Statistical Methods for Transmission Image Reconstruction with Nonlocal Edge-Preserving Regularization

by

Feng Yu

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

Doctoral Committee:

Associate Professor Jeffrey A Fessler, Chair
Professor Alfred O Hero III
Associate Professor Douglas C Noll
Professor W Leslie Rogers

ABSTRACT

Statistical Methods for Transmission Image Reconstruction with Nonlocal Edge-Preserving Regularization

by

Feng Yu

Chair: Jeffrey A Fessler

Tomographic image reconstruction using statistical methods can improve image quality over the conventional filtered backprojection (FBP) method. The effectiveness of a statistical image reconstruction method depends on its three principal components: the statistical measurement model, the regularization method, and the iterative algorithm for maximizing the objective function. This dissertation contributes new methodology and/or analysis to each of these three components, emphasizing PET and SPECT transmission scans, which are essential for accurate attenuation correction in emission tomography.

The first part considers edge-preserving regularization. We propose an objective function that incorporates *nonlocal* boundary information. We use an alternating minimization scheme with deterministic annealing to minimize our new objective function; we use variational techniques implemented using level sets to perform boundary extraction. We compare the bias/variance tradeoff of this novel algorithm with a penalized likelihood (with local Huber roughness penalty) algorithm.

The second part analyzes the effect of deadtime on the counting statistics of detectors. We present a new way of analyzing the moments of the counting process for a counter system affected by various models of deadtime related to PET and SPECT imaging. We derive simple and exact expressions for the first and second moments of the number of

recorded events under various models, for both singles counting and coincidence counting. From this analysis, we study the suitability of the Poisson statistical model assumed in most statistical image reconstruction algorithms.

The third and final part considers the problem of reconstructing images for a certain transmission imaging geometry, where the transmitted "beams" of photons overlap on the detector, such that a detector element may record photons that originated in different sources or source locations and thus traversed different paths through the object. We propose a new algorithm for statistical image reconstruction of attenuation maps that explicitly accounts for overlapping beams in transmission scans. The algorithm is guaranteed to monotonically increase the objective function at each iteration. The availability of this algorithm enables the possibility of deliberately designing systems with increased beam overlap so as to increase count rates.

I dedicate this thesis to my family for their love, patience, and support during my doctoral

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 to and respecting my ideas throughout my doctoral work. It was his financial and professional support that enabled me to finish this work.

I also extend my gratitude to my colleagues: Idris Elbakri, Hakan Erdogan, Jeongtae Kim, Robinson Piramuthu, Web Stayman, Nan Sotthivirat, and Mehmet Yavuz with whom I shared the same or neighboring offices.

In addition, I would also like to thank the Mathematics department at the University of Michigan, for providing so many interesting and challenging courses. In particular, I would like to thank Prof. Igor Dolgachev and Prof. Dror Varolin of the University of Michigan, and Prof. Dr. Dietmar Vogt of the University of Wuppertal, Germany, for providing valuable guidance on mathematical matters.

Finally, I thank my family, my parents, my aunts, and my grandma for their loving support and patience. I thank them for the sacrifice they made during this hard-working, yet fun-filled period in my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

APPENDIX

CHAPTER 1

Introduction/Contributions

In statistical image reconstructions for ECT (emission computed tomography), one must consider the following three important aspects of the problem: the statistical model, the regularization method, and the iterative maximization/minimization algorithm. An accurate statistical model is a prerequisite for a good reconstruction. Secondly, the user usually has some *a priori* information about the image to be reconstructed; this information can be incorporated into the regularization method to produce "reasonable-looking" images. Finally, a good algorithm is necessary to ensure that a good reconstructed image can be obtained in a tolerable amount of time.

An important physical factor one must consider when reconstructing emission maps is the effect of attenuation. Transmission scans are usually performed to estimate the attenuation parameters, which are then used to correct the emission scans for attenuation effects.

In the first part of our work (Chapter 3), we focus on the regularization problem in ECT transmission reconstructions. Because this reconstruction problem is usually ill-posed, a roughness penalty is imposed on the solution. Conventional penalty functions penalize the differences between the values of neighboring pixels; the larger the difference, the higher the penalty (one such penalty is the square of the difference, henceforth called the "quadratic penalty"); thus this penalty encourages a globally smooth image. However, the image in the transmission reconstruction problem is not always globally smooth. Instead, it usually consists of smooth regions separated by sharper boundaries. The global roughness penalty over-smoothes the edges. Thus various *edge-preserving* penalties have been

proposed. The main idea behind these penalties is: large differences between values of neighboring pixels are usually due to a sharp edge, while small differences are assumed to be primarily due to noise. Thus these edge-preserving penalties penalize large differences between neighboring pixels less severely than the quadratic penalty. But these schemes still only rely on information from *local* neighborhood to determine the existence of edges, not using the *nonlocal* information that the image is made up of regions separated by smooth boundaries. In this part of our work, we propose an objective function that incorporates *nonlocal* boundary information into our regularization. We use an alternating minimization scheme with deterministic annealing to minimize our new objective function. Because of the nonconvexity of our objective function, we use deterministic annealing in the hope that we will reach a "better" local minimum.

In the second part of our work (Chapters 4 and 5), we focus on the effect of deadtime on the counting statistics of detectors. The consideration is especially important in post-injection transmission scans for PET. Post-injection protocols for PET are more convenient for the patient than conventional PET protocols. However, one has to deal with the contamination of the transmission scan with undesirable emission counts. Because of this contamination, the singles rate will greatly increase under the post-injection protocol, thus deadtime becomes an even more important issue. Since the singles rate is higher than in the conventional protocols, the percentage of lost events will be higher, too. Traditional methods of correcting for deadtime mainly use methods of moments, *i.e.*, the true counts are estimated from the observed counts by solving the equation relating the mean of the true counts and mean of the observed counts; the estimated true counts are then used to reconstruct the image. From a statistical point of view, the estimated true counts are not Poisson distributed, and using a Poisson likelihood in the reconstruction may be suboptimal. We propose a new approach for deriving the moments of the counting process, and analyze how non-Poisson the counting process really is from its first and second moments.

In the third part of our work (Chapter 6), we design a reconstruction algorithm for a source/detector configuration encountered in SPECT transmission scans. The system configuration has several fixed-position collimated line sources opposing a parallel-beam collimator. The main difficulty with such a configuration is that a photon registered by one

detector may originate from two or more sources, and there is no previous algorithm that properly models this effect. We propose a statistical model that takes this into account, and design a reconstruction algorithm for this model that is monotonic and convergent (to a local minimum).

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

- Proposed a novel nonlocal regularization method for ECT transmission scans, which incorporates region and boundary information, and preserves edges.

- Proposed a new approach for deriving the moments of the counting process by realistic detectors, *i.e.*, affected by deadtime.

- For single photon counting, derived exact and closed-form expressions for mean and variance of the counting process affected by type II or type III deadtime.

- For coincidence photon counting, derived exact (or approximate) and closed-form expressions for mean and variance of the coincidence counting process, for both ideal detectors and non-ideal (type II deadtime) detectors.

- Analyzed the suitability of the commonly-assumed Poisson model for various types of photon counting systems.

- Designed a reconstruction algorithm for a source/detector configuration where there are several fixed-position collimated line sources.

- Analyzed the optimal collimation angle in such a system.

These contributions have been summarized in three manuscripts submitted for publication[93, 94, 95], which form the basis of this dissertation.

CHAPTER 2

Overview of PET Physics

Emission Computed Tomography (ECT) is a medical imaging technique that gives functional information about physiological processes, as opposed to CT or MRI which usually give structural and anatomical information. In ECT, a radioactive compound (radio-tracer) is injected into the body; this compound radio-labels certain substances, such as glucose. After the radio-tracer has distributed throughout the body, an image that indicates where the radio-tracer has distributed in higher or lower concentration can be made. In the case of glucose, the result will be an image related to glucose metabolism. The most popular emission tomography techniques are Positron Emission Tomography (PET) and Single Photon Emission Tomography (SPECT).

In PET, the radioisotope emits a positron which annihilates with an electron after traveling up to a few millimeters. This annihilation event produces two 511 keV gamma photons which travel at nearly $180°$ from one another. If both these photons travel coplanar with the coincidence detector ring that is around the body and both survive until reaching the detectors, the sinogram bin corresponding to the line joining the two detectors (line of response or LOR) is incremented. The detector circuits check for whether the two photons have arrived within a coincidence window of a short duration Δ.

The detectors in most current PET cameras are made from blocks of large crystals of BGO (bismuth-gernamate) coupled to a group of photo-multiplier tubes (PMTs). The detector efficiency for a detector pair is defined as the probability that a photon pair arriving at the crystals actually get detected. This efficiency varies from detector to detector. For

example, detector efficiency is lower for a crystal located at 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. Furthermore, it takes some time (about 2 μs) to process a photon detection, during which any new events are not detected. The period is called "dead time".

The measurements are usually organized in an array with the projection angles in the vertical axis and the projection bins in the horizontal axis. This array is called the sinogram, since a single point in the original image approximately traces a sine wave in the projection domain. Each sinogram bin corresponds to a line of response.

2.1 Attenuation, Scatter, and Randoms

At 511 keV, the photons suffer from two different interactions: photoelectron absorption and Compton scatter. Although absorption is very small at this energy, scattering causes the photon not to arrive at its original destination. Most scattered photons are not detected at all. Thus, these interactions affect the detection probability of annihilations; this effect, called attenuation, is one of the most important physical factors to be corrected for. The probability that a photon pair will survive along a LOR (survival probability) is:

$$\alpha_i = e^{-\int_{LOR_i} \mu(x)dx},$$ (2.1)

where $\mu(x)$ is the linear attenuation coefficient at 511 keV as a function of distance along the line joining the two detectors; it is nearly 0 in the air. The attenuation correction factor (ACF) is defined as $1/\alpha_i$. Since the survival probability is independent of the location of annihilation along the LOR, the scan count for each LOR can be corrected (by multiplication) by the ACF.

Detected single photons (by detectors) are called *singles*. Sometimes, however, one or both of the annihilation photons can be Compton scattered and still be detected. These events are called scattered events. They are false detections since the photon paths are not collinear. Energy discrimination is used to reject some of these events, since scattered photons usually have less energy. But not all the scattered photons can be rejected. Figure 2.1a illustrates a scattered event.

If two pairs of photons are emitted within the same coincidence window of duration Δ, and one photon out of each pair is absorbed or scattered out of the plane, another type of false detection occurs. This event, called a random event, along with scatter, will contribute to measurement error. Figure 2.1b illustrates a random event. There is a relationship between the singles rate and the randoms rate for a given pair of detectors:

$$R = 2\Delta S_1 S_2, \tag{2.2}$$

where R is the randoms rate, S_1 and S_2 are the singles rate for each detector, and Δ is the coincidence window duration. (A random event is caused by two unrelated singles striking two detectors in the same time interval.)

Randoms can be estimated by observing "delayed" coincidences over detectors: if two single events detected by two elements of a detector pair are within neighboring coincidence windows (e.g., they have been detected with a time difference of 2Δ), they are called the delayed coincidences. Delayed coincidences can be caused only by randoms; neither true annihilations nor scattered events can cause randoms because the difference in arrival time is greater than Δ. Since single event occurrences are uniformly distributed over time in a PET scan, the number of delayed coincidences for a detector pair is an estimate of the number of random events in that particular pair of detectors. The delayed coincidence randoms estimates are usually smoothed in the sinogram domain, since randoms are known to be almost constant no matter what object is in the FOV. In this work, we assume that the mean number of randoms for every detector pair is known.

2.2   Transmission Scans in PET

For attenuation correction, an accurate estimate of the survival probabilities α_i or attenuation correction factors $1/\alpha_i$ needs to be made. Early methods of attenuation correction employed simple geometric approximations to the attenuating medium (e.g., a uniform ellipse). While this approximation may be satisfactory in some cases, e.g., the brain, it is not suitable for every case; one example is the thorax, where the two lungs inside the body have very different attenuation coefficients than the soft tissue outside the lungs.
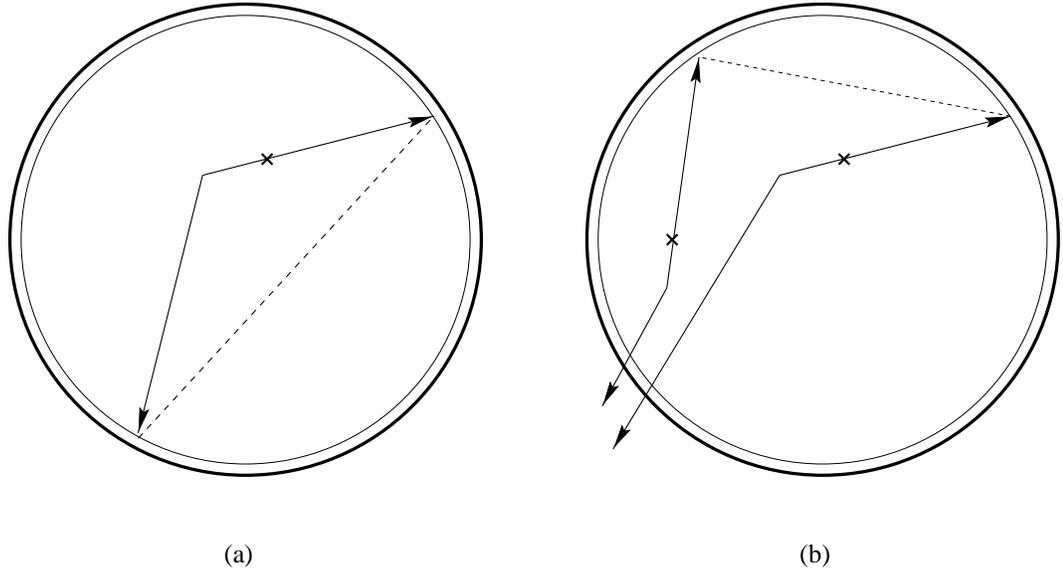
Figure 2.1: (a) scatter. (b) random. Dashed line denotes the detected LOR.

Measured attenuation correction methods provide more accurate estimates of ACF's. Typically, a transmission scan using only external radioactive sources is done to measure the ACF's before the body is injected with the radio-tracer. For calibration purposes, another transmission scan should be done when the patient is not in the scanner. This "blank scan" can be made relatively noise free because long scanning times can be used.

The statistical model of the transmission scan is:

$$\tilde{Y}_i \sim \text{Poisson}\{b_i \alpha_i + r_i\}, \tag{2.3}$$

where b_i is the blank scan rate (deadtime corrected), r_i is the background noise such as randoms and scatters, and α_i is the survival probability as given in (2.1). We can estimate α_i's directly using:

$$\hat{\alpha}_i = \text{smooth}\left(\frac{y_i - r_i}{b_i}\right). \tag{2.4}$$

Smoothing is done to reduce noise.

The ACF's can also be calculated from an attenuation map reconstructed from the transmission scan. Since we only have finite-dimensional measurements, it is only possible to reconstruct finite-dimensional attenuation maps. It is natural to represent this map in terms of pixels. If we let μ_j be the attenuation coefficient of the jth pixel, the statistical model

for the transmission scan is:

$$\tilde{Y}_i \sim \text{Poisson}\{b_i e^{-\sum_j g_{ij}\mu_j} + r_i\}, \tag{2.5}$$

$$\hat{\alpha}_i = e^{-\sum_j g_{ij}\hat{\mu}_j}, \tag{2.6}$$

where $\tilde{Y}_i$ is a random variable measuring the photon count of the ith ray, and g_{ij} is an entry in the system matrix $\mathbf{G}$. In matrix form, the statistical model is:

$$\tilde{Y}_i \sim \text{Poisson}\{b_i e^{-[\mathbf{G}\mu]_i} + r_i\}. \tag{2.7}$$

We can use Taylor expansion to approximate the log-likelihood around the point $l = \mathbf{G}\tilde{\mu}$ ($\tilde{\mu}$ is an initial estimate, e.g., an FBP reconstruction) to obtain the following quadratic log-likelihood function which is easy to minimize [74]:

$$L(\mu; l) = v'(\mathbf{G}\mu) + \frac{1}{2}(l - \mathbf{G}\mu)'\mathbf{W}(l - \mathbf{G}\mu), \tag{2.8}$$

where $v_i = (\frac{y_i}{b_i e^{-l}+r_i} - 1)b_i e^{-l}$, and $\mathbf{W} = \text{Diag}[(1 - \frac{y_i r_i}{(b_i e^{-l}+r_i)^2})b_i e^{-l}]$. Or we can estimate the ACF's directly by maximizing the Poisson log-likelihood:

$$L(\mu) = \sum_i y_i \log(b_i e^{-[\mathbf{G}\mu]_i} + r_i) - (b_i e^{-[\mathbf{G}\mu]_i} + r_i). \tag{2.9}$$

A new efficient algorithm for maximizing this objective is given in [30, 31], and used for the results in Chapters 3 and 6.

2.3   Emission Scan Model

After the transmission scan is done to measure the ACF's (if we choose to do the transmission scan), the body is injected with radio-tracer and the emission scan is done. The statistical model for the emission scan is:

$$\tilde{Y}_i \sim \text{Poisson}\{\sum_j a_{ij}\lambda_j + r_i\}, \tag{2.10}$$

where r_i is the background events (e.g., randoms and scatters), Y_i is the emission measurement of the ith ray, and $a_{ij} = c_i g_{ij}$ is an entry in the system matrix $\mathbf{A}$ (c_i is ray-dependent calibration factors, such as previously mentioned detector efficiency; g_{ij} is the probability

that a detection is made on ray i given that a photon is emitted from pixel j). In matrix form, the statistical model is:

$$\tilde{Y}_i \sim \text{Poisson}\{[\mathbf{A}\lambda + \underline{r}]_i\}. \tag{2.11}$$

For the first part of this thesis (Chapter 3), however, we will not focus on the "likelihood" part of the problem (but rather, as we will discuss later, on the "regularization" part). Thus we will simply use $L(f; y)$ (f represents the image or image parameters to be estimated; y is the measurement) to denote the suitable likelihood in our objective.

CHAPTER 3

Edge-Preserving Tomographic Reconstruction With Nonlocal Regularization

3.1 Maximum Penalized Likelihood Reconstruction with Local Regularization

The problem of reconstructing an unknown image f from a measurement vector y is usually ill-posed in the sense of Hadamard [5, 6]. Knowledge of the direct model is rarely sufficient to determine a satisfactory solution. If we obtain the maximum likelihood estimate (MLE) of the image by maximizing the log-likelihood function $L(f; y)$, then the resulting image is very noisy. Thus it is necessary to regularize the solution by imposing *a priori* assumptions. One simple regularization method supposes that images are globally smooth, and enforces a roughness penalty on the solution by adding a quadratic function to the negative log-likelihood. Such a "penalized" likelihood objective function has the following form:

$$\Phi(f) = -L(f; y) + \beta V(f), \tag{3.1}$$

where

$$V(f) = \int |\nabla f(\vec{x})|^2 d\vec{x} \tag{3.2}$$

is a measure of image roughness[1]. The image estimate is obtained by:

$$\hat{f} = \arg\min_f \Phi(f), \tag{3.3}$$

where the minimization with regard to f is often restricted to the nonnegative values. The function given in (3.2) is often unsatisfactory, due to the fact that many images are not globally smooth. They have region boundaries across which the image values can vary rapidly. The quadratic regularization in (3.2) causes edges to become blurred. In many images, small differences between neighboring pixels are often due to noise, while large differences are due to the presence of edges. This assumption has formed the basis for many edge-preserving regularization schemes proposed in the literature. Although a few region-based Baysian priors have been proposed, *e.g.* [8], most edge-preserving regularization methods rely on information from *local* neighborhood to determine the presence of edges.

One such scheme is to replace the quadratic penalty function in (3.2) with a non-quadratic function ψ that increases less rapidly than the quadratic function for sufficiently large arguments, such as the Huber function [39, 37]:

$$V(f) = \int \psi(|\nabla f(\vec{x})|)d\vec{x} \tag{3.4}$$

$$\text{or } V(f) = \int \psi\left(\left|\frac{df}{dx_1}\right|\right)d\vec{x} + \ldots + \int \psi\left(\left|\frac{df}{dx_n}\right|\right)d\vec{x}, \tag{3.5}$$

$$\text{where } \psi(t) = \begin{cases} t^2/2, & |t| \le \delta \\ \delta|t| - \delta^2/2, & |t| > \delta. \end{cases} \tag{3.6}$$

This function[2] increases linearly, instead of quadratically, for arguments larger than δ. Thus the objective function penalizes large differences between neighboring pixels less severely than the quadratic penalty, while maintaining the same level of penalty for small differences. This property allows sharper edges in the reconstructed image. If the log-likelihood is concave, then the objective function incorporating the regularization term given in (3.6) will be globally convex, so it is usually possible to find a global minimum of the objective function by using suitable iterative algorithms [39, 25, 16].

---

[1]In practice, a discretized solution is usually implemented; an example of such an implementation is $V(f) = \sum_j \sum_k w_{jk}(f_j - f_k)^2$ where w_{jk} is nonzero only for neighboring pixels; $d\vec{x}$ is a shorthand for $dx_1 \ldots dx_n$, where n is the dimension of the "image".

[2]We observe that $V(f)$ described by (3.4) is rotationally invariant, while $V(f)$ described by (3.5) is not.

One could also use a nonconvex penalty function, such as the broken-parabola function used by Lee *et al* [60, 47, 59]:

$$\phi(t) = \begin{cases} \gamma t^2, & \gamma t^2 \leq \alpha \\ \alpha, & \gamma t^2 > \alpha. \end{cases} \tag{3.7}$$

The broken-parabola function makes the objective function non-convex and non-differentiable, thereby ruling out gradient-based descent methods. Instead, a deterministic annealing algorithm is proposed in [60], so that a series of objective functions using different ϕ_γ's as the penalty function are minimized:

$$\phi_\gamma(t) = -\frac{1}{\gamma}\log(e^{-\gamma c t^2} + e^{-\gamma \alpha}). \tag{3.8}$$

This ϕ_γ function resembles a quadratic function when γ is small; it approaches the broken-parabola function as $\gamma \to \infty$. Thus the minimization of a nonconvex nondifferentiable objective function is transformed into a series of minimizations of nonconvex but differentiable objective functions, during which γ is increased for each minimization, so that the final penalty function is close to the broken-parabola. During each minimization over a certain γ, auxiliary variables that represent the local strength of an edge are introduced. The minimization method alternates between updating these auxiliary variables and updating pixel intensities. This approach also uses information from the *local* neighborhood. Figure 3.1 compares the quadratic, Huber, and the broken-parabola function.

Both edge-preserving methods, *i.e.*, replacing the quadratic penalty function with the Huber function or the broken-parabola function, modify the local penalty function ψ so that it will penalize large differences between neighboring pixels less than the quadratic penalty. A related approach to these methods is the "total variation" method, where one attempts to remove noise from an image by solving a nonlinear minimization problem involving a total variation criterion[27, 28, 85]. All the approaches reviewed above rely on information from local neighborhood to determine the presence of an edge locally. This chapter describes a method for including *nonlocal* information into the regularization method, thus hopefully achieving better results under certain cases. In the specific case of emission computed tomography, accurate attenuation correction is usually necessary for a satisfactory emission
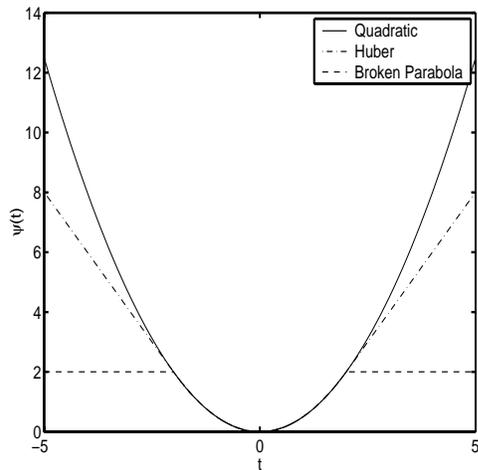
Figure 3.1: A comparison of quadratic, Huber, and the broken-parabola penalty function

reconstruction[51]. Accurate attenuation correction requires an accurate map of attenuation coefficients. A PET attenuation map consists of a small number of regions, *i.e.*, lungs, spine, body tissue, etc. The attenuation coefficients within each region are fairly uniform, but they vary a great deal between neighboring regions and the transition between regions can be fairly rapid, *e.g.*, across a few pixels. A regularization method that incorporates this additional prior information should be able to outperform a purely *local* regularization method. In this chapter, we propose a new objective function for image reconstruction that incorporates boundary and region information into its regularization. Section 3.3 describes our new reconstruction algorithm; Section 3.5 compares the proposed algorithm to a "conventional" statistical reconstruction algorithm and FBP; Section 3.7 summarizes the results and outlines possible future work.

3.2   Image Segmentation Techniques

Our new regularization approach to image reconstruction exploits boundary and region information. The ideas we use are derived from the image segmentation community. We will now review a few image segmentation techniques, and present an attractive implementation technique for certain snake models called the "level set technique". Finally in this section, we will review a few transmission attenuation map segmentation techniques. Please note that our goal is to incorporate segmentation techniques into our reconstruction problem, rather than a 2-step method of "first reconstruct then segment".

3.2.1   General Techniques

Image segmentation is a critical problem of early vision and it has been extensively studied. Approaches to image segmentation can be roughly divided into four types:

1. Local Filtering approaches such as the Canny edge detector [11].
2. Boundary models, such as Snakes [53] and Balloon methods [20];
   also parametric models such as used in [83, 80]
3. Region growing techniques [2].
4. Global optimization approaches (for labels) based on some criteria, such as
   Bayesian or Minimum Description Length (MDL)[7, 46, 45, 58].

Local Filtering Techniques

Usually local filtering approaches do not have a criteria which is to be minimized by the method. Since in our case, we do not have an image to start with, it is very difficult to incorporate a local filtering technique into our method. Furthermore, local filtering techniques usually only consider local information, thus it would not suit our objective of incorporating nonlocal information into the regularization. Thus we will not consider local filtering techniques.

Boundary Models

These methods can be divided into two broad categories: parametric and non-parametric. Staib *et al* [80] uses the Fourier basis to describe the boundary curve:

$$\left[\begin{array}{c} x(t) \\ y(t) \end{array} \right] = \left[\begin{array}{c} a_0 \\ c_0 \end{array} \right] + \sum_{k=1}^{n} \left[\begin{array}{cc} a_k & b_k \\ c_k & d_k \end{array} \right] \left[\begin{array}{c} \cos kt \\ \sin kt \end{array} \right] \tag{3.9}$$

One drawback of this approach is that the initialization must be fairly close to the actual boundary; furthermore, changing any one parameter in the representation affects the global shape of the boundary. Titus [83] used B-Splines to represent the boundary curve radially relative to a fixed center point. The number of knots in the curve in fixed *a priori*. Then estimating the boundary is turned into a problem of estimating the parameters describing the spline function. This approach can achieve a good local fit for shapes whose behavior in one region is not necessarily related to their behavior everywhere else. The drawback in both these two parametric approaches is that the number of parameters must be decided in advance and calculating the derivatives of the objective function with regard to these parameters (so that a gradient descent method can be used to minimize the objective function) is very cumbersome. Also, depending on the choice of parameterization, the shape of the object may be restricted.

The non-parametric approach to the boundary model simply deals with boundary curve directly. One of the most popular non-parametric approaches is the snake model [53]. A snake is an active contour defined by the parametric mapping

$$\Gamma(p) = (x(p), y(p)), \tag{3.10}$$

where $p \in [0, 1]$ parameterizes the contour, $\Gamma(0) = \Gamma(1)$, and all derivatives match at 0 and 1. Instead of considering a single curve, we consider a smooth family of closed curves $\Gamma(p, t) = (x(p, t), y(p, t))$ where t parameterizes the family and p the given curve, $p \in [0, 1]$. A typical energy for a snake is:

$$E[\Gamma(\cdot, t)] = \int_0^1 \left(\frac{1}{2} (\alpha |\Gamma_p|^2 + \beta |\Gamma_{pp}|^2) - \lambda |\nabla I|^2 \right) dp, \tag{3.11}$$

where $\Gamma_p(p, t) \triangleq (x_p(p, t), y_p(p, t))$. To find the curve that minimizes this energy, one can perform steepest descent (in the sense that the snake evolves in the direction that decreases

the energy most rapidly, *i.e.*, in the direction of the negative functional derivative $\frac{\delta E}{\delta \Gamma}$):

$$\frac{\partial \Gamma}{\partial t} = -\frac{\delta E}{\delta \Gamma(\cdot, t)} = -\alpha \Gamma_{pp} + \beta \Gamma_{pppp} + \lambda \nabla |\nabla I|^2. \tag{3.12}$$

In computer implementation, an approximation to the derivatives is made and a finite time step (Δt) is taken:

$$\Gamma(\cdot, t + \Delta t) \approx \Gamma(\cdot, t) - \frac{\delta E}{\delta \Gamma(\cdot, t)} \Delta t. \tag{3.13}$$

For the classical snake model to work satisfactorily, the initial snake must be fairly close to the actual boundary, as the only force pushing the snake to the boundary is the third term, $\lambda \nabla |\nabla I|^2$, which is only effective near the boundary. To increase the "capture range" of snake models, the balloon models introduce an additional force $\nu \vec{\mathcal{N}}$ (ν is a constant) which pushes the contour out in the normal direction (no matter where the contour is). This is not very satisfying since even if the contour has reached the true boundary, the force acting on the contour will not go to zero; so if the additional force is not carefully controlled, the curve may expand indefinitely. A more satisfying way of increasing the capture range is proposed in [86]; in their work, rather than a constant force, a spatially variant "gradient vector flow" field is generated from the image; then the contour is put under the influence of this force and the internal forces which aim to keep the contour smooth. This snake model has a fairly large capture range and does not expand out uncontrollably (unlike if a constant expansion force is added) after it reaches the true boundary.

Region Growing Techniques

The goal of snake models is to find closed boundary contours; in contrast, the goal of the region growing (merging) is to divide the image Ω into, say M, of regions, so that $\Omega = \cup_{i=1}^{M} R_i$, $R_i \cap R_j = \emptyset$ if $i \neq j$, and each R_i satisfies a homogeneity criterion. Region merging builds up complicated regions by combining smaller regions using a statistical similarity test, *e.g.*, Fisher's test. If there are two adjacent regions R_1 and R_2 where n_1, n_2, $\hat{\mu}_1$, $\hat{\mu}_2$, $\hat{\sigma}_1^2$, and $\hat{\sigma}_2^2$ are the sizes, sample means, and sample variances of R_1 and R_2, respectively. Then, to decide whether or not to merge them, the squared Fisher distance is evaluated:

$$\frac{(n_1 + n_2)(\hat{\mu}_1 - \hat{\mu}_2)^2}{n_1 \hat{\sigma}_1^2 + n_2 \hat{\sigma}_2^2} = \frac{n \hat{\sigma}^2}{n_1 \hat{\sigma}_1^2 + n_2 \hat{\sigma}_2^2} - 1, \tag{3.14}$$

16

where $n = n_1 + n_2$ and $\hat{\sigma}^2$ is the sample variance of the mixture region. If this statistic is below a certain threshold then the regions are merged. Region growing can be seen as a special case of region merging, where R_1 is a region and R_2 is a single pixel at the boundary of R_1. Although region growing algorithms are very intuitive, it is rarely know what, if any, global cost function they are minimizing, let alone whether the algorithm converges to the minimum of such a global cost function, and the resulting regions often end up with jagged boundaries.

Both Bayes and MDL specify ways for segmenting images using global cost function criteria. They have different motivations but are actually equivalent to each other, in the sense that one can be transformed to the other. Under the Bayes approach, the observed image is modeled as a degraded version of an ideal image assumed to be piecewise smooth. For example, Mumford and Blake use the following criteria:

$$E[f, \Gamma] = \mu \iint_\Omega (f - I)^2 dx dy + \lambda \iint_{R-\Gamma} |\nabla f|^2 dx dy + \nu |\Gamma|, \qquad (3.15)$$

where I is the input image, f is the output image, and Γ labels the discontinuities. Leclerc suggests that segmenting the images according to the above Baysian model should be equivalent to obtaining its minimum description length in terms of a previously specified description language. A typical MDL criterion occurs in (3.16); it differs from (3.15) by letting the σ's be unknown variables which are assumed to be constant within each region. It is usually very difficult to minimize this type of energy functional; algorithms such as simulated annealing, graduated nonconvexity, and deterministic annealing are sometimes successful.

Global Optimization Approaches

All these approaches have their advantages and disadvantages. The filtering approach only takes into account local information and cannot guarantee closed edge contours. Snake models only make use of information along the boundary and require good initialization (or adding ballooning force) to yield satisfactory results. Region growing techniques take into account region information, but the boundaries they generate are often not smooth. Global optimization techniques impose a global criterion for segmenting the image, but it is often

difficult and time-consuming to find their minima.

In [96], the region competition algorithm was presented. Region competition minimizes the following generalized Bayes/MDL criterion using the variational principle:

$$E[\Gamma, \{\alpha_i\}] = \sum_{i=1}^{M} \left(\frac{\mu}{2} \int_{\delta R_i} ds - \log P(\{I_{(x,y)} : (x,y) \in R_i\} | \alpha_i) + \lambda \right), \qquad (3.16)$$

where the first term is the length of the boundary curve for region R_i; the second term is the sum of the cost for coding the intensity of every pixel (x, y) inside region R_i according to a distribution $P(\{I_{(x,y)} : (x,y) \in R_i\} | \alpha_i)$; λ is the code length needed to describe the distribution and code system for region R_i. They employed alternating minimization: with Γ fixed,

$$\hat{\alpha}_i = \arg\min_{\alpha_i} \left(-\iint_{R_i} \log P(\alpha_i | I_{(x,y)}) dx dy \right). \qquad (3.17)$$

With α_i's fixed, they employed steepest descent (for any point $\vec{v} = (x, y)$):

$$\frac{d\vec{v}}{dt} = -\frac{\delta E[\Gamma, \{\alpha_i\}]}{\delta \vec{v}} = \sum_{k \in Q(\vec{v})} \left(-\frac{\mu}{2} \kappa_{k(\vec{v})} \vec{\mathcal{N}}_{k(\vec{v})} + \log P(I_{\vec{v}} | \alpha_k) \vec{\mathcal{N}}_{k(\vec{v})} \right), \qquad (3.18)$$

where $Q(\vec{v}) = \{k | \vec{v} \text{ lies on } \Gamma_k\}$, *i.e.*, the summation is done over those regions R_k for which $\vec{v}$ is on Γ_k. It turns out there is a simple intuitive interpretation for the above equation: the first term maintains the smoothness of the contour; the second term is the statistic force, $f = \log P(I | \alpha) \vec{\mathcal{N}}$; this force always tries to compress the region, and the better the point $\vec{v}$ satisfies the homogeneity requirement the weaker the statistic force; hence the neighboring regions compete to get the points along its common boundary. The region competition algorithm combines some of the most attractive aspects of snakes, region growing, and global cost function approaches. The segmentation is parameterized in terms of boundaries, but the energy function also takes into account region information. This technique is also not very time-consuming (in the sense that the number of parameters is only of $O(n)$, rather than $O(n^2)$ if we assign every pixel a label, and minimize a cost function dependent on these labels).

3.2.2   Level Set Approach to Front Propagation

Now we describe the level set implementation technique for active contours. This approach has many advantages when compared to traditional snake implementations; we

use this implementation throughout our work. The traditional approach to the problem of a closed curve (snakes) $C(s,t)$, moving in the plane along its normal vector field with curvature-dependent speed, is to take the equations describing the motion for the position vector $C(s,t)$, and then discretize the parameterization with a set of discrete marker particles lying on the moving front. These discrete markers are updated in time by approximating the spatial derivatives in the equations of motion, thereby advancing their positions. This approach has several problems [76]. First, small errors in the computed particle positions are tremendously amplified by the curvature term, leading to instability unless an extremely small time step is employed. Second, topological changes are difficult to manage when the evolving interface breaks or merges. Third, extension of this technique to three dimensions is difficult.

The level set technique was proposed by Osher and Sethian [72, 77] to deal with these problems; in the level set technique, the evolving front $C(t)$ (a function of time) is represented as the level set $\{\Psi = 0\}$ of a function Ψ (called the hypersurface in this setting); the main idea is to embed a 1-D curve in a 2-D function. Thus given an initial closed contour $C_{t=0}(p) = (x(p), y(p))^T$ in $\mathbb{R}^2$ where $0 \leq p \leq 1$, the function Ψ at $t = 0$ will be:

$$\Psi(x, y, t = 0) = \begin{cases} 0, & \text{if } (x, y) \text{ is on } C \\ d, & \text{otherwise,} \end{cases} \tag{3.19}$$

where d is the closest distance to initial level 0 of Ψ (it can be some other function of (x, y), and the plus (minus) sign is chosen if the point (x, y) outside (inside) level 0). Hence we have the initial function $\Psi(x, y, t = 0) : \mathbb{R}^2 \to \mathbb{R}$ with the property that

$$C_{t=0}(p) = ((x(p), y(p))^T | \Psi(x(p), y(p), t = 0) = 0). \tag{3.20}$$

Figure 3.2 illustrates an expanding circle and its level set representation.

A relationship must be established between evolution of the contour C and evolution of the function Ψ, so that $\forall t \geq 0, C_t(p) = ((x(p), y(p))^T | \Psi(x(p), y(p), t) = 0)$. Let $C(p, t) : S^1 \times [0, \infty) \to \mathbb{R}^2$ be a family of curves[3] satisfying the following (curvature-dependent) evolution equation:

$$\frac{\partial C}{\partial t} = \nu(\kappa)\vec{\mathcal{N}} \tag{3.21}$$

---

[3]In three dimensions, $C(p, t) : S^2 \times [0, \infty) \to \mathbb{R}^3$ denotes a family of surfaces; refer to [64, 90] for more details.
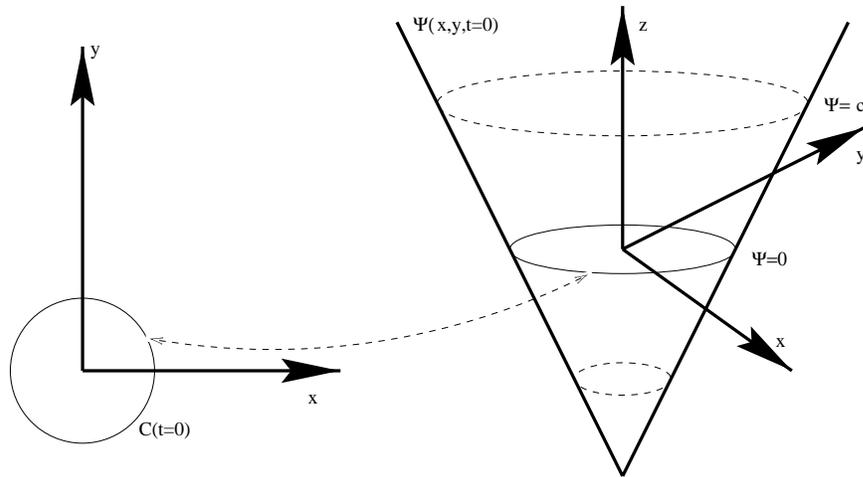
(a)
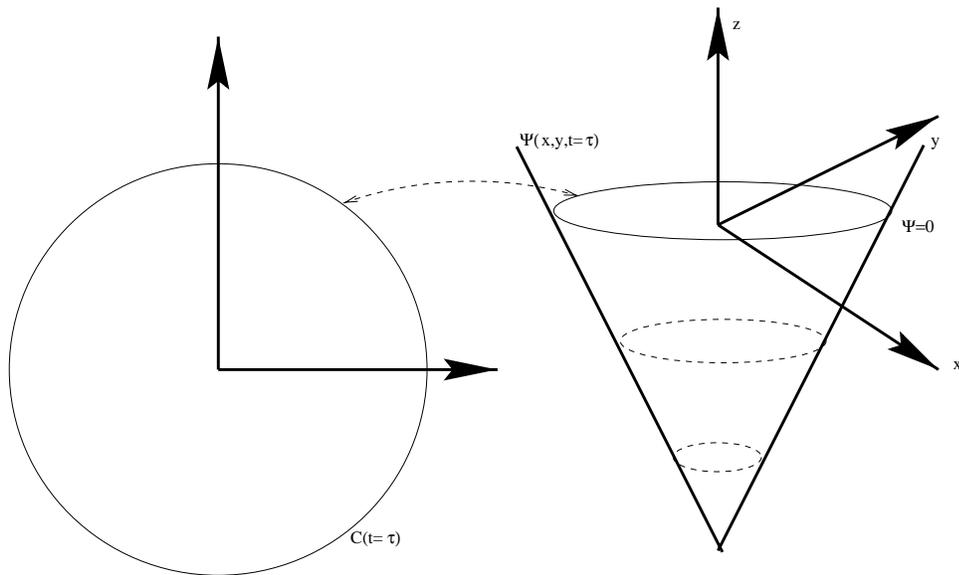
(b)

Figure 3.2: Level set formulation of equation of motion. (a) shows the curve C at $t = 0$ and its level set representation. (b) shows the curve C, after expanding, at $t = \tau$.

The equation that evolves the function Ψ so that the contour C is always the zero level set of Ψ must satisfy:

$$\Psi(C(t), t) = 0. \tag{3.22}$$

By the chain rule,

$$\Psi_t + \nabla\Psi(C(t), t) \cdot C'(t) = 0. \tag{3.23}$$

If ν is the speed in the inward normal direction, then $C'(t) \cdot \vec{\mathcal{N}} = \nu$ where $\vec{\mathcal{N}} = -\frac{\nabla\Psi}{|\nabla\Psi|}$, and this yields the evolution equation for Ψ:

$$\Psi_t = \nu(\kappa)\|\nabla\Psi\| \tag{3.24}$$

The curvature can be calculated directly from function Ψ using the following formula [68]:

$$\kappa = \operatorname{div}\left(\frac{\nabla\Psi}{\|\nabla\Psi\|}\right). \tag{3.25}$$

There are several advantages in this formulation. The first is that the evolving function always remains a function even if the contour C changes topology, breaks, or merges. The second is that Ψ can be approximated using a discrete grid; finite difference approximation can be made to calculate the derivatives. It is also easy to determine the embedded contour from the level set Ψ. Finally, going to three or higher dimensions requires no significant change. Because of these advantages, the level set approach, since its introduction, has been used in a wide collection of problems involving evolving interfaces.

3.2.3 Shape Recovery with Front Propagation

Malladi *et al* [64] proposed an application of the front propagation using the level set approach to the problem of shape recovery. A constant inflation (or deflation) term F_0 is added to the curvature-dependent speed term, and an image-dependent speed function $\phi(x, y)$ is also incorporated; the equation describing the evolution of the level set becomes:

$$\frac{\partial\Psi}{\partial t} = \phi(x, y)\|\nabla\Psi\|\left(\operatorname{div}\left(\frac{\nabla\Psi}{\|\nabla\Psi\|}\right) + F_0\right). \tag{3.26}$$

The constant F_0 is added to keep the curve moving in the proper direction (either expanding with a negative F_0 or shrinking with a positive F_0). The image-dependent speed function

$\phi(x,y)$ is defined in a way such that it has very small magnitude in the vicinity of an edge and so acts to stop the evolution when the contour gets close to an edge. The following speed function can be used:

$$\phi = \frac{1}{1 + \|\nabla G_\sigma * I\|^n} \text{ or } \phi = e^{-\|\nabla G_\sigma * I\|}, \tag{3.27}$$

where G_σ is a Gaussian smoothing filter which is used to smooth out noise. Both of these definitions make ϕ close to 0 if the magnitude of the local gradient is large, and close to 1 if the magnitude of the local gradient is close to zero. Thus the initial curve will shrink or expand until it meets a boundary of the image. Figure 3.3 shows a noisy image and its corresponding ϕ function.

