4 for lungs, soft tissue and heart, respectively. The pixel size is 4.22 mm. We simulated PET transmission and emission scans with 160 radial bins and 192 angles uniformly spaced over 180° . The g_{ij} factors corresponded to 3.375 mm wide strip integrals with 3.375 mm center to center spacing, which is an approximation to ideal line integral that accounts for finite detector width.

We set the number of counts of transmission scan to 2 million and of emission scan to 1 million. The randoms rate were 10% in both scans. There was emission contamination of 10% (k_i) in the transmission window.

We generated $M = 100$ realizations of pseudorandom Poisson transmission and emission measurements according to the models, then reconstructed images using the listed methods. For the PL reconstructions, we used the grouped ascent algorithm for transmission [42], and SAGE algorithm for emission reconstructions [45].

6.4.2 Results

We present the average bias versus standard deviation estimate graphs for both attenuation and emission image reconstructions in Figures 6.4, 6.5 and 6.6. In these figures, the horizontal axis values are obtained from the sample mean image of M reconstructions. We take the average of absolute differences between the true image and the sample mean within a region of interest (W). The values are normalized by the average value of the true image in that region. Let $\theta^n, n = 1..M$ denote the

reconstructions obtained from M realizations, and let $\bar{\theta}$ denote their sample mean, then:

$$b(\theta) = \frac{\sum_{j \in W} |\bar{\theta}_j - \theta_j^{true}|}{\sum_{j \in W} |\theta_j^{true}|} \times 100\%, \quad (6.5)$$

is the estimate of the average bias. We choose the window W to be a central rectangular region containing both lungs for bias estimates.

The vertical axis values are the average standard deviation estimate found from M realizations in another region of interest, i.e:

$$\sigma(\theta) = \frac{1}{|W|} \sum_{j \in W} \sqrt{\frac{1}{M-1} \sum_{n=1}^M (\theta_j^n - \bar{\theta}_j)^2}. \quad (6.6)$$

The window W for the attenuation map is the same as the bias window. For emission standard deviation estimates, we choose a smaller region around the cardiac activity area.

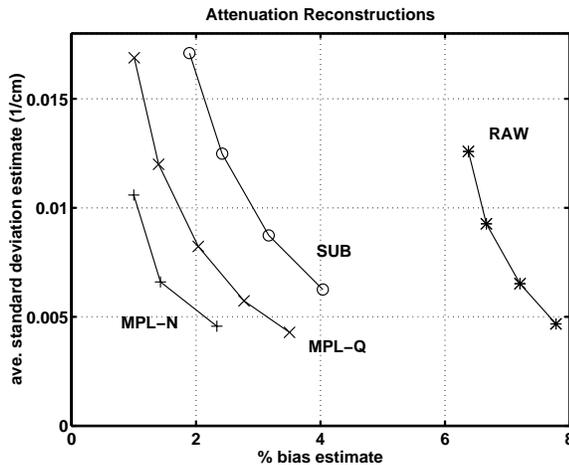


Figure 6.4: Bias vs standard deviation trade-offs for attenuation map reconstructions. Labels indicate different reconstruction methods.

The plot in Figure 6.4 clearly indicates that both MPL methods have better performance than the subtraction method for transmission processing. The raw estimate is severely biased as expected. Also, MPL-N has a better bias-variance trade-off than MPL-Q reconstruction. Non-quadratic penalties appear to be preferable for transmission reconstruction.

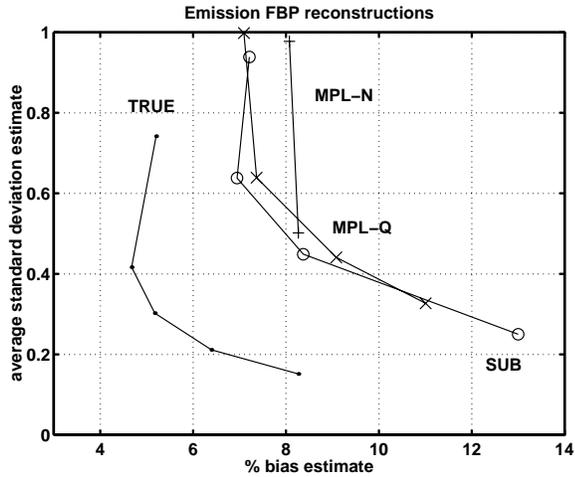


Figure 6.5: Bias vs standard deviation trade-offs for FBP emission reconstructions. Labels indicate the attenuation map that is used.

In Figure 6.6, we used an emission MPL-Q reconstruction with fixed $\beta = 2^{6.4}$ giving a resolution of about 12 mm. Thus, the difference in the bias values are only due to different transmission reconstructions. The bias increases as we use smoother attenuation maps, but the standard deviation estimates do not go down as much. It can be seen that MPL-N transmission followed by MPL-Q emission reconstruction seems to give the best result.

Figure 6.5 presents a similar plot for FBP emission reconstructions. In this case, the resolutions of emission data are matched to attenuation resolution and no further smoothing is done. Thus, actually initial points on the graph are very noisy which results in the strange curves in the plot. This is due to the fact that, even the mean images corresponding to these reconstructions are noisy, which show up in the bias estimate. Actually, in the ideal case, we should put error bars to show the accuracy of our bias and standard deviation estimates. However, since the resolutions are matched, there should not be any contribution from systematic artifacts at the edges unlike MPL-Q estimates. The bias values in this case are proportional to FWHM values for reasonably smooth reconstructions. The plot indicates that FBP method is inferior to MPL-Q for low values of bias. But, for higher values of bias ($\geq 13\%$), FBP seems comparable to MPL-Q because of the resolution mismatch problem in the

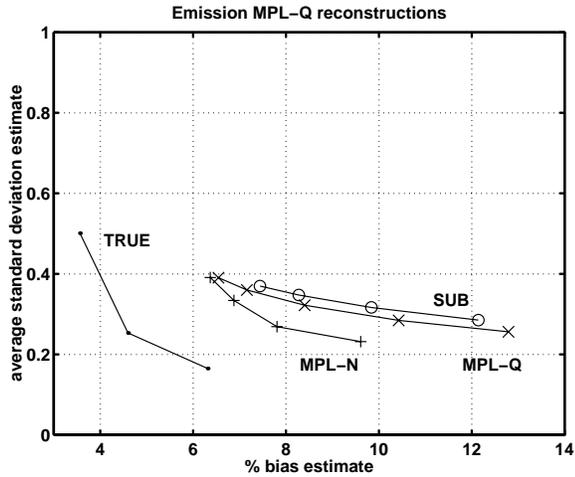


Figure 6.6: Bias vs standard deviation trade-offs for quadratic penalty emission reconstructions. Labels indicate the attenuation map that is used.

statistical method.

We also present sample reconstruction images from a single realization. In Figure 6.7, the attenuation maps can be observed. The MPL-N looks much better than MPL-Q or SUB reconstructions. Here, note that the resolutions of the last two estimates are almost matched, but the first one is sharper at the edges. Even then, MPL-N looks less noisy.

Emission images are presented in Figures 6.8 and 6.9. Overall, MPL-Q estimates look better qualitatively than FBP ones which contain disturbing streak artifacts. The reconstructions with RAW ACFs have systematic negative bias. Visually, the best one is the combination of MPL-N for attenuation and MPL-Q for emission, which seems to reduce the noise inside the lungs appearing as hot spots in other reconstructions. This noise is apparently coming from noisy transmission data, which is successfully reduced in the MPL-N attenuation map estimate.

6.5 Conventional versus Joint Methods

We have done simulations to test the performance of joint statistical estimation method with the conventional method for post-injection scans. We have used the

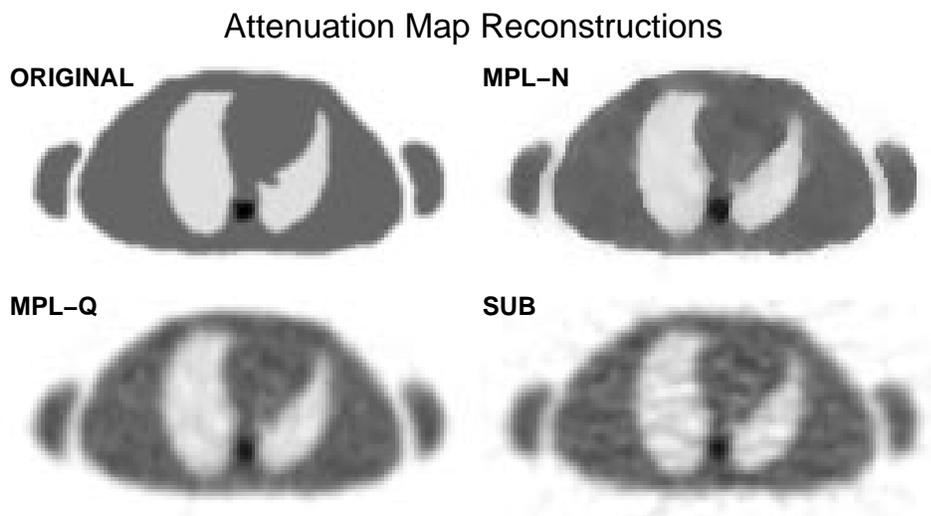


Figure 6.7: Attenuation Maps obtained through various reconstructions from a single noisy measurement realization.

models (6.2) and (6.3) to generate Poisson data using the dimensions and parameters of the ECAT EXACT scanner. We assumed the emission contamination factor to be $k_i = 0.2$. The total transmission and emission counts were nine million and six million respectively. The randoms were assumed to be uniform with 5% and 10% of the transmission and emission counts respectively. The deadtime and scatter effects were ignored in the simulations. The images reconstructed by joint and conventional methods are shown in Figure 6.10. The smoothing parameters were chosen to yield similar images. Nonquadratic Lange's penalty [68] was used for the attenuation image and quadratic penalty for the emission image. Further investigation of bias and variance properties of the joint methods are necessary to evaluate these methods. We leave this as future work.

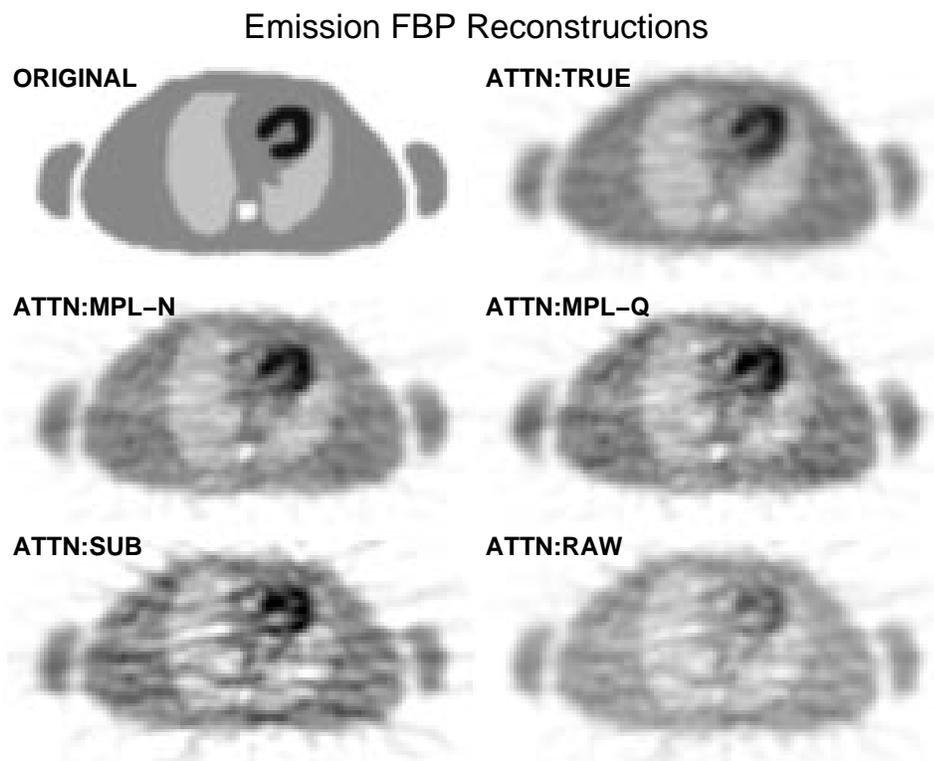


Figure 6.8: FBP reconstructions obtained from a single realization.

