WAVELET PHASE FILTER FOR DENOISING IN TOMOGRAPHIC IMAGE RECONSTRUCTION

BY

BIQUAN LIN

Submitted in partial fulfillment of the
requirements of the degree of
Doctor of Philosophy in Mathematics
in the Graduate School of the
Illinois Institute of Technology

Approved ______________________
Advisor

Chicago, Illinois
December, 1994

ACKNOWLEDGMENT

I would like to extend sincere thanks to my advisor, Dr. Elwood T. Olsen, for his guidance and friendship during the course of my graduate studies at IIT. I am truly thankful to him for sharing his great insight and perspectives with me, for guiding me when I needed direction, and for allowing me to pursue research topics of my own interest.

I would also like to thank the members of my thesis committee, Dr. Platon C. Deliyannis, Dr. Nikolas Galatsanos, Dr. Bao G. Nguyen and Dr. Miles Wernick for many instructive discussions on wavelet methods, Fourier analysis and image processing. Especially, I appreciate the wavelet and signal processing seminar initiated by Dr. Bao G. Nguyen and Dr. Nikolas Galatsanos, which stimulated my interest in wavelet methods and applications.

I am greatly thankful to the Department of Mathematics of IIT, which supported my graduate studies by offering an appointment of teaching assistantship, and to the Mathematics and Computer Science Division of Argonne National Laboratory, which offered me a research aide position in my final school year.

Thanks are also due to Dr. Henry Stark of the ECE Department of IIT, who kindly allowed me to use computers at his laboratory, and Ms. Jane Wurster, PhD candidate in the ECE Department of IIT, who helped me get familiar with many medical imaging programs. I would also like to thank Dr. Man K. Kwong and Dr. Peter Tang of the MCS Division of Argonne Lab for many fruitful discussions and continuous encouragement.

Finally, I would like to express my special gratitude to my wife and my parents for their understanding, support and encouragement in my pursuing graduate studies.

B.L.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Image reconstruction techniques such as position emission tomography (PET) can be used to study the physiology and biochemistry of the body in a pictorial manner. However, some of these techniques are characterized by a high level of noise. Wavelet methods can smooth out noise while preserving significant features of images. Mallat et al., for example, proposed a denoising scheme exploiting wavelet modulus maxima, but the scheme is sensitive to the variance of noise. In this study, we explore the properties of wavelet phase, with a focus on statistical tomographic image restoration. In comparison with conventional filtering methods, the wavelet phase method improved the quality of reconstructed statistical tomographic images. Specifically, we prove that the wavelet phase of regular Poisson noise under symmetric Haar transform converges in distribution to a random variable uniformly distributed on $[0, 2\pi)$. Using such property, we propose three wavelet-phase-based denoising schemes: edge tracking, local phase variance thresholding, and scale phase variation thresholding. The numerical experiments conducted indicate that the wavelet phase is a useful component for wavelet based denoising methods. The computational costs of the algorithms are bounded by the costs of the Fourier reconstruction algorithms.

CHAPTER I
INTRODUCTION

Positron emission tomography (PET) is a promising technique for studying physiology and biochemistry of the body in a pictorial manner. Even thought some idea of the technique can be rooted to as early as 50's, the first practical PET device was developed by Phelps, Hoffman, Ter-Pogossian and their co-workers in 1972. One important issue in PET is how to process the measurements of observed photons into an easily grasped visual image which indicates the concentration of a marked chemical substance in the body. This process is called PET image reconstruction. In the beginning, the image reconstruction algorithms were developed based on deterministical projection model and based on Fourier transform techniques. The convolution back projection (CBP) and direct Fourier reconstruction (DFR) methods are two typical examples in this approach. These algorithms were later found to be sensitive to the perturbation of measurements ("noise"). One way to reduce noise sensitivity is to model the emission process as a random process and develop the reconstruction algorithms based on the statistical model and by using statistical estimation techniques. The maximum likelihood (ML) and maximum a posteriori (MAP) methods are two typical statistical methods. However, among other problems, these statistical algorithms are generally very computationally demanding. Another approach to reduce noise sensitivity is to filter measurement data before employing a deterministical model based reconstruction algorithm. Some conventional filtering schemes have been investigated and found to be unsatisfactory. Designing a feasible filtering scheme is the goal of this dissertation.

Wavelet theory, as a unified framework for several techniques which have been developed independently in the area of applied mathematics and signal processing, has recently attracted a lot of attention in both mathematics and signal processing communities. Some researchers in the field are involved in exploiting characteristics of wavelets by introducing various wavelet bases. Others are trying to develop and apply wavelet based methods to some specific applications such as data compression, modeling of stochastic processes and noise reduction. Wavelet theory and applications are developing very rapidly. However, most studies are on one-dimensional wavelets. Even though some researchers extended their studies to multi-dimensional cases, they are more concerned with some characteristics such as magnitude and zero-crossing which are common for both one-dimensional and multi-dimensional cases. This dissertation studies phase of a two-dimensional wavelet, the characteristic of a two-dimensional wavelet without one-dimensional counterpart.

A good noise filtering scheme should remove noise in an image as much as possible and at the same time preserve features ("edges") of the image as much as possible. Mallat, Zhong and Hwang proposed a multiscale edge representation by using modulus maxima of wavelet transforms. They also studied the behavior of noise under wavelet transform at different scales and developed a noise filtering scheme using observations of noise behavior and incorporating edge information. Their method is promising and works well in some cases. However, their approach, like other magnitude-based methods, is sensitive to the variance of noise. Since the variance of noise occurring in PET applications is typically high, the wavelet modulus maxima based noise filtering scheme does not perform well. In this dissertation, we design a PET noise filtering scheme using both modulus and phase information.

The remainder of this dissertation is organized as follows: Chapter II reviews PET image reconstruction methods including convolution back projection (CBP), direct Fourier reconstruction (DFR), maximum likelihood (ML) and maximum a posteriori (MAP). This chapter also gives a statistical model for PET process in a rigorous way. Chapter III reviews several conventional noise filtering schemes including lowpass filtering, Wiener filtering and adaptive filtering. This chapter also apply the statistical model established in Chapter I to analyze the noise occurring in PET raysum data. Chapter IV discusses several edge detection schemes including the gradient magnitude method, the Laplacian method, the gradient angular method and the multiscale edge detection method. Chapter V introduces wavelet theory and its applications to edge detection and image restoration. Chapter VI gives the main results of the dissertation. In this chapter, we derive some theorems which exploit the properties of wavelet phases under Haar basis, propose PET noise filtering schemes by incorporating phase information, describe numerical experiments and analyze the algorithms for their costs and performances. Chapter VII summarizes the achievements in this thesis and gives some suggestions for further studies.