(a)   (b)
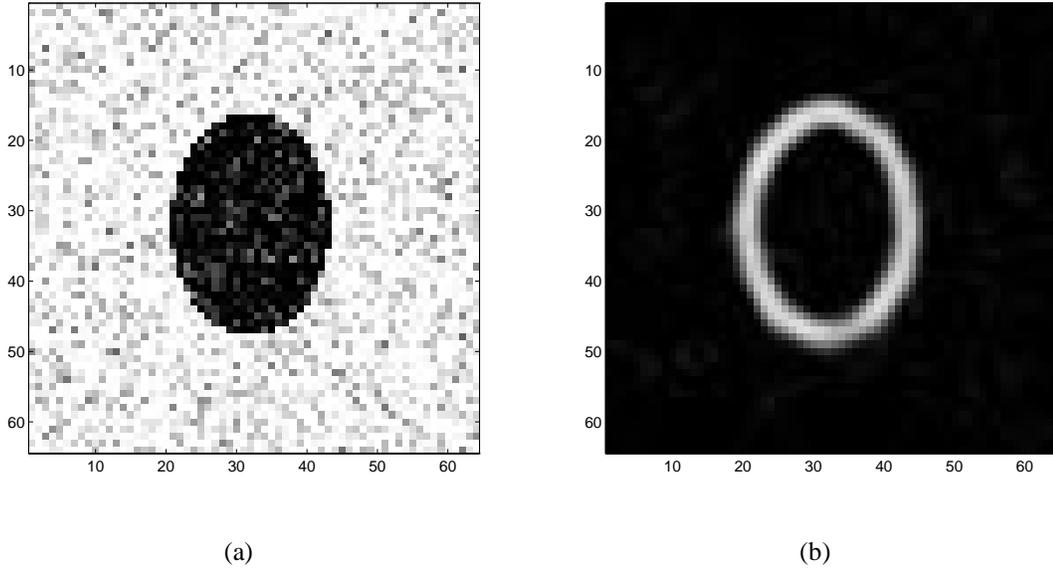
Figure 3.3: (a) noisy image. (b) its corresponding ϕ function.

The image-dependent speed term only describes the speed of the evolving curve, *i.e.*, the level 0 of the hypersurface; it does not describe the motion of other levels of the hypersurface, thus it has no meaning on other levels. However, Ψ is defined over the entire hypersurface, thus the speed term on nonzero levels must be defined, in order to evolve the level set according to (3.26). The image-dependent speed function ϕ is *extended* from level 0 to all other levels in the following manner: let $\hat{\phi}$ at a point P lying on a level set $\{\Psi = C\}$ be the value of ϕ at a point Q, such that Q is the closest to P and lies on the level

(a)   (b)
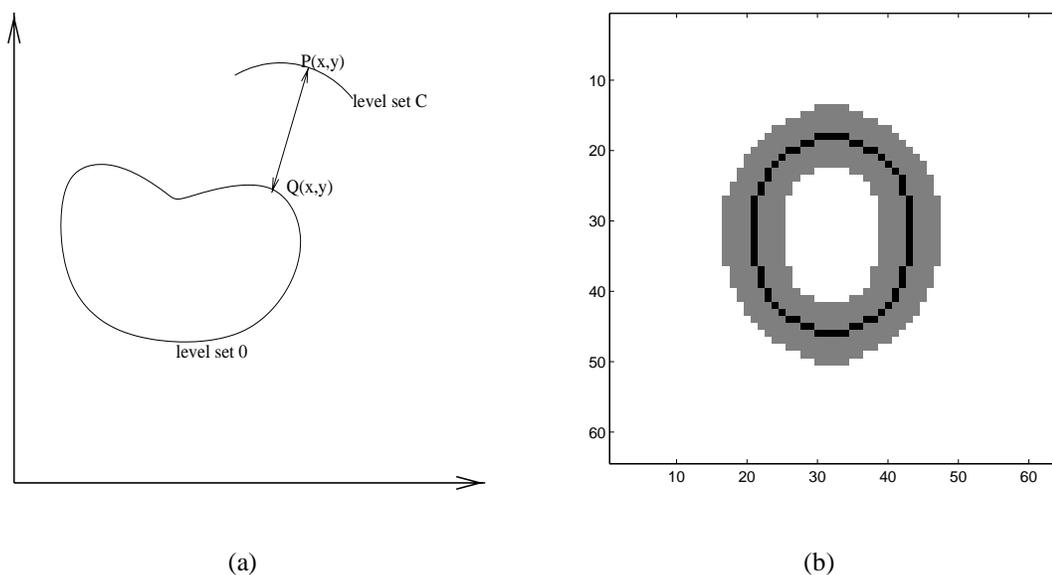
Figure 3.4: (a) extension of the speed term. (b) gray area is the narrow band constructed around the black contour.

set $\{\Psi = 0\}$. This extension of the image-dependent speed term has the important property that the level sets evolving under such a speed term will not collide. This is, however, an expensive operation of $O(N^3)$ (if there are $N \times N$ pixels in the image), because a search for the smallest distance to level 0 must be carried out for every pixel in the hypersurface. Figure 3.4a illustrates the extension of the speed term.

To save time, the "narrow-band" extension can be used [1]. Instead of extending the image-dependent speed function to the entire hypersurface, only extension to a band around level 0 is performed. Once the front is about to expand or shrink out of this band, a new band is constructed around the new front, and the new band is used until the new front is about to get out of this band. This requires re-initialization of the band and the hypersurface every certain number of iterations. In this re-initialization, a search for the smallest distance to level 0 is performed, and those that are within a certain threshold are made into the new band. Figure 3.4b illustrates the narrow band approach.

Every time an extension of the speed term is performed, level 0 must first be re-acquired. This is not trivial since after a few iterations, there will be no values on the hypersurface

23

exactly equal to 0. Thus a contour must be extracted[4] from hypersurface.

This method of shape recovery works well. Figure 3.5 and 3.6 show successful boundary extraction of two images. One major disadvantage is that it still requires a small time step, because a large time step will risk the contour expand out of (or into) the actual boundary. Once that happens, it will keep expanding uncontrollably. This problem is remedied in the work done by Yezzi *et al* described in the next subsection.

---

[4]We use Matlab's contour program to extract a discrete version of the level-0 curve, *i.e.*, the output of the contour program consists of samples of the continuous curve.

(a)   (b)   (c)

(d)   (e)   (f)

(g)   (h)   (i)
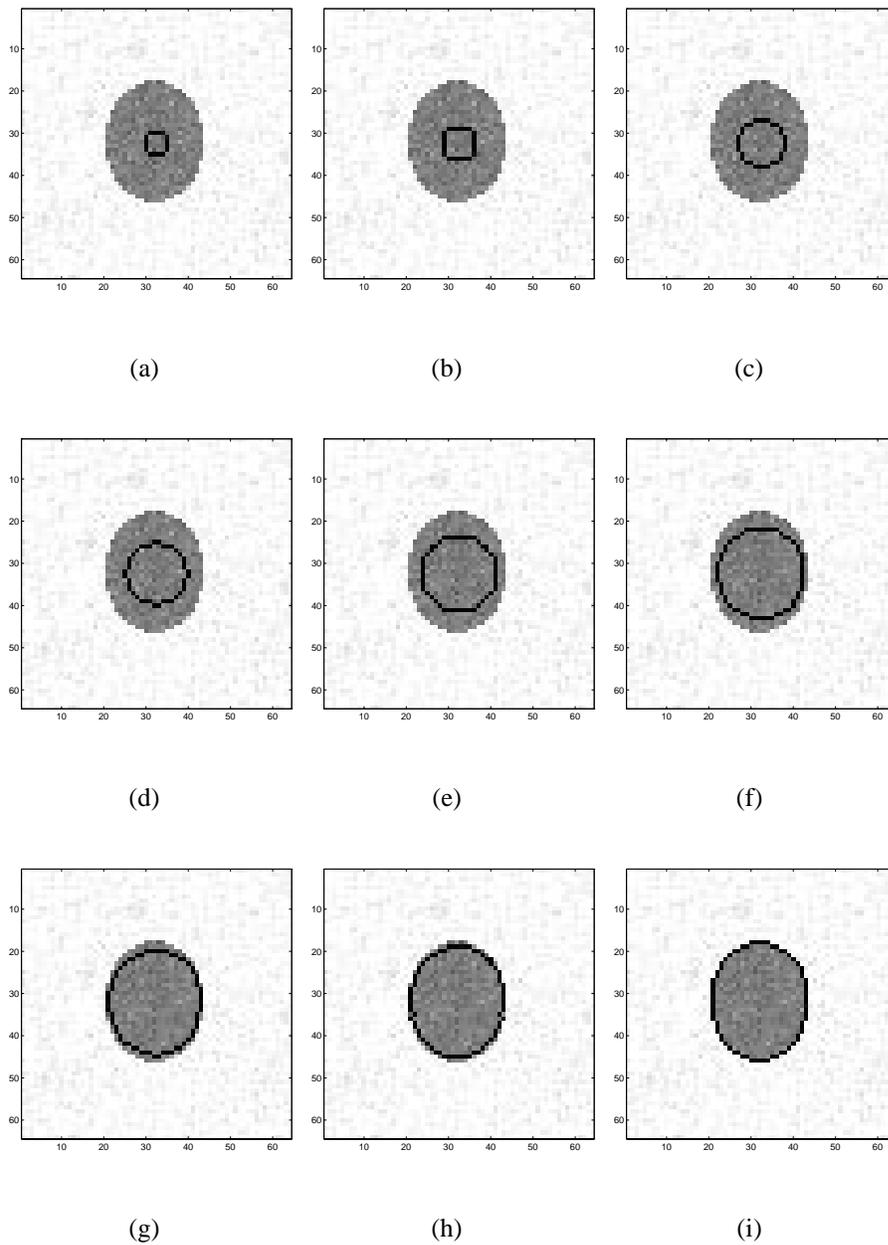
Figure 3.5: Boundary detection of an ellipse. Black contour indicates the propagating front.
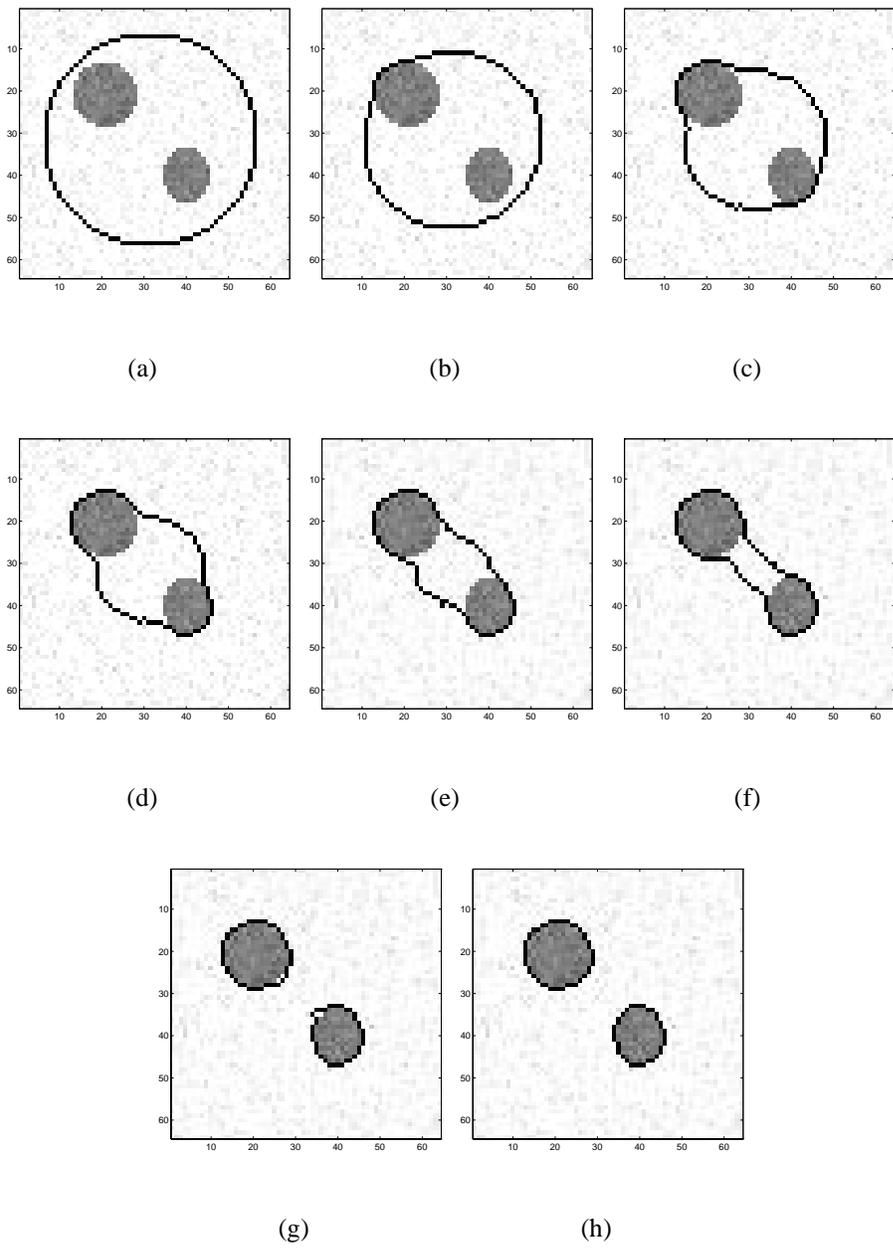
Figure 3.6: Boundary detection of 2 objects with one initial contour

3.2.4 Shape Recovery Using Energy Minimization

Yezzi *et al* [90] proposed an improvement to the above approach by modifying the length functional from $L(t) = \int_0^1 \|\frac{\partial C}{\partial p}\| dp$ to:

$$L_\phi(t) = \int_0^1 \left|\frac{\partial C}{\partial p}\right| \phi dp, \tag{3.28}$$

where ϕ is the image-dependent speed function as defined in (3.27). This new length functional is the "shortest" when the contour lies right on the boundary of an object where the magnitude of local gradient is very large and consequently the image-dependent speed function is very close to 0. Thus when $L_\phi(t)$ is minimized, the contour is on the boundary of the object. Figure 3.7 shows the same contour will result in different lengths measured by this new length functional.

Taking first variation of the usual length functional $L(t) = \int_0^1 \|\frac{\partial C}{\partial p}\| dp$, we have:

$$-L'(t) = \int_0^{L_t} \langle \frac{\partial C}{\partial t}, \kappa \vec{\mathcal{N}} \rangle ds. \tag{3.29}$$

Hence the direction in which $L(t)$ is decreasing most rapidly is when

$$\frac{\partial C}{\partial t} = \kappa \vec{\mathcal{N}}. \tag{3.30}$$

A simple closed curve converges to "round" points when evolved according to (3.30) without developing singularities [48]. Applying the same technique to the new length functional and taking first variation of (3.28), and using integration by parts:

$$-L'_\phi(t) = \int_0^{L_\phi(t)} \langle \frac{\partial C}{\partial t}, \phi \kappa \vec{\mathcal{N}} - (\nabla \phi \cdot \vec{\mathcal{N}}) \vec{\mathcal{N}} \rangle ds, \tag{3.31}$$

which means that the direction in which the L_ϕ perimeter is shrinking as fast as possible is given by:

$$\frac{\partial C}{\partial t} = (\phi \kappa - \nabla \phi \cdot \vec{\mathcal{N}}) \vec{\mathcal{N}}. \tag{3.32}$$

This is the gradient flow corresponding to the minimization of the length functional L_ϕ. The level set implementation of this flow can be obtained using the same technique outlined in (3.22), (3.23), and (3.24):

$$\frac{\partial \Psi}{\partial t} = \phi(x, y) \|\nabla \Psi\| \operatorname{div} \left(\frac{\nabla \Psi}{\|\nabla \Psi\|} \right) + \nabla \phi \cdot \nabla \Psi. \tag{3.33}$$
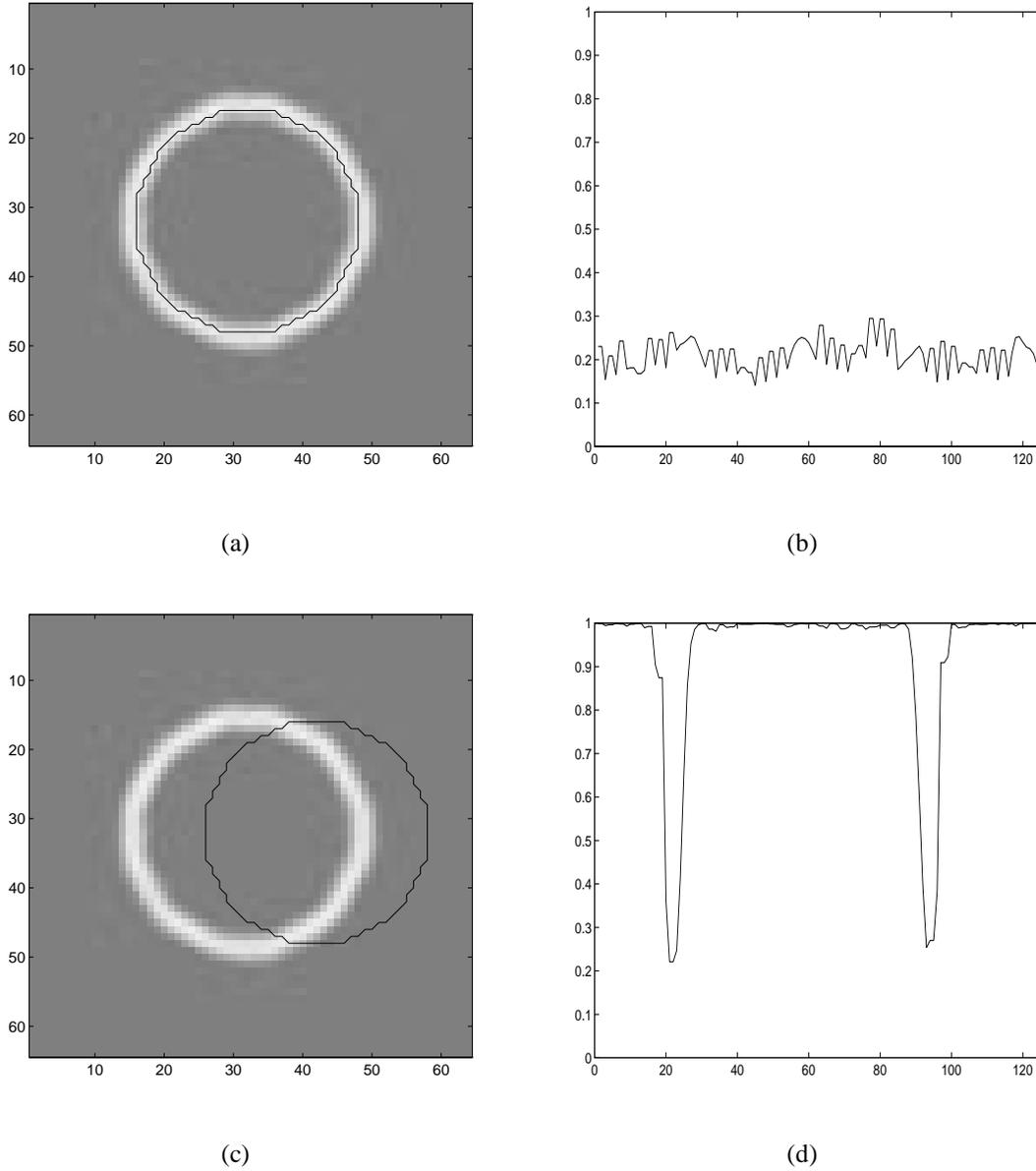
Figure 3.7: the contour is (a) on the boundary (c) not on the boundary; background is the speed function. (b) (d) ϕ function on the contour. In (d) ϕ function is 1 at most places.

In actual implementation, a constant inflation term F_0 is added (proposed by Malladi *et al* in [64]), making the level set implementation:

$$\frac{\partial \Psi}{\partial t} = \phi(x,y)\|\nabla\Psi\| \left(\mathrm{div}\left(\frac{\nabla\Psi}{\|\nabla\Psi\|} \right) + F_0 \right) + \nabla\phi \cdot \nabla\Psi. \qquad (3.34)$$

This modified level set approach has all the advantages of Malladi [64], but has the added benefit of being able to tolerate a larger time step. For a ϕ as defined in (3.27), $\nabla\phi$

will look like a "doublet" near an edge. It attracts the evolving contour as it approaches an edge, and pushes it back if it should pass the edge [54]. Thus a larger time step can be used. Furthermore, when the contour rests on the edge, the energy functional (3.28) is actually minimized.

3.2.5   Segmentation Techniques for PET Attenuation Correction

Instead of using a reconstructed attenuation map for PET attenuation correction, many authors have proposed applying various segmentation techniques to the reconstructed attenuation map, then assigning some (often constant) values to each region.

Xu *et al* [87] segmented the transmission image into anatomic regions by first calculating the threshold between regions using the histogram of the attenuation values; then the image is segmented using these thresholds into different regions; every pixel in every region is then assigned the average attenuation values within that region. Meikle *et al* [66] used a more sophisticated method of thresholding for thorax transmission images: first, the histogram obtained from a reconstruction with no smoothing at all (but after median-filtering) is used to fit three Gaussian PDF's (corresponding to lung, air and soft tissue); then every pixel's new value is calculated using:

$$\mu' = \frac{1}{N}(\mu_{\mathrm{air}}P(\mu_{\mathrm{air}}|\mu) + \mu_{\mathrm{lung}}P(\mu_{\mathrm{lung}}|\mu) + \mu_{\mathrm{st}}P(\mu_{\mathrm{st}}|\mu)), \qquad (3.35)$$

where

$$N = P(\mu_{\mathrm{air}}|\mu) + P(\mu_{\mathrm{lung}}|\mu) + P(\mu_{\mathrm{st}}|\mu). \qquad (3.36)$$

So a continuous range of μ values is possible in the "segmented" image. Then a median filter is applied to the "segmented" image.

Instead of reconstructing the attenuation map then segmenting this map, some authors have proposed calculating directly the boundaries from projection data. Tomitani [84] used this approach to find the boundary of the human skull. First, the boundary contours on the projection data is estimated by thresholding the sinogram; these contours are often jagged, thus only the lower order Fourier coefficients of the contours are retained; finally the boundary contours are transformed into the image domain. Although this approach

is only applied to human skull transmission images, it may also be possible to use this approach for the thorax.

Instead of first reconstructing the image and then segmenting the reconstructed image, Fessler [36] presented a "unified reconstruction/segmentation method" based on a penalized weighted least-squares objective function. But the parameterization is discrete, *i.e.*, there are only a discrete set of possible attenuation coefficients (the lungs, soft tissue, bone, etc.). Even though this method minimizes a single objective, the parameters are still discrete, and the different pixels within a "region" cannot take on different values. This is not the case for lungs. Furthermore, this method also relies only on local information.

Most of these segmentation techniques impose a constant value on a certain region, which is not true under certain cases, *e.g.*, the lung region usually does not have uniform attenuation coefficients and different patients have different attenuation coefficients in this region. Most of these approaches assign a constant number to a certain region, while the others assign some number decided using some *ad hoc* formula to every pixel, so that the attenuation coefficients have a continuous range. We will apply our new nonlocal regularization approach to this problem, so that we take into account boundary information while still estimating the attenuation coefficient at every point directly from the projection data.

3.3 A New Approach to Edge-Preserving Regularization

3.3.1 A New Objective Function

Our development of the new objective function was motivated by the specific application of PET and SPECT transmission tomography, but its use is not restricted to attenuation map reconstructions. We assume that the actual object to be reconstructed is everywhere differentiable (and thus continuous). We also assume that the object consists of regions that are piecewise smooth everywhere except very close to the region boundaries where the object intensity changes rapidly but continuously to values in its neighboring region(s). Thus an edge-preserving penalty function should penalize local gradients that are within each region more than local gradients that are very close to a boundary. Furthermore, we assume that the boundaries separating the regions are smooth. Figure 3.8 shows an example object (representative of an attenuation map of a thorax image at 511 KeV) and one of its vertical profiles. This object is piecewise smooth, but not piecewise constant, due to variations in lung density.

(a)   (b)
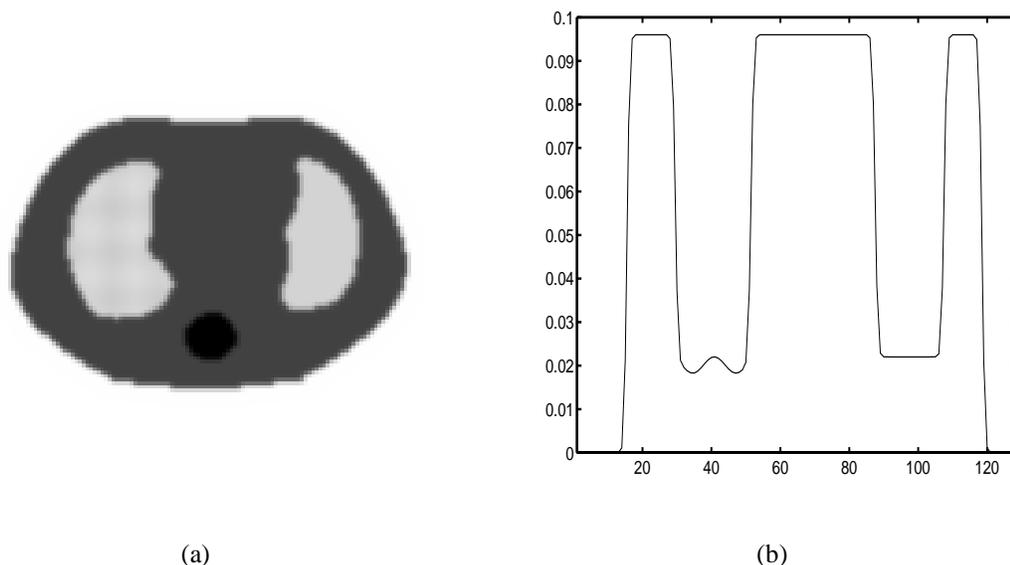
Figure 3.8: (a) Phantom (b) A profile of the phantom: row 60.

Like (3.1), the objective function we propose also consists of a data-fit term and a penalty term. However, our penalty considers not only the image values but also the char-

acteristics of region boundaries within the image. Following the convention in PDE-based image analysis literature, we present a non-discretized formulation. Let f denote the object, $\Gamma_k \in \mathcal{G}$ denote the kth boundary curve, and Ω denote the domain of the image. Let $R_m(\Gamma) \subset \mathbb{R}^2$ denote the mth region[5], where each R_m does not include its boundary, hence they are open sets. The regions are separated by boundary curves Γ, where $\Gamma = \cup_{k=1}^{K} \Gamma_k$. We assume that the number of boundary curves K is fixed (and known *a priori*) and that the boundary curves do not touch each other, *i.e.* $\Gamma_{k_1} \cap \Gamma_{k_2} = \emptyset$ if $k_1 \neq k_2$, which is reasonable for transmission tomography. Hence the number of regions is $M = K + 1$ and $R_{m_1} \cap R_{m_2} = \emptyset$ if $m_1 \neq m_2$.

We propose the following objective function of the object f and the boundaries Γ [92]:

$$J(f, \Gamma) = -L(f; y) + V(f, \Gamma) \tag{3.37}$$

$$V(f, \Gamma) = \sum_{k=1}^{K} \left[\beta J_2(f, k, \Gamma) + \mu J_1(\Gamma_k) \right] \tag{3.38}$$

$$J_1(f, \Gamma_k) = \int_{\Gamma_k} ds \tag{3.39}$$

$$J_2(f, k, \Gamma) = \int_{\Omega} h_k(\delta_k(\vec{x}, \Gamma)) |\nabla f(\vec{x})|^2 d\vec{x}. \tag{3.40}$$

The first term $-L(f; y)$ is the negative log-likelihood that measures the "faithfulness" of the reconstructed object to the measured data. The term $\int_{\Gamma_k} ds$ penalizes the length of the boundary, so that the boundary curves remain smooth. The term $J_2(f, k, \Gamma)$, which is rotationally invariant, penalizes local gradients inside each region more than local gradients close to the boundary; $\delta_k : \mathbb{R}^2 \times \mathcal{G}^k \to \mathbb{R}$ is the signed distance of $\vec{x}$ to Γ if the closest point on Γ to $\vec{x}$ lies on Γ_k, otherwise, the function is zero (hence all locations x where $\delta_k(x, \Gamma)$ is nonzero are necessarily in the two regions that are separated by Γ_k). The function $h_k : \mathbb{R} \to [0, 1]$ maps small arguments to values near zero and larger arguments to values near unity. For simplicity, we only use h_k's that belong to $C^{\infty}(-\infty, +\infty)$. The J_2 term has a similar effect on the reconstructed image as the penalty described by Eqn (3.5); but in J_2, how much the local gradients at a specific location is penalized is decided by where this location is with regard to the boundary, hence the penalty is "nonlocal". An example of h_k is shown in Figure 3.9. Figure 3.9a shows a one-dimensional object. Figure 3.9b

---

[5]We present our model in 2D, but extension to 3D should be straightforward.
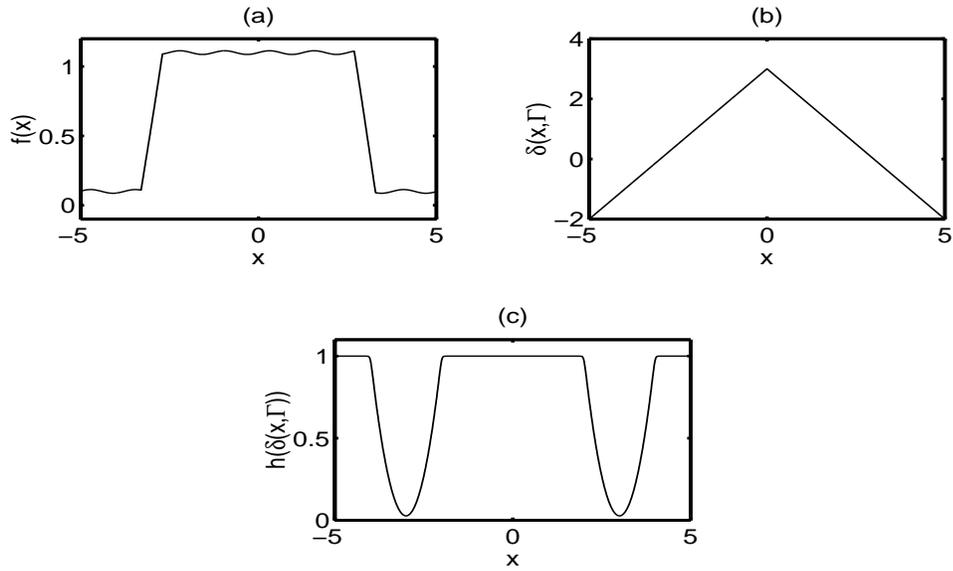
Figure 3.9: (a) Example of a 1-D object (b) Signed distance to the boundary (c) One possible h function.

shows the signed distance to the boundary, in this case, the boundary points are at -3 and $+3$. Figure 3.9c illustrates the type of h function we may want to use; the value of $h_k(\delta_k(\cdot))$ is 1 well inside each region, but gradually decreases close to the boundary. This approach allows larger gradients in the reconstructed object close to the boundary. In two dimensions (or three dimensions), the value of $h_k(\delta_k(\cdot))$ at every point is determined by the signed distance between the point and the boundary curves.

The weighting parameters, β, μ, and the functions h_k must be chosen carefully to avoid over-smoothing of the reconstructed image or the boundary curves. If we know certain regions are more nonuniform than other regions, *e.g.*, the lungs in PET transmission scans, then we should design the corresponding h_k's such that the local gradients are less penalized in the lung regions. What h function one uses, *e.g.*, how "wide" the non-unity part of h is, depends on the degree of certainty in the accuracy of boundary extraction; see Section 3.3.5 for more details on how one chooses the h function. Furthermore, the curve length term (3.39), which acts to keep the boundary smooth, will cause a shorter curve to be favored against a longer curve even though the region roughness penalty J_2 hopefully keeps this force in check. Nevertheless, the associated parameter μ should always be small

33

to avoid excess shrinking of the boundary curve.

3.3.2   An Alternating Minimization Scheme

We use alternating minimization to jointly minimize the objective function given in (3.37) over f and Γ, *i.e.*, we first hold f constant and minimize Φ with regard to Γ, then using the most recent estimate of Γ, we minimize Φ with regard to f; we alternate between these two steps until convergence. When f is fixed, the second and third terms depend on Γ. We must minimize the following objective:

$$J_f(\Gamma) \;\; = \;\; \sum_{k=1}^{K} \left(\beta J_2(f, k, \Gamma) + \mu \int_{\Gamma_k} ds \right) \tag{3.41}$$

$$\Gamma^{n+1} \;\; = \;\; \arg\min_{\Gamma} J_{f^n}(\Gamma), \tag{3.42}$$

where J_2 was defined in (3.40). As is common in PDE-based image analysis, we perform steepest descent with respect to Γ. For any point $\vec{v} = (x, y)$ on the boundary Γ we evolve that point according to the following differential equation:

$$\frac{d\vec{v}}{dt} = -\frac{\delta J_f(\Gamma)}{\delta \vec{v}}, \tag{3.43}$$

where the right-hand side is the negative functional derivative of the objective. Since it is difficult to analytically derive the functional derivatives of J_2, we evaluate its functional derivatives numerically. The functional derivative of J_2 must point in the normal direction of the curve, as any movement in the tangential direction would not change the curve. We use a scheme similar to the central difference method for evaluating local derivatives. (Central differences are usually accurate to a higher order than one-sided differences.) For a given point $\vec{p_0}$ on the contour Γ, we define a function z which is zero except in the neighborhood of $\vec{p_0}$ and for which $\Gamma + z$ differs from Γ only in the normal direction. (We can imagine some force being exerted on the curve; this force is nonzero only in the neighborhood of $\vec{p_0}$; if we exert this force in the normal direction of the curve at $\vec{p_0}$ for an infinitesimal period of time, then we will have a small perturbation of the curve at $\vec{p_0}$ in the normal direction). Using this idea, we approximate the functional derivative of $J_2(\Gamma)$ at $\vec{v} = \vec{p_0}$ as follows:

$$\frac{\delta J_2}{\delta \vec{v}} \;\; \approx \;\; \frac{1}{2} \frac{J_2(\Gamma + z) - J_2(\Gamma - z)}{\Delta \sigma} \tag{3.44}$$

34

$$= \frac{1}{2} \left(\frac{J_2(\Gamma + z) - J_2(\Gamma)}{\Delta \sigma} - \frac{J_2(\Gamma - z) - J_2(\Gamma)}{\Delta \sigma} \right) \quad (3.45)$$

$$\triangleq \frac{1}{2}(u_1 - u_2), \quad (3.46)$$

where $\Delta \sigma$ is the area lying between the curve Γ and the perturbed curve $\Gamma + z$ [44].

For the last term in $J_f(\Gamma)$, the direction in which the curve length decreases most rapidly is when [90]:

$$\frac{\partial \Gamma}{\partial t} = -\kappa \vec{\mathcal{N}}, \quad (3.47)$$

i.e., the speed of the evolution at any point is the curvature of the boundary at that point, and the curve evolves in the inward normal direction. Combining (3.43), (3.45), and (3.47), we evolve the boundary using the level set method[64, 90, 54] according to the following:

$$\frac{d\vec{v}}{dt} = -(\mu \kappa + \frac{\beta}{2}(u_1 - u_2))\vec{\mathcal{N}}. \quad (3.48)$$

Evolving the curve according to (3.48) yields a curve that approximately minimizes J_f; we call this step the "boundary estimation" step.

The force exerted by J_2 in (3.41) is nearly zero in smooth regions, and is only significant close to the actual boundary where local gradients are large. Figure 3.10 illustrates this property in one dimension. Let f' denote the derivative of f in the x direction; let Γ_o denote the old boundary at 0.3 and Γ_n denote the new boundary at 0.35. Moving the boundary Γ from 0.3 to 0.35 would change $h(\delta(\cdot, \Gamma))$ from $h(\delta(\cdot, \Gamma_n))$ to $h(\delta(\cdot, \Gamma_n))$, *i.e.*, the "valley" of $h(\delta(\cdot, \Gamma))$ is moved from 0.3 to 0.35, but the change in the roughness penalty, *i.e.*, $\int (h(\delta(x, \Gamma_n)) - h(\delta(x, \Gamma_o)))|f'|dx$, would be very small. Thus evolution according to (3.48) alone would require a fairly close initialization to the actual boundary. As discussed near (3.55) below, we circumvent this problem by using the initialization procedure for the boundary that employs another force (from a global measure) which ensures that the boundary moves no matter where the contour is.

For the second stage of the minimization, we hold Γ fixed at its previous estimate Γ^n and minimize with regard to f. When Γ is held fixed, the relevant terms in the objective function (3.37) are the following:

$$J_\Gamma(f) = -L(f; y) + \beta \sum_{k=1}^{K} J_2(f, k, \Gamma) \quad (3.49)$$

35
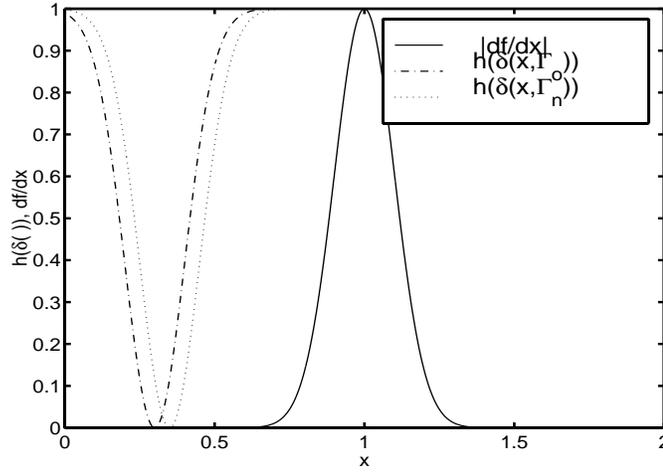
Figure 3.10: Plot of change in J_2 when the "curve" is very far from the actual boundary

$$J_2(f, k, \Gamma) = \int_\Omega h_k(\delta_k(\vec{x}, \Gamma))|\nabla f(\vec{x})|^2 d\vec{x}. \tag{3.50}$$

Hence we minimize $J_\Gamma(f)$ with regard to f as follows:

$$f^{n+1} = \arg\min_f J_{\Gamma^n}(f). \tag{3.51}$$

When updating the boundary using (3.42), the h function in J_2 pushes the boundary toward image locations where the gradient is large; when updating the objective f using (3.51), the h function imposes a space-varying weighting of the penalty on local gradients. This weighting depends on the signed distance from each pixel to the nearest estimated boundary. Every term in (3.49) is quadratic in f, except possibly the log-likelihood term, which involves logarithms in the case of Poisson measurements. Therefore, the minimization problem (3.51) is a standard penalized likelihood problem, and we can minimize J_Γ over f using methods such as the conjugate gradient method [73, 41] (if quadratic) or the paraboloid surrogates/coordinate ascent method (if not) [30].

We iteratively alternate between the two steps (3.42) and (3.51). Both these two steps will, under ideal circumstances[6], monotonically decrease the objective as defined in (3.37). In addition, the objective is bounded below, so the algorithm will presumably converge toward a local minimum.

---

[6]Under realistic circumstances, where dt is taken to be finite, the minimization of (3.41) according to a discretized version of (3.43) may not be exactly monotonic. Such effects are inevitable when continuous formulations are discretized.

3.3.3 Discretization

We discretize the image f using the usual square grid. We discretize the level sets, which embed the boundary curves, into a square grid using the same sampling grid as the image. However, the sampling spacing of the boundary curve itself may be finer than the image pixel spacing. Equation (3.48) describes the evolution of the boundary. The first term is the smoothing term which is independent of the image; its implementation using the level set method is as described in [64]. In addition to smoothing term, the evolution of the boundary is determined by the functional derivatives of J_2.

To implement the approximation described in (3.48), we observe that for a point $\vec{p}_1$ on the boundary curve to be the closest to a point $\vec{p}_0$ in the image, the line connecting $\vec{p}_0$ and $\vec{p}_1$ must be perpendicular to the tangent line of the boundary curve at $\vec{p}_1$ (see Figure 3.11). Thus if we make a small enough perturbation of the boundary curve in a neighborhood of $\vec{p}_1$, the only points in the image that are possibly affected by this change of the boundary curve (in terms of their distances to the curve) will be restricted to a narrow band perpendicular to the boundary curve (see Figure 3.12). Since this band can be made arbitrarily narrow by making the boundary curve perturbation small enough, we can make the approximation that the image values remain constant in the lateral direction of the band and evaluate the functional derivative of J_2 using equally spaced points on the line perpendicular to the boundary curve at $\vec{p}_1$; we use bi-cubic interpolation from the neighboring points [73] where ∇f is not available. At every time step of the evolution of the boundary curve, we also keep a record of which point on the boundary each image point is closest to. When we evaluate the functional derivative of J_2 at a point $\vec{p}_1$, we make the approximation that a very small movement in the curve at $\vec{p}_1$ will not cause any image point that is not closest to $\vec{p}_1$ previously to be closest to $\vec{p}_1$ after the movement of the boundary curve, hence we only need to evaluate the change in J_2 on those points that are already closest to $\vec{p}_1$. In total, the evaluation of the functional derivative of J_2 is an $O(n)$ operation, where the image is of size $n \times n$.