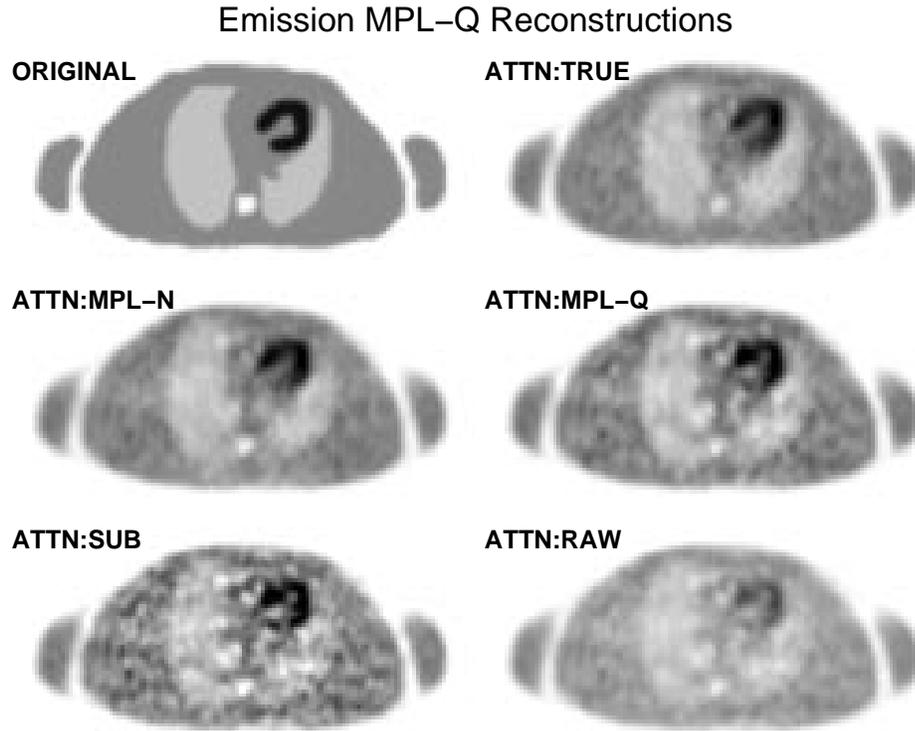


Figure 6.9: MPL-Q reconstructions obtained from the same noisy data as the FBP reconstructions.

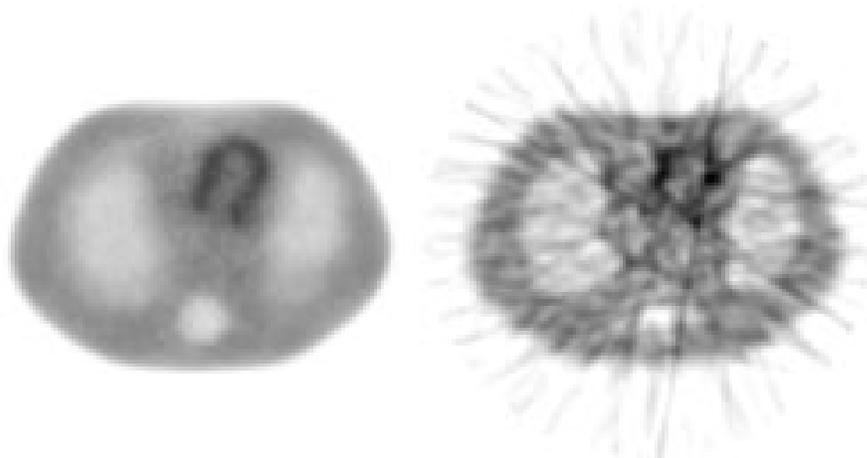


Figure 6.10: Images obtained from joint and conventional image reconstruction methods.

CHAPTER 7

Noise Analysis and Scan Time Optimization

7.1 Introduction

For PET reconstruction, one has to do two sets of scans, namely transmission and emission scans. One uses the attenuation correction information obtained from the former scan to aid in estimating the radiotracer emission image from the latter one. Conventional methods of reconstruction are based on linear processing of the transmission and emission data, multiplicative correction of attenuation factors in the sinogram domain followed by FBP to reconstruct the emission image. This approach ignores Poisson nature of the data. Recently, there is growing interest on reconstruct/reproject methods for attenuation correction in which one reconstructs the attenuation map and, after possibly some processing in the image domain, this map is reprojected to be used in the attenuation correction factors (ACF) computation. The use of statistical methods for reconstructing attenuation maps as well as emission images is becoming attractive in the medical research community, especially due to faster computers and faster algorithms. In this chapter, we reconstruct ACFs using both conventional and penalized-likelihood reconstruct/reproject (PL) methods for post-injection transmission scans. For brevity, we reconstruct emission images with FBP only. Resolution matching is critical in attenuation correction, so we add a post-filtering step to statistical reconstructions to yield approximately Gaussian point spread functions which reduces artifacts from point spread function

mismatches. This post-filter reduces the negative sidelobes from the point spread function of penalized-likelihood reconstructions [46].

In this chapter, we study the effects of emission and transmission scans on the variance of the reconstructed emission image for different reconstruction methods. Particularly we are interested in the optimum scan time fractions under a fixed total scan time constraint, which would result in the smallest variance in a region of interest in the final emission image estimate. Previous studies of scan time optimization [3] were based on NEC criteria with multiple acquisitions of emission and transmission data and focused on conventional reconstructions. Some of the intermediate (co)variance approximations developed here might also be useful for other purposes such as determining the weights in a 3D weighted least-squares image reconstruction [18]. We analyze both the conventional and statistical reconstruction cases. We give approximate analytical formulas for conventional and quadratic penalty attenuation map reconstructions and compare empirical results with the analytical predictions. Our analysis is based on Poisson statistics and mathematical approximations [37].

Let $y^E = [y_1^E \dots y_N^E]'$ and $y^T = [y_1^T \dots y_N^T]'$ be emission and post-injection transmission scan count vectors, and let $\mu = [\mu_1 \dots \mu_p]'$ and $\lambda = [\lambda_1 \dots \lambda_p]'$ be attenuation map and emission image pixel value vectors respectively.

We define the survival probabilities as follows:

$$\alpha_i(\mu) = e^{-l_i(\mu)},$$

where $l_i(\mu)$ represents the line integral along projection i of the attenuation map μ . We also define the emission contamination count rate

$$\kappa_i(\lambda, \mu) = k_i \varepsilon_i \alpha_i(\mu) p_i(\lambda).$$

Here k_i is the fraction of emission counts contaminating the transmission data (the portion in the transmission window for rotating rod sources), $p_i(\lambda)$ represents the geometric projections of the true emission image λ , and ε_i contains the detector efficiencies and a scaling factor that accounts for emission scan count rate. We assume that the emission scan measurements y^E and the transmission scan measurements y^T

are independent Poisson measurements with corresponding means:

$$\bar{y}_i^T(\mu, \lambda) = \tau^T (b_i \alpha_i(\mu) + r_i^T + \kappa_i(\lambda, \mu)), \quad (7.1)$$

$$\bar{y}_i^E(\lambda, \mu) = \tau^E (\varepsilon_i \alpha_i(\mu) p_i(\lambda) + r_i^E). \quad (7.2)$$

Here, τ^T and τ^E are transmission and emission scan times respectively. $l_i(\mu) = [\mathbf{G}\mu]_i \triangleq \sum_{j=1}^p g_{ij} \mu_j$ and $p_i(\lambda) = \sum_{j=1}^p g_{ij} \lambda_j$ are geometric tomographic projections of parameters μ and λ . b_i , r_i^T and r_i^E are blank scan, transmission scan randoms and emission scan randoms count rates respectively. We assume $\{b_i\}$, $\{r_i^T\}$, $\{\varepsilon_i\}$, $\{r_i^E\}$ and $\{g_{ij}\}$ are known constants throughout this chapter.

7.2 ACF Estimation

Attenuation correction is a must for quantitatively accurate emission image reconstruction. We define attenuation correction factors (ACFs) $\gamma_i(\mu) = e^{l_i(\mu)} = 1/\alpha_i(\mu)$. This is the multiplicative factor that corrects for the effects of attenuation in the emission data. We consider two different ways of estimating the ACFs: 1) Conventional smoothing method and 2) Reconstruct/reproject penalized-likelihood (PL) method.

In the non-statistical conventional method, we estimate the emission contamination by:

$$\hat{\kappa}_i = \text{smooth} \left\{ k_i \left(y_i^E / \tau^E - r_i^E \right) \right\}, \quad (7.3)$$

and we estimate the ACFs by reciprocating the survival probabilities, that is $\hat{\gamma}_i = 1/\hat{\alpha}_i$, where

$$\hat{\alpha}_i = \text{smooth} \left\{ \left(y_i^T / \tau^T - r_i^T - \hat{\kappa}_i \right) / b_i \right\}. \quad (7.4)$$

The smoothing operation is often used to reduce noise in the ACFs. We also use smoothing to reduce noise in the emission contamination estimate in (7.3).

In a statistical reconstruction, one estimates the ACFs by $\hat{\gamma}_i = e^{l_i(\hat{\mu})}$ where $\hat{\mu}$ is the attenuation map estimate computed by the reconstruction algorithm. The emission contamination estimate (7.3) is included in the model. The statistical reconstruction is considered in detail in Section 7.5.

7.3 Emission Image Reconstruction

For brevity, we consider here the conventional FBP method to reconstruct emission images. We define the attenuated emission projections function as

$$z_i(\lambda, \mu) = p_i(\lambda)\alpha_i(\mu).$$

A linear unbiased estimate of this function is the following:

$$\hat{z}_i = \text{smooth}\{(y_i^E/\tau^E - r_i^E)/\varepsilon_i\}. \quad (7.5)$$

Then, an estimate of the projections $p_i(\lambda)$ can be obtained by:

$$\hat{p}_i = \hat{\gamma}_i \hat{z}_i.$$

The emission image is reconstructed by standard FBP method. We use the ramp filter only because the estimate \hat{p}_i is already a smooth estimate of $p_i(\lambda)$. Thus,

$$\hat{\lambda} = \text{FBP}_{\text{ramp}}\{\hat{p}_i\}.$$

7.4 Emission Covariance Estimates

The covariance of the emission image estimate vector $\hat{\lambda}$ obtained by the above procedure can be written as follows:

$$\text{Cov}\{\hat{\lambda}\} = \mathbf{P} \text{Cov}\{\hat{p}\} \mathbf{P}', \quad (7.6)$$

where the matrix \mathbf{P} represents the linear FBP operation with a ramp filter. We need to find the covariance of the random vector $\hat{p} = [\hat{z}_i \hat{\gamma}_i]_{i=1}^N$. The computation of the exact covariance of this expression is computationally intensive and is not desirable. Instead, we prefer to evaluate this covariance as a separable sum of the covariances of the vectors \hat{z} and $\hat{\gamma}$. For this purpose, we consider the Taylor series expansion of $\hat{z}_i \hat{\gamma}_i$ in the neighborhood of $\bar{z}_i \bar{\gamma}_i$ where \bar{z} and $\bar{\gamma}$ are mean values of \hat{z} and $\hat{\gamma}$ respectively. Then:

$$\hat{p}_i = \hat{z}_i \hat{\gamma}_i \approx \bar{z}_i \bar{\gamma}_i + \bar{\gamma}_i (\hat{z}_i - \bar{z}_i) + \bar{z}_i (\hat{\gamma}_i - \bar{\gamma}_i) \quad (7.7)$$

$$= \bar{z}_i \hat{\gamma}_i + \bar{\gamma}_i (\hat{z}_i - \bar{z}_i), \quad (7.8)$$

and consequently,

$$\text{Cov}\{\hat{p}\} \approx \mathcal{D}\{\bar{z}_i\} \text{Cov}\{\hat{\gamma}\} \mathcal{D}\{\bar{z}_i\} + \mathcal{D}\{\bar{\gamma}_i\} \text{Cov}\{\hat{z}\} \mathcal{D}\{\bar{\gamma}_i\}. \quad (7.9)$$

The ACFs $\hat{\gamma}_i$ are not *linearly* related to variables with known covariances. In the conventional method, $\hat{\gamma}_i = 1/\hat{\alpha}_i$. In the statistical method $\hat{\gamma}_i = e^{l_i(\hat{\mu})}$. These are both nonlinear functions. Since the covariance of $\hat{\alpha}$ can be found exactly for the conventional method and the covariance of $\hat{\mu}$ can be approximated for the statistical method, we can linearize these formulas around $\bar{\alpha}_i$ and \bar{l}_i to get an estimate of the covariance of $\hat{\gamma}$. This linearization was the method used in [107] to estimate the variances of the ACFs. But, this linearization is not very accurate for especially the conventional method, because the function $f(x) = 1/x$ cannot be closely approximated by a linear function especially when the denominator (survival probabilities) is close to zero and the variance of the denominator is high.