CHAPTER II
TOMOGRAPHIC IMAGE RECONSTRUCTION

2.1 Introduction

PET image reconstruction is a process which reconstructs an image of the concentration of a marked chemical compound in an object region from photons detected in detectors surrounding the region. Several reconstruction methods have been developed and heavily studied since the later 70's. Among them, the convolution back projection (CBP) method was the first method introduced in the field and has been employed in the most commercial PET imaging systems. The method is based on the central-slice projection theorem, which establishes a relation between the two-dimensional Fourier transform of an image and the one-dimensional Fourier transform of the projections ("raysums", or line integrals of the concentration) of the image. The direct Fourier reconstruction (DFR) method, developed by Stark *et al* [STAR79, 81a, 81b]*, is another reconstruction method based on this theorem. Instead of taking "convolution" and "back projection" as the CBP method, the DFR method interpolates the Fourier transforms of raysums. Since both CBP and DFR methods use techniques in Fourier analysis such as fast Fourier transform (FFT), they are referred to as Fourier methods in literature. In 1982, Shepp and Vardi [SHEP82] introduced a Poisson process model for PET emission process. Based on this model and by using the maximum likelihood (ML) method in statistical estimation theory, they derived a so-called Expectation Maximization (EM) algorithm for PET image reconstruction. By generalizing the model proposed by Shepp and Vardi and by using Bayesian estimation techniques, Derin [DERI84], Geman and Geman [GEMA84], Besag [BESA86], Sauer and Bouman [SAUE91] and others have made contributions in developing maximum a posteriori (MAP) methods for image reconstruction applications. The ML and MAP methods are referred to as statistical methods since they are based on the statistical models of PET process. This chapter reviews all these methods.

---

* Codes in brackets refer to coded references in the bibliography.

In this chapter, we first introduce the projection model for PET image reconstruction and the central-slice projection theorem, the theoretical foundation for Fourier reconstruction methods. Next, we discuss the basic ideas of two Fourier methods, the CBP method and the DFR method, and compare their advantages and disadvantages. After that, we establish statistical models for PET problem. Then, we discuss ML and MAP methods, which are based on the statistical models. Finally, we make some additional remarks.

2.2 Central-slice Projection Theorem

Let f(x,y) denote an image in a two dimensional space. For mathematical simplicity and at the same time without losing generality of real life images, we assume f(x,y) $\in \mathbf{L^1}(\mathbf{R^2}) \cap \mathbf{L^2}(\mathbf{R^2})$. Note that a real image is bounded and has compact support and thus always satisfies this assumption. Under this assumption, the existence of integrals which will be encountered in this section is obvious. Let $\hat{f}(u,v)$ denote the Fourier transform of f(x,y). Given an angle θ, we can define a rotated coordinate system (s,t) by

$$\begin{bmatrix} s \\ t \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}, \tag{2.2.1}$$

or equivalently,

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} s \\ t \end{bmatrix}. \tag{2.2.2}$$

The line integral of f(x,y) along a line parallel to the t axis and with a distance s from it is called the projection (or raysum) of f(x,y) at (θ, s), denoted by

$$p_\theta(s) \equiv \int_{-\infty}^{+\infty} f(x(s,t), y(s,t))dt. \tag{2.2.3}$$

Physically, the raysums are interpreted as the integrals of the concentration of the marked chemical substance alone specified lines. The purpose of an image reconstruction algorithm is to reconstruct an image of the concentration from the raysum data. This type of approach is referred to as a projection modeled approach, since the line integrals (2.2.3) are referred to in literature as “projections”.

Let $\hat{p}_\theta(\omega)$ denote the one-dimensional Fourier transform of $p_\theta(s)$. Then, we have the following so-called Central-slice Projection Theorem.

Theorem : The one dimensional Fourier transform of a raysum $p_\theta(s)$ is equal to a "central slice" of the two-dimensional Fourier transform $\hat{f}$ of an image at angle θ:

$$\hat{p}_\theta(\omega) = \hat{f}(\omega \cos\theta, \omega \sin\theta) \qquad \omega \in R,\ 0 \le \theta < \pi. \tag{2.2.4}$$

Proof. Taking Fourier transform for both sides of (2.2.3) for fixed θ, we obtain

$$\hat{p}_\theta(\omega) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x(s,t), y(s,t)) e^{-i\omega s} dt ds$$

Changing variables by (2.2.1) for the preceding integral, it follows that

$$\hat{p}_\theta(\omega) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y) e^{-i\omega(x \cos\theta + y \sin\theta)} dx dy$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y) e^{-i(\omega \cos\theta)x - i(\omega \sin\theta)y} dx dy$$

$$= \hat{f}(\omega \cos\theta, \omega \sin\theta).$$

QED

Note that since $\omega(\cos(\theta + \pi), \sin(\theta + \pi)) = -\omega(\cos(\theta), \sin(\theta))$, we obtain all possible values of $\hat{f}$ from one-dimensional Fourier transforms $\hat{p_\theta}$ by restricting $\omega \in [0, \infty), \theta \in [0, 2\pi)$ or alternatively by restricting $\omega \in (-\infty, \infty), \theta \in [0, \pi)$ in (2.2.4).

The convolution back projection (CBP) method and direct Fourier reconstruction (DFR) method are based on the central-slice projection theorem. The next two sections discuss the basis ideas of these two methods.

2.3 Convolution Back Projection Method

Knowing the raysum $p_\theta(s)$ for all θ, $0 \le \theta < 2\pi$, we can obtain the one-dimensional Fourier transform $\hat{p}_\theta(\omega)$. Due to the Central-slice Projection Theorem and the inverse Fourier transform formula, we have:

$$f(x,y) = \frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{+\infty} \hat{p}_\theta(\omega)e^{i\omega(x\cos\theta + y\sin\theta)}\omega d\omega d\theta. \tag{2.3.1}$$

Define the convolution operator C by

$$C[p_\theta(s)] = \frac{1}{2\pi}\int_0^{+\infty} \hat{p}_\theta(\omega)e^{i\omega s}\omega d\omega \tag{2.3.2}$$

and the back projection operator B by

$$B[g] = \frac{1}{2\pi}\int_0^{2\pi} g(\theta)d\theta. \tag{2.3.3}$$

Since s = xcosθ + ysinθ, it follows that

$$f(x,y) = B[C[p_\theta(s)]]. \tag{2.3.4}$$

Because the integral in (2.3.2) may diverge and because some smoothing of the image is often desired in any event, it is a common practice to replace (2.3.2) by

$$C[p_\theta(s)] = \frac{1}{2\pi}\int_0^{+\infty} \hat{p}_\theta(\omega)e^{i\omega s}\omega R(\omega)d\omega \tag{2.3.5}$$

where $R(\omega)$ is a window function with compact support and usually is a lowpass filter. The formula (2.3.4) gives a way to reconstruct f(x,y) from the raysum $p_\theta(s)$. The above inversion techniques are known as convolution back projection (CBP) method. Since a filter is employed in the formula (2.3.5), the method is occasionally called filtered convolution back projection (FCBP) method.