Using these ideas, we arrived at the following conjecture for the expression of the func-
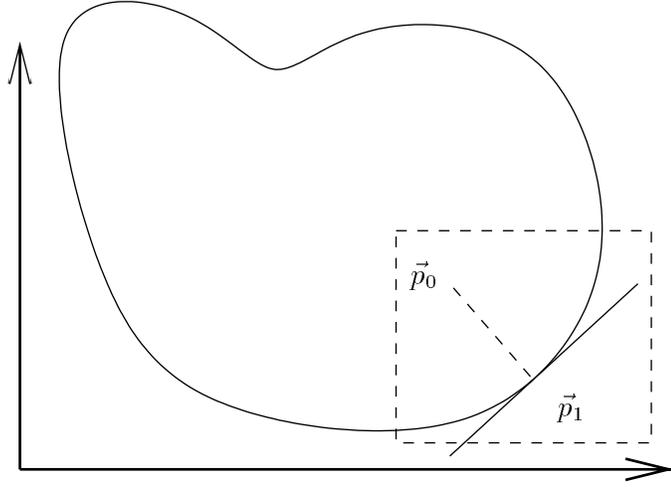
Figure 3.11: $\vec{p}_1$ is the closest point to $\vec{p}_0$ on the curve.

tional derivative of J_2:

$$\frac{\delta J_2}{\delta \vec{v}} = \left(\int_l (h_k \circ \delta_k)'(r) |\nabla f(l(r))|^2 I_{(l(r),\Gamma_k)}(r) dr \right) \vec{\mathcal{N}}, \qquad (3.52)$$

where $l(r)$ denotes the line perpendicular to the boundary curve Γ_k at the point $\vec{v}$ (l is parameterized by r such that r increases in the same direction as the outward normal vector of Γ_k at $\vec{v}$), $(h_k \circ \delta_k)'$ denotes the derivative of the function $(h_k \circ \delta_k)$ with regard to its first parameter, and $I_{(l(r),\Gamma_k)}$ is an indicator function:

$$I_{(l(r),\Gamma_k)}(r) = \begin{cases} 1 & \vec{v} \text{ is the closest point in } \Gamma \text{ to } l(r) \\ 0 & \text{otherwise.} \end{cases} \qquad (3.53)$$

A proof of this conjecture has been elusive. The difficulty mainly lies determining the size of "critical points" with regard to a closed curve, *i.e.*, whether it has Lebesgue measure 0; we define a critical point is to be a point in $\mathbb{R}^2$ that has more than one closest point on the curve Γ. In Figure 3.13, the solid line denotes the curve Γ and the dotted line denotes the set of critical points. With this simple shape, the set of critical points on the line perpendicular to every point has Lebesgue measure 0 in $\mathbb{R}$. But for a more complex shape, the author does not know how the set of critical points will behave. The ϵ-neighborhood theorem states[49]: for a compact boundaryless manifold Γ in $\mathbb{R}^M$ and a positive number ϵ, let Γ^ϵ be the open set of points in $\mathbb{R}^M$ with distance less than ϵ from Γ; if ϵ is sufficiently small, then each point $w \in \Gamma^\epsilon$ possesses a unique closest point in Γ. However, the theorem only

38
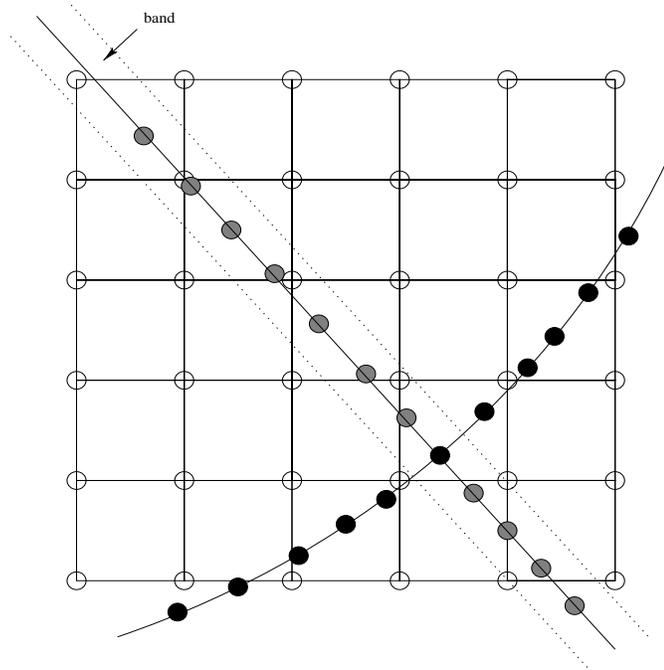
Figure 3.12: White dots denote image points; black dots denote boundary points; J_2 evaluated on points represented by shaded dots.

works in an ϵ-neighborhood of the curve, and how large ϵ can be depends on how high the maximum curvature of the curve is; the author does not know of any theorem that characterizes number of closest points (on Γ) for all points in $\mathbb{R}^M$.

Even though we cannot prove the correctness of (3.52), the boundary curves seem to evolve to a local minimum of (3.41) when we use a discretized version of (3.52) to implement our method.
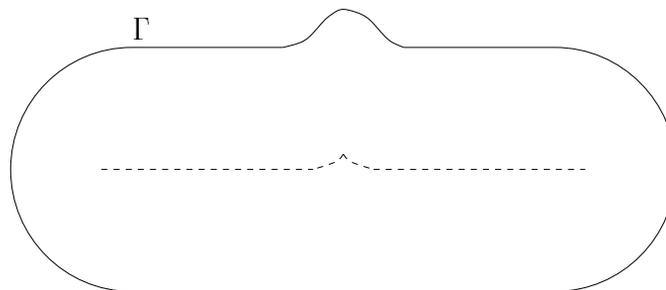
Figure 3.13: The solid line denotes the curve Γ; the dotted line denotes the set of critical points in $\mathbb{R}^2$.

3.3.4 Initialization

We first use a local penalty such as the space-invariant quadratic penalty (as described in (3.1) and (3.2)) to reconstruct an initial image estimate f^0. As discussed in Section 3.3.2, performing steepest descent with respect to the objective given in (3.41) may not push the curve toward the boundaries if the initial curve is too far away from the boundary. To address this limitation, we add a third penalty J_3 (within-region homogeneity) to J_f:

$$\tilde{J}_f(\Gamma) \triangleq \sum_{k=1}^{K} \left(\beta J_2(f, R_k) + \mu \int_{\Gamma_k} ds \right) + \sum_{m=1}^{M} \gamma J_3(f, R_m(\Gamma)) \qquad (3.54)$$

$$J_3(f, R_m) \triangleq \int_{R_m} \left| f(\vec{x}) - \frac{\int_{R_m} f(\vec{x}')d\vec{x}}{\int_{R_m} d\vec{x}} \right|^2 d\vec{x}, \qquad (3.55)$$

J_3 penalizes the difference between every pixel value and the average pixel value of its region. This is a global measure which exerts a force on the curve no matter how close the boundary estimate is to the image gradients. The evolution of the curve, as determined by J_3, is essentially a competition between bordering regions. Each pixel on the curve borders two regions; each of these two regions exerts a force trying to pull the pixel inside; the boundary will evolve towards whichever region exerts a stronger force, as determined by (3.55). We gradually reduce γ to zero. Eventually, we only rely on J_2 to move the curve to a local minimum of $J_f(\Gamma)$[7]. The implementation of the differential equation as described in (3.43) using the level set method ("narrow band" approach), taking into account of all the forces, is[64, 90]:

$$\frac{\partial \Psi}{\partial t} = \|\nabla \Psi\| \left(\mu \text{div} \left(\frac{\nabla \Psi}{\|\nabla \Psi\|} \right) + F_0 \right), \qquad (3.56)$$

where Ψ is the function that embeds the curve Γ_k, and $F_0(\vec{v}) = \beta \frac{\delta J_2}{\delta \vec{v}} + \gamma \sum_{m=1}^{M} \frac{\delta J_3(f, R_m)}{\delta \vec{v}}$ is only meaningful on level 0; the extension of the speed term F_0 has been described in Section 3.2.3.

---

[7] In practice, we run a fixed number of iterations determined sufficient for the boundary curve to converge to a local minimum. One could also stop the curve evolution when the maximum force exerted by J_2 falls below a preset threshold.

3.3.5 Deterministic Annealing

To form an initial estimate of the image f prior to applying boundary estimation step, we perform penalized likelihood tomographic reconstruction using a standard space-invariant penalty. There is no h function as given in (3.40) in such a reconstruction, but we can think of h as simply being a constant, say unity (Figure 3.14a), *i.e.*, h is independent of the boundary. But for the reconstructed image to have sharp boundaries, we must assign small weights to differences in pixel pairs close to the boundary, *e.g.*, as shown in Figure 3.14b. Thus to avoid getting stuck in a poor local minimum, we must change the shape of h function gradually from the initial constant function to the desired h function, *i.e.*, we employ deterministic annealing.

<center>(a)   (b)</center>
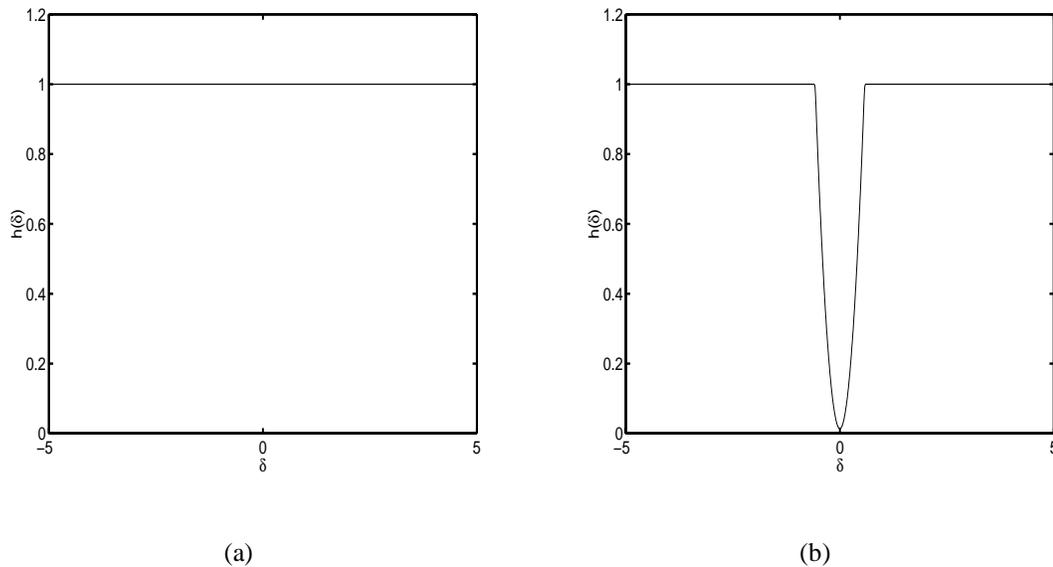
Figure 3.14: (a) Implicit h function used in local regularization. (b) Edge-preserving h function as a function of signed distance to the boundary.

Instead of going from the implicit constant function to the desired h function in one step, we take several steps. Suppose that from empirical experience with a given category of images with similar noise levels (*e.g.*, 3-minute PET transmission scans of the thorax), we have found that the initial boundary is within, say, five pixels. Then we assign small weights (via the h function) to all pixel pairs within a distance of five or six pixels to the detected boundary, and assign large weights (unity) to all other pixel pairs; thus neighboring

<center>41</center>

pixels that are more than five or six pixels away from the detected boundary will be coupled, while the boundary is allow to shift to within those pixels between which the weights are very small. We gradually evolve the h function from the constant function toward the final desired h function, as shown in Figure 3.15. Hopefully, the final boundary will eventually be a very good local minimum. The functions used in Figure 3.15 are [60]:

$$h(t) = -\log\left(\frac{e^{-b\lambda t^2} + e^{-b}}{2}\right),\qquad(3.57)$$

with $(b, \lambda) = (2.5, 1/6),\ (5, 1/4),\ (15, 2/5),\ (25, 1),\ \text{and}\ (50, 3).$
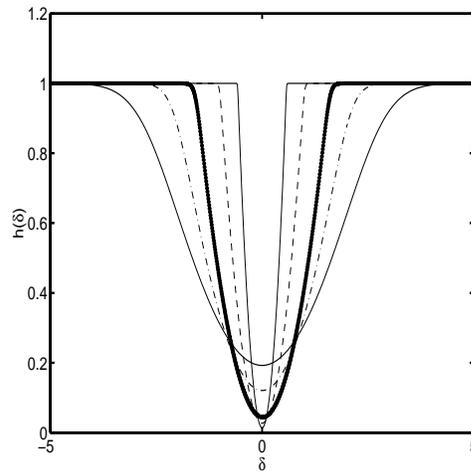
Figure 3.15: The evolution of h function

Table 3.1 shows an outline of the algorithm.

Reconstruct an initial image $f^{(0)}$ using a local penalty

for each cycle $m = 1, \ldots, M$

      choose h_k^m's, $k = 1, \ldots, k$, *e.g.*, using (3.57)

      evolve the boundaries according to (3.42) and (3.48) using the level set method

      update the image according to (3.51)

end

Table 3.1: Algorithm outline

Evolving h_k^{initial} to h_k^{desired} usually involves 4-5 "cycles"; one cycle consists of two stages, *i.e.*, the "image reconstruction" stage, and the "boundary estimate" stage[8].

---

[8]The h_k^m's need not be different for every cycle; one may hold any or all of h_k's fixed for a few cycles.

3.4 Accuracy of the Estimated Boundary

3.4.1 A Metric For the Accuracy of Estimated Boundary

Defining a metric for the accuracy of our estimated boundaries is not a trivial task. We would like this metric to have some physical meaning while at the same time, have the property of being rotational invariant. Fourier descriptors satisfy this requirement, but a change in any Fourier coefficients will lead to a change in the shape of the contour everywhere. Using B-splines would avoid this problem, but we still would have to decide how many parameters to use; and the variance of the each coefficient in the representation do not necessarily have clear physical interpretations. Thus we propose the following non-parametric way to evaluate the accuracy of our estimated boundaries: for any point on the true boundary x_0, we find the point on our estimated boundary x that has the shortest distance (among all points on our estimated boundary) to x_0, *i.e.*,

$$x = \arg \min_{x' \in \Gamma} |x' - x_0|, \tag{3.58}$$

where Γ is the estimated boundary; the vector from x_0 to x (*i.e.*, $x - x_0$) is defined as the "deformation" of our true boundary at point x_0 (see figure 3.16). Using this metric, we can average the deformation vectors across all realizations to obtain an average deformation for a given point on the true boundary; from that, we can plot our average boundary, by deforming our true boundary according to the average deformation at every point. We can also investigate average magnitude of our deformation and other properties of our boundary estimation process.

3.4.2 Accuracy of the Boundary Extraction

As discussed in Section 3.3.1, the curve length term J_1 in (3.39) favors a shorter curve. However, even if we set the weighting parameter for J_1 to zero, there would still be boundary curve shrinkage simply due to the fact that we discretize the differential equation. This effect is commonly known as *numeric diffusion*[61]: by approximating the differential equation in the discrete domain, we introduce diffusion terms into the differential equation; this new differential equation with the diffusion terms included is commonly known
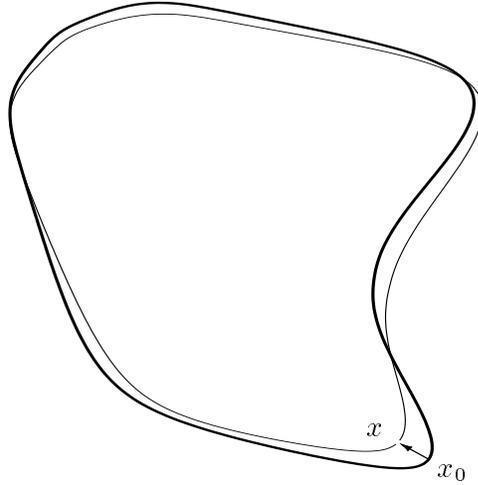
Figure 3.16: Deformation of point x_0 on the true boundary (thick line) to point x on the estimated boundary (thin line).

as the *modified equation*. Another way to put it is: we are approximating the modified equation better than the original differential equation. The effect of numerical diffusion can be reduced by using finer discrete grids; however, as the grid gets finer, the computational cost goes up at least linearly.

To quantify how much the estimated boundary curve shrink, we focus on the boundary extraction part of the proposed method, *i.e.*, with image f known, we evolve the curve according to (3.43). First, we apply the proposed method, with the weight of J_1 set to zero, to a noiseless image. We use the output of Matlab's contour program as the ideal boundary extraction ("curve 1"). The noiseless image and the boundary extraction obtained by Matlab's contour program are shown in Figure 3.17a; the boundary extraction obtained with the proposed method ("curve 2") is shown in Figure 3.17b. The deformation of the ideal boundary extraction (arrows point from "curve 1" to "curve 2"), scaled by 10, is shown in Figure 3.18. The maximum inward deformation is 0.2767 pixels; the maximum outward deformation is 0.2765 pixels; the average signed deformation is 0.0155 pixels inward; the average magnitude of deformation of all the pixels on the boundary curve is 0.0995 pixels. The deformation shown here is solely due to discretization, since the weight assigned to J_1 is set to zero.

Then, we applied the proposed method to an image contaminated by Gaussian noise.

(a)   (b)
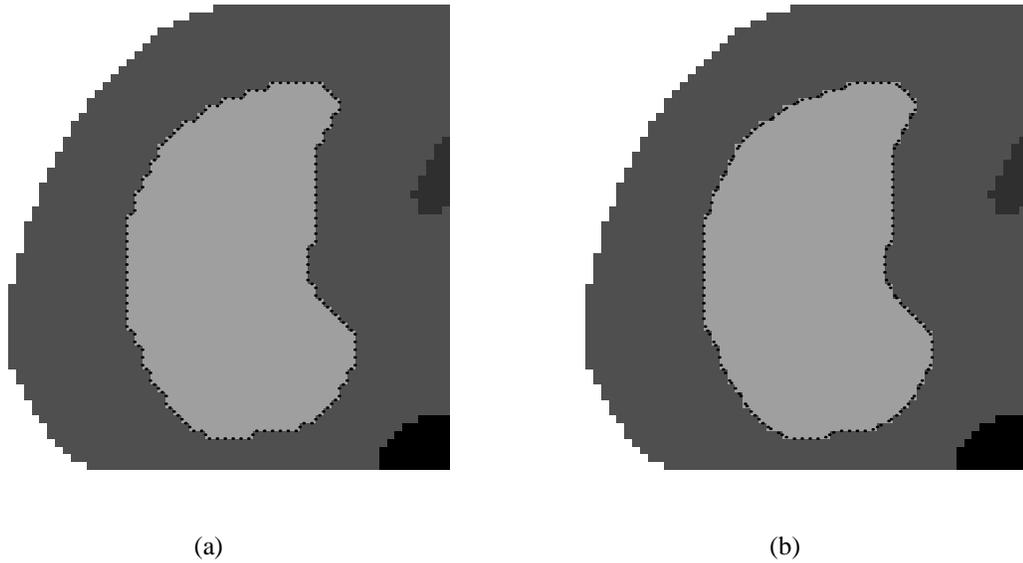
Figure 3.17: Noiseless image (a) Boundary extraction by Matlab's contour program (b) Boundary extraction with the proposed method

Figure 3.19 shows boundary extraction of such an image; this time, the weight assigned to J_1 is not set to zero. We performed 50 realizations, and obtained the average boundary curve. We compare the average boundary curve to the curve extracted from the noiseless image by the proposed method ("curve 2"). The deformation of "curve 2", scaled by 10, is shown in Figure 3.20. The maximum inward deformation is 0.1401 pixels, the maximum outward deformation is 0.1377 pixels; the average signed deformation is 0.0160 pixels inward; the average magnitude of deformation of all the pixels on the boundary curve is 0.0576 pixels.
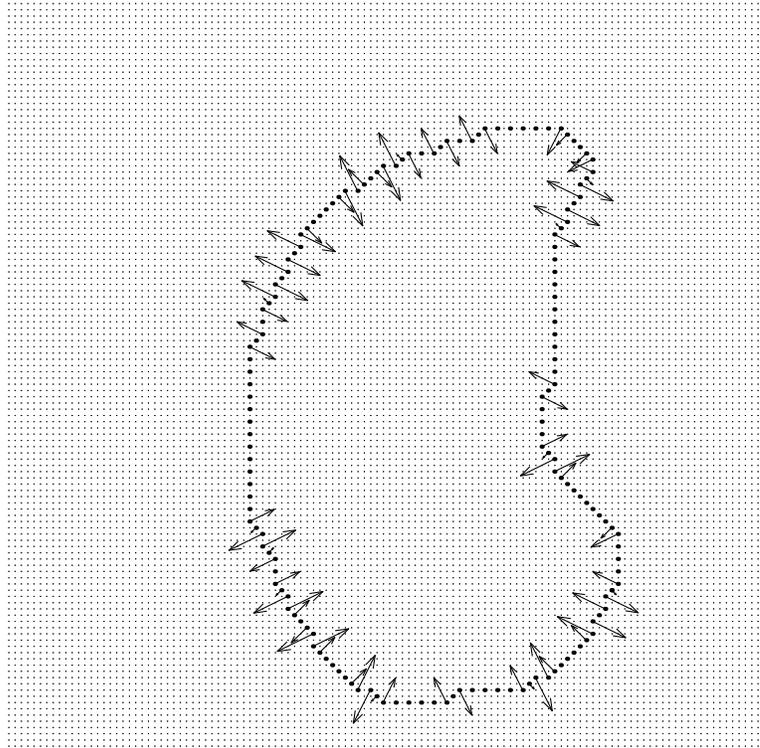
Figure 3.18: Effect of discretization: deformation of the boundary curve extracted by Matlab's contour program ("curve 1"), scaled by 10
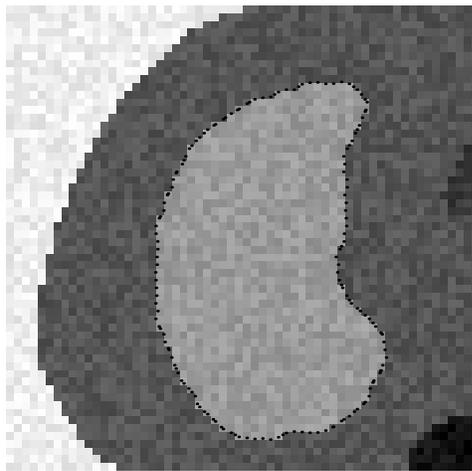
Figure 3.19: Noisy image: boundary extraction with the proposed method
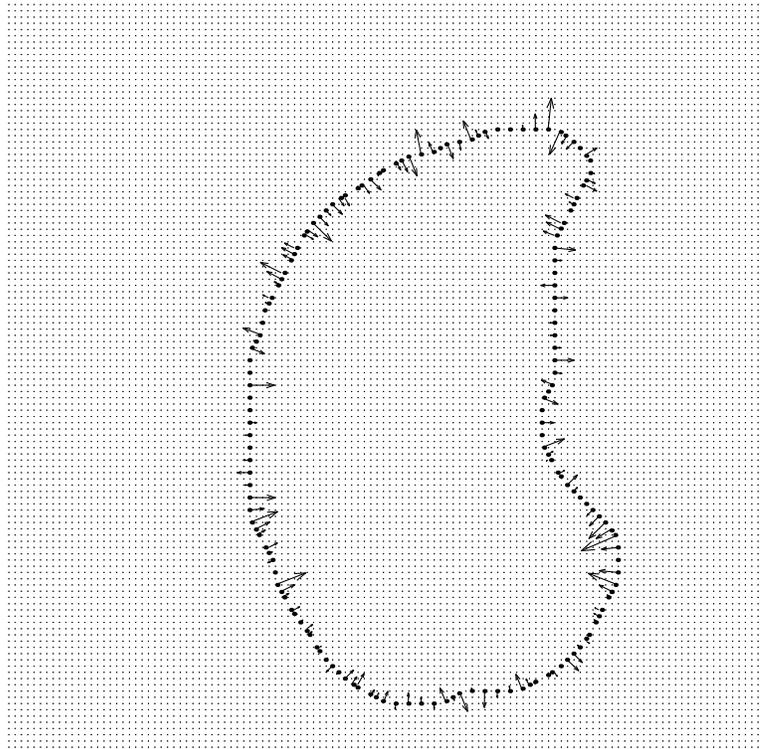
Figure 3.20: Effect of image smoothing and J_1 term: deformation of the boundary curve extracted from the noiseless image by the proposed method ("curve 2"), scaled by 10

3.5 Statistical Results

In this section, we compare the proposed algorithm to the edge-preserving reconstruction method described in [31] which is based on local regularization on L, hereafter referred to as the "Huber method", in terms of bias/variance tradeoffs. We simulated PET transmission scans of a digital phantom that resembles the human thorax. The body has attenuation coefficient $0.096/\mathrm{cm}$. Inside the body, there are two "lungs"; the left lung has nonuniform attenuation coefficients; the right lung has uniform attenuation coefficients; both lungs have average attenuation coefficient $0.022/\mathrm{cm}$. The "spine" has uniform attenuation coefficients $0.13/\mathrm{cm}$. Figures 3.8 shows the phantom and a profile of the phantom. The image consists of 128×128 pixels, each of size $0.42 \times 0.42\mathrm{cm}^2$. The sinogram consisted of 192 radial samples and 160 angular samples, similar to the CTI ECAT921 PET scanner; the ray spacing is 0.3375 cm. Random coincidences account for about 5% of the recorded counts. We performed two studies, one with 1,000,000 counts, the other with 300,000 counts (about equivalent to a 10 and 3-minute scans, respectively). Our initial image was reconstructed using conventional space-invariant quadratic penalty over first-order neighbors for the proposed method. The boundaries were initialized manually[9] as shown in Figure 3.21.

For the transmission reconstructions, we considered two regions of interest (ROI) in the reconstructed images: the left lung (region 1) and the right lung (region 2). We performed 50 realizations for this study. For each ROI, we calculated the average attenuation coefficient from each group of 50 ROI values. For the Huber penalty, we plotted ROI bias versus standard deviation as a function of the regularization parameter β, for three δ: 0.002, 0.004, and 0.008, where 0.004 is about $1/10$ of minimum contrast; for the proposed penalty, we manually selected β and h to cover a range of bias-variance tradeoffs.

For the high count case, one realization from FBP, Huber, and proposed reconstructions are shown in Figure 3.22. A comparison of two profiles (row 65) from the Huber and the proposed reconstruction methods are shown in Figure 3.23. We selected images of similar bias to be shown, *i.e.*, the average attenuation coefficients of the left lung region (and the right lung region) from images reconstructed with all three methods have similar

---

[9]A procedure can be easily developed to obtain better initial boundaries.
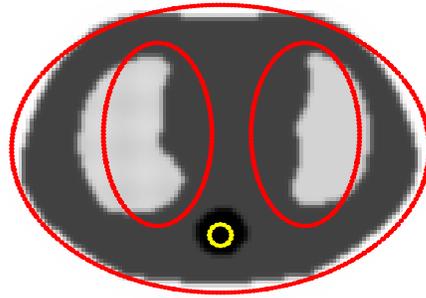
Figure 3.21: Initial contours

bias. Judging the variability in the Huber and the proposed reconstructions, we can see that the proposed method yields less variance at this bias level[10]. This initial observation is confirmed quantitatively by plotting the bias against the variance of the two ROI's we defined above (Figure 3.24, error bars in the bias direction are too small to be shown).

---

[10]The unevenness of the reconstruction using the proposed penalty near the boundary, as shown in Figure 3.22c, is due to the fact that the h function is near zero at these locations. A simple remedy to this problem would be to increase the h function near the boundary curves. One could also use a "directional" penalty, *i.e.*, penalize more heavily in the tangential direction of the boundary curve, while decrease the penalty in the perpendicular direction.
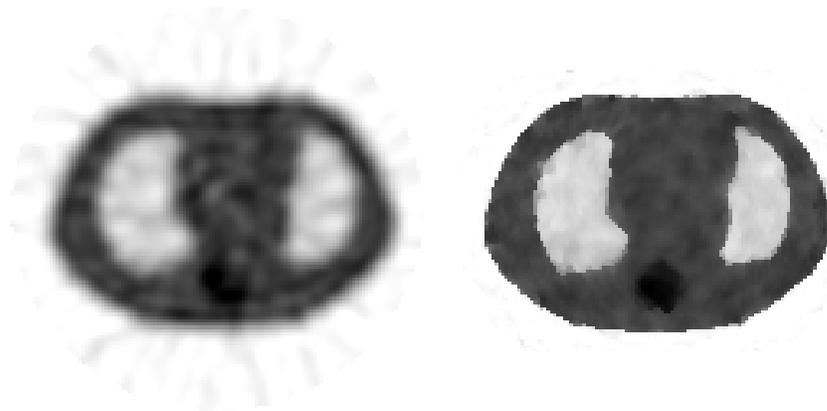
(a)

(b)

(c)

Figure 3.22: 1,000,000 counts transmission reconstruction (a) FBP reconstruction (b) reconstruction with Huber penalty (c) Proposed penalty
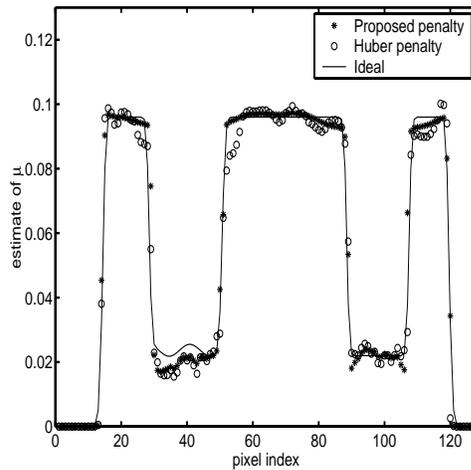
Figure 3.23: 1,000,000 counts transmission reconstruction: Comparison of Huber penalty and proposed penalty
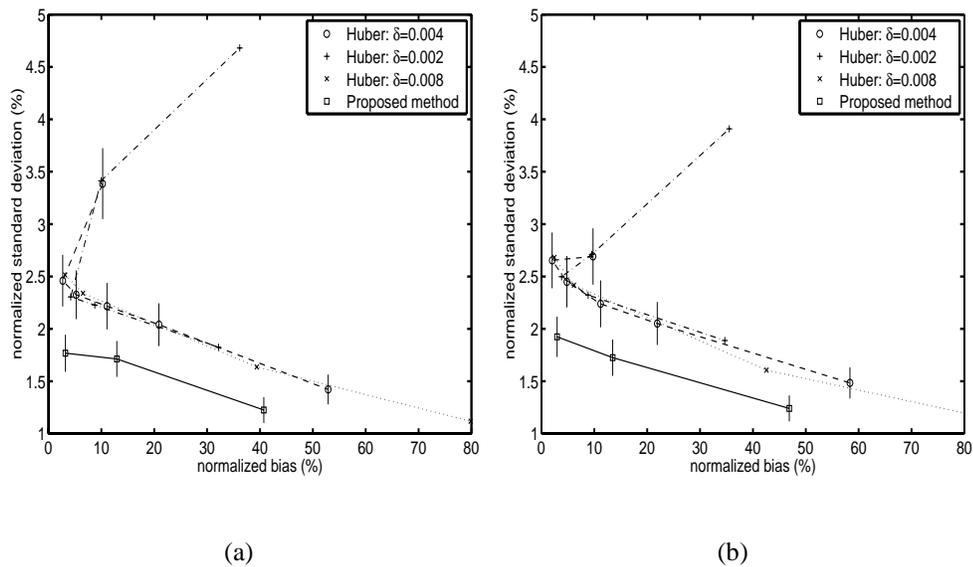
Figure 3.24: 1,000,000 counts transmission reconstruction, bias-variance plot of (a) Region 1 (b) Region 2

To investigate how the proposed non-local regularization performs when the transmission map is applied to the attenuation correction of emission reconstructions (for details, see [89]), we simulated PET emission scans of the digital phantom shown in Figure 3.25. The relative activities of the lungs, spine, heart, and body were 12, 9, 40, and 22, respectively. The emission projections included the effects of nonuniform attenuation corresponding to the attenuation map in Figure 3.21. We reconstructed emission images using FBP with attenuation correction based on two sets of transmission maps: one set regularized by the proposed non-local penalty, and the other set by the Huber penalty.
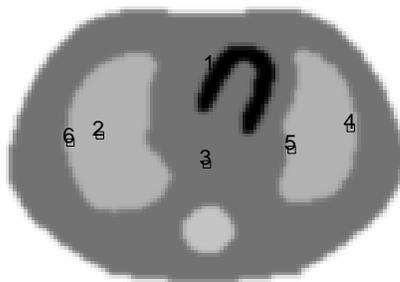
Figure 3.25: Emission phantom

Figure 3.27 shows emission reconstructions using transmission maps reconstructed with the Huber penalty and the proposed penalty. To compare the effect of the two different penalties on the emission reconstruction, we selected two groups of reconstructed transmission maps, one using the Huber penalty, and the other using the proposed penalty. Then [17] we smoothed the projection of the emission phantom (with ideal attenuation applied) and the transmission map, so that the resolution of the final emission reconstructions using Huber penalty and the proposed penalty matched each other[11] We use the follow-

---

[11]The smoothing was done because the reconstructed attenuation maps using the proposed penalty have fairly sharp edges. Because of noise, the boundary detection is not perfect and severe artifacts will result if we do not smooth the reconstructed attenuation map. We also attempted to reconstruct attenuation maps with blurry edges directly, *i.e.*, use "broader" h functions. We are, however, less successful with this approach, since the reconstruction with space-invariant h functions have highly spatially varying resolution. To achieve uniform spatial resolution, one would most likely need space-varying h functions.

ing simple method to determine the resolution of a set of reconstructions. Given the ideal image μ^{true} and the average reconstructed image $\check{\mu}$, the resolution of $\check{\mu}$ is:

$$\arg\min_\sigma \sum_{j \in \mathcal{M}} \left|[G_\sigma \mu^{\mathrm{true}}]_j - \check{\mu}_j\right|^2 \tag{3.59}$$

where G_σ represents a Gaussian smoothing filter with FWHM σ, and $\mathcal{M}$ denotes a mask used during reconstruction. The normalized[12] standard deviation (in units of %) for the mean of the six regions of 3×3 pixels, with a fixed spatial resolution of 4.2 pixels (computed according to (3.59)), are shown in Table 3.2.

	Interior			Boundary		
Region	1	2	3	4	5	6
Huber penalty	4.8 ± 0.5	11.0 ± 1.1	6.7 ± 0.7	14.8 ± 1.5	16.4 ± 1.6	15.9 ± 1.6
Proposed penalty	1.1 ± 0.1	8.1 ± 0.8	1.9 ± 0.2	14.4 ± 1.4	17.1 ± 1.7	18.1 ± 1.8

Table 3.2: Normalized standard deviation (%) for the six ROI's using 1,000,000 counts transmission reconstructions

We also calculated $\sigma_{\mathrm{proposed}}/\sigma_{\mathrm{Huber}}$ for all pixels within the image; Figure 3.26 shows a histogram of these standard deviation ratios. For 85.8% of pixels, the proposed penalty produced lower standard deviations than the Huber penalty. The median reduction in the standard deviations is 47.1%.

---

[12]Normalized by the mean of each region.
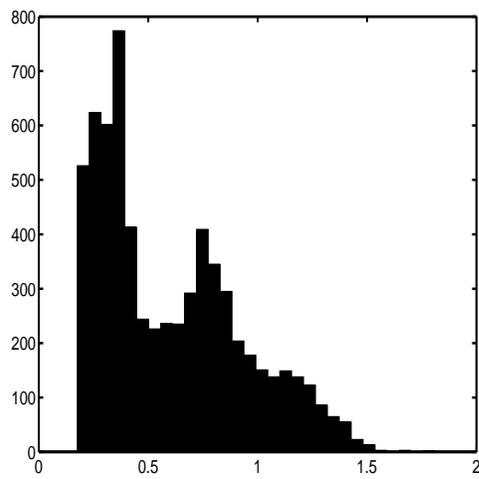
Figure 3.26: Histogram of $\sigma_{\mathrm{proposed}}/\sigma_{\mathrm{Huber}}$ for emission reconstruction using transmission scans with 1,000,000 counts
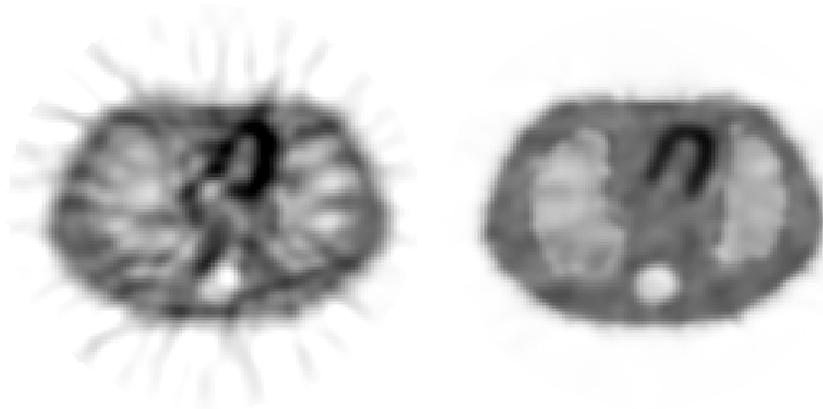
(a)   (b)

(c)

Figure 3.27: Emission reconstruction using transmission scans (1,000,000 counts) reconstructed by (a) FBP (b) Huber penalty (c) Proposed penalty

We performed similar studies for the low count case. One realization from the FBP, Huber, and proposed transmission reconstructions are shown in Figure 3.28. (The streaks in the FBP reconstruction in this case are caused by the fact that some rays recorded zero counts). Figure 3.29 compares two profiles (row 65) from the Huber and proposed reconstruction methods; these profiles came from images of similar bias. Figure 3.30 shows the bias/variance tradeoffs for the two lung regions.

(a)   (b)

(c)
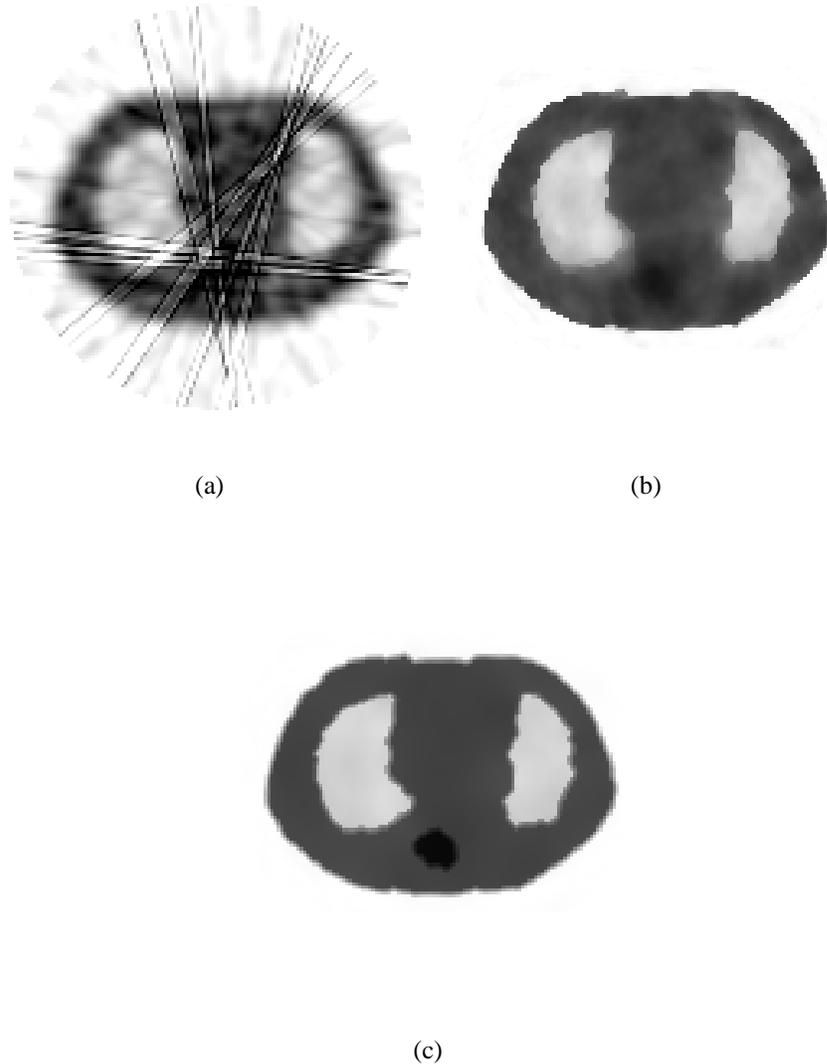
Figure 3.28: 300,000 counts transmission reconstruction (a) FBP reconstruction (b) Reconstruction with Huber penalty (c) Proposed penalty
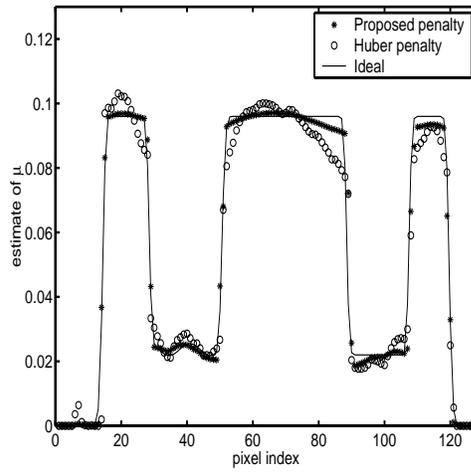
Figure 3.29: 300,000 counts transmission reconstruction: Comparison of Huber penalty and proposed penalty

(a)   (b)
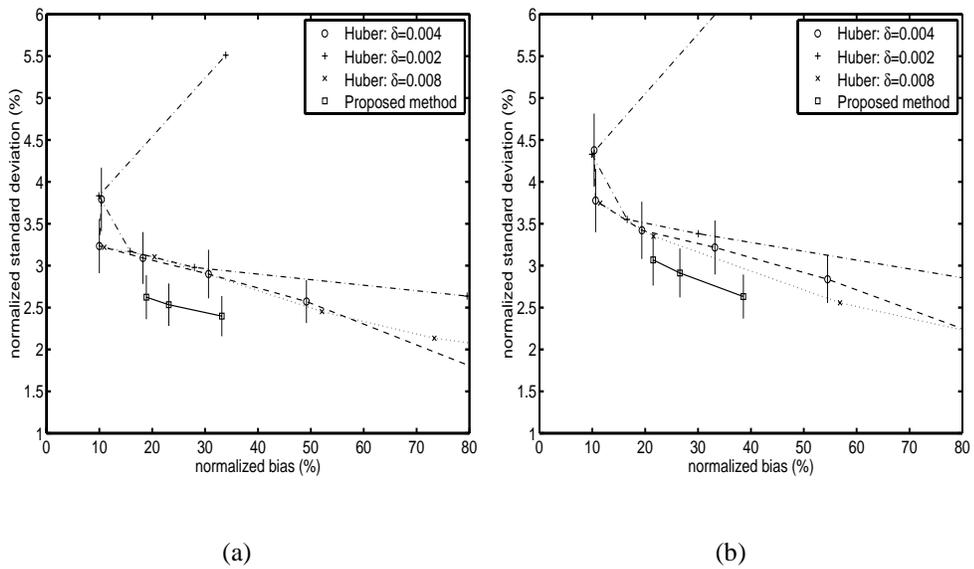
Figure 3.30: 300,000 counts transmission reconstruction, bias-variance plot of (a) Region 1 (b) Region 2

The emission reconstructions from noiseless emission data corrected using transmission maps regularized by the proposed penalty and the Huber penalty are shown in Figure 3.32. We observe that the heart region is much more uniform in reconstructed images using the proposed penalty than using the Huber penalty. The histogram of $\sigma_{\mathrm{proposed}}/\sigma_{\mathrm{Huber}}$ for all pixels within the image is shown in Figure 3.31. For 81.7% of pixels, the proposed penalty produced lower standard deviations than the Huber penalty. The median reduction in the standard deviations is 34.6%. The normalized standard deviation (in units of %) for the mean of the six regions of 3×3 pixels as defined in Figure 3.25, with a fixed spatial resolution of 6.4 pixels (computed according to (3.59)), are shown in Table 3.3.