To overcome this problem, we propose an approximation for the probability distribution function of the ACFs. We assume that $\hat{\gamma}_i$ are lognormal distributed. A random variable is lognormal distributed if its logarithm is normally distributed. We believe this is a very accurate assumption because $\hat{\gamma}_i$ is an estimate of $e^{l_i(\mu)}$ and the projections of any random variable (here $l_i(\hat{\mu})$) should be approximately Gaussian due to the Central Limit Theorem. This provides us extra information about the ACFs. With this assumption, one can compute the mean and variance of $\hat{\gamma}_i$'s directly in terms of mean and variance of $\hat{\alpha}_i$ in the conventional method and in terms of mean and variance of \hat{l}_i in the statistical method.

So, for the conventional method, we get:

$$\bar{\gamma}_i = \frac{\bar{\alpha}_i^2 + \sigma_{\hat{\alpha}_i}^2}{\bar{\alpha}_i^3} \quad (7.10)$$

and

$$\sigma_{\hat{\gamma}_i}^2 = \frac{(\bar{\alpha}_i^2 + \sigma_{\hat{\alpha}_i}^2)^2 \sigma_{\hat{\alpha}_i}^2}{\bar{\alpha}_i^8}. \quad (7.11)$$

Even with the lognormality assumption, the covariance matrix of $\hat{\gamma}$ is not easy to compute directly. But, the diagonal of the matrix is known. So, we propose this

approximation for the covariances:

$$\text{Cov}\{\hat{\gamma}_i, \hat{\gamma}_j\} \approx \frac{\sigma_{\hat{\gamma}_i} \sigma_{\hat{\gamma}_j}}{\sigma_{\hat{\alpha}_i} \sigma_{\hat{\alpha}_j}} \text{Cov}\{\hat{\alpha}_i, \hat{\alpha}_j\} \quad (7.12)$$

$$= \sigma_{\hat{\gamma}_i} \sigma_{\hat{\gamma}_j} \rho(\hat{\alpha}_i, \hat{\alpha}_j), \quad (7.13)$$

where $\rho(\hat{\alpha}_i, \hat{\alpha}_j)$ represents the correlation coefficient of the vector $\hat{\alpha}$. In matrix form:

$$\text{Cov}\{\hat{\gamma}\} \approx \mathbf{D}_1 \text{Cov}\{\hat{\alpha}\} \mathbf{D}_1,$$

where

$$\mathbf{D}_1 = \mathcal{D} \left\{ \frac{\sigma_{\hat{\gamma}_i}}{\sigma_{\hat{\alpha}_i}} \right\} = \mathcal{D} \left\{ \frac{\bar{\alpha}_i^2 + \sigma_{\hat{\alpha}_i}^2}{\bar{\alpha}_i^4} \right\}.$$

We make sure that the diagonal of the covariance matrix of $\hat{\gamma}$ matches the variances we get from the lognormal assumption. This formula assumes that the correlation coefficient of $\hat{\gamma}$ is largely determined by the smoothing operator \mathbf{B} and is the same as the correlation coefficient for $\hat{\alpha}$.

Plugging in the approximation (7.9) for $\hat{\gamma}$ and writing $\bar{z}_i \approx \bar{\alpha}_i \text{smooth}\{p_i(\lambda^{\text{true}})\}$, we get the following

$$\text{Cov}\{\hat{p}\} \approx \mathbf{D}_E \text{Cov}\{\hat{z}\} \mathbf{D}_E + \mathbf{D}_T \text{Cov}\{\hat{\alpha}\} \mathbf{D}_T, \quad (7.14)$$

where

$$\mathbf{D}_E \triangleq \mathcal{D} \left\{ \frac{\bar{\alpha}_i^2 + \sigma_{\hat{\alpha}_i}^2}{\bar{\alpha}_i^3} \right\}$$

and

$$\mathbf{D}_T \triangleq \mathbf{D}_E \mathcal{D} \left\{ \text{smooth}\{p_i(\lambda^{\text{true}})\} \right\}.$$

The mean and covariance of \hat{z} can be found easily from the expression (7.5) since it is linearly related to y^E . A simple analysis yields: $\bar{z}_i = \text{smooth}\{p_i(\lambda^{\text{true}})e^{-l_i(\mu^{\text{true}})}\}$ and from (7.5) and (7.2):

$$\text{Cov}\{\hat{z}\} = \frac{1}{\tau^E} \mathbf{B} \mathcal{D} \{q_i\} \mathbf{B}', \quad (7.15)$$

where

$$q_i \triangleq (\varepsilon_i \alpha_i(\mu^{\text{true}}) p_i(\lambda^{\text{true}}) + r_i^E) / \varepsilon_i^2, \quad (7.16)$$

and \mathbf{B} is a smoothing convolution matrix along the radial direction of the projection space. It was argued in [20] that angular smoothing is not desirable in attenuation correction, so we smooth only in radial direction.

For conventional ACF computation, ignoring the noise in the emission contamination estimate, the covariance of $\hat{\alpha}$ can be found from (7.4) and (7.1).

$$\text{Cov}\{\hat{\alpha}\} = \frac{1}{\tau^T} \mathbf{B} \mathbf{D} \{s_i\} \mathbf{B}^T, \quad (7.17)$$

where

$$s_i \triangleq (b_i \alpha_i(\mu^{\text{true}}) + r_i^T + \bar{\kappa}_i) / b_i^2.$$

Here, \mathbf{B} is the same smoothing matrix as in (7.15). The same operator \mathbf{B} is used to obtain both \hat{z} and $\hat{\alpha}$ to avoid artifacts from resolution mismatch [16, 28]. We used Gaussian smoothing as suggested in [16] which avoids artifacts in the reconstructed image. The mean of emission contamination can be determined from (7.3) as

$$\bar{\kappa} = \mathbf{B} \left[k_i \varepsilon_i \alpha_i(\mu^{\text{true}}) p_i(\lambda^{\text{true}}) \right]_{i=1}^N.$$

The variance of $\hat{\alpha}_i$ can be found from (7.17) as

$$\sigma_{\hat{\alpha}_i}^2 = s_i / \tau^T \sum_k B_{ik}^2.$$

Using (7.4), one can find the mean values of $\hat{\alpha}$ as

$$\bar{\alpha} = \mathbf{B} [\alpha_i(\mu^{\text{true}})]_{i=1}^N. \quad (7.18)$$

The variance of the sum over a region of interest in the emission image can be found from (7.6), (7.14), (7.15) and (7.17) as

$$\text{Cov}\{u' \hat{\lambda}\} = u' \text{Cov}\{\hat{\lambda}\} u = \frac{1}{\tau^E} v^E + \frac{1}{\tau^T} v^T, \quad (7.19)$$

with

$$v^E = \sum_{i=1}^N q_i (w_i^E)^2 \text{ and } v^T = \sum_{i=1}^N s_i (w_i^T)^2,$$

and where u is a vector of ones in the region of interest and zeros elsewhere and we define the w^E and w^T vectors as follows:

$$w^E \triangleq \mathbf{B}' \mathbf{D}_E \mathbf{P}' u, \quad w^T \triangleq \mathbf{B}' \mathbf{D}_T \mathbf{P}' u.$$

7.5 Penalized-Likelihood Attenuation Reconstruction

While conventional method of ACF computation has been used for some time, reconstruct/reproject methods have gained some interest recently. In a statistical reconstruct/reproject method for ACF computation, an attenuation map estimate $\hat{\mu}$ is found from noisy transmission data by maximizing the penalized-likelihood objective function

$$\Phi(\mu; y^T) = L(\mu; y^T) - \beta R(\mu),$$

where $L(\mu, y^T)$ is the log-likelihood function and $R(\mu)$ is a regularizing roughness penalty function. After estimating the attenuation map $\hat{\mu}$, we estimate the ACFs by:

$$\hat{\gamma}_i = e^{l_i(\hat{\mu})},$$

where $l_i(\hat{\mu}) = [\mathbf{G}\hat{\mu}]_i$ is the geometric projection of the attenuation map estimate $\hat{\mu}$. If one uses FBP for emission reconstruction, then \hat{z} should be smoothed to yield similar resolution with the $\hat{\gamma}$ [36] in order to reduce resolution mismatch artifacts.

7.5.1 Resolution

Penalized likelihood (PL) or penalized weighted least squares (PWLS) methods are very attractive image reconstruction methods due to their superb noise reduction properties. The variance weighting in PWLS method reduces the variance of the estimates as compared to penalized unweighted least squares (PULS) or FBP reconstructions, because it makes use of the statistical information in the measurements. However, attenuation maps reconstructed with PL or PWLS methods have non-uniform resolution [46] even with a quadratic penalty. This non-uniform resolution is caused by the variance weighting in PWLS (or PL) method and hence does not exist in a PULS reconstruction. Due to this non-uniform resolution, ACF computation by PL method from a real transmission scan causes resolution mismatch between the emission data and reconstructed ACFs. This mismatch reveals itself as artifacts in the final reconstructed emission image.

Fessler's certainty based penalty [36] yields more uniform resolution in terms of the average FWHM of the point spread function over the image. But, it still has non-uniform resolution in that the psf is not circularly symmetric but the level contours look like ellipses whose orientation are image dependent and space-variant. Stayman and Fessler have recently proposed a new modification to the quadratic penalty [106] which yields more circularly symmetric uniform resolution properties. We used this modification in our reconstructions. This modification makes the resolution properties of the PL method close to PULS method. Quadratic PULS method was shown to be essentially equivalent to FBP method with the following constrained least-squares (CLS) filter defined in spatial frequency domain by (equation (50) in [36])

$$F_p(u; \beta) = \frac{\text{sinc}(ku) / \text{sinc}(u)}{\text{sinc}(ku)^2 + c\beta u^3}, \quad u \in [0, 0.5] \quad (7.20)$$

where u denotes spatial frequency, k is the ratio of the detector strip width to the pixel size of the system model, and c is a constant dependent on system geometry. This CLS filter has high negative sidelobes in the space domain. The filters that smooth the ACFs and emission data have to be matched. So, the emission data should be blurred with the same filter (7.20). But, due to high negative sidelobes of filter in (7.20), after dividing the appropriately blurred emission data to computed survival probabilities from reconstructions, we get artifacts especially for higher blurring amounts (higher β s) around the boundaries of the image. So, we conclude that the results in [16] only hold for Gaussian smoothing.

To overcome this problem, we first reconstruct a higher resolution image using a smaller β value than desired and then we filter the projections with the following filter to obtain a Gaussian smoothing effect:

$$F_2(u) = \frac{F_p(u; \beta)F_g(u; w)}{|F_p(u; \beta)|^2 + 0.1(1 - \cos(2\pi u))}, \quad u \in [0, 0.5],$$

where $F_g(u; w)$ is the desired Gaussian filter with desired FWHM w . Now, the emission data is also filtered with the Gaussian shaped filter $F_g(u; w)$. This approach reduces artifacts and yields acceptable images. The ACF computation in this case is done as follows:

$$\hat{\mu} = \arg \max_{\mu} \Phi(\mu; y^T),$$

$$\begin{aligned}\hat{l} &= \mathbf{B}_2 \mathbf{G} \hat{\mu}, \\ \hat{\gamma}_i &= e^{\hat{l}_i},\end{aligned}\tag{7.21}$$

where \mathbf{B}_2 is the convolution matrix corresponding to $F_2(u)$ above.

7.5.2 Covariance Approximations

The covariance formula in (7.9) is still valid in PL transmission reconstruction. We use the following first order Taylor series expansion for the ACFs:

$$\hat{\gamma} = e^{\hat{l}} \approx e^{\bar{l}} + e^{\bar{l}}(\hat{l} - \bar{l}),\tag{7.22}$$

where $\bar{l} = \mathbf{B}_2 \mathbf{G} \check{\mu}$ is the mean projection vector where $\check{\mu} = \arg \max_{\mu \geq 0} \Phi(\mu; \bar{y}^T)$ is the image reconstructed with noiseless data. $\check{\mu}$ is a very good approximation for the mean of $\hat{\mu}$ [37]. We do not use the lognormality assumption here, because we believe that the above approximation is accurate enough and lognormal assumption leads to much more computation. From (7.22) and (7.21),

$$\text{Cov}\{\hat{\gamma}\} = \mathcal{D}\{e^{\bar{l}}\} \mathbf{B}_2 \mathbf{G} \text{Cov}\{\hat{\mu}\} \mathbf{G}' \mathbf{B}_2' \mathcal{D}\{e^{\bar{l}}\}.$$

To find the covariance of the implicitly defined estimator $\hat{\mu}$, we use the formulas introduced in [37].