In practice, the CBP reconstruction method is implemented for discretized images and by using discrete approximation of the above formulas. The resulting algorithm of CBP method can be summarized as following:

Step 1. Take 1D FFT for discretized raysums $p_{\theta_i}(s_j)$, i=0, 1, 2,..., $N_\theta - 1$, j=0, 1, 2,..., $N_s - 1$, where N_θ is the number of views and N_s is the number of samples in each view;

Step 2. Filter the DFT of raysums obtained in Step 1 by a discrete lowpass filter $\omega_i R(\omega_i)$, i=0, 1, ..., $N_\omega - 1$, where N_ω is the size of the filter;

Step 3. Take inverse FFT for the filtered DFT of raysums obtained in Step 2;

Step 4. Take back projection for the data obtained in Step 3, that is, use a numerical quadrature technique to compute the integral (2.3.3).

Note that (2.3.2) and (2.3.5) are not exact inverse Fourier transforms, since the integration is from 0 to $+\infty$ rather than $-\infty$ to $+\infty$. Therefore, in the preceding Step 3, we in fact need to pad with zeros for negative frequency ω before employing an inverse FFT algorithm.

2.4 Direct Fourier Reconstruction Method

The direct fourier reconstruction (DFR) method is another effective PET image reconstruction method based on the central slice projection theorem. The key in this approach is using a polar sampling theorem, which can be derived from the Shannon sampling theorem. The polar sampling theorem says that the Fourier transform of an image can be obtained by interpolating its polar samples. This section discusses the principle and the algorithm of the direct Fourier reconstruction (DFR) method.

From the above central-slice projection theorem, we know that the 1D Fourier transform (FT) of each projection is a "central-slice" of the 2D Fourier transform of the image, *i.e.*,

$$\hat{p}_\theta(\omega) = \hat{f}(\omega\cos\theta, \omega\sin\theta).$$

Theoretically, if we knew $\hat{f}$ for all (ω, θ), then we could easily convert this FT from polar to Cartesian coordinates and reconstruct f(x,y) by the discrete approximation of the following inverse Fourier transform formula:

$$f(x, y) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\hat{f}(u, v)e^{ixu+iyv}\,du\,dv \tag{2.4.1}$$

where $\hat{f}(u, v)$ is the 2D FT of f(x,y) and is converted from

$$\hat{f}(\omega, \theta) = \hat{f}(\omega \cos \theta, \omega \sin \theta).$$

However, we know $\hat{\mathrm{f}}$ only on a finite discrete set of "polar points" $(\omega_{\mathrm{n}}, \theta_{\mathrm{k}})$. Hence, the problem of reconstructing PET images becomes to how to interpolate the known sampled values at the polar points and obtain the values at a rectangular Cartesian grid which are required to reconstruct f(x,y) by using a discrete form of (2.4.1). To solve this problem, Stark *et al* [STAR79, 81a, 81b] derived a result called the polar sampling theorem from Shannon sampling theorem.

Definition: f(x,y) is said to be space-limited to A if f(x,y)=0 for $\sqrt{x^2 + y^2} > A$. f(x,y) is said to be angularly band-limited to B if f(r, θ), the polar transform of f(x,y), is band-limited to B with respect to θ, *i.e.*, $\hat{f}_r(\alpha) = 0$ for $|\alpha| > B$ where $\hat{f}_r(\alpha)$ is the 1D Fourier transform of f(r, θ) with respect to θ.

Polar Sampling Theorem Let f(x,y) be space-limited to 2A, and let its Fourier transform, $\hat{\mathrm{f}}$, be angularly band-limited to $\frac{M}{2\pi}$. Then $\hat{\mathrm{f}}$ can be reconstructed from its polar samples $(\omega_{\mathrm{n}}, \theta_{\mathrm{m}}) = (\frac{\mathrm{n}}{2\mathrm{A}}, \frac{\pi \mathrm{m}}{\mathrm{M+1}})$ via

$$\hat{f}(\omega, \theta) = \sum_{n=-\infty}^{\infty} \sum_{m=0}^{2M+1} \hat{f}(\frac{n}{2A}, \frac{\pi m}{M+1}) \mathrm{sinc}[2A(\omega - \frac{n}{2A})] \sigma(\theta - \frac{\pi m}{M+1}), \quad (2.4.2)$$

where $\omega \geq 0$, $\mathrm{sinc}(x) = \frac{\sin(x)}{x}$ and $\sigma(x) = \frac{\sin(Nx/2)}{N \sin(N/2)}$.

The formula (2.4.2) enables us to compute $\hat{f}(\omega, \theta)$ at any point (ω, θ) if we know the values at countably many polar samples. In practice, however, we do not actually have the values at all the countably many polar samples and therefore we have to truncate n up to some bound N and use the truncated version of (2.4.2). After truncation, we can compute the discrete Fourier transforms $\hat{f}(u_k, v_l)$ with Cartesian coordinates by changing variables as

$$\begin{cases} \omega_{kl} = \sqrt{u_k^2 + v_l^2} \\ \theta_{kl} = \arccos \frac{u_k}{\omega_{kl}} \end{cases} \quad (2.4.3)$$

From these DFT's with Cartesian coordinates, we can reconstruct the discretized object image $f(x_i, y_j)$ by inverse DFT, the discrete form of formula (2.4.1). This is the basic idea of DFR methods.

The numerical algorithm of DFR method is as follows. Let N_θ denote the number of views and N_s denote the number of samples in each view. Assume N_s is even for simplicity. The observed raysum data is denoted by $p(n\Delta s, m\Delta\theta)$ where $\Delta\theta$ is the angle of view, Δs is the sampling increment in each view and n=$-N_s/2, ..., N_s/2-1$, $m = 0, 1, , ..., N_\theta - 1$. The DFR algorithm can be divided into three steps. First, take 1D Fast Fourier transform of raysum $p(n\Delta s, m\Delta\theta)$ for each m. Second, for each $(u_k, v_l) = (k\Delta u, l\Delta v)$, compute $(\omega_{kl}, \theta_{kl})$ by (2.4.3) and computer $\bar{f}(\omega_{kl}, \theta_{kl})$ by truncating n to some bound N in (2.4.2). Therefore, we obtain $\hat{f}(k\Delta u, l\Delta v)$, which is the discrete Fourier transform of the image to be reconstructed. Third, take 2-D inverse fast Fourier transform of $\hat{f}(k\Delta u, l\Delta v)$ to obtain the reconstructed image f(i,j).

Since f(x,y) is assumed to be space-limited and its Fourier transform $\hat{\text{f}}$ is assumed to be angularly band-limited, the Fourier transform $\hat{\text{f}}$ must be infinitely differentiable and spread over the whole radial direction ω_n [SLEP83]. The truncation in the radial direction n will introduce a truncation error, but the error can be controlled by choosing sufficiently large N. The following theorem is for the one-dimensional case [DYM72, pp.129], but it supports the above assertion.