	Interior			*Boundary*		
Region	1	2	3	4	5	6
Huber penalty	5.0 ± 0.5	14.1 ± 1.4	9.0 ± 0.9	14.0 ± 1.4	15.7 ± 1.6	15.3 ± 1.5
Proposed penalty	1.7 ± 0.2	8.5 ± 0.9	3.9 ± 0.4	13.4 ± 1.3	18.0 ± 1.8	14.9 ± 1.5

Table 3.3: Normalized standard deviation (%) for the six ROI's using 300,000 counts transmission reconstructions
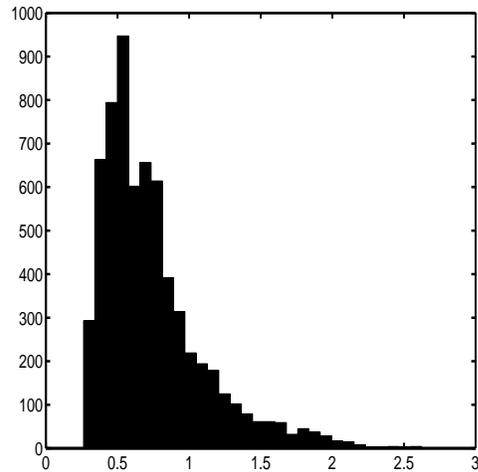
Figure 3.31: Histogram of $\sigma_{\mathrm{proposed}}/\sigma_{\mathrm{Huber}}$ for emission reconstruction using transmission scans with 300,000 counts

(a)   (b)

(c)
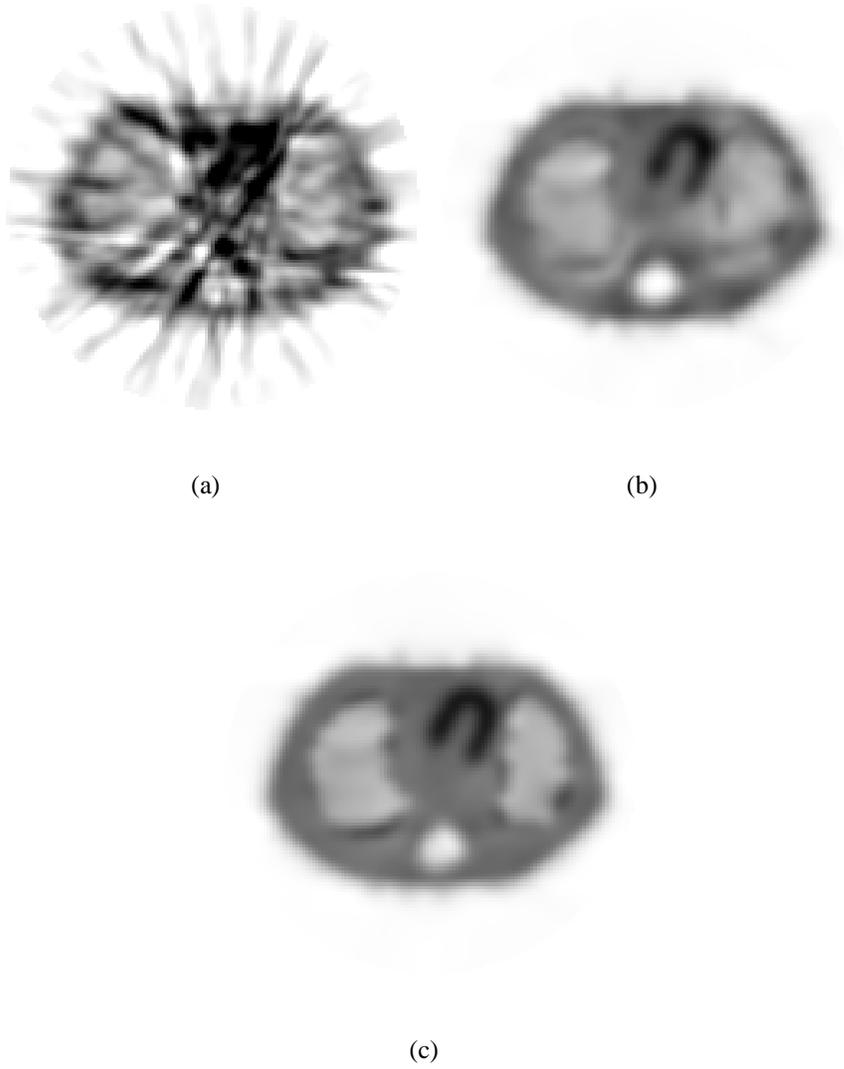
Figure 3.32: Emission reconstruction using transmission scans (300,000 counts) recon-structed by (a) FBP (b) Huber penalty (c) Proposed penalty

3.6 Extension to Three-Dimensional Images

In this section, we formulate our 3-D nonlocal edge-preserving regularization. Conceptually, this is a simple extension of our 2-D formulations. We simply replace the penalty on the curve length term, J_1 in (3.39), with a surface area functional, and the penalty J_2 in (3.40) remains the same, except $\vec{x}$ now denotes 3-D spatial coordinates. Let Γ_k denote the kth boundary surface (the number of boundary surfaces is still assumed to be fixed and known), $f(\vec{x})$ denote the 3-D object, and y denote the 3-D measurement[13].

$$J(f, \Gamma) = -L(f; y) + V(f, \Gamma) \tag{3.60}$$

$$V(f, \Gamma) = \sum_{k=1}^{K} \left[\beta J_2(f, k, \Gamma) + \mu J_1(\Gamma_k) \right] \tag{3.61}$$

$$J_1(f, \Gamma_k) = \int_{\Gamma_k} dS \tag{3.62}$$

$$J_2(f, k, \Gamma) = \int_{\Omega} h_k(\delta_k(\vec{x}, \Gamma)) |\nabla f(\vec{x})|^2 d\vec{x}. \tag{3.63}$$

To minimize $J(f, \Gamma)$ with f fixed, we still perform steepest descent on Γ as in (3.43). The conjecture we had about the expression of the functional derivative of J_2, as given in (3.52), can also be applied to (3.63); in fact, the expression of the functional derivative does not change at all, except that $\vec{v}$ now denotes a point on the surface, rather than a curve. The direction in which J_1 decreases rapidly is [90]:

$$\frac{\partial \Gamma}{\partial t} = H \vec{\mathcal{N}}, \tag{3.64}$$

where H denotes the mean curvature, *i.e.*, the arithmetic mean of the principal curvatures. Implementation of (3.64) using level sets is also given in [90]. There is a subtle theoretical difference between 2-D curve evolution and 3-D surface evolution; whereas the geometric heat equation for curves (3.47) will shrink a simple closed curve to a round point without developing singularities, the surface evolution equation (3.64) may cause the development of singularities[90]. This abstract concern is of little importance to us, since our surface is

---

[13]The 3-D measurement can be a stack of 2-D measurements, *i.e.*, there is no inter-slice rays as in the case of 2.5-D transmission scans; or it can be truly 3-D.

mainly influenced by the other penalty term J_2; however, a possible remedy to this problem has been given in [90].

We obtained a 10-minute transmission scan of a real patient on a CTI 921 PET scanner, then we thinned this data to the equivalent of a 3-minute transmission scan. The image consists of $134 \times 134 \times 47$ pixels; the sinogram has 47 slices, each consisting of 192 radial samples and 160 angular samples. The reconstructions using the proposed penalty are shown in Figure 3.33 and 3.34. A comparison of the 6th slice of the reconstruction using FBP, the Huber penalty, and the proposed penalty is shown in Figure 3.35. The boundary surfaces extracted using the proposed penalty are shown in Figures 3.36 and 3.37.

Figure 3.33: Three-dimensional transmission reconstruction with proposed penalty: slice 1 to 25

Figure 3.34: Three-dimensional transmission reconstruction with proposed penalty: slice 26 to 47
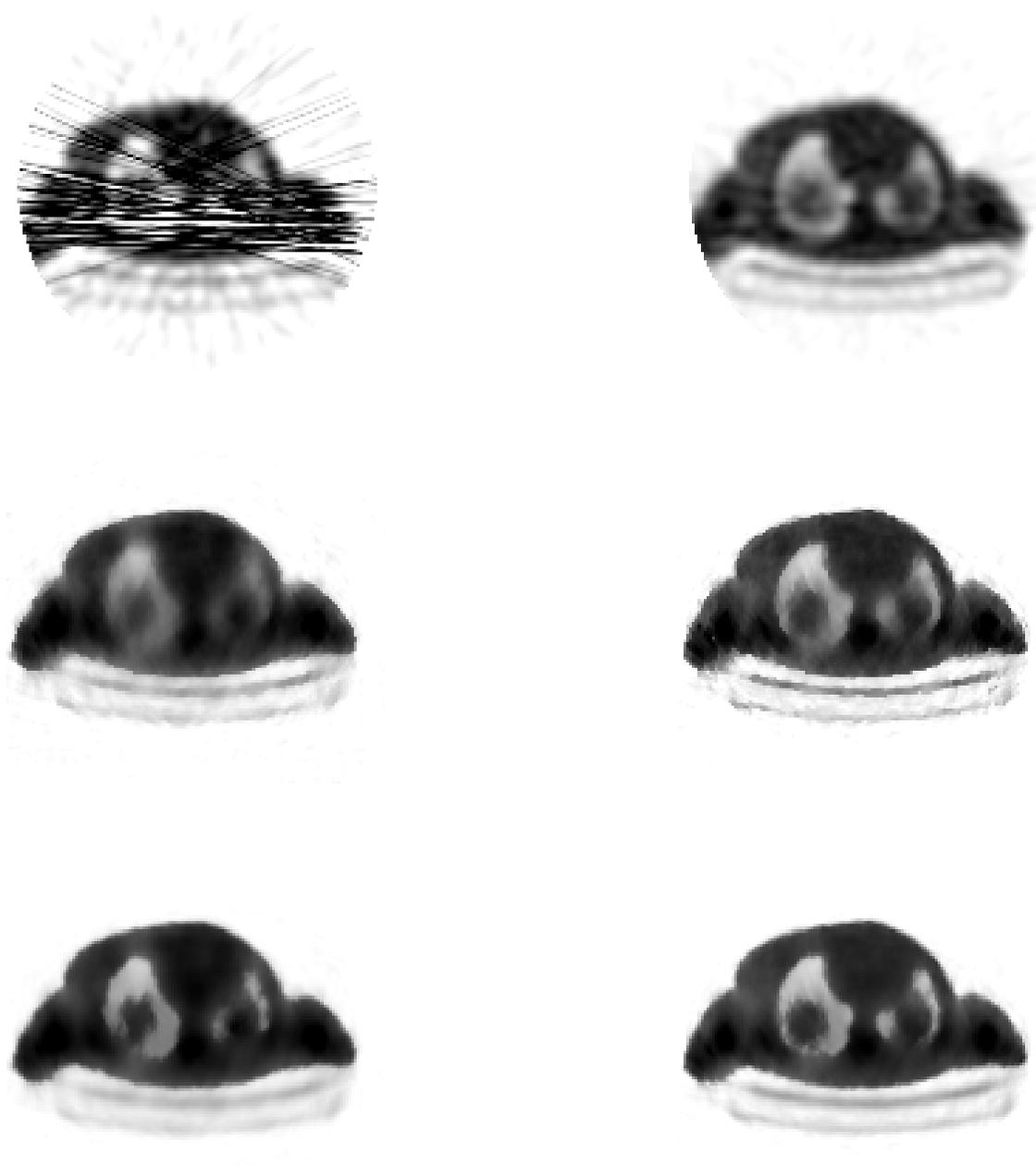
Figure 3.35: Slice No. 6: left column, 3-minute scan; right column, 10-minute scan; top row, FBP reconstruction; middle row, 3-D Huber penalty; bottom row, proposed penalty
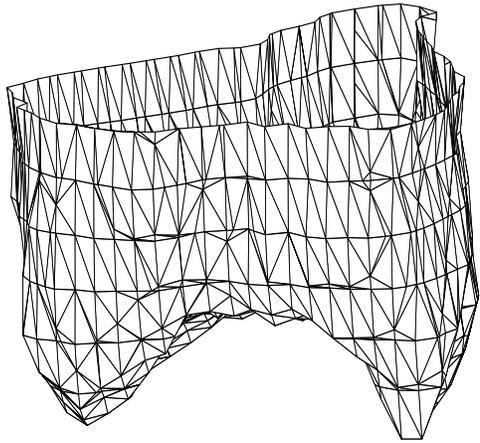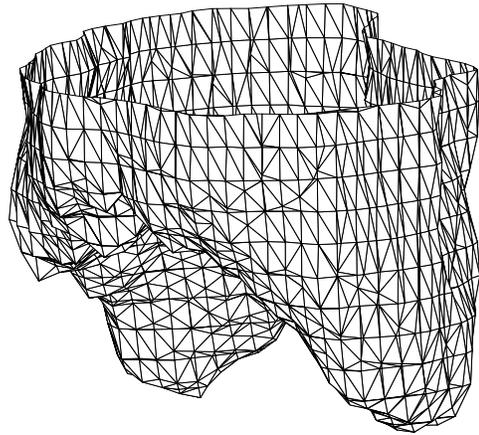
(a)   (b)

Figure 3.36: Boundary surfaces extracted by the proposed penalty: the lungs
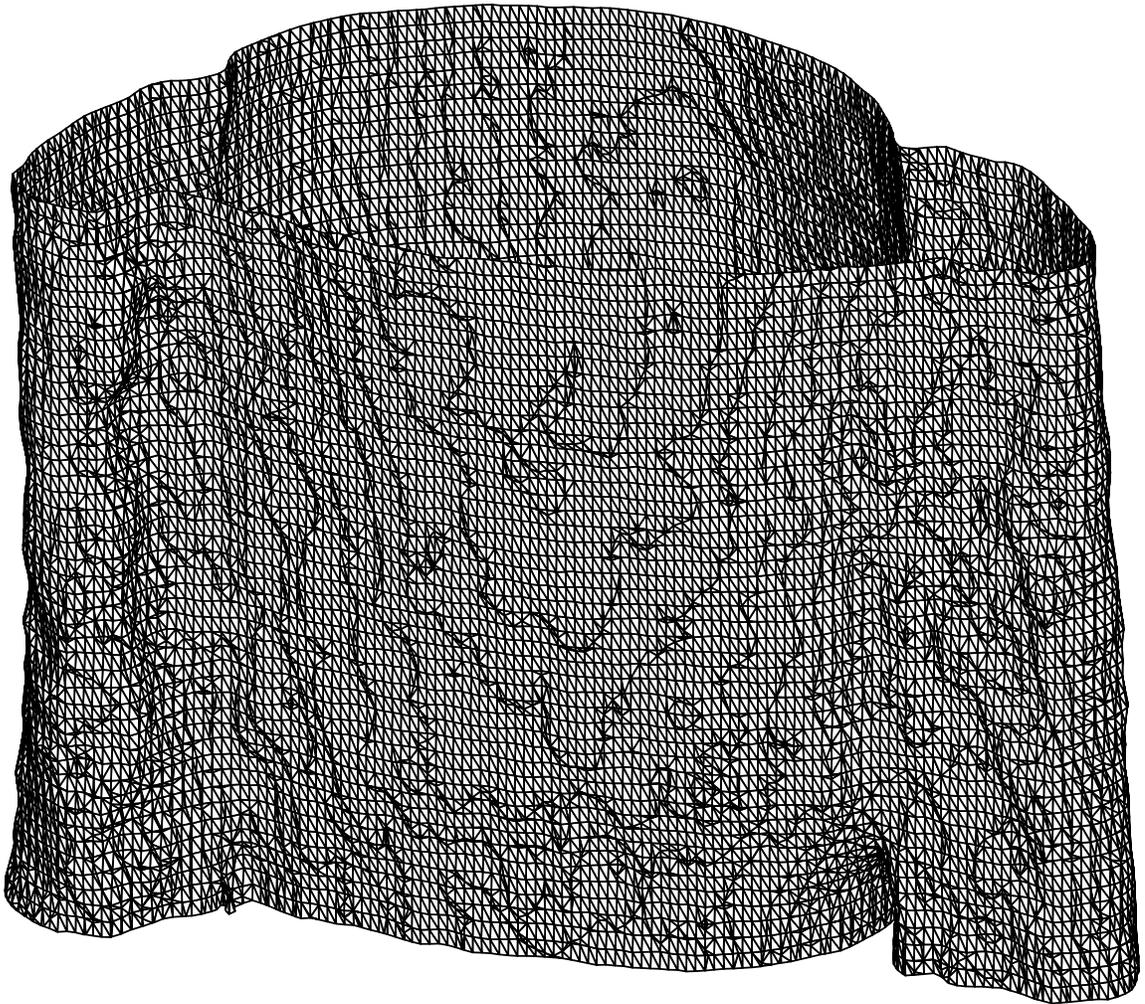
Figure 3.37: Boundary surfaces extracted by the proposed penalty: the body

(a)   (b)
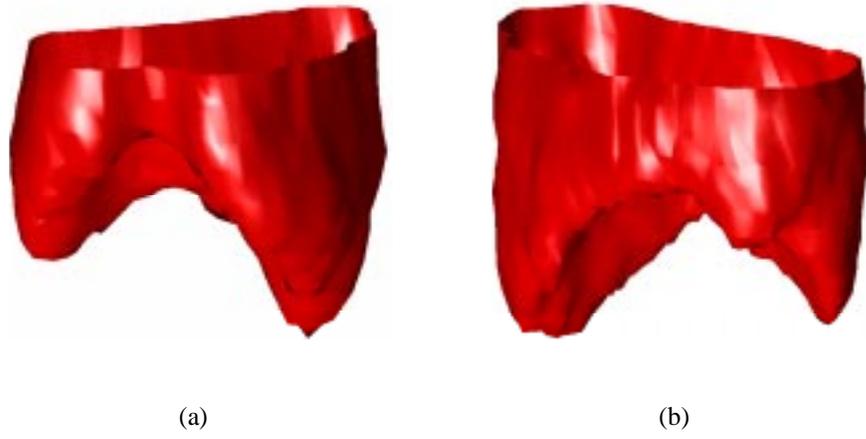
Figure 3.38: Boundary surfaces extracted by the proposed penalty: the lungs

Figure 3.39: Boundary surfaces extracted by the proposed penalty: the body

3.7 Discussion and Future Work

We have presented a new regularization method for tomographic image reconstruction based on a nonlocal penalty function. The nonlocal penalty produces transmission reconstructions with better ROI bias/variance tradeoffs than a local Huber penalty. When these transmission reconstructions are applied to ideal emission data, the nonlocal penalty used for transmission reconstruction produces emission images with smaller variances (for a fixed spatial resolution) for most pixels in the image, *i.e.*, from 80% to 85%; the median standard deviation in the image is reduced by 35% to 50%.

However, reconstruction using the proposed penalty is more time consuming than using "conventional" local penalties. Each cycle of the annealing process requires one "stage" of image reconstruction. Thus if one runs 5 cycles for the annealing process, the time devoted to updating the image is up to 5 times that of the local penalties, though for "reconstruction" stage of the second to fifth annealing cycles, we do not need to run as many iterations as in the first annealing cycle. The "boundary estimate" stage is also most time-consuming during the first cycle of the annealing process. Since we used initial boundaries that were far away from the true boundary (to demonstrate the robustness of our boundary estimate algorithm), it took about 60 time steps for the two lung boundary curves to converge (the body and spine boundary curves converges much faster). Each time step of curve evolution for each of the two lungs takes about 0.6 of the time needed for a single iteration of Huber reconstruction. The total time needed by the "boundary estimate" stages in all annealing cycles greatly depends on initialization of curves, size of the time steps, etc. In the present implementation, the boundary estimate stage of the first annealing cycle takes about twice the time of an "image reconstruction" stage. Subsequent boundary estimations were much faster since we evolved from the curve estimated by the previous cycle in the annealing process. In total, computing the boundary estimates takes about the time of three to four reconstructions[14].

A related but fundamentally different approach to the problem of reconstructing atten-

---

[14]In the present implementation, the "boundary estimate" stage is written in Matlab code, as compared to fairly optimized C code for the image reconstruction stage; hence significant improvement can be made by converting the code for "boundary estimate" to C.

uation maps based on region information has been proposed by Debreuve *et al*[24]. They proposed the following functional of the boundary curves C:

$$J(C) = \int_0^\lambda W(C(p))|C'(p)|dp,$$

where p parameterizes C and W is a data consistency term. They minimized this functional by evolving a PDE to obtain a "segmentation" of the attenuation map without actually "reconstructing" the attenuation map. Their approach assumed that the attenuation coefficient inside each region is constant, whereas our approach allows for nonuniform regions which may arise in clinical situations.

Currently, the h functions are chosen experimentally, *i.e.*, trial and error. A more systematic approach in choosing h functions, so that the transition in pixel values between neighboring regions can be carefully controlled, will make this method much easier to use. Furthermore, the choice of different h functions at various stages of the deterministic annealing process should have a great deal of effect on how "good" a local minimum the algorithm eventually converges to. A study on this effect will be also very useful toward applying this algorithm in clinical situations.

CHAPTER 4

Singles Counting Statistics Affected by Deadtime

4.1 Introduction

Every photon counting system exhibits a characteristic called *deadtime*. Since the pulses produced by a detector have finite time duration, if a second pulse occurs before the first has disappeared, the two pulses will overlap to form a single distorted pulse [78]. Depending on the system, one or both arrivals will be lost. In PET or SPECT scanners, the length of pulse resolving time, often just called "deadtime", denoted τ, is around 2μs. Counting systems are usually classified into two categories: nonparalyzable (type I) or paralyzable (type II). In a nonparalyzable system, each recorded photon produces a deadtime of length τ; if an arrival is recorded at t, then any arrival from t to $t + \tau$ will not be recorded. In a paralyzable system, each photon arrival, whether recorded or not, produces a deadtime of length τ; if there is an arrival at t, then any arrival from t to $t + \tau$ will not be recorded. In some SPECT systems [29], we encounter a third model that is similar to the paralyzable model: if two photons arrive within τ of each other, then neither photon will be recorded (*e.g.*, due to pulse pile-up); we call this the type III model. The asymptotic moments of the nonparalyzable model are well known [35]. For the paralyzable model, the exact expression for the mean of the number of recorded events from time 0 to t, denoted $Y(t)$, has been derived previously [12]. However, for the type III model, only an approximate expression for the mean number of recorded events has been derived [29]. In this chapter, we derive the exact mean and variance expressions of $Y(t)$ for both type II and type III models.
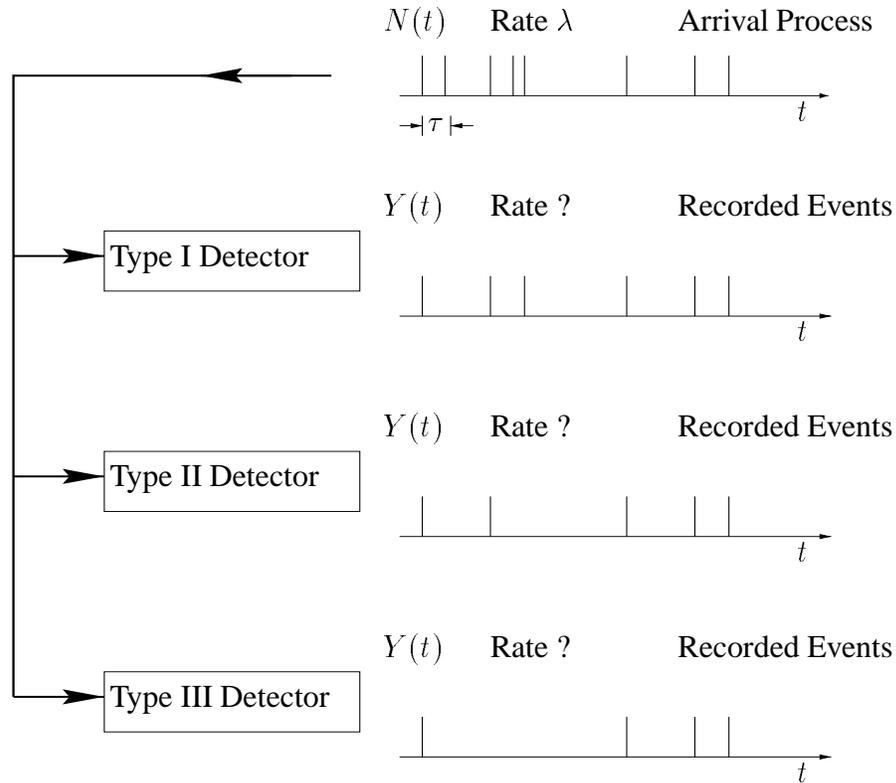
Figure 4.1: Illustration of systems affected by three types of deadtime

This investigation of deadtime statistics was originally motivated by the goal of finding appropriate statistical models for image reconstruction of PET and SPECT scans with high deadtime losses. There are four natural choices for dealing with deadtime in image reconstruction:

(i) ignore it altogether;

(ii) correct the number of recorded events for deadtime losses and plug the corrected data into the reconstruction algorithm;

(iii) incorporate deadtime losses into the system matrix of the usual Poisson statistical model;

(iv) develop reconstruction algorithms based on the exact statistics of the counting process.

For a quantitatively accurate reconstruction, we must correct for the effect of deadtime to avoid underestimation of source activity. This consideration rules out the first choice.

Previous work [81, 22, 65, 88, 82] in this field usually involves the second choice, *i.e.*, using the method of moments to correct the sinograms for deadtime losses, and reconstructing the image using these corrected counts. In statistical image reconstruction, it is generally assumed that the number of recorded events at a detector is Poisson distributed. However, in the presence of deadtime, the fact that there can be no recorded events within τ of each other makes the counting process non-Poisson[55]. However, if the process is approximately Poisson, then a simple modification of the system matrix, *i.e.*, correct the elements of the system matrix, a_{ij}, by the deadtime loss factor, should suffice. This is the third choice as listed above, which would yield estimates with lower variance than plugging the *corrected* counts into a statistical reconstruction algorithm with an *uncorrected* system matrix. But simply correcting the number of recorded events or building this as a "loss factor" in the system model while assuming that the number of recorded events is Poisson distributed may be suboptimal. In this work, we investigate not only the mean, but also the variance of the number of recorded events. If the mean and variance disagree significantly, then reconstructions based on Poisson statistical model would have suboptimally large variances. We discuss this further in Section 4.6 after we derive the exact mean and variance for the counting process.

4.2   Statistical Analysis of Deadtime

We define a "photon arrival" to mean a photon interacting with the scintillator with sufficient deposited energy to trigger detection. The photon arrival process $N(t)$ counts the number of arrivals during the time interval $(0, t]$, and the photon recording process $Y(t)$ counts the number of recorded events. For simplicity, we assume that $N(t)$ is a homogeneous Poisson process with constant rate λ (photon arrivals per unit time) *i.e.*, we neglect radio-isotope decay and other physical or physiological effects that may cause variable arrival rate (see Appendix C for a generalization). We first review a few simple and useful facts about the Poisson process [35]. The increment $N(t_2) - N(t_1)$, which is the number of photon arrivals during the time interval $(t_1, t_2]$, is Poisson distributed with mean $(t_2 - t_1)\lambda$. $N(t)$ has stationary and independent increments. If T_n denotes the time of the nth pho-

ton arrival, then the waiting time (or inter-arrival time) $W_n = T_n - T_{n-1}$ is exponentially distributed with mean $1/\lambda$.

For simplicity, we also assume that the deadtime τ is known and deterministic. Most systems can be adequately modeled to have a constant deadtime, independent of count rate.

4.2.1 Asymptotic Analysis via Renewal Theory

The counting processes in all three types of systems discussed above are examples of "renewal processes" [35], and renewal theory has been the classical basis for deadtime analysis[63, 70, 69, 34]. A renewal process involves recurrent patterns connected with repeated trials. Roughly speaking, if after each occurrence of a pattern $\mathcal{E}$, the random process starts from scratch in the sense that the trials following an occurrence of $\mathcal{E}$ form a replica of the whole process, then the process qualifies as a renewal process. If we define $\mathcal{E}$ to be the state[1] of "the counter is ready to record the next photon arrival", then after each occurrence of $\mathcal{E}$, the counting process is statistically equivalent. A very useful random variable to define is $T_{\mathcal{E}}$, the waiting time between one renewal and the next (renewal here means return to $\mathcal{E}$). Note that in the context of photon counting system, with $\mathcal{E}$ defined as above, the number of renewals from 0 to t is almost[2] exactly the number of recorded events from 0 to t. If $T_{\mathcal{E}}$ has ensemble mean $\mu_{\mathcal{E}}$ and variance $\sigma_{\mathcal{E}}^2$, then the number of renewals from 0 to t, $\tilde{Y}(t)$, is asymptotically Gaussian distributed [21][35] with the following moments:

$$E[\tilde{Y}(t)] \sim t/\mu_{\mathcal{E}}, \mathrm{Var}[\tilde{Y}(t)] \sim t\sigma_{\mathcal{E}}^2/\mu_{\mathcal{E}}^3, \qquad (4.1)$$

where $\sim$ indicates that the ratio of the two sides tends to unity as $t/\mu_{\mathcal{E}} \to \infty$. Hence asymptotically, the mean and variance of the waiting time between renewals forms a sort of "duality" relationship with the mean and variance of the number of renewals. We observe that when $\tau = 0$, *i.e.*, no deadtime, $T_{\mathcal{E}}$ is exponentially distributed with mean $1/\lambda$ and variance $1/\lambda^2$; thus $E[\tilde{Y}(t)] \sim \lambda t$ and $\mathrm{Var}[\tilde{Y}(t)] \sim \lambda t$, as expected since $\tilde{Y}(t)$ would

---

[1]For type III deadtime, we define renewal as "return to $\mathcal{E}$ after recording an event".

[2]Almost since we have to consider photons arriving shortly before time 0 (or t) but renewal occurring shortly after time 0 (or t). If one redefines the time of a recorded event to be τ after the the photon arrives at the detector, then the number of recorded events and the number of renewals during $(0, t]$ would be exactly the same. For stationary increment processes, which definition one adopts makes absolutely no difference in terms of the statistics of the process.

be Poisson distributed with mean λt when there is no deadtime. In realistic cases where deadtime loss becomes significant, $\mu_{\mathcal{E}}$ is usually very small when compared to t, hence the Gaussian approximation is often very good.

For the nonparalyzable deadtime model (type I model), it is easy to derive the asymptotic mean and variance of $\tilde{Y}(t)$ from the moments of $T_{\mathcal{E}}$. After each recording of an event, the "deadtime" when the system cannot record any incoming arrival is simply τ. Thus $T_{\mathcal{E}} = T + \tau$, where T is an exponentially distributed random variable with mean $1/\lambda$. Hence,

$$\mu_{\mathcal{E}} = 1/\lambda + \tau = \frac{1 + \lambda\tau}{\lambda} \tag{4.2}$$

$$\sigma_{\mathcal{E}} = 1/\lambda. \tag{4.3}$$

Thus from (4.1), the counting process for a nonparalyzable (type I) system is asymptotically Gaussian distributed with:

$$E[\tilde{Y}(t)] \sim \frac{\lambda t}{1 + \lambda\tau}, \operatorname{Var}[\tilde{Y}(t)] \sim \frac{\lambda t}{(1 + \lambda\tau)^3}. \tag{4.4}$$

Figure 4.2 shows the mean and variance of the counting process of systems affected by nonparalyzable deadtime. When $\lambda\tau > 0.1$, the mean and variance of $\tilde{Y}(t)$ differ by at least 20%. Carloni *et al* derived the following exact expression for the ith moment for the nonparalyzable deadtime model [13]:

$$E[Y^i(t)] = \sum_{k=0}^{\lfloor t/\tau \rfloor} [(k+1)^i - k^i]\Gamma(k+1, \lambda(t - k\tau))/k!, \tag{4.5}$$

where $i = 1, 2, \ldots$, and $\Gamma(j, t) = \int_0^t e^{-x}x^{j-1}dx$ is the incomplete gamma function. From this expression, they derived asymptotic expressions for the mean and variance of the number of recorded events that coincide with (4.4).

For the other two deadtime models, if we try to derive $E[Y(t)]$ from $E[T_{\mathcal{E}}]$, it is much more difficult to obtain a simple closed form expression because if we try to derive $E[T_{\mathcal{E}}]$, we get an infinite sum and it is not easy to obtain every term in this sum, let alone a closed-form expression for $E[T_{\mathcal{E}}]$. The variance of $T_{\mathcal{E}}$ is even more complicated. Therefore, in the following section, we describe a new approach for deriving the moments of counting processes.
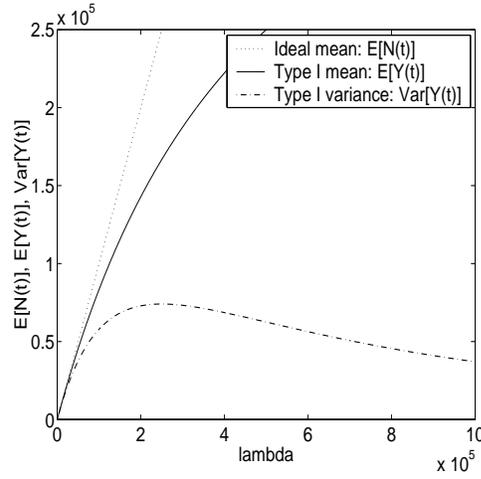
Figure 4.2: Mean and variance for nonparalyzable (type I) systems, with $t = 1\mathrm{s}, \tau = 2\mu\mathrm{s}$.

4.2.2 Exact Mean and Variance of Counting Processes

We first consider a general counting process Y where $Y(t_1, t_2)$ denotes the number of recorded events during the time interval $(t_1, t_2]$ and $Y(t)$ is a shorthand for $Y(0, t)$. We define the instantaneous rate $\gamma : \mathbb{R} \to [0, \infty)$ of the process $Y(t)$ as:

$$\gamma(s) \triangleq \lim_{\delta \to 0} E[Y(s + \delta) - Y(s)]/\delta, \tag{4.6}$$

and the instantaneous second moment $\alpha : \mathbb{R} \to [0, \infty)$ as:

$$\alpha(s) \triangleq \lim_{\delta \to 0} E[(Y(s + \delta) - Y(s))^2]/\delta. \tag{4.7}$$

We also define the correlation function $\beta : \mathbb{R}^2 \to [0, \infty)$ as:

$$\beta(s_1, s_2) \triangleq \lim_{\delta_1, \delta_2 \to 0} E[(Y(s_1 + \delta_1) - Y(s_1))(Y(s_2 + \delta_2) - Y(s_2))]/(\delta_1 \delta_2). \tag{4.8}$$

We assume that the following regularity conditions hold[3]

(i) γ and α are well-defined μ-almost everywhere, and β is well defined μ_2-almost everywhere, and γ and β are integrable with respect to μ and μ_2 over any finite interval and rectangle, respectively;

(ii) $E[Y(s, s + \delta)]/\delta$ and $E[Y^2(s, s + \delta)]/\delta$ are uniformly bounded for all s and $\delta \in (0, 1)$;

[3]μ and μ_2 denote Lebesgue measures on $\mathbb{R}$ and $\mathbb{R}^2$, respectively.

76

(iii) $E[Y(s_1, s_1 + \delta_1)Y(s_2, s_2 + \delta_2)]/(\delta_1 \delta_2)$ is uniformly bounded for all s_1, s_2, and $\delta_1, \delta_2 \in (0, 1)$ such that $(s_1, s_1 + \delta_1) \cap (s_2, s_2 + \delta_2) = \emptyset$.

These assumptions hold for a wide variety of counting processes, including any homogeneous Poisson process with finite intensity. Furthermore, for an arbitrary random process Y, if $E[Y(s, s + \delta)]/\delta$, $E[Y^2(s, s + \delta)]/\delta$, and $E[Y(s_1, s_1 + \delta_1)Y(s_2, s_2 + \delta_2)]/(\delta_1 \delta_2)$ are respectively uniformly bounded above by those of a homogeneous Poisson process, then assumption (ii) and (iii) hold for Y. Specifically, if a random process results from some form of selection from a Poisson process with bounded intensity, then assumptions (ii) and (iii) hold.

For analysis purposes, we artificially divide the time interval $[0, t]$ into n segments of length δ each, *i.e.*, $t = n\delta$. We have

$$Y(t) = \sum_{i=0}^{n-1} Y(i\delta, (i + 1)\delta), \tag{4.9}$$

$$E[Y(t)] = \sum_{i=0}^{n-1} E[Y(i\delta, (i + 1)\delta)], \tag{4.10}$$

$$= \int_{\mathbf{R}} f_\delta(s) ds, \tag{4.11}$$

where we define the following piecewise constant function:

$$f_\delta(s) \triangleq \begin{cases} E[Y(j\delta, (j + 1)\delta)]/\delta, & \text{if } s \in (j\delta, (j + 1)\delta], 0 \le j \le n - 1 \\ 0, & \text{otherwise.} \end{cases} \tag{4.12}$$

Since $\gamma(t)$ is well-defined almost everywhere in the interval $[0, t]$ and $E[Y(s, s + \delta)]/\delta$ is uniformly bounded, by the Lebesgue Dominated Convergence theorem (LDCT)[10],

$$\lim_{\delta \to 0} \int_{\mathbf{R}} f_\delta(s) d\mu(s) = \int_{\mathbf{R}} \lim_{\delta \to 0} f_\delta(s) d\mu(s)$$

$$= \int_0^t \gamma(s) ds. \tag{4.13}$$

Hence, we have the following simple general expression for the mean of the counting process in terms of its instantaneous rate[4]:

$$E[Y(t)] = \int_0^t \gamma(s)ds. \tag{4.14}$$

We consider the second moment by a similar argument:

$$
\begin{aligned}
E[Y^2(t)] &= E[(\sum_{i=0}^{n-1} Y(i\delta, (i+1)\delta))^2] \\
&= \sum_{i=0}^{n-1} E[Y^2(i\delta, (i+1)\delta))] + \sum_{i=0}^{n-1} \sum_{j=0, j\neq i}^{n-1} E[Y(i\delta, (i+1)\delta)Y(j\delta, (j+1)\delta)] \\
&= \sum_{i=0}^{n-1} E[Y^2(i\delta, (i+1)\delta))] \\
&\quad + 2 \sum_{i=0}^{n-2} \sum_{j=i+1}^{n-1} E[Y(i\delta, (i+1)\delta)Y(j\delta, (j+1)\delta)] \\
&= \int_{\mathbf{R}} g_\delta(s)d\mu(s) + 2\int_{\mathbf{R}^2} h_\delta(s_1, s_2)d\mu_2(s_1, s_2), \tag{4.15}
\end{aligned}
$$

where we define the following piecewise constant functions:

$$
g_\delta(s) \triangleq
\begin{cases}
E[Y^2(j\delta, (j+1)\delta)]/\delta, & \text{if } s \in (j\delta, (j+1)\delta] \text{ and } 0 \le j \le n-1 \\
0, & \text{otherwise,}
\end{cases} \tag{4.16}
$$

and

$$
h_\delta(s_1, s_2) \triangleq
\begin{cases}
E[Y(i\delta, (i+1)\delta)Y(j\delta, (j+1)\delta)]/\delta^2, & \text{if } s_1 \in (i\delta, (i+1)\delta], \\
& s_2 \in (j\delta, (j+1)\delta], \\
& 0 \le i \le n-2, \\
& \text{and } i+1 \le j \le n-1 \\
0, & \text{otherwise.}
\end{cases} \tag{4.17}
$$

Since β is well-defined almost everywhere in $[0, t] \times [0, t]$ and $E[Y(s_1, s_1 + \delta)Y(s_2, s_2 + \delta)]/\delta^2$ is uniformly bounded, by LDCT and Fubini's Theorem[10],

$$\lim_{\delta \to 0} \int_{\mathbf{R}^2} h_\delta(s_1, s_2)d\mu_2(s_1, s_2) = \int_{\mathbf{R}^2} \lim_{\delta \to 0} h_\delta(s_1, s_2)d\mu_2(s_1, s_2)$$

[4]If $E[Y(t)]$ is differentiable for all t, then $\gamma(t) = \frac{dE[Y(t)]}{dt}$, and (4.14) results from the fundamental theorem of calculus. However, $E[Y(s)Y(t)]$ is not everywhere differentiable even for very simple random processes, $e.g.$, for the Poisson process N with intensity λ, $E[N(s)N(t)] = \lambda \min(s, t) + \lambda^2 st$. So a similar argument involving the fundamental theorem of calculus runs into difficulties for the second moment.

$$= \int_0^t \int_{s_1}^t \beta(s_1, s_2) ds_2 ds_1$$

$$= \int_0^t \int_{s_2}^t \beta(s_1, s_2) ds_1 ds_2. \tag{4.18}$$

Similarly, one can show that

$$\lim_{\delta \to 0} \int_{\mathbf{R}} g_\delta(s) d\mu(s) = \int_0^t \alpha(s) ds. \tag{4.19}$$

Thus using (4.15), (4.18), and (4.19), we have the following general expression for the second moment of $Y(t)$:

$$E[Y^2(t)] = \int_0^t \alpha(s) ds + 2 \int_0^t \int_{s_1}^t \beta(s_1, s_2) ds_2 ds_1. \tag{4.20}$$

In the context of counting processes with deadtime, which includes all random processes considered in this work, the process satisfies this additional assumption:

(iv) there exists a positive δ_0 such that $\forall \delta \in (0, \delta_0)$, $Y(s, s + \delta) \leq 1$.

If we pick $\delta_0 < \tau$, then assumption (iv) holds. For $\delta < \delta_0$, since $0^2 = 0$ and $1^2 = 1$,

$$E[Y^2(s, s + \delta)] = E[Y(s, s + \delta)], \tag{4.21}$$

so

$$\alpha(s) = \gamma(s). \tag{4.22}$$

Thus we obtain the following corollary of (4.20) for random processes satisfying assumptions (i) to (iv):

$$E[Y^2(t)] = E[Y(t)] + 2 \int_0^t \int_{s_1}^t \beta(s_1, s_2) ds_2 ds_1. \tag{4.23}$$

Furthermore, if $Y(t)$ has stationary increments, then $\gamma(s)$ is constant and $\beta(s_1, s_2) = \beta(0, s_2 - s_1)$ and we can further simplify the results (4.14) and (4.23) to the following:

$$E[Y(t)] = \gamma t \tag{4.24}$$

$$E[Y^2(t)] = \gamma t + 2 \int_0^t (t - s) \beta(0, s) ds. \tag{4.25}$$

The above general approach used to find the second moment of $Y(t)$ could be extended to higher order moments. However, as the order gets higher, the expressions get more complicated.