The general form of penalized-likelihood estimates is $\hat{\mu} = \arg \max_{\mu} \Phi(\mu; y^T)$, where μ is the parameter vector and y^T is the measurement vector. This defines an implicit function $\hat{\mu} = h(y^T)$. A first order Taylor expansion of the equation $\nabla \Phi(\mu; y^T) = 0$ around $(\check{\mu}, \bar{y}^T)$ yields the following approximation [37]:

$$\text{Cov}\{\hat{\mu}\} \approx \mathbf{Q} \text{Cov}\{y^T\} \mathbf{Q}',\tag{7.23}$$

where

$$\mathbf{Q} = \left[-\nabla^2 \Phi(\check{\mu}, \bar{y}^T) \right]^{-1} \nabla \Phi(\check{\mu}, \bar{y}^T).$$

We use this formula to evaluate the covariance of the penalized-likelihood estimate of the attenuation map μ . We again ignore the noise in the emission contamination estimate and use the mean value for it in our approximations. The formula yields:

$$\text{Cov}\{\hat{\mu}\} \approx \frac{1}{\tau^T} \mathbf{H}^{-1} \mathbf{G}' \mathcal{D} \left\{ \frac{(b_i e^{-\bar{l}_i})^2 (b_i e^{-l_i^{\text{true}}} + \tilde{r}_i^T)}{(b_i e^{-\bar{l}_i} + \tilde{r}_i^T)^2} \right\} \mathbf{G} \mathbf{H}^{-1},$$

where

$$\mathbf{H} = \mathbf{G}'\mathcal{D} \left\{ b_i e^{-\bar{l}_i} \left(1 - \frac{(\tilde{r}_i^T)(b_i e^{l_i^{true}} + \tilde{r}_i^T)}{(b_i e^{-\bar{l}_i} + \tilde{r}_i^T)^2} \right) \right\} \mathbf{G} + \beta \mathbf{R}.$$

Here \mathbf{R} is the Hessian of the penalty function and includes the modified penalty weights [106] and $\tilde{r}_i^T = r_i^T + \bar{\kappa}_i$.

In this case, the variance of the sum over a region can be predicted with a formula similar to (7.19). The emission part of the formula is now

$$w^E = \mathbf{B}'\mathcal{D} \{ e^{\bar{l}_i} \} \mathbf{P}'u,$$

and q_i remains the same as (7.16). The transmission part changes a lot due to statistical method as term w^T should be changed to:

$$w^T = \mathbf{G}\mathbf{H}^{-1}\mathbf{G}'\mathbf{B}'_2\mathcal{D} \{ \bar{z}_i e^{\bar{l}_i} \} \mathbf{P}'u,$$

and the s_i term should be

$$s_i = \frac{(b_i e^{-\bar{l}_i})^2 (b_i e^{-l_i^{true}} + \tilde{r}_i^T)}{(b_i e^{-\bar{l}_i} + \tilde{r}_i^T)^2}.$$

The most computationally intensive part in this computation is the part where $\mathbf{H}^{-1}v^*$ should be computed for $v^* = \mathbf{G}'\mathbf{B}'_2\mathcal{D} \{ \bar{z}_i e^{\bar{l}_i} \} \mathbf{P}'u$. This operation can be performed by solving the equation:

$$\mathbf{H}x = v^* \tag{7.24}$$

using iterative methods such as conjugate gradient. Also, we assume the mean for $\hat{\gamma}$ is now, $\bar{\gamma} = e^{\bar{l}_i}$.

These variance predictions are useful, because they do not require hundreds of empirical reconstructions of data [37]. However, they require knowing the true parameters and noiseless sinograms. For real data, these are not known, but one can still get a good approximation of variances by replacing the true parameters by their noisy counterparts [37].

Finally, the optimal time fraction for the emission scan can be found by minimizing the variance in (7.19) with respect to the emission scan time when total scan time is fixed. For the QPL method, the simple analysis yields

$$\tau_{opt}^E = \tau^{total} \frac{v^E - \sqrt{v^E v^T}}{v^E - v^T}.$$

Note that for the conventional method, the above formula is invalid because the v^T term is not independent from the scan time duration τ^T .

7.6 Results

We have done a series of simulations to test the proposed variance predictions and to find the optimal scan times under a total scan time constraint. We used two 2-D images corresponding to an attenuation map and an emission image to generate noisy transmission and emission data with 150000 and 50000 counts per minute respectively. The true images are shown in Figure 5.4. The transmission scan had 5% randoms and an emission contamination of 5%. Emission scan had 10% randoms. Randoms rates were assumed to be constant. The total scan time was 20 minutes. To obtain empirical standard deviations, 300 realizations were generated for each scan time distribution. The emission images were reconstructed with FBP with a smoothing filter that yields about 9 mm FWHM resolution in the image domain. ACFs were computed using the conventional and quadratic penalized-likelihood statistical methods. The resolutions for the ACFs were matched for these two methods. The standard deviations of these parameters in the reconstructed emission images were found empirically and predicted analytically using the derived formulas. 35 iterations were used to solve (7.24). The predicted and empirical standard deviations are plotted versus emission scan time fraction in the corresponding figures: 1) Figure 7.1: The sum over the heart region; 2) Figure 7.2: The sum over a rectangular region in the heart; 3) Figure 7.3: A single pixel in the heart. The predicted and empirical results agree to a certain level, but there are still discrepancies in the predictions which we believe are due to nonlinearities and nonnegativity constraints in the image reconstruction process. Better strategies should be developed for more accurate predictions. Note that the prediction will be more challenging for nonquadratic penalties which result in multimodal probability distribution functions for image estimates [40]. The statistical method not only reduces the overall variance, but also yields a larger optimum emission scan time fraction (about 40%) as compared to the conventional

method (about 30%). The standard deviation is reduced by about 15-20% in the statistical method as compared to the conventional method.

The predictions seem to match the empirical data for the statistical reconstruction, but the predictions for the conventional method seem to underestimate the standard deviations. We conjecture that, the approximations used in deriving the variance formulas causes the mismatch. We are currently working on improving our approximations. Due to highly nonlinear processing of data however, it is likely that there will be some discrepancy between predicted and empirical standard deviation estimates.

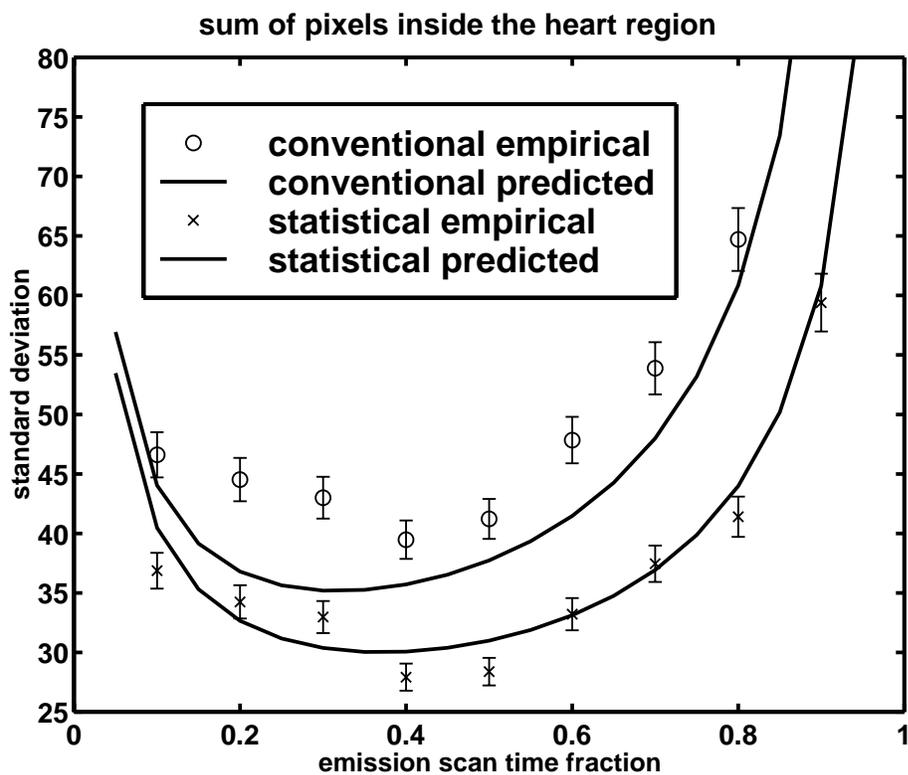


Figure 7.1: Standard deviation of the sum over the heart region estimates versus emission scan time fraction for conventional and statistical ACF computations.

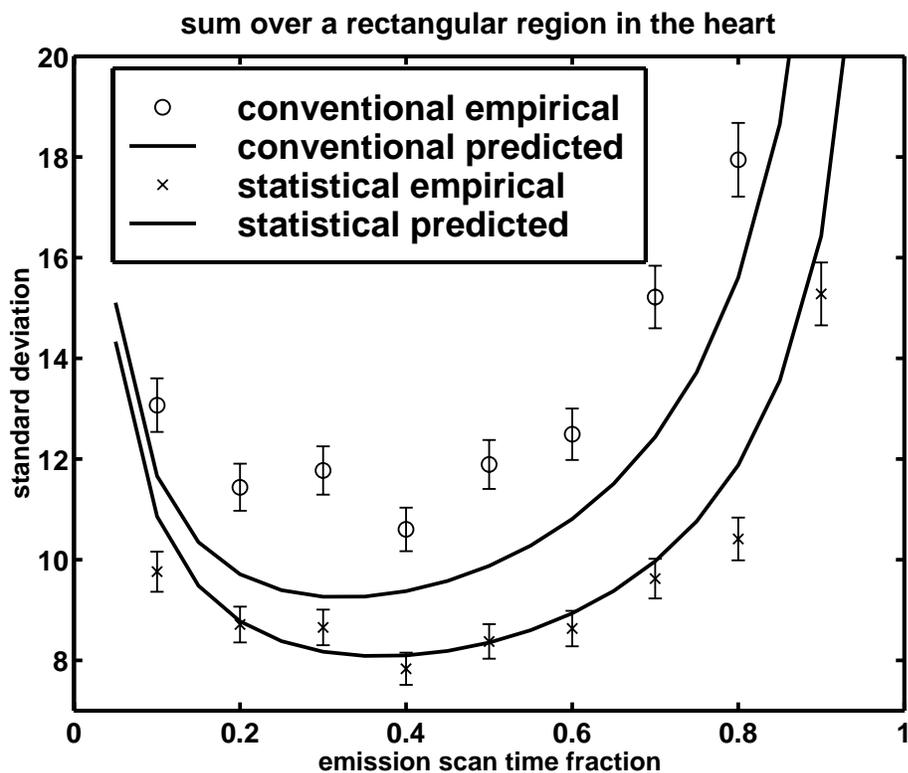


Figure 7.2: Standard deviation estimates of the sum over the a rectangular region in the heart versus emission scan time fraction for conventional and statistical ACF computations.

7.7 Conclusion

We presented new approximate formulas for covariances of reconstructed emission images with conventional and statistical ACF computation for post-injection scans. These formulas can be used to predict the variance of the sum over a region of interest in the final reconstructed emission image instead of expensive empirical reconstructions. These formulas can also be used to determine optimal scan times devoted to emission and transmission scans under a total scan time constraint. Results show that, statistical ACF computation not only reduces the overall standard deviation but also yields higher optimum emission scan time fraction than the conventional method.