Theorem: (1). Suppose that f(t) is band-limited to B. Then,

$$f(t) = \sum_{n=-\infty}^{+\infty} f(\frac{n}{2B}) \frac{\sin \pi(2Bt - n)}{\pi(2Bt - n)}. \tag{2.4.4}$$

(2). In addition to the condition in (1), let f(t) be normalized, *i.e.*,

$$\int_{-\infty}^{+\infty} |f(t)|^2 dt = 1,$$

and have energy α for interval [-A, A], *i.e.*,

$$\int_{|t| \le A} |f(t)|^2 dt = \alpha. \tag{2.4.5}$$

Denote $f_\delta(t) = f(t+\delta)$ where δ, $0 < \delta \leq 1/(2B)$, is referred to as time lag. If we approximate f(t) by

$$\tilde{f}(t) = \sum_{|n|\leq 2AB} f(\frac{n}{2B})\frac{\sin\pi(2Bt-n)}{\pi(2Bt-n)}. \tag{2.4.6}$$

then the total error energy resulted from the time lag δ is bounded by $\frac{1-\alpha}{2B}$, *i.e.*,

$$\int_0^{1/(2B)} ||f_\delta - \tilde{f}_\delta||_2^2 d\delta \leq \frac{1-\alpha}{2B}. \tag{2.4.7}$$

Remark. The formula (2.4.4) is called Shannon sampling formula and it says that a bandlimited signal can be reconstructed from its countable and equally spaced samples. But in practice, the signal value can only be computed by truncating to some finite number of samples and the formula (2.4.6) is the truncation version of (2.4.4). For each time, the truncating error can be measured in L_2 norm and be denoted by $||f-\tilde{f}||_2$. Suppose we take two sets of samples, each sample location in the first set is shifted by δ from that in the second set. The original signal can be approximated by either set of samples via (2.4.6) but the approximation error vary with δ. However, this theorem gives a bound for the total L_2 square error for $delta < 1/(2B)$.

Proof. The formula (2.4.4) can be derived by inserting a Fourier series expansion of $\hat{f}$ (such expansion exists since f(t) is bandlimited) into an inverse Fourier transform formula giving f from $\hat{f}$. Next, we prove (2.4.7). Since $\hat{f}(\omega)$ can be expanded into a Fourier series as

$$\hat{f}(\omega) = \sum_{n=-\infty}^{+\infty} \frac{1}{2B} f(\frac{n}{2B}) e^{\frac{-i\pi\omega n}{B}}, \tag{2.4.8}$$

it follows that

$$\begin{aligned} ||f||_2^2 = ||\hat{f}||_2^2 &= \int_{-B}^{B} |\hat{f}(\omega)|^2 d\omega \\ &= \frac{1}{2B}\sum_{n=-\infty}^{n=+\infty} |f(\frac{n}{2B})|^2. \end{aligned} \tag{2.4.9}$$

Since f(t) is band-limited, $f_\delta(t)$ is also bandlimited. Using (2.4.6), it is easy to verify that $\tilde{f}_\delta(t)$ is also bandlimited. Therefore, $f_\delta(t) - \tilde{f}_\delta(t)$ is bandlimited. Since

$$||f_\delta - \tilde{f}_\delta||_2^2 = ||f_\delta - \sum_{|n|\leq 2AB} f_\delta(\frac{n}{2B})\frac{\sin \pi(2Bt-n)}{\pi(2Bt-n)}||_2^2$$

$$= ||\sum_{|n|>2AB} f_\delta(\frac{n}{2B})\frac{\sin \pi(2Bt-n)}{\pi(2Bt-n)}||_2^2,$$

applying (2.4.9) to $f_\delta - \tilde{f}_\delta$, we obtain

$$||f_\delta - \tilde{f}_\delta||_2^2 = \frac{1}{2B}\sum_{m=-\infty}^{+\infty} |\sum_{|n|>2AB} f_\delta(\frac{n}{2B})\frac{\sin \pi(m-n)}{\pi(m-n)}|^2$$

$$= \frac{1}{2B}\sum_{|m|>2AB} |f_\delta(\frac{m}{2B})|^2.$$

Thus,

$$\int_0^{1/(2B)} ||f_\delta - \tilde{f}_\delta||_2^2 d\delta$$

$$= \frac{1}{2B}\int_0^{1/(2B)} \sum_{|m|>2AB} |f_\delta(\frac{m}{2B})|^2 d\delta$$

$$= \frac{1}{2B}\sum_{|m|>2AB}\int_{m/(2B)}^{(m+1)/(2B)} |f(\delta)|^2 d\delta$$

$$\leq \frac{1}{2B}\int_{|t|>A} |f(t)|^2 dt$$

$$= \frac{1-\alpha}{2B}.$$

QED

As with CBP, a lowpass filter or other restoration scheme can be used to smooth the image. The scheme is applied to the noisy raysum data before employing the DFR algorithm. The simulation results by using several restoration schemes including lowpass filtering and a POCS method, can be seen in Yang's MS thesis [YANG91].

2.5 Comparison of CBP with DFR

In this section, we make some comments regarding to the computational cost and performance of the CBP and DFR methods.

First, let us consider the number of operations. Suppose $N_\theta = N_s = N$. It was pointed out in [STAR81] that DFR requires $O(N^2 \log N)$ operations while CBP requires $O(N^3)$ operations. It is obvious that DFR method has a speed advantage over CBP method.

Next, let us consider the performances of the methods in the presence of noise. For a non-noise case, Yang's[YANG91] numerical experiments demonstrated that the quality of the reconstructed image provided by the DFR method is at least as good as that by the CBP method. If noise is present, the image reconstructed by the DFR method without filtering is not as good as that reconstructed by the CBP method. But DFR can perform as well as CBP by carefully choosing a filter. For both noiseless and noisy situations, the mean square errors of these two methods are not significantly different.

The common drawback of these two methods is that they are very sensitive to noise. If noise is present, the quality of the reconstructed images is not satisfactory even after employing a filter. This can be seen in Yang's simulation results. The premise of this thesis is that the reason why the filters considered by Yang do not effectively ameliorate the image is that they do not sufficiently consider the features of PET noise. Since the PET process is characterized by its high level noise, it is desirable to analyze the noise more carefully and to design a sophisticated filter specifically for the PET noise. This is the approach we use here. After the filter has been applied to the raysums, we will use the DFR or CBP method to reconstruct the image.

In summary, the Fourier reconstruction methods perform well when the raysum data is of good quality. However, if the raysum data has a high noise level, these methods, which are very sensitive to noise, perform less well. This drawback might be inherent since they originally were developed from a deterministic model. In the following sections, we first introduce statistical models for PET problem. Next, we discuss the Poisson model based maximum likelihood (ML) method and the Cox model based maximum a posteriori (MAP) method. After that, we compare these statistical methods with Fourier methods.

2.6 Statistical Model of PET Problem

2.6.1 Modeling PET Emission Process. Positron Emission Tomography (PET) is a medical technique for measuring the spatial variation in concentration of a marked chemical substance in an organ of a human or an animal. The technique is essentially as follows: a position emitting substance is injected into the regions of interest. Surrounding the region, a group of detectors is installed. When a positron is emitted from the radioactive substance, it finds a nearby electron and they annihilate each other. As the result of the annihilation, two photons are generated and travel in nearly opposite directions. It is assumed here that the detectors lie in a plane surrounding the region. A pair of the detectors detects the two photons almost coincidently. From these measurements by using a PET image reconstruction algorithm, one obtains an image showing the density of the marked chemical substance in the plane of the detectors.