4.3 Single Photon Counting

4.3.1 Mean and Variance of Recorded Singles Counts, Model Type II

First we consider the paralyzable model in which if the waiting time for a photon arrival is less than τ, then this photon is not recorded. We derive the mean and variance of $Y(t)$, the number of recorded events from time 0 to time t. We observe that $Y(t)$ inherits the stationary increment property of the arrival process $N(t)$. We first derive $E[Y(0, \delta)]$, where we pick $\delta < \tau$ such that the number of recorded events during $(0, \delta]$ is either 0 or 1. Let T_1 denote the time of the first photon arrival after time 0; it is exponentially distributed. If there is an arrival at $T_1 = s$, $0 < s < \delta$, and there is no arrival between $s - \tau$ and s (in fact, we only need to make sure there is no arrival between $s - \tau$ and 0, i.e., $N(0) - N(s - \tau) = 0$, since the first arrival after 0 occurs at s), then there will be a recorded event during the interval $(0, \delta]$. Thus

$$
\begin{aligned}
E[Y(0, \delta)] &= \mathrm{P}[Y(0, \delta) = 1] \\
&= \int_0^\infty \mathrm{P}[Y(0, \delta) = 1 | T_1 = s] f_{T_1}(s) ds \\
&= \int_0^\delta \mathrm{P}[\text{no arrival during}(s - \tau, 0) | T_1 = s] f_{T_1}(s) ds \\
&= \int_0^\delta \mathrm{P}[N(s - \tau, 0) = 0 | T_1 = s] f_{T_1}(s) ds \\
&= \int_0^\delta e^{-\lambda(\tau - s)} \lambda e^{-\lambda s} ds = \int_0^\delta \lambda e^{-\lambda \tau} ds = \lambda \delta e^{-\lambda \tau}.
\end{aligned}
\tag{4.26}
$$

Hence by the definition given in (4.6), the instantaneous rate of $Y(t)$ is

$$
\gamma = \lambda e^{-\lambda \tau},
\tag{4.27}
$$

and by (4.24), we easily obtain the following result (*e.g.*, [78]),

$$
E[Y(t)] = \lambda t e^{-\lambda \tau},
\tag{4.28}
$$

i.e., the recorded/arrival ratio for type II systems, denoted ξ_2, is

$$
\xi_2 \triangleq \frac{E[Y(t)]}{E[N(t)]} = e^{-\lambda \tau}.
\tag{4.29}
$$

The variance of $Y(t)$ for the type II model is (see Appendix A):

$$\mathrm{Var}[Y(t)] = \lambda t e^{-\lambda\tau}(1 - (2\lambda\tau - \lambda\tau^2/t)e^{-\lambda\tau}). \tag{4.30}$$

We can compute numerically that $\max_{\lambda\tau}(2\lambda\tau e^{-\lambda\tau}) \approx 0.74$, hence $\mathrm{Var}[Y(t)]$ will always be positive. To compare the variance and the mean, we note that

$$\lim_{t\to\infty} \frac{\mathrm{Var}[Y(t)]}{E[Y(t)]} = 1 - 2\lambda\tau e^{-\lambda\tau} = 1 - 2\xi_2 \log\xi_2. \tag{4.31}$$

Figure 4.3 shows the mean and variance of the singles count for a detector affected by deadtime of type II. Since the mean and variance can differ greatly, $Y(t)$ is not Poisson.

(a)   (b)
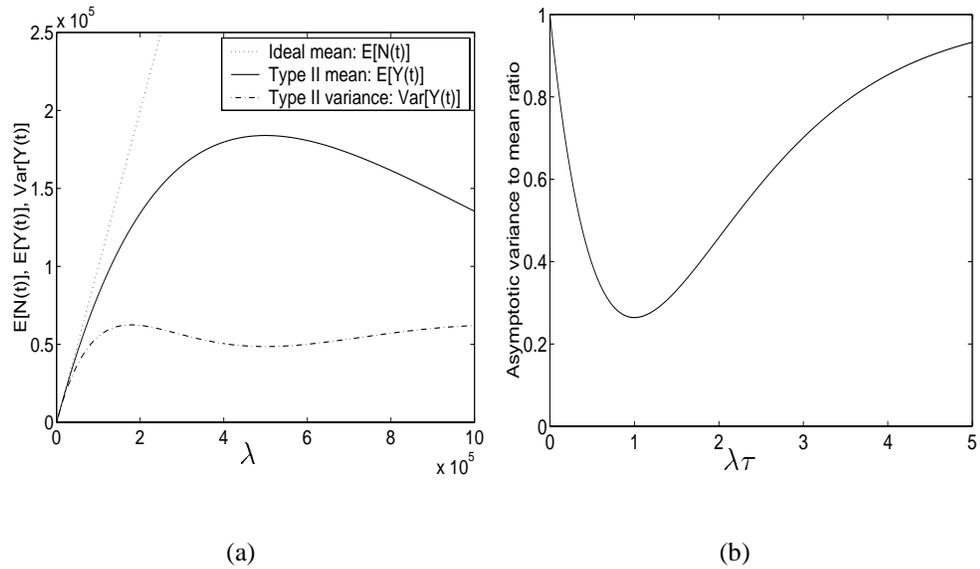
Figure 4.3: Mean and variance for paralyzable (type II) systems, with $t = 1\mathrm{s}, \tau = 2\mu\mathrm{s}$.

4.3.2   Mean and Variance of Recorded Singles Counts, Model Type III

Now we turn to the type of system described in [29], in which if the waiting time for a photon arrival is less than τ, then neither this photon nor the previous photon will be recorded. We again observe that $Y(t)$ inherits the stationary increment property of the arrival process $N(t)$. We first derive $E[Y(0,\delta)]$, where we pick $\delta < \tau$ such that the number of recorded events during $(0,\delta]$ is still either 0 or 1. Hence,

$$E[Y(0,\delta)] = \mathrm{P}[Y(0,\delta) = 1]$$

$$= \int_0^\delta \mathrm{P}[Y(0,\delta)=1|T_1=s]f_{T_1}(s)ds$$

$$= \int_0^\delta \mathrm{P}[N(s-\tau,0)=0]\mathrm{P}[(s,s+\tau)=0]f_{T_1}(s)ds$$

$$= \int_0^\delta e^{-\lambda(\tau-s)}e^{-\lambda\tau}\lambda e^{-\lambda s}ds = \int_0^\delta \lambda e^{-\lambda 2\tau}ds = \lambda\delta e^{-\lambda 2\tau}. \qquad (4.32)$$

Hence for this system, the instantaneous rate as defined in (4.6) is

$$\gamma = \lambda e^{-\lambda 2\tau}, \qquad (4.33)$$

and by (4.24), the expected number of recorded events for a type III system is exactly:

$$E[Y(t)] = \lambda t e^{-\lambda 2\tau}. \qquad (4.34)$$

The type III system was analyzed using approximations in [29]. To compare our exact result (4.34) with the approximate analysis presented in [29], we note that the mean waiting time between recorded events is:

$$\mu_{\mathcal{E}} = t/E[Y(t)] = \frac{1}{\lambda}e^{\lambda 2\tau} \qquad (4.35)$$

$$= \frac{1}{\lambda}(1 + 2\lambda\tau + 2(\lambda\tau)^2 + \frac{4}{3}(\lambda\tau)^3 + \frac{2}{3}(\lambda\tau)^4 + O(\lambda\tau)^5). \qquad (4.36)$$

Comparing this exact expansion to the approximate mean waiting time derived in [29, eqn. 16], we find that the approximation in [29] is accurate to 2nd order.

The variance of $Y(t)$ for the type III model is (see Appendix B):

$$\mathrm{Var}[Y(t)] = \lambda t e^{-\lambda 2\tau} + 2e^{-3\lambda\tau}(\lambda t - \lambda\tau - 1)$$

$$+ e^{-4\lambda\tau}(4\lambda^2\tau^2 - 4\lambda^2 t\tau + 2 - 2\lambda t + 4\lambda\tau). \qquad (4.37)$$

To compare the variance and the mean, we observe that

$$\lim_{t\to\infty}\frac{\mathrm{Var}[Y(t)]}{E[Y(t)]} = 1 - 2(1 + 2\lambda\tau - e^{\lambda\tau})e^{-2\lambda\tau}. \qquad (4.38)$$

To simplify this expression, we observe that when $\lambda\tau \ll 1$, $e^{\lambda\tau} - 1 \approx \lambda\tau$, and

$$\lim_{t\to\infty}\frac{\mathrm{Var}[Y(t)]}{E[Y(t)]} \approx 1 - 2\lambda\tau e^{-2\lambda\tau} = 1 - \xi_3\log\xi_3, \qquad (4.39)$$

where $\xi_3 \triangleq E[Y(t)]/E[N(t)] = e^{-2\lambda\tau}$. Figure 4.4 shows the (exact) mean and variance of the singles count $Y(t)$ for type III systems. Again $Y(t)$ is not Poisson, but the difference between the variance and the mean is much smaller than type I or type II systems.

$$\text{(a)} \qquad\qquad\qquad\qquad \text{(b)}$$
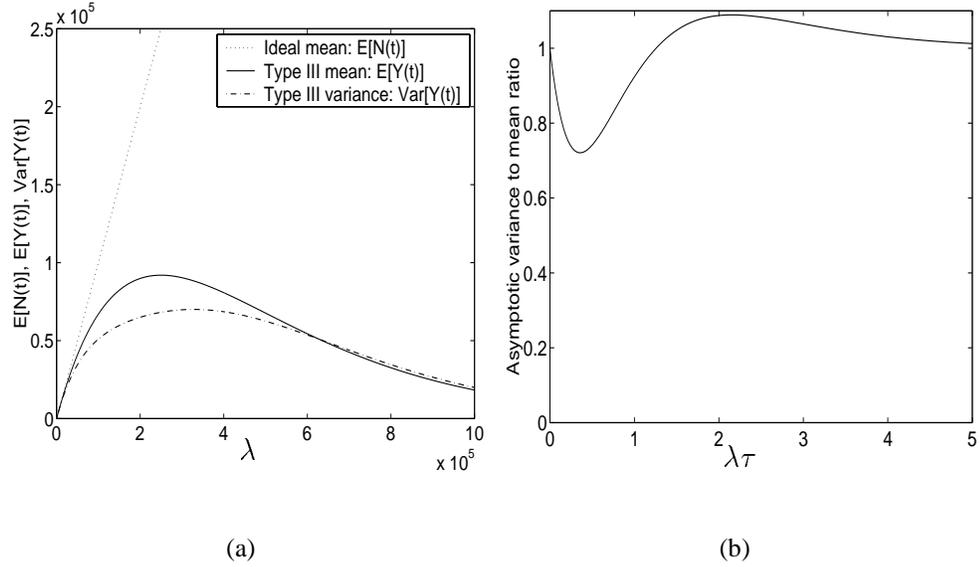
Figure 4.4: Mean and variance for type III systems, with $t = 1\text{s}, \tau = 2\mu\text{s}$.

4.4 Recorded Singles Counts by Block Detectors

In many photon counting systems, several detectors are grouped into a "block"; examples include block PET detectors and Anger cameras. When a photon arrives at any detector in the block, the whole block goes dead for τ, *i.e.*, no detector in the block can record any photon for τ. For analysis purposes, we can initially treat the block of detectors as a single big detector. Let $\lambda_1, \ldots, \lambda_l$ denote the incident photon arrival rates for each of the l detectors in the block. Let $Y_j(t)$ denote the number of events recorded by the jth detector, and let $Z(t)$ denote the total number of events recorded by all detectors in the block ($Z = \sum_{j=1}^{l} Y_j$). We have derived above the exact first and second moments of $Z(t)$ for detector blocks affected by type II and type III deadtime, and in each case, the mean and the variance of $Z(t)$ can differ greatly. However, what is of greater interest in image reconstruction is the mean and variance of the number of events recorded by each detector in the block. Given that $Z(t)$ events are recorded by the entire block, the conditional distribution of the number of events recorded by any individual detector is multinomial where the fraction of events allotted to the jth detector is $\eta_j \triangleq \lambda_j/\lambda$. Thus from [3, p. 99],

$$E[Y_j(t)] \quad = \quad \eta_j E[Z(t)] \tag{4.40}$$

83

$$\text{Var}[Y_j(t)] \quad = \quad \eta_j(1 - \eta_j)E[Z(t)] + \eta_j^2 \text{Var}[Z(t)]. \tag{4.41}$$

We observe that the variance to mean ratio is

$$\frac{\text{Var}[Y_j(t)]}{E[Y_j(t)]} \quad = \quad 1 - \eta_j(1 - \text{Var}[Z(t)]/E[Z(t)]) \tag{4.42}$$

$$\geq \quad 1 - \eta_j. \tag{4.43}$$

For a system with say, 64 detectors in a block, $\eta_j \approx 1/64$ (assuming that the count rates λ_j's are nearly uniform), so from (4.43) the mean and the variance of the number of recorded events by a single detector will differ by less than 2%, regardless of count rates and dead-time losses. Furthermore, since $E[Z(t)]$ must be quite large for deadtime to have a significant effect, when η_j is small, the distribution of $Y_j(t)$ will be approximately Poisson by the usual binomial argument. The only case where the variance to mean ratio is significantly less than 1 would be when η_j is large (*i.e.* the count rates λ_j's are very heterogeneous) and $\text{Var}[Z(t)]/E[Z(t)]$ is small (*i.e.*, the total count rate $\sum_{j=1}^{l} \lambda_j$ is large). In all other cases, the mean and the variance would be approximately equal. However, the *covariance* between the measurements recorded by different elements within the block can be nonzero[3, p. 101]:

$$\text{Cov}(Y_i(t), Y_j(t)) = \eta_i \eta_j (\text{Var}[Z(t)] - E[Z(t)]). \tag{4.44}$$

Thus in the presence of deadtime, the assumption that the measurements are independent (which is made ubiquitously in statistical reconstruction methods) is incorrect. However, when η_i and η_j are small, so is the covariance between individual detector elements, so the impact of this dependence may be small.

4.5   Count Rate Correction for System Type III

For a quantitatively accurate reconstruction, we must correct for the effect of dead-time to avoid underestimation of source activity. For type III systems, Engeland *et al* [29] proposed the following correction formula,

$$\hat{\lambda} = \frac{Y}{t}\left(1 + \frac{2Y}{t}\tau + \frac{6Y}{t^2}\tau^2\right), \tag{4.45}$$

which they obtained by solving an approximate mean waiting time expression up to second order in τ by means of the expansion $\lambda = a + b\tau + c\tau^2$. We propose to estimate the true count rate by solving numerically our exact expression (4.34), *i.e.*, solve

$$\frac{Y}{t} = \hat{\lambda}e^{-2\hat{\lambda}\tau} \qquad (4.46)$$

for $\hat{\lambda}$ given Y and t. One could solve analytically the exact mean waiting time expression (4.35) up to second order in τ, which yields exactly the same estimator as (4.45), but this estimator does not solve (4.35) exactly. Figure 4.5 compares our new estimator (4.46) and the estimator proposed in [29]. It shows that our new estimator is unbiased even at very high count rates. The error bars are not shown in the figure as they are smaller than the plotting symbols. When t is large, the standard deviation is very small when compared to the mean of $Y(t)$, thus these estimates have extremely small standard deviations. By solving (4.46) numerically, we obtain essentially perfect deadtime correction for a type III system.
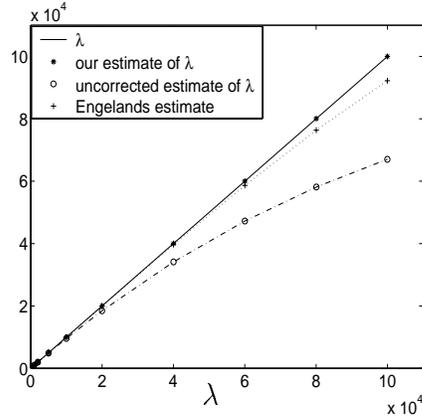
Figure 4.5: 20 realizations, with $t = 10\text{s}, \tau = 2\mu\text{s}$.

4.6   Discussion

We have analyzed the mean and variance of the recorded singles counts for three distinct models of deadtime. In all three cases, the variance can be significantly less than the mean, indicating that the counting statistics are not Poisson in the presence of deadtime. Deadtime losses can be significant in practical SPECT and PET systems, particularly in fully 3D

PET imaging and in SPECT transmission measurements with a scanning line source. The count rates for a detector block (PET) or detector zone (SPECT) can be significant enough to yield non-Poisson statistics for the total counts recorded by the block or zone. However, in the practical situations that we are aware of, the count rates for individual detector elements within the block or zone are usually not high enough to correspond to significant differences between the mean and the variance. As we have shown in Section 4.4, even though the variance of the counts recorded by a block can be significantly lower than the mean, the variance of the counts recorded by an individual detector within a block is nevertheless quite close to the mean and likely to be well approximated by a Poisson distribution. Furthermore, the correlation between individual detectors will be fairly small. Thus it appears that statistical image reconstruction based on Poisson models, while certainly not optimal, should be adequate in practice even under fairly large deadtime losses, provided the deadtime loss factor is included in the system matrix. We must add one caveat to this conclusion however. Although pairs of individual detectors have small correlation, the correlation coefficient between the *sum* of one group of detectors and the *sum* of all other detectors in a block may not be small in the presence of deadtime. The effect of such correlations on image reconstruction algorithms is unknown and may deserve further investigation. Another natural extension of this work would be to consider systems with random resolving times τ. As long as the minimum resolving time is greater than zero, assumption (iv) would still hold and the derivations would be similar.

CHAPTER 5

Coincidence Counting Statistics Affected by Deadtime

The statistics of the coincidence counting process by detectors affected by deadtime is of fundamental importance to the problem of statistical image reconstruction. However, to our knowledge, not even the mean of the random coincidence process has been derived *rigorously*. In this chapter, we give a rigorous derivation of the mean and variance of the coincidence counting process under various scenarios, and analyze the suitability of the commonly assumed Poisson statistical model used in image reconstruction.

5.1   General Result

We assume the following model with no attenuation of the true coincidences (see Figure 5.1):

$$
\begin{aligned}
X_1(t) &= X(t) + N_1(t), \\
X_2(t) &= X(t) + N_2(t);
\end{aligned}
\tag{5.1}
$$

there are two detectors each recording single photons according to deadtime model type II, with deadtime τ known and deterministic; the arrival processes at the two detectors are denoted $X_1(t)$ and $X_2(t)$; $X(t)$ denotes photons originating from the true coincidence source, $N_1(t)$ and $N_2(t)$ denote photons originating from the random singles sources arriving at the two detectors; $X(t)$, $N_1(t)$, and $N_2(t)$ are independent and all are homogeneous Poisson processes, *i.e.*, we neglect radio-isotope decay for simplicity; λ_1, λ_2, and λ_X denote the
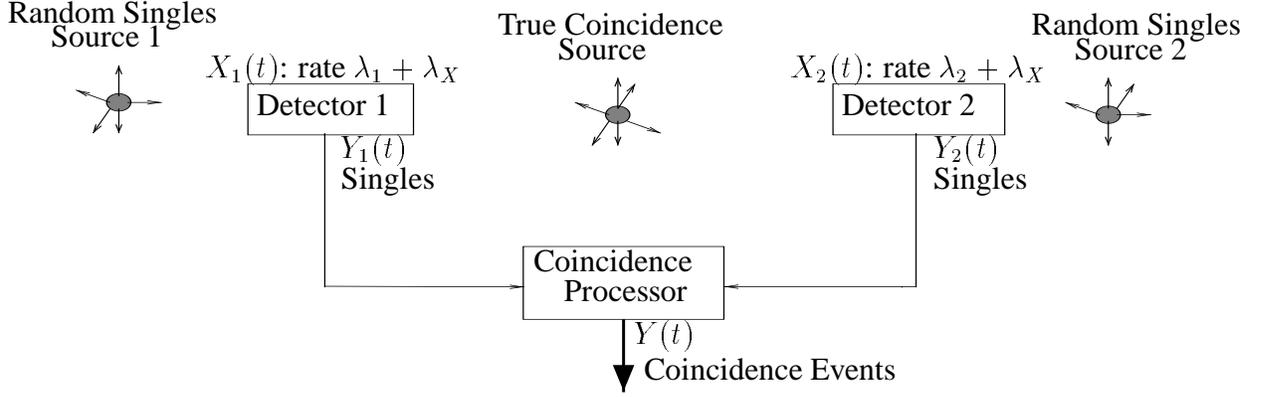
Figure 5.1: Model for coincidence counts

intensity of N_1, N_2, and X, respectively; X_1 and X_2 are statistically dependent due to X, unless $\lambda_X = 0$ (*i.e.*, the random coincidence case); $Y(t)$ denote the number of recorded coincidence events. For simplicity, we assume that pairs of X photons are recorded by their respective detectors at exactly the same time; we also ignore uncertainty in the time-stamping of recorded singles [19]. Let $Y_1(t)$ and $Y_2(t)$ denote the number of recorded singles at detectors 1 and 2, respectively. Let ϵ denote the length of the coincidence timing window; we assume $2\epsilon < \tau$. For a pair of photons to be recorded as a coincidence event, both photons must first be recorded by their respective detectors, and if one photon is recorded by detector 1 at time t_1, and one photon is recorded by detector 2 at time t_2, and $|t_1 - t_2| < \epsilon$, then this pair of photons is recorded as a coincidence event. To avoid ambiguity, we define the time of coincidence to be the arrival time of the later photon. Thus if there is one photon recorded by detector 1 at time t_1 and no photon is recorded by detector 2 at time t_1, then the number of coincidences at time t_1 is the number of photons recorded by detector 2 during $(t_1 - \epsilon, t_1]$. If there is one photon recorded by both detectors both at time t_1, then the number of coincidences at time t_1 is the sum of the number of photons recorded by detector 1 and detector 2 during $(t_1 - \epsilon, t_1]$.

In addition to the four assumptions made in Section 4.2.2, we also make the following additional assumptions about the singles process Y_i:

(v) $\lim_{\delta \to 0} \frac{Pr[Y_i(s, s+\delta) = k]}{\delta} = 0, \forall s, \forall k \geq 2$;

(vi) $\exists \delta_0 > 0, \sum_{k=2}^{\infty} k^i \bar{p}(k, \delta_0) < \infty$ for $i = 1, 2,$ or $3,$

where $\bar{p}(k, \delta_0) \triangleq \sup\{P[Y_i(s, s + \delta) = k]/\delta : s \in [0, t), \delta \in (0, \delta_0)\}$. These assumptions hold for a wide variety of (singles) counting processes, including all (singles) processes considered hereafter. Specifically with regard to assumption (vi), if Y_i is a homogeneous Poisson process with intensity λ, then $P[Y_i(s, s + \delta) = k]/\delta = e^{-\lambda\delta}\lambda^k\delta^{k-1}/k! < \lambda^k/k!$ for $\delta < 1$, and $\sum_{k=2}^{\infty} k^i\lambda^k/k! < \infty$ for any integer i.

5.2 Random Coincidences

In this section, we derive the mean and variance of the random coincidence process (assuming $\lambda_X = 0$), first for the case of ideal detectors (no deadtime), and then for realistic detectors (with deadtime). In the next section, we derive the mean and variance (bounds) of the counting process having both true and random coincidences. Let $Y(t)$ denote the number of recorded coincidence events during $(0, t]$. Let $Y_a(t)$ and $Y_b(t)$ denote the number of recorded coincidence events during $(0, t]$ that have the later singles event arriving at detector 1 and detector 2, respectively. Then

$$Y(t) = Y_a(t) + Y_b(t). \tag{5.2}$$

5.2.1 Mean of Random Coincidence Counts

For the purely random coincidence process, We derive the mean of $Y(t)$ for a general class of (recorded singles) processes, *i.e.*, we do not assume any particular model such as Poisson. For this derivation, we only assume that the singles processes $Y_1(t)$ and $Y_2(t)$ are independent, have stationary increments with rates $\gamma_i \triangleq \frac{E[Y_i(s, s+\delta)]}{\delta}, i = 1, 2$, and satisfy assumptions (i) - (iii), (v), and (vi). For an arbitrary time interval $(0, \delta]$,

$$\begin{aligned} E[Y_a(0, \delta)] &= \sum_{k=0}^{\infty} E[Y_a(0, \delta)|Y_1(0, \delta) = k]P[Y_1(0, \delta) = k] \\ &= 0 + E[Y_a(0, \delta)|Y_1(0, \delta) = 1]P[Y_1(0, \delta) = 1] \\ &\quad + \sum_{k=2}^{\infty} E[Y_a(0, \delta)|Y_1(0, \delta) = k]P[Y_1(0, \delta) = k]. \end{aligned}$$

Given that there is *one* arrival at detector 1 during $(0, \delta]$, the number of recorded coincidence events with the later singles event recorded by detector 1 is the number of singles

89

event recorded by detector 2 during a time interval of length ϵ. Furthermore, if there are k arrivals at detector 1 during $(0, \delta]$, then the number of recorded coincidence events with the later singles event recorded by detector 1 can be no more than $kX_2(-\epsilon, \delta)$. Hence

$$\lim_{\delta \to 0} \frac{1}{\delta} \sum_{k=2}^{\infty} E[Y_a(0, \delta)|Y_1(0, \delta) = k]P[Y_1(0, \delta) = k]$$

$$\leq \lim_{\delta \to 0} \frac{1}{\delta} \sum_{k=2}^{\infty} \gamma_2(\epsilon + \delta)kP[Y_1(0, \delta) = k]. \tag{5.3}$$

Using assumption (vi) and applying the Lebesgue Dominated Convergence Theorem (LDCT) [10], we obtain:

$$\lim_{\delta \to 0} \frac{1}{\delta} \sum_{k=2}^{\infty} E[Y_a(0, \delta)|Y_1(0, \delta) = k]P[Y_1(0, \delta) = k] = \sum_{k=2}^{\infty} \lim_{\delta \to 0} \frac{\gamma_2(\epsilon + \delta)k}{\delta}P[Y_1(0, \delta) = k]$$

$$= 0,$$

by assumption (v). Hence

$$\lim_{\delta \to 0} \frac{1}{\delta} E[Y_a(0, \delta)] = \lim_{\delta \to 0} \frac{1}{\delta} \gamma_2 \epsilon P[Y_1(0, \delta) = 1] + 0. \tag{5.4}$$

Furthermore, since

$$\begin{aligned} \gamma_1 &= \lim_{\delta \to 0} \frac{E[Y_1(0, \delta)]}{\delta} \\ &= \lim_{\delta \to 0} \frac{P[Y_1(0, \delta) = 1] + \sum_{k=2}^{\infty} kP[Y_1(0, \delta) = k]}{\delta} \\ &= \lim_{\delta \to 0} \frac{P[Y_1(0, \delta) = 1]}{\delta}, \end{aligned} \tag{5.5}$$

we have

$$\lim_{\delta \to 0} \frac{E[Y_a(0, \delta)]}{\delta} = \epsilon \gamma_1 \gamma_2. \tag{5.6}$$

By symmetry,

$$\lim_{\delta \to 0} \frac{E[Y_b(0, \delta)]}{\delta} = \epsilon \gamma_2 \gamma_1. \tag{5.7}$$

From (5.2), we conclude that the random coincidence rate is:

$$\gamma = \lim_{\delta \to 0} \frac{E[Y(0, \delta)]}{\delta} = 2\epsilon \gamma_1 \gamma_2. \tag{5.8}$$

It is easy to verify using (5.3) that $E[Y(0, \delta)]/\delta$ is uniformly bounded for all $\delta \in (0, 1)$, hence applying (4.14),

$$E[Y(t)] = 2\epsilon\gamma_1\gamma_2 t. \tag{5.9}$$

This result is well known (when the recorded singles process is Poisson, *i.e.*, recorded with no deadtime) [50, 75] but has not been derived formally to our knowledge. Remarkably, the result holds for a fairly broad class of singles processes, as we have shown.

5.2.2 Ideal Detectors (No Deadtime)

We now show that the (random) coincidence process is not exactly Poisson[1] even when the recorded singles processes are Poisson (i.e., in the hypothetical case of no deadtime losses) by showing $\mathrm{Var}[Y(t)] \neq E[Y(t)]$ using (4.20). We first find $\beta(0, s)$. For $s > \epsilon$, $Y(0, \delta)$ and $Y(s, s + \delta)$ are independent for $\delta < \min(s, s - \epsilon)$, hence

$$\beta(0, s) = (2\epsilon\gamma_1\gamma_2)^2. \tag{5.10}$$

For $0 < s < \epsilon$, we show in Appendix A that

$$\beta(0, s) = (2\epsilon\gamma_1\gamma_2)^2 + (\gamma_1 + \gamma_2)\gamma_1\gamma_2(2\epsilon - s). \tag{5.11}$$

It is easy to verify that Y satisfies assumption (iii) by using ideas similar to those used for (D.8) and (D.12). However, Y does not satisfy assumption (iv) in the absence of deadtime, hence we must use (4.23) to derive its second moment. We have

$$
\begin{aligned}
E[Y^2(\delta)] &= E[(Y_a(\delta) + Y_b(\delta))^2] \\
&= E[Y_a^2(\delta)] + E[Y_b^2(\delta)] + 2E[Y_a(\delta)Y_b(\delta)],
\end{aligned}
\tag{5.12}
$$

and

$$
\begin{aligned}
E[Y_a^2(\delta)] &= \sum_{k=0}^{\infty} E[Y_a^2(0, \delta)|X_1(0, \delta) = k]\mathrm{P}[X_1(0, \delta) = k] \\
&= 0 + E[Y_a^2(0, \delta)|X_1(0, \delta) = 1]\mathrm{P}[X_1(0, \delta) = 1] \\
&\quad + \sum_{k=2}^{\infty} E[Y_a^2(0, \delta)|X_1(0, \delta) = k]\mathrm{P}[X_1(0, \delta) = k].
\end{aligned}
$$

---

[1]In the absence of deadtime, the coincidence process will be Poisson if the two coincidence photons always arrive at the two detectors at exactly the same time and there is no uncertainty in the time-stamping of recorded photons. It appears that coincidence processes are exactly Poisson only in this *highly* idealized case.

If $X_1(0, \delta) = k$, then $Y_a(0, \delta) \leq k X_2(-\epsilon, \delta)$, so

$$\lim_{\delta \to 0} \frac{1}{\delta} \sum_{k=2}^{\infty} E[Y_a^2(0, \delta) | X_1(0, \delta) = k] \mathrm{P}[X_1(0, \delta) = k]$$

$$\leq \lim_{\delta \to 0} \frac{1}{\delta} \sum_{k=2}^{\infty} (\gamma_2(\epsilon + \delta) + (\gamma_2(\epsilon + \delta))^2) k^2 \mathrm{P}[X_1(0, \delta) = k]$$

$$= 0. \tag{5.13}$$

Thus by similar argument made in (D.10):

$$\lim_{\delta \to 0} \frac{1}{\delta} E[Y_a^2(\delta)] = (\epsilon \gamma_2 + (\epsilon \gamma_2)^2) \gamma_1. \tag{5.14}$$

Furthermore,

$$\lim_{\delta \to 0} \frac{1}{\delta} E[Y_a(\delta) Y_b(\delta)]$$

$$= \lim_{\delta \to 0} \frac{1}{\delta} \sum_{k,l=1}^{\infty} E[Y_a(\delta) Y_b(\delta) | X_1(0, \delta) = k, X_2(0, \delta) = l] \cdot$$

$$\mathrm{P}[X_1(0, \delta) = k, X_2(0, \delta) = l]$$

$$\leq \lim_{\delta \to 0} \frac{1}{\delta} \sum_{k,l=1}^{\infty} (k + \gamma_1(\delta + \epsilon))(l + \gamma_2(\delta + \epsilon)) k l \mathrm{P}[X_1(0, \delta) = k, X_2(0, \delta) = l]$$

$$= \lim_{\delta \to 0} \frac{1}{\delta} \left(\sum_{k=1}^{\infty} (k + \gamma_1(\delta + \epsilon)) k \mathrm{P}[X_1(0, \delta) = k] \right) \left(\sum_{l=1}^{\infty} (l + \gamma_1(\delta + \epsilon)) l \mathrm{P}[X_2(0, \delta) = l] \right)$$

$$= 0. \tag{5.15}$$

Hence applying (5.14) and (5.15) to (5.12) and using symmetry:

$$\alpha(s) = \epsilon \gamma_1 \gamma_2 (2 + \epsilon(\gamma_1 + \gamma_2)). \tag{5.16}$$

Using ideas leading to (5.13) and (5.15), one can easily verify that $E[Y^2(0, \delta)]/\delta$ is uniformly bounded. Hence from (5.11), (5.10), (5.16), and (4.20), for $t > \epsilon$, the variance of $Y(t)$ is:

$$\mathrm{Var}[Y(t)]$$

$$= \int_0^t \alpha(s) ds + 2 \int_0^t (t - s)(2 \epsilon \gamma_1 \gamma_2)^2 ds + 2 \int_0^\epsilon (t - s)(\gamma_1 + \gamma_2) \gamma_1 \gamma_2 (2\epsilon - s) ds - (\gamma t)^2$$

$$= \epsilon^2 \gamma_1 \gamma_2 (\gamma_1 + \gamma_2) t + 2 \epsilon \gamma_1 \gamma_2 t + (2 \epsilon \gamma_1 \gamma_2 t)^2 + \epsilon^2 (\gamma_1 + \gamma_2) \gamma_1 \gamma_2 (3t - \frac{4}{3}\epsilon) - (2 \epsilon \gamma_1 \gamma_2 t)^2$$

$$= 2 \epsilon \gamma_1 \gamma_2 t (1 + 2\epsilon(\gamma_1 + \gamma_2)(1 - \epsilon/3t)). \tag{5.17}$$

The variance (5.17) is "inflated" relative to the mean (5.9) by the factor $2\epsilon(\gamma_1 + \gamma_2)(1 - \epsilon/3t)$; hence the random coincidence process is not Poisson even in the absence of deadtime.

5.2.3 Non-ideal Detectors (With Deadtime)

We now derive the variance of $Y(t)$ in the presence of deadtime with $\tau > 2\epsilon$. We proceed as in the case of recorded singles process being Poisson. We first derive $\beta(0, s)$. Since we assume $\tau > 2\epsilon$, forming two coincidence events would require two recorded pairs of photons, each pair forming a coincidence event. The minimum time separation for the two later-recorded photons is at least τ, hence[2] $\beta(0, s) = 0$ for $0 < s < \tau$. If $\tau + \epsilon < s < t$, then $Y(0, \delta)$ and $Y(s, s + \delta)$ are independent for $\delta < s - \tau - \epsilon$, hence $\beta(0, s) = \gamma^2$. The most complicated case is when $\tau < s < \tau + \epsilon$; we show in Appendix B that

$$\beta(0, s) = \lambda_1^2 \lambda_2^2 e^{-(\lambda_1 + \lambda_2)\tau}(\epsilon^2 + (s - \tau)(4\epsilon - s + \tau)) \tag{5.18}$$

under this case. We observe that $0 \leq (s - \tau)(4\epsilon - s + \tau) \leq 3\epsilon^2$ when $0 < s - \tau < \epsilon$, hence $0 < \beta(0, s) < \gamma^2$ for $\tau < s < \tau + \epsilon$. Thus for $t > \tau + \epsilon$:

$$\text{Var}[Y(t)]$$
$$= \gamma t + 2 \int_{\tau+\epsilon}^{t} (t - s)(2\epsilon\gamma_1\gamma_2)^2 ds + 2 \int_{\tau}^{\tau+\epsilon} (t - s)(\gamma_1\gamma_2)^2(\epsilon^2 + (s - \tau)(4\epsilon - s + \tau))ds - (\gamma t)^2$$
$$= 2\epsilon\gamma_1\gamma_2 t + (\gamma_1\gamma_2)^2 \frac{1}{6}\epsilon^2(5\epsilon^2 - 16\epsilon(t - \tau) + 24(t - \tau)^2) - (2\epsilon\gamma_1\gamma_2 t)^2$$
$$= \gamma t(1 - \gamma\tau(2 - \tau/t)) + \frac{\gamma^2}{24}(5\epsilon^2 - 16\epsilon(t - \tau)), \tag{5.19}$$

where $\gamma = 2\epsilon\gamma_1\gamma_2 = 2\epsilon\lambda_1\lambda_2 e^{-(\lambda_1 + \lambda_2)\tau}$ from (5.8) and the mean of $Y(t)$ is γt. When $\epsilon \ll \tau \ll t$, we can approximate (5.19) by $\gamma t(1 - 2\gamma\tau)$, hence the variance is "deflated" relative to the mean by approximately $2\gamma\tau$ which is extremely small under normal circumstances.

5.3 Prompt Coincidences

Now we generalize to the case in which there are both random coincidences and true coincidences, *i.e.*, $\lambda_X \neq 0$. Since we assume $\epsilon < \tau/2$, the event $[Y(\delta) \geq 2]$ can never

---

[2]Note that if $\epsilon < \tau < 2\epsilon$, then $\beta(0, s) \neq 0$ if $s < \epsilon$.
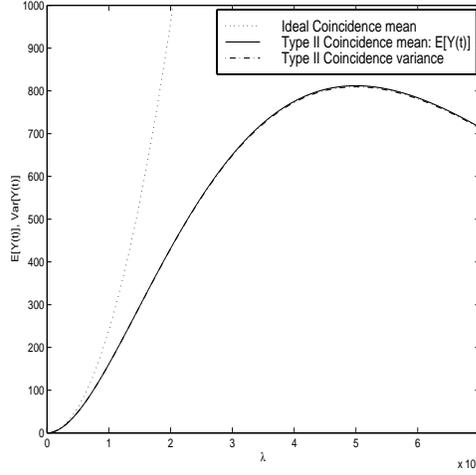
Figure 5.2: Mean and variance for paralyzable systems, random coincidence counts, with $t = 1\mathrm{s}, \tau = 2\mu s, \epsilon = 12ns, \lambda_1 = \lambda_2$.

occur when $\delta < \epsilon$.

5.3.1 Mean

We derive the mean of the prompt coincidences using (4.24). In the presence of dead-time, the event that there is one recorded coincidence event during $(0, \delta]$ consists one of the following 4 mutually exclusive events:

(E_1) one N_1 photon and one N_2 photon form a coincidence event;

(E_2) one X photon at detector 1, and one N_2 photon at detector 2;

(E_3) one X photon at detector 2, and one N_1 photon at detector 1;

(E_4) A pair of X photons is recorded.

We will derive the probability for each of these 4 events.

We split E_1 into two disjoint sub-events: E_{1a} and E_{1b}, where E_{1a} denotes the event that one N_1 photon and one N_2 photon form a coincidence event and furthermore the later photon is recorded by detector 1. Let T_1 denote the time of the first N_1 photon arrival after time 0, and $T_2|T_1$ denote the time of the first N_2 photon arrival after $T_1 - \epsilon$:

$$
\begin{aligned}
\mathrm{P}[E_{1a}] &= \int_0^\delta \int_{s_1 - \epsilon}^{s_1} \mathrm{P}[N_1(s_1 - \tau, 0) = 0, N_2(s_2 - \tau, s_1 - \epsilon) = 0, \\
&\quad X(s_2 - \tau, s_1) = 0 | T_1 = s_1, T_2 = s_2] f_{T_2|T_1}(s_2|s_1) f_{T_1}(s_1) ds_2 ds_1
\end{aligned}
$$

94

$$= \int_0^\delta \int_{s_1-\epsilon}^{s_1} e^{-\lambda_1(\tau-s_1)} e^{-\lambda_2(s_1-\epsilon+\tau-s_2)} e^{-\lambda_X(s_1-s_2+\tau)} \lambda_2 e^{-\lambda_2(s_2-s_1+\epsilon)} \lambda_1 e^{-\lambda_1 s_1} ds_2 ds_1$$

$$= \epsilon \lambda_1 \lambda_2 \delta e^{-(\lambda_1+\lambda_2+\lambda_X)\tau} \frac{1-e^{-\lambda_X \epsilon}}{\lambda_X \epsilon}. \tag{5.20}$$

By symmetry, $\mathrm{P}[E_{1b}] = \mathrm{P}[E_{1a}]$, so

$$\mathrm{P}[E_1] = 2\epsilon \lambda_1 \lambda_2 \delta e^{-(\lambda_1+\lambda_2+\lambda_X)\tau} \frac{1-e^{-\lambda_X \epsilon}}{\lambda_X \epsilon}. \tag{5.21}$$

For E_2, the X photon arriving at detector 2 must be lost due to deadtime caused by an N_2 photon arriving at detector 2 within ϵ before. It cannot be due to an X photon arrival at detector 2 because otherwise, the X photon at detector 1 would also be lost. If an N_2 photon arriving at detector 2 and an X photon arriving at detector 1 form a coincidence event, the N_2 photon arriving at detector 2 must arrive before the X photon because otherwise the N_2 will be lost due to the X photon at detector 2; hence the X photon arriving at detector 1 will be the later arriving photon. Let T_X denote the time of the first X photon arrival after time 0, and $T_2|T_X$ denote the time of the first N_2 photon arrival after $T_X - \epsilon$:

$$\begin{aligned}
\mathrm{P}[E_2] &= \int_0^\delta \int_{s_1-\epsilon}^{s_1} \mathrm{P}[X(s_2-\tau,0)=0, N_1(s_1-\tau,s_1)=0, \\
&\qquad N_2(s_2-\tau,s_1-\epsilon)=0|T_X=s_1,T_2=s_2]f_{T_2|T_X}(s_2|s_1)f_{T_X}(s_1)ds_2 ds_1 \\
&= \int_0^\delta \int_{s_1-\epsilon}^{s_1} e^{-\lambda_X(\tau-s_2)} e^{-\lambda_1 \tau} e^{-\lambda_2(s_1-\epsilon-s_2+\tau)} \lambda_2 e^{-\lambda_2(s_2-s_1+\epsilon)} \lambda_X e^{-\lambda_X s_1} ds_2 ds_1 \\
&= \delta(1-e^{-\lambda_X \epsilon})\lambda_2 e^{-(\lambda_1+\lambda_2+\lambda_X)\tau}. \tag{5.22}
\end{aligned}$$

By symmetry,

$$\mathrm{P}[E_3] = \delta(1-e^{-\lambda_X \epsilon})\lambda_1 e^{-(\lambda_1+\lambda_2+\lambda_X)\tau}. \tag{5.23}$$

For E_4, the simplest of all four cases, we only need to make sure there is at least one coincidence arrival during $(0,\delta]$, and there is no coincidence or random arrival less than τ before:

$$\begin{aligned}
\mathrm{P}[E_4] &= \int_0^\delta \mathrm{P}[X(s-\tau,0)=0, N_1(s-\tau,s)=0, N_2(s-\tau,s)=0|T_X=s]f_{T_X}(s)ds \\
&= \int_0^\delta e^{-\lambda_X(\tau-s)} e^{-\lambda_1 \tau} e^{-\lambda_2 \tau} \lambda_X e^{-\lambda_X s} ds \\
&= e^{-(\lambda_1+\lambda_2)\tau} \lambda_X \delta e^{-\lambda_X \tau} \\
&= \lambda_X \delta e^{-(\lambda_X+\lambda_1+\lambda_2)\tau}. \tag{5.24}
\end{aligned}$$

Thus in total,

$$
\begin{aligned}
E[Y(0,\delta)] &= \mathrm{P}[E_1] + \mathrm{P}[E_2] + \mathrm{P}[E_3] + \mathrm{P}[E_4] \\
\gamma &= \left[\left(\frac{2\lambda_1\lambda_2}{\lambda_X} + \lambda_1 + \lambda_2 \right) \left(1 - e^{-\lambda_X \epsilon} \right) + \lambda_X \right] e^{-(\lambda_X + \lambda_1 + \lambda_2)\tau} \quad (5.25)
\end{aligned}
$$

$$
E[Y(t)] = t \left[\left(\frac{2\lambda_1\lambda_2}{\lambda_X} + \lambda_1 + \lambda_2 \right) \left(1 - e^{-\lambda_X \epsilon} \right) + \lambda_X \right] e^{-(\lambda_X + \lambda_1 + \lambda_2)\tau} \quad (5.26)
$$

We observe that since ϵ is very small, $\frac{1 - e^{-\lambda_X \epsilon}}{\lambda_X \epsilon} \approx 1$ and since λ_X is usually very small, $(\lambda_1 + \lambda_2)(1 - e^{-\lambda_X \epsilon}) \approx 0$. Hence

$$
E[Y(t)] \approx t(2\epsilon\lambda_1\lambda_2 + \lambda_X)e^{-(\lambda_X + \lambda_1 + \lambda_2)\tau}. \tag{5.27}
$$

5.3.2 Variance

Trying to find the exact expression for variance would be a painful exercise in this case; instead, we choose to bound it. As argued previously, $\beta(0,s) = 0$ for $0 < s < \tau$, and $\beta(0,s) = \gamma^2$ for $\tau + \epsilon < s < t$. For $\tau < s < \tau + \epsilon$, an event recorded at time s could be lost due to the deadtime effect of an event recorded at time 0, so we can bound $\beta(0,s)$ between 0 and γ^2. Hence $\int_0^t (t-s)\beta(0,s)ds$ in (4.25) can be bounded between $\int_{\tau+\epsilon}^t (t-s)\beta(0,s)ds$ and $\int_\tau^t (t-s)\beta(0,s)ds$. Hence

$$
\gamma t(1 - \gamma(\tau + \epsilon)(2 - (\tau + \epsilon)/t)) \leq \mathrm{Var}[Y(t)] \leq \gamma t(1 - \gamma\tau(2 - \tau/t)), \tag{5.28}
$$

where γ is as given in (5.25). This bound is very tight since usually $\epsilon \ll \tau$.