We considered only 2D smoothing for the transmission reconstructions. In real

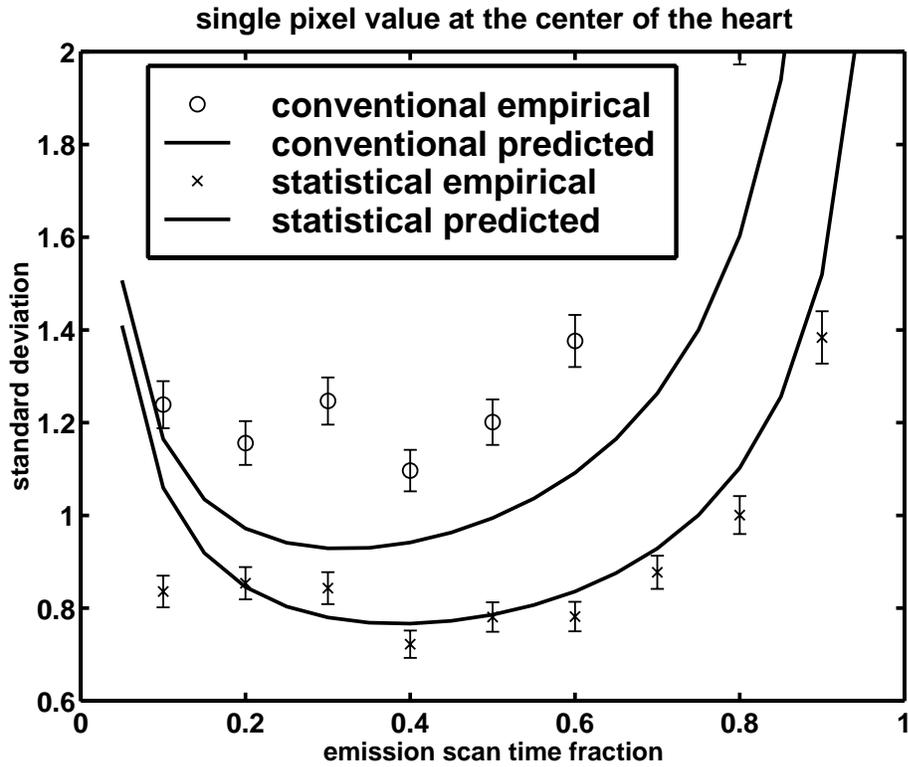


Figure 7.3: Standard deviation estimates of a single pixel value versus emission scan time fraction for conventional and statistical ACF computations.

PET transmission scans, usually axial smoothing is employed to further reduce the noise in the attenuation correction factors. In addition, 3D regularization penalties can be used in PL reconstructions. If such z-smoothing or 3D regularization is used, then the optimal scan time fractions and variance levels will change possibly reducing the noise contribution from the transmission scan and increasing the time that should be devoted to the emission scan.

There is some discrepancy between the empirical and predicted variance approximations. The differences might be due to linearization of nonlinear operations such as division and exponentiation, slow convergence of (7.24) and nonnegativity constraints for PL image reconstruction and others. The covariance approximations for statistical method work well for quadratic penalties, but not for non-quadratic penalties. The extension to non-quadratic penalties is left as future work.

CHAPTER 8

Conclusions and Future Work

8.1 Summary

We introduced new algorithms for penalized-likelihood transmission image reconstruction based on paraboloidal surrogates principle. The PSCD algorithms introduced in Chapter 4 converge very fast to the optimum. We introduced the optimum curvature which yields fastest convergence while still guaranteeing monotonicity. Pre-computed curvatures do not guarantee monotonicity, but practically they almost always converge with a benefit of reducing the CPU time further. The SPS algorithm described in Chapter 5 is a separable simultaneous update algorithm which can also be made monotonic by choosing the optimum curvature. SPS is parallelizable and easy to implement. The OSTR algorithm which is obtained by applying the ordered subsets principle to SPS algorithm accelerates it for both maximum likelihood and penalized likelihood problems. OSTR is a fast and easy to implement algorithm for transmission tomography. We recommend to use monotonic PSCD algorithms for guaranteed convergence to the PL solution. However, if very fast methods are required, then parallelizable OSTR algorithms can be used to reconstruct acceptable images very close to the true optimum.

We analyzed post-injection transmission and emission scans in Chapter 6. The statistical model including the emission contamination is developed and image reconstruction methods are described for this problem. The results indicate statistical

methods outperform conventional methods in bias-variance tests conducted by Monte Carlo simulations.

A noise analysis on emission images reconstructed using the ACFs from noisy transmission data and noisy emission data is described in Chapter 7. The effects of transmission and emission noise on the emission image variance is attempted to be separated to enable scan time optimization. The variance predictions give a general idea but they are not very accurate due to nonlinear processing and nonnegativity constraints.

8.2 Conclusions

- We conclude that penalized-likelihood reconstruction of attenuation maps followed by reprojection to correct the emission data for attenuation yields much better emission images as compared to ones that are reconstructed using conventional attenuation correction methods. Especially for low count transmission scans, PL reconstruction is much more preferable to conventional methods.
- One can get fast converging algorithms by making use of the optimization transfer principle in optimization. Iteratively optimizing a complex objective function by finding simpler surrogate functions at each iteration which can be optimized easily results in much faster algorithms.
- In a post-injection transmission scan, including the emission contamination in the statistical model and making use of statistical techniques is superior to subtraction based conventional methods. Statistical methods result in lower noise images.
- The noise propagating from transmission scans to emission images affect reconstructions considerably. Sometimes transmission noise is more dominant than the emission noise in the final image. To reduce this noise, edge preserving penalized-likelihood methods should be employed for transmission scans.

8.3 Future Work

The research presented in this thesis is only part of a long term effort to make statistical image reconstruction methods practical for PET. We focused on developing fast and stable algorithms for transmission tomography. Our efforts can be extended by further analyzing the convergence properties of our algorithms, by applying similar ideas to emission tomography and improving our ideas to find faster algorithms. There are many improvements and analysis to be made for the post-injection problem and noise analysis chapter. We itemize some possible future work as follows:

- Penalized likelihood objective for transmission tomography is nonconvex when there are background counts in the data. However, it is likely that for a certain range of background parameters, the objective function might be unimodal which would imply global convergence for our monotonic PSCD algorithm. Our practical experience suggests that the objective function is most probably unimodal for small background percentages. This problem requires further investigation to determine a range of background values at which the objective function remains unimodal.
- The algorithms introduced in Chapter 4 and Chapter 5 of this thesis were specifically targeted towards transmission scans. However, it is possible to apply the ideas to emission scans as well. Both PSCD and OSTR algorithms can be adopted to emission image reconstruction [41]. There are some difficulties when $r_i^E = 0$ in the emission scans which can be overcome with a modification of the algorithms.
- Kudo *et al.* [67] claim to prove convergence for a general class of algorithms that cover OSTR as a special case when appropriate relaxation schemes are used as mentioned in Chapter 5. It might be possible to obtain a convergent OSTR algorithm by using their idea.
- An ideal algorithm for transmission or emission tomography should be fast, simple to implement and convergent. It should enforce nonnegativity easily,

work with any type of system matrix and be parallelizable. This ideal algorithm is still a future goal for research in this area.

- Further analysis of alternating sequential and joint estimation approaches introduced in Chapter 6 for post-injection scans is required. The effects of the regularization parameters are different in joint estimation than sequential estimation. A better understanding of their effects will enable one to choose those parameters optimally. Bias-variance analysis and a thorough comparison of joint and sequential methods for post-injection scans should be performed for better assessment.
- Resolution mismatch between survival probabilities and emission data result in artifacts in the reconstructed images as mentioned in Chapter 6. Even when penalized-likelihood attenuation maps are used, these artifacts exist. In noisy reconstructions, these artifacts are hard to notice, but they might result in misdiagnosis. It is desirable to find new methods to overcome this problem.
- Variance predictions described in Chapter 7 can be improved. The following three reasons cause problems in variance estimations: 1) propagation of variance through nonlinear operations such as division, multiplication and exponentiation, 2) nonnegativity constraints, and 3) nonquadratic penalties in a PL reconstruction. Techniques focusing on each one of these problems are required for more accurate variance predictions. The predictions can be generalized to include the effects of axial post-smoothing and 3D regularization penalty as well.

APPENDICES

APPENDIX A

Proof of Maximum Curvature

We prove in this appendix that the maximum second derivative of $h_i(l)$ for $l \geq 0$ is given by (4.14). We drop the subscript i for simplicity. We assume $b > 0$, $y \geq 0$, and $r \geq 0$ throughout these appendices.

The form of the h functions is critical in the following. The second and third derivatives of the function h in (4.3) are:

$$\ddot{h}(l) = \left(1 - \frac{yr}{(be^{-l} + r)^2}\right) be^{-l}, \quad (\text{A.1})$$

$$h^{(3)}(l) = \left(yr \left[\frac{-be^{-l} + r}{(be^{-l} + r)^3}\right] - 1\right) be^{-l}. \quad (\text{A.2})$$

The intuition behind the maximum curvature is as follows. The generalized mean value theorem for twice continuously differentiable functions which is given in [17] at page 228, states that the maximum second derivative satisfies (4.9). It is possible to visualize this by considering the first derivative of a twice continuously differentiable function. The first derivative of $q(l; l^c)$ function is a line which has a slope that is equal to the maximum slope in $\dot{h}(l)$ for $l > 0$. When plotted together, we see that $\dot{q}(l; l^c)$ lies above $\dot{h}(l)$ for $l > l^c$ and lies below it for $l < l^c$. This is illustrated in Figure A.1. Furthermore, for the $h(l)$ function we consider for transmission tomography, the second derivative is either a decreasing function (lemma 1) or it has a single local minimum (lemma 2) so that the maximum value is attained either at $l = 0$ or $l = \infty$. The second derivative at ∞ is zero, so the maximum second derivative can be given by $[\ddot{h}(0)]_+$. Next, we provide the details.

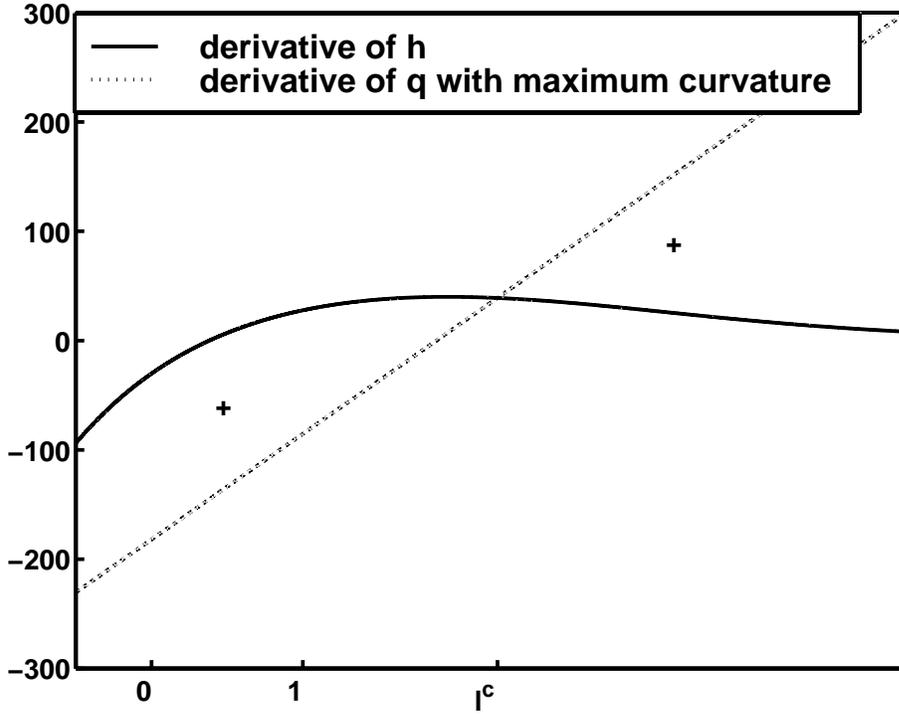


Figure A.1: Plot of $\dot{q}(l; l^c)$ and $\dot{h}(l)$ for maximum curvature.

First, we prove two lemmas about properties of the h functions. These lemmas are used for the proofs in Appendix B as well.

Lemma 1 *The following are equivalent for $h(l)$ defined in (4.3):*

- (E1) $r = 0$ or $r \geq y$,
- (E2) h is strictly convex,
- (E3) \dot{h} is strictly concave,
- (E4) \dot{h} is monotonically increasing,
- (E5) \ddot{h} is monotonically decreasing.

Proof: Since h is three times continuously differentiable, h is strictly convex if and only if $\ddot{h} > 0$ and \dot{h} is strictly concave if and only if $h^{(3)} < 0$. Clearly, $\ddot{h} > 0$ if and only if \dot{h} is monotonically increasing. So, (E2) \iff (E4). For similar reasons (E3) \iff (E5).

If $r = 0$ or $r \geq y$, then $yr < (be^{-l} + r)^2$, so from (A.1) $\ddot{h}(l) \geq 0$, $\forall l$. Thus, (E1) \Rightarrow (E2).