In order to apply mathematical and statistical methods to the PET problem, we need to model the above PET physics. For the purpose of balancing simplicity and feasibility, we impose several restrictions and assumptions to the PET model:

First, we model the emission process as a two-dimensional problem. Even though a true emission process occurs in a three-dimensional space, the restriction to the plane where the detectors lie can simplify the problem and presumably characterize the real emission process.

Second, we neglect some details of PET emission process such as the attenuation by body tissues and the halflife of the radioactive chemical substance.

Third, while some photons emitted are not detected, we model only those photon pairs which are detected.

Fourth, we assume that two emitted photons travel in the exactly opposite directions. By such assumption, the photons detected in a pair of detectors are only from the tube connecting these two detectors.

Finally, we assume that the emissions occurring in the region follows a Poisson model. We formulate this assumption as follows:

Let I denote the total number of pixels in the region and A_i, i=1,2, ... I denote the pixels. Let D denote the total number of detectors and d_i, i=1,2, ... D denote the detectors. Let (d_i, d_k) denote the cylindrical tube connecting two detectors d_i and d_k where i $\neq$ k. (d_i, d_k) is called a detector tube or a detector pair. There are a total of $\frac{D(D-1)}{2}$ of such detector pairs. For notational simplicity, we let $J = \frac{D(D-1)}{2}$ and $P_j = (d_i, d_k), j = 1, 2, ...J$. Let $N_{i,j}$ be number of photons emitted from a pixel A_i and detected by detector a pair P_j during a period of time Δt. We assume:

(i) For pixels A_i which do not intersect the detector tube P_j, $N_{i,j} = 0$ (this must be the case if photons travel in exactly opposite directions and is a reasonable assumption so long as photons travel in nearly opposite directions); for pixels A_i which intersect the detector tube P_j, each $N_{i,j}$ is a random variable following Poisson distribution with mean $\lambda_{i,j}$, that is, for each (i,j),

$$P(N_{i,j} = n_{i,j}) = e^{-\lambda_{i,j}} \frac{\lambda_{i,j}^{n_{i,j}}}{n_{i,j}!};$$

(ii) $\{N_{i,j}, i = 1, 2, ..., I, j = 1, 2, ..., J\}'s$ are independent, that is,

$$P(N_{i,j} = n_{i,j}, i = 1, 2, ...I, j = 1, 2, ..., J) = \prod_{i=1}^{I} \prod_{j=1}^{J} P(N_{i,j} = n_{i,j}).$$

The above (i) and (ii) is our Poisson model assumptions. The Poisson model is commonly used for PET. We make some comments on the model. In the assumption (i), for pixels A_i which intersect the detector tube P_j, $N_{i,j}$ is assumed to follow Poisson distribution with mean $\lambda_{i,j}$. This assumption is reasonable for the following reasons. Let N_i denote the number of photons emitted from pixel A_i for a given time interval of observation, then N_i can be modeled as a Poisson distributed random variable if we assume that (a) the number of photons emitted in nonoverlapping subintervals are independent; (b) the probability of exact one photon pair emitted in a sufficiently short subinterval of length h is approximately proportional to h (the halflife of the radioactive chemical substance is ignored in this model); (c) the probability of two or more photon pairs emitted in a sufficiently short subinterval is zero. Since $N_{i,j}$ can be treated as a thinning process of N_i, that is, for given i, $N_{i,j}$'s are independent and $\sum_j N_{i,j} = N_i$, it means that $N_{i,j}$'s are also Poisson random variables as long as N_i is a Poisson random variable. Therefore, the assumption (i) is reasonable if we assume the PET emission process follows (a), (b) and (c). The mean $\lambda_{i,j}$ in the assumption (i) is proportional to (1) the concentration of marked chemical substance; (2) the duration of interval of observation and (3) the area of intersection between pixel A_i and tube P_j. The (i) and (ii) assumptions are consistent with what is known about β-decay process; the event that a marked atom undergoes β-decay seems to be independent of anything in the environment, including the β-decay of nearby atoms. In summary, we conclude that the Poisson model is a reasonable assumption for PET emission process.

For MAP estimation, a more general model called a Cox model is often used. In the Poisson model, each mean $\lambda_{i,j}$ in above (i) is a real number (not a random variable). In the Cox model, each $\lambda_{i,j}$ is assumed to be a random variable. The idea is that the $\lambda_{i,j}$'s change in " random" ways as the identity of the patient changes but $N_{i,j}$'s, given the identity of the patient, exhibit the Poisson distribution and independence assumptions stated forth above. In essence, a Cox model generalizes a Poisson model by assuming that each $\lambda_{i,j}$ be a random variable (so it will be changed to $\Lambda_{i,j}$ to denote a random variable as convention) and by replacing "independence" in above (ii) by " conditional independence" given $\Lambda_{i,j} = \lambda_{i,j}$. Formally, a Cox model replaces (i) and (ii) above by:

(i') For pixels A_i not located in the detector tube P_j, $N_{i,j} = 0$; for pixels A_i located in the detector tube P_j, the conditional distribution of $N_{i,j}$ given $\Lambda_{i,j} = \lambda_{i,j}$ is

$$P(N_{i,j} = n_{i,j} | \Lambda_{i,j} = \lambda_{i,j}) = e^{-\lambda_{i,j}} \frac{\lambda_{i,j}^{n_{i,j}}}{n_{i,j}!};$$

(ii') The set $\{N_{i,j}, i = 1, 2, ..., I, j = 1, 2, ..., J\}$ are conditionally independent given $\Lambda_{i,j} = \lambda_{i,j}$, that is,

$$P(N_{i,j} = n_{i,j} | \Lambda_{i,j} = \lambda_{i,j}, i = 1, 2, ..., I, j = 1, 2, ..., J)$$

$$= \prod_{i=1}^{I} \prod_{j=1}^{J} P(N_{i,j} = n_{i,j} | \Lambda_{i,j} = \lambda_{i,j}).$$

As noted above, the Cox model is used in MAP method, which will be discussed in Section 2.8. Now, we return to the Poisson model.

2.6.2 The Distribution of Number of Detected Photons. One important result which can be drawn from the Poisson model is that the number of photons detected in detector pairs are independent, each of them is either a constant zero or a Poisson distributed random variable. This subsection is to derive this result.

Let $N_{.,j}, j = 1, 2, ...J$, denote the number of photons detected at the detector pair j. Then,

$$N_{.,j} = \sum_{i=1}^{I} N_{i,j}.$$

If $A_i \cap P_j = \emptyset$ for all i, $N_{.,j} = 0$ follows by assumption (i). Otherwise, since $N_{i,j}$'s are independent Poisson random variables and since the sum of independent Poisson random variables is a Poisson random variable and since the mean of the sum of random variables is equal to the sum of the means of random variables, it follows that $N_{.,j}, j = 1, 2, ...J$ is a Poisson random variable with mean $\sum_{A_i \cap P_j \neq \emptyset} \lambda_{i,j}$.