In PET, X photons can only originate from a narrow strip that connects the two detectors, while N_1 and N_2 photons can come from all other directions. Hence under normal circumstances, *i.e.*, when $\lambda_1 \approx \lambda_2 \gg \lambda_X$, the loss ratio of variance over mean:

$$
\xi \triangleq 1 - \mathrm{Var}[Y(t)]/E[Y(t)], \tag{5.29}
$$

which is approximately the expected number of recorded coincidences from 0 to 2τ, is extremely small. Only when λ_X is large enough to cause significant deadtime effect of its own, will the loss ratio be significant; the behavior of the system would then resemble the singles case where we have shown there would be significant loss of variance[94].

96

5.3.3 Attenuation Effects

The model (5.1) is unrealistic because it assumes both X photons are always recorded. In practice, due to attenuation, the probability that an X photon is recorded is less than 1. We assume the following model to include the effect of attenuation:

$$
\begin{aligned}
X_1(t) &= M_1(t) + N_1(t), \\
X_2(t) &= M_2(t) + N_2(t),
\end{aligned}
\tag{5.30}
$$

where $M_1(t)$ and $M_2(t)$ are independently binomial-thinned from the same Poisson process $X(t)$ of intensity λ_X, *i.e.*, given that $X(\delta) = 1$, the event $[M_1(\delta) = 1]$ and the event $[M_2(\delta) = 1]$ are independent. The survival probability for M_1 and M_2 are denoted p_1 and p_2, respectively. Let $\tilde{X}$ denote the (independently) binomial-thinned true coincidence process; it has intensity $p_1 p_2 \lambda_X$; let $\tilde{M}_1$ denote the number of pairs of X photons that only survived to reach detector 1, but not 2, and $\tilde{M}_2$ those that only survived to reach detector 2, but not 1. Then $\tilde{M}_1$ and $\tilde{M}_2$ have intensity $p_1(1 - p_2)\lambda_X$ and $p_2(1 - p_1)\lambda_X$, respectively. Furthermore, $\tilde{X}$, $\tilde{M}_1$, and $\tilde{M}_2$ are statistically independent, and, $\tilde{M}_1$ and $\tilde{M}_2$ contributes to the singles counts at detector 1 and 2, respectively. Hence this case reduces to the no-attenuation case (5.1), but now X in (5.1) has intensity $p_1 p_2 \lambda_X$, N_1 has intensity $\lambda_1 + p_1(1 - p_2)\lambda_X$, and N_2 has intensity $\lambda_2 + p_2(1 - p_1)\lambda_X$. Hence we can obtain the exact expression of $E[Y(t)]$ by plugging the appropriate values into (5.26). When λ_X is very small compared to λ_1 or λ_2, we can approximate $E[Y(t)]$ by

$$
E[Y(t)] \approx t(2\epsilon\lambda_1\lambda_2 + p_1 p_2 \lambda_X)e^{-(p_1 p_2 \lambda_X + \lambda_1 + \lambda_2)\tau}.
\tag{5.31}
$$

5.4 Delayed Coincidences

In PET systems, delayed coincidences are usually recorded for use in estimating the number of random coincidences recorded in the prompt window, so that coincidences in the prompt window can be corrected to reflect only "true" coincidences, *i.e.*, X photons [50, 79]. The inherent assumption in this scheme is that the mean of the coincidences in the delayed window is equal to the mean of the random coincidences in the prompt window.

Now we investigate the validity of this assumption. We denote d as the delay; we assume $3\epsilon/2 < d < \tau - \epsilon/2$. For a pair of photons to be recorded as a *delayed* coincidence event, both photons must first be recorded by their respective detectors and if one photon is recorded by detector 1 at time t_1, and one photon is recorded by detector 2 at time t_2, and $||t_1 - t_2| - d| < \epsilon/2$, then this pair of photons is recorded as a delayed coincidence event. From this definition, no coincidence event can be counted as a delayed coincidence event. To avoid ambiguity, we define the time of a delayed coincidence to be the arrival time of the later photon. Thus if there is one photon recorded by detector 1 at time t_1 and no photon is recorded by detector 2 at time t_1, then the number of delayed coincidences at time t_1 is the number of photons recorded by detector 2 during $(t_1 - d - \epsilon/2, t_1 - d + \epsilon/2)$.

Let $Y(t)$ denote the number of recorded delayed coincidence events during $(0, t]$. As in the coincidence case, we split $Y(t)$ into $Y_a(t)$ and $Y_b(t)$, where $Y_a(t)$ and $Y_b(t)$ denote the number of recorded coincidence events during $(0, t]$ that have the later singles event arriving at detector 1 and detector 2, respectively. Let T_1 denote the time of the first photon arrival at detector 1 after time 0, and T_2 denote the time of the first photon arrival at detector 2 after time $T_1 - d - \epsilon/2$. We derive $E[Y_a(0, \delta)]$ for $\delta < \tau$. For this derivation, we assume there is no attenuation; extension to the attenuation-case is straightforward. We have:

$$
\begin{aligned}
E[Y_a(0,\delta)] &= \int_0^\infty \mathrm{P}[Y_a(0,\delta) = 1 | T_1 = s_1] f_{T_1}(s_1) ds_1 \\
&= \int_0^\delta \mathrm{P}[Y_2(s_1 - d - \epsilon/2, s_1 - d + \epsilon/2) = 1, Y_1(0,\delta) = 1 | T_1 = s_1] f_{T_1}(s_1) ds_1 \\
&= \int_0^\delta \int_{s_1-d-\epsilon/2}^{s_1-d+\epsilon/2} \mathrm{P}[N_2(s_2 - \tau, s_1 - d - \epsilon/2) = 0, N_1(s_1 - \tau, 0) = 0, \\
&\qquad X(s_2 - \tau, s_1) = 0 | T_1 = s_1, T_2 = s_2] f_{T_2|T_1}(s_2|s_1) f_{T_1}(s_1) ds_2 ds_1 \\
&= \int_0^\delta \int_{s_1-d-\epsilon/2}^{s_1-d+\epsilon/2} e^{-\lambda_2(s_1-d-\epsilon/2-s_2+\tau)} e^{-\lambda_1(\tau-s_1)} e^{-\lambda_X(s_1+\tau-s_2)} \\
&\qquad \lambda_2 e^{-\lambda_2(s_2-s_1+d+\epsilon/2)} \lambda_1 e^{-\lambda_1 s_1} ds_2 ds_1 \\
&= \int_0^\delta \int_{s_1-d-\epsilon/2}^{s_1-d+\epsilon/2} \lambda_2 e^{-\lambda_2 \tau} \lambda_1 e^{-\lambda_1 \tau} e^{-\lambda_X(s_1+\tau-s_2)} ds_2 ds_1 \\
&= \lambda_2 e^{-\lambda_2 \tau} \lambda_1 e^{-\lambda_1 \tau} e^{-\lambda_X(\tau+d+\epsilon/2)} \delta \frac{(e^{\lambda_X \epsilon} - 1)}{\lambda_X}.
\end{aligned}
\tag{5.32}
$$

Hence

$$\lim_{\delta \to 0} \frac{1}{\delta} E[Y_a(0, \delta)] = \epsilon \gamma_1 \gamma_2 \frac{e^{\lambda_X \epsilon} - 1}{\lambda_X \epsilon} e^{-\lambda_X(\tau + d + \epsilon/2)}. \tag{5.33}$$

Using symmetry and the fact that $Y = Y_a + Y_b$, we have the following expression for the mean number of delayed coincidences:

$$E[Y(t)] = 2\epsilon \gamma_1 \gamma_2 t \frac{e^{\lambda_X \epsilon} - 1}{\lambda_X \epsilon} e^{-\lambda_X(\tau + d + \epsilon/2)}. \tag{5.34}$$

Hence in the presence of deadtime, the mean number of delayed coincidences is not exactly the same as the mean number of prompt random coincidences unless $\lambda_X = 0$. Nevertheless, since ϵ, d, and λ_X are usually very small, $E[Y(t)] \approx 2\epsilon \gamma_1 \gamma_2 t$.

5.5 Discussion

We have analyzed the mean and variance of recorded coincidence counts under various scenarios. Under all these scenarios, the coincidence counting process is not Poisson, even in the case of random coincidence counts recorded by ideal detectors. Nevertheless, we have shown that the variance is very close to the mean for any detector pair. For non-ideal detectors, the loss ratio of variance over mean ξ is approximately the expected number of recorded events during the time interval $(0, 2\tau]$. Unless the true coincidence rate is so high that $\xi \gg 0$, the ratio of variance over mean will be very close to 1. For coincidence counts recorded by two detector elements in two different detector blocks, ξ is even smaller, by similar analyses as in Section 4.4. Hence it appears that statistical image reconstruction based on Poisson models, while certainly not optimal, should be adequate in practice even under fairly large deadtime losses, provided the deadtime loss factor is included in the system matrix.

It is interesting to compare the coincidence counting process to the singles counting process. Whereas deadtime causes the single photon counting process to be significantly non-Poisson, in the sense that the variance of the process is significantly less than mean, the variance of the coincidence counting process is extremely close to the mean from previous analyses. The reason for this lies in the fact that the loss ratio of variance over mean, ξ, is

primarily determined by γ (in addition to τ), the instantaneous rate of the counting process itself. In the single photon counting process, γ is relatively large, but in the coincidence counting process, γ is usually very small even though γ_1 and γ_2 (the rates of recorded random single photons) may be very large. Thus deadtime causes significant loss in the mean of the process, but not in the ratio of the variance over the mean.

A possible extension to this work is to take into account the fact that two X photons do not necessarily arrive at the two detectors at exactly the same time, and furthermore, there is uncertainty in the time-stamping of recorded singles[19]. It is plausible that due to this uncertainty, some coincidences in the delayed window would be counted as coincidences in the prompt window, and *vice versa*, though any sensibly designed system should make the probability of such events extremely low.

CHAPTER 6

Maximum Likelihood Transmission Image Reconstruction for Overlapping Transmission Beams

6.1   Introduction

To reconstruct quantitatively accurate images of radioisotope emission distributions in SPECT, one must compensate for the effects of photon absorption or attenuation. Accurate attenuation correction requires good attenuation maps, and one can reconstruct such maps from transmission scan measurements obtained either prior to or simultaneously with the SPECT emission scan.

Several source/detector configurations for SPECT transmission scans have been investigated[9], including a single fixed line source opposite a symmetric fan-beam collimator, used in triple-head SPECT cameras, a scanning line source for orthogonal dual-head cameras, and offset line sources opposite asymmetric fan-beam collimators. Cellar *et al* [14] describe an alternative geometry based on several fixed-position collimated line sources opposing a parallel-beam collimator. In that system design, the source collimation was selected to minimize overlap on the detector of the transmitted "fan-beams." They then applied the filtered back-projection (FBP) algorithm to reconstruct the attenuation map (an ART algorithm was also mentioned without details). This source collimation has the undesirable consequence of very nonuniform count profiles, as shown in Figure 4 of [14]. It is natural to expect that higher and more uniform count profiles could lead to better reconstructed

attenuation maps *if the overlap can be properly modeled by the reconstruction method*.

In both the scanning line source geometry and the geometry of Celler *et al* [14], there can be overlap of the beam footprints, as illustrated in Figure 6.2. Previously published statistical algorithms for transmission tomography, e.g. [30, 31, 56, 74, 57, 71, 39], are inapplicable to the multiple source problem when the beams overlap. In this chapter we formulate a statistical model for multiple-source transmission measurements with arbitrary overlapping beams, and then derive an iterative algorithm for maximizing the likelihood (or a regularized variant thereof). The log-likelihood is not necessarily globally concave, which usually precludes proofs of convergence to a global maximum. The algorithm that we present is guaranteed to increase the likelihood at every iteration, and the set of fixed points of the algorithm is the same as the set of stationary points of the objective function. The algorithm also satisfies the continuity conditions of Meyer [67]. Therefore, by the convergence results in [67], the proposed algorithm produces a sequence of estimates that converge from any nonnegative initial image to a stationary point of the objective, provided the set of stationary points is not a continuum. This is nearly as strong of a convergence result as one might expect for a possibly nonconcave objective function. In Section 6.2, we give the statistical model and the proposed maximum-likelihood algorithm; in Section 6.3, we present some simulation results; in Section 6.4, we present the results from our preliminary study on optimal source collimation angles for a fixed system setup; and finally Section 6.5 is discussion and conclusion.

6.2 Statistical Model

Let Y_i denote the number of photons counted by the ith detector element[1] during the transmission scan, for $i = 1, \ldots, N$, where N is the number of measurement elements. Each detector element conceivably may count photons that originated in any of the $M \geq 1$ sources. We assume that separate blank scans are available for each of the sources (or source positions for a scanning line source). (This information is essential for unscrambling the multiplexing of overlapping beams.) Let b_{im} denote the mean number of photons that would be observed during a transmission scan by the ith detector originating in the mth source in the absence of any patient in the scanner. Typically the b_{im}'s would be determined by a periodically acquired calibrating "blank scan", performed separately for each of the M sources, and then scaled by the relative durations of the blank scan and transmission scans. However, we ignore any statistical uncertainty in the b_{im}'s and treat them as known constants. This assumption is reasonable provided the blank scans are sufficiently lengthy.

Let $\underline{\mu}^{\text{true}} = [\mu_1^{\text{true}}, \ldots, \mu_p^{\text{true}}]'$ denote the vector of unknown attenuation coefficients for each of the p pixels or voxels in the attenuation map (or the coefficients of some other basis for the attenuation distribution such as B-splines [25]). The line integral between the mth source and the ith detector location through the attenuating object is approximated by the following sum:

$$[\boldsymbol{A}^m \underline{\mu}]_i = \sum_{j=1}^{p} a_{ij}^m \mu_j,$$

where $\boldsymbol{A}^m = \{a_{ij}^m\}$ is a $N \times p$ matrix with nonnegative elements and the a_{ij}^m's represent line-lengths or normalized strip-intersection areas[2]. Thus by Beer's law the "survival probability" for a photon transmitted from the mth source in the direction of the i detector is (approximately) $\exp(-[\boldsymbol{A}^m \underline{\mu}]_i)$.

We assume the Y_i's have independent Poisson distributions:

$$Y_i \sim \text{Poisson}\left\{\bar{y}_i(\underline{\mu}^{\text{true}})\right\},$$

---

[1]Each "detector element" corresponds to a unique radial position and view angle, i.e., for typical 2D reconstruction $N = N_r N_\theta$ where N_r is the number of radial samples along the detector and N_θ is the number of view angles or "steps."

[2]Normalized by strip width

where the means are given by

$$\bar{y}_i(\underline{\mu}) = \left[\sum_{m=1}^{M} b_{im} \exp\left(-[\boldsymbol{A}^m \underline{\mu}]_i\right)\right] + r_i. \tag{6.1}$$

The r_i's are nonnegative constants that one can include to account for the mean contributions of scatter, room background, and emission crosstalk[43]. We treat these r_i's as known constants, though in practice they must be determined experimentally. However, since scatter is a spatially smooth function, one can safely smooth scatter estimates fairly heavily, so generally the uncertainty in the r_i's can be made much smaller than that of the Y_i's.

The summation over m in (6.1) allows for arbitrary overlap of the beams transmitted from each source. Non-overlapping beams would correspond to the assumption that if $b_{im} \neq 0$, then $b_{ik} = 0$ for all $k \neq m$, *i.e.* $b_{im}b_{ik} = 0$ for all $k \neq m$.

Under the above statistical model, given a particular measurement realization $Y = [y_1, \ldots, y_N]'$, we can write the log-likelihood for $\underline{\mu}$ in the following convenient form:

$$L(\underline{\mu}, Y) = \sum_{i=1}^{N} h_i\left(\sum_{m=1}^{M} u_{im}(\underline{\mu})\right), \quad h_i(t) = y_i \log t - t, \tag{6.2}$$

$$u_{im}(\underline{\mu}) = b_{im} \exp\left(-[\boldsymbol{A}^m \underline{\mu}]_i\right) + r_i/M,$$

ignoring constants independent of $\underline{\mu}$. Since the form of this log-likelihood is sufficiently different from the usual models for emission tomography and transmission tomography [56], previously derived algorithms for maximum likelihood estimation are not directly applicable to this problem.

One could easily derive an expectation-maximization (EM) algorithm [26] that would monotonically increase the likelihood $L(\underline{\mu}, Y)$ for this problem, generalizing [56]. However, the convergence would be as painfully slow and the M-step as difficult as the usual transmission EM algorithm. Instead, we propose an algorithm based on an extension of our recent work on paraboloidal surrogates methods [30, 31]. For even faster "convergence" one could apply ordered subsets ideas [33].

Because of the ill-posedness of the reconstruction problem, a penalty term is usually added to the likelihood to encourage piecewise smoothness in the reconstructed image,

104

resulting in the following objective function:

$$\Phi(\mu, Y) = L(\mu, Y) - \beta R(\mu). \tag{6.3}$$

Our goal is to produce a penalized-likelihood estimate:

$$\hat{\mu} = \arg\max_{\mu \geq \underline{0}} \Phi(\mu, Y). \tag{6.4}$$

Most roughness penalties $R(\mu)$ can be expressed in the following general form:

$$R(\mu) = \sum_{k=1}^{K} \psi_k([C\mu]_k), \tag{6.5}$$

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 here are convex, symmetric, nonnegative and differentiable [31].

6.3 Algorithm

We focus on the unregularized maximum-likelihood problem; the regularized approach easily follows from [31]. Since maximizing the log-likelihood directly is difficult to do, we apply the principle of optimization transfer [57, 4] and define a "surrogate function" $Q(\underline{\mu}; \underline{\mu}^n)$ that is easier to maximize. Since this surrogate function depends on the previous estimate $\underline{\mu}^n$ at the nth iteration, the algorithm consists of repeatedly maximizing the surrogate function, *i.e.*

$$\underline{\mu}^{n+1} = \arg\max_{\underline{\mu} \geq \underline{0}} Q(\underline{\mu}; \underline{\mu}^n). \tag{6.6}$$

Note that the maximization is constrained to enforce the nonnegativity constraint. The key algorithm design requirement is to choose Q functions that satisfy the following conditions:

$$Q(\underline{\mu}^n; \underline{\mu}^n) = L(\underline{\mu}^n, Y), \ \forall \underline{\mu}^n \geq \underline{0} \tag{6.7}$$

$$\left.\frac{\partial Q}{\partial \mu_j}(\mu; \mu^n)\right|_{\mu=\mu^n} = \left.\frac{\partial L}{\partial \mu_j}(\mu)\right|_{\mu=\mu^n}, \ \forall j = 1, \ldots, p \tag{6.8}$$

$$Q(\underline{\mu}; \underline{\mu}^n) \leq L(\underline{\mu}, Y), \ \forall \underline{\mu} \geq \underline{0}. \tag{6.9}$$

These conditions ensure that the proposed iteration monotonically increases the likelihood.

A difficulty in maximizing L is the sum over m within the logarithm in (6.2). To move the summation outside of the logarithm, we first adapt De Pierro's multiplicative convexity trick [23]. Because h_i is concave:

$$h_i\left(\sum_{m=1}^{M} u_{im}(\underline{\mu})\right) = h_i\left(\sum_{m=1}^{M} \frac{u_{im}^n}{\bar{y}_i^n} \frac{u_{im}(\underline{\mu})}{u_{im}^n} \bar{y}_i^n\right) \tag{6.10}$$

$$\geq \sum_{m=1}^{M} \frac{u_{im}^n}{\bar{y}_i^n} h_i\left(\frac{u_{im}(\underline{\mu})}{u_{im}^n} \bar{y}_i^n\right), \tag{6.11}$$

where $u_{im}^n \triangleq u_{im}(\underline{\mu}^n)$, and $\bar{y}_i^n \triangleq \bar{y}_i(\underline{\mu}^n)$. This inequality leads to our first surrogate function:

$$Q_1(\underline{\mu}; \underline{\mu}^n) \triangleq \sum_{i=1}^{N} \sum_{m=1}^{M} \frac{u_{im}^n}{\bar{y}_i^n} h_i\left(\frac{u_{im}(\underline{\mu})}{u_{im}^n} \bar{y}_i^n\right) \tag{6.12}$$

$$= \sum_{i=1}^{N} \sum_{m=1}^{M} \frac{u_{im}^n}{\bar{y}_i^n} g_{im}^n([A^m \mu]_i), \tag{6.13}$$

where

$$b_{im}^n \triangleq \frac{\bar{y}_i^n}{u_{im}^n} b_{im},$$

$$r_{im}^n \triangleq \frac{\bar{y}_i^n}{u_{im}^n} \frac{r_i}{M}$$

$$g_{im}^n(l) \triangleq y_i \log\left(b_{im}^n e^{-l} + r_{im}^n\right) - \left(b_{im}^n e^{-l} + r_{im}^n\right).$$

The surrogate function Q_1 remains too difficult to maximize directly because the argument of each h_i still depends on u_{im}, which has a complicated exponential form. However, it follows easily from the results in [30, 31] that the following paraboloidal function is a valid surrogate for Q_1:

$$Q_2(\underline{\mu}; \underline{\mu}^n) \triangleq \sum_{i=1}^{N} \sum_{m=1}^{M} \frac{u_{im}^n}{\bar{y}_i^n} q_{im}^n([\boldsymbol{A}^m \underline{\mu}]_i) \triangleq Q(\underline{\mu}; \underline{\mu}^n) \tag{6.14}$$

where

$$q_{im}^n(l) \triangleq g_{im}^n(l_{im}^n) + \dot{g}_{im}^n(l_{im}^n)(l - l_{im}^n) - \frac{1}{2} c_{im}^n(l - l_{im}^n)^2 \tag{6.15}$$

and

$$\dot{g}_{im}^n(l) \triangleq \frac{d}{dl} g_{im}^n(l)$$

$$\ddot{g}^n_{im}(l) \quad \triangleq \quad \frac{d^2}{dl^2} g^n_{im}(l)$$

$$l^n_{im} = [A^m \mu^n]_i \quad = \quad \sum_{j=1}^{p} A^m_{ij} \mu^n_j.$$

To ensure (6.9), we must choose the curvatures $\{c^n_{im}\}$ appropriately[30, 31]. As discussed in [31], for the fastest convergence rate, we would like to choose the curvatures as small as possible, subject to the constraint that the surrogate function q^n_{im} lies below the functions g^n_{im}. For completeness, we include the following formula for the optimum curvature, *i.e.*, $c^n_{im} = \min\{c \geq 0 : g^n_{im}(l) \leq q^n_{im}(l) \ \forall l \geq 0\}$ derived in [31]:

$$c^n_{im} = \begin{cases} \left[-2 \dfrac{g^n_{im}(0) - g^n_{im}(l^n_{im}) + \dot{g}^n_{im}(l^n_{im})(l^n_{im})}{(l^n_{im})^2} \right]_+ , & l^n_{im} > 0 \\[2ex] \left[-\ddot{g}^n_{im}(0) \right]_+ , & l^n_{im} = 0 \end{cases} \tag{6.16}$$

$$= \begin{cases} \left[(2/(l^n_{im})^2) \left\{ b^n_{im}(1 - e^{-l^n_{im}}) - y_i \log \dfrac{b^n_{im} + r^n_{im}}{\bar{y}^n_{im}} + b^n_{im} l^n_{im} e^{-l^n_{im}} \left(\dfrac{y_i}{\bar{y}^n_{im}} - 1 \right) \right\} \right]_+ , & l^n_{im} > 0 \\[3ex] \left[1 - \dfrac{y_i r^n_{im}}{(b^n_{im} + r^n_{im})^2} \right]_+ , & l^n_{im} = 0, \end{cases} \tag{6.17}$$

where

$$\bar{y}^n_{im} \stackrel{\triangle}{=} b^n_{im} e^{-l^n_{im}} + r^n_{im}.$$

Other curvature choices that lead to even faster convergence (but do not guarantee monotonicity) can be found in [31].

Since our second surrogate Q_2 is a quadratic functional, it is easily maximized by a variety of algorithms, including the coordinate ascent algorithm[74, 37]. Adding a penalty function is straightforward. However, Q_2 is not separable and if we want to apply the ordered subsets idea, we must have a separable surrogate function.

6.3.1 Separable Paraboloidal Surrogate Algorithm

Now we derive a separable paraboloidal surrogate algorithm; we do not use this algorithm for the simulation described in this chapter since it converges very slowly even though it is guaranteed to be monotonic. As noted by De Pierro in [23]:

$$[A^m \underline{\mu}]_i = \sum_{j=1}^{p} z^m_{ij} \left[\frac{a^m_{ij}}{z^m_{ij}} (\mu_j - \mu^n_j) + [A^m \underline{\mu}]_i \right] , \tag{6.18}$$

107

provided $\sum_{j=1}^{p} z_{ij}^m = 1$ for all i and m. Thus if $z_{ij}^m \geq 0$ then it follows from the concavity of q_{im}^n that

$$q_{im}^n([\boldsymbol{A}^m \underline{\mu}]_i) \geq \sum_{j=1}^{p} z_{ij}^m q_{im}^n \left(\frac{a_{ij}^m}{z_{ij}^m}(\mu_j - \mu_j^n) + [\boldsymbol{A}^m \underline{\mu}]_i \right). \tag{6.19}$$

If we choose $z_{ij}^m = a_{ij}^m / \gamma_i^m$, where $\gamma_i^m = \sum_{j=1}^{p} a_{ij}^m$ (such that $\sum_{j=1}^{p} a_{ij}^m / \gamma_i^m = 1$ for all i and m), then from (6.12) and (6.19) we obtain our third surrogate function as follows:

$$Q_3(\mu; \mu^n) = \sum_{i=1}^{N} \sum_{m=1}^{M} \sum_{j=1}^{p} \frac{u_{im}^n a_{ij}^m}{\bar{y}_i^{\,n} \gamma_i^m} q_{im}^n (\gamma_i^m(\mu_j - \mu_j^n) + [\boldsymbol{A}^m \mu^n]_i). \tag{6.20}$$

This surrogate function is a quadratic form, and one that is trivial to maximize because it is a separable function. The partial derivatives of Q_3 are given as follows:

$$-\frac{\partial^2}{\partial \mu_j^2} Q_3(\mu; \mu^n) = \sum_{i=1}^{N} \sum_{m=1}^{M} \frac{u_{im}^n a_{ij}^m \gamma_i^m}{\bar{y}_i^{\,n}} c_{im}^n \triangleq d_j(\mu^n) \tag{6.21}$$

$$\frac{\partial}{\partial \mu_j} Q_3(\mu; \mu^n) = \sum_{i=1}^{N} \sum_{m=1}^{M} \frac{u_{im}^n a_{ij}^m}{\bar{y}_i^{\,n}} \dot{g}_{im}(l_{im}^n) - \sum_{i=1}^{N} \sum_{m=1}^{M} \frac{u_{im}^n a_{ij}^m \gamma_i^m}{\bar{y}_i^{\,n}} c_{im}^n(\mu_j - \mu_j^n) \tag{6.22}$$

Note that (6.8) is satisfied since

$$\frac{\partial}{\partial \mu_j} Q_3(\mu; \mu^n) \bigg|_{\mu_j = \mu_j^n} = \sum_{i=1}^{N} \sum_{m=1}^{M} \frac{u_{im}^n a_{ij}^m}{\bar{y}_i^{\,n}} \dot{g}_{im}(l_{im}^n) = \frac{\partial}{\partial \mu_j} Q_2(\mu; \mu^n) \bigg|_{\mu_j = \mu_j^n} = \frac{\partial}{\partial \mu_j} L(\mu, Y) \bigg|_{\mu = \mu^n} \tag{6.23}$$

From the above expressions for the partial derivatives of Q_3, the unconstrained maximizer of $Q_3(\mu; \mu^n)$ with regard to μ_j is given by:

$$\mu_j^n + \frac{1}{d_j(\mu^n)} \frac{\partial}{\partial \mu_j} Q_3(\mu; \mu^n) \bigg|_{\mu_j = \mu_j^n}. \tag{6.24}$$

Since Q_3 is separable and concave, the iterative algorithm for the maximization of L is:

$$\mu^{n+1} = [\mu^n + \boldsymbol{D}(\mu^n)^{-1} \nabla' L(\mu^n, Y)]_+, \tag{6.25}$$

where $\nabla' = [\frac{\partial}{\partial \mu_1} \ldots \frac{\partial}{\partial \mu_p}]'$ denotes the (column) gradient operator, $[x]_+ = x$ for $x > 0$ and zero otherwise, and $\boldsymbol{D}(\mu^n)$ is a $p \times p$ diagonal matrix with jth diagonal entry $d_j(\mu^n)$. The ordered subsets idea could also be easily applied to this algorithm [52, 32] and adding regularization is straightforward as given in [32].

The iteration (6.25) monotonically increases the likelihood, however, it has very slow convergence due to the small curvatures of the surrogate parabolas. Applying the ordered subsets idea leads to faster "convergence" but monotonicity is no longer guaranteed.

6.3.2 Coordinate Ascent Algorithm

To obtain a monotonic algorithm that converges relatively quickly, we can apply coordinate ascent to the surrogate Q defined in (6.14), *i.e.*, sequentially update one pixel at a time while holding all other pixels fixed. First, we obtain the likelihood surrogate parabola for a particular pixel with every other pixel value fixed:

$$
\begin{aligned}
\hat{Q}_j^n(\mu_j) &\triangleq Q([\hat{\mu}_1, \ldots, \hat{\mu}_{j-1}, \mu_j, \hat{\mu}_{j+1}, \ldots, \hat{\mu}_p]; \mu^n) \\
&= Q(\hat{\mu}; \mu^n) + \dot{Q}_j^n(\hat{\mu}_j)(\mu_j - \hat{\mu}_j) - \frac{1}{2} d_j^n(\mu_j - \hat{\mu}_j)^2,
\end{aligned}
\tag{6.26}
$$

where $\hat{\mu}$ denotes the most recent estimate of μ, $\hat{\mu}_j$ denotes the jth entry of $\hat{\mu}$, μ^n denotes the estimate for μ after the nth iteration (with which $Q(\cdot, \mu^n)$ is constructed), and $\hat{Q}_j^n(\mu_j)$ is treated as a function of μ_j only. The derivative of the likelihood surrogate parabola at $\hat{\mu}_j$ is:

$$
\dot{Q}_j^n(\hat{\mu}) \triangleq \frac{\partial}{\partial \mu_j} Q_j^n(\mu_j) \bigg|_{\mu_j = \hat{\mu}_j} = \sum_{i=1}^N \sum_{m=1}^M \frac{u_{im}^n}{\bar{y}_i^n} a_{ij}^m [\dot{g}_{im}^n(l_{im}^n) - c_{im}^n(\hat{l}_{im} - l_{im}^n)],
\tag{6.27}
$$

where

$$
\hat{l}_{im} \triangleq [\mathbf{A}^m \hat{\mu}]_i = \sum_{j=1}^p \mathbf{A}_{ij}^m \hat{\mu}_j
$$

$$
\dot{g}_{im}^n(l) = \left(\frac{y_i}{b_{im}^n e^{-l} + r_{im}^n} - 1 \right) b_{im}^n e^{-l}.
$$

The curvature of the parabola $\hat{Q}_j^n(\mu_j)$ is

$$
d_j^n \triangleq \sum_{i=1}^N \sum_{m=1}^M \frac{u_{im}^n}{\bar{y}_i^n} (a_{ij}^m)^2 c_{im}^n.
\tag{6.28}
$$

Extension to the penalized likelihood case is straightforward following the methods in[30, 31], so we omit the details and only comment on notations: $\hat{R}_j(\mu_j)$ denotes the penalty surrogate parabola for μ_j, and $\hat{p}_j$ denotes its curvature. Combining the likelihood surrogate parabolas in (6.26) and the penalty surrogate parabolas, the maximization step of the coordinate ascent for pixel j is:

$$
\hat{\mu}_j^{\text{new}} = \arg \max_{\mu_j \geq 0} \hat{Q}_j^n(\mu_j) - \beta \hat{R}_j(\mu_j) = \left[\hat{\mu}_j + \frac{\dot{Q}_j^n(\hat{\mu}) - \beta \dot{R}_j(\hat{\mu})}{d_j^n + \beta \hat{p}_j} \right]_+.
\tag{6.29}
$$

Because of our construction based on surrogate functions that satisfy (6.9), this update is guaranteed to monotonically increase the value of Φ. One iteration is finished when all pixels are updated via (6.29) in a sequential order. We update the paraboloidal surrogate function after one iteration of coordinate ascent (CA), although one could also perform more than one CA iteration per surrogate. An outline of this algorithm is given in Table 6.1, where ψ_k denote the potential function used in the penalty acting as a norm on the "soft constraints" $C\mu \approx 0$, K is the number of such constraints, and $\hat{\psi}_k$ denotes the surrogate function used for ψ_k; see [31] for details about possible surrogate functions used for ψ_k.

Initialize: $\hat{\mu} = \mathrm{FBP}\{\log((\sum_{m=1}^{M} b_{im})/(y_i - r_i))\}_{i=1}^{N}$ and $\hat{l}_{im} = \sum_{j=1}^{p} a_{ij}^{m}\hat{\mu}_j, i = 1, \ldots, N$

for each iteration $n = 1, \ldots,$ niter

$$
\begin{aligned}
u_{im} &= b_{im}e^{-[\boldsymbol{A}\hat{\mu}]_i} + r_i/M \\
\bar{y}_i &= \sum_{m=1}^{M} b_{im}e^{-[\boldsymbol{A}\hat{\mu}]_i} + r_i \\
b_{im}^{n} &= \frac{\bar{y}_i}{u_{im}}b_{im}, r_{im}^{n} = \frac{\bar{y}_i}{u_{im}}\frac{r_i}{M}
\end{aligned}
$$

Compute c_{im} according to (6.16)

$$
\dot{q}_{im} = \dot{g}_{im} = \left(\frac{y_i}{b_{im}^{n}e^{-\hat{l}_{im}} + r_{im}^{n}} - 1\right) b_{im}^{n}e^{-\hat{l}_{im}}, \text{for } i = 1, \ldots, N
$$

for $j = 1, \ldots, p$

$$
\begin{aligned}
\dot{Q}_j &= \sum_{m=1}^{M}\sum_{i=1}^{N} \frac{u_{im}}{\bar{y}_i}a_{ij}^{m}\dot{q}_{im} \\
d_j &= \sum_{m=1}^{M}\sum_{i=1}^{N} \frac{u_{im}}{\bar{y}_i}(a_{ij}^{m})^2 c_{im} \\
\mu_j^{\mathrm{old}} &= \mu_j \\
\hat{\mu}_j &:= \left[\hat{\mu}_j + \frac{\dot{Q}_j - \beta\sum_{k=1}^{K} c_{kj}\dot{\psi}_k([\boldsymbol{C}\hat{\mu}]_k)}{d_j + \beta\sum_{k=1}^{K} c_{kj}^2\ddot{\psi}_k([\boldsymbol{C}\hat{\mu}]_k)}\right]_+ \\
\dot{q}_{im} &:= \dot{q}_{im} - a_{ij}^{m}c_{im}(\hat{\mu}_j - \mu_j^{\mathrm{old}})
\end{aligned}
$$

end

$$
\hat{l}_{im} := \hat{l}_{im} + \frac{\dot{q}_{im} - \dot{g}_{im}}{c_{im}}, \text{for } i = 1, \ldots, N
$$

end

Table 6.1: Algorithm outline

6.4   Simulation Results

We compared the proposed reconstruction algorithm to the "conventional" reconstruction algorithms (statistical and FBP) that treat the transmission measurements simply as ideal normalized parallel "strip-integrals". The system geometry corresponded fairly closely to the Siemens $\mathrm{Profile}^{\mathrm{TM}}$ system (Hoffman Estates, IL) [42]. The sources for the simulated system consisted of a multiple line source array with 14 sources, unequally spaced, located on a line parallel to the detector and 110 cm away from the detector plane. The detector plane was located 22 cm away from the center of rotation. We simulated a range of source collimation angles, from $1.6°$ (with almost no overlap in the transmission beams) to $6.6°$, and disregarded detector collimation. The image consisted 128×128 pixels of size $3.56 \times 3.56\mathrm{mm}^2$. The sinogram size was 128×60 with detector bins of width $4.8\mathrm{mm}$ (*i.e.*, the simulated detector response was rectangular with width $4.8\mathrm{mm}$). We performed the simulation for two levels of transmitted counts, one corresponding to a system whose center rods have just been replaced (the new source case), and the other with sources that have all decayed by one half-life (the old source case). At $1.6°$, we simulated 321,000 transmitted counts for the new source case (160,000 counts for the old source case), and 263,000 background counts (on average). As the source collimation angle increases, the number of transmitted counts increases, naturally, but the number of background counts remains the same; at $2.6°$, there are 523,000 transmitted counts for the new source case (261,000 counts for the old source case); and at $6.6°$, there are 1,396,000 transmitted counts for the new source case. For simplicity, we used a space-invariant quadratic penalty over first-order neighbors throughout our simulations. The phantom used in our simulations, the ROI used for the evaluation of variance in Section 6.5 (outlined by solid lines), and the large region used for the evaluation of spatial resolution (outlined by dashed lines) are shown in Figure 6.1. Figure 6.2 illustrates the system setup. The radial distribution of blank counts (at any projection angle) is shown in Figure 6.3.
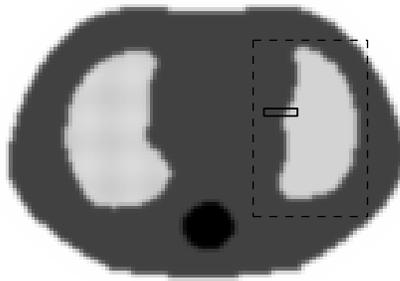
Figure 6.1: Digital Phantom used in our simulations and the ROI used for collimation angle optimization.
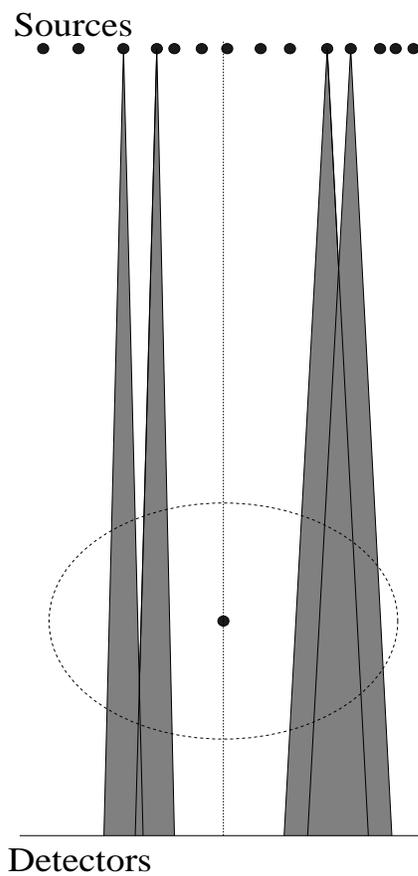
Figure 6.2: Scaled illustration of the system setup; the two fan-beams on the left have collimation angle $2.6°$; the two fan-beams on the right side have collimation angle $5.6°$.

(a)   (b)
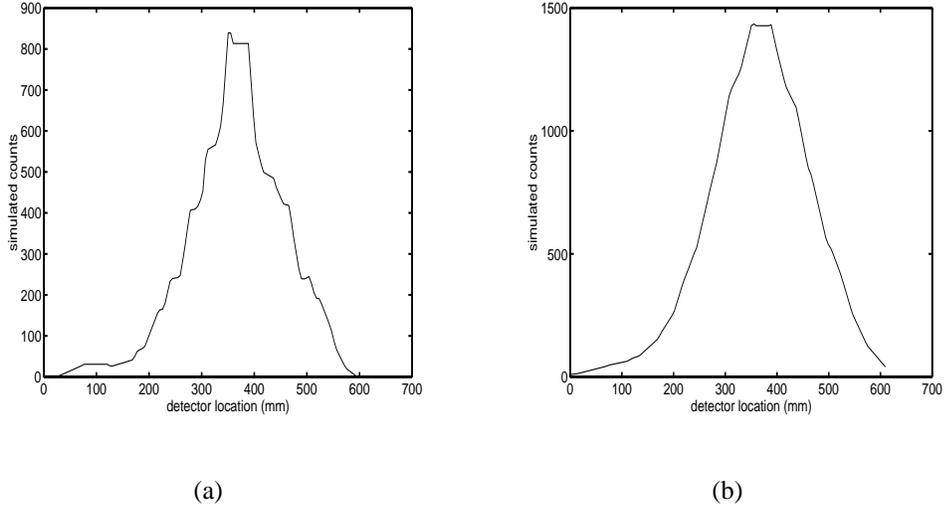
Figure 6.3: Distribution of blank counts (a) collimation angle $2.6°$ (b) collimation angle $5.6°$.

Figures 6.4 and 6.5 show reconstructions of noisy data using FBP, the parallel algorithm[3], and the proposed algorithm, with new sources. Figure 6.4 shows reconstructions with $2.6°$ source collimation and Figure 6.5 shows reconstructions with $3.6°$ source collimation. Figures 6.6 and 6.7 show the same reconstructions with old sources. The spatial resolution of the images in these figures is 4.7 pixels on the top row and 6.8 pixels on the bottom row. We use the following simple method to determine the resolution of a particular (noiseless) reconstruction. Given the ideal image μ^{true} and the reconstructed image $\check{\mu}$ using ideal data:

$$\check{\mu} = \arg \max_{\mu \geq \underline{0}} \Phi\big(\mu, \bar{y}_i(\mu^{\text{true}})\big), \tag{6.30}$$

we define the resolution of $\check{\mu}$ to be:

$$\arg \min_{\sigma} \sum_{j \in \mathcal{R}} \big|[G_\sigma \mu^{\text{true}}]_j - \check{\mu}_j\big|^2 \tag{6.31}$$

where G_σ represents a Gaussian smoothing filter with FWHM σ, and the $\mathcal{R}$ is a large region encompassing both the right lung and our ROI as illustrated in Figure 6.1.