To prove (E1) \Rightarrow (E3), from (A.2), it suffices to show that $(be^{-l} + r)^3 > yr(-be^{-l} + r)$. But this is trivial since $r^3 \geq yr^2$ under the conditions (E1).

To prove the opposite, if $r \neq 0$ and $y > r$, then one can easily show that $\ddot{h}(l)$ and $-h^{(3)}(l)$ can take negative values for sufficiently large l considering (A.1) and (A.2). So, (E2) \Rightarrow (E1) and (E3) \Rightarrow (E1). \blacksquare

Lemma 2 *When $y > r$ and $r \neq 0$, the nonconvex function \dot{h} has the following properties:*

- (P1) \dot{h} is continuously differentiable,
- (P2) \dot{h} has exactly one critical point l^* , i.e. $\ddot{h}(l^*) = 0$ and l^* is a local maximizer of $\dot{h}(l)$,
- (P3) \dot{h} is strictly concave and monotone increasing for $l < l^*$,
- (P4) \dot{h} is monotone decreasing for $l > l^*$,
- (P5) \ddot{h} has exactly one critical point l^z , i.e. $h^{(3)}(l^z) = 0$ and l^z is a local minimizer of $\ddot{h}(l)$.

Proof: (P1) is obvious from (4.20) and (A.1).

In the nonconvex case, the equation $\ddot{h}(l) = 0$ has exactly one solution in \mathbb{R} , $l^* = \log\left(\frac{b}{\sqrt{yr} - r}\right)$. Since $h^{(3)}(l^*) = -2\frac{(\sqrt{yr} - r)^2}{\sqrt{yr}} < 0$, l^* is a local maximum, proving (P2).

Solutions to the equation $h^{(3)}(l) = 0$ are the roots of a cubic polynomial in the variable $t = be^{-l}$ which has only one real solution. The real root is negative when h is convex resulting no solution for l . But, in the nonconvex case the real root is positive and results in exactly one solution $l^z = \log(b/(a/3 - yr/a - r))$ where $a = \sqrt[3]{27yr^2 + 3\sqrt{3y^3r^3 + 81y^2r^4}}$. So, $\ddot{h}(l)$ has exactly one critical point. We have shown above that $h^{(3)}(l^*) < 0$ and one can easily see that $h^{(3)}(l) \approx \left(\frac{y}{r} - 1\right) be^{-l} > 0$ for large l . Thus $l^z > l^*$ and $h^{(3)}(l) < 0$ for $l < l^z$. So, $\ddot{h}(l)$ is monotonically decreasing

for $l < l^z$. Also for $l > l^z$, $h^{(3)}(l) > 0$ and $\ddot{h}(l)$ is monotonically increasing. This proves that l^z is a local minimum for $\ddot{h}(l)$. Hence, (P5) is proven.

To prove (P3), we have to show $h^{(3)}(l) < 0$ and $\ddot{h}(l) > 0$ for $l < l^*$. But, as we found above $l^* < l^z$ and $h^{(3)}(l) < 0$ for $l < l^z$. Also, $\ddot{h}(l) > 0$ for $l < l^*$ since l^* is the only critical point and local maximizer of \dot{h} due to (P2). So, (P3) is also proven.

The function $\ddot{h}(l)$ has exactly one zero crossing l^* from (P2) which is a local maximizer of \dot{h} . Then, $\ddot{h}(l)$ has to be always negative for $l > l^*$ proving (P4). To verify, one can easily see that, $\ddot{h}(l) \approx \left(1 - \frac{y}{r}\right) b e^{-l} < 0$ for large l values. So $\ddot{h}(l) < 0 \forall l > l^*$. ■

The following result follows from (E5) of Lemma 1 for the convex case and from (P5) of Lemma 2 for the nonconvex case.

Corollary 1 *The maximum value for \ddot{h} in the region $[0, \infty)$ is achieved at the end points, i.e.*

$$\begin{aligned} c_i(l_i^n) &= \max_{l \in [0, \infty)} \{\ddot{h}(l)\} \\ &= \max\{\ddot{h}(\infty), \ddot{h}(0)\} \\ &= [\ddot{h}(0)]_+, \\ &= \left[\left(1 - \frac{yr}{(b+r)^2}\right) b \right]_+. \end{aligned}$$

The result follows since $\lim_{l \rightarrow \infty} \ddot{h}(l) = 0$.

APPENDIX B

Proof of Optimum Curvature

In this appendix, we prove that the curvature defined in (4.33) is the optimum curvature that satisfies (4.32), which in turn implies from (4.31) that the choice (4.33) yields the fastest convergence rate.

The intuition behind the optimum curvature is that it should be possible to obtain better (smaller) curvatures than the conservative maximum curvature for this problem. A visualization for the optimum curvature is given in Figure B.1. The slope of the first derivative $\dot{q}(l; l^c)$ is adjusted to make it possible to have $q(l; l^c)$ lie above $h(l)$ for $l > 0$. This is achieved when the areas between the $\dot{q}(l; l^c)$ function and the $\dot{h}(l)$ function for $l < l^c$ cancels each other so that the difference integral (which is the difference $q(l; l^c) - h(l)$) remains positive for $l < l^c$. We provide details in the following.

We first prove two lemmas about strictly concave functions.

Lemma 3 *A one-dimensional line $l(x) = ax + b$ can intersect a strictly concave (or strictly convex) function $f(x)$ at most twice.*

Proof: Suppose $l(x_i) = f(x_i)$ at points $x_1 < x_2 < x_3$. Then since $f(x)$ is strictly concave, $f(x) > l(x)$ for $x \in (x_1, x_3)$, which contradicts the initial assumption that $f(x_2) = l(x_2)$. ■

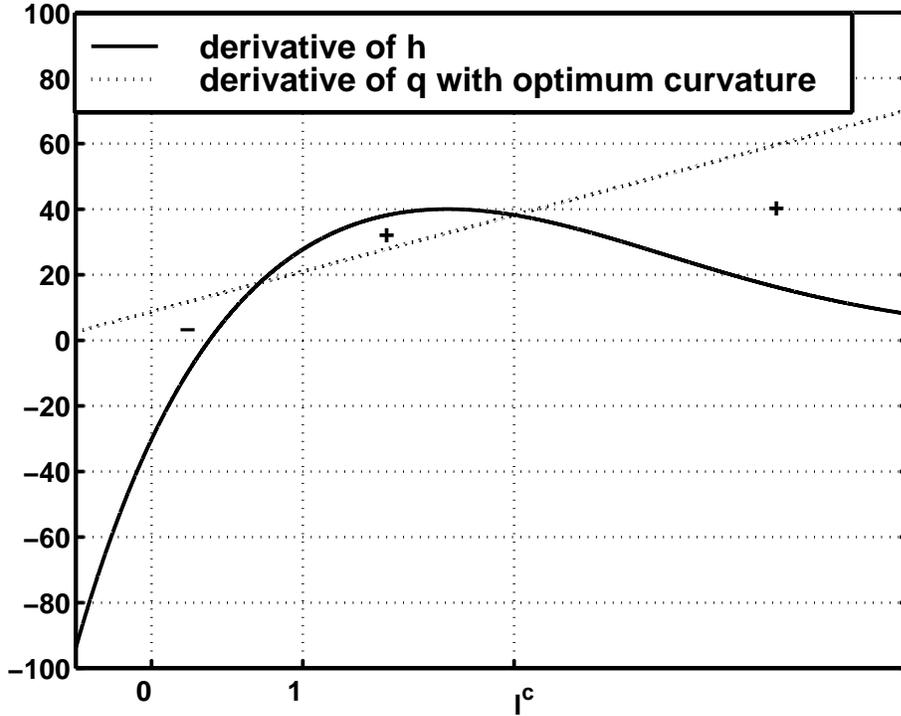


Figure B.1: Plot of $\dot{q}(l; l^c)$ and $\dot{h}(l)$ for optimum curvature.

Lemma 4 *Let $f(x)$ be a one-dimensional strictly concave function, and let $l(x) = ax + b$ be a line that intersects $f(x)$ at the two points $x_1 < x_2$. Then*

$$f(x) < l(x) \text{ for } x \in (-\infty, x_1) \cup (x_2, \infty).$$

Proof: Suppose there exists an $x_3 > x_2$ such that $f(x_3) \geq l(x_3)$. Consider the new line $m(x)$ that intersects $f(x)$ at x_1 and x_3 . Since $m(x_1) = l(x_1)$ and $m(x_3) = f(x_3) \geq l(x_3)$, it follows from the affine form of $l(x)$ and $m(x)$ that $m(x_2) \geq l(x_2) = f(x_2)$, which contradicts the assumption that $f(x)$ is strictly concave. The case $x_3 < x_1$ is similar. ■

For simplicity in this appendix, we drop the subscript i and the dependence on n for the variables. Let $h(l)$ be the marginal negative log-likelihood function defined in (4.3) with derivatives presented in (4.20), (A.1) and (A.2) and let $q(l)$ be the parabolic surrogate function defined in (4.8) with the “optimum curvature” c defined in (4.33). We use l^c to denote the current projection value l_i^n . The reader may visualize the

following proofs by considering the plots of \dot{h} and \dot{q} functions shown in Fig. 4.3.

We define the difference function by:

$$\delta(l) \triangleq q(l) - h(l). \quad (\text{B.1})$$

To show that $q(l) \geq h(l)$ for $l \geq 0$ as required by (4.9), it suffices to show that $\delta(l) \geq 0$. When $l^c = 0$, it is obvious from Appendix A that $\delta(l) \geq 0$. Thus we focus on the case $l^c > 0$ in the following.

Lemma 5 *The following conditions are sufficient to ensure $\delta(l) \geq 0, \forall l \in [0, \infty)$.*

- (C1) $\delta(0) \geq 0$ and $\delta(l^c) = 0$,
- (C2) $\dot{\delta}(l) \geq 0$ for $l \geq l^c$, and
- (C3) either
 - (C31) $\dot{\delta}(l) < 0, \forall l \in [0, l^c)$, or
 - (C32) $\exists l^p \in [0, l^c)$ such that $\dot{\delta}(l) \geq 0$ for $l \in [0, l^p]$ and $\dot{\delta}(l) \leq 0$ for $l \in (l^p, l^c]$.

Proof: Since $\delta(l^c) = 0$

$$\delta(l) = \int_{l^c}^l \dot{\delta}(t) dt. \quad (\text{B.2})$$

- Case $l \geq l^c$. The integrand in (B.2) is nonnegative due to (C2), so $\delta(l) \geq 0$.
- Case $l \in [0, l^c]$. If (C31) is true, then $\delta(l) = \delta(l^c) - \int_l^{l^c} \dot{\delta}(t) dt \geq \delta(l^c) = 0$.

If (C32) holds and $l \in [0, l^p]$, then $\delta(l) = \delta(0) + \int_0^l \dot{\delta}(t) dt \geq \delta(0) \geq 0$ by (C1).

Likewise if (C32) holds and $l \in (l^p, l^c]$, then $\delta(l) = \delta(l^c) - \int_l^{l^c} \dot{\delta}(t) dt \geq \delta(l^c) = 0$ again by (C1).

Hence, $\delta(l) \geq 0 \forall l \geq 0$ under the above conditions. ■

We now establish the conditions of Lemma 5. (C1) follows directly from the definition (4.32), so we focus on (C2) and (C3) below. We first treat the case where $h(l)$ is strictly convex.

Lemma 6 *If $h(l)$ is a convex function and $\dot{h}(l)$ is concave for $l \geq 0$, then the difference function $\delta(l)$ in (B.1) with the curvature c defined in (4.33) satisfies conditions (C2) and (C32) in Lemma 5. Furthermore, $c > 0$.*

Proof: It is trivial to show that the conditions (E2) through (E5) of Lemma 1 hold in this case for $l \geq 0$. First we prove $c > 0$. Suppose $c = 0$, so \dot{q} is a constant. Since $\dot{h}(l)$ is increasing by (E4) in Lemma 1 and $\dot{q}(l^c) = \dot{h}(l^c)$, it is obvious that $\dot{q}(l) > \dot{h}(l)$, $\forall l \in [0, l^c)$, so $\delta(0) = -\int_0^{l^c} \dot{\delta}(t) dt < 0$ contradicting (C1). So, $c > 0$ in this case and $\delta(0) = 0$ by design.