Next, we show that $\{N_{.,j}, j = 1, 2, ...J\}$ are independent random variables, that is,

$$P(N_{.,1} = n_{.,1}, N_{.,2} = n_{.,2}, ..., N_{.,J} = n_{.,J}) = \prod_{j=1}^{J} P(N_{.,J} = n_{.,J}).$$

Observe that a zero random variable is independent with any collection of random variables. It will not lose generality by assuming all $N_{.,j_k}$'s are Poisson distributed. In fact,

$$\begin{aligned}
&P(N_{.,1} = n_{.,1}, N_{.,2} = n_{.,2}, ..., N_{.,J} = n_{.,J}) \\
&= P(\sum_{i=1}^{I} N_{i,1} = n_{i,1}, ..., \sum_{i=1}^{I} N_{i,J} = n_{.,J})) \\
&= \sum_{\{\sum n_{i_j,j} = n_{.,j}, j=1,2,...,J\}} P(N_{i_1,1} = n_{i_1,1}, ..., N_{i_J,J} = n_{i_J,J}) \\
&= \sum_{\{\sum n_{i_j,j} = n_{.,j}, j=1,2,...,J\}} P(N_{i_1,1} = n_{i_1,1})...P(N_{i_J,J} = n_{i_J,J}) \\
&= \sum_{\{\sum n_{i_1,1} = n_{.,1}\}} P(N_{i_1,1} = n_{i_1,1}) \sum_{\{\sum n_{i_2,2} = n_{.,2}\}} P(N_{i_2,2} = n_{i_2,2})... \\
&\qquad \sum_{\{\sum n_{i_J,J} = n_{.,J}\}} P(N_{i_J,J} = n_{i_J,J}) \\
&= P(N_{.,1} = n_{.,1})P(N_{.,2} = n_{.,2})...P(N_{.,J} = n_{.,J}).
\end{aligned}$$

In conclusion, we have shown that the number of photons detected in detector pairs are independent, each of them is either a Poisson random variable or a zero random variable.

Similarly, we can show that the number of photons emitted from pixels and detected by detector pairs are also independent, each of them is either a constant zero or a Poisson distributed random variable. Formally, let $N_{i,.}$ denote the number of photons emitted from pixel i and detected by detector pairs. Then,

$$N_{i,.} = \sum_{j=1}^{J} N_{i,j}.$$

If $N_{i,j} = 0$ for all j=1,2,...J, then $N_{i,.}$=0. Otherwise, $N_{i,.}$ is a Poisson distributed random variable with mean $\lambda_{i,.} = \sum_{j=1}^{J} \lambda_{i,j}$, where $\lambda_{i,j}$ is the mean of $N_{i,j}$.

2.6.3 Statistical Analysis of Projection Model. The model underlying Fourier methods reflects the physics of X-ray transmission tomography. The algorithms based on the model are very effective in computerized tomography (CT) applications. Even though the physics of transmission tomography is entirely different from that of emission tomography, including PET, Fourier methods have been used in almost all PET image reconstruction systems. The reason for that is that Fourier methods achieve reasonably good reconstruction at much lower cost, compared with a maximum likelihood (ML) method or maximum a posteriori (MAP) method. In transmission tomography, the term projection (or raysum) is used to refer to the accumulated photons along each mutually parallel line for each angle, which is rotated from 0 to 180 degree. In order to utilize Fourier methods for emission tomography such as PET, we need to interpret the raysum in a reasonable way and hope that it could be used to reflect the measured number of photon pairs simultaneously detected by each pair of detectors. In this subsection, we explore the relation between raysum data and the number of detected photons, and will find that the raysum data is in fact a reasonable approximation of the number of detected photons and thus follows Poisson distribution. This may partly answer the question why Fourier methods can achieve a reasonably good image reconstruction.

Recall in Section 2.2 that the raysum of a density function f(x,y) at angle θ is defined by (2.2.3). In practice, we cannot compute raysum for infinitely many parallel lines for infinitely many angles, which are from 0 to 180 degree. A discretization is necessary. Let N_θ denote the total number of discretized projection angles and $\theta_l, l = 1, 2, ..., N_\theta$ denote each discretized angle. Let N_s denote the total number of discretized parallel lines and s_k, $k = 1, 2, ..., N_s$ denote each line. By the same reason, the object region (the area through which parallel X-rays project in transmission tomography or the area from which emission occurs in emission tomography) is discretized into I pixels. Let A_i, $i = 1, 2, ...I$, denote each pixel. By discretizing (2.2.3), the raysum for (θ_l, s_k) is computed by

$$p(l,k) = \sum_{A_i \cap S_{l,k} \neq \emptyset} f(A_i) \tag{2.6.1}$$

where $S_{l,k}$ denotes the parallel strip which is divided equally by N_s and is associated with angle θ_l and s_k. Note that the Equ. (2.6.1) is just a discretized version of (2.2.3) and we want to see whether the number of detected photons in the Poisson model satisfies such equation under some reasonable assumption so that the number of detected photons can be interpreted as raysums.

In the previous subsection, it was derived that number of photons detected in detector pairs $P_j, j = 1, 2, ...J$, are independent Poisson random variables with mean

$$\lambda_{.,j} = \sum_i \lambda_{i,j}. \tag{2.6.2}$$

where $\lambda_{i,j}$ is set to be 0 if $A_i \cap P_j = \emptyset$ Suppose that any photon pair must be detected by one of the detector pairs. Then $N_{i,.}$ defined above is the total number of photons emitted from pixel A_i. Assume that the random variable indicating the number of photons detected at P_j and from A_i has the same mean, that is $\lambda_{i,j} = c\lambda_{i,.}$ for $A_i \cap P_j \neq \emptyset$, where c is a constant. Under the assumption, the equation (2.6.2) become

$$\lambda_{.,j} = c \sum_{A_i \cap P_j \neq \emptyset} \lambda_{i,.}. \tag{2.6.3}$$

Observed that equation (2.6.1) and (2.6.3) are essentially the same by corresponding $\lambda_{.,j}$ to $p(l,k)$, $c\lambda_{i,.}$ to $f(A_i)$ and P_j to $S_{l,k}$. This means that the raysum data can be treated as the mean of the number of photons detected in detector pairs under the above assumptions.