We observe from Figures 6.4-6.7 that the proposed algorithm consistently produces less noisy reconstructions than both the parallel algorithm and FBP (this conclusion from

[3]We refer to the penalized-likelihood reconstruction assuming ideal normalized parallel "strip-integrals" (and ignoring beam overlap) as the parallel algorithm.

anecdotal evidence will be confirmed in Section 6.5). The noise reduction is especially significant when the collimation angle is large (*i.e.*, $3.6°$) and the desired spatial resolution is high (*i.e.*, 4.7 pixels). In such cases, FBP simply cannot produce a reconstructed image of the desired resolution even with an unapodized ramp filter. Since the parallel algorithm is based on an incorrect system and statistical model, one expects artifacts due to model mismatch. The absence of apparent artifacts in Figures 6.4-6.7 is due to regularization and noise. Figure 6.8 shows the reconstructed images from noiseless data ($4.6°$ collimation angle) using the parallel and proposed algorithms with almost no regularization ($\beta = 2^{-10}$). The reconstructed image from noiseless data using the parallel algorithm shows severe artifacts resulting from model mismatch, which are absent in the reconstructed image from noiseless data using the proposed algorithm (Figure 6.8c).

As the collimation angle increases, the artifacts generated by the parallel algorithm worsen. In fact, even the noiseless reconstruction has a spatial resolution of about 5.6 pixels when the collimation angle is $4.6°$ (Figure 6.8a)[4], and at $6.6°$, the noiseless reconstruction has a spatial resolution of 7.7 pixels. Thus for system setups with significantly overlapping transmission beams, the parallel algorithm simply cannot produce a high-resolution reconstruction, no matter how many counts one collects.

---

[4]The noiseless reconstruction (Figure 6.8c) using the proposed algorithm has a spatial resolution of 1.4 pixels.

FBP **Parallel algorithm** **Proposed algorithm**
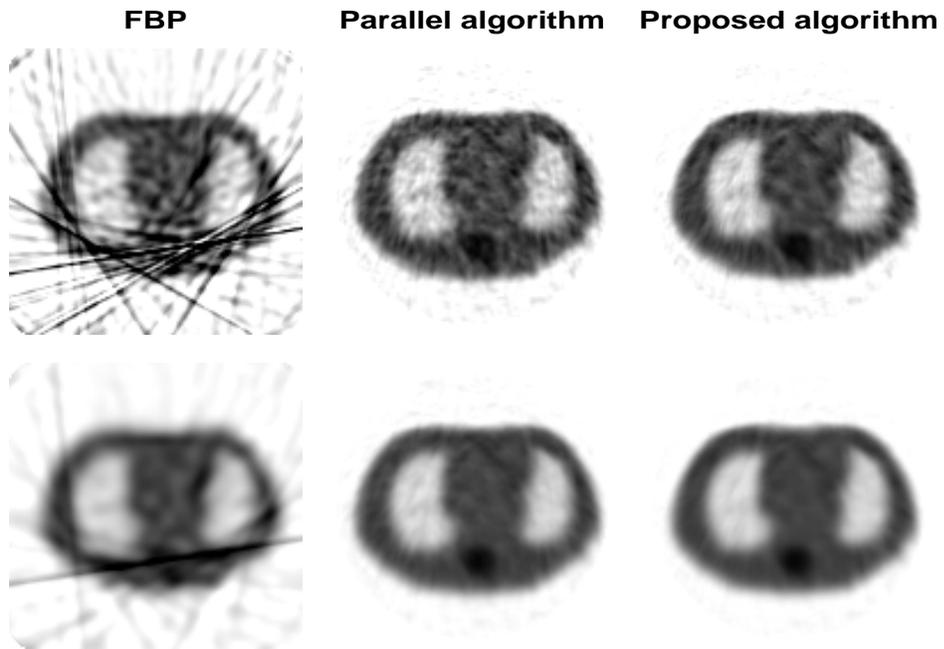
Figure 6.4: New sources; collimation angle: $2.6°$; 785,000 counts; top row: resolution 4.7 pixels; bottom row, resolution 6.8 pixels.

FBP **Parallel algorithm** **Proposed algorithm**
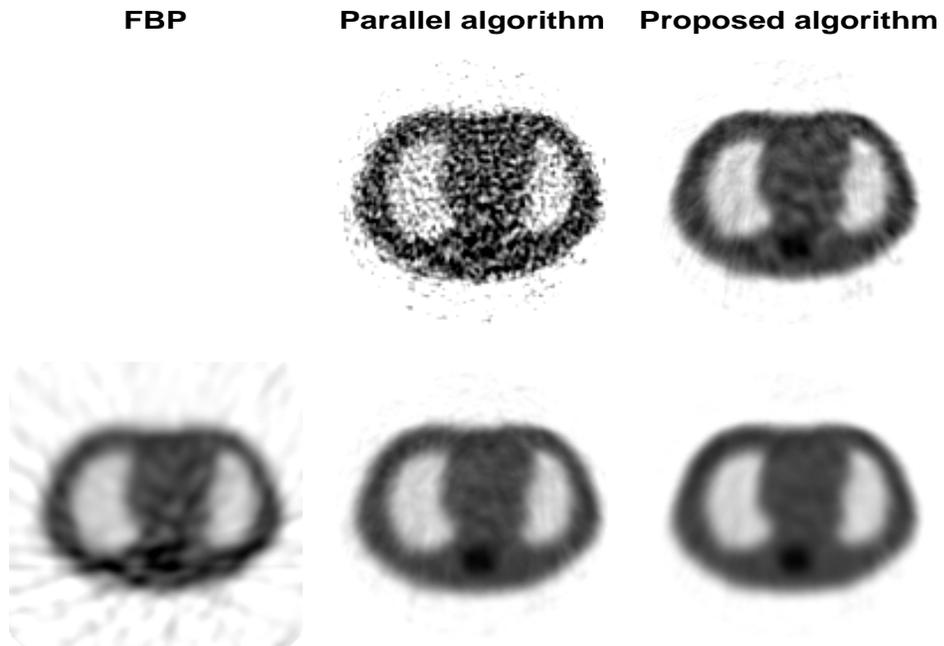
Figure 6.5: New sources; collimation angle: $3.6°$; 994,000 counts; top row: resolution 4.7 pixels; bottom row, resolution 6.8 pixels. A resolution of 4.7 pixels was not achievable with FBP in this case.
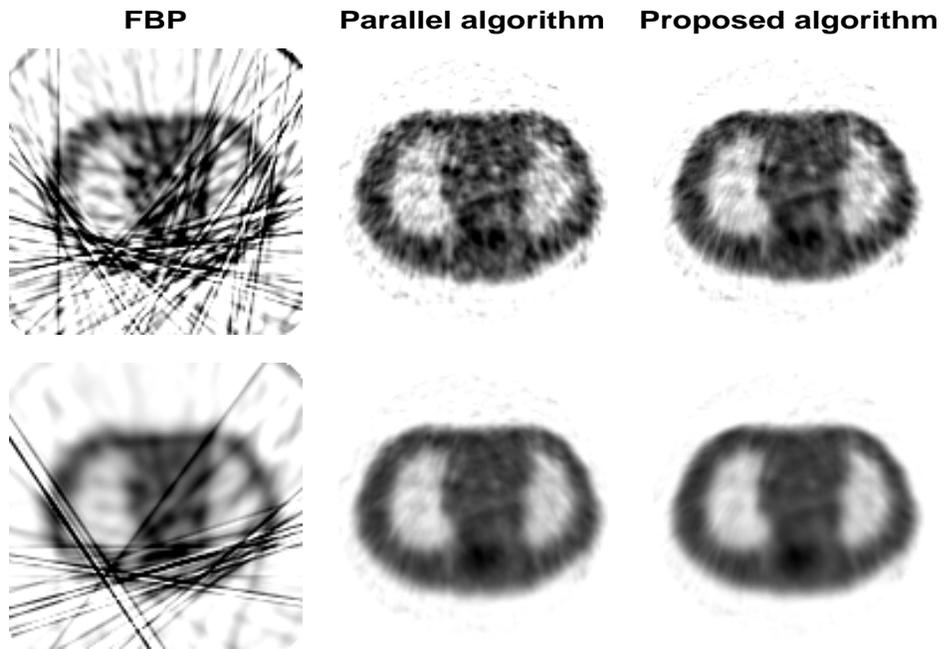
Figure 6.6: Old sources; collimation angle: $2.6°$; 392,000 counts; top row: resolution 4.7 pixels; bottom row, resolution 6.8 pixels.
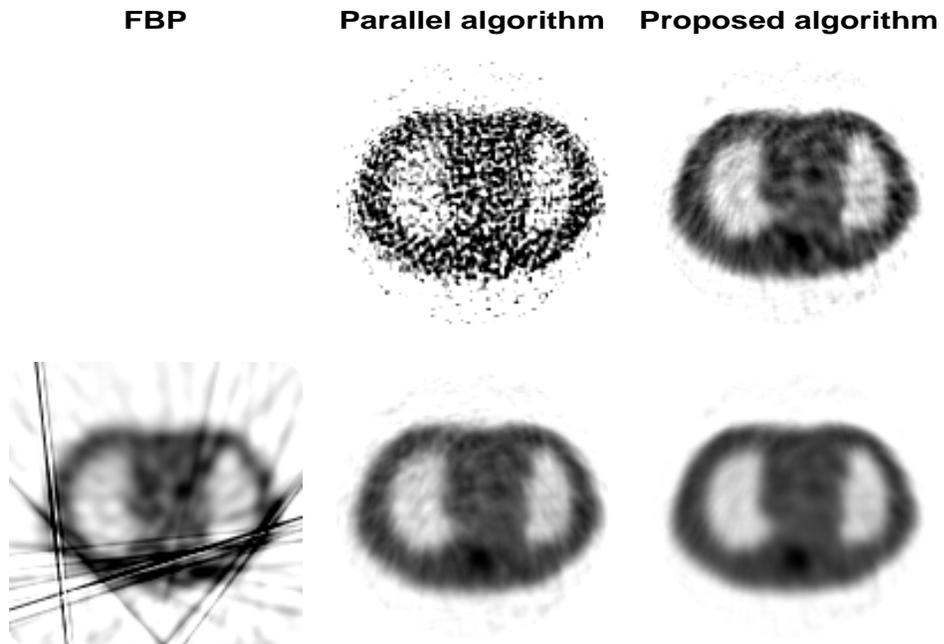
Figure 6.7: Old sources; collimation angle: $3.6°$; 497,000 counts; top row: resolution 4.7 pixels; bottom row, resolution 6.8 pixels. A resolution of 4.7 pixels was not achievable with FBP in this case.
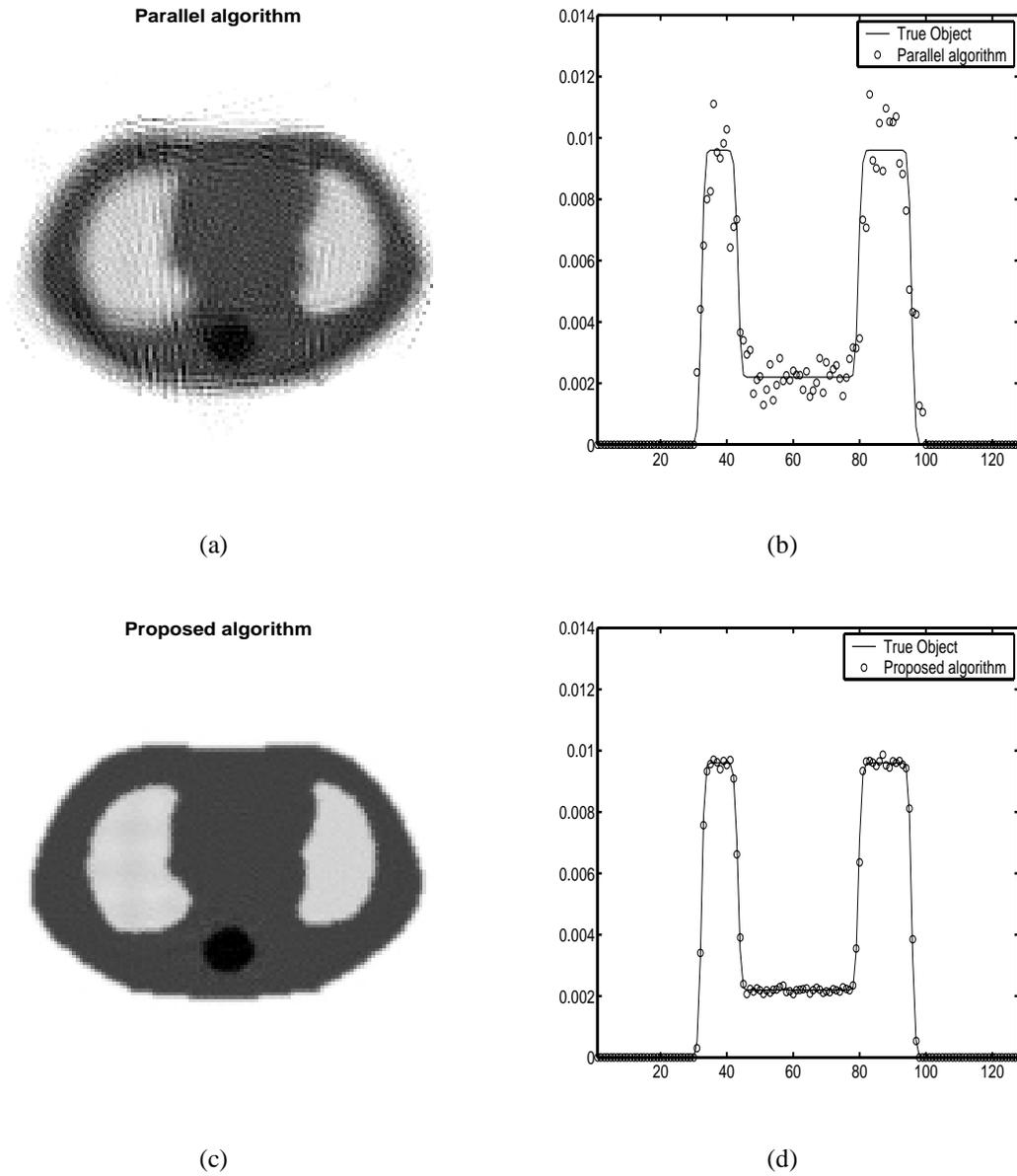
Figure 6.8: Reconstruction using the parallel and proposed algorithms with almost no regularization; collimation angle $4.6°$ (a) (b) Parallel algorithm (c) (d) Proposed algorithm.

6.5 Collimation Angle Optimization

We performed a preliminary study of the optimal source collimation angle given that the system configuration, source strength, and background counts remain constant. As source collimators open up, *i.e.*, the collimation angle increases, there would be more counts, but there would also be more overlap of transmission beams. Hence initially, when the transmission beams widen from no overlap to some overlap, we expect better resolution/variance tradeoffs; however, as the transmission beams open up more and more, we expect less and less improvements, and eventually worse resolution/variance tradeoff since eventually each detected photon hitting will yield very little information about where it originated. We want to obtain the collimation angle that minimizes a region of interest (ROI) variance for a fixed spatial resolution. We investigated the proposed algorithm, the parallel algorithm (*i.e.*, the conventional statistical algorithm), and FBP (derived assuming ideal parallel "line-integrals"). For the statistical algorithms, instead of performing numerical simulations, we used the approach outlined in [38] to compute the variance of the ROI. For FBP, since numerical simulations are relatively inexpensive, we performed 2000 realizations for each data point.

The approximate covariance for an implicitly defined biased estimator is[38]:

$$\mathrm{Cov}(\hat{\mu}) \approx [-\nabla^{20}\Phi(\check{\mu},\bar{y})]^{-1}\nabla^{11}\Phi(\check{\mu},\bar{y})\mathrm{Cov}\{Y\}[\nabla^{11}\Phi(\check{\mu},\bar{y})]'[-\nabla^{20}\Phi(\check{\mu},\bar{y})]^{-1}, \quad (6.32)$$

where $\check{\mu}$ is defined in (6.30). We first derive $\nabla^{20}\Phi(\mu,Y)$ and $\nabla^{11}\Phi(\mu,Y)$, which are needed for computing the variance of a ROI for the proposed algorithm. Defining

$$p_{im}(\mu) \quad \overset{\triangle}{=} \quad b_{im}e^{-\sum_j a_{ij}^m \mu_j} \quad (6.33)$$

$$\bar{y}_i(\mu) \quad = \quad \sum_{m=1}^{M} p_{im}(\mu) + r_i, \quad (6.34)$$

then

$$\frac{\partial p_{im}(\mu)}{\partial \mu_j} = -a_{ij}^m p_{im}(\mu) = \frac{\partial \bar{y}_i(\mu)}{\partial \mu_j}. \quad (6.35)$$

Since

$$L(\mu,Y) \quad = \quad \sum_{i=1}^{N} y_i \log \bar{y}_i(\mu) - \bar{y}_i(\mu), \quad (6.36)$$

119

we have

$$\frac{\partial L(\mu, Y)}{\partial \mu_j} = \sum_{i=1}^{N} \frac{y_i}{\bar{y}_i(\mu)} \frac{\partial \bar{y}_i(\mu)}{\partial \mu_j} - \frac{\partial \bar{y}_i(\mu)}{\partial \mu_j} = \sum_{i=1}^{N} \left(1 - \frac{y_i}{\bar{y}_i(\mu)} \right) \sum_{m=1}^{M} a_{ij}^m p_{im}(\mu) \quad (6.37)$$

and, applying the chain rule:

$$\frac{\partial^2 L(\mu, Y)}{\partial \mu_j \partial y_i} = -\frac{\sum_{m=1}^{M} a_{ij}^m p_{im}(\mu)}{\bar{y}_i(\mu)} \quad (6.38)$$

$$\frac{\partial^2 L(\mu, Y)}{\partial \mu_j \partial \mu_k} = \frac{\partial}{\partial \mu_k} \sum_{i=1}^{N} \left(1 - \frac{y_i}{\bar{y}_i(\mu)} \right) \sum_{m=1}^{M} a_{ij}^m p_{im}(\mu)$$

$$= -\sum_{i=1}^{N} \left(1 - \frac{y_i}{\bar{y}_i(\mu)} \right) \sum_{m=1}^{M} a_{ij}^m a_{ik}^m p_{im}(\mu)$$

$$-\sum_{i=1}^{N} \frac{y_i}{\bar{y}_i^2(\mu)} \left(\sum_{m=1}^{M} a_{ij}^m p_{im}(\mu) \right) \left(\sum_{m=1}^{M} a_{ik}^m p_{im}(\mu) \right). \quad (6.39)$$

Let $\mathbf{C}(\mu) \stackrel{\triangle}{=} \{c_{ij}(\mu)\}$, where $c_{ij}(\mu) \stackrel{\triangle}{=} \sum_{m=1}^{M} a_{ij}^m p_i^m(\mu)$ so $\mathbf{C}(\mu) = \sum_{m=1}^{M} \text{diag}\{p_i^m(\mu)\} \boldsymbol{A}^m$.
Then we obtain:

$$\nabla^{11}\Phi(\mu, Y) = -\mathbf{C}(\mu)'\text{diag}\{1/\bar{y}_i(\mu)\} \quad (6.40)$$

$$-\nabla^{20}\Phi(\mu, Y) = \sum_{m=1}^{M} (\boldsymbol{A}^m)'\text{diag}\{(\bar{y}_i(\mu) - r_i)(1 - y_i/\bar{y}_i(\mu))\}\boldsymbol{A}^m$$

$$+ T\mathbf{C}(\mu)'\text{diag}\{y_i/\bar{y}_i^2(\mu)\}\mathbf{C}(\mu) + \beta \mathbf{R}(\mu), \quad (6.41)$$

where $\mathbf{R}(\mu) = \nabla^2 R(\mu)$. (For a detailed derivation of the penalty part, see [38].) We use the following recipe to compute the approximate estimator (using the proposed algorithm) variance of a ROI:

(i) Compute $\check{\mu}$ by applying the proposed algorithm to noise-free data $\{\bar{y}_i(\mu^{\text{true}})\}$;

(ii) Forward project $\check{\mu}$ to compute $\bar{y}_i(\check{\mu}) = \sum_{m=1}^{M} p_{im}(\check{\mu}) + r_i$;

(iii) Use an iterative method such as conjugate gradient [91] or Gauss-Siedel [18] to solve

$[-\nabla^{20}\Phi(\check{\mu}, \bar{y})]u^{\text{ROI}} = e^{\text{ROI}}$, where e^{ROI} is a vector with entries $1/n_{\text{ROI}}$ for pixels inside the region and zeros for pixels outside; n_{ROI} denotes the number of pixels in the ROI;

120

(iv) Compute $(u^{\mathrm{ROI}})'\nabla^{11}\Phi(\check{\mu},\bar{y})\mathrm{Cov}\{Y\}[\nabla^{11}\Phi(\check{\mu},\bar{y})]'u^{\mathrm{ROI}}$ by first forward projecting u^{ROI} to compute $v = \mathrm{diag}\{1/\bar{y}_i(\check{\mu})\}\mathbf{C}(\check{\mu})u^{\mathrm{ROI}}$, and then summing:

$$\mathrm{Var}(\hat{\mu}^{\mathrm{ROI}}) \approx \sum_{i=1}^{N} v_i^2 y_i(\mu^{\mathrm{true}}). \qquad (6.42)$$

For the parallel algorithm, the computation of the approximate variance has been outlined in [38], except that we use $\mathrm{Cov}(Y) = \mathrm{diag}\{\bar{y}_i(\mu^{\mathrm{true}})\}$ in (23) of [38], where $\bar{y}_i$ is based on the overlapping beam model (6.34) rather than the parallel strip-integral model of [38].

We analyzed the resolution/variance tradeoffs for the ROI illustrated in Figure 6.1: it is a 2×9 region that goes across the boundary of the soft tissue, the lung, and the heart. We performed the analysis at two levels of transmitted counts, one with new sources (the high count case) and the other with old sources (the low count case), *i.e.*, the same as was done for Section 6.4. The background count level for both cases remains the same throughout. We only discuss the high count case; the low count case gives similar results. Figures 6.9 and 6.10 show the resolution/variance curves of the proposed algorithm and the parallel algorithm, respectively, for various collimation angles ranging from $1.6°$ to $6.6°$. Based on these two figures, we calculated (using cubic spline interpolation when needed) the variances of the ROI versus collimation angle at two fixed resolutions: 4.7 pixels and 6.8 pixels, as shown in Figures 6.11 and 6.12. The proposed algorithm outperforms the parallel algorithm which in turn outperforms FBP at both resolutions, and at all collimation angles. The performance gain of the proposed algorithm over the parallel algorithm is more impressive at larger collimation angles, *e.g.*, more than $2.5°$. Using the proposed algorithm, it seems that the optimal collimation angle for a resolution of 4.7 pixels is around $5°$, and $> 7°$ for a resolution of 6.8 pixels, both much larger than typical collimation angles found on SPECT cameras. However, if the suboptimal parallel algorithm were used for reconstruction, then the optimal collimation angle would be $2.6°$ for a resolution of 4.7 pixels, and $4.1°$ for a resolution of 6.8 pixels. Even though the optimal collimation angle for the parallel algorithm is closer to what is typically found on SPECT cameras than the proposed algorithm, the proposed algorithm outperforms the parallel algorithm at all collimation angles. Furthermore, for collimation angles larger than $3.5°$, a resolution requirement of 4.7

121

pixels would probably mean very high variances in the parallel reconstruction. Figure 6.13 shows the optimal collimation angle for the proposed and parallel algorithms at different desired spatial resolutions. Naturally, as the desired spatial resolution improves, the optimal collimation angle decreases. Figure 6.14 shows the minimum achievable normalized standard deviation for the the proposed and parallel algorithms at different desired spatial resolutions. As expected, the proposed algorithm consistently outperforms the parallel algorithm, resulting in reduction in the standard deviation by as much as 40%.
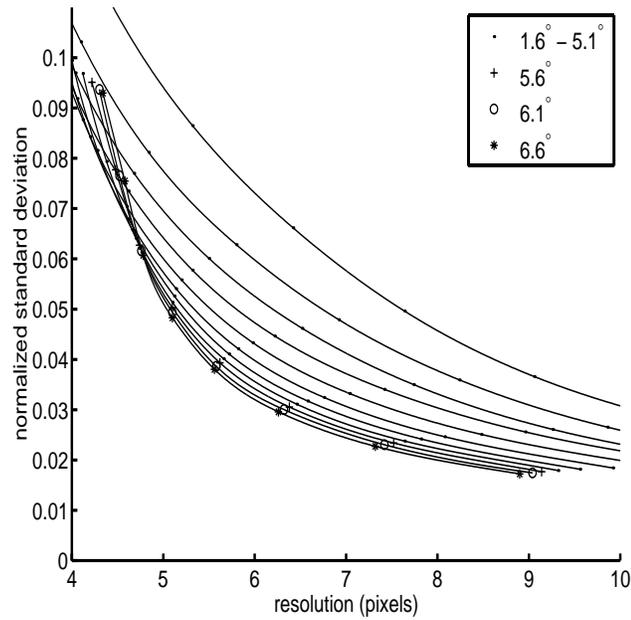
Figure 6.9: High count case: ROI resolution/variance curves for the proposed algorithm.
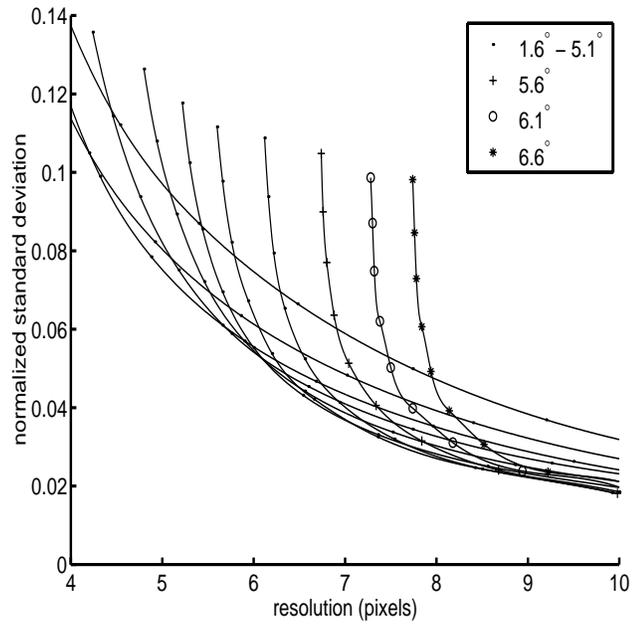
Figure 6.10: High count case: ROI resolution/variance curves for the parallel algorithm.
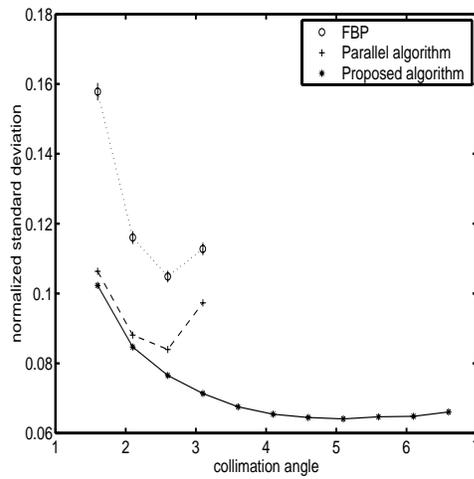
Figure 6.11: High count case: ROI variances at desired spatial resolution 4.7 pixels.
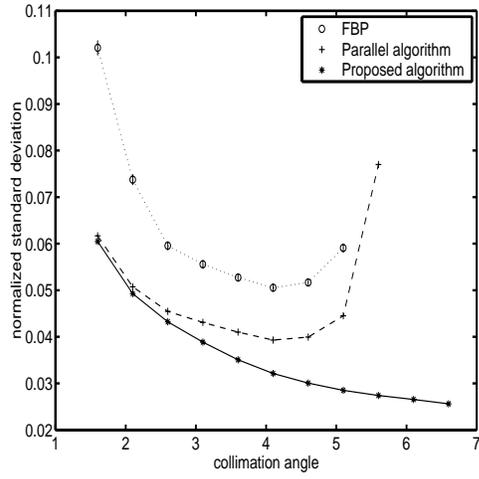
Figure 6.12: High count case: ROI variances at desired spatial resolution 6.8 pixels.
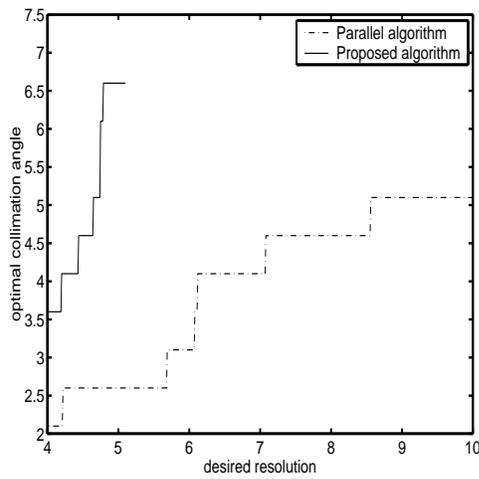
Figure 6.13: High count case: A comparison of optimal collimation angles, at different resolutions for the proposed and parallel algorithms.
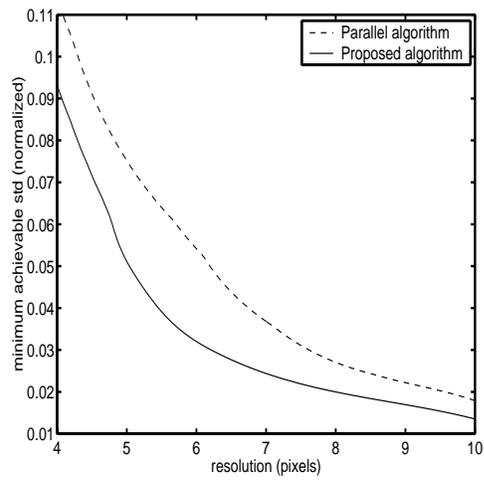
Figure 6.14: High count case: A comparison of minimum achievable normalized standard deviation, at different resolutions for the proposed and parallel algorithms.

6.6   Discussion and Conclusion

We have presented a new algorithm for statistical image reconstruction of attenuation maps that explicitly accounts for overlapping beams in transmission scans; an example of such a system can be found in [15]. The algorithm is guaranteed to monotonically increase the objective function at each iteration, and consistently achieves better variance/resolution tradeoffs than "conventional" image reconstruction algorithms, both statistical (the parallel algorithm) and non-statistical (FBP).

From our preliminary study on the optimal collimation angle, we find that it is desirable to open up the source collimators and allow beam overlap — provided the overlap is modeled appropriately in the reconstruction algorithm. (This conclusion applies to our ideal simulation settings where system geometry is known perfectly. Robustness of the proposed algorithm to model mismatch needs investigation.) However, detector collimation currently limits usable source collimation angles to $\geq 2.6°$, so alternatives to widening source collimation are needed to improve the counting statistics on this system.

The proposed algorithm is more time-consuming and uses more memory than conventional statistical algorithms. For our simulations, we used 14 system matrices (one for each source) with appropriate collimation angles. The system matrices with collimation angle $2.6°$ occupy 20 MBytes of disk space, and the system matrices with collimation angle $7°$ occupy 50 MBytes of disk space, compared to 8 MBytes occupied a single system matrix used by the parallel algorithm. It also takes longer to project or backproject an image in the proposed algorithm than the parallel algorithm; in fact, for collimation angle $2.6°$, each iteration of the proposed algorithm takes about 1.9 seconds, compared to about 0.8 seconds needed for the parallel algorithm, on a Sun Ultra2 workstation. Furthermore, because of the overlap between transmission beams, the convergence rate of the proposed algorithm is also slower than the parallel algorithm; as the overlap between transmission beams increases, *i.e.*, the collimation angle gets larger, the curvatures c_{im} as given in (6.16) get larger, hence the convergence rate becomes even slower. In fact, if one goes from a collimation angle of $1.6°$ to $5.6°$, the average curvature of the surrogate parabolas increases by a factor of 3. With regard to the number of iterations necessary for convergence, *i.e.*, the smallest

126

n such that $\Phi(\mu^n) - \Phi(\mu^0) > 0.999\left[\Phi(\mu^*) - \Phi(\mu^0)\right]$, where $\Phi(\mu^*)$ is the largest objective value obtained in 40 iterations, the parallel algorithm required 22 iterations and the proposed algorithm required 30 iterations for collimation angle $2.6°$ and a desired spatial resolution of 4.7 pixels. Hence the total amount time required by the proposed algorithm would be 3.3 times that of the parallel algorithm for collimation angle $2.6°$ and a desired spatial resolution of 4.7 pixels.

CHAPTER 7

Conclusion and Future Work

7.1 Edge-Preserving Tomographic Reconstruction With Nonlocal Regularization

In this part of our work, we presented a new regularization method for tomographic image reconstruction based on a nonlocal penalty function. The nonlocal penalty produces transmission reconstructions with better ROI bias/variance tradeoffs than a local Huber penalty. When these transmission reconstructions are applied to ideal emission data, the nonlocal penalty used for transmission reconstruction produces emission images with smaller variances (for a fixed spatial resolution) for most pixels in the image, *i.e.*, from 80% to 85%; the median standard deviation in the image is reduced by 35% to 50%.

The future work in this part should focus on the criterion upon which the h functions are selected, during both the intermediate and the final stage(s) of the deterministic annealing process. The final h function one uses controls the transition in pixel values between neighboring regions. Different h functions will result in very different resolutions in the reconstructed image, at least near the region boundaries. Knowledge of the resolutions of the reconstructed image is critical to the application of the attenuation coefficient maps to the emission reconstruction. Secondly, the selection of h functions during the intermediate stages of the annealing process will affect how fast the algorithm converges and also how "good" a local minimum the algorithm finds. Presumably, the slower the annealing pro-

cess, the "better" the local minimum reached by the algorithm. However, one must balance between the desire for a good local minimum and the practical concern of how long in time the algorithm will be run. A study of the what sort of h functions and how fast to evolve these functions will be useful if this algorithm is eventually to be used in clinical situations.

As we discussed in Section 3.3.1, the arc-length penalty term J_1 in (3.39) favors a shorter curve, and hence causes a bias in the boundary extraction. Even though our simulation showed this bias is negligible, a more theoretically satisfying way to resolve this issue is to replace the curve-length term with a penalty that does not favor a shorter curve, *e.g.*[96],

$$J_1(f, \Gamma_k) = \int_{\Gamma_k(p)} \frac{1}{2} (\mu_1 |(\Gamma_k)_p|^2 + \mu_2 |(\Gamma_k)_{pp}|^2) dp, \tag{7.1}$$

which can be minimized by steepest descent:

$$\frac{d\Gamma_k}{dt} = -\frac{\delta J_1}{\delta \Gamma} = -\mu_1 (\Gamma_k)_{pp} + \mu_2 (\Gamma_k)_{pppp}, \tag{7.2}$$

even though the implementation of the fourth derivative will be numerically problematic.

Finally, an analysis of the resolution/noise property of the proposed penalty, though difficult, will be very useful toward clinical application of the proposed penalty.

7.2 Counting Statistics Affected by Deadtime

In this part of our work, we made original contributions to the fundamental understanding of the random process, by proposing a new method of deriving its moments. However, this method, in practice, is only useful for the first and second moments; the derivation of higher moments turns out to be very time-consuming (even prohibitive once 4th or 5th moment is reached) for any reasonable complicated deadtime models. Nevertheless, we derived *rigorously*, the first and second moments of a number of counting processes, both single and coincidence; most of these expressions are hitherto unknown. Of course, we have just began to scratch the surface of this very difficult problem; a great deal more, including completely new methodologies, may be needed before we can achieve a good understanding of the nature of the process.

Most of the possible future work in the part involves applying the proposed method to more complicated deadtime models, *e.g.*, where the deadtime τ is stochastic instead of pre-determined, or the time-stamping of recorded singles has random errors. Furthermore, as we pointed out in Section 4.4, the effect of possibly-not-so-small correlation coefficient between the sum of one group of detectors and the sum of all other detectors in the block on the reconstructed images should be studied.

7.3 Maximum Likelihood Transmission Image Reconstruction for Overlapping Transmission Beams

In this final part of our work, we presented a new algorithm for statistical image reconstruction of attenuation maps that explicitly accounts for overlapping beams in transmission scans. The algorithm is guaranteed to monotonically increase the objective function at each iteration, and consistently achieves better variance/resolution tradeoffs than "conventional" image reconstruction algorithms, both statistical (the parallel algorithm) and non-statistical (FBP). We also found that it is desirable to open up the source collimators and allow beam overlap — provided the overlap is modeled appropriately in the reconstruction algorithm. However, the convergence speed of the proposed algorithm depends on the amount of the overlap between transmission beams – this effect should be studied. One should also study the robustness of the proposed algorithm to model mismatch. Because the proposed algorithm allows for wider source collimation (but detector collimation currently limits how wide the source collimators open up), a system design that involves transmission sources with energies high enough to penetrate the detector collimators should be studied. Finally, the collimation angle optimization in Section 6.5 are performed assuming quadratic penalties; a similar study on the optimal collimation angle should be done assuming the non-quadratic penalties, or even the non-local penalty proposed in Chapter 3.

This work can also be applicable to the problem of the sourceless attenuation correction[40], *i.e.*, completely avoid transmission scans and extract the attenuation map directly from the *noisy* emission projections. Because of the extreme ill-posedness of this problem, a strong

regularizing penalty, such as the non-local penalty proposed in Chapter 3, may be needed.

APPENDICES

APPENDIX A

Deadtime Analysis: Variance Derivation for Deadtime Model II

We derive the variance of $Y(t)$ for deadtime model II, the paralyzable model. We first derive $\beta(0, s)$. We consider two cases.

<u>CASE 1</u>: $0 < s < \tau$

We pick δ such that $0 < \delta < s < s + \delta < \tau$. Two recorded events cannot correspond to photons that arrived within τ of each other. Hence for $0 < s < \tau$, $E[Y(0, \delta)Y(s, s + \delta)] = 0$, and by the definition given in (4.6): $\beta(0, s) = 0$.

<u>CASE 2</u>: $\tau < s < t$

We pick δ such that $\delta < \tau$ and $s + \delta < t$ and $\delta < s - \tau$. For $s > \tau$, $Y(0, \delta)$ and $Y(s, s + \delta)$ are statistically independent, since the event "there is an arrival during $(0, \delta]$" is statistically independent from the event "there is an arrival during $(s, s + \delta]$", because they are at least τ apart in time[1]. Hence by (4.26),

$$E[Y(0, \delta)Y(s, s + \delta)] = E^2[Y(0, \delta)] = (\lambda \delta e^{-\lambda \tau})^2, \tag{A.1}$$

and

$$\beta(0, s) = (\lambda e^{-\lambda \tau})^2. \tag{A.2}$$

---

[1]If there is one arrival each during $(0, \delta]$, $(s/2, s/2 + \delta]$, and $(s, s + \delta]$, then $Y(0, \delta)Y(s, s + \delta) = 0$; but loss of the photon that arrived during $(s, s + \delta]$ is due to the arrival during $(s/2, s/2 + \delta]$; whether there is any arrival during $(0, \delta]$ is independent of whether the arrival during $(s, s + \delta]$ is recorded.

Combining the above two cases and using (4.25) yields

$$
\begin{aligned}
E[Y^2(t)] &= \gamma t + 2 \int_\tau^t (t-s)(\lambda e^{-\lambda \tau})^2 ds \\
&= \lambda t e^{-\lambda \tau} + [(t-\tau)(\lambda e^{-\lambda \tau})]^2.
\end{aligned}
\tag{A.3}
$$

Using $\mathrm{Var}[Y(t)] = E[Y^2(t)] - E^2[Y(t)]$, with (4.28) and (A.3), and simplifying yields (4.30).

APPENDIX B

Deadtime Analysis: Variance Derivation for Deadtime Model III

We derive the variance of $Y(t)$ for the type III deadtime model. Again, we first derive the correlation function $\beta(0, s)$. This derivation is more complicated than the type II model, due to the fact that if two photons arrive at times s_1 and s_2 respectively and $\tau < s_2 - s_1 < 2\tau$, then $(s_1 - \tau, s_1 + \tau) \cap (s_2 - \tau, s_2 + \tau) \neq \emptyset$ and $Y(s_1, s_1 + \delta)$ and $Y(s_2, s_2 + \delta)$ would both depend on what happens during $(s_2 - \tau, s_1 + \tau)$.

CASE 1: $0 < s < \tau$

We pick δ such that $0 < \delta < s < s + \delta < \tau$. Two recorded events cannot correspond to photons that arrived within τ of each other. Hence for $0 < s < \tau$, $E[Y(0, \delta)Y(s, s + \delta)] = 0$, and $\beta(0, s) = 0$.

CASE 2: $\tau < s < 2\tau$

We pick δ such that $s + \delta < 2\tau$ (hence $\delta < \tau$) and $\delta < s - \tau$. As discussed above, for $\tau < s < 2\tau$, $Y(0, \delta)$ and $Y(s, s + \delta)$ will be statistically dependent. If there is exactly one photon arrival each during $(0, \delta]$ and $(s, s + \delta]$ at time s_1 and s_2 respectively, then both events will be recorded if and only if there is no arrival during $(s_1 - \tau, s_1)$, (s_1, s_2), or $(s_2, s_2 + \tau]$ (since $\tau < s_2 - s_1 < 2\tau$, $(s_1, s_1 + \tau] \cup (s_2 - \tau, s_2) = (s_1, s_2)$.) Hence,

$$
E[Y(0, \delta)Y(s, s + \delta)]
$$
$$
= \ P[Y(0, \delta) = 1, Y(s, s + \delta) = 1]
$$
$$
= \ \int_0^\delta \int_0^\delta P[\text{no arrival during}(s_1 - \tau, 0), \text{ or } (s_1, s), \text{ or } (s_2, s_2 + \tau]]
$$

$$f_{T_1}(s_1)f_{T_1}(s_2)ds_1ds_2$$

$$= \int_0^\delta \int_0^\delta e^{-\lambda(\tau-s_1)} e^{-\lambda(s-s_1)} e^{-\lambda\tau} \lambda e^{-\lambda s_1} \lambda e^{-\lambda s_2} ds_1 ds_2$$

$$= e^{-\lambda 2\tau} \lambda^2 e^{-\lambda s} \int_0^\delta \int_0^\delta e^{\lambda s_1} e^{-\lambda s_2} ds_1 ds_2$$

$$= (e^{\lambda\delta} - 1)^2 e^{-\lambda(2\tau+s+\delta)}, \tag{B.1}$$

and

$$\beta(0, s) = \lambda^2 e^{-\lambda(2\tau+s)}. \tag{B.2}$$

<u>CASE 3: $2\tau < s < t$</u>

We pick δ such that $\delta < 2\tau$ and $s + \delta < t$ and $\delta < s - 2\tau$. For $2\tau < s < t$, $Y(0, \delta)$ and $Y(s, s+\delta)$ are statistically independent, since the event "there is an arrival during $(0, \delta]$" is statistically independent from the event "there is an arrival during $(s, s + \delta]$", because they are at least 2τ apart in time. Thus

$$E[Y(0, \delta)Y(s, s+\delta)] = E^2[Y(0, \delta)] = (\lambda\delta e^{-\lambda 2\tau})^2, \tag{B.3}$$

and

$$\beta(0, s) = (\lambda e^{-\lambda 2\tau})^2. \tag{B.4}$$

Combining the above three cases and using (4.25) yields

$$\begin{aligned} E[Y^2(t)] &= \gamma t + 2 \int_\tau^{2\tau} (t - s)\lambda^2 e^{-\lambda(2\tau+s)} ds + 2 \int_{2\tau}^t (t - s)(\lambda e^{-\lambda 2\tau})^2 ds \\ &= \lambda t e^{-\lambda 2\tau} + 2e^{-4\lambda\tau}(1 - \lambda t + 2\lambda\tau) + 2e^{-3\lambda\tau}(\lambda t - \lambda\tau - 1) \\ &\quad + [(t - 2\tau)(\lambda e^{-\lambda 2\tau})]^2. \end{aligned} \tag{B.5}$$

Simple algebra leads to (4.37).