To prove (C32), consider \dot{h} . The line \dot{q} cannot intersect the strictly concave \dot{h} at more than two points due to Lemma 3. We know that $\dot{\delta}(l^c) = 0$, thus l^c is an intersection point. We have $\delta(0) = 0$ and $\delta(l^c) = 0$ by definition. From mean value theorem, there must be another intersection point $l^p \in [0, l^c)$ such that $\dot{\delta}(l^p) = 0$. We know by Lemma 3 that there cannot be any additional points where $\dot{\delta}(l) = 0$. $\dot{\delta}(l) < 0$ for $l \in (l^p, l^c)$ due to concavity of \dot{h} and $\dot{\delta}(l) > 0$ for $l \in [0, l^p)$ due to Lemma 4. (C32) is proven.

To prove (C2), apply Lemma 4 to the strictly concave function \dot{h} with two points l^p and l^c as the intersection points of the line with the curve. ■

We now consider the realistic nonconvex case.

Lemma 7 *Let $h(l)$ be a nonconvex function with its derivative \dot{h} satisfying properties (P1), (P2) and (P3) in Lemma 2. The difference function $\delta(l)$ defined in (B.1) with the curvature defined in (4.33) satisfies (C2) and (C3) in Lemma 5.*

Proof:

The reader can refer to Fig. 4.3 for representative plots of h and its first derivative. Note that in Lemma 2, (P2) \Rightarrow (P4) directly.

Consider these two cases where l^* is defined as in Lemma 2:

- CASE $l^c < l^*$.

In this case, by (P3) of Lemma 2, l^c is in a concave increasing region. By Lemma 6, (C32) holds as well as the fact that $c > 0$. To prove (C2), we use property

(P4), that \dot{h} is a decreasing function for $l > l^*$. So, since $\dot{q}(l^*) > \dot{h}(l^*)$ (as for (C2) in Lemma 6 again) and $c \geq 0$, $\dot{q}(l) > \dot{h}(l), \forall l \geq l^c$.

- CASE $l^c \geq l^*$. Since by (C1), $\delta(l^c) - \delta(0) = \int_0^{l^c} \dot{\delta}(t) dt \leq 0$, $\dot{\delta}(l) = \dot{q}(l) - \dot{h}(l)$ cannot always be nonnegative over the interval $[0, l^c]$. So, either $\dot{q}(l) < \dot{h}(l), \forall l \in [0, l^c]$ or \dot{q} intersects \dot{h} ($\dot{\delta}(l) = 0$) at least once in $[0, l^c]$. If the former case occurs, (C31) holds by definition. If the latter case occurs, then we have to prove that (C32) holds, *i.e.* there is no more than one point at which \dot{q} intersects \dot{h} in $[0, l^c]$. Since $c \geq 0$ and \dot{h} is decreasing in the region $l > l^*$, the intersection point(s) $l^p < l^*$. We cannot apply Lemma 3 here to prove that there is no other intersection point, but we can use Lemma 4 to prove it. Assume there is another intersection point. Then, the function $\dot{q} > \dot{h}$ in the concave region outside the interval between two intersection points by Lemma 4 which implies $\dot{\delta}(l^*) > 0$ and $\dot{\delta}(l) > 0$ for $l > l^*$. But this would contradict the fact that $\dot{\delta}(l^c) = 0$. So, (C32) must hold.

In this case, the fact that $c \geq 0$ is enough to prove (C2), since \dot{h} is decreasing in this region. ■

Theorem 1 *Let $h(l)$ be a one-dimensional function that satisfies either of the following:*

- (H1) $h(l)$ is strictly convex and $\dot{h}(l)$ is strictly concave in the feasible region $l \geq 0$, or
- (H2) $\dot{h}(l)$ satisfies (P1), (P2) and (P3) of Lemma 2.

Then the curvature defined in (4.33) satisfies the optimality condition in (4.32).

Proof: For h functions that satisfy conditions (H1), Lemma 6 with Lemma 5 prove that the curvature (4.33) satisfies (4.9) for $l^c > 0$. For h functions satisfying conditions in (H2), Lemma 7 and Lemma 5 similarly prove that the curvature (4.33) satisfies (4.9) for $l^c > 0$. The rest of the proof applies to both cases (H1) and (H2).

For $l^c = 0$, c in (4.33) is the maximum second derivative in $[0, \infty)$, and (4.9) is satisfied by mean value theorem as mentioned in Section 4.3.1.

We need to prove that no other nonnegative curvature less than (4.33) satisfies (4.9).

Assume $0 \leq c^* < c$, and let

$$q^*(l) = h(l^c) + \dot{h}(l^c)(l - l^c) + \frac{1}{2}c^*(l - l^c)^2.$$

Obviously c^* can exist only when $c > 0$ since $c = 0$ is the minimum curvature we allow. With $c > 0$, it is obvious from (4.33) that $q(0) = h(0)$. If $l^c > 0$, this clearly implies that $q^*(0) < q(0) = h(0)$ which shows that c^* cannot satisfy (4.9). If $l^c = 0$, then a curvature $c^* < c$ would force \dot{q} to lie under \dot{h} for some small values of l . That is, $\exists \epsilon > 0$ such that $q(l) < h(l)$ for $\epsilon > l > 0$. Thus c^* does not satisfy (4.9) even for $l^c = 0$. ■

Corollary 2 *The “optimum curvature” defined in (4.33) using the marginal negative log-likelihood function $h_i(l)$ defined in (4.3) for the transmission tomography problem satisfies the optimality condition in (4.32) for $b_i > 0, y_i \geq 0, r_i \geq 0$.*

Proof: The function $h_i(l)$ defined in (4.3) satisfies the conditions (H1) or (H2) of Theorem 1 depending on the values of y_i and r_i as shown in Lemmas 1 and 2. Hence Theorem 1 is directly applicable to the transmission tomography problem. ■

BIBLIOGRAPHY

- [1] 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.
- [2] J. Besag, "On the statistical analysis of dirty pictures," *J. Royal Stat. Soc. Ser. B*, vol. 48, no. 3, pp. 259–302, 1986.
- [3] 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.
- [4] D. Böhning and B. G. Lindsay, "Monotonicity of quadratic approximation algorithms," *Ann. Inst. Stat. Math.*, vol. 40, no. 4, pp. 641–63, 1988.
- [5] C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Tr. Im. Proc.*, vol. 2, no. 3, pp. 296–310, July 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] J. A. Browne and T. J. Holmes, "Developments with maximum likelihood X-ray computed tomography," *IEEE Tr. Med. Im.*, vol. 12, no. 2, pp. 40–52, March 1992.
- [8] J. A. Browne and A. R. D. Pierro, "A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomography," *IEEE Tr. Med. Im.*, vol. 15, no. 5, pp. 687–99, October 1996.
- [9] 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.
- [10] C. Byrne, E. Soares, T. S. Pan, S. Glick, M. King, and A. Del-Guerra, "Accelerating the EM algorithm using rescaled block-iterative methods," in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 3, pp. 1752–6, 1996.
- [11] C. L. Byrne, "Block-iterative methods for image reconstruction from projections," *IEEE Tr. Im. Proc.*, vol. 5, no. 5, pp. 792–3, May 1996.
- [12] C. L. Byrne, "Convergent block-iterative algorithms for image reconstruction from inconsistent data," *IEEE Tr. Im. Proc.*, vol. 6, no. 9, pp. 1296–1304, September 1997.

- [13] L. R. Carroll, P. Kretz, and G. Orcutt, "The orbiting rod source: Improving performance in PET transmission correction scans," in *Emission Computed Tomography: Current Trends*, Esser, editor, pp. 235–247, Society of Nuclear Medicine, New York, 1983.
- [14] R. E. Carson, M. E. Daube-Witherspoon, and M. V. Green, "A method for postinjection PET transmission measurements with a rotating source," *J. Nuc. Med.*, vol. 29, no. 9, pp. 1558–1567, September 1988.
- [15] B. Chan, M. Bergström, M. R. Palmer, C. Sayre, and B. D. Pate, "Scatter distribution in transmission measurements with positron emission tomography," *J. Comp. Assisted Tomo.*, vol. 10, no. 2, pp. 296–301, March 1986.
- [16] 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.
- [17] P. G. Ciarlet, *Introduction to numerical linear algebra and optimisation*, Cambridge, Cambridge, 1982.
- [18] C. Comtat, P. E. Kinahan, M. Defrise, C. Michel, and D. W. Townsend, "Fast reconstruction of 3D PET data with accurate statistical modeling," *IEEE Tr. Nuc. Sci.*, vol. 45, no. 3, pp. 1083–9, June 1998.
- [19] M. Dahlbom and E. J. Hoffman, "An evaluation of a two-dimensional array detector for high resolution PET," *IEEE Tr. Med. Im.*, vol. 7, no. 4, pp. 264–272, December 1988.
- [20] M. E. Daube-Witherspoon and R. E. Carson, "Investigation of angular smoothing of PET data," in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 3, pp. 1557–61, 1996.
- [21] M. E. Daube-Witherspoon, R. E. Carson, and M. V. Green, "Post-injection transmission attenuation measurements for PET," *IEEE Tr. Nuc. Sci.*, vol. 35, no. 1, pp. 757–761, February 1988.
- [22] M. E. Daube-Witherspoon, R. E. Carson, Y. Yan, and T. K. Yap, "Scatter correction in maximum likelihood reconstruction of PET data," in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 2, pp. 945–947, 1992.
- [23] 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.
- [24] 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.
- [25] A. R. De Pierro and M. E. B. Yamagishi. *Fast EM-like methods for maximum 'a posteriori' estimates in emission tomography*, 1998.
- [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] M. E. Daube-Witherspoon, and R. E. Carson, "Investigation of angular and axial smoothing of PET data," in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, 1996.
- [28] H. Erdoğan and J. A. Fessler, "Statistical image reconstruction methods for simultaneous emission/transmission PET scans," in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 3, pp. 1579–83, 1996.
- [29] H. Erdoğan and J. A. Fessler, "Accelerated monotonic algorithms for transmission tomography," in *Proc. IEEE Intl. Conf. on Image Processing*, volume 2, pp. 680–4, 1998.
- [30] H. Erdoğan and J. A. Fessler, "Fast monotonic algorithms for transmission tomography," *IEEE Tr. Med. Im.*, 1998. To appear.
- [31] H. Erdoğan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Phys. Med. Biol.*, 1999.
- [32] H. Erdoğan, G. Gualtieri, and J. A. Fessler, "An ordered subsets algorithm for transmission tomography," in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, 1998. Inadvertently omitted from proceedings. Available from web page.
- [33] 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.
- [34] J. A. Fessler, "ASPIRE 3.0 user's guide: A sparse iterative reconstruction library," Technical Report 293, Comm. and Sign. Proc. Lab., Dept. of EECS, Univ. of Michigan, Ann Arbor, MI, 48109-2122, July 1995. Available on WWW from <http://www.eecs.umich.edu/~fessler>.
- [35] 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.
- [36] 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.
- [37] 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.
- [38] J. A. Fessler, "Grouped coordinate descent algorithms for robust edge-preserving image restoration," in *Proc. SPIE 3071, Im. Recon. and Restor. II*, pp. 184–94, 1997.
- [39] 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.
- [40] J. A. Fessler, H. Erdoğan, and W. B. Wu, "Exact distribution of edge-preserving MAP estimators for linear signal models with Gaussian measurement noise," *IEEE Tr. Im. Proc.*, 1998. Submitted.

- [41] J. A. Fessler and H. Erdoğan, "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.
- [42] 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.
- [43] 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.
- [44] 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.
- [45] 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.
- [46] 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.
- [47] E. P. Ficaro, J. A. Fessler, P. D. Shreve, J. N. Kritzman, P. A. Rose, and J. R. Corbett, "Simultaneous transmission/emission myocardial perfusion tomography: Diagnostic accuracy of attenuation-corrected 99m-Tc-Sestamibi SPECT," *Circulation*, vol. 93, no. 3, pp. 463–73, February 1996.
- [48] 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.
- [49] G. Germano and E. J. Hoffman, "Investigation of count rate and deadtime characteristics of a high resolution PET system," *J. Comp. Assisted Tomo.*, vol. 12, no. 5, pp. 836–846, September 1988.
- [50] M. Goitein, "Three-dimensional density reconstruction from a series of two-dimensional projections," *Nucl. Instr. Meth.*, vol. 101, no. 15, pp. 509–18, June 1972.
- [51] 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.
- [52] P. J. Green, "On use of the EM algorithm for penalized likelihood estimation," *J. Royal Stat. Soc. Ser. B*, vol. 52, no. 3, pp. 443–452, 1990.
- [53] 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. 458–4662, 1989.