For Fourier methods, the raysum is estimated by the observed number of photons detected in each detector pair. Due to the low count rate intrinsic to PET, this estimate is a high variance estimate. The heuristic argument is as follows. Let N_k denote the number of photons detected in detector pair k for a period time of Δt. From the discussion, we know that N_k's are independent Poisson random variables. Let λ be the mean of the total number of photons detected in all detector pairs, *i.e.*, $\lambda = E(\sum_{k=1}^{K} N_k)$. λ reflects the count rate of the PET process. The count rate λ can be increased by increasing observation time Δt as long as the radioactivity caused by the radioactive substance is still alive. To compare the variance of raysum at different count rate, we normalize the number of photons detected in each detector pair, N_k, by a factor λ. Let Y_k, $Y_k = \frac{N_k}{\lambda}$, denote raysum image. IF Y_k has higher variance, the image appears more noisy. For each Y_k,

$$E(Y_k) = \frac{\lambda_k}{\lambda}$$

and

$$V(Y_k) = V(\frac{N_k}{\lambda}) = \frac{\lambda_k}{\lambda^2} = \frac{E(Y_k)}{\lambda}.$$

Here we see that the variance of Y_k is proportional to the inverse of the count rate. For a low count rate, the variance is typically high. Here we also know that the image will get less noisy if we wait longer time Δt since the count rate increases with Δt. Therefore, the reconstructed image by Fourier method is not satisfactory for a statistically modeled PET problem. In order to improve the quality of reconstructed images, some filtering is needed before employing Fourier methods.

2.7 Maximum Likelihood Method

The maximum likelihood (ML) method in PET image reconstruction is a method based on the Poisson model of the emission. It maximizes the likelihood function of the observed data. Shepp and Vardi [SHEP82] developed an iterative algorithm called EM algorithm to implement the ML method for PET. We summarize their approach in the following paragraph.

The maximum likelihood method is often used to estimate nonrandom parameters based on the observation of random variables. The probability of an observation varies with the parameters. The estimated values of the parameters are so chosen so that probability (or likelihood) of the observation is maximized. Let α denote a parameter and $\{x_1, x_2, ..., x_n\}$ denote an observation of random variables $\{X_1, X_2, ..., X_n\}$. A likelihood function $\mathrm{L}(\alpha)$ is defined by $P(x_1, x_2, ..., x_n|\alpha)$. Sometimes, the parameter values can be determined by using elementary calculus techniques. For example, assuming that $\{X_1, X_2, ..., X_n\}$ are i.i.d. random variables with known variance σ and unknown mean θ, the ML method is to estimate the mean by maximizing

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\theta)^2}{2\sigma^2}}.$$

It is easy to verify that the ML estimator of θ is $\frac{1}{n}\sum_{i=1}^{n} x_i$.

Recall in the preceding section that the number of photons detected in the detector pair k, $N_{.,k}$ (for notational simplicity, which is also denoted by N(k)), k=1,2, ..., K, are independent Poisson distributed random variables with mean $\lambda(k) = \lambda_{.,k}$ where K is the total number of non-trivial pairs (a pair is said to be trivial if the number of photons detected in the pair is a constant zero random variable). Therefore, the likelihood function of the observed data is

$$L(\lambda) = P(\mathbf{N}|\lambda) = \prod_{k=1}^{K} e^{-\lambda(k)} \frac{\lambda(k)^{n(k)}}{n(k)!}$$

where n(k) is the observed number of photons in tube k. Recall that the number of photons emitted from pixel l, U(l)=$N_{l,.}$, l=1,2,...,L, are are also independent Poisson distributed random variables with mean $u(l) = \lambda_{l,.}$ where L is the total number of non-trivial pixels (a pixel is said to be trivial if the number of photons emitted from this pixel is a constant zero random variable). Let $p(l,k) = \frac{\lambda_{l,k}}{\lambda_{l,.}}$. Since $\lambda(k) = \sum_{l=1}^{L} u(l)p(l,k)$, $k = 1, 2, ..., K$, it follows that we can write

$$L(\mathbf{u}) = \prod_{k=1}^{K} e^{-\sum_{l=1}^{L} u(l)p(l,k)} \frac{(\sum_{l=1}^{L} u(l)p(l,k))^{n(k)}}{n(k)!}. \tag{2.7.1}$$

for the likelihood function for the mean densities u(l). Taking logarithm of (2.7.1), we obtain

$$\ln L(\mathbf{u}) = \sum_{k=1}^{K}(-\sum_{l=1}^{L} u(l)p(l,k) + n(k)\ln(\sum_{l=1}^{L} u(l)p(l,k)) - \ln n(k)!).$$

Taking first derivative with respect to each u(k) and letting it be zero, we derive that the ML estimate $\hat{\mathbf{u}}$ of $\mathbf{u}$ satisfies

$$1 = \sum_{k=1}^{K} \frac{n(k)p(l,k)}{\sum_{j=1}^{L} \hat{u}(j)p(j,k)}, l = 1, 2, ...L.$$

One can find a complete argument in [VARD86] showing that such $\hat{\mathbf{u}}$ is the maximum point.

Multiplying $\hat{u}(l)$ to both sides of the preceding identity, we obtain

$$\hat{u}(l) = \hat{u}(l) \sum_{k=1}^{K} \frac{n(k)p(l,k)}{\sum_{j=1}^{L} \hat{u}(j)p(j,k)}, l = 1, 2, ...L \tag{2.7.2}$$

The $\hat{u}(l)$ in (2.7.2) is mathematically redundant, but it suggests that $\hat{u}(l)$ might be obtained iteratively via

$$\hat{u}^{i+1}(l) = \hat{u}^{i}(l) \sum_{k=1}^{K} \frac{n(k)p(l,k)}{\sum_{j=1}^{L} \hat{u}^{i}(j)p(j,k)}, l = 1, 2, ...L \tag{2.7.3}$$

from a given initial data $\hat{u}^0(l), l = 0, 1, 2, ...L$. This iterative scheme is called EM algorithm.

The EM algorithm can be shown to converge to a global maximum of the likelihood function (for the proof, see Lange and Carson [LANG84]). In the numerical experiment by Shepp and Vardi [SHEP82], the quality of the image reconstructed by the EM method seems better than that by the conventional back projection method.

However, there are two main drawbacks of this EM algorithm. First, it is computationally much more expensive than CBP or DFR. Suppose that the image size is N^2 and that the number of detectors is N. Then both K and L in formula (2.7.3) are $O(N^2)$. Therefore, from (2.7.3) it is easy to see that the EM algorithm requires $O(N^4)$ operations for each iteration while CBP requires $O(N^3)$

operations and DFR requires $O(N^2 \ln N)$ operations. Since the EM method usually takes 50 iterations to achieve sufficient precision [VARD85], the time cost of the EM algorithm is much higher than that by either CBP or DFR method. Even though some acceleration techniques such as Lewitt and Muehllehner's[LEWI86] relaxation method have been proposed to speed up the convergence of the iteration, the EM algorithm still remains computationally very expensive.

Second, the EM algorithm demonstrates instability during the iteration process. Even though the convergence of the EM algorithm is guaranteed by Lange and Carson's arguments, Y. Vardi, L. A. Shepp and L. Kaufman[VARD85] observed that as the EM iteration process goes on, the reconstructed image become more oscillatory and some artifacts near the edge of the region of interest start to appear. This phenomenon became very strong after 64 iterations in their numerical experiments. Snyder *et al* [SNYD87] analyzed the causes of this instability. According to [SNYD87], the artifacts are not resulted from the EM algorithm itself but due to the quantization error of (2.7.1) and finite machine precision. They also provided a sieve method which applies a Gaussian filter in the iteration process to reduce the artifacts. However, Liow and Strother [LIOW91] pointed out that the EM algorithm with sieves provides image with better quality compared to CBP only if a large number (100-200) of iterations has been processed. Due to the huge computational cost of the EM algorithm, this smoothing technique is also not practical. Even though some other techniques have been appended to improve the EM algorithm, this maximum likelihood method still has not been used in a commercial PET device.

2.8 Maximum A Posteriori Method

In this section, we explore how to apply maximum a posteriori (MAP) method to the Cox model of PET emission process and then make some general comments on MAP method in its application to image reconstruction. The essence of MAP estimation is to find a better estimate by exploiting the prior information of random mean measure.

We recall that a Cox model can be generalized from a Poisson model by replacing deterministic mean measure with random mean measure. If we assume that the mean measure of PET emission is not a collection of deterministic functions u(l) but a collection of random variables U(l) with some probability distributions, then the PET emission process is modeled as a Cox model. In this setting, the number of photons detected in tubes are conditionally independent Poisson process, *i.e.*,

$$P(n|u) = P(N(1) = n(1), ..., N(K) = n(K)|U(1) = u(1), ..., U(L) = u(L))$$

$$= \prod_{k=1}^{K} e^{-\sum_{l=1}^{L} u(l)p(l,k)} \frac{(\sum_{l=1}^{L} u(l)p(l,k))^{n(k)}}{n(k)!}. \tag{2.8.1}$$

The reconstruction problem becomes to estimate random mean measure U(l), i=1,2,...,L, from random variable N(k), k=1,2,...,K. The MAP method for PET reconstruction is such a method that estimate U by maximizing the posteriori density $P(u|n)$. Observe

$$P(u|n) = \frac{P(n|u)P(u)}{P(n)}, \tag{2.8.2}$$

Since maximizing $P(u|n)$ is equivalent to maximizing $\log P(u|n)$ it follows that the MAP estimate of U is an estimate which maximizes $\log P(n|u) + \log P(u) - \log P(n)$. Since $\log P(n)$ does not depend on u, it follows that the MAP estimate $\hat{U}$ of U is an estimate which maximizes $\log P(n|u) + \log P(u)$. $P(n|u)$ is based on a model, as in (2.8.1) for PET image reconstruction problem. The MAP estimate also depends on the distribution P(u), which reflects our prior knowledge about random variable U. The choice of such prior distribution is a critical issue in MAP estimation In fact, most studies in MAP image estimation are devoted to the development of feasible numerical algorithms under different prior assumptions.

We make some comments on the MAP methods. First, we discuss the relation between the MAP estimate and the ML estimate discussed in the previous section. Suppose that P(u) is a a constant over a set Γ in the parameter space, *i.e.*, P(u)=c for u $\in \Gamma$ and P(u)=0 otherwise. Then it can be seen from (2.8.2) that maximizing P(u|n) is equivalent to maximizing P(n|u) by considering the cases of $u \in \Gamma$ and $u \notin \Gamma$. It means that MAP estimate is equal to ML estimate for this case.

Therefore, the ML estimate discussed in the previous section is a special case of the MAP method.

Second, we discuss the prior distribution $P(u)$. Assume that P(n|u) is Gaussian distributed (Note that this assumption differs from that in PET as (2.8.1)). If p(u) is also Gaussian, then the MAP estimate is equal to E(U|N), the minimum mean square error estimate of U [TREE68]. For the applications of image reconstruction, however, the Gaussian prior distribution results in excessively blurred reconstruction [BOUN91]. Generally, a Gaussian assumption assumes that every pixels are correlated. This may be the cause for the above blurring phenomenon. To incorporate the local nature of images, a Markov random field (MRF) assumption is commonly adopted for prior distribution. A random field is a collection of random variables with index in a subset of the two-dimensional lattice, denoted by $U(i,j),(i,j) \in \Gamma$ where Γ is a subset of the two-dimensional lattice. A random field is called a Markov random field if given the distributions of all pixels, the conditional distribution of U(i,j) depends only on the distributions of its neighbor pixels, that is,

$$P(U(i,j) = u(i,j) | \{U(k,l),(k,l) \in \Gamma\})$$

$$= P(U(i,j) = u(i,j) | \{U(k,l),(k,l) \in O(i,j)\})$$

where $O(i,j)$ denotes the neighbor of (i,j). The term "neighbor" above deserves to be defined rigorously even though it can be easily grasped from geometry. Neighbor $O(i,j)$ of (i,j) is a collection of pixels satisfying:

(i) $(i,j) \in O(i,j)$;

(ii) If $(k.l) \in O(i,j)$, then $(k,l)) \in O(k,l)$);

(iii) If $(k,l) \in O(i,j)$ and if the distance between pixel (m,n) and (i,j) is less than or equal to that between (k,l) and (i,j), then $(m,n) \in O(i,j)$.

For example, let $O_1(i,j) = \{(k,l), |k-i| + |l-j| \leq 1\}$. Then $O_1(i,j)$ satisfies the above (i) to (iii) and thus it defines a neighbor of (i,j) (In literature, it is referred to as order one neighbor). Another example of neighbor of (i,j) is $O_2(i,j) = \{(k,l), |k-i| + |l-j| \leq \sqrt{2}\}$, which is referred to as order two neighbor of (i,j) in literature.

Finally, we discuss the pros and cons of the MAP method. Theoretically, the MAP approach yields the most likely image for a given observation data. This attribute is also indicated by many experimental studies. However, the MAP approach requires prior knowledge about the image, which is not always available in reality. Another disadvantage dominated in all MAP approaches is their high computational costs. Recently, Sauer and Bouman [SAUE91] speeded up the convergence for some cases by using a local update scheme similar to Gauss-Seidel iterations. But this improvement did not generally occur and thus it remains to be further investigated.

2.9 Concluding Remarks

From the discussion in this chapter, we conclude that if the prior distribution of a PET image is known then the MAP method gives the best possible reconstruction; otherwise, the ML method gives the best possible reconstruction for the PET image. But these two methods are usually expensive.

Both the ML method and the MAP method are based on statistical models and intended to estimate a image by maximizing likelihood functions. The ML method emphasizes the global maximum of the likelihood function, the local features of image might be ignored. However, a MAP method can take the local features into account by incorporating them into a prior distribution of MRF. Because of this, MAP methods are thought to provide better reconstruction than ML methods [SAUE91].

Generally, statistical methods enjoy accuracy advantage over Fourier methods but suffer a high computational cost. Failure to take noise explicitly into consideration is, according to the view expounded here, the basic reason the noise sensitivity of Fourier method. ML and MAP methods improve the reconstruction by modeling the emission process statistically and by establishing algorithms based on the statistical models. The goal of this thesis is to achieve reasonably good reconstruction by improving inexpensive CBP and DFR methods. The next chapter will apply the statistical models to analyze the noise occurring in PET data and subsequent chapters will attempt to exploit the analysis to design effective filters.