APPENDIX C

Deadtime Analysis: Mean Derivation for Inhomogeneous Arrival Process

Due to the decay of an isotope photon source, the photon arrival process is not exactly homogeneous. In medical imaging, the arrival rates are inhomogeneous due to radiotracer dynamics. In this section, we derive $E[Y(t)]$ for paralyzable deadtime model[1], assuming only that the instantaneous photon arrival rate $\lambda(t)$ is continuous. This relaxes the assumption made in Section 4.2 that λ is constant. For an inhomogeneous process, $E[Y(s, s + \delta)] \neq E[Y(0, \delta)]$ in general. First we observe that the waiting time for the first photon arrival after time s, denoted T_1, has the following distribution:

$$F_{T_1}(r) = \mathrm{P}[T_1 \leq r] = 1 - \mathrm{P}[T_1 > r] = 1 - \mathrm{P}[N(s, r) = 0]$$

$$= 1 - e^{-\int_s^r \lambda(q)dq}. \tag{C.1}$$

Hence for $r > s$,

$$f_{T_1}(r) = \frac{d}{dr} F_{T_1}(r) = \lambda(r)e^{-\int_s^r \lambda(q)dq}. \tag{C.2}$$

For $0 < \delta < \tau$, we have:

$$
\begin{aligned}
E[Y(s, s + \delta)] &= \mathrm{P}[Y(s, s + \delta) = 1] \\
&= \int_s^{s+\delta} \mathrm{P}[Y(s, s + \delta) = 1 | T_1 = r] f_{T_1}(r)dr \\
&= \int_s^{s+\delta} \mathrm{P}[N(r - \tau, s) = 0] f_{T_1}(r)dr
\end{aligned}
$$

---

[1]Extension to the type III deadtime model is straightforward.

$$= \int_s^{s+\delta} e^{-\int_{r-\tau}^s \lambda(q_1)dq_1} \lambda(r) e^{-\int_s^r \lambda(q_2)dq_2} dr$$

$$= \int_s^{s+\delta} \lambda(r) e^{-\int_{r-\tau}^r \lambda(q)dq} dr. \tag{C.3}$$

Since λ is continuous, and $e^{-\int_{r-\tau}^r \lambda(q)dq}$ is continuous in r, we conclude:

$$\gamma(s) = \lambda(s) e^{-\int_{s-\tau}^s \lambda(q)dq}. \tag{C.4}$$

Hence[2]

$$E[Y(t)] = \int_0^t \lambda(s) e^{-\int_{s-\tau}^s \lambda(q)dq} ds. \tag{C.5}$$

If τ is small relative to variations in λ, then $\int_{s-\tau}^s \lambda(q)dq \approx \lambda(s)\tau$, so

$$E[Y(t)] \approx \int_0^t \lambda(s) e^{-\lambda(s)\tau} ds. \tag{C.6}$$

This approximation can be applied to other deadtime models as well. Similarly, the second moment of Y is:

$$E[Y^2(t)] = E[Y(t)] + 2 \int_0^t \int_{s_1+\tau}^t \gamma(s_1)\gamma(s_2) ds_2 ds_1. \tag{C.7}$$

---

[2]In fact, it is unnecessarily restrictive to limit λ to continuous functions; all we need is that λ is integrable and bounded over $[0, t]$. If λ is integrable on $[0, t]$, then $\lambda_1(r) = \lambda(r) e^{-\int_{r-\tau}^r \lambda(q)dq}$ is integrable on $[0, t]$; then almost every point of $[0, t]$ is a Lebesgue point of λ_1 [10, Theorem 7.40]; and if s is a Lebesgue point of λ_1, then $\lim_{\delta \to 0} \frac{\int_s^{s+\delta} \lambda_1(q)dq}{\delta} = \lambda_1(s)$ [10, Theorem 7.39].

APPENDIX D

Deadtime Analysis: Partial Derivation of Variance for the Coincidence Process with Ideal Detectors

We derive $\beta(0, s)$ (for $0 < s < \epsilon$) of the random coincidence process with ideal detectors. We have

$$
\begin{aligned}
E[Y(0, \delta)Y(s, s + \delta)] &= E[(Y_a(0, \delta) + Y_b(0, \delta))(Y_a(s, s + \delta) + Y_b(s, s + \delta))] \\
&= E[Y_a(0, \delta)Y_a(s, s + \delta)] + E[Y_a(0, \delta)Y_b(s, s + \delta)] \\
&\quad + E[Y_b(0, \delta)Y_a(s, s + \delta)] + E[Y_b(0, \delta)Y_b(s, s + \delta)]. \text{ (D.1)}
\end{aligned}
$$

We pick δ such that

$$
0 < \delta < s < s + \delta < \epsilon, \tag{D.2}
$$

hence $\forall s_1 \in (0, \delta], \forall s_2 \in (s, s + \delta], 0 < s_2 - s_1 < \epsilon$. We derive $E[Y_a(0, \delta)Y_a(s, s + \delta)]$:

$$
\begin{aligned}
&E[Y_a(0, \delta)Y_a(s, s + \delta)|X_1(0, \delta) = 1, X_1(s, s + \delta) = 1] \\
&= \int_0^\delta \int_s^{s+\delta} E[Y_a(0, \delta)Y_a(s, s + \delta)|X_1(0, \delta) = 1, X_1(s, s + \delta) = 1, \\
&\quad T_{1,1} = s_1, T_{1,2} = s_2]/\delta^2 ds_2 ds_1, \tag{D.3}
\end{aligned}
$$

where $T_{1,1}$ and $T_{1,2}$ denote the time of the recorded singles events at detector 1 during $(0, \delta]$ and $(s, s + \delta]$, respectively. For $s_1 \in (0, \delta]$ and $s_2 \in (s, s + \delta]$, since $s_1 > s_2 - \epsilon$, we have (see Figure D.1):

$$
E[Y_a(0, \delta)Y_a(s, s + \delta)|X_1(0, \delta) = 1, X_1(s, s + \delta) = 1, T_{1,1} = s_1, T_{1,2} = s_2]
$$

$$
\begin{aligned}
&= E[X_2(s_1 - \epsilon, s_1)X_2(s_2 - \epsilon, s_2)] \\
&= E[(X_2(s_1 - \epsilon, s_2 - \epsilon) + X_2(s_2 - \epsilon, s_1))(X_2(s_2 - \epsilon, s_1) + X_2(s_1, s_2))] \\
&= E[X_2(s_1 - \epsilon, s_2 - \epsilon)X_2(s_2 - \epsilon, s_1) + X_2(s_1 - \epsilon, s_2 - \epsilon)X_2(s_1, s_2) \\
&\quad + X_2^2(s_2 - \epsilon, s_1) + x_(s_2 - \epsilon, s_1)X_2(s_1, s_2)] \\
&= \gamma_2^2(s_2 - s_1)(\epsilon - s_2 + s_1) + \gamma_2^2(s_2 - s_1)^2 \\
&\quad + (\gamma_2^2(\epsilon - s_2 + s_1)^2 + \gamma_2(\epsilon - s_2 + s_1)) + \gamma_2^2(\epsilon - s_2 + s_1)(s_2 - s_1) \\
&= (\gamma_2\epsilon)^2 + \gamma_2(\epsilon - (s_2 - s_1)).
\end{aligned}
\tag{D.4}
$$
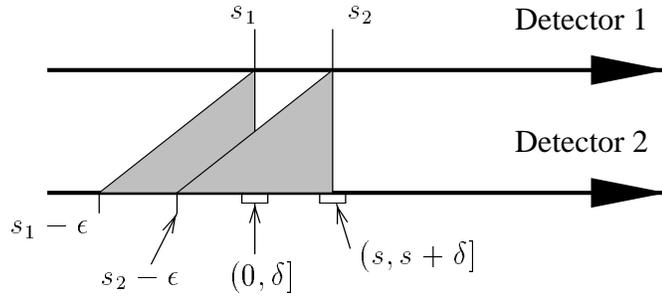
Figure D.1: Graphic aid for $E[Y_a(0,\delta)Y_a(s, s+\delta)|X_1(0,\delta) = 1, X_1(s, s+\delta) = 1, T_{1,1} = s_1, T_{1,2} = s_2]$

Integrating (D.3) yields:

$$
E[Y_a(0,\delta)Y_a(s, s+\delta)|X_1(0,\delta) = 1, X_1(s, s+\delta) = 1] = (\gamma_2\epsilon)^2 + \gamma_2(\epsilon - s). \tag{D.5}
$$

By total probability:

$$
\begin{aligned}
E[Y_a(0,\delta)Y_a(s, s+\delta)] &= \sum_{k,l=0}^{\infty} E[Y_a(0,\delta)Y_a(s, s+\delta)|X_1(0,\delta) = k, X_1(s, s+\delta) = l] \\
&\quad \times \mathrm{P}[X_1(0,\delta) = k, X_1(s, s+\delta) = l].
\end{aligned}
\tag{D.6}
$$

When $k + l \geq 3$,

$$
E[Y_a(0,\delta)Y_a(s, s+\delta)|X_1(0,\delta) = k, X_1(s, s+\delta) = l] \leq \gamma_2^2(s + \delta + \epsilon)^2 kl, \tag{D.7}
$$

hence,

$$
\sum_{k,l=0,k+l\geq 3}^{\infty} E[Y_a(0,\delta)Y_a(s, s+\delta)|X_1(0,\delta) = k, X_1(s, s+\delta) = l]
$$

$$\times P[X_1(0,\delta) = k, X_1(s, s+\delta) = l]$$

$$
\begin{aligned}
= \quad & E[Y_a(0,\delta)Y_a(s,s+\delta)|X_1(0,\delta)=2, X_1(s,s+\delta)=1]P[X_1(0,\delta)=2, X_1(s,s+\delta)=1] \\
& + \sum_{k=1}^{\infty}\sum_{l=2}^{\infty} E[Y_a(0,\delta)Y_a(s,s+\delta)|X_1(0,\delta)=k, X_1(s,s+\delta)=l] \\
& \times P[X_1(0,\delta)=k, X_1(s,s+\delta)=l] \\
\leq \quad & 2\gamma_2^2(s+\delta+\epsilon)^2 P[X_1(0,\delta)=2]P[X_1(s,s+\delta)=1] \\
& + \sum_{k=1}^{\infty}\sum_{l=2}^{\infty} \gamma_2^2(s+\delta+\epsilon)^2 kl P[X_1(s,s+\delta)=l].
\end{aligned}
$$

Since both $\sum_{k=1}^{\infty}\gamma_2(s+\delta+\epsilon)kP[X_1(0,\delta)=k]$ and $\sum_{l=2}^{\infty}\gamma_2(s+\delta+\epsilon)lP[X_1(s,s+\delta)=l]$ converge absolutely, by Merten's Theorem [62],

$$
\begin{aligned}
& \sum_{k,l=0,k+l\geq 3}^{\infty} E[Y_a(0,\delta)Y_a(s,s+\delta)|X_1(0,\delta)=k, X_1(s,s+\delta)=l] \\
& \times P[X_1(0,\delta)=k, X_1(s,s+\delta)=l] \\
\leq \quad & 2\gamma_2^2(s+\delta+\epsilon)^2 P[X_1(0,\delta)=2]P[X_1(s,s+\delta)=1] + \left(\sum_{k=1}^{\infty}\gamma_2(s+\delta+\epsilon)k \right. \\
& \times \left. P[X_1(0,\delta)=k]\right)\left(\sum_{l=2}^{\infty}\gamma_2(s+\delta+\epsilon)lP[X_1(s,s+\delta)=l]\right).
\end{aligned}
\tag{D.8}
$$

Applying LDCT to (D.8), we have by similar argument as (5.3):

$$
\begin{aligned}
\lim_{\delta\to 0}\frac{1}{\delta^2}\sum_{k,l=0,k+l\geq 3}^{\infty} & E[Y_a(0,\delta)Y_a(s,s+\delta)|X_1(0,\delta)=k, X_1(s,s+\delta)=l] \\
& \times P[X_1(0,\delta)=k, X_1(s,s+\delta)=l] = 0.
\end{aligned}
\tag{D.9}
$$

Hence by (D.5) and (D.6), for $0 < s < \epsilon$,

$$
\begin{aligned}
\lim_{\delta\to 0}\frac{E[Y_a(0,\delta)Y_a(s,s+\delta)]}{\delta^2} & = \lim_{\delta\to 0}\frac{[(\gamma_2\epsilon)^2 + \gamma_2(\epsilon-s)]P[X_1(s,s+\delta)=X_1(0,\delta)=1]}{\delta^2} + 0 \\
& = \lim_{\delta\to 0}\frac{[(\gamma_2\epsilon)^2 + \gamma_2(\epsilon-s)][e^{-\delta\gamma_1}(\delta\gamma_1)]^2}{\delta^2} \\
& = (\epsilon\gamma_1\gamma_2)^2 + \gamma_1^2\gamma_2(\epsilon-s).
\end{aligned}
\tag{D.10}
$$

Now we derive $E[Y_b(0,\delta)Y_a(s,s+\delta)]$ by similar methods. Let $T_{2,1}$ and $T_{1,2}$ denote the time of the recorded singles events by detector 2 during $(0,\delta]$ and by detector 1 during $(s, s+\delta]$, respectively. We have (see Figure D.2):

$$E[Y_b(0,\delta)Y_a(s,s+\delta)|X_2(0,\delta)=1, X_1(s,s+\delta)=1, T_{2,1}=s_1, T_{1,2}=s_2]$$

$s_1 - \epsilon$

s_1

s_2

Detector 1

Detector 2

$s_2 - \epsilon$

$(0, \delta]$

$(s, s + \delta]$
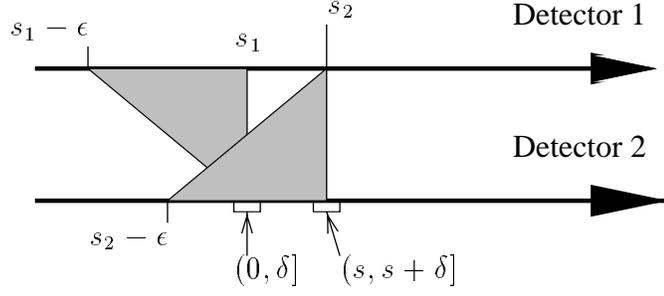
Figure D.2: Graphic aid for $E[Y_b(0, \delta)Y_a(s, s + \delta)|X_1(0, \delta) = 1, X_1(s, s + \delta) = 1, T_{2,1} = s_1, T_{1,2} = s_2]$

$$
\begin{aligned}
&= E[X_1(s_1 - \epsilon, s_1)X_2(s_2 - \epsilon, s_2)|X_2(0, \delta) = 1, X_1(s, s + \delta) = 1, T_{2,1} = s_1, T_{1,2} = s_2] \\
&= E[X_1(s_1 - \epsilon, s_1)|X_1(s, s + \delta) = 1, T_{1,2} = s_2] \cdot E[X_2(s_2 - \epsilon, s_2)|X_2(0, \delta) = 1, T_{2,1} = s_1] \\
&= (\gamma_1 \epsilon)E[X_2(s_2 - \epsilon, 0) + 1 + X_2(\delta, s_2)] \\
&= (\gamma_1 \epsilon)(1 + \gamma_2(\epsilon - \delta)). \quad (D.11)
\end{aligned}
$$

Since this is independent of s_1 and s_2, the integral as in (D.3) is identical to (D.11). Furthermore, since

$$
\begin{aligned}
\sum_{k,l=0,k+l\geq 3}^{\infty} &E[Y_b(0, \delta)Y_a(s, s + \delta)|X_1(0, \delta) = k, X_1(s, s + \delta) = l] \\
&\times P[X_1(0, \delta) = k, X_1(s, s + \delta) = l] \\
\leq \sum_{k,l=0,k+l\geq 3}^{\infty} &(k + \gamma_1(s + \delta + \epsilon))^2 kl P[X_1(0, \delta) = k, X_1(s, s + \delta) = l], \quad (D.12)
\end{aligned}
$$

by a similar argument made in (D.8), we have

$$
\begin{aligned}
\lim_{\delta \to 0} \frac{1}{\delta^2} \sum_{k,l=0,k+l\geq 3}^{\infty} &E[Y_b(0, \delta)Y_a(s, s + \delta)|X_1(0, \delta) = k, X_1(s, s + \delta) = l] \\
&\times P[X_1(0, \delta) = k, X_1(s, s + \delta) = l] \\
= 0. \quad (D.13)
\end{aligned}
$$

Thus following (D.10),

$$
\lim_{\delta \to 0} \frac{E[Y_b(0, \delta)Y_a(s, s + \delta)]}{\delta^2} = (\epsilon\gamma_1\gamma_2)^2 + \gamma_1^2\gamma_2\epsilon. \quad (D.14)
$$

By symmetry, from (D.10) and (D.14):

$$
\lim_{\delta \to 0} \frac{E[Y_b(0, \delta)Y_b(s, s + \delta)]}{\delta^2} = (\epsilon\gamma_1\gamma_2)^2 + \gamma_2^2\gamma_1(\epsilon - s) \quad (D.15)
$$

142

$$\lim_{\delta \to 0} \frac{E[Y_a(0, \delta)Y_b(s, s + \delta)]}{\delta^2} = (\epsilon\gamma_1\gamma_2)^2 + \gamma_2^2\gamma_1\epsilon. \tag{D.16}$$

Hence combining (D.10), (D.14), (D.15), and (D.16), we obtain for $0 < s < \epsilon$,

$$\beta(0, s) = (2\epsilon\gamma_1\gamma_2)^2 + (\gamma_1 + \gamma_2)\gamma_1\gamma_2(2\epsilon - s). \tag{D.17}$$

APPENDIX E

Deadtime Analysis: Partial Derivation of Variance for the Coincidence Process with Non-ideal Detectors

We derive $\beta(0, s)$ (for $\tau < s < \tau + \epsilon$) of the random coincidence process with non-ideal detectors. We pick $\delta < \min(s, \tau + \epsilon - s, s - \tau)$, which respectively ensures that $(0, \delta] \cap (s, s + \delta] = \emptyset$, $s_2 - s_1 < \tau + \epsilon$, and $s_2 - s_1 > \tau$, $\forall s_1 \in (0, \delta]$, $\forall s_2 \in (s, s + \delta]$. Following (D.1) and (D.3),

$$E[Y_a(0, \delta)Y_a(s, s + \delta)|Y_1(0, \delta) = 1, Y_1(s, s + \delta) = 1]$$
$$= \int_0^\delta \int_s^{s+\delta} E[Y_a(0, \delta)Y_a(s, s + \delta)|Y_1(0, \delta) = 1, Y_1(s, s + \delta) = 1,$$
$$T_{1,1} = s_1, T_{1,2} = s_2]/\delta^2 ds_2 ds_1,$$

where $T_{1,1}$ and $T_{1,2}$ are defined as in (D.3). We define $T_{2,1}$ to be the time of the first photon arrival after $s_1 - \epsilon$, and $T_{2,2}$ to be the time of the first photon arrival after $s_2 - \epsilon$, both at detector 2. (If $T_{2,1} > s_2 - \epsilon$, then $T_{2,1} = T_{2,2}$.) We have (see Figure E.1):

$$E[Y_a(0, \delta)Y_a(s, s + \delta)|Y_1(0, \delta) = 1, Y_1(s, s + \delta) = 1, T_{1,1} = s_1, T_{1,2} = s_2]$$
$$= \text{P}[Y_2(s_1 - \epsilon, s_1) = 1, Y_2(s_2 - \epsilon, s_2) = 1]$$
$$= \int_{s_1-\epsilon}^{s_1} \int_{s_2-\epsilon}^{s_2} \text{P}[Y_2(s_1 - \epsilon, s_1) = 1, Y_2(s_2 - \epsilon, s_2) = 1|T_{2,1} = s_3, T_{2,2} = s_4]$$
$$\times f_{T_{2,2}}(s_4) f_{T_{2,1}}(s_3) ds_4 ds_3$$
$$= \int_{s_1-\epsilon}^{s_1} \int_{s_2-\epsilon}^{s_2} \text{P}[N_2(s_3 - \tau, s_1 - \epsilon) = 0, N_2(s_4 - \tau, s_2 - \epsilon) = 0|T_{2,1} = s_3, T_{2,2} = s_4]$$
$$\times f_{T_{2,2}}(s_4) f_{T_{2,1}}(s_3) ds_4 ds_3$$

$$= \int_{s_1-\epsilon}^{s_2-\epsilon-\tau} \int_{s_2-\epsilon}^{s_2} P[N_2(s_3-\tau, s_1-\epsilon)=0, N_2(s_4-\tau, s_2-\epsilon)=0|T_{2,1}=s_3, T_{2,2}=s_4]$$

$$\times f_{T_{2,2}}(s_4) f_{T_{2,1}}(s_3) ds_4 ds_3 + \int_{s_2-\epsilon-\tau}^{s_1} \int_{s_2-\epsilon}^{s_2} P[N_2(s_3-\tau, s_1-\epsilon)=0,$$

$$N_2(s_4-\tau, s_2-\epsilon)=0|T_{2,1}=s_3, T_{2,2}=s_4] f_{T_{2,2}}(s_4) f_{T_{2,1}}(s_3) ds_4 ds_3$$

$$= \int_{s_1-\epsilon}^{s_2-\epsilon-\tau} \int_{s_2-\epsilon}^{s_2} P[N_2(s_3-\tau, s_1-\epsilon)=0] P[N_2(s_4-\tau, s_2-\epsilon)=0]$$

$$\times f_{T_{2,2}}(s_4) f_{T_{2,1}}(s_3) ds_4 ds_3 + \int_{s_2-\epsilon-\tau}^{s_1} \int_{s_3+\tau}^{s_2} P[N_2(s_3-\tau, s_1-\epsilon)=0]$$

$$\times P[N_2(s_4-\tau, s_2-\epsilon)=0] f_{T_{2,2}}(s_4) f_{T_{2,1}}(s_3) ds_4 ds_3$$

$$= \int_{s_1-\epsilon}^{s_2-\epsilon-\tau} \int_{s_2-\epsilon}^{s_2} e^{-\lambda_2(s_1-\epsilon-s_3+\tau)} e^{-\lambda_2(s_2-\epsilon-s_4+\tau)} \lambda_2 e^{-\lambda_2(s_3-s_1+\epsilon)} \lambda_2 e^{-\lambda_2(s_4-s_2+\epsilon)} ds_4 ds_3$$

$$+ \int_{s_2-\epsilon-\tau}^{s_1} \int_{s_3+\tau}^{s_2} e^{-\lambda_2(s_1-\epsilon-s_3+\tau)} e^{-\lambda_2(s_2-\epsilon-s_4+\tau)} \lambda_2 e^{-\lambda_2(s_3-s_1+\epsilon)} \lambda_2 e^{-\lambda_2(s_4-s_2+\epsilon)} ds_4 ds_3$$

$$= \int_{s_1-\epsilon}^{s_2-\epsilon-\tau} \int_{s_2-\epsilon}^{s_2} \lambda_2^2 e^{-2\lambda_2\tau} ds_4 ds_3 + \int_{s_2-\epsilon-\tau}^{s_1} \int_{s_3+\tau}^{s_2} \lambda_2^2 e^{-2\lambda_2\tau} ds_4 ds_3$$

$$= \lambda_2^2 e^{-2\lambda_2\tau} \frac{\epsilon^2 + (s_2-s_1-\tau)(2\epsilon-(s_2-s_1-\tau))}{2}. \tag{E.1}$$

Hence, by similarly tedious calculations akin to those for (D.10):

$$\lim_{\delta\to 0} \frac{E[Y_a(0,\delta) Y_a(s,s+\delta)]}{\delta^2} = \lambda_1^2 \lambda_2^2 e^{-2(\lambda_1+\lambda_2)\tau} \frac{\epsilon^2 + (s-\tau)(2\epsilon-(s-\tau))}{2}. \tag{E.2}$$
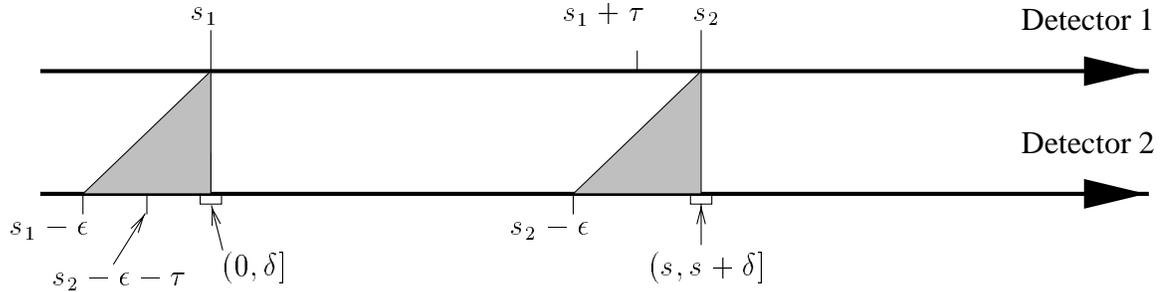
Figure E.1: Graphic aid for $E[Y_a(0,\delta) Y_a(s,s+\delta)|Y_1(0,\delta)=1, Y_1(s,s+\delta)=1, T_{1,1}=s_1, T_{1,2}=s_2]$

And for $E[Y_b(0,\delta) Y_a(s,s+\delta)]$ (see Figure E.2):

$$E[Y_b(0,\delta) Y_a(s,s+\delta)|Y_2(0,\delta)=1, Y_1(s,s+\delta)=1, T_{2,1}=s_1, T_{1,2}=s_2]$$

$$= E[Y_b(0,\delta)|Y_2(0,\delta)=1, T_{2,1}=s_1] \cdot E[Y_a(s,s+\delta)|Y_2(0,\delta)=1,$$

$$Y_1(s,s+\delta)=1, T_{2,1}=s_1, T_{1,2}=s_2]$$

$$= \lambda_1 \epsilon e^{-\lambda_1 \tau} \cdot \int_{s_1+\tau}^{s_2} \mathrm{P}[N_2(s_3 - \tau, s_1 + \tau) = 0 | T_{2,2} = s_3] f_{T_{2,2}}(s_3) ds_3$$

$$= \lambda_1 \epsilon e^{-\lambda_1 \tau} \cdot \int_{s_1+\tau}^{s_2} \lambda_2 e^{-\lambda_2 \tau} ds_3$$

$$= \epsilon(s_2 - s_1 - \tau)\lambda_1 \lambda_2 e^{-(\lambda_1 + \lambda_2)\tau}. \tag{E.3}$$

Hence by similar calculations made for (D.10) and (E.2):

$$\lim_{\delta \to 0} \frac{E[Y_b(0, \delta)Y_a(s, s + \delta)]}{\delta^2} = \epsilon(s - \tau)\lambda_1^2 \lambda_2^2 e^{-(\lambda_1 + \lambda_2)\tau}. \tag{E.4}$$
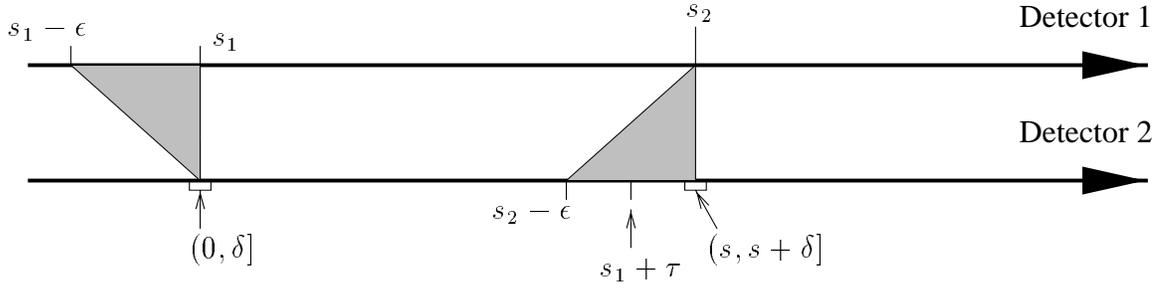
Figure E.2: Graphic aid for $E[Y_b(0, \delta)Y_a(s, s + \delta) | Y_2(0, \delta) = 1, Y_1(s, s + \delta) = 1, T_{2,1} = s_1, T_{1,2} = s_2]$

Using symmetry and combining (E.2) and (E.4), we obtain when $\tau < s < t + \epsilon$,

$$\beta(0, s) = \lambda_1^2 \lambda_2^2 e^{-(\lambda_1 + \lambda_2)\tau} (\epsilon^2 + (s - \tau)(4\epsilon - s + \tau)). \tag{E.5}$$

146

BIBLIOGRAPHY

[1] D. Adalsteinsson and J. Sethian, "A fast level set method for propagating interfaces," *Journal of Computational Physics*, vol. 118, no. 2, pp. 269–277, May 1995.

[2] R. Adams and L. Bischof, "Seeded region growing," *IEEE Tr. Patt. Anal. Mach. Int.*, vol. 16, , 1994.

[3] H. H. Barrett and W. Swindell, *Radiological imaging: the theory of image formation, detection, and processing*, Academic, New York, 1981.

[4] M. P. Becker, I. Yang, and K. Lange, "EM algorithms without missing data," *Stat. Meth. Med. Res.*, vol. 6, no. 1, pp. 38–54, 1997.

[5] M. Bertero, C. De Mol, and E. R. Pike, "Linear inverse problems with discrete data, i: General formulation and singular system analysis," *Inverse Prob.*, vol. 1, no. 4, pp. 301–330, November 1985.

[6] M. Bertero, T. Poggio, and V. Torre, "Ill posed problems in early vision," *Proc. IEEE*, vol. 76, no. 8, pp. 869–889, August 1988.

[7] A. Blake and A. Zisserman, *Visual reconstruction*, MIT Press, Cambridge, MA, 1987.

[8] J. E. Bowsher, V. E. Johnson, T. G. Turkington, R. J. Jaszczak, C. E. Floyd, and R. E. Coleman, "Bayesian reconstruction and use of anatomical a priori information for emission tomography," *IEEE Tr. Med. Im.*, vol. 15, no. 5, pp. 673–86, October 1996.

[9] J. E. Bowsher, M. P. Tornai, J. Peter, D. E. Gonzalez Trotter, A. Krol, D. R. Gilland, and R. J. Jaszczak. *Multiple transmission sources and modeling the axial resolution of a line source*, 1999. Preprint.

[10] A. M. Bruckner, J. B. Bruckner, and B. S. Thomson, *Real Analysis*, Prentice Hall, 1997.

[11] J. Canny, "A computational approach to edge detection," *IEEE Tr. Patt. Anal. Mach. Int.*, vol. 8, no. 6, pp. 679–698, November 1986.

[12] F. Carloni, A. cCorberi, M. Marseguerra, and C. M. Porceddu, "The asympototic method of dead-time correction in Poissonian distribution," *Nucl. Instr. Meth.*, vol. 78, pp. 70–6, 1970.

[13] F. Carloni, A. Corberi, M. Marseguerra, and C. M. Porceddu, "Asymptotic method of dead-time correction in Poissonian distribution," *Nucl Instrum Methods*, vol. 78, no. 1, pp. 70–76, February 1970.

[14] A. Celler, A. Sitek, and R. Harrop, "Reconstruction of multiple line source attenuation maps," *IEEE Tr. Nuc. Sci.*, vol. 44, no. 4, pp. 1503–8, August 1997.

[15] A. Celler, A. Sitek, E. Stoub, P. Hawman, R. Harrop, and D. Lyster, "Multiple line source array for SPECT transmission scans: Simulation, phantom and patient studies," *J. Nuc. Med.*, vol. 39, no. 12, pp. 2183–9, December 1998.

[16] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Tr. Im. Proc.*, vol. 6, no. 2, , February 1997.

[17] 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.

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

[19] N. H. Clinthorne, N. A. Petrick, W. L. Rogers, and A. O. Hero, "A fundamental limit on timing performance with scintillation detectors," *IEEE Tr. Nuc. Sci.*, vol. 37, no. 2, pp. 658–63, April 1990.

[20] L. D. Cohen and I. Cohen, "Finite-element methodws for active contour models and balloons for 2-d and 3-d images," *IEEE Tr. Patt. Anal. Mach. Int.*, vol. 15, no. 11, pp. 1131–47, November 1993.

[21] D. R. Cox, *Renewal Theory*, Methuen and Co., 1962.

[22] M. E. Daube-Witherspoon and R. E. Carson, "Unified deadtime correction model for PET," *IEEE Tr. Med. Im.*, vol. 10, no. 2, pp. 267–275, September 1991.

[23] 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.

[24] E. Debreuve, M. Barlaud, G. Aubert, and J. Darcourt, "Attenuation map segmentation without reconstruction using a level set method in nuclear medicine imaging," in *Proc. IEEE Intl. Conf. on Image Processing*, volume 1, pp. 34–38, 1998.

[25] A. H. Delaney and Y. Bresler, "Globally convergent edge-preserving regularized reconstruction: an application to limited-angle tomography," *IEEE Tr. Im. Proc.*, vol. 7, no. 2, pp. 204–221, February 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] D. Dobson and O. Scherzer, "Analysis of regularized total variation penalty methods for denoising," *Inverse Prob.*, vol. 12, no. 5, pp. 601–17, October 1996.

[28] D. C. Dobson and C. R. Vogel, "Convergence of an iterative method for total variation denoising," *SIAM J. Numer. Anal.*, vol. 34, no. 5, pp. 1779–91, October 1997.

[29] U. Engeland, T. Striker, and H. Luig, "Count-rate statistics of the gamma camera," *Phys. Med. Biol.*, vol. 43, no. 10, pp. 2939–47, October 1998.

[30] 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.

[31] H. Erdoğan and J. A. Fessler, "Monotonic algorithms for transmission tomography," *IEEE Tr. Med. Im.*, vol. 18, no. 9, pp. 801–14, September 1999.

[32] H. Erdoğan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Phys. Med. Biol.*, vol. 44, no. 11, pp. 2835–51, November 1999.

[33] 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.

[34] G. Faraci and A. R. Pennisi, "Experimental dead-time distortions of Poisson processes," *Nucl. Instr. Meth.*, vol. 212, no. 1, pp. 307–310, July 1983.

[35] W. Feller, *An Introduction to Probability Theory and Its Applications*, John Wiley and Sons, 1968.

[36] J. A. Fessler, "Segmented attenuation correction for PET using ICM," in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 2, pp. 1182–1184, 1992.

[37] 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.

[38] 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.

[39] 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.

[40] J. A. Fessler, "Maximum likelihood sourceless attenuation correction for spect," *Draft*, 1999.

[41] J. A. Fessler and S. D. Booth, "Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction," *IEEE Tr. Im. Proc.*, vol. 8, no. 5, pp. 688–99, May 1999.

[42] E. P. Ficaro, J. A. Fessler, J. N. Kritzman, P. C. Hawman, S. W. DeBruin, and J. R. Corbett, "Multiple line source array TCT/ECT system: A technical phantom evaluation," *J. Nuc. Med. (Abs. Book)*, vol. 40, no. 5, pp. 4, May 1999.

[43] 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.

[44] I. M. Gelfand and S. V. Fomin, *Calculus of variations*, Prentice-Hall, NJ, 1963. Translation by R A Silverman.

[45] D. Geman, S. Geman, C. Graffigne, and P. Dong, "Boundary detection by constrained optimization," *IEEE Tr. Patt. Anal. Mach. Int.*, vol. 12, no. 7, pp. 609–628, July 1990.

[46] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and Bayesian restoration of images," *IEEE Tr. Patt. Anal. Mach. Int.*, vol. 6, no. 6, pp. 721–741, November 1984.

[47] G. Gindi, A. Rangarajan, M. Lee, P. J. Hong, and I. G. Zubal, "Bayesian reconstruction for emission tomography via deterministic annealing," in *Information Processing in Medical Im.*, H. H. Barrett and A. F. Gmitro, editors, volume 687 of *Lecture Notes in Computer Science*, pp. 322–338, Springer Verlag, Berlin, 1993.

[48] M. Grayson, "The heat equation shrinks embedded plane curves to round points," *J. Differential Geometry*, vol. 26, pp. 285–314, 1987.

[49] V. Guillemin and A. Pollack, *Differential Topology*, Prentice Hall, 1974.

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

[51] 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.

[52] 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.

[53] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: active contour models," *Intl. J. Comp. Vision*, vol. 1, pp. 321–31, 1987.

[54] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi, "Conformal curvature flows: from phase transitions to active vision," *Archive for Rational Mechanics and Analysis*, vol. 134, no. 3, pp. 275–301, 1996.

[55] G. F. Knoll, *Radiation detection and measurement*, Wiley, New York, 2 edition, 1989.

[56] 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.

[57] 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.

[58] Y. G. Leclerc, "Constructing simple stable descriptions for image partitionin," *Int'l J computer Vision*, vol. 3, pp. 73–102, 1989.

[59] M. Lee, A. Rangarajan, G. I. Zubal, and G. Gindi, "A continuation method for emission tomography," *IEEE Tr. Nuc. Sci.*, vol. 40, no. 6, pp. 2049–58, December 1993.

[60] 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.

[61] R. J. LeVeque, *Finite Difference Methods for Differential Equations*, AMath585-6 Coursepack, University of Washington, 1998.

[62] J. Lewin and M. Lewin, *An introduction to mathematical analysis*, McGraw-Hill, 1988.

[63] J. Libert, "Statistique de comptage avec distribution uniofmre de temps mort," *Nucl. Instr. Meth.*, vol. 151, no. 3, pp. 555–61, May 1978.

[64] R. Malladi, J. A. Sethian, and B. C. Vemuri, "Shape modeling with front propagation: a level set approach," *IEEE Tr. Patt. Anal. Mach. Int.*, vol. 17, no. 2, pp. 158–76, February 1995.

[65] B. M. Mazoyer, M. S. Roos, and R. H. Huesman, "Dead time correction and counting statistics for positron tomography," *Phys. Med. Biol.*, vol. 30, no. 5, pp. 385–399, 1985.

[66] 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.

[67] R. R. Meyer, "Sufficient conditions for the convergence of monotonic mathematical programming algorithms," *J. Comput. System. Sci.*, vol. 12, pp. 108–21, 1976.

[68] F. Morgan, *Riemannian Geometry*, Jones and Bartlett, 1993.

[69] J. W. Müller, "Dead-time problems," *Nucl. Instr. Meth.*, vol. 112, no. 1, pp. 47–57, September 1973.

[70] J. W. Müller, "Some formulae for a dead-time-distorted Poisson process," *Nucl. Instr. Meth.*, vol. 117, no. 2, pp. 401–4, March 1974.

[71] 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.

[72] S. Osher and J. A. Sethian, "Fronts propagation with curvature dependent speed: Algorithms based on Hamilton-Jacobi formulations," *Journal of Computational Physics*, vol. 79, pp. 12–49, 1988.

[73] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical recipes in C*, Cambridge Univ. Press, 1988.

[74] 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.

[75] L. I. Schiff, "Statistcal analysis of counter data," *Physical Review*, vol. 50, pp. 88–96, July 1936.

[76] J. Sethian, "Curvature and the evolutions of fronts," *Commun. Math. Phys.*, vol. 101, pp. 487–99, 1985.

[77] J. A. Sethian, *Level set methods: Evolving interfaces in geometry, fluid mechanics, computer vision and materials sciences*, Cambridge University Press, Cambridge, 1996.

[78] J. A. Sorenson and M. E. Phelps, *Physics in nuclear medicine*, Saunders, Philadelphia, 2 edition, 1987.

[79] T. J. Spinks, T. Jones, M. C. Gilardi, and J. D. Heather, "Physical performance of the latest generation of commercial positron scanner," *IEEE Tr. Nuc. Sci.*, vol. 35, no. 1, pp. 721–725, February 1988.

[80] L. H. Staib and J. S. Duncan, "Boundary finding with parametrically deformable models," *IEEE Tr. Patt. Anal. Mach. Int.*, vol. 14, no. 11, pp. 1061–1075, November 1992.

[81] C. W. Stearns, D. A. Chesler, J. E. Kirsch, and G. L. Brownell, "Quantitative imaging with the MGH analog ring positron tomograph," *IEEE Tr. Nuc. Sci.*, vol. 32, no. 1, pp. 898–901, February 1985.

[82] Y.-C. Tai, A. Chatziioannou, M. Dahlbom, and E. J. Hoffman, "Investigation on deadtime characteristics for simultaneous emission-transmission data acquisition in PET," in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 2, pp. 1489–93, 1997.

[83] 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.

[84] T. Tomitani, "An edge detection algorithm for attenuation correction in emission CT," *IEEE Tr. Nuc. Sci.*, vol. 34, no. 1, pp. 309–213, February 1987.

[85] C. R. Vogel and M. E. Oman, "Iterative methods for total variation denoising," *SIAM J. Sci. Comp.*, vol. 17, no. 1, pp. 227–38, January 1996.

[86] C. Xu and J. L. Prince, "Snakes, shapes, and gradient vector flow," *IEEE Tr. Im. Proc.*, vol. 7, no. 3, pp. 359–69, March 1998.

[87] 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.

[88] S. Yamamoto, M. Amano, S. Miura, H. Iida, and I. Kan, "Deadtime correction method using random coincidence for PET," *J. Nuc. Med.*, vol. 27, no. 12, pp. 1925–1928, December 1986.

[89] M. Yavuz and J. A. Fessler, "Penalized-likelihood estimators and noise analysis for randoms-precorrected PET transmission scans," *IEEE Tr. Med. Im.*, vol. 18, no. 8, pp. 665–74, August 1999.

[90] A. Yezzi, S. Kichenassamy, A. Kumar, P. Olver, and A. Tannenbaum, "A geometric snake model for segmentation of medical imagery," *IEEE Tr. Med. Im.*, vol. 16, no. 2, pp. 199–210, April 1997.

[91] D. M. Young, *Iterative solution of large linear systems*, Academic Press, New York, 1971.

[92] D. F. Yu and J. A. Fessler, "Edge-preserving tomographic reconstruction with nonlocal regularization," in *Proc. IEEE Intl. Conf. on Image Processing*, volume 1, pp. 29–33, 1998.

[93] D. F. Yu and J. A. Fessler, "Edge-preserving tomographic reconstruction with nonlocal regularization," *IEEE Tr. Im. Proc.*, 1999. Submitted.

[94] D. F. Yu and J. A. Fessler, "Mean and variance of singles photon counting with deadtime," *Phys. Med. Biol.*, 1999. Accepted.

[95] D. F. Yu, J. A. Fessler, and E. P. Ficaro, "Maximum likelihood transmission image reconstruction for overlapping transmission beams," *IEEE Tr. Med. Im.*, 1999. Accepted.

[96] S. C. Zhu and A. Yuille, "Region competition: unifying snakes, region growing, and Bayes/MDL/energy for multiband image segmentation," *IEEE Tr. Patt. Anal. Mach. Int.*, vol. 18, no. 9, pp. 884–900, September 1996.