- [54] 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.
- [55] T. Hebert, R. Leahy, and M. Singh, "Fast MLE for SPECT using and intermediate polar representation and a stopping criterion," *IEEE Tr. Nuc. Sci.*, vol. 35, no. 1, pp. 615–619, 1988.
- [56] 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.
- [57] P. K. Hooper, S. R. Meikle, S. Eberl, and M. J. Fulham, "Validation of post injection transmission measurements for attenuation correction in neurologic FDG PET studies," *J. Nuc. Med.*, vol. 37, pp. 128–36, 1996.
- [58] 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.
- [59] P. J. Huber, *Robust statistics*, Wiley, New York, 1981.
- [60] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Tr. Med. Im.*, vol. 13, no. 4, pp. 601–609, December 1994.
- [61] 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.
- [62] 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.
- [63] A. C. Kak and M. Slaney, *Principles of computerized tomographic imaging*, IEEE Press, New York, 1988.
- [64] C. Kamphuis and F. J. Beekman, "Accelerated iterative transmission CT reconstruction using an ordered subsets convex algorithm," *IEEE Tr. Med. Im.*, vol. 17, no. 6, pp. 1101–1105, December 1998.
- [65] J. S. Karp, G. Muehlehner, H. Qu, and X. H. Yan, "Singles transmission in volume-imaging PET with a ^{137}Cs source," *Phys. Med. Biol.*, vol. 40, no. 5, pp. 929–44, May 1995.
- [66] J. T. Kent and C. Wright, "Some suggestions for transmission tomography based on the EM algorithm," in *Stochastic Models, Statistical Methods, and Algorithms in Im. Analysis*, M. P. P. Barone, A. Frigessi, editor, volume 74 of *Lecture Notes in Statistics*, pp. 219–232, Springer, New York, 1992.
- [67] H. Kudo, H. Nakazawa, and T. Saito, "Convergent block-iterative method for general convex cost functions," in *Proc. of 1999 Int. Meeting on Fully Three Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, pp. 247–250, 1999.

- [68] 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, June 1991.
- [69] K. Lange, M. Bahn, and R. Little, "A theoretical study of some maximum likelihood algorithms for emission and transmission tomography," *IEEE Tr. Med. Im.*, vol. 6, no. 2, pp. 106–114, June 1987.
- [70] 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.
- [71] 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.
- [72] R. Leahy and X. H. Yan, "Incorporation of anatomical MR data for improved functional imaging with PET," in *Information Processing in Medical Im.*, A. Colchester and D. Hawkes, editors, pp. 105–20, Wiley-Liss, 1991.
- [73] S.-J. Lee, A. Rangarajan, and G. Gindi, "Bayesian image reconstruction in SPECT using higher order mechanical models as priors," *IEEE Tr. Med. Im.*, vol. 14, no. 4, pp. 669–80, December 1995.
- [74] E. M. Levitan and G. T. Herman, "A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography," *IEEE Tr. Med. Im.*, vol. 6, no. 3, pp. 185–192, September 1987.
- [75] W. K. Luk, W. M. Digby, and M. E. Casey, "An analysis of correction methods for emission contamination in PET post-injection transmission measurement," in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 4, pp. 1633–37, 1995.
- [76] A. Macovski, *Medical imaging systems*, Prentice-Hall, New Jersey, 1983.
- [77] S. H. Manglos, G. M. Gagne, A. Krol, F. D. Thomas, and R. Narayanaswamy, "Transmission maximum-likelihood reconstruction with ordered subsets for cone beam CT," *Phys. Med. Biol.*, vol. 40, no. 7, pp. 1225–41, July 1995.
- [78] S. R. Meikle, D. L. Bailey, P. K. Hooper, S. Eberl, B. F. Hutton, W. F. Jones, R. R. Fulton, and W. J. Fulham, "Simultaneous emission and transmission measurements for attenuation correction in whole body PET," *J. Nuc. Med.*, vol. 36, pp. 1680–8, 1995.
- [79] 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.
- [80] S. R. Meikle, M. Dahlbom, S. R. Cherry, and A. Chatziioannou, "Attenuation correction in whole body PET," *J. Nuc. Med. (Abs. Book)*, vol. 33, no. 5, pp. 862, May 1992.
- [81] S. R. Meikle, D. L. Bailey, P. K. Hooper, S. Eberl, B. F. Hutton, W. F. Jones, R. R. Fulton, and M. J. Fulham, "Simultaneous emission and transmission measurements

- for attenuation correction in whole body PET,” *J. Nuc. Med.*, vol. 36, no. 9, pp. 1680–1688, September 1995.
- [82] C. C. Meltzer, R. N. Bryan, H. H. Holcomb, A. W. Kimball, H. S. Mayberg, B. Sadzot, J. P. Leal, H. N. Wagner, Jr., and J. J. Frost, “Anatomical localization for PET using MR imaging,” *J. Comp. Assisted Tomo.*, vol. 14, no. 3, pp. 418–426, May 1990.
- [83] E. U. Mumcuoglu, *Statistical modeling and fast bayesian reconstruction in positron tomography*, PhD thesis, Univ. of Souther California, Los Angeles, CA., December 1994.
- [84] 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.
- [85] 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.
- [86] W. Niethammer, “A note on the implementation of the successive overrelaxation method for linear complementarity problems,” *Numerical Algorithms*, vol. 4, no. 1, pp. 197–200, January 1993.
- [87] J. Nuyts, B. D. Man, P. Dupont, M. Defrise, P. Suetens, and L. Mortelmans, “Iterative reconstruction for helical CT: A simulation study,” *Phys. Med. Biol.*, vol. 43, no. 4, pp. 729–37, April 1998.
- [88] J. M. Ollinger, “Reconstruction-Reprojection processing of transmission scans and the variance of PET Images,” in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 3, pp. 1696–1699, 1991.
- [89] J. M. Ollinger, “Maximum likelihood reconstruction of transmission images in emission computed tomography via the EM algorithm,” *IEEE Tr. Med. Im.*, vol. 13, no. 1, pp. 89–101, March 1994.
- [90] J. M. Ollinger, “Model-based scatter correction for fully 3D PET,” *Phys. Med. Biol.*, vol. 41, no. 1, pp. 153–76, January 1996.
- [91] J. M. Ollinger and J. A. Fessler, “Positron emission tomography,” *IEEE Sig. Proc. Mag.*, vol. 14, no. 1, pp. 43–55, January 1997.
- [92] C. A. Pelizzari, G. T. Y. Chen, D. R. Spelbring, R. R. Weichselbaum, and C. T. Chen, “Accurate three-dimensional registration of CT, PET, and/or MR images of the brain,” *J. Comp. Assisted Tomo.*, vol. 13, no. 1, pp. 20–26, January 1989.
- [93] M. E. Phelps, E. J. Hoffman, N. A. Mullani, and M. M. Ter-Pogossian, “Application of annihilation coincidence detection to transaxial reconstruction tomography,” *J. Nuc. Med.*, vol. 16, no. 3, pp. 210–224, March 1975.
- [94] 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.

- [95] A. J. Rockmore and A. Macovski, "A maximum likelihood approach to emission image reconstruction from projections," *IEEE Tr. Nuc. Sci.*, vol. 23, pp. 1428–1432, 1976.
- [96] A. J. Rockmore and A. Macovski, "A maximum likelihood approach to transmission image reconstruction from projections," *IEEE Tr. Nuc. Sci.*, vol. 24, no. 3, pp. 1929–1935, June 1977.
- [97] S. Saquib, J. Zheng, C. A. Bouman, and K. D. Sauer, "Provably convergent coordinate descent in statistical tomographic reconstruction," in *Proc. IEEE Intl. Conf. on Image Processing*, volume 2, pp. 741–4, 1996.
- [98] 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.
- [99] K. D. Sauer, S. Borman, and C. A. Bouman, "Parallel computation of sequential pixel updates in statistical tomographic reconstruction," in *Proc. IEEE Intl. Conf. on Image Processing*, volume 3, pp. 93–6, 1995.
- [100] 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.
- [101] B. W. Silverman, M. C. Jones, J. D. Wilson, and D. W. Nychka, "A smoothed EM approach to indirect estimation problems, with particular reference to stereology and emission tomography," *J. Royal Stat. Soc. Ser. B*, vol. 52, no. 2, pp. 271–324, 1990.
- [102] R. J. Smith, F. Benard, J. S. Karp, G. Muehllhner, and E. Gualtieri, "Singles transmission scans performed post-injection for quantitative whole body FDG-PET," in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 3, pp. 1527–31, 1996.
- [103] R. J. Smith, J. S. Karp, and G. Muehllhner, "Post injection transmission scanning in a volume imaging PET camera," *IEEE Tr. Nuc. Sci.*, vol. 41, no. 4, pp. 1526–1531, August 1994.
- [104] 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.
- [105] J. A. Sorenson and M. E. Phelps, *Physics in nuclear medicine*, Saunders, Philadelphia, 2 edition, 1987.
- [106] W. Stayman and J. A. Fessler, "Spatially-variant roughness penalty design for uniform resolution in penalized-likelihood image reconstruction," in *Proc. IEEE Intl. Conf. on Image Processing*, volume 2, pp. 685–9, 1998.
- [107] C. Stearns and D. C. Wack, "A noise equivalent counts approach to transmission imaging and source design," *IEEE Tr. Med. Im.*, vol. 12, no. 2, pp. 287–292, June 1993.
- [108] S. C. Strother, M. E. Casey, and E. J. Hoffman, "Measuring PET scanner sensitivity: relating countrates to image signal-to-noise ratios using noise equivalent counts," *IEEE Tr. Nuc. Sci.*, vol. 37, no. 2, pp. 783–788, April 1990.

- [109] M. M. Ter-Pogossian, M. E. Phelps, M. E. Hoffman, and N. A. Mullani, "A positron-emission transaxial tomograph for nuclear imaging (PETT)," *Radiology*, vol. 114, pp. 89–98, 1975.
- [110] M. M. Ter-Pogossian, M. E. Raichle, and B. E. Sobel, "Positron-emission tomography," *Scientific American*, vol. 243, pp. 171–181, October 1980.
- [111] C. J. Thompson, "The problem of scatter correction in positron volume imaging," *IEEE Tr. Med. Im.*, vol. 12, no. 1, pp. 124–132, March 1993.
- [112] 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.
- [113] C. J. Thompson, N. Ranger, A. C. Evans, and A. Gjedde, "Validation of simultaneous PET emission and transmission scans," *J. Nuc. Med.*, vol. 32, no. 1, pp. 154–160, January 1991.
- [114] C. J. Thompson, N. T. Ranger, and A. C. Evans, "Simultaneous transmission and emission scans in positron emission tomography," *IEEE Tr. Nuc. Sci.*, vol. 36, no. 1, pp. 1011–1016, February 1989.
- [115] S. R. Titus, *Improved penalized likelihood reconstruction of anatomically correlated emission data*, PhD thesis, Univ. of Michigan, Ann Arbor, MI, 48109-2122, Ann Arbor, MI., October 1996.
- [116] 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.
- [117] T. G. Turkington, R. E. Coleman, S. F. Schubert, and A. Ganin, "An evaluation of post-injection transmission measurement in PET," *IEEE Tr. Nuc. Sci.*, vol. 41, no. 4, pp. 1538–1544, August 1994.
- [118] 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.
- [119] K. Wienhard, L. Eriksson, S. Grootoink, M. Casey, U. Pietrzyk, and W. D. Heiss, "Performance evaluation of a new generation positron scanner ECAT EXACT," *J. Comp. Assisted Tomo.*, vol. 16, no. 5, pp. 804–813, September 1992.
- [120] E. Z. Xu, N. A. Mullani, K. L. Gould, and W. L. Anderson, "A segmented attenuation correction for PET," *J. Nuc. Med.*, vol. 32, no. 1, pp. 161–165, January 1991.
- [121] 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.
- [122] 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.

- [123] D. M. Young, *Iterative solution of large linear systems*, Academic Press, New York, 1971.
- [124] D. F. Yu and J. A. Fessler, “Mean and variance of coincidence photon counting with deadtime,” *Phys. Med. Biol.*, 1999.
- [125] J. Zheng, S. Saquib, K. Sauer, and C. Bouman, “Functional substitution methods in optimization for Bayesian tomography,” *IEEE Tr. Im. Proc.*, March 1997. Submitted to *IEEE Tr. Image Proc.*
- [126] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, “Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization,” *ACM Tr. Math. Software*, vol. 23, no. 4, pp. 550–60, December 1